



HAL
open science

Étude in silico et caractérisation fonctionnelle des complexes de TWIST1

Charlotte Bouard

► **To cite this version:**

Charlotte Bouard. Étude in silico et caractérisation fonctionnelle des complexes de TWIST1. Cancer. Université de Lyon, 2016. Français. NNT : 2016LYSE1116 . tel-01535047

HAL Id: tel-01535047

<https://theses.hal.science/tel-01535047v1>

Submitted on 8 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N°d'ordre NNT : 2016LYSE1116

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

Opérée au sein de
L'Université Claude Bernard Lyon 1

Ecole Doctorale N° accréditation
Biologie Moléculaire Intégrative et Cellulaire

Spécialité de doctorat : Biochimie
Discipline : Cancérologie

Soutenue publiquement le 13 juillet 2016, par :
Charlotte BOUARD

Etude *in silico* et caractérisation fonctionnelle des complexes de TWIST1

Devant le jury composé de :

Monsieur le Professeur Alain PUISIEUX, président

Monsieur le Docteur Pascal AUFFINGER, rapporteur
Madame le Docteur Hélène JAMET, rapporteur

Monsieur le Professeur Raphaël TERREUX, examinateur
Monsieur le Professeur Ahcène BOUMENDJEL, examinateur

Madame le Professeur Léa PAYEN-GAY, directeur de thèse

"Comme beaucoup de domaines des connaissances humaines, la cancérologie ne constitue pas une "vraie" science au sens épistémologique du terme. C'est plutôt un extraordinaire champ de convergences et d'applications de concepts, de connaissances et de technologies diverses issues de la génétique, de la biologie cellulaire, de la biochimie, de l'immunologie et de la physique. Dans un monde de plus en plus complexe, l'homme isolé, tout imbu soit-il de son maigre savoir, est perdu et impuissant. Au contraire, la cancérologie me semble constituer aujourd'hui l'un des modèles humains les plus achevés de gestion collective de la complexité. Elle mériterait à cet égard un intérêt de plus en plus grand non seulement de la part de nos décideurs et de nos politiques, mais aussi chez ceux des scientifiques dont le thème d'étude est précisément la maîtrise cohérente d'un savoir qui s'étend exponentiellement."

Thomas Tursz, La Nouvelle Médecine du cancer : Histoire et espoir (Cerami et al., 2012; Tursz, Septembre 2013)

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	M. le Professeur Frédéric FLEURY
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directeur Général des Services	M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. ETIENNE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. le Professeur Y. MATILLON
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y. VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E. PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurance	Directeur : M. N. LEBOISNE

Résumé

La réactivation aberrante du gène *TWIST1* embryonnaire a été identifiée dans le laboratoire comme un mécanisme récurrent d'inactivation des voies dépendantes de P53 et RB dans de nombreuses tumeurs. En diminuant la sénescence et l'induction de l'apoptose, TWIST1 coopère avec des protéines oncogéniques dans la transformation cellulaire *in vitro* et intervient dans l'initiation et la progression tumorale *in vivo*. Comme TWIST1 est très faiblement exprimée dans la plupart des cellules adultes différenciées, elle constitue une cible attrayante pour des thérapeutiques futures. Récemment, l'hétérodimère TWIST1/E2A (E12 ou E47, deux produits d'épissage alternatif du gène TCF3) joue un rôle pro-métastatique dans le cancer de la prostate (Gajula et al., 2015), alors que ce complexe est la forme oncogénique de TWIST1 dans des cellules épithéliales mammaires humaines (Jacqueroud et al., 2016). L'hétérodimérisation par l'intermédiaire des domaines HLH est une condition *sine qua non* pour la formation de domaine de liaison à l'ADN (Massari & Murre, 2000). Une approche *in silico* basée sur la modélisation par homologie des complexes de TWIST1 met en évidence le rôle déterminant des boucles inter-hélicales dans le maintien de complexes de TWIST1 à l'ADN en étudiant plusieurs insertions de TWIST1 observées chez les patients atteints du syndrome Saethre-Chotzen (Bouard et al., 2014). Ensuite des approches *in silico* et *in vitro* nous ont permis de comprendre les mécanismes moléculaires sous-jacents à la reconnaissance des séquences E-box des promoteurs de gènes cibles par l'hétérodimère TE et leur stabilisation sur l'ADN (Bouard et al., 2016). Nous avons décrit trois états différents du complexe TWIST1/E12 lié à l'ADN, à des séquences E-box fonctionnelles et modifiées, en fonction de l'affinité de reconnaissance des E-box par TWIST1/E12 (Bouard et al., 2016). Et enfin, cette approche *in silico* nous a permis de montrer l'impact de la phosphorylation de TWIST1 sur la dimérisation et sur la liaison à l'ADN des complexes de TWIST1. Cette dernière étude met en exergue l'importance des régulations post-traductionnelles dans l'étude de l'activité de la protéine TWIST1 et dans la recherche d'inhibiteurs.

Mots clefs :

TWIST1/E12, facteur de transcription bHLH, liaison à l'ADN, E-box, cancer, stratégie *in silico*.

Abstract

Aberrant reactivation embryonic TWIST1 gene has been identified in the laboratory as a recurrent inactivation mechanism dependent pathways of P53 and RB in many tumors. Decreasing senescence and the induction of apoptosis, TWIST1 cooperates with oncogenic proteins in cell transformation *in vitro* and is involved in tumor initiation and progression *in vivo*. As TWIST1 is very weakly expressed in most adult differentiated cells, it is an attractive target for future therapies. Recently, the heterodimer TWIST1 / E2A (E12 or E47, two alternative splice product of the TCF3 gene) plays a pro-metastatic role in prostate cancer (Gajula et al., 2015), while the complex is the oncogenic form of TWIST1 in human mammary epithelial cells (Jacqueroud et al., 2016). The heterodimerization through HLH areas is a prerequisite for DNA binding domain of training (Massari & Murre, 2000). An *in silico* approach that based homology modeling of complex TWIST1 highlights the key role of inter-hélicales loops in the DNA TWIST1 complex retention student TWIST1 several insertions observed in patients Saethre-Chotzen syndrome (Bouard et al., 2014). Then approaches *in silico* and *in vitro* have enabled us to understand the molecular mechanisms underlying the recognition of E-box sequences of the promoters of target genes by TE heterodimer and stabilize DNA (Bouard et al., 2016). We have described three different states of the complex TWIST1 / E12 bound to DNA to functional E-box sequences and modified according to the affinity recognition by E-box TWIST1 / E12 (Bouard et al., 2016). Finally, this approach *in silico* allowed us to show the impact of TWIST1 phosphorylation on dimerization and binding to DNA TWIST1 complexes. This latest study highlights the importance of post-translational regulation in the study of the activity of the protein and TWIST1 in the search for inhibitors.

Key words:

TWIST1/E12, bHLH transcription factor, DNA binding, E-box, cancer, *in silico* strategy.

Title :

In silico study and functional characterization of TWIST1 complexes.

Remerciements

Je tiens tout d'abord à remercier l'ensemble des membres de mon jury. Je remercie particulièrement mes rapporteurs Madame le Docteur Hélène JAMET et Monsieur le Professeur Pascal AUFFINGER qui ont accepté de prendre le temps de relire mon manuscrit et qui prendront le temps de discuter de mon sujet au cours de ma soutenance de thèse. Je remercie également mes examinateurs Messieurs les Professeurs Raphaël TERREUX et Ahcène BOUMENDJEL ainsi que Monsieur le Professeur Alain PUISIEUX d'avoir accepté d'être le président de mon jury de thèse. Je remercie également Madame Léa PAYEN-GAY d'être membre de mon jury de thèse en tant que directeur de thèse. Je souhaite également profiter de ces remerciements pour remercier les membres de mon comité de suivi de thèse Monsieur le Professeur Ahcène BOUMENDJEL et Monsieur le Docteur Pierre FALSON. Je suis ravie d'avoir partagé mes travaux avec vous tous malgré vos plannings respectifs.

Au-delà de sa contribution au sein de mon jury de thèse, je remercie Monsieur le Professeur Alain PUISIEUX qui m'a accueilli au sein de son laboratoire à la fin de mon master. Toutes ces années dans ce laboratoire m'ont permis de mener à bien mes travaux de recherche et ont été animées par de nombreuses interactions avec d'autres scientifiques.

Léa, je voudrais te remercier pour ses années très enrichissantes d'un point de vue scientifique et humain. Merci d'avoir été une directrice de thèse en or qui a su jongler avec ces nombreuses occupations pour toujours être disponible. Merci de m'avoir fait confiance et de m'avoir donné autant de liberté tout en étant très impliquée dans mes travaux. Comme dirait Annelou Dupuis : « Tu me consacres du temps, c'est le plus généreux de tous les cadeaux. ».

Un grand merci à toute l'équipe « Happy Team », même si le nom est vintage je le trouve très sympa : Anne-Pierre, Mamibou, Christelle Julie, Geoffrey et la « Melanome team » qui grandit de jour en jour. Merci aussi aux anciens de l'équipe Stéphane A. et Guigui à qui je souhaite tout le meilleur. Un merci particulier à Caroline sans qui je n'aurais peut-être jamais croisé Léa et ça aurait quand même été ballot. Un merci tout spécial à Madame Agnès pour ces différents échanges plus ou moins scientifiques et sportifs mais toujours constructifs. Merci aussi aux gameleuses du midi qui sont toujours pleines de vie et d'énergie et rendait ainsi le quotidien plus léger. Un Merci à fond les ballons à mes copines de thèse : Ebticem, Freddie, Roxane, Maglou, Clothilde (même si tu es loin je pense à toi ☺) et Mélanie: merci pour tous ces moments partagés et ce bonheur quotidien de voir vos bouilles. Merci aussi à Ebti pour ces heures sportives et potins qui redonnaient toujours de l'énergie.

Un petit merci à ma Villières qui a su m'écouter, aussi bien dans mes répétitions de TWIST que dans mes moments de doute, me soutenir et m'encourager pour que je retrouve mon énergie et que je crois en l'avenir. Merci aussi à mes cops de jeunesse qui sont là au quotidien (Mon Chamcham, Grégoire, Janjan, Jess, Kim, Emilie C.).

Merci à ma famille : mes grands-parents, mes parents et à mes frères pour leur soutien et leur relativité à toutes épreuves. Merci à Jaklou et Mamou pour leur écoute et leurs relectures des derniers instants. Merci évidemment à Nicolas, mon futur mari, pour tout !

Merci aussi à tous (même ceux que j'ai oublié) pour toutes ces choses qui ne se décrivent pas et qu'aucun mot ne peut résumer. Donc « Merci la vie » pour cette thèse de rêve et pour ces belles rencontres. Cela fait beaucoup de « merci », mais comme dirait Pablo Neruda « Un seul mot, usé, mais qui brille comme une vieille pièce de monnaie : merci ! ».

Table des matières

CHAPITRE 1 : INTRODUCTION GENERALE	25
1 DE LA PREHISTOIRE A NOS JOURS	27
2 DONNEES EPIDEMIOLOGIQUES EN FRANCE	31
2.1 INCIDENCE DE TOUS LES CANCERS	31
2.2 MORTALITE DE TOUS LES CANCERS	32
2.3 SURVIE	33
2.4 PREVALENCE DE TOUS LES CANCERS	34
3 REGARD ELEMENTAIRE DE LA TUMORIGENESE	35
3.1 ORIGINE DU CANCER ET HETEROGENEITE TUMORALE	35
3.2 PRINCIPALES CARACTERISTIQUES BIOLOGIQUES, FONCTIONNELLES ET MOLECULAIRES DES CELLULES CANCEREUSES	38
3.2.1 <i>Indépendance vis-à-vis des signaux de prolifération</i>	39
3.2.2 <i>Insensibilité aux signaux antiprolifératifs</i>	41
3.2.3 <i>Stimulation de l'angiogenèse</i>	42
3.2.4 <i>Potentiel de réplication illimité</i>	45
3.2.5 <i>Résistance à la mort cellulaire</i>	47
3.2.6 <i>Pouvoir d'invasion et de dissémination métastatique</i>	51
4 APPROCHES THERAPEUTIQUES	55
4.1 DE LA MOLECULE AU MEDICAMENT	55
4.2 L'ACTION DES MEDICAMENTS	56
4.3 APPORT DE L' <i>IN SILICO</i> DANS LA DECOUVERTE DES MEDICAMENTS	57
4.4 PERSPECTIVES THERAPEUTIQUES	59
CHAPITRE 2 : MODELISATION DE STRUCTURES TRIDIMENSIONNELLES DE PROTEINES	61
<hr/>	
1 GENERALITE SUR LA MODELISATION MOLECULAIRE	63
2 DIFFERENTES ENERGIES EN MECANIQUE MOLECULAIRE	65
3 NOTION DE CHAMP DE FORCE	67
4 METHODES DE MODELISATION MOLECULAIRE	69
4.1 METHODE D'ENFILAGE DES REPLIEMENTS	69
4.2 ALPHABETS STRUCTURAUX	69
4.3 METHODE AB INITIO	70

4.4	MODELISATION PAR HOMOLOGIE	70
4.4.1	<i>Choix du support</i>	70
4.4.2	<i>Alignement entre la séquence et les empreintes</i>	71
4.4.3	<i>Construction du modèle</i>	71
5	DYNAMIQUE MOLECULAIRE	73
5.1	DYNAMIQUES MOLECULAIRES D'UN POINT DE VUE THEORIQUE	73
5.1.1	<i>Equations de mouvements</i>	73
5.1.2	<i>Intégrations des équations</i>	73
5.2	DYNAMIQUES MOLECULAIRES D'UN POINT DE VUE PRATIQUE : DEROULEMENT D'UNE SIMULATION	73
6	OPTIMISATION DES STRUCTURES TRIDIMENSIONNELLES	75
CHAPITRE 3 : TWIST1 DANS LES CANCERS		77
1	TWIST1, FACTEUR DE TRANSCRIPTION DE LA FAMILLE DES BHLH	79
1.1	LA GRANDE FAMILLE DES BHLH	79
1.2	TWIST, FACTEUR DE TRANSCRIPTION BHLH	82
1.2.1	<i>Du gène à la protéine</i>	82
1.2.2	<i>TWIST1 au cours de l'évolution</i>	82
1.2.3	<i>Structure des protéines TWIST1 et TWIST2</i>	83
1.2.4	<i>Partenaires de dimérisation de TWIST1 via son domaine HLH</i>	86
1.2.5	<i>Les E-box reconnues par TWIST1</i>	88
1.2.6	<i>Autres domaines de liaison de TWIST1</i>	90
1.3	PATHOLOGIES ASSOCIEES AUX MUTATIONS DES PROTEINES TWIST	92
1.3.1	<i>Pathologie associée aux mutations de TWIST1 : Le syndrome de Saethre- Chotzen</i>	93
1.3.2	<i>Pathologie associée aux mutations de TWIST2 : Le syndrome de Setleis</i>	98
1.4	INCIDENCE DES MUTATIONS DE TWIST1 ASSOCIEES AU SYNDROME DE SAETHRE-CHOTZEN SUR LE DEVELOPPEMENT DE CANCERS.	100
1.4.1	<i>Les mutations de TWIST1 et de TWIST2 retrouvées dans les cancers.</i>	101
2	TWIST D'UN POINT DE VUE FONCTIONNEL	107
2.1	FONCTION EMBRYONNAIRE DE TWIST DE LA DROSOPHILE AUX VERTEBRES	107
2.1.1	<i>TWIST chez la drosophile</i>	107
2.1.2	<i>TWIST chez les vertébrés</i>	107
2.2	FONCTION DE TWIST A L'AGE ADULTE	109

3	IMPACT DES MODIFICATIONS POST-TRADUCTIONNELLES SUR LES FONCTIONS DE TWIST1	111
3.1	PHOSPHORYLATION	111
3.1.1	<i>Phosphorylation des Sérines 18 et 20 de TWIST1</i>	112
3.1.2	<i>Phosphorylation de la Sérine 42 de TWIST1</i>	113
3.1.3	<i>Phosphorylation de la Sérine 68 de TWIST1</i>	114
3.1.4	<i>Phosphorylation des résidus de Thréonine 121 et Sérine 123 de TWIST1</i>	115
3.2	UBIQUITINATION	117
4	DES FONCTIONS EMBRYONNAIRES AUX FONCTIONS ONCOGENIQUES DE TWIST1	119
4.1	FORTE EXPRESSION DE TWIST1 DANS DE NOMBREUX CANCERS	119
4.2	LES FONCTIONS ONCOGENIQUES DE TWIST1	119
4.2.1	<i>TWIST1/E12 est une forme oncogénique de TWIST1</i>	120
4.2.2	<i>TWIST1 et mécanismes de sauvegarde</i>	120
4.2.3	<i>TWIST1 est impliqué dans la TEM</i>	122
4.3	TWIST1 : UNE CIBLE THERAPEUTIQUE	123
	CHAPITRE 4 : OBJECTIFS DE LA THESE	125
	CHAPITRE 5 : TRAVAUX REALISES	129
1	TRAVAIL DE SYNTHESE : PLASTICITE DES CELLULES CANCEREUSES ET DISSEMINATION METASTATIQUE	133
2	TRAVAIL 1 : ELABORATION DE MODELES <i>IN SILICO</i> DE TWIST1/TWIST1 ET DE TWIST1/E12	147
3	TRAVAIL 2 : IDENTIFICATION DU COMPLEXE ONCOGENIQUE DE TWIST1 DANS UN CONTEXTE CANCEREUX MAMMAIRE HUMAIN	167
4	TRAVAIL 3 : ETUDE DE L'INTERACTION DE TWIST1/E12 AVEC DIFFERENTES SEQUENCES E-BOX	187
5	TRAVAIL 4 : IMPACT DE LA MUTATION R154P SUR LA DIMERISATION DU COMPLEXE TWIST1/E12	213
	CHAPITRE 6 : PARTICIPATION COLLABORATIVE AUX TRAVAUX DE L'EQUIPE	231
1	TRAVAIL COLLABORATIF 1 : ETUDE DE L'ACTIVITE INHIBITRICE DU MBLI-141 SUR LES FONCTIONS DE TRANSPORT D'ABCG2 (ETUDES PRECLINIQUES)	235

2	TRAVAIL COLLABORATIF 2 : ETUDE DE LA TOLERANCE ET DE LA TOXICO- CINETIQUE D'UN PRODUIT A USAGE INDUSTRIEL (TMTZ)	253
	CHAPITRE 7 : RESUME DE MES TRAVAUX DE RECHERCHE	265
1	REVUE BIBLIOGRAPHIQUE : « PLASTICITE DES CELLULES CANCEREUSES ET DISSEMINATION METASTATIQUE »	268
2	ETUDE <i>IN SILICO</i> DES DIMÈRES TWIST1/TWIST1 ET TWIST1/E12	268
2.1	CONSTRUCTION DES MODELES IN SILICO DES HOMO ET HETERODIMERES DE TWIST1	269
2.2	IDENTIFICATION DES ELEMENTS CLEFS DE TWIST1/E12 QUI INFLUENCENT LA RECONNAISSANCE DES MOTIFS ADN CIBLES DES PROMOTEURS DE GENES	270
2.3	RECHERCHE D'INHIBITEURS SPECIFIQUES DE TWIST1/E12	271
3	IMPLICATIONS DANS DES TRAVAUX DE PHARMACO-TOXICOLOGIE	271
3.1	ETUDES PRECLINIQUES NON REGLEMENTAIRES: TOLERANCE ET EFFICACITE D'INHIBITEURS DE NOUVELLES CIBLES THERAPEUTIQUES (<i>IN VIVO</i>)	271
3.2	MESURE DE LA TOXICITE DE MOLECULES D'INTERET INDUSTRIEL.	272
	CHAPITRE 8 : PERSPECTIVES	275
	REFERENCES BIBLIOGRAPHIQUES	283

Liste des abréviations

ABCG2 : ATP-binding cassette sub-family G member 2
ADN : Acide DésoxyriboNucléique
ADC : Advanced Chemistry Development
ADME : Absorption, Distribution, Métabolisme, Excrétion
ADMET : Administration, distribution, métabolisme, élimination, toxicité
AKT : RAC-alpha serine/threonine-protein kinase (PKB)
AHR : récepteur aux hydrocarbures aromatiques
ALT : Alternative Lengthening of Telomeres
AMBER : Assisted Model Building and Energy Refinement
AMM : Autorisation de Mise sur le Marché
ARF : Alternative Reading Frame
ARN : Acide RiboNucléique
ARNm : Acide RiboNucléique messenger
ATM : Ataxia Telangiectasia Mutated
ATP : Adénosine TriPhosphate
Å : angström
BAX : BCL-2-Associated X Proteins
BAK : BCL-2 homologous antagonist/killer
BCL-2 : B-cell Chronic Lymphocytic leukemia/lymphoma-2
BETA2 : Beta-cell E-box transcriptional activator 2 (NeuroD1)
bHLH : basic Helix-Loop-Helix
BMI1 : B Lymphoma Mo-MLV insertion region 1 homolog (mouse)
BPL : Bonne Pratiques de Laboratoire françaises et européennes
B-RAF : Rapidely Accelerated Fibrosarcome homologue B
CDK2 : Caséine Kinase 2
CHARMM : Chemistry at HARvard Molecular Maachanics
CSC : Cellules Souches Cancéreuses
CXCL12 : C-X-C motif chemokine 12
CXCL14 : C-X-C motif chemokine 14
C-terminale : Carboxy-terminale
E-box : Enhancer-box
E-Cadhérine : Epithelial-Cadherin
ELF3 : E74 like ETS transcription factor 3
ELF5 : E74 like ETS transcription factor 5

EMT : Epithelio-Mesenchymal Transition
EMT-TFs : Epithelio-Mesenchymal Transition Transcription Factors
ERBB2 : erb-b2 receptor tyrosine kinase 2
ERK1 et 2 : Extracellular signal Regulated Kinase 1 et 2
EZH2 : Enhancer of Zeste Homolog 2
FGFR2 : Fibroblast Growth Factor Receptor 2
GCN5/pCAF : GCN5 related N-acetyltransferases
GRHL2 : GRainyHead Like transcription factor 2
HAND1 : Heart And Neural crest Derivatives expressed 1
HGF : Hepatocyte Growth Factor
HIF-1 α : Hypoxia-inducible factor 1-alpha
HAT : Histones Acetyl-Transférases
HDAC2 : Histone deacetylase 2
HLH : Helix-Loop-Helix
HOX5 : Homeobox-leucine zipper protein HOX5
Id : DNA-binding protein inhibitor ID-1
IL-6 : interleukine 6
INCa : Institut National du Cancer
IRM : imagerie par résonance magnétique
JNK : c-Jun N-terminal Kinase
kcal : kilocalorie
kDa : kilodalton
lymphocytes T « helpers 1 » LTH1
MAD : Mitotic Arrest Deficient protein
MAX : Myc-Associated factor X
MAPK : Mitogen Activated Protein Kinase
MEC : matrice extracellulaire
MEF2 : Myocyte-specific Enhancer Factor 2
Mi2 β : Mi-2 autoantigen 218 kDa protein
NuRD : NUCleosome Remodeling and Deacetylase
miR-200 : micro RNA 200
miR-3 : micro RNA 3
MMFF94 : Merck Molecular Force Field
MM2, MM3 : Allinger Molecular Mechanics
MTA2 : MeTastasis Associated 1 family member 2
mTOR : Mechanistic target Of Rapamycin
MDM2 : Double Minute 2 protein

MYOD1 : MYOblast Determination protein 1
MYC : Myc proto-oncogene protein (c-MYC)
MTT : 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) tetrazolium
N-Cadhérine : Neural-Cadherin
N-terminale : Amino-terminale
NCBI : National Center for Biotechnology Information
NF- κ B : Nuclear Factor kappa B
NEUROD1 : Neurogenic differentiation factor 1
NLS : Nuclear Localization Signal
NOXA : phorbol-12-myristate-13-acetate-induced protein 1
pCAF : p300/CBP-Associated Factor
PET scan : Tomographie par Émission de Positrons ou TEP
PDB : Protein Data Bank
PGC-1 α : Peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PPARGC1A)
PI3K : Phosphatidyl Inositol 3-Kinase
PIP2 : Phosphatidyl Inositol 4,5 Phosphate
PIP3 : Phosphatidyl Inositol (3,4,5) Phosphate
PKA : Protéine Kinase A
Ppa : Partner of paired
PPAR γ : Peroxisome Proliferator-Activated Receptor gamma
PRRX1 : Paired mesoderm homeobox protein 1
PTEN : Phosphatase and TENsin homolog
PUMA : P53 Upregulated Modulator of Apoptosis
P21: Cyclin-dependent kinase inhibitor 1 (CDK-1)
p53 ou TP53: Phosphoprotein 53
p62 : sequestosome 1(SQSTM1)
QSAR : Quantitative Structure-Activity Relationship
RAS :RAt Sarcoma
RB: RetinoBlastoma
RbAp46 : Retinoblastoma-binding protein p46
RING-finger SCF : Skp1-Cullin1-F-box protein
RMN : Résonance Magnétique Nucléaire
RMSD : Root Mean Square Deviation
RNA : Ribonucleotid acid
RUNX2 : Runt related transcription factor 2
RX : Rayons X
siRNA : small interfering RNA

shRNA : small hairpin RNA or short hairpin RNA
SNAI1 : snail family transcriptional repressor 1
SNAI2 : snail family transcriptional repressor 2
SNAI3 : snail family transcriptional repressor 3
SOV : Structural Overlap
SREBP-1 : Sterol regulatory element-binding protein 1
SCS : syndrome de Saethre-Chotzen
STAT4 : Signal transducer and activator of transcription 4
TE : TWIST1/E12
TEM : Transition Epithélio-Mésenchymateuse
TME : Transition Mésenchymo-Epithéliale
TGF-béata de type 2 :
TMTZ : (E)-1,1,4,4-tétraméthyl-2-tétrazène
TSP-1 : Thrombospondin-1
TT : TWIST1/TWIST1
UTR : UnTranslated Region
VEGF : Vascular endothelial growth factor
VEGF-A : Vascular endothelial growth factor A
ZEB1/2 : Zinc finger E-box-binding homeobox 1/2
3D : tridimensionnel

Table des figures

FIGURE 1: CHIFFRES CLEFS DE L'INCIDENCE DU CANCER EN 2012. _____	31
FIGURE 2 : EVOLUTION DE L'INCIDENCE ET DE LA MORTALITE EN FRANCE DE 1980 A 2012 EN FONCTION DU SEXE. _____	32
FIGURE 3 : CHIFFRES CLEFS DE LA MORTALITE PAR CANCER EN 2012. _____	32
FIGURE 4 : SURVIE NETTE A 10 ANS SELON LA LOCALISATION CANCEREUSE ET LE SEXE. _____	33
FIGURE 5 : SURVIE NETTE A 10 ANS SELON LE PRONOSTIC ET LE SEXE. _____	33
FIGURE 6 : NOMENCLATURE GENERALE DES TUMEURS EN FONCTION DE L'ORIGINE CELLULAIRE. _____	35
FIGURE 7 : REPRESENTATION DE L'HETEROGENEITE INTRA-TUMORALE. _____	36
FIGURE 8 : LES MECANISMES DE TRANSFORMATION CONTRIBUANT A L'HETEROGENEITE TUMORALE. _____	37
FIGURE 9 : CARACTERISTIQUES DES CELLULES CANCEREUSES. _____	38
FIGURE 10 : STRATEGIES ALTERNATIVES DE LA CELLULE CANCEREUSE POUR LE MAINTIEN DE LA PROLIFERATION. _____	40
FIGURE 11 : MUTATIONS DE LA VOIE PI3K/AKT/MTOR ASSOCIEES A LA TUMORIGENESE. _____	41
FIGURE 12 : RB ET P53, FACTEURS SUPPRESSEURS DE TUMEUR. _____	42
FIGURE 13: FORMATION DE VAISSEAUX SANGUINS ET LA COMMUTATION ANGIOGENIQUE. _____	44
FIGURE 14 : STRUCTURE ET FONCTION DES TELOMERES. _____	46
FIGURE 15 : SENESCENCE REPLICATIVE ET CATASTROPHE GENETIQUE, DES BARRIERES PROLIFERATIVES. _____	47
FIGURE 16 : DEUX VOIES PRINCIPALES DE L'APOPTOSE. _____	48
FIGURE 17: AUTOPHAGIE DANS LE CONTEXTE TUMORAL. _____	50
FIGURE 18: CARACTERISTIQUES CELLULAIRES ASSOCIEES A LA NECROSE. _____	50

FIGURE 19 : ETAPES CLEFS DE LA TUMORIGENESE. _____	52
FIGURE 20: BALANCE EMT/MET. _____	53
FIGURE 21 : DIFFERENTS MODES DE MIGRATION MIS EN ŒUVRE PAR LES CELLULES TUMORALES AU COURS DE L'INVASION. _____	54
FIGURE 22 : DEVELOPPEMENT D'UN MEDICAMENT. _____	56
FIGURE 23 : ACTION DES MEDICAMENTS EN TROIS PHASES. _____	57
FIGURE 24: EXEMPLE DE THERAPIES CIBLEES DES CARACTERISTIQUES CLEFS DES CELLULES CANCEREUSES. _____	60
FIGURE 25: ILLUSTRATIONS DES POTENTIELS IMPLIQUES DANS L'ENERGIE POTENTIELLE. _____	66
FIGURE 26 : RELATION ENTRE LE TAUX D'IDENTITE DE SEQUENCE ET LES METHODES DE MECANIQUE MOLECULAIRE UTILISEES. _____	69
FIGURE 27 : REPRESENTATION DES MINIMA ENERGETIQUES LOCAUX ET GLOBAUX PAR UNE APPROCHE DE MINIMISATION CLASSIQUE. _____	75
FIGURE 28 : CLASSIFICATION STRUCTURELLE DES PROTEINES BHLH. _____	80
FIGURE 29 : DEUX DOMAINES CLEFS DES FACTEURS DE TRANSCRIPTION BHLH. _	81
FIGURE 30 : ALIGNEMENT DE PROTEINES TWIST1 ISSUES DE DIFFERENTES ESPECES AVEC CLUSTAL OMEGA (SIEVERS & HIGGINS, 2014). _____	83
FIGURE 31: STRUCTURE DES PROTEINES TWIST1 ET TWIST2. _____	85
FIGURE 32: BALANCE ENTRE HOMODIMERE ET HETERODIMERE DE TWIST1. ____	88
FIGURE 33 : DIMERISATION DE TWIST1 EST UN PREREQUIS A SON ACTIVITE TRANSCRIPTIONNELLE. _____	89
FIGURE 34 : PRINCIPALES E-BOX DECRITES POUR TWIST1. _____	90
FIGURE 35 : DOMAINES D'INTERACTIONS DE TWIST1 AVEC CES PARTENAIRES. __	92
FIGURE 36 : SIGNES CLINIQUES ASSOCIES AU SYNDROME DE SAETHRE-CHOTZEN. 94	
FIGURE 37: MUTATIONS DU GENE CODANT POUR TWIST1 ASSOCIEES AU SCS. ____	97
FIGURE 38 : SIGNES CLINIQUES ASSOCIES AU SYNDROME DE SETLEIS. _____	98
FIGURE 39 : MUTATIONS DE TWIST2 DECRITES DANS LE SYNDROME DE SETLEIS. 100	
FIGURE 40: ANALYSE DES ALTERATIONS SOMATIQUES DE TWIST1 SUR CBIOPORTAL. _____	103

FIGURE 41: ANALYSE DES ALTERATIONS SOMATIQUES DE TWIST2 SUR CBIOPORTAL.	105
FIGURE 42: RECAPITULATIF DES PHENOTYPES OBSERVES SELON LE GENOTYPE ET LES ESPECES.	109
FIGURE 43 : SITES DE PHOSPHORYLATION DE TWIST1 (HORNBECK ET AL., 2015).	112
FIGURE 44 : PHOSPHORYLATION DE TWIST1 PAR CDK2 PERMET LA STABILISATION DE LA PROTEINE TWIST1.	113
FIGURE 45: PHOSPHORYLATION DE LA SERINE 42 PAR PKB FAVORISE LES EFFETS ANTI-APOPTOTIQUES DE TWIST1.	114
FIGURE 46 : PHOSPHORYLATION DE LA SERINE 68 STABILISE LA PROTEINE TWIST1.	115
FIGURE 47: IMPACT DE LA PHOSPHORYLATION DE LA THREONINE 121 ET DE LA SERINE 123 SUR LA BALANCE HOMO-HETERODIMERE.	116
FIGURE 48: TWIST1 INHIBE LES MECANISMES DE SAUVEGARDES CELLULAIRES.	122
FIGURE 49: OBJECTIFS DE LA THESE	127
FIGURE 50: SPECIFICITE D'INTERACTION DE TROIS MOLECULES CHIMIQUES.	279
FIGURE 51 : PERSPECTIVES DE MES TRAVAUX DE THESE.	281

CHAPITRE 1 : INTRODUCTION GENERALE

1 De la préhistoire à nos jours

Les pathologies cancéreuses existent depuis l'origine de la vie. Les premières descriptions connues concernent les squelettes d'un dinosaure retrouvé dans le Wyoming et du pithécantrophe de Java qui portaient des signes d'ostéosarcomes (-700 000 avant Jésus-Christ). Plus tard, des traces d'ostéosarcomes ont également été mises en évidence sur des guerriers néolithiques, l'homme de Mixnitz en Autriche et l'homme de Münsingen en Suisse (-10 000 avant Jésus-Christ).

Plus récemment, la description de cancer, sans que le terme cancer ne soit utilisé, a été retrouvée dans des Papyrus Égyptiens datant de -3500 avant Jésus-Christ. Le papyrus Edwin Smith, traité de pathologie interne et de chirurgie osseuse, décrit huit cas de tumeurs ou d'ulcères de la poitrine enlevés par cautérisation. Ce traité spécifie qu'il s'agit de maladie sans traitement. Un autre papyrus, datant de -1552 avant Jésus-Christ, déconseille d'enlever une grosse tumeur de la cuisse pour éviter les problèmes infectieux qui pourraient être fatals. Il y a cependant très peu de descriptions précises de cancer avant Hippocrate (460-370 avant Jésus-Christ, Grec). Ce dernier est un médecin grec considéré comme le père de la médecine. Il décrit longuement l'extension locale et la généralisation de la maladie en utilisant les termes : "*carcinosis*" (crabe), "*skirros*" (tumeur dure) et "*carcinoma*", qui permettent d'introduire clairement la notion de gradation et d'évolution progressive. Il y eut peu d'évolution pendant les 3 siècles suivants. De -25 avant Jésus-Christ à 50 après Jésus-Christ, le médecin Romain Aulus Cornelius Celsus influencé par les médecines grecques et égyptiennes, traduisit "*carcinosis*" en cancer pour décrire des ulcérations malignes avec pénétration profonde et "*carcinoma*" (néologie du grec) pour décrire des lésions pré-malignes et malignes superficielles. Déjà à cette époque, Aulus Cornelius Celsus pensait que seuls les premiers degrés pouvaient être traités montrant donc que le traitement du cancer était associé à l'avancement de la maladie. Par la suite de 130 à 201 après Jésus-Christ, Claude Galien (129-216, grec) proposa des théories valables pendant un millénaire. Il considérait que les tumeurs "*oncos*" (terme utilisé pour les spécialistes du cancer : les oncologues) étaient dues à un excès d'humeur (de bile noire) qui se solidifiait dans certaines parties du corps. Son traitement consistait en l'administration de purges pour dissoudre la bile « solidifiée » et si la lésion ne régressait pas alors il pratiquait l'excision.

Au Moyen Age, la médecine arabe se distingua notamment au niveau de la lutte contre le cancer. Avicenne (980-1037, Perse) et Albucasis (1013-1106, Espagnol) préconisaient l'excision des cancers débutants et accessibles tout en cautérisant les tissus avoisinants. Ils conseillaient de ne rien faire en cas de lésions avancées. De plus, Albucasis purgeait les patients de leur bile noire en guise de traitement préopératoire. C'est également au Moyen Age que les chirurgiens tels que John of Arderne (1307-1390, Anglais) et Guy de Chauliac (1300-1370, Français) contribuèrent à une meilleure connaissance des cancers. Le premier a décrit les symptômes du carcinome du rectum avec ses

hémorragies et obstructions. Le second a réalisé de larges excisions de lésions débutantes et soignait les lésions graves avec des onguents de manière palliative.

A l'époque de la Renaissance, l'éveil de l'esprit critique qui traversait le monde, permit une évolution de la médecine. D'une part, la levée d'interdiction de la pratique d'autopsie a permis une meilleure connaissance de l'anatomie. Gabriele Falloppio (1523-1562, Italien) décrit plusieurs types de cancers, pratiqua l'excision et la cautérisation des lésions récentes et traita les lésions avancées avec différents onguents dont l'Arsenic. Gaspare Aselli (1581-1626, Italien) découvre le système lymphatique qui joue un rôle clef dans la dissémination métastatique. Jean Pecquet (1622-1674, Français) découvre le collecteur principal des vaisseaux lymphatiques qu'est le canal thoracique. D'autre part l'observation clinique des différentes formes de cancers ainsi que l'enregistrement des cas sont écrits par Wilhelm Fabricius Hildanus (1560-1634, Allemand), Marco Aurelio Severino (1580-1656, Italien), qui fut le premier à illustrer ses descriptions avec des dessins et enfin Johann Schultes (1595-1645, Allemand).

Au XVII^{ème} siècle, Daniel Sennert (1572-1637, Allemand) affirme que le cancer est contagieux. C'est ainsi que les cancéreux sont exclus de nombreux hôpitaux pendant près de 2 siècles. En 1740, le Chanoine et mécène Jean Godinot (1661-1749, Français) fit construire un des premiers hôpitaux exclusivement dédiés aux pathologies tumorales à Reims.

Au cours des XVIII^{ème} et XIX^{ème} siècles, la théorie des humeurs de Claude Galien va largement influencer les médecins. Henri François le Dran (1685-1770, Français) décrit que le cancer commence localement et peut ensuite s'étendre par les canaux lymphatiques vers les ganglions lymphatiques ce qui explique comment un cancer du sein, par exemple, peut atteindre ensuite les poumons. Il préconise donc l'excision de la tumeur mais également des ganglions lymphatiques axillaires. Il conclut que si ces derniers sont envahis alors le cancer était grave. Xavier Bichat (1771-1802, Français) découvre que les différentes localisations de cancer ne sont en réalité qu'une seule et même maladie mais dans différents organes. C'est le début de la notion de métastase. C'est également au cours de ces deux siècles que le traitement du cancer va être associé à la pathologie. En effet Giovanni Battista Morgagni (1682-1771, Italien) fut le créateur de l'Anatomie Pathologique *via* des dissections réalisées dans le but de connaître les lésions ayant provoqué la mort du sujet. Ce dernier liait donc les résultats d'autopsie vue à l'œil nu avec l'évolution clinique de la maladie. Percivall Pott (1714-1788, Anglais), médecin anglais, est le précurseur de l'épidémiologie des cancers. Il constata que plusieurs garçons ramoneurs mourraient vers l'âge de vingt ans de cancers du scrotum mettant en avant un lien entre cancer et une substance chimique contenue dans la suie. Bernard Peyrilhe (1737-1804, Français) contribua à la recherche expérimentale dans le traitement du cancer en prélevant le liquide par une ponction au niveau d'une lésion de cancer du sein et en l'injectant ensuite dans la cavité péritonéale d'un chien. Il confirma la théorie métastatique *via* les canaux lymphatiques mise en évidence par Henri François le

Dran. John Hunter (1728-1793, Britannique) a décrit précisément la conduite à tenir des chirurgiens pour l'opération de cancers et fut le premier à évoquer des prédispositions au cancer comme l'hérédité, l'âge et le climat. Rudolf Virchow (1821-1902, Allemand) est considéré comme l'un des fondateurs de l'anatomie pathologique moderne. Il est également le fondateur de la théorie de la pathologie cellulaire d'après laquelle les maladies trouvent leurs origines dans les altérations des cellules. L'utilisation du microscope moderne lui a permis, sur les traces de Giovanni Battista Morgagni, de corréler l'observation des lésions micro- et macroscopiques aux signes cliniques et à la pathologie. Cette approche permet une meilleure compréhension des dommages et aide le développement de la chirurgie du cancer. Joseph Lister (1827-1912, Britannique) appliqua la théorie de Louis Pasteur en introduisant l'antisepsie et l'asepsie en chirurgie ce qui permit de réduire considérablement le taux de mortalité des patients par infection post-opératoire.

Au cours des XX et XXIème siècle, les connaissances et les traitements du cancer ont progressé énormément grâce à la multidisciplinarité. En effet, les rayons X ont été découverts en 1895 par Wilhelm Röntgen (1845-1923, Allemand), puis la radioactivité en 1898 par Pierre (1859-1906, Français) et Marie Curie (1867-1934, Polono-Française) à partir du phénomène découvert par Henri Becquerel (1852-1908, Français). Ces découvertes bouleversent les thérapeutiques. Puis la chimiothérapie, qui est l'utilisation de médicaments pour tuer les cellules malignes, est mise en place dès 1946 avec l'utilisation de moutardes azotées (agents alkylants) qui diminue la croissance des cellules leucémiques. En 1971, le taxol est utilisé pour inhiber la croissance de cellules leucémiques. En 1977, les prémices de la thérapie hormonale voient le jour puisque des essais sur des modèles animaux approuvent l'utilisation du tamoxifène (modulateur des récepteurs aux œstrogènes) dans le traitement du cancer du sein, ce qui est en accord avec les travaux de George Thomas Beatson qui montrait le lien entre le développement des cancers du sein et les oestrogènes (1848-1933, Britannique). Plus récemment, l'immunothérapie, impliquant l'utilisation d'anticorps monoclonaux ou de cytokines, stimule la réponse immunitaire des patients afin de détruire les cellules cancéreuses. Les marqueurs biologiques et le progrès de l'imagerie médicale (IRM, PET scan) permettent la détection de certains cancers à des stades très précoces et assurent ainsi une meilleure prise en charge. Dès 1928, les mutations génétiques sont proposées comme pouvant être à l'origine des cancers. Après la découverte de l'ADN (Acide DésoxyriboNucléique) avec Watson et Crick en 1953 (Watson & Crick, 1953), Alfred George Knudson (né en 1922, Américain) qui est un généticien spécialisé dans la génétique du cancer, révèle en 1971 que le cancer nécessite au moins deux mutations. Cette approche génétique permet une meilleure connaissance de la biologie cellulaire et moléculaire avec la découverte de l'implication d'enzymes et de protéines clefs dans le cancer. Cette approche met en évidence également la transmission héréditaire de certains cancers, avec aujourd'hui plus de 70 gènes de prédisposition décrits.

Aujourd'hui la prise en charge des patients atteints de cancer est en pleine transformation. Après l'ère de la thérapeutique empirique (de l'antiquité à la fin du XVIIIème siècle) et fonctionnelle (début du XXème siècle), nous entrons dans l'ère de la médecine personnalisée. En effet, le développement des techniques d'analyse en génétique, en protéomique, en imagerie médicale modifie l'approche thérapeutique en favorisant aujourd'hui les thérapies ciblées c'est à dire des thérapies qui s'attaquent à des cibles spécifiques des cellules cancéreuses limitant ainsi leur action sur les cellules saines. L'ère de la médecine personnalisée permet à chaque malade d'avoir un traitement qui tient compte des caractéristiques de chaque tumeur afin de préciser le diagnostic, d'identifier les anomalies en cause et de les traiter tout en considérant l'âge et l'environnement du patient, et en limitant les effets indésirables des traitements.

La rédaction de cette partie a été réalisée à partir de différentes sources bibliographiques ((Canada); Bordet; Tour) (Society; Tursz, Septembre 2013).

2 Données épidémiologiques en France

Ces données sont issues du rapport de l'Institut National du Cancer (INCa) « les cancers en France » édition 2014 (cancer, Février 2015). Il est à noter que les chiffres sont ceux de 2012 et qu'il n'y a actuellement pas de chiffres plus récents. « Tous les cancers » permettent d'appréhender l'incidence et la mortalité de manière globale, toutefois ils regroupent un ensemble très hétérogène qui peut masquer certaines évolutions.

2.1 Incidence de tous les cancers

Le nombre de nouveaux cas de cancer a considérablement augmenté entre 1980 et 2012 avec une augmentation de +107,6% chez l'homme et de +111,4% chez la femme. Cette hausse peut s'expliquer par l'augmentation et le vieillissement de la population, la majorité des cas survenant chez les personnes âgées (65 ans et plus). En 2012, l'INCa estime en France le nombre de nouveaux diagnostics de cancer et montre que la survenue des cancers est plus fréquente chez l'homme. L'âge médian au moment du diagnostic est variable selon le type de cancer. Chez l'homme, les trois cancers les plus fréquents sont ceux de la prostate, du poumon et du colon-rectum alors que chez les femmes, ce sont les cancers du sein, du colon-rectum et du poumon. Ce dernier connaît une progression constante chez les femmes en raison du tabagisme, avec une multiplication par 7 du nombre de nouveaux cas au cours des 30 dernières années (Figure 1).

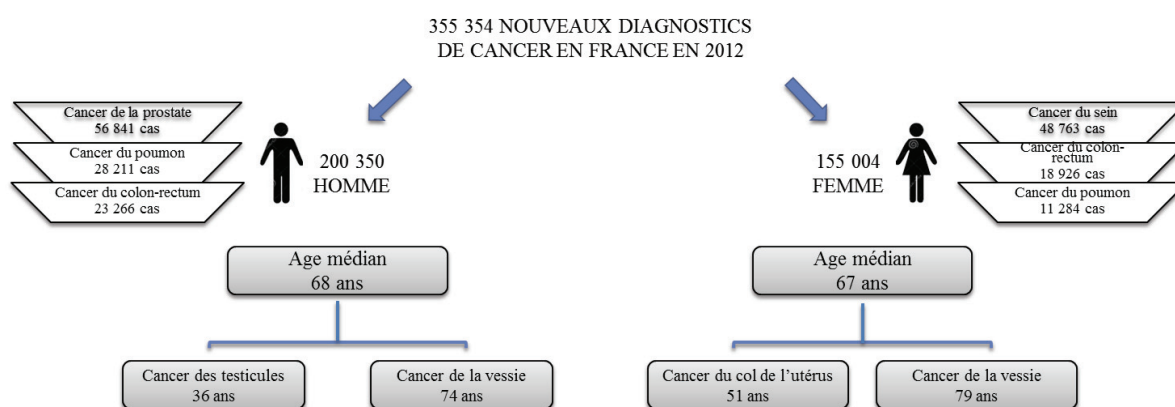
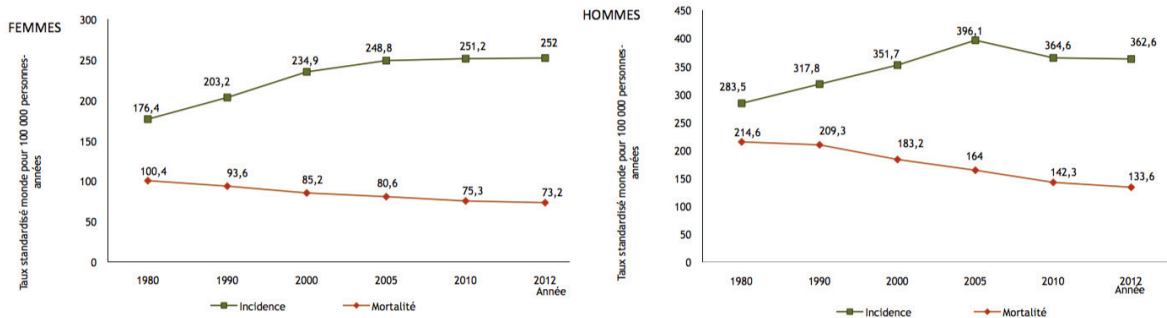


Figure 1: chiffres clés de l'incidence du cancer en 2012.

Le taux d'incidence des cancers rapporte le nombre de cas pour 100 000 personnes par an standardisé en utilisant comme population de référence la population mondiale. En 2012, il est estimé à 362,6 chez l'homme et à 252,0 chez la femme. En 1980, les taux d'incidence étaient respectivement de 283,5 chez les hommes et de 176,4 chez les femmes. En termes de variation du taux d'incidence de 1980-2012 (standardisés sur la population mondiale), une augmentation de +0,8% par an chez l'homme et de +1,1% par an chez les femmes est observée. Cependant, les tendances semblent s'être inversées sur la période 2005-2012 avec une diminution de -1,3% par an chez l'homme et une

stabilisation de l'augmentation chez la femme avec une augmentation +0,2% par an, au lieu de +1,6% entre 1980 et 2005 chez la femme (Figure 2). Ces tendances sont liées à celle du cancer de la prostate chez l'homme et du cancer du sein chez la femme.



Sources : Binder-Foucard F, 2013. Traitement : INCA 2013

Figure 2 : évolution de l'incidence et de la mortalité en France de 1980 à 2012 en fonction du sexe.

Source INCa, "les cancers en France" édition 2014 (cancer, Février 2015)

2.2 Mortalité de tous les cancers

La prise en charge des pathologies tumorales est un enjeu de santé publique majeur, puisque depuis 2004, les cancers constituent la première cause de mortalité en France, avant les maladies cardio-vasculaires. Ils représentent 33% de l'ensemble des décès masculins et 24% de l'ensemble des décès féminins.

Les trois cancers à plus haut taux de mortalité chez l'homme sont le cancer du poumon, le cancer colorectal et le cancer de la prostate alors que chez la femme ce sont le cancer du sein, le cancer du poumon et le cancer colorectal (Figure 3). Néanmoins, le taux de mortalité par cancer du sein chez la femme diminue depuis 15 ans.

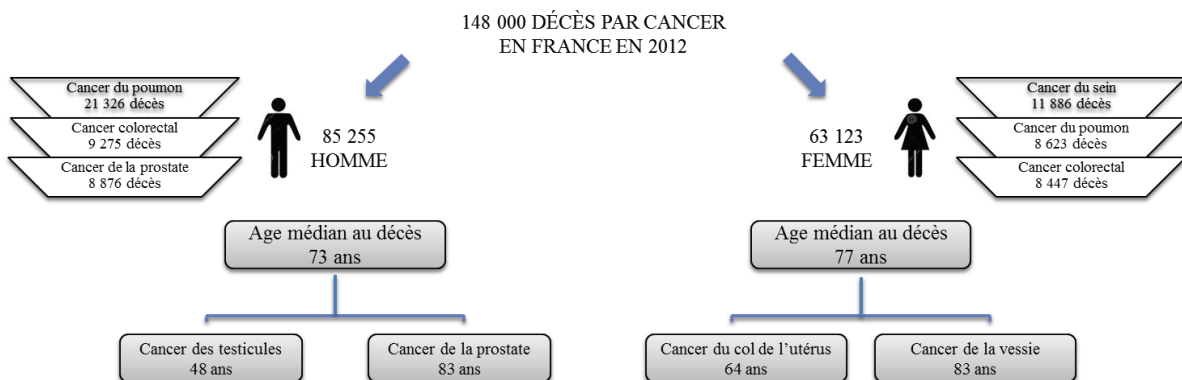


Figure 3 : chiffres clés de la mortalité par cancer en 2012.

En 2012, les taux de mortalité (standardisés sur la population mondiale) des cancers sont estimés à 133,6 pour 100 000 personnes-années chez l'homme et à 73,2 chez la femme. Le nombre de décès par cancer a augmenté de +11,0% chez l'homme et de +20,3% chez la femme entre 1980 et 2012. L'augmentation constatée peut être attribuée à l'évolution démographique (augmentation et vieillissement de la population). Le taux standardisé de mortalité (sur la population mondiale) a diminué en moyenne de -1,5% par an chez les hommes et de -1% chez les femmes au cours de la période 1980-2012. Les taux de mortalité restent cependant toujours plus élevés chez les hommes que chez les femmes, néanmoins ils diminuent plus rapidement chez les hommes (diminution des consommations d'alcool et de tabac ; Figure 2). La baisse du taux de mortalité peut s'expliquer par le développement et l'amélioration des méthodes de dépistages avec des diagnostics plus précoces, mais aussi des approches thérapeutiques avec des traitements plus efficaces. S'ajoute à cela, la diminution globale du risque de développer un cancer.

2.3 Survie

La survie que l'on observe si le cancer est la seule cause de décès possible, des personnes atteintes de cancers (diagnostiqués entre 1989 et 2004) varie selon la localisation cancéreuse. La survie nette à 10 ans varie en effet de 1% à 93% chez l'homme et de 5% à 92% chez la femme (Figure 4).

Localisation cancéreuse	Survie nette à 10 ans chez l'homme	Localisation cancéreuse	Survie nette à 10 ans chez la femme
Mésothéliome de la plèvre	1%	Pancréas	5%
Testicule	93%	Thyroïde	92%

Figure 4 : survie nette à 10 ans selon la localisation cancéreuse et le sexe.

Les cancers de mauvais pronostic et les cancers de bon pronostic peuvent être classés selon la survie nette à 10 ans. Les hommes ont un pourcentage plus élevé de cas de cancers de mauvais pronostics alors que les femmes ont un pourcentage plus élevé de cancers de bon pronostic (Figure 5). Pour la majorité des localisations cancéreuses, la survie nette est meilleure chez la femme que chez l'homme, et elle est plus favorable chez les plus jeunes entre 15 et 44 ans que chez les plus âgés de 75 ans et plus.

		Cancers de mauvais pronostics	Cancers à pronostics intermédiaires	Cancers de bon pronostics
Survie nette à 10 ans		<33%	33 à 65%	>66%
Pourcentage de cas	Homme	40%	32%	28%
	Femme	16%	32%	52%

Figure 5 : survie nette à 10 ans selon le pronostic et le sexe.

A l'exception du cancer du col de l'utérus et du cancer des poumons, la majorité des cancers diagnostiqués entre 1989 et 2004 ont montré une amélioration de la survie des patients. Cette amélioration peut s'expliquer par l'augmentation de la durée d'observation de la maladie, l'amélioration des dépistages (plus précoces et plus précis), les progrès thérapeutiques. D'autres paramètres, comme l'âge et l'état de santé "hors cancer" au moment du diagnostic, sont également des facteurs d'influence. De manière générale, plus un cancer est détecté tôt, meilleures sont les chances de survie.

2.4 Prévalence de tous les cancers

Le nombre de personne de 15 ans et plus en vie en 2008 ayant eu un cancer (toutes localisations confondues) au cours de leur vie est de l'ordre de 3 millions en France, avec 1 570 000 hommes (6,4% de la population masculine de 15 ans et plus) et 1 412 000 femmes (5,1% de la population féminine de 15 ans et plus). Chez l'homme, la moitié des cas de prévalence à 5 ans est attribuable au cancer de la prostate alors que chez la femme, la moitié des cas de prévalence à 5 ans est attribuable au cancer du sein. Ces données sont bien en accord avec l'incidence que nous avons vu précédemment (Figure 1).

3 Regard élémentaire de la tumorigenèse

Le terme « cancer » englobe un grand groupe de maladies très différentes caractérisées par la multiplication et la propagation anarchique de cellules anormales. Les cancers peuvent être classés en fonction de la nature de la cellule ou du tissu dans lequel ils se développent (Figure 6).

Cellules ou tissus d'origine	Exemple de tumeurs malignes
Tissus épithéliaux (Carcinome) 85% des nouveaux cas de cancers	Carcinome épidermoïde (épithélium malpighien) Adénocarcinome (épithélium glandulaire)
Tissus conjonctifs (Sarcome) Moins de 1% des nouveaux cas de cancers	Fibrosarcome (Fibroblastes) Liposarcome (Adipocytes)
Tissus hématopoïétiques	Syndrome myéloprolifératifs (Myéloïdes) Lymphome (Lymphocytes)
Autres tissus	Mélanome (Mélanocytes) Neuroblastome (Tissus embryonnaire nerveux)

Figure 6 : nomenclature générale des tumeurs en fonction de l'origine cellulaire.

3.1 Origine du cancer et hétérogénéité tumorale

Une cellule saine évolue progressivement vers un état néoplasique suite à l'acquisition successive de propriétés caractéristiques. Une des caractéristiques majeures des tumeurs humaines est leur hétérogénéité au niveau biologique, morphologique et de l'expression de marqueurs moléculaires.

- Il y a tout d'abord l'hétérogénéité inter-tumorale qui correspond à l'hétérogénéité entre des tumeurs issues de différents patients. Dans l'étude des altérations géniques impliquées soit dans les cancers du sein soit dans les cancers colorectaux, l'hétérogénéité inter-tumorale a permis d'identifier certains gènes couramment mutés dits « mountains » d'autres moins souvent mutés dits « hills » montrant ainsi la variabilité des mutations génétiques impliquées dans un type de cancer donné (Wood et al., 2007). L'étude génomique de 100 tumeurs du sein montre une variabilité du nombre de mutations somatiques avec 73 combinaisons différentes de gènes mutés possibles (Stephens et al., 2012).
- Il y a également l'hétérogénéité intra-tumorale qui correspond à une hétérogénéité entre les cellules d'une même tumeur. L'hétérogénéité intra-tumorale a été mise en évidence *via* plusieurs approches : des études de caryotypes, de gènes individuels ou de l'ensemble du génome (Hoglund, Gisselsson, Sall, & Mitelman, 2002; Nik-Zainal et al., 2012; Shibata, Schaeffer, Li, Capella, & Perucho, 1993). Au vu de ces hétérogénéités décrites, Marco Gerlinger et son équipe ont cherché à savoir si une seule biopsie était suffisante pour connaître l'ensemble des altérations génétiques d'une tumeur. Il a réalisé des études sur des biopsies de tumeurs rénales issues de quatre patients. Le profil d'altération génétique des différentes régions de la tumeur primaire diffère significativement de celui des métastases.

De plus, il existe différentes populations de cellules malignes au sein d'une tumeur primaire dont une pourrait d'ailleurs être à l'origine de la formation des métastases (Région 4 cf. Figure 7) (Gerlinger et al., 2012). Deux autres types d'hétérogénéité tumorale ont été décrits : l'hétérogénéité inter-métastatique et l'hétérogénéité intra-métastatique (Vogelstein et al., 2013).

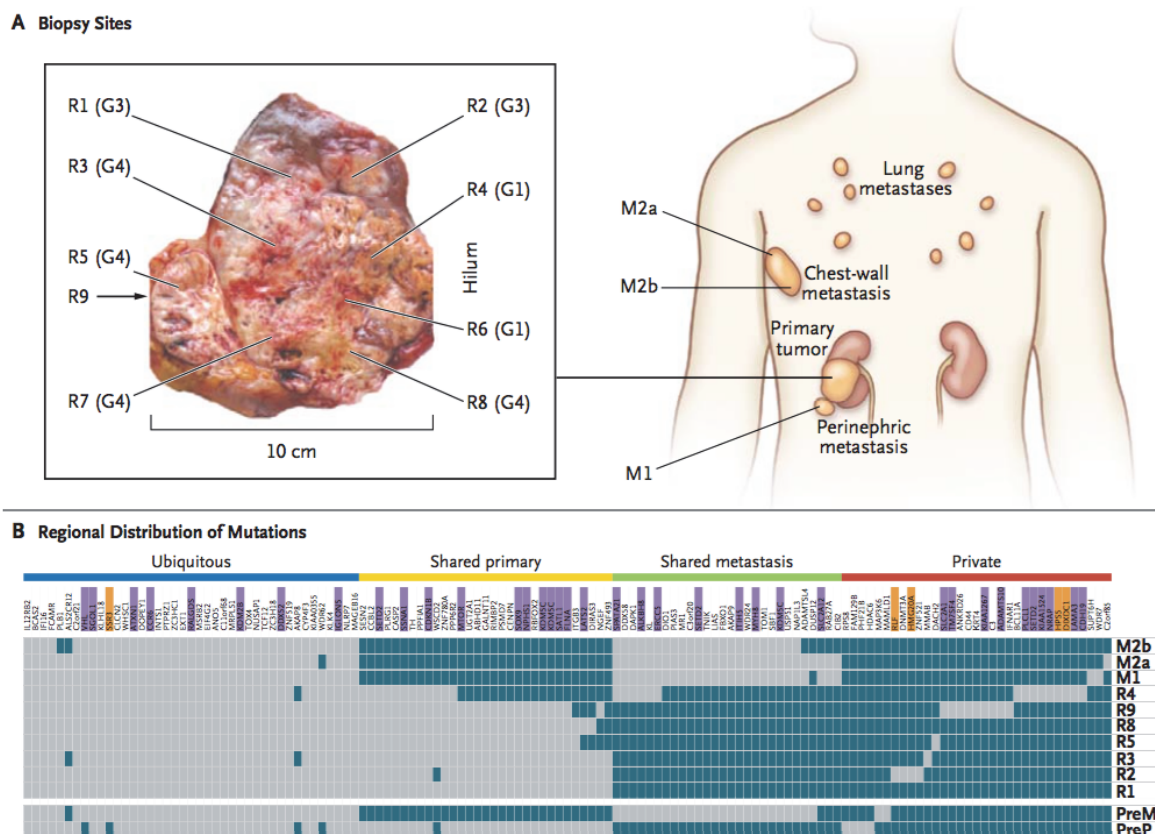


Figure 7 : représentation de l'hétérogénéité intra-tumorale.

(Gerlinger et al., 2012)

A: Localisation des différents sites de biopsies au niveau de la tumeur primaire ou des métastases situées au niveau de sites secondaires réalisées chez un patient (G représente le grade de la tumeur).
B: L'analyse du séquençage à haut débit met en évidence la présence en gris ou l'absence en bleu d'altérations génétiques selon la localisation des sites de biopsies décrits ci-dessus. Les barres de couleur au-dessus de l'analyse de séquençage classent les mutations selon qu'elles soient ubiquitaires, au niveau de la tumeur primaire, au niveau des sites métastatiques, ou uniques à la région décrite.

L'hétérogénéité tumorale a longtemps été expliquée par le modèle stochastique de transformation maligne qui est le modèle Darwinien de l'expansion clonale (Figure 8). La succession de mutations aléatoires conférant un avantage sélectif aboutit à la formation d'une tumeur maligne (Hanahan & Weinberg, 2011; Nowell, 1976). Les mutations conférant un avantage de croissance

concernent les gènes dits « drivers » (140 gènes drivers ont été décrits) et conduisent à l'émergence de nouvelles sous populations cellulaires (Vogelstein et al., 2013). Au cours des nombreux cycles réplicatifs subits par les cellules lors de la progression tumorale, les cellules peuvent acquérir de nombreuses altérations annexes ne conférant pas d'avantage sélectif et dites « passagers » (Stratton, Campbell, & Futreal, 2009). En résumé, une tumeur se compose de différentes sous-populations cellulaires qui se distinguent par le profil d'altérations spécifiques qu'elles ont sélectionnées au cours du processus de tumorigénèse (Bernards & Weinberg, 2002). Dans certains cas, les nouvelles mutations acquises conduisent à la perte des propriétés tumorigéniques et sont donc perdues au cours du développement tumoral.

Depuis quelques années, l'hétérogénéité est associée à un autre modèle de transformation tumorale qu'est le modèle hiérarchique basé sur le concept de cellules souches cancéreuses (CSC) (Figure 8)(Reya, Morrison, Clarke, & Weissman, 2001). Une CSC présente des caractéristiques d'une cellule souche normale, à savoir une capacité d'auto-renouvellement et de multi-potentialité (processus de différenciation). L'essor du concept des CSC est dû à la capacité d'une faible proportion de cellules tumorales présentes au sein de leucémies myéloïdes chroniques de générer des tumeurs chez des souris immunodéficientes (Bonnet & Dick, 1997; Lapidot et al., 1994). Des CSC ont été identifiées puis isolées dans le cancer du sein et dans d'autres types de tumeurs solides (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003; Qiu, Fang, Luo, & Ouyang, 2015). Selon le type de tumeur, les CSC expriment des marqueurs spécifiques (Qiu et al., 2015). Les CSC sont impliquées dans l'initiation, la croissance, les métastases, la résistance aux traitements et la récurrence des cancers humains (Hanahan & Weinberg, 2011) (Qiu et al., 2015).

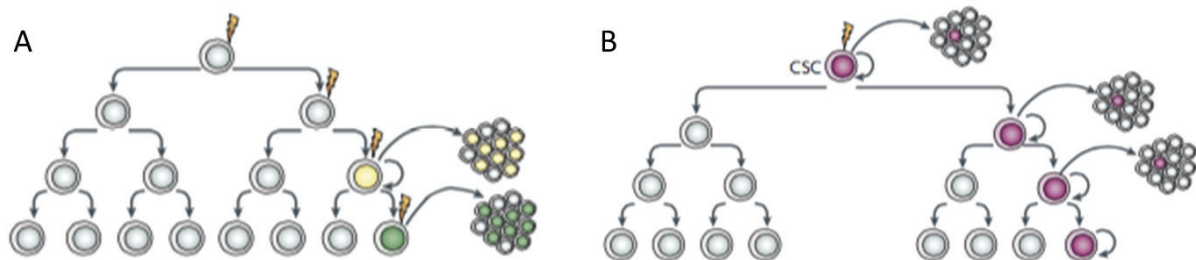


Figure 8 : les mécanismes de transformation contribuant à l'hétérogénéité tumorale.

D'après Visvader and Lindeman, 2008 (Visvader & Lindeman, 2008)

A: Modèle stochastique Darwinien. Ce premier modèle d'hétérogénéité repose sur l'acquisition d'altérations génétiques ou épigénétiques aléatoires conférant un avantage prolifératif et/ou de survie aux cellules tumorales qui conduit donc à l'émergence d'une nouvelle sous-population (cellules jaunes ou vertes). B: Modèle hiérarchique basé sur le concept de cellules souches cancéreuses. Ce second modèle repose sur le concept que seule une proportion restreinte de cellules possédant certaines propriétés de cellules souches serait à l'origine de la tumeur. Ces cellules seraient capables de se multiplier et de donner naissance à une progéniture différenciée contribuant par ce biais à l'hétérogénéité intra-tumorale.

Ces deux théories sont bien différentes, mais elles ne sont pas mutuellement exclusives et il semblerait que certains types de tumeurs suivent préférentiellement l'un ou l'autre modèle (Shackleton, Quintana, Fearon, & Morrison, 2009). Il est important de noter également que le microenvironnement aurait un impact sur l'hétérogénéité intra-tumorale en favorisant les instabilités génomiques (Radisky & Bissell, 2006). Les cellules tumorales auraient, selon leur localisation au sein de la tumeur, un accès différent aux signaux du microenvironnement ce qui influencerait leur phénotype (Bissell & Hines, 2011; Magee, Piskounova, & Morrison, 2012; Polyak, Haviv, & Campbell, 2009).

3.2 Principales caractéristiques biologiques, fonctionnelles et moléculaires des cellules cancéreuses

Les cellules cancéreuses adaptent les grandes fonctions cellulaires pour leur permettre de croître et de résister à différentes conditions environnementales. Malgré la diversité et la complexité des pathologies cancéreuses, Douglas Hanahan et Robert Weinberg ont montré que les cellules tumorales partageaient des propriétés communes qui les différenciaient des cellules normales (Hanahan & Weinberg, 2000, 2011). Dans un premier temps, ils ont décrit six propriétés caractéristiques pour les cellules cancéreuses (Figure 9) (Hanahan & Weinberg, 2000). Dans un second temps, ils ont complété la description de ces propriétés en y associant d'une part l'état inflammatoire des lésions précancéreuses et malignes liées aux cellules immunitaires, et d'autre part l'acquisition d'une instabilité génomique des cellules cancéreuses avec l'apparition de mutations aléatoires et de réarrangements chromosomiques (Hanahan & Weinberg, 2011).

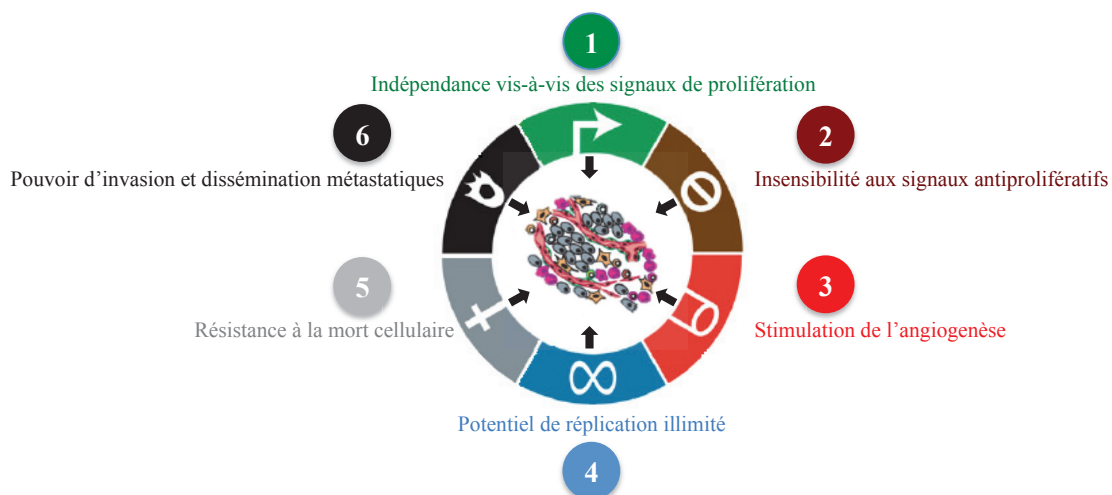


Figure 9 : caractéristiques des cellules cancéreuses.

D'après Hanahan and Weinberg, 2000 (Hanahan & Weinberg, 2000).

3.2.1 Indépendance vis-à-vis des signaux de prolifération

Les tissus normaux assurent le maintien de l'homéostasie cellulaire alors que les cellules cancéreuses cherchent à devenir maîtresses de leur propre destin avec une capacité de prolifération infinie. Cette propriété est souvent liée à l'activation de proto-oncogènes impliqués dans les différentes voies de signalisation de la prolifération ou de la croissance cellulaire. En fait, un oncogène est une version altérée dominante d'un gène dit proto-oncogène qui est normalement impliqué dans ces voies et favorise ainsi la transformation maligne. Par exemple, les facteurs de croissance, les récepteurs à la surface de la membrane plasmique, les protéines des voies de signalisation, les facteurs de régulation de la transcription peuvent avoir des fonctions oncogéniques.

Les cellules cancéreuses maintiennent la signalisation proliférative par l'intermédiaire de différentes stratégies alternatives plus ou moins indépendantes des signaux mitogènes. Les cellules cancéreuses peuvent produire des facteurs de croissance, ligands des récepteurs de la cellule cancéreuse elle-même, on parle alors de stimulation proliférative autocrine. Par exemple, il a été montré que l'interleukine 6, molécule inflammatoire, était capable de favoriser la prolifération de cellules tumorales de l'endomètre et du sein (Che et al., 2014; Iliopoulos, Hirsch, & Struhl, 2009).

Mais les cellules cancéreuses peuvent également envoyer des signaux aux cellules stromales qui vont pouvoir en retour alimenter en facteurs de croissance les cellules tumorales, on parle alors de stimulation paracrine (Bhowmick, Neilson, & Moses, 2004). Il a été montré notamment que les chémokines CXCL14 et CXCL12 (C-X-C motif chemokine 14 et C-X-C motif chemokine 12), surexprimées respectivement dans les cellules myoépithéliales et les cellules myofibroblastiques, lient un récepteur des cellules épithéliales en favorisant ainsi la prolifération, la migration et l'invasion dans le cancer du sein (Allinen et al., 2004). Une délétion au niveau du récepteur TGF-béta de type 2 des fibroblastes est corrélée à une augmentation de la sécrétion du facteur de croissance Hepatocyte Growth Factor (HGF) par les fibroblastes. L'HGF ainsi sécrété va se fixer à son récepteur de l'HGF localisé au niveau de la surface des cellules épithéliales et favorise ainsi l'invasion et la prolifération des cellules épithéliales (N. Cheng, Chytil, Shyr, Joly, & Moses, 2008).

Les cellules cancéreuses peuvent permettre à leurs récepteurs d'augmenter leur sensibilité vis à vis du facteur de croissance afin de les rendre « hyperactifs ». L'étude des tumeurs de Kyllikki Haapala a montré que dans 33% des cas l'amplification du gène codant pour le récepteur aux androgènes, entraînant une augmentation de l'expression de son récepteur, est associée à une augmentation de prolifération puisque ces récepteurs deviennent alors sensibles aux faibles doses d'androgènes (Haapala et al., 2007).

Les cellules cancéreuses peuvent également altérer la structure de leurs récepteurs ou d'un élément en aval de leurs voies de signalisation afin qu'ils soient actifs de manière constitutive. Par exemple, 40% des mélanomes humains ont une mutation activatrice de la protéine B-RAF (Rapidly Accelerated Fibrosarcome homologue B) résultant en l'activation constitutive de la voie des MAPK (Mitogen Activated Protein Kinase) (Davies & Samuels, 2010). D'autre part, le gène codant un récepteur de facteur de croissance avec une partie intracellulaire à activité Tyrosine Kinase ERBB2 (erb-b2 receptor tyrosine kinase 2) se trouve dans une région du génome qui est amplifiée dans de nombreux cancers. La surexpression de cette protéine entraîne l'activation constitutive de la phosphorylation de la partie intracellulaire de celui-ci et favorise ainsi la prolifération cellulaire *via* notamment l'activation d'AKT (RAC-alpha serine/threonine-protein kinase PKB) (Figure 10)(B. P. Zhou et al., 2000) (D. Yu & Hung, 2000) (X. Zhou et al., 2004) (Stoica et al., 2003).

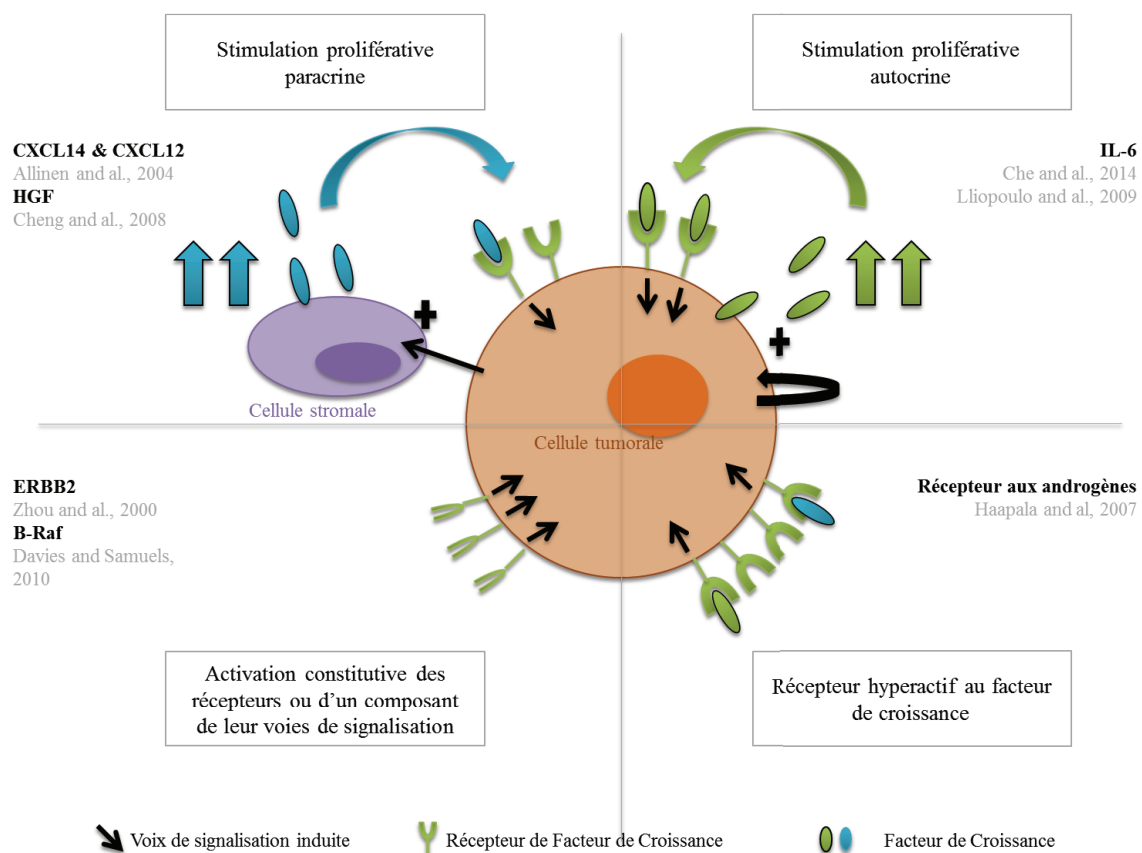


Figure 10 : stratégies alternatives de la cellule cancéreuse pour le maintien de la prolifération.

Une voie de signalisation majeure dans le maintien de l'homéostasie cellulaire est la voie PI3K/AKT/mTOR (Phosphatidyl Inositol 3-Kinase ; Mechanistic target Of Rapamycin) qui est souvent dérégulée dans les cellules cancéreuses. Cette voie est impliquée dans des fonctions régulatrices au niveau de l'apoptose, du métabolisme, de l'angiogenèse, de la croissance et du cycle cellulaire (Carnero, Blanco-Aparicio, Renner, Link, & Leal, 2008; Dreyer, Sablin, Faivre, & Raymond, 2009). La PI3K est une kinase intracellulaire permettant la conversion du PIP2 en PIP3

(Phosphatidyl Inositol 4,5 Phosphate en Phosphatidyl Inositol 3,4,5 Phosphate). La phosphatase PTEN quant à elle permet la dégradation du PIP3 et agit donc comme un inhibiteur de la PI3K. Ainsi, les mutations associées à une perte de fonction de la phosphatase PTEN sont très fréquentes et favorisent la tumorigenèse *via* l'amplification de la voie de la PI3K (Carnero et al., 2008; Jiang & Liu, 2009; Leslie & Longy, 2016; Salmena, Carracedo, & Pandolfi, 2008; Yuan & Cantley, 2008). A l'inverse, des mutations activatrices de la PI3K ou de la sérine/thréonine kinase AKT en aval favorisent également le processus de tumorigenèse (Carnero et al., 2008). Une autre kinase, nommée mTOR, intervient dans cette voie de signalisation en aval de la PI3K et d'AKT. Cette dernière joue un rôle important dans la synthèse des protéines, l'autophagie, la croissance et la prolifération cellulaire (Sudarsanam & Johnson, 2010). Des études récentes ont permis d'identifier des mutations activatrices de mTOR favorisant également le processus tumoral (Figure 11)(Gerlinger et al., 2012; Grabiner et al., 2014; Sato, Nakashima, Guo, Coffman, & Tamanoi, 2010).

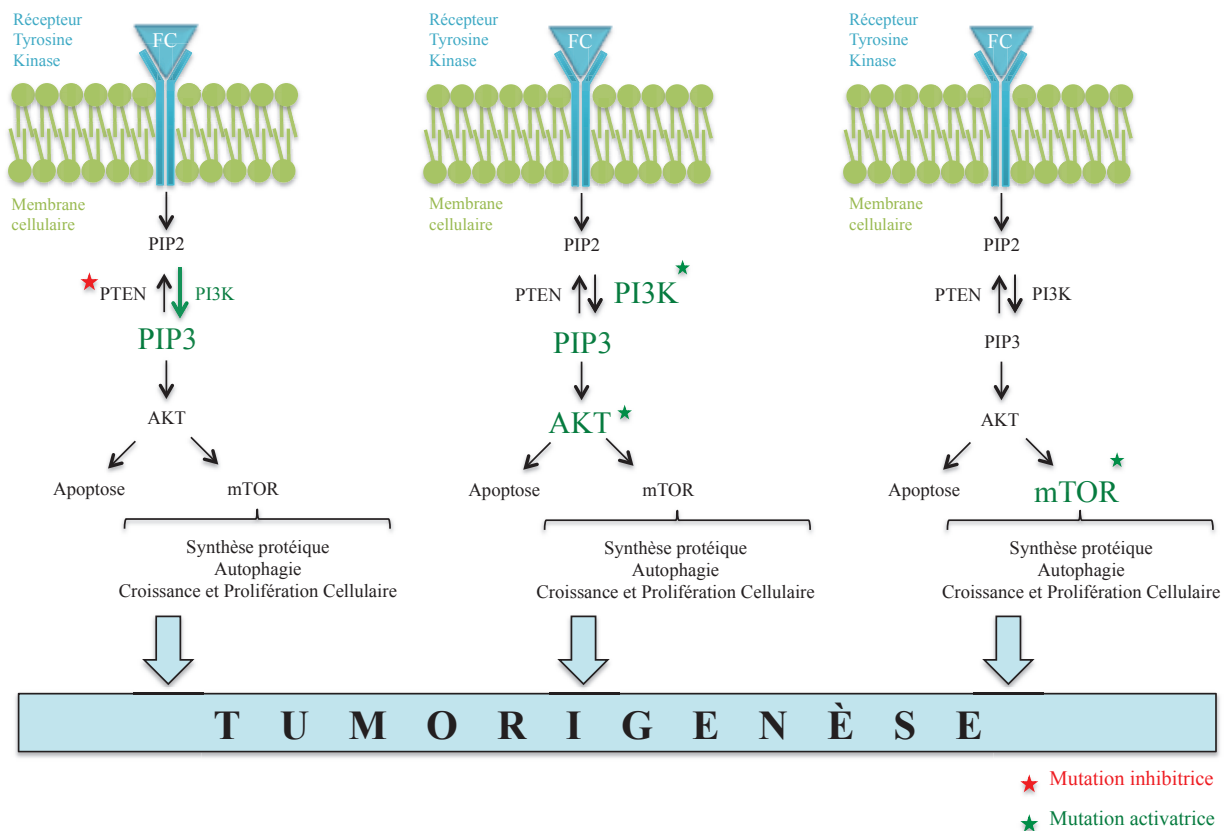


Figure 11 : mutations de la voie PI3K/AKT/mTOR associées à la tumorigenèse.

3.2.2 Insensibilité aux signaux antiprolifératifs

Au-delà de l'indépendance vis-à-vis des signaux de prolifération, les cellules cancéreuses contournent les fonctions des gènes dits suppresseurs de tumeurs. Ces gènes peuvent exercer des fonctions de contrôle négatif du cycle cellulaire (comme par exemple *P16/INK4A*, *P21* (Cyclin-dependent kinase inhibitor 1 (CDK-1)), *RB*), du maintien de la stabilité du génome (comme par

exemple *p53*, *ATM*), de la réparation de l'ADN ou encore des gènes pro-apoptotiques (exemple *p53*). Parmi ces gènes, les plus connus sont ceux codant pour les protéines RB et P53 (Figure 12). La protéine RB exerce un contrôle négatif sur le cycle cellulaire et est une protéine de séquestration de la protéine E2F (Deshpande, Sicinski, & Hinds, 2005; Sherr & McCormick, 2002). Une perte de fonction de la protéine RB dans des tumeurs humaines ou dans des modèles murins contribue à l'initiation et à la progression tumorale (Burkhart & Sage, 2008). La protéine P53 est considérée comme le gardien de l'intégrité du génome (Biegging & Attardi, 2012; Meek, 2015). C'est un facteur de transcription qui répond à différents stressés génotoxiques et cytotoxiques en provoquant l'arrêt du cycle cellulaire, l'apoptose, la réparation de l'ADN, la différenciation cellulaire ou encore la sénescence (Figure 12)(Evan & Vousden, 2001; Stommel & Wahl, 2005; Vassilev, 2005). Toutefois la plupart des mutants de P53 associés à différents phénotypes cancéreux sont caractérisés par une perte de fonction suppresseur de tumeur d'une part, et un gain de fonction oncogénique contribuant à la progression maligne d'autre part (P. A. Muller & Vousden, 2014).

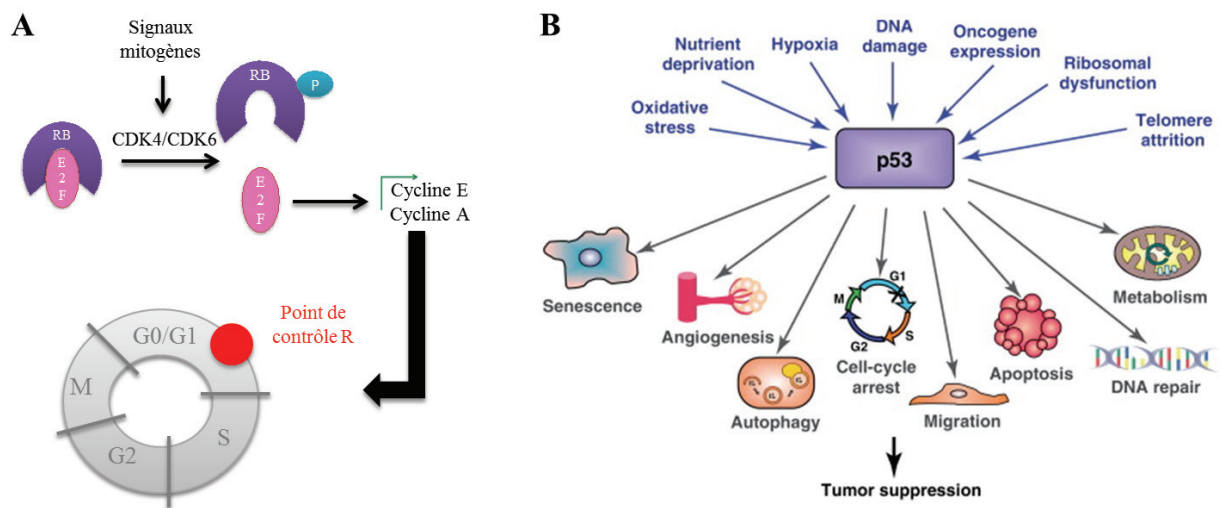


Figure 12 : RB et P53, facteurs suppresseurs de tumeur.

A. Rôle de la séquestration de RB au cours du cycle cellulaire (À partir de Knudsen and Knudsen, 2008)(Knudsen & Knudsen, 2008). B. P53 est le gardien du génome (Biegging & Attardi, 2012).

3.2.3 Stimulation de l'angiogenèse

L'angiogenèse est le processus dynamique aboutissant à la formation de nouveaux vaisseaux sanguins à partir de vaisseaux préexistants, le plus souvent des capillaires et des veinules (Folkman & Shing, 1992; Scoazec, 2000). Au cours de l'embryogenèse, le développement du système vasculaire implique deux mécanismes distincts. D'une part, la vasculogenèse qui consiste en la formation de vaisseaux *de novo* à partir de cellules précurseurs endothéliales (formation des premiers vaisseaux embryonnaires mais également l'extension du réseau vasculaire) (Swift & Weinstein, 2009). D'autre part, l'angiogenèse qui consiste en la formation de nouveaux vaisseaux à partir de ceux existants *via*

deux types de processus d'angiogenèse (Carmeliet & Jain, 2000; Risau, 1997). Le premier type est l'angiogenèse par bourgeonnement avec les « tip cells » qui sont des cellules endothéliales motiles et invasives créant des extensions qui guident la formation de nouvelles ramifications (R. H. Adams & Alitalo, 2007). Le second type est l'angiogenèse par intussusception où un vaisseau existant peut se scinder en deux vaisseaux par compartimentation (Figure 13A). Chez l'adulte une fois que le réseau vasculaire définitif est établi, les cellules endothéliales demeurent essentiellement dans un état quiescent à l'exception des processus de néo-vascularisation qui apparaissent au cours d'évènements physiologiques ou pathologiques. En effet chez l'adulte, dans le cadre de processus physiologiques tels que la cicatrisation, l'appareil reproducteur femelle durant les cycles menstruels et la gestation, l'angiogenèse survient, mais uniquement de façon transitoire (Ferrara et al., 1998; Goede, Schmidt, Kimmina, Kozian, & Augustin, 1998).

En revanche, dès les stades précoces de la tumorigenèse, il y a une commutation angiogénique (« switch angiogenic ») en faveur d'une stimulation permanente de ce processus favorisant ainsi la formation continue de nouveaux vaisseaux dont l'organisation semble avoir de multiples anomalies structurelles et fonctionnelles (Baluk, Hashizume, & McDonald, 2005; Hanahan & Folkman, 1996; Raica, Cimpean, & Ribatti, 2009). Outre les mécanismes classiques d'angiogenèse et de vasculogenèse, les cellules tumorales utilisent d'autres processus pour la vascularisation tumorale : le phénomène des vaisseaux co-optés où les cellules tumorales entourent les vaisseaux sanguins préexistants (Holash et al., 1999), le phénomène de mimétisme vasculaire où les cellules tumorales forment des vaisseaux sanguins tumoraux à la place des cellules endothéliales (Bittner et al., 2000; Maniotis et al., 1999), et le phénomène de différenciation où la formation d'un endothélium tumoral se fait à partir des cellules souches tumorales qui acquièrent un phénotype pseudo-endothélial et tapissent la paroi interne des vaisseaux. (Figure 13B). D'autre part, la vascularisation tumorale semble impliquer également l'artério-veinogenèse qui consiste en un remodelage des artères et veines préexistantes (Nagy, Chang, Shih, Dvorak, & Dvorak, 2010).

D'un point de vue moléculaire, la commutation angiogénique, étape indispensable au développement tumoral, est régulée par des facteurs activateurs, tel que VEGF, et inhibiteurs, tel que la Thrombospondine-1 (TSP-1), de l'angiogenèse (Figure 13C). Ces facteurs pouvant être sécrétés à la fois par les cellules tumorales et par les cellules du microenvironnement (Baeriswyl & Christofori, 2009; Bergers & Benjamin, 2003). Le facteur de croissance de l'endothélium vasculaire A (VEGF-A) semble être la plus puissante des molécules angiogéniques et cible les cellules endothéliales en se fixant à des récepteurs de surface de type tyrosine kinase (Ferrara & Henzel, 1989). A l'inverse des cellules adultes normales, les cellules tumorales sont très dépendantes du VEGF-A. L'expression du VEGF-A est induite par l'hypoxie et par la signalisation oncogénique (Carmeliet, 2005; Ferrara, 2009; Mac Gabhann & Popel, 2008). Certains facteurs pro-angiogéniques de la famille des facteurs de

croissance fibroblastiques (FGF) sont impliqués dans l'angiogenèse tumorale suite à leur activation chronique (Baeriswyl & Christofori, 2009). A l'inverse la TSP-1, l'angiostatine et l'endostatine (fragment C-terminal du collagène XVIII (Carboxy-terminal)) peuvent agir comme des inhibiteurs endogènes de l'angiogenèse et sont donc des obstacles naturels de l'angiogenèse tumorale (Folkman, 2002, 2006; Kazerounian, Yee, & Lawler, 2008; Nyberg, Xie, & Kalluri, 2005; Ribatti, 2009).

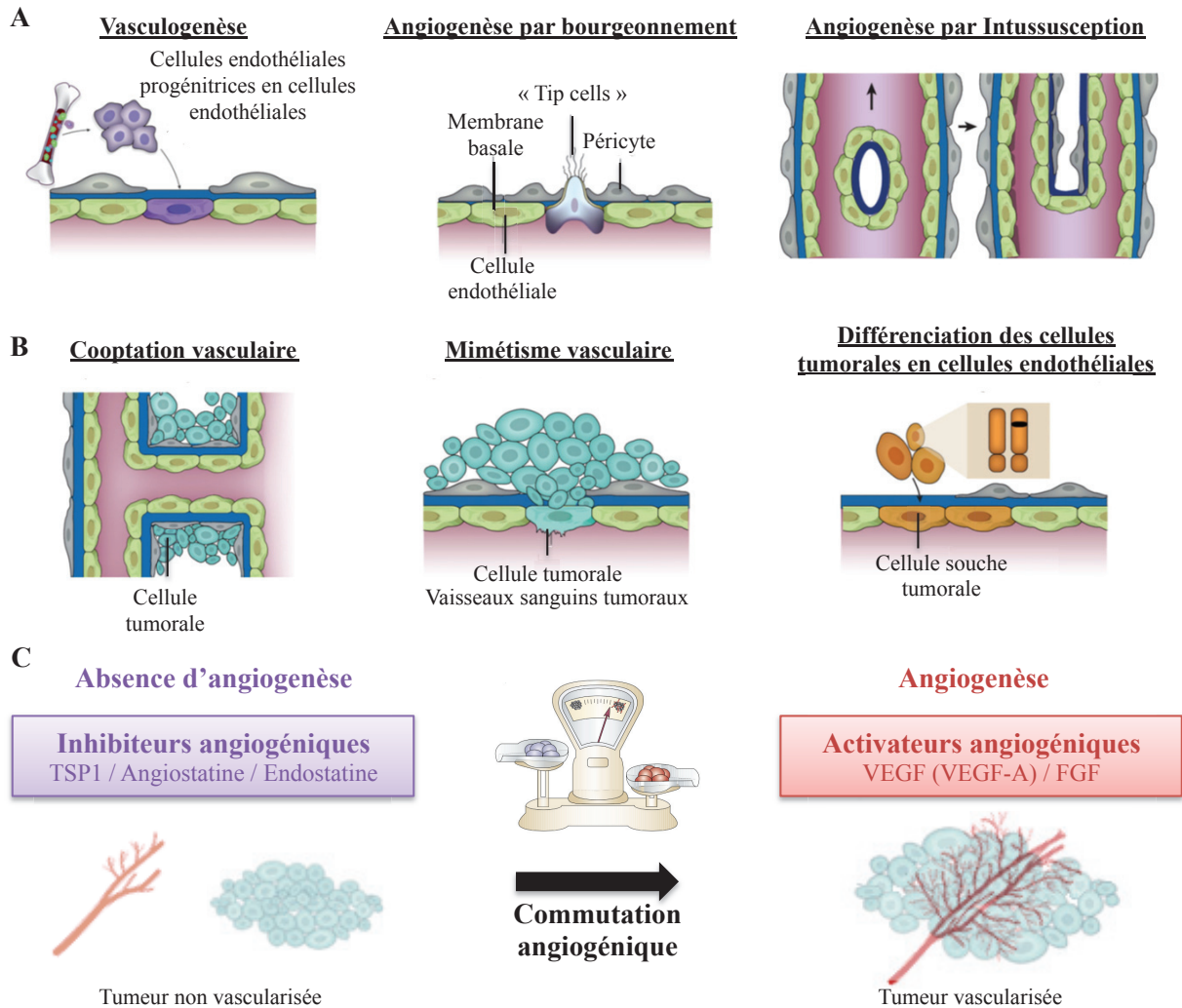


Figure 13: formation de vaisseaux sanguins et la commutation angiogénique.

A: La formation des vaisseaux sanguins dans les tissus normaux et tumoraux. B: La formation des vaisseaux sanguins dans les tumeurs (D'après Jain and Carmeliet 2012) (Jain & Carmeliet, 2012) C: La commutation angiogénique est dépendante de la balance entre des facteurs pro- et anti-angiogéniques et favorise le développement tumoral (D'après Bergers and Benjamin, 2003) (Bergers & Benjamin, 2003).

En effet, il existe des tumeurs très fortement angiogéniques et donc très vascularisées (Turner, Harris, Melmed, & Wass, 2003; Zee et al., 2010). Toutefois, certaines tumeurs y compris des tumeurs très agressives comme les adénocarcinomes canauxiers pancréatiques sont hypo-vascularisées et sont caractérisées par la présence de déserts vasculaires au niveau du stroma (Olive et al., 2009).

3.2.4 Potentiel de réplication illimité

La plupart des cellules normales ont un nombre de cycles cellulaires limités. En effet la prolifération des cellules est freinée par deux barrières : la sénescence réplivative (Hayflick & Moorhead, 1961) et la catastrophe génétique, autrement nommée « la crise » (Wright & Shay, 1992).

La sénescence réplivative est un arrêt de prolifération cellulaire mis en place en réponse à un raccourcissement des télomères. Les télomères préservent l'intégrité du matériel génétique au cours des cycles cellulaires et participent à l'architecture fonctionnelle des chromosomes (Zakian, 1995). En effet, ce sont de multiples répétitions d'hexa-nucléotides en tandem composées de protéines spécifiques qui protègent les extrémités des chromosomes en évitant les fusions par leurs extrémités et la dégradation des séquences contenues dans la partie interne du chromosome (Blackburn & Gall, 1978; Blasco, 2005; de Lange, 2006; H. Muller, 1938; Shay & Wright, 2000). A chaque cycle de division cellulaire, les télomères se raccourcissent progressivement ce qui rend les extrémités chromosomiques dépourvues de protection (Harley, Futcher, & Greider, 1990; Hastie et al., 1990). Carol Greiner et Elisabeth Blackburn ont découvert une enzyme capable de synthétiser et d'allonger les télomères, la télomérase (Greider & Blackburn, 1985). Une activité télomérase reconstituée dans des cellules fibroblastiques permettrait de stabiliser la longueur des télomères de ces cellules et ainsi leur permettrait d'échapper à la sénescence (Bodnar et al., 1998). Chez l'Homme, cette enzyme est présente dans les cellules de la lignée germinale et dans certaines cellules souches somatiques, mais elle est absente ou présente en très faible quantité dans les cellules somatiques adultes (K. Masutomi et al., 2003). Cette enzyme est en revanche réactivée dans la majorité des cancers (Kim et al., 1994; Kenkichi Masutomi & Hahn, 2003). La présence de la télomérase permet de maintenir la longueur des télomères malgré les cycles de division cellulaire que la cellule subit, et confère donc aux cellules une résistance à l'induction de la sénescence alors que, la suppression de la télomérase entraîne le raccourcissement des télomères et donc l'activation de l'un ou l'autre de ces obstacles prolifératifs (Figure 14) (Shay & Wright, 2000). La sénescence réplivative est une réponse cellulaire apparentée à la réponse aux dommages de l'ADN. En effet, l'entrée en sénescence est dépendante des voies suppresseurs de tumeurs P53/RB et présente de nombreuses similitudes avec la réponse aux cassures de l'ADN (d'Adda di Fagagna et al., 2003). Il est important de préciser que dans certains cas, la modification des télomères peut déclencher une mort par apoptose dépendante de P53 (Deng, Chan, & Chang, 2008).

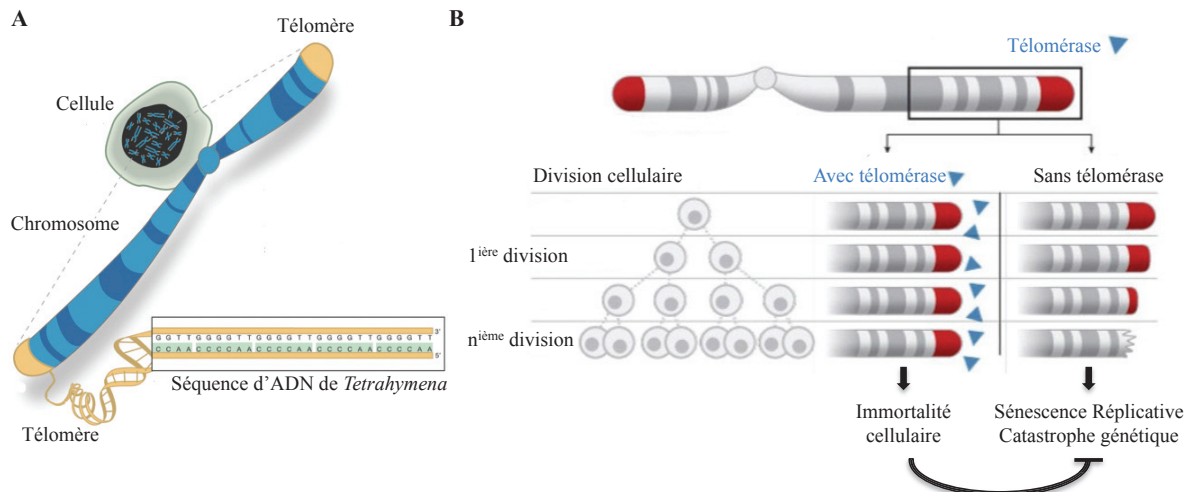


Figure 14 : structure et fonction des télomères.

D'après "The nobel Committe for physiology or medicine" 2009 / Annika Röhl.

En revanche, lors de l'inactivation de P53 et de RB, les cellules échappent à la sénescence et continuent de proliférer en accumulant des altérations chromosomiques (Counter et al., 1992; Ducray, Pommier, Martins, Boussin, & Sabatier, 1999; Shay, Van Der Haegen, Ying, & Wright, 1993) (Klingelhutz, Barber, Smith, Dyer, & McDougall, 1994). En accord avec ces données, il a été montré sur des souris qu'une déficience de P53 associée à une altération de la fonction des télomères favoriserait la progression tumorale (Artandi et al., 2000; Artandi & DePinho, 2010; L. Chin et al., 1999). La prolifération des cellules est alors limitée par une seconde barrière proliférative, appelée la « crise ». En effet, lorsque l'instabilité du génome devient trop importante les cellules entrent en crise et meurent massivement. En revanche certaines cellules sont capables d'échapper à cette seconde barrière proliférative. Celles-ci ont stabilisé leurs télomères soit par la réactivation de la télomérase soit par un processus de recombinaison dit Alternative Lengthening of Telomeres (ALT) (Counter et al., 1998) (Deng et al., 2008; Muntoni & Reddel, 2005; J. Zhu, Wang, Bishop, & Blackburn, 1999). L'équipe de Christophe Raynaud a étudié les modifications de la longueur de télomères et l'activation de la voie de réponse aux dommages de l'ADN à trois étapes différentes de la carcinogénèse mammaire humaine. Les lésions précancéreuses auraient des niveaux d'expression de la télomérase non significatifs et semblent marquées par le raccourcissement des télomères. En revanche, les carcinomes déclarés semblent au contraire avoir un niveau d'expression élevé, en accord avec les télomères plus allongés. Toutefois l'acquisition de la télomérase au cours du processus tumoral, permet le maintien des phénotypes acquis au stade des lésions précancéreuses avec des mutations favorisant le développement tumoral (Figure 15)(K. Chin et al., 2004; Raynaud et al., 2010).

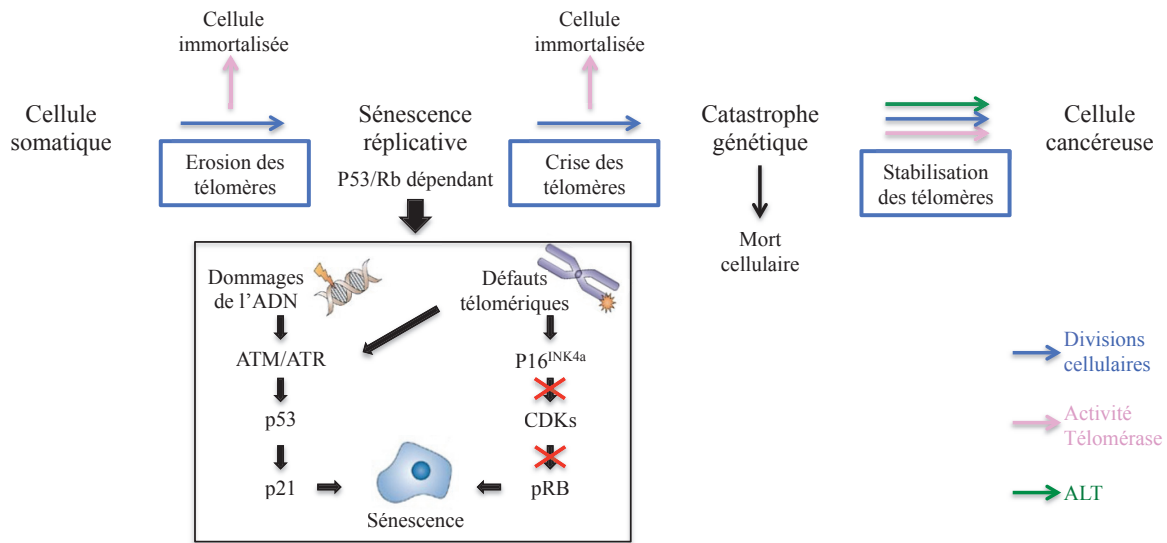


Figure 15 : sénescence répliquative et catastrophe génétique, des barrières prolifératives.

D'après « La sénescence » de Véronique Gire 2005 (Gire, 2005).

3.2.5 Résistance à la mort cellulaire

Dans le contexte cancéreux, la croissance des tumeurs n'est pas uniquement la conséquence d'une prolifération excessive mais c'est le résultat d'un déséquilibre entre la prolifération et la mort cellulaire (Martin & Green, 1995). En effet l'apoptose, considérée comme une barrière naturelle au développement tumoral, serait diminuée dans certaines tumeurs augmentant ainsi la malignité et la résistance aux traitements (Cory & Adams, 2002) (J. M. Adams & Cory, 2007; Lowe, Cepero, & Evan, 2004). Il existe deux voies majeures de l'apoptose qui mènent à l'activation des caspases, protéases à cystéine intracellulaires, puis à l'activation de la cascade de démolition cellulaire (J. M. Adams, 2003) (Vermeulen, Van Bockstaele, & Berneman, 2005). La première voie est dite extrinsèque ou voie des récepteurs de morts impliquant par exemple, les Fas Ligand/ Fas Receptor. Cette première voie n'implique pas ou très peu la mitochondrie. Elle est activée suite à un signal extracellulaire *via* la fixation des ligands spécifiques aux récepteurs de mort. La seconde voie est dite intrinsèque ou voie apoptotique mitochondriale, celle-ci implique plusieurs protéines de la famille des BCL-2 (pro-apoptotiques et anti-apoptotiques ; B-cell Chronic Lymphocytic leukemia/lymphoma-2) ainsi que des protéines mitochondriales (Vermeulen et al., 2005; J. Yu & Zhang, 2005). L'entrée en apoptose de la cellule *via* cette seconde voie est déterminée par l'interaction de 3 protéines de la famille des BCL-2 avec l'intervention de protéines pro et anti-apoptotiques (J. M. Adams & Cory, 2007). La sous famille des BCL-2 lie les protéines BAX (BCL-2-Associated X Proteins) et BAK (BCL-2 homologous antagonist/killer) par l'intermédiaire de leurs motifs BH3, dans la membrane externe mitochondriale afin d'annihiler leur effet pro-apoptotique. Lorsque l'inhibition anti-apoptotique est levée, BAX et BAK perturbent l'intégrité de la membrane mitochondriale externe pour

libérer le cytochrome C pro-apoptotique. Ensuite le Cytochrome C relâché par la mitochondrie va activer la cascade des caspases qui induiront de nombreux changements cellulaires associés à l'apoptose (diminution du volume cellulaire, condensation et fragmentation de la chromatine, bourgeonnements cellulaires et formation de corps apoptotiques par exemple). Au-delà de ces deux voies principales d'activation, il existe de nombreux facteurs capables de déclencher ou non l'apoptose comme notamment le suppresseur de tumeur P53 (Junttila & Evan, 2009) (Lowe et al., 2004). En effet cette protéine est capable d'enclencher la voie apoptotique par l'intermédiaire de deux protéines de la famille BCL-2 (BH3 only ; B-cell Chronic Lymphocytic leukemia/lymphoma-2) que sont NOXA et PUMA (P53 Upregulated Modulator of Apoptosis) en réponse à des anomalies de l'ADN et à des cassures chromosomiques (Figure 16). Parmi les stratégies mises en place par les cellules tumorales pour contourner l'apoptose, il y a la perte de la protéine P53. Les tumeurs peuvent également atteindre des objectifs similaires en augmentant l'expression des régulateurs anti-apoptotiques (BCL-2), en diminuant l'expression des facteurs pro-apoptotiques (BAX, PUMA) ou en court-circuitant la voie extrinsèque d'apoptose par le ligand.

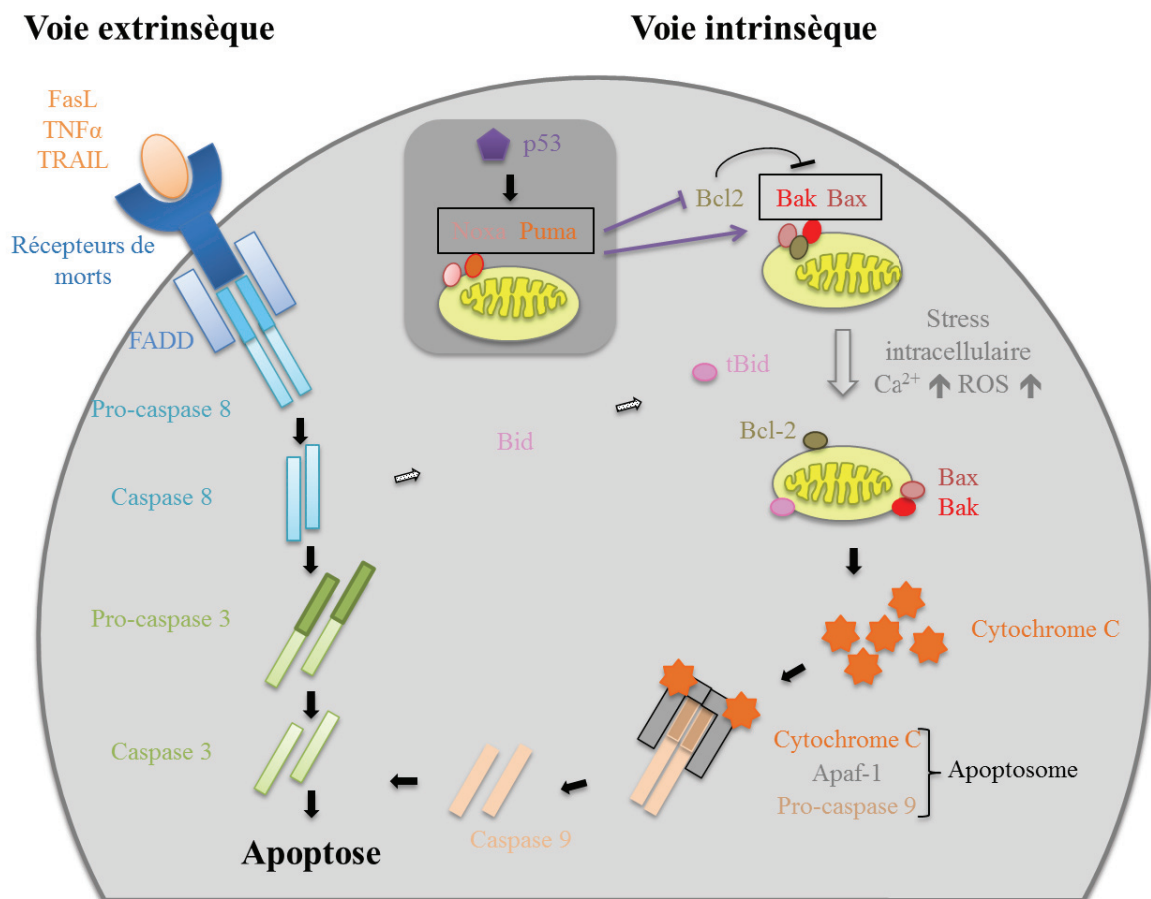


Figure 16 : deux voies principales de l'apoptose.

A ce jour, l'apoptose n'est plus la seule mort envisagée comme barrière contre le cancer, puisqu'il existe d'autres formes de mort cellulaire telles que l'autophagie, la nécrose et la sénescence. L'autophagie est une voie majeure du catabolisme lysosomal qui permet la dégradation, en plusieurs étapes, des macromolécules et des organites cellulaires. Il y a la formation d'une vacuole à double membranes, nommée phagophore, dans laquelle est immergée une partie du cytoplasme. Puis la membrane du phagophore s'allonge jusqu'à ce que les deux extrémités se rejoignent pour former l'autophagosome qui fusionnera ensuite avec les lysosomes (Figure 17A)(Mizushima, 2007). Lorsque le contexte physiologique est normal, les niveaux d'autophagie sont faibles. Dans certains états de stress, le plus évident semble être la carence en nutriments, l'autophagie peut être fortement induite (Levine & Kroemer, 2008; Mizushima, 2007). Au cours de ces dernières années, le rôle de l'autophagie a été très controversé. Cette voie de dégradation lysosomale semble agir comme un régulateur clef de la tumorigenèse en empêchant le développement de cancer précoce et en facilitant la progression de la tumeur avancée (Kenific & Debnath, 2015). L'équipe de Guido Kroemer avance l'hypothèse que ce phénomène implique une perte temporaire des compétences autophagiques notamment au cours des stades précoces du développement tumoral. Ainsi à des stades précoces, des défauts du processus autophagique permettraient aux cellules saines d'acquérir des caractéristiques malignes *via* l'accumulation de dommages cellulaires (Galluzzi et al., 2015; Xu, Zhang, Wang, Xu, & Liu, 2015). Cette hypothèse est en accord avec l'absence d'expression de certains gènes responsables de l'autophagie et avec l'activation de voies comme PI3K/AKT/mTOR qui inhibent l'autophagie dans des tumeurs malignes humaines (Ahn et al., 2007; Maiuri et al., 2009). De plus, dans les modèles de souris l'inactivation de gènes spécifiques de l'autophagie (comme Becline-1) semblent augmenter la susceptibilité de développer un cancer (Levine & Kroemer, 2008; Qu et al., 2003; White & DiPaola, 2009; H. Zhu et al., 2007). Alors qu'à des stades plus avancés, c'est à dire lorsque la malignité est établie, la restauration des réponses autophagiques serait essentielle pour soutenir la survie, la prolifération et la croissance des cellules cancéreuses (Galluzzi et al., 2015). L'autophagie pourrait également aider à la survie des cellules tumorales en favorisant le recyclage intracellulaire pour fournir ainsi des substrats pour le métabolisme cellulaire et donc promouvoir la tumorigenèse (Figure 17B)(Galluzzi et al., 2015; Kimmelman, 2011; White, 2012; Xu et al., 2015). Toutefois, les raisons du « switch autophagique » au cours du processus tumoral ne sont pas établies. D'autre part, les voies autophagique et apoptotique semblent interconnectées *via* la voie PI3K/AKT/mTOR et la Becline-1 entre autre. En fonction de l'état physiologique de la cellule et des signaux cellulaires, les voies autophagique et/ou d'apoptose peuvent être induites (Levine & Kroemer, 2008; Mathew, Karantza-Wadsworth, & White, 2007; Sinha & Levine, 2008; White & DiPaola, 2009).

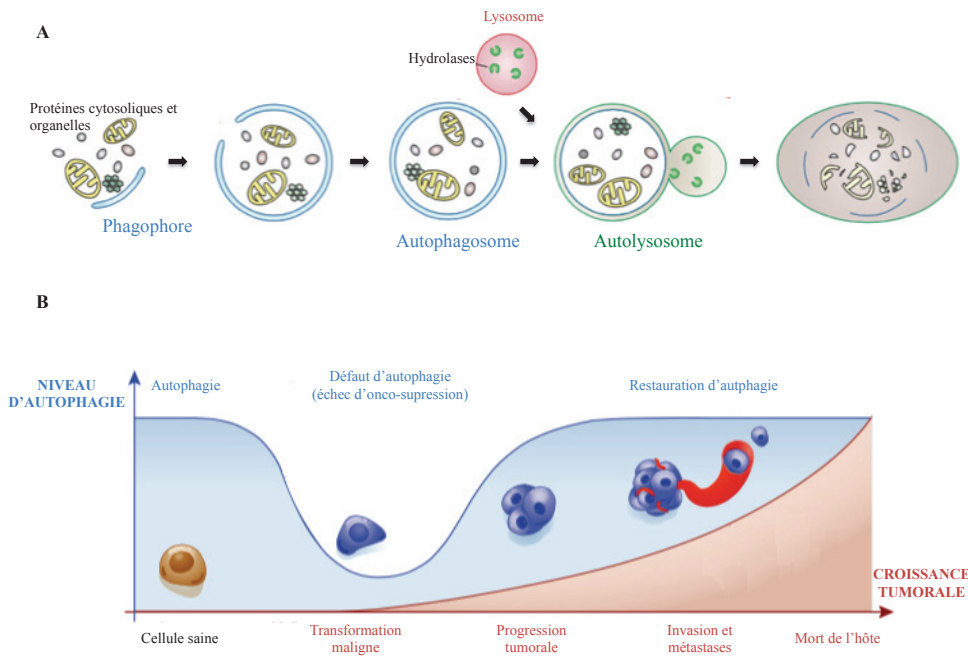


Figure 17: autophagie dans le contexte tumoral.

A: Rappel sur les étapes clés de l'autophagie (D'après Xie and Klionsky, 2007) (Xie & Klionsky, 2007). B: Autophagie dans la progression tumorale et la transformation maligne (D'après de Galluzzi and al., 2015) (Galluzzi et al., 2015).

Dans le cas de la mort par nécrose, la cellule gonfle et explose en libérant son contenu dans le microenvironnement local (Figure 18). Lorenzo Galluzzi et Guido Kroemer ont montré que la nécrose était, comme l'apoptose, un mécanisme régulé et non pas un évènement cellulaire accidentel ou aléatoire (Galluzzi & Kroemer, 2008; Hitomi et al., 2008; Zong & Thompson, 2006). Contrairement à l'apoptose et à l'autophagie, la nécrose entraîne la libération de signaux pro-inflammatoires dans le microenvironnement et ainsi les cellules nécrotiques sont capables de recruter des cellules inflammatoires du système immunitaire (Galluzzi & Kroemer, 2008; Grivennikov, Greten, & Karin, 2010). Les cellules nécrotiques peuvent libérer des facteurs de régulation comme l'IL-1 qui stimulent la prolifération des cellules viables et ainsi faciliter la prolifération néoplasique (Grivennikov et al., 2010). C'est ainsi que la nécrose favorise l'hyper-prolifération et le développement tumoral. Les cellules cancéreuses utilisent la mort nécrotique pour le développement tumoral en tolérant la nécrose afin d'utiliser les cellules inflammatoires qui apportent des facteurs de croissances pour stimuler la croissance des cellules survivantes (Hanahan & Weinberg, 2011).

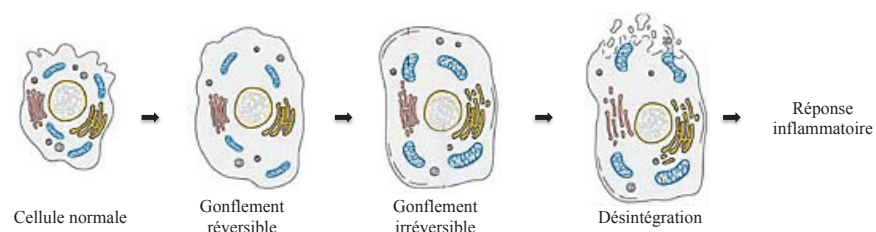


Figure 18: caractéristiques cellulaires associées à la nécrose.

La sénescence, quant à elle, est un processus physiologique qui entraîne une lente dégradation des fonctions de l'organisme suite à un arrêt de réplication qui aboutit à la mort de la cellule. Le premier type de sénescence cellulaire découvert est lié au raccourcissement des télomères et est appelé sénescence répllicative. Nous en avons discuté dans le chapitre 3.2.4 au sujet du potentiel de réplication illimité. Aujourd'hui d'autres types de sénescence, indépendants des télomères, ont été décrits. La sénescence prématurée est induite par différents stress (endommagement de l'ADN, modification de la structure de la chromatine, signaux mitogéniques, activation d'oncogènes) aussi bien dans les cellules normales que dans les cellules tumorales. Au-delà du rôle de protection de la sénescence, ce processus pourrait dans certaines conditions favoriser l'émergence des cellules tumorales et promouvoir la croissance des cellules pré-néoplasiques.

3.2.6 Pouvoir d'invasion et de dissémination métastatique

La tumorigenèse est l'ensemble des événements impliqués dans le développement d'un cancer à partir d'une cellule saine et comprend la formation de la tumeur primaire d'une part et la cascade métastatique d'autre part. Cette cascade métastatique est initiée par l'invasion locale qui consiste en la migration des cellules issues de la tumeur primaire dans la matrice extracellulaire environnante. Néanmoins cette première étape peut être dissociée des suivantes puisqu'une tumeur invasive localement ne devient pas forcément métastatique. La seconde étape du processus est l'intravasation, permettant aux cellules tumorales de pénétrer dans la circulation *via* le franchissement de la paroi des vaisseaux sanguins ou lymphatiques. Cette seconde étape initie donc la dissémination dans la circulation supposant que les cellules tumorales deviennent alors circulantes pour une durée indéterminée et survivent en suspension dans le sang ou dans la lymphe jusqu'à leur extravasation qui consiste en la traversée de la paroi des vaisseaux sanguins ou lymphatiques pour atteindre et coloniser le tissu d'un organe cible (troisième et quatrième étape). Il s'en suit donc la dernière étape de la progression métastatique : la formation des métastases allant de la formation de micrométastases à la formation de métastases cliniquement décelables (De Craene & Berx, 2013; Talmadge & Fidler, 2010; Thiery, Acloque, Huang, & Nieto, 2009). Pour cela il faut que les cellules tumorales s'implantent dans l'organe cible, survivent plus ou moins longtemps, c'est ce qu'on appelle la « dormance » et acquièrent à nouveau la capacité à proliférer pour donner naissance à une micro-métastase, puis une macro-métastase (Figure 19). De plus, le développement des métastases est un processus très inefficace car seulement 2% des cellules tumorales circulantes survivant dans le flux vasculaire sont à l'origine de micro-métastases et seulement 1% de ces micro-métastases progressent jusqu'à un stade macroscopique (Luzzi et al., 1998).

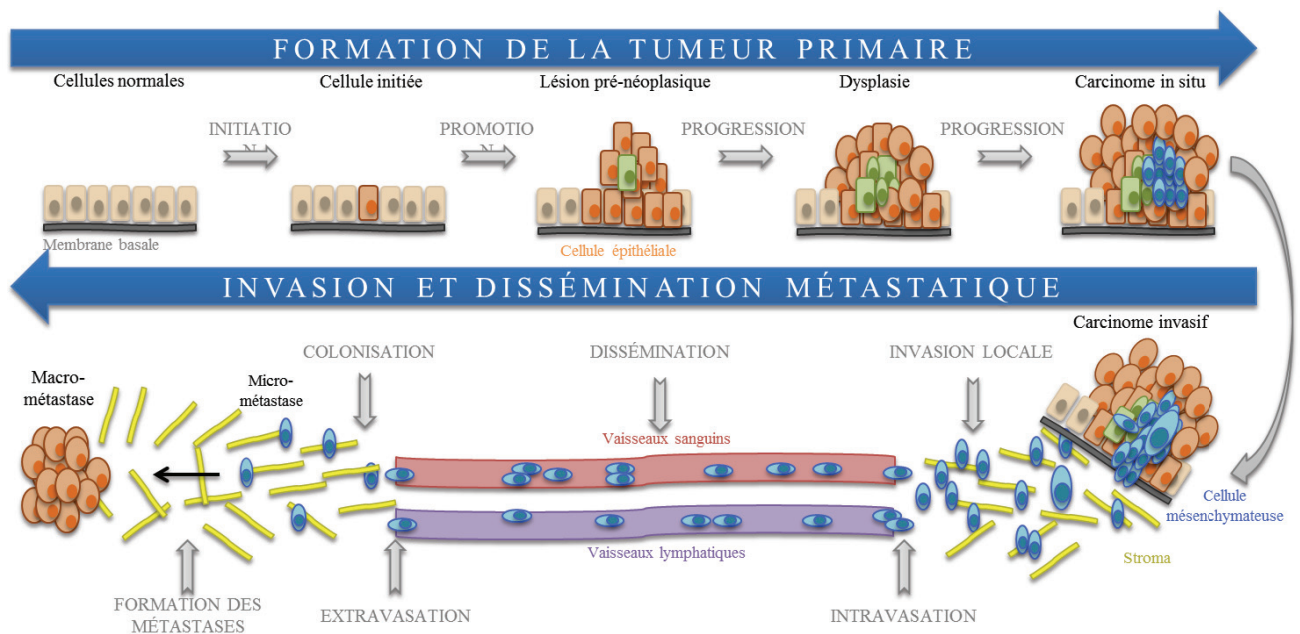


Figure 19 : étapes clés de la tumorigénèse.

Les cellules impliquées dans ce processus métastatique acquièrent de nouvelles propriétés. Elles ont des modifications de leurs contacts intercellulaires et cellule-matrice extracellulaire (MEC) mais également une diminution de l'expression des molécules favorisant un état cytotatique. A l'inverse, dans ces cellules il y a une augmentation de l'expression des molécules favorisant la migration cellulaire. En effet dans des tumeurs épithéliales humaines, différents mécanismes d'inactivation de la E-Cadhérine (Epithelial-Cadherin), une glycoprotéine transmembranaire impliquée dans les interactions intercellulaires qui permettrait de limiter l'invasion tumorale, ont été rapportés. Au contraire, la N-Cadhérine (Neural Cadherin), activateur de la nucléation du cytosquelette d'actine, va être surexprimée dans des cellules de carcinomes invasifs (Figure 20)(Behrens, Mareel, Van Roy, & Birchmeier, 1989; Berx & van Roy, 2009; Cavallaro & Christofori, 2004; Gavard, Marthiens, Monnet, Lambert, & Mege, 2004).

De nombreuses études ont montré l'implication de la Transition Epithélio-Mésenchymateuse (TEM ou EMT pour Epithelio-Mesenchymal Transition en anglais) au cours de la cascade métastatique de cellules épithéliales transformées, mais également au cours de l'initiation tumorale (De Craene & Berx, 2013; Hugo et al., 2007; Morel et al., 2012). La TEM est un processus embryonnaire fondamental et réversible qui implique de profonds changements cellulaires, morphologiques et comportementaux. De manière intéressante, il a été montré que la TEM conférait aux cellules tumorales des propriétés de cellules souches, favorisant leur capacité d'adaptation dans un environnement différent de celui du tissu d'origine (Mani et al., 2008; Morel et al., 2008). Par ailleurs cette transition, permettant la conversion de cellules épithéliales polarisées et jointives en cellules mésenchymateuses individuelles et motiles, est régulée par des facteurs de transcription dits « EMT-TFs » (Epithelio-Mesenchymal Transition Transcription Factors)(Morel et al., 2012; Puisieux et al.,

2014). Certains de ces facteurs de transcription comme SNAI1 (snail family transcriptional repressor 1), SNAI2 (SLUG ; snail family transcriptional repressor 2), SNAI3 (SMUC ; snail family transcriptional repressor 3), ZEB1/2 (Zinc finger E-box-binding homeobox 1/2) répriment directement la transcription de la E-Cadhérine contrairement à d'autres de ces facteurs comme TWIST1 et TWIST2 qui la répriment indirectement (Puisieux et al., 2014; Thiery et al., 2009). Par ailleurs, il a été montré que la répression des facteurs de polarité cellulaire apico-basale pouvait être impliquée dans l'activation de voies oncogéniques (Aigner et al., 2007; Cordenonsi et al., 2011). De plus les EMT-TFs pourraient induire l'expression de métalloprotéases de la MEC, l'augmentation de la motilité et la résistance à l'apoptose (Aigner et al., 2007; Spaderna et al., 2008; Tsai, Donaher, Murphy, Chau, & Yang, 2012). En revanche, les dernières étapes de la progression métastatique implique une transition, opposée à la TEM, dite Mésoenchymo-Epithéliale (TME ou MET en anglais) montrant ainsi la plasticité des cellules tumorales (De Craene & Berx, 2013; Jolly et al., 2015). Une cellule tumorale disséminée ayant acquis un phénotype mésoenchymateux doit subir cette TME pour former des tumeurs secondaires avec un phénotype similaire à celui retrouvé au niveau de tumeur primaire (D. Gao, Vahdat, Wong, Chang, & Mittal, 2012; Hugo et al., 2007; Jolly et al., 2015). En effet cette transition est associée, entre autre, à une augmentation de la E-Cadhérine. Certains facteurs comme GRHL2 (GRainyHead Like transcription factor 2) et ELF3/ELF5 (E74 like ETS transcription factor 3 ; E74 like ETS transcription factor 5) sont sous exprimés au cours de la TEM alors qu'ils induisent une TME lorsqu'ils sont surexprimés dans des cellules mésoenchymateuses (De Craene & Berx, 2013). Toutefois la TEM peut être partielle et ainsi permettre aux cellules d'acquérir de nouveaux traits mésoenchymateux tout en exprimant des traits épithéliaux résiduels (Hanahan & Weinberg, 2011). De manière intéressante, la cascade métastatique met en évidence l'existence d'une balance Epithélio-Mésoenchymateuse qui est régie selon l'expression de facteurs de transcription en faveur de la TME comme par exemple, GRLH2, ELF5, miR-200 (micro RNA 200), miR-3 (micro RNA 3) ou en faveur de la TEM comme par exemple, TWIST1, SNAI1, ZEB1, PRRX1 (Figure 20)(Jolly et al., 2015).

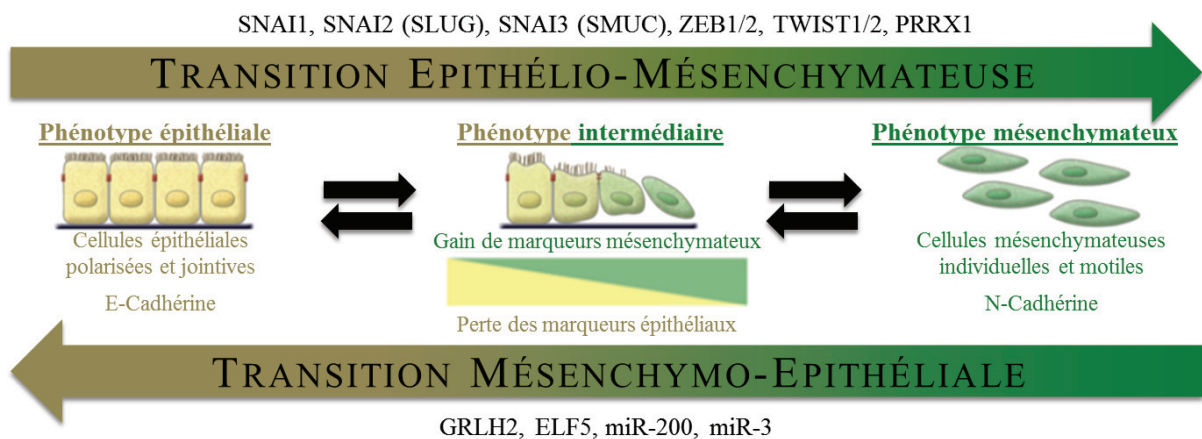


Figure 20: balance EMT/MET.

D'après Kalluri and Weinberg, 2009(Kalluri & Weinberg, 2009).

D'autres mécanismes que la TEM (autrement nommée invasion unicellulaire mésenchymateuse) sont mis en jeu par les cellules néoplasiques pour l'acquisition de propriétés invasives. Certains de ces programmes d'invasion impliquent des cellules isolées. C'est par exemple, le cas des mouvements amiboïdes (proches de celui des organismes unicellulaires amibes) impliquant une morphologie cellulaire ronde avec deux modes de migrations amiboïdes distincts : le mode protrusif pseudopodial et le mode contractile dit *blebby* à cause de la présence importante de bulles membranaires (*bleb*) (Friedl & Wolf, 2010). D'autres programmes d'invasion sont multicellulaires et impliquent des mouvements cellulaires massifs en cohortes et/ou en structures coordonnées tout en préservant l'adhésion cellule-cellule au cours de la migration. Ces stratégies ne nécessitent vraisemblablement pas de TEM dans sa forme complète (Figure 21)(Clark & Vignjevic, 2015).

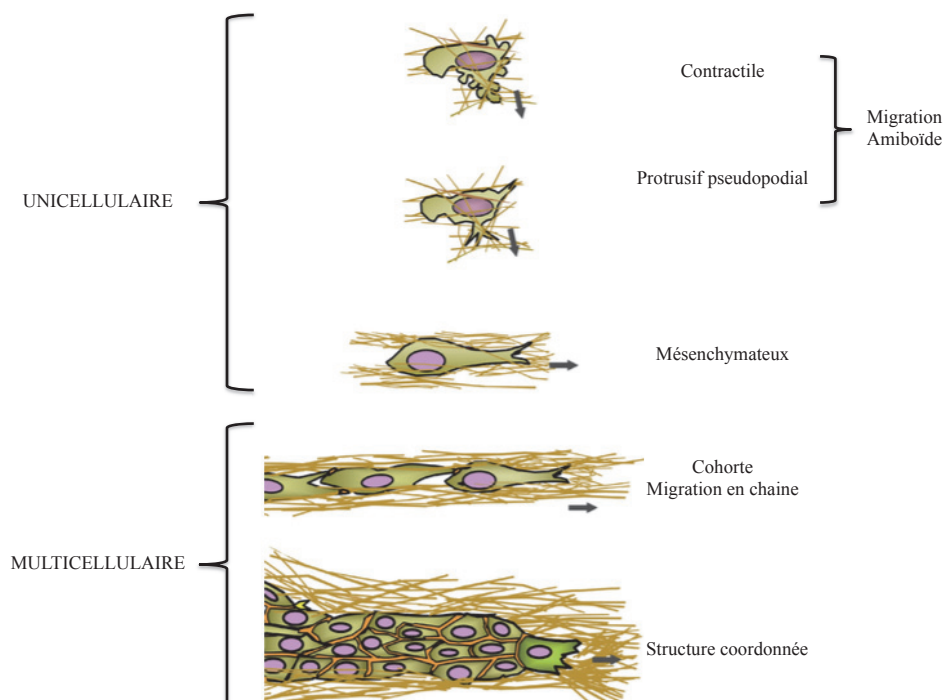


Figure 21 : différents modes de migration mis en œuvre par les cellules tumorales au cours de l'invasion.

D'après Clark and Vignjevic, 2015(Clark & Vignjevic, 2015).

4 Approches thérapeutiques

4.1 De la molécule au médicament

Le développement d'un nouveau médicament est long et onéreux. De manière générale, deux approches principales sont à l'origine d'un médicament. La première approche peut être le fruit du hasard. La découverte d'un principe actif peut être issue des connaissances de la chimie médicinale suite au fractionnement des extraits de plantes ou suite au criblage d'extraits de plantes ou de bibliothèques de composés chimiques. Cette première approche consiste en réalité à identifier une substance qui a l'effet pharmacologique recherché lors de son administration à un être humain, à un animal ou à des cellules sans en connaître le mode d'action. La seconde approche se fait suite à la découverte d'une cible moléculaire. Elle consiste en la recherche de composés capables de se fixer à la cible moléculaire déterminée afin d'en moduler ses propriétés. Ainsi, lorsque la structure tridimensionnelle de la cible est connue, nous pouvons, par des approches *in silico*, prédire l'affinité de fixation de certaines molécules chimiques pour la cible. Toutefois, la prédiction de l'affinité de fixation demande une compréhension des différentes interactions mises en jeu entre la molécule chimique et la cible, certes, mais également entre la molécule chimique et son véhicule au cours de son acheminement vers la cible (Figure 22).

Au cours des recherches précliniques, les candidats médicaments vont être testés *in vitro* sur des modèles cellulaires, puis *in vivo* chez l'animal afin de déterminer les paramètres pharmaco-toxicocinétiques et les constantes d'utilisations pharmacologiques et toxicologiques. Ces connaissances vont permettre de connaître le devenir d'une molécule dans un organisme vivant et d'estimer l'efficacité du produit potentiel dans des systèmes physiologiques allant de l'échelle cellulaire à l'échelle animale. L'étude de l'efficacité comprend des études d'effet-dose et de cinétique mais également des études sur la cinétique d'élimination du produit. Ensuite ces approches permettent également de faire des études toxicologiques du médicament candidat afin d'évaluer les effets secondaires éventuels (génotoxicité, tératogénicité, cardiotoxicité, hépatotoxicité). Parmi les animaux utilisés au cours de ces tests précliniques, il y a les modèles rongeurs, le chien, le porc (proximité biologique avec l'homme) et les primates selon les organes cibles ou le système ciblé par la molécule candidate (Figure 22). Les recherches précliniques mènent à la recherche clinique organisée en quatre phases : les phases I à IV. C'est au cours de ces différentes phases que l'intérêt thérapeutique, la dose optimale d'utilisation et le rapport bénéfice/risque du médicament potentiel seront déterminés. Ainsi le médicament potentiel peut faire l'objet d'une autorisation de mise sur le marché (AMM) qui est un accord obligatoire pour la commercialisation d'un médicament. Suite à l'AMM, la pharmacovigilance permet de détecter des effets indésirables très rares (Figure 22).

De nombreux composés ont des effets significatifs sur des modèles cellulaires et ou animaux mais ne seront jamais des médicaments. En effet, les candidats médicaments doivent être de puissants modulateurs de leurs cibles et ne doivent pas moduler d'autres fonctions physiologiques que celles ciblées. De plus, les molécules candidates ne doivent pas avoir de toxicité dans les différents environnements testés. De plus, les candidats médicaments doivent être faciles à administrer et doivent être assez stables dans le temps pour atteindre leur cible en concentration suffisante pour être efficaces.

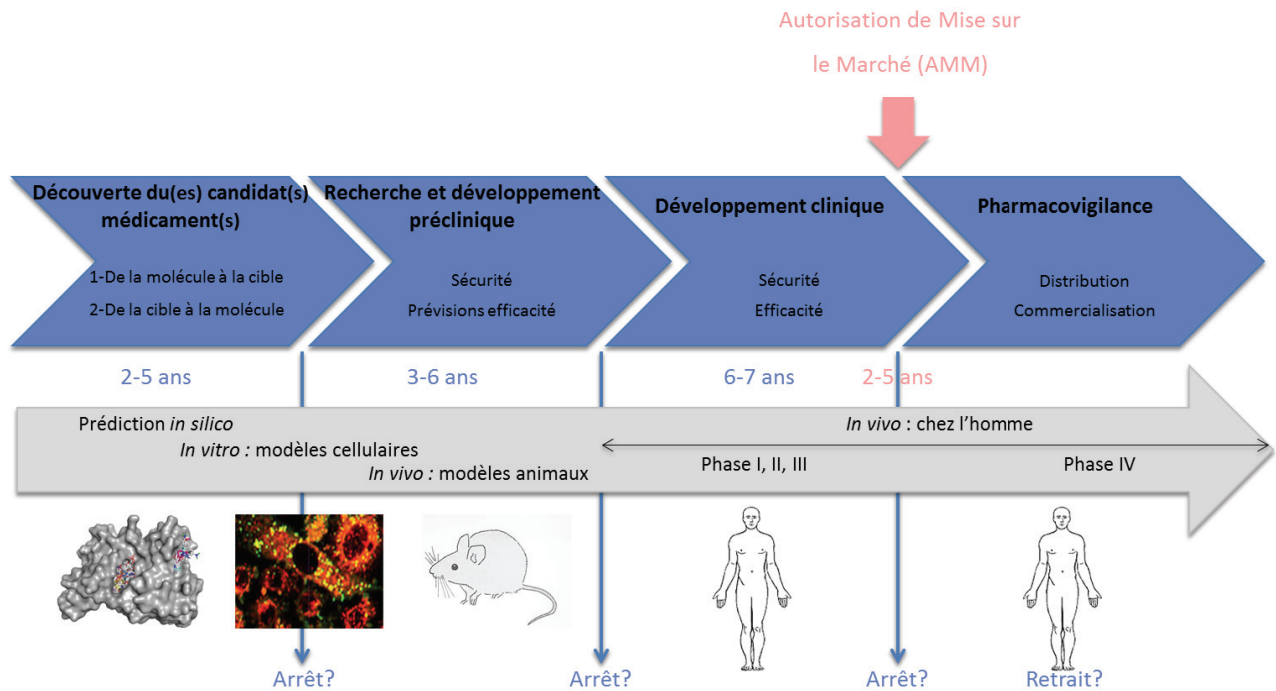


Figure 22 : développement d'un médicament.

4.2 L'action des médicaments

Lorsqu'une molécule entre dans l'organisme, elle est confrontée à différents obstacles et son action suit trois phases : une phase de biopharmaceutique, qui correspond à la mise à disposition du principe actif dans l'organisme, une seconde phase de pharmacocinétique, qui correspond au devenir du médicament dans l'organisme, et enfin une phase de pharmacodynamie qui correspond aux effets du médicament sur l'organisme. La première phase de biopharmaceutique concerne les formes orales solides et comprend la libération du principe actif dans l'organisme et la dissolution du principe actif (solubilisation ou passage à travers les membranes selon la cible). La seconde phase, la pharmacocinétique, est en lien avec son absorption, sa distribution, son métabolisme et son excrétion. Ces quatre notions de pharmacocinétique sont regroupées sous le terme de propriétés ADME (Absorption, Distribution, Métabolisme, Excrétion). L'absorption est le passage du site

d'administration à la circulation générale ou directement au site d'action. La distribution est la distribution du principe actif dans les différents tissus de l'organisme. Le métabolisme est la transformation chimique du principe actif dans l'organisme. L'élimination est l'excrétion ou la métabolisation du principe actif. Et enfin, la pharmacodynamie est l'étude des effets biochimiques et physiologiques du médicament sur l'organisme, aussi bien des effets thérapeutiques que des effets indésirables (Figure 23).

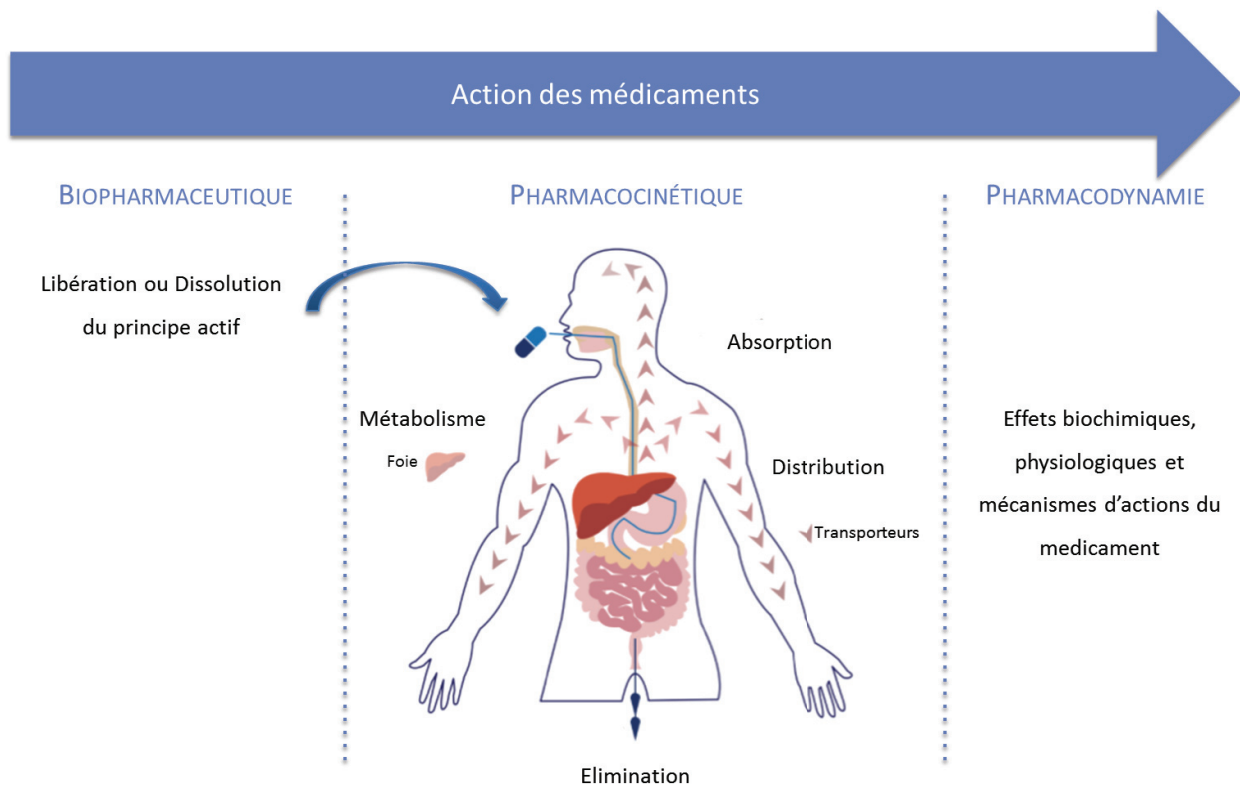


Figure 23 : action des médicaments en trois phases.

4.3 Apport de l'*in silico* dans la découverte des médicaments

Les champs d'applications *in silico* concernent le développement des outils informatiques pour le traitement de systèmes très variés. Ils concernent la biologie, la chimie, les polymères, les matériaux, les nanotechnologies et la découverte d'agents thérapeutiques. Ce champ est basé sur les lois de la physique et de la chimie qui elles-mêmes reposent sur les lois mathématiques. Les approches *in silico* permettent donc de simuler ou de modéliser un phénomène biologique à l'aide d'outils informatiques. Il existe deux grandes familles de méthodes *in silico*.

La première est basée sur les relations « structure-activité » plus communément appelées QSAR (Quantitative Structure-Activity Relationship). Celle-ci repose sur une relation mathématique couplée à des méthodes d'analyse de données reliant des propriétés moléculaires à un effet expérimental pour

une série de composés chimiques similaires. La relation QSAR peut être utilisée pour la prédiction de propriétés de molécules candidates et donnent des informations sur la nature des interactions entre ligands et récepteurs. En effet dans la découverte d'agents thérapeutiques, les approches *in silico* concernent l'optimisation, le développement et la biodisponibilité et offrent une alternative non coûteuse et rapide aux expérimentations biologiques (Nendza et al., 2013). Cette approche permet d'affiner le choix des molécules candidates pour sélectionner les molécules les plus spécifiques et les moins toxiques. Ces outils, comme par exemple le logiciel ADC/Percepta (Advanced Chemistry Development), permettent de prédire les effets biologiques potentiels et les propriétés ADMET des molécules candidates (Administration, distribution, métabolisme, élimination, toxicité) en se basant sur les relations « structure-activité ». Ces méthodes sont en effet utilisées comme des outils prédictifs, mais plus récemment elles ont été utilisées comme des outils confirmatifs, pour évaluer les propriétés ADMET de médicaments putatifs au cours des premières étapes de la recherche et de développement. Outre les avantages économiques de cette approche, celle-ci permet l'identification de « hits » et leur optimisation structurelle afin d'augmenter leur biodisponibilité, leur activité sans ignorer leur sécurité (Tao et al., 2015; Y. Wang et al., 2015).

Cette première approche ne permet toutefois pas de quantifier la force d'une liaison ni les atomes impliqués dans cette liaison contrairement à la seconde grande famille de méthodes *in silico*. Celle-ci, est basée sur les propriétés atomistiques de l'entité ou du système étudié. En effet, les méthodes de modélisation à l'échelle atomique reposent sur des approches de mécanique quantiques et moléculaires (cf. chapitre 2). Les méthodes quantiques sont capables de calculer les propriétés électroniques des systèmes. Les méthodes moléculaires permettent de déterminer la structure tridimensionnelle, l'énergie ainsi que les propriétés physico-chimiques des molécules et des systèmes biologiques à partir de leur composition atomique. Ces approches de modélisation moléculaires sont à la base des études d'amarrage moléculaire et de criblage *in silico*. Les études d'amarrage moléculaire consistent à rechercher les positions et les orientations les plus probables de molécules chimiques en interaction avec une cible et à déterminer l'énergie d'interaction entre la molécule chimique et la cible. Le criblage *in silico* permet de réaliser rapidement et à moindre coût des prédictions de l'activité des molécules chimiques. Le criblage virtuel consiste à cibler *in silico* une bibliothèque de molécules chimiques afin d'éliminer les composés supposés inactifs ou les molécules indésirables tout en priorisant les composés les plus susceptibles d'être actifs.

L'explosion du nombre de publications recensées au cours des vingt dernières années sur ces différentes approches *in silico* montre le réel intérêt de ces stratégies. Ces publications s'intéressent à ces méthodes, parce qu'elles facilitent la recherche et le développement dans la conception de

nouveaux médicaments pour évaluer leur toxicité, mais également aux techniques permettant d'évaluer leur fiabilité (Ramasamy & Selvam, 2015).

4.4 Perspectives thérapeutiques

Aujourd'hui, la biologie des cancers ne peut pas être comprise en énumérant simplement les caractéristiques intrinsèques des cellules cancéreuses compte tenu de l'importance des interactions entre les cellules précancéreuses et cancéreuses et leur microenvironnement au cours des différentes étapes de la tumorigenèse. A la différence des chimiothérapies classiques qui ciblent les cellules en division (cancéreuses ou non), les thérapies ciblées bloquent les mécanismes spécifiques des cellules cancéreuses et exploitent ainsi les différences biologiques et moléculaires existant entre les cellules cancéreuses et les cellules saines de l'organisme. Pour exemple, des médicaments interférant avec chacune des capacités acquises nécessaires à la croissance et la progression tumorale, sont représentés et utilisés en clinique dans le traitement de certaines formes de cancer humain (Figure 24). La thérapie ciblée est possible aujourd'hui en raison des nombreuses évolutions notamment en termes de génétique, de protéomique et d'outils de détections qui permettent de préciser les caractéristiques biologiques et moléculaires de chaque tumeur. En effet, l'identification de biomarqueurs permettra de guider l'accès à une thérapie ciblée. Les biomarqueurs sont des molécules ou des protéines surexprimées ou anormalement absentes dans certains types de tumeurs. Ces marqueurs peuvent être prédictifs d'une réponse positive à une thérapie ciblée ou encore de l'évolution d'une tumeur et/ou de sa sévérité. Les biomarqueurs permettent d'établir un diagnostic précis, de choisir un traitement ciblé quand il est disponible, de suivre l'efficacité d'un traitement ou au contraire la résistance à une chimiothérapie, d'établir un pronostic sur l'évolution d'une tumeur, d'évaluer le risque d'apparition de métastases.

Ces stratégies ciblées conduisent à une médecine personnalisée en traitant chaque patient de manière individuelle en fonction des spécificités génétiques et biologiques de sa tumeur afin d'améliorer la performance des soins, d'éviter les traitements inutiles et d'améliorer la qualité de vie des patients. A ce jour, plus de 80% des autorisations de mise sur le marché concernent des thérapies ciblées.

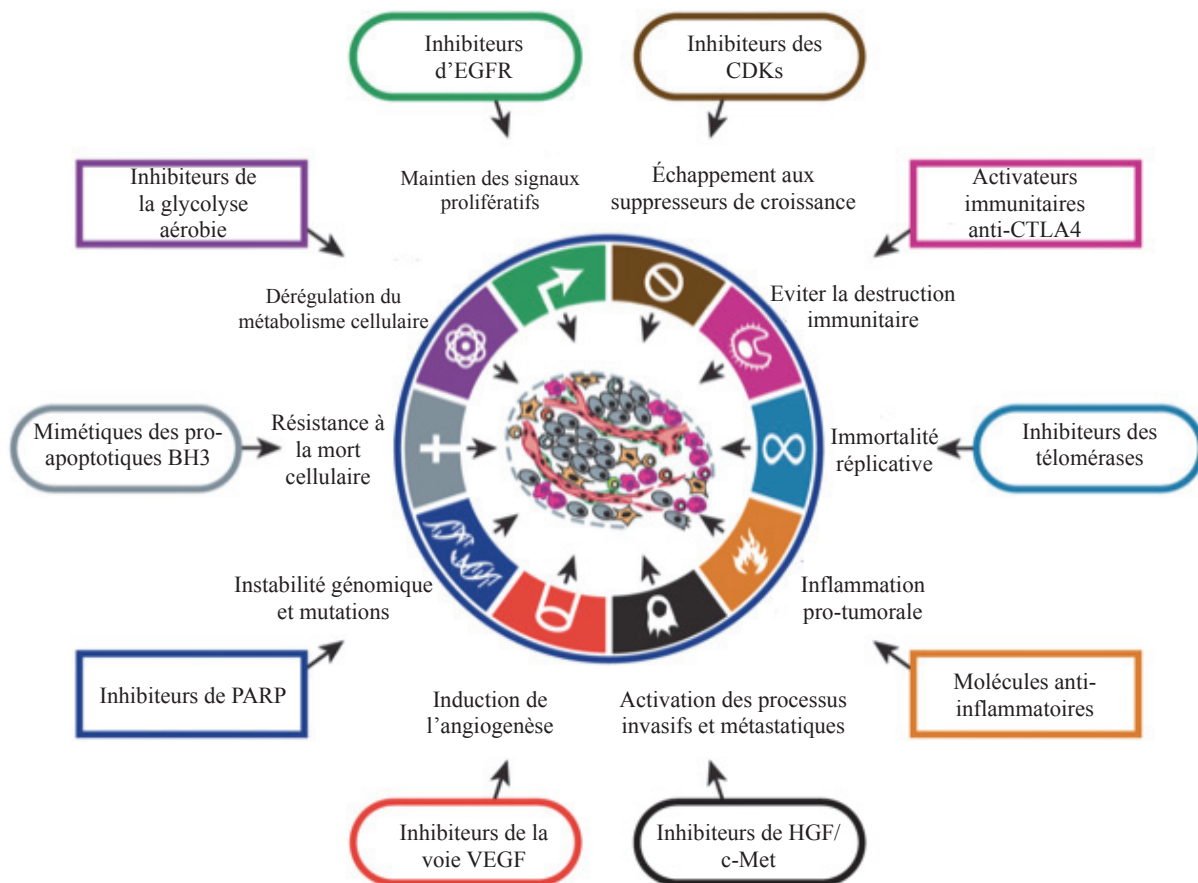


Figure 24: exemple de thérapies ciblées des caractéristiques clés des cellules cancéreuses.

D'après Hanahan and Weinberg, 2011 (Hanahan & Weinberg, 2011).

CHAPITRE 2 : MODELISATION DE STRUCTURES TRIDIMENSIONNELLES DE PROTEINES

La cristallographie aux rayons X, la résonance magnétique nucléaire et la cryo-microscopie électronique sont trois méthodes permettant de déterminer par une approche expérimentale la structure tridimensionnelle de macromolécules biologiques. Ces structures 3D expérimentales sont contenues dans des banques de données Protein Data Bank (PDB) (Berman et al., 2000; PDB). Ces fichiers PDB ne contiennent pas d'informations concernant la connexion entre les atomes et généralement ne contiennent pas les coordonnées des atomes d'hydrogènes. Ainsi, la « connectivité » entre les atomes doit être calculée avant de visualiser les structures. Le RMSD (Root Mean Square Deviation) permet de comparer la ressemblance structurale entre deux protéines et correspond à l'écart quadratique moyen des distances interatomiques entre les atomes de deux structures.

1 Généralité sur la modélisation moléculaire

Lorsque la structure tridimensionnelle de ces macromolécules biologiques n'a pas été déterminée de manière expérimentale, des approches de bio-informatique structurale permettent de la construire *in silico*. Un des objectifs de la bio-informatique structurale est la modélisation moléculaire. Cette dernière regroupe un ensemble de méthodes permettant de simuler le comportement d'un système allant de la molécule diatomique aux macromolécules biologiques. Ainsi, ces approches permettent de décrire le comportement électronique des atomes et des molécules afin de mieux comprendre leur réactivité, l'importance de certains acides aminés et le processus de repliement d'une protéine. Les approches de modélisation moléculaire reposent sur l'énergie d'interaction des atomes du système moléculaire en fonction de leurs positions relatives et reposent ainsi sur les lois de la mécanique quantique et de la mécanique moléculaire.

Les méthodes quantiques dont les calculs sont basés sur l'équation de Schrödinger, distinguent les noyaux et les électrons dont les interactions déterminent la distribution spatiale et l'énergie associée. Cette méthode est particulièrement bien adaptée pour les petites molécules avec un nombre d'atomes inférieur à 300.

Les méthodes moléculaires sont des méthodes « tout atomes ». Les atomes sont alors représentés sous forme de sphères chargées. Les interactions sont considérées comme des « ressorts » et modélisées par des oscillations autour d'une position d'équilibre définie par un référentiel : le champ de force. Les interactions déterminent la distribution spatiale des particules atomiques et leurs énergies. Toutefois, cette approche de l'étude du mouvement des noyaux ne prend pas en compte directement celui des électrons mais le prend en compte indirectement comme une « perturbation » au mouvement du noyau. Ainsi cette approche permet de calculer la conformation de la molécule sans donner aucune information directe sur la configuration électronique.

2 Différentes énergies en mécanique moléculaire

La mécanique moléculaire a pour but de calculer l'énergie potentielle (énergie associée à une ou plusieurs molécules d'un système biologique) d'une molécule ou d'un système de molécules en fonction des coordonnées des atomes. La définition de cette énergie potentielle dépend du champ de force appliqué sur lequel nous reviendrons ultérieurement. Elle est représentée par la somme des contributions internes, entre atomes liés chimiquement, et externes, entre atomes non liés lorsque le champ de force associé est CHARMM (Chemistry at HARvard Molecular Mechanics ; Figure 25).

Les interactions entre atomes liés représentent l'énergie covalente du système et mettent en évidence la contribution énergétique de la déformation de la structure chimique en maintenant les centres de chiralité de la molécule. Le potentiel entre atomes liés comprend (Figure 46):

- les **liaisons covalentes** E_{Liaison} qui évaluent l'énergie d'élongation d'une liaison chimique :
 - K_r est analogue à une constante de force autrement nommée constante de raideur du ressort en kcal/Å² (kilocalorie par angström carré);
 - r correspond à la distance interatomique mesurée ;
 - r_0 correspond à la distance « idéale » définie comme l'équilibre selon le champ de force.
- les **angles de valence** E_{Valence} qui évaluent l'énergie de déformation angulaire entre deux liaisons chimiques :
 - K_θ est la constante de force angulaire en kcal/rad² ;
 - θ correspond à l'angle de valence mesuré ;
 - θ_0 correspond à l'angle de valence « idéal » définie comme l'équilibre selon le champ de force.
- les **angles dièdres** $E_{\text{Dièdres}}$ qui représentent l'énergie de torsion des liaisons chimiques:
 - K_ϕ est la constante de force angulaire en kcal ;
 - n est la périodicité ;
 - δ est l'angle de rotation possible parmi N possibilités ;
 - ϕ est l'angle dièdre mesuré.
- l'**énergie impropre** $E_{\text{DièdreImpropre}}$ qui permet, par exemple, aux atomes d'un cycle de rester dans le même plan :
 - K_ω est la constante de planarité angulaire en kcal/rad² ;
 - ω est l'angle de planarité
 - ω_0 est l'angle de planarité « idéal » définie comme l'équilibre selon le champ de force.

Le potentiel entre atomes non liés considère les atomes de deux molécules distinctes ou alors deux atomes appartenant à une même molécule mais séparés par au moins trois liaisons. Ce second potentiel permet d'évaluer la contribution non covalente de l'énergie du système. Le potentiel entre atomes non liés comprend (Figure 25):

- les **interactions électrostatiques** E_{Elec} qui décrivent les interactions entre particules chargées :
 - o ij une paire d'atomes ;
 - o q_i et q_j représentent les charges ponctuelles de i et j ;
 - o ϵ_0 correspond à la permittivité du vide ;
 - o ϵ_r correspond à une constante diélectrique ;
 - o r_{ij} correspond à la distance entre les charges.
- les **interactions de van der Waals** E_{VdW} qui correspondent aux forces dites de dispersion :
 - o ϵ_{ij} correspond à une constante diélectrique ;
 - o le potentiel « r^{-12} » correspond au terme répulsif et représente le principe d'exclusion de Pauli et la répulsion internucléaire à courte distance ;
 - o le potentiel « r^{-6} » correspond au terme attractif et représente les forces de London.

Comme nous l'avons dit précédemment la définition de l'énergie potentielle dépend du choix du champ de force. Ainsi, selon le champ de force, l'énergie potentielle ne considère pas les mêmes contributions internes. L'énergie potentielle associée au champ de force AMBER (Assisted Model Building and Energy Refinement) par exemple, ne considère pas l'énergie impropre ($E_{DièdreImpropre}$) qui permet par exemple aux atomes d'un cycle de rester dans le même plan mais en revanche considère un potentiel permettant de mesurer la participation des liaisons hydrogènes du système.

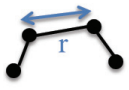

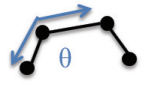
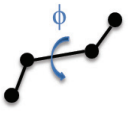
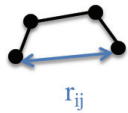

$E = E_{Liée} + E_{Non Liée}$	
$E_{Liée} = E_{Liaison} + E_{Valence} + E_{Dièdres} + E_{DièdresImpropre}$	$E_{Non Liée} = E_{Elec} + E_{VdW}$
$E_{liaison} = \sum_{liaisons} K_r (r - r_0)^2$ 	$E_{Elec} = \sum_{\substack{ij \text{ atomes} \\ \text{non liés}}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}$ 
$E_{valence} = \sum_{angles} K_\theta (\theta - \theta_0)^2$ 	
$E_{dièdres} = \sum_{dièdres} K_\phi^n (1 + \cos(n\Phi - \delta))$ 	$E_{VdW} = \sum_{\substack{ij \text{ atomes} \\ \text{non liés}}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right]$ 
$E_{dièdresImpropre} = K_\omega (\omega - \omega_0)^2$ 	

Figure 25: illustrations des potentiels impliqués dans l'énergie potentielle.

3 Notion de champ de force

Le champ de force est l'outil de base de la mécanique moléculaire. Il correspond à l'ensemble des paramètres décrivant un atome et permettant de résoudre l'expression analytique de l'énergie potentielle du système. Autrement dit, le champ de force correspond à un modèle mathématique représentant, aussi bien que possible, les variations de l'énergie potentielle tout en tenant compte de la géométrie moléculaire. Ainsi plusieurs champs de force existent selon le domaine d'application :

- AMBER : adapté aux protéines et acides nucléiques (Weiner, 1981),
- CHARMM (Chemistry at HARvard Molecular Mechanics) : adapté aux protéines, aux acides nucléiques et aux lipides (Brooks, 1983),
- MMFF94 (Merck Molecular Force Field) : champ de force spécialisé dans le traitement des petites molécules organiques (Halgren & BNachbar, 1996),
- MM2, MM3 (Allinger Molecular Mechanics) : adapté aux petites molécules organiques (N., 1977) (Norman L. Allinger, Fanbing Li, Liqun Yan, & Tai, 1990).

Les champs de force couramment utilisés pour la modélisation de biomolécules biologiques sont AMBER et CHARMM. La paramétrisation des champs de force est faite à partir de données expérimentales (Résonance Magnétique Nucléaire (RMN), RX) sur le plus grand nombre possible de molécules.

4 Méthodes de modélisation moléculaire

Comme nous l'avons vu la modélisation moléculaire est donc un des objectifs de la bio-informatique structurale qui consistent à prédire la structure tridimensionnelle de protéines. Il existe quatre grandes classes de méthodes, qui sont mises en œuvre selon les taux d'identité de séquences obtenus, que sont la méthode d'enfilage des repliements, les alphabets structuraux, la méthode *ab initio* et la modélisation par homologie. De manière générale, les modèles obtenus sont toujours faux et il convient de les confronter à l'expérience afin de les valider.

Méthodes	Enfilage des repliements	Alphabets structuraux	Ab initio	Modélisation par homologie
Taux identités de séquences	30%	30%	<15%	>30%

Figure 26 : relation entre le taux d'identité de séquence et les méthodes de mécanique moléculaire utilisées.

4.1 Méthode d'enfilage des repliements

Cette méthode s'applique lorsque la séquence à modéliser présente une identité de séquence de l'ordre de 30% avec une structure connue. La méthode d'enfilage consiste à « enfiler » une séquence de structure inconnue dans une structure 3D connue et définie dans la PDB. Ensuite, un potentiel empirique permet de vérifier la compatibilité du repliement et de la séquence « enfilée ». Cette approche est souvent utilisée pour exclure les repliements improbables.

4.2 Alphabets structuraux

Le principe de cette seconde approche repose sur la décomposition des structures tridimensionnelles en fragments. Au lieu de considérer la structure entière ou des fragments de structures secondaires (information insuffisante pour la construction d'une structure tridimensionnelle), on considère des « alphabets structuraux » qui correspondent à des éléments structuraux essentiels pour la structure tridimensionnelle. Autrement dit, les alphabets structuraux sont des bibliothèques de petits prototypes qui approximent des parties de structures protéiques. Ainsi cette approche permet l'approximation de la structure tridimensionnelle des protéines avec un pouvoir prédictif plus ou moins fort selon le nombre d'éléments d'alphabets structuraux nécessaires dépendants de la taille de la protéine considérée.

4.3 Méthode ab initio

Cette méthode est particulièrement bien adaptée et fiable pour les molécules de petites tailles lorsque les identités de séquences sont inférieures à 15%. Elle permet de donner une structure tridimensionnelle à une protéine à partir de l'information contenue dans sa séquence en se basant sur des calculs d'énergies d'interactions au sein des macromolécules biologiques et en cherchant la structure de plus basse énergie.

4.4 Modélisation par homologie

Cette approche est basée sur le fait que des protéines homologues ont des structures 3D proches puisque les structures tridimensionnelles sont mieux conservées que les séquences d'une espèce à l'autre (Chothia et Lesk, 1986). Ainsi, si la structure 3D expérimentale d'une protéine homologue est définie alors elle servira d'empreinte pour modéliser la structure 3D proche de l'empreinte en incluant la séquence d'intérêt de la protéine à modéliser. Cette stratégie nécessite un taux d'identité de séquence au moins supérieur ou égale à 30% entre la séquence de la protéine à modéliser et la séquence de la protéine « empreinte ». Cette modélisation est celle qui permet l'obtention de la structure tridimensionnelle ayant la meilleure résolution après les approches expérimentales. Les régions structurellement conservées sont directement modélisées à partir de l'empreinte alors que les zones d'insertions/de délétions et les boucles, qui sont des zones très flexibles restent difficiles à modéliser par cette approche.

Cette approche nécessite plusieurs étapes : le choix de l'empreinte, l'alignement des séquences et des structures secondaires, et enfin la construction qui consiste en le remplacement de la séquence de l'empreinte par la séquence de la protéine dont la structure 3D est à déterminer. La qualité de la structure dépend donc de la qualité et de la résolution de la structure dite « modèle » mais également de la qualité de l'alignement fourni.

4.4.1 Choix du support

La première étape consiste donc en la détermination de la structure tridimensionnelle qui servira d'empreinte et donc de modèle pour la construction de la structure tridimensionnelle de la protéine d'intérêt. Lorsque deux séquences de longueur comparable présentent plus de 30% d'identité répartie sur plus de 70% de la longueur de la séquence la plus longue, alors la relation d'homologie est dite sans risques. Ensuite, il peut ne pas exister de protéine homologue dont la structure est résolue ce qui rend la modélisation par homologie impossible. Enfin lorsque le taux d'identité de séquence est plus faible, on parle d'homologie distante. Dans ce cas, une approche alternative à la transitivité d'homologie basée sur la comparaison des structures secondaires permet également d'optimiser le choix du support malgré les faibles taux d'identité de séquences. Le SOV (*Structural Overlap*) permet

de mesurer l'accord entre un élément de structure prédit et observé et est ainsi un bon indicateur pour la ressemblance des éléments de structures secondaires entre deux protéines.

4.4.2 Alignement entre la séquence et les empreintes

L'alignement entre la séquence à modéliser et la séquence de l'empreinte est important pour localiser les zones de modifications structurales. Le modélisateur joue un rôle clef puisque l'alignement de séquence optimal n'est pas forcément celui qui est pertinent biologiquement. Ainsi c'est le modélisateur qui doit optimiser au mieux l'alignement en limitant les empêchements stériques qui sont compliqués à résoudre.

4.4.3 Construction du modèle

Il y a plusieurs stratégies pour la construction. La première méthode, rarement utilisée de nos jours, consiste à construire d'abord le cœur de la protéine qui est bien structuré et commun à l'empreinte et la protéine d'intérêt. Les autres méthodes sont des approches substitutives ou globales.

Il existe en effet une approche dite substitutive qui consiste à faire des mutations ponctuelles pour transformer l'empreinte en modèle de la protéine d'intérêt. Cette stratégie est applicable si les deux séquences sont proches et présentent peu de zones de réarrangements structurales (peu d'insertions, de délétions et/ou de réorientations des chaînes latérales des acides aminés). Lorsque les séquences sont un peu plus éloignées, il existe une seconde approche qui consiste à construire d'abord les blocs structurellement très conservés avant de les relier par les boucles modélisées soit *de novo* soit en cherchant dans la PDB les boucles, déjà construites, pouvant s'ajuster au modèle. L'inconvénient de ces méthodes est qu'elles utilisent une seule empreinte alors que l'utilisation de plusieurs empreintes chevauchantes permettrait l'obtention d'un modèle plus grand et plus qualitatif même avec des taux d'identités faibles.

Ainsi une méthode dite globale ou géométrique a été mise en place afin de pouvoir utiliser toutes les empreintes disponibles. Cette approche calcule, pour chaque atome, une distance moyenne entre les différentes empreintes avec un écart-type associé permettant de définir un intervalle de variation possible pour la position des atomes correspondants et communs avec la séquence d'intérêt. Une liste des atomes et des angles communs aux empreintes et à la protéine d'intérêt est établie et sera imposée *via* une modélisation sous contraintes. Enfin, il y a une étape d'optimisation de la géométrie qui dépend du logiciel utilisé.

5 Dynamique moléculaire

La dynamique moléculaire a pour objectif de simuler le comportement d'une molécule au cours du temps dans des conditions bien définies. Pour cette approche de dynamique moléculaire on part d'une conformation déterminée en mécanique moléculaire et on cherche à augmenter l'énergie cinétique du système *via* l'augmentation de la température afin de déterminer l'énergie cinétique d'une molécule ou d'un système moléculaire. La dynamique moléculaire peut être utilisée pour étudier le comportement dynamique d'une molécule au cours du temps (simulation des mouvements intra-moléculaires) ou pour faire de la recherche conformationnelle (optimiser les structures).

5.1 Dynamiques moléculaires d'un point de vue théorique

5.1.1 Equations de mouvements

La dynamique moléculaire permet de calculer la force exercée sur chaque atome et fournit ainsi différentes informations sur leurs trajectoires : vitesse et position des atomes. Cette approche est décrite par les principes de la mécanique classique newtonnienne. La force $F_i(t)$ exercée sur chaque atome i de coordonnées $r_i(t)$ au temps t est déterminée par dérivation de la fonction énergie potentielle E et définie telle que: $F_i(t) = -\frac{\delta E}{\delta r_i(t)}$. En considérant un pas de temps très court, on peut intégrer les équations de mouvement et obtenir une trajectoire de chaque atome en fonction du temps.

5.1.2 Intégrations des équations

Il existe plusieurs méthodes de dynamiques moléculaires mais généralement la dynamique repose sur la résolution des équations de mouvement de Newton pour chacun des atomes constituant le système moléculaire. La résolution de ces équations se fait par des méthodes d'intégrations numériques. Il en existe plusieurs et le choix de la méthode d'intégration est importante pour la qualité des calculs et donc la trajectoire définie *in silico* du système moléculaire. Ce type de méthode a pour but de calculer les différentes coordonnées ainsi que la vitesse des différents atomes à l'instant $t + \Delta t$ à partir de l'instant t . Au-delà du choix de la méthode d'intégration, le choix du pas Δt est important pour la qualité des calculs. En effet, selon le système étudié, le pas d'intégration Δt devra être approprié. D'autre part, l'énergie totale du système doit être conservée durant toute la simulation.

5.2 Dynamiques moléculaires d'un point de vue pratique : déroulement d'une simulation

Une simulation de dynamique moléculaire comprend plusieurs étapes. La première étape consiste en la détermination de l'état initial qui est une conformation optimisée par la mécanique moléculaire. La seconde étape consiste en une phase de mise en température. Au cours de cette étape

le système passe du zéro absolu à la température définie pour la simulation *via* des paliers de température. A chaque palier de température la vitesse et la position des atomes sont recalculées. Il s'en suit ainsi la troisième étape qui est la simulation proprement dite où le mouvement des atomes est recalculé jusqu'à ce que le système atteigne un état d'équilibre. A cet état d'équilibre, l'énergie cinétique est redistribuée en énergie potentielle pour qu'à l'équilibre l'énergie potentielle et l'énergie cinétique soient égales.

La notion de température est fondamentale puisque la vitesse est proportionnelle à la racine carrée de la température. Ainsi les dynamiques moléculaires sont effectuées soit à température constante, soit à volume constant, soit à pression constante.

6 Optimisation des structures tridimensionnelles

D'un point de vue énergétique, la conformation optimale est celle pour laquelle l'énergie est la plus basse. Ainsi les approches de minimisation de l'énergie potentielle permettent d'optimiser la structure des protéines en appréhendant au mieux ces minima énergétiques permettant donc à la molécule, ou à plusieurs molécules, d'atteindre une structure dite « relâchée » optimisant ainsi la géométrie du système. Au-delà de son utilisation au cours de la modalisation moléculaire, la minimisation de l'énergie potentielle peut être utilisée pour les approches expérimentales (RMN par exemple).

Cette approche de minimisation consiste à trouver les coordonnées des atomes qui permettent de minimiser la fonction d'énergie potentielle du système étudié sachant que les positions atomiques du système sont variables. Toutefois, l'importance des degrés de liberté fait fluctuer le système autour d'un ensemble de conformation stable avec différents minima. Cependant, l'état de plus basse énergie, appelé minimum global par opposition aux minima locaux, n'est généralement pas atteint avec les méthodes de minimisation d'énergie classiques, comme *la plus grande pente* par exemple, car elles ne permettent pas le franchissement des barrières d'énergie. De plus un autre inconvénient de cette méthode est que le minimum local obtenu pour deux configurations initiales différentes n'est pas identique.

Dès 1983, les méthodes de recuit simulé ont été développées. Cette méthode est un algorithme basé sur le processus physique de refroidissement d'un liquide ou de cristallisation d'un métal. Cette méthode de recuit simulé peut être réalisée par la méthode de dynamique moléculaire. Cette approche repose sur un « chauffage virtuel » *via* un apport énergétique de la molécule. Cette étape de chauffage permet à la molécule de franchir les barrières énergétiques. Ceci permet donc d'aller au-delà des minima locaux au voisinage de la structure initiale afin d'explorer un nouvel espace conformationnel pour atteindre des minima plus lointains de la structure initiale ayant un état de plus basse énergie (Figure 27).

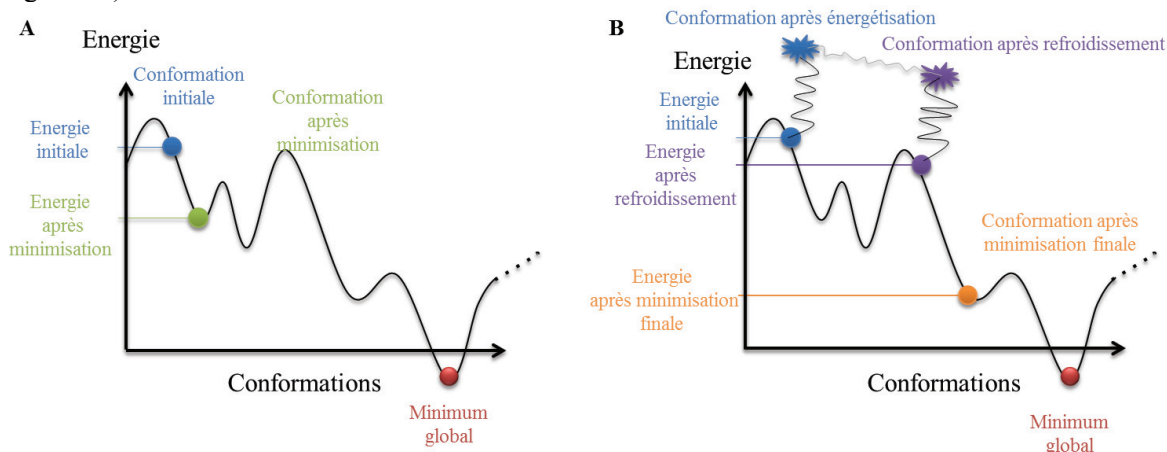


Figure 27 : représentation des minima énergétiques locaux et globaux par une approche de minimisation classique.

CHAPITRE 3 : TWIST1 DANS LES CANCERS

1 TWIST1, facteur de transcription de la famille des bHLH

1.1 La grande famille des bHLH

La grande famille des facteurs de transcriptions bHLH (basic Helix-Loop-Helix) regroupe plus de deux cents membres extrêmement conservés entre les espèces (PFAM : PF00010)(S. Jones, 2004; Murre et al., 1994). Ces facteurs de transcription ont des fonctions cellulaires variées. En effet, certains jouent un rôle dans la régulation de la croissance cellulaire, la différenciation en tant que régulateurs positifs ou négatifs, l'apoptose, la prolifération, la myogenèse, la neurogenèse, la détermination sexuée et la lymphomagenèse (Jan & Jan, 1993; S. Jones, 2004; Puisieux et al., 2006; Teng & Li, 2014). Les membres de cette superfamille ont un motif protéique d'une soixantaine d'acides aminés composé de deux domaines hautement conservés et fonctionnellement distincts (Figure 28 ; Figure 29)(Murre et al., 1994). Le premier domaine est un domaine basique à l'extrémité amino-terminale qui permet aux facteurs de transcription de lier l'ADN au niveau de séquence hexanucléotidique nommé E-box (Enhancer-box)(Ellenberger, Fass, Arnaud, & Harrison, 1994; Murre, McCaw, Vaessin, et al., 1989). Le second domaine est un domaine Hélice-Boucle-Hélice riche en résidus hydrophobes au niveau de l'extrémité carboxy-terminale qui facilite les interactions avec les autres protéines pour former des homodimères et des hétérodimères (Murre, McCaw, Vaessin, et al., 1989). Ainsi différents dimères peuvent se former et se lier avec différentes affinités au niveau des séquences E-box de promoteurs de gènes conduisant donc à une régulation transcriptionnelle (Figure 29B, C et D) (S. Jones, 2004).

A partir de ces caractéristiques, différentes approches de classification ont été mises en œuvre. La plus commune des classifications est celle de Murre qui tient compte de la distribution tissulaire, des capacités de dimérisation et des capacités de liaison à l'ADN (Murre et al., 1994) (Massari & Murre, 2000). Les protéines HLH sont ainsi classées en 7 grandes familles. La classe I regroupe les protéines dites E et inclut E12, E47. Ce sont des protéines HLH, exprimées dans différents tissus, capables de former des homodimères et des hétérodimères liant l'ADN au niveau de E-box. La classe II contient des membres comme MYOD1 (MYOblast Determination protein 1), NeuroD1/BETA2 (Neurogenic Differentiation factor 1), HAND1 (Heart And Neural crest Derivatives expressed 1) et TWIST1 qui ont une expression tissu spécifique. Ils sont capables de former des homodimères mais forment préférentiellement des hétérodimères avec des protéines E. Les protéines ayant un motif Leucine Zipper adjacent au motif HLH appartiennent à la classe III et comprennent, entre autre, des protéines de la famille des facteurs de transcription MYC (Myc proto-oncogene protein), SREBP-1 (Sterol regulatory element-binding protein 1) (Alt et al., 1986; Zhao, Zhao, Zhou, Mattei, & de Crombrughe, 1993). La classe IV est composée des protéines MAD (Mitotic Arrest Deficient protein), MAX (MYC-Associated factor X) qui peuvent se dimériser avec MYC ou avec d'autres

protéines (Ayer, Kretzner, & Eisenman, 1993; Blackwood & Eisenman, 1991). Les protéines HLH sans domaine basique appartiennent à la classe V (Benezra et al., 1990). Ce sont des régulateurs négatifs des bHLH de classe I et II. La protéine Id (DNA-binding protein inhibitor ID) appartient à cette classe de par l'absence de domaine basique (Benezra et al., 1990). Les protéines HLH de la classe VI sont définies par une proline caractéristique au niveau de leur domaine basique. La protéine de drosophile hairy appartient par exemple à cette classe VI (Rushlow et al., 1989). La dernière classe VII regroupe des protéines avec un domaine bHLH-PAS avec, par exemple, le récepteur aux hydrocarbures aromatiques (AHR) ou encore HIF-1 α (Hypoxia-inducible factor 1-alpha) (Figure 28B)(Crews, 1998).

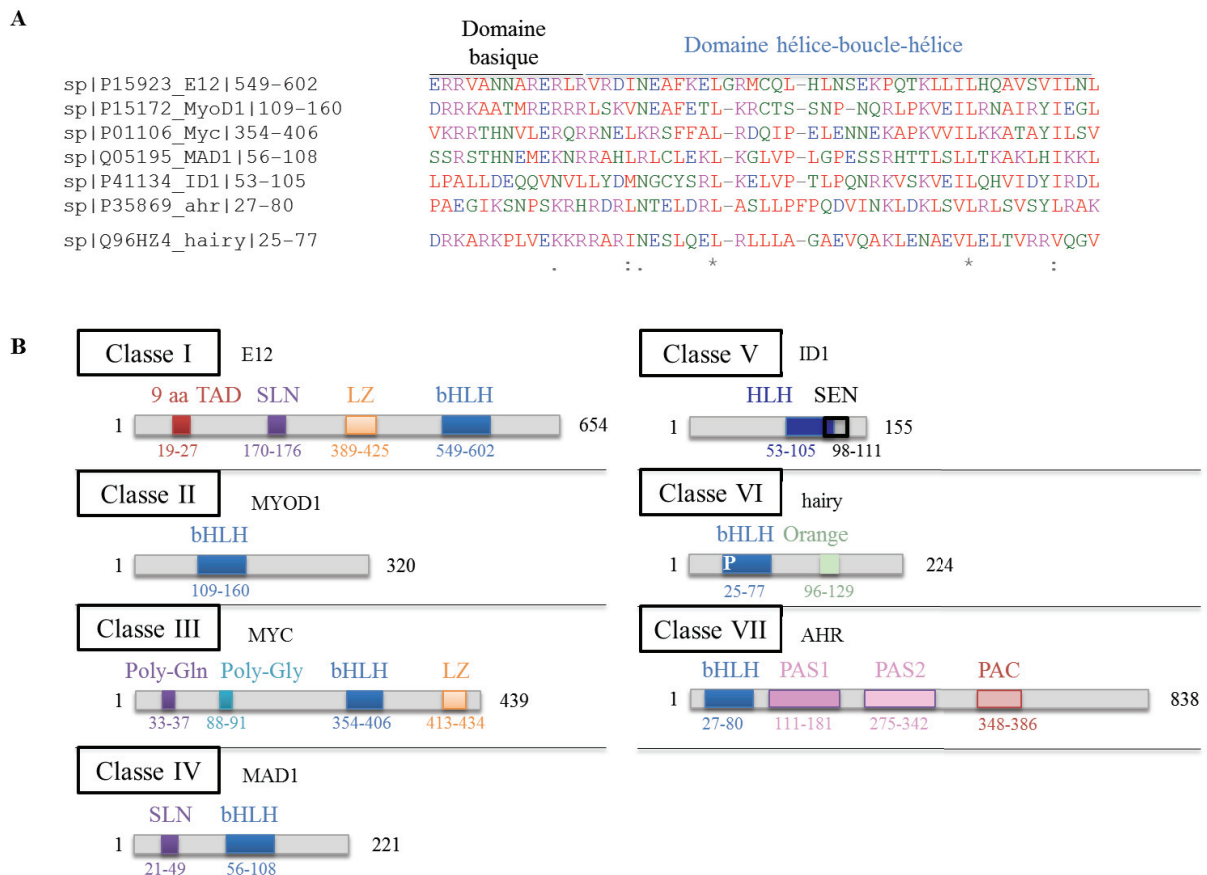


Figure 28 : classification structurelle des protéines bHLH.

A: Alignement des séquences bHLH (*Muscle Software*). B: Classification des protéines bHLH (D'après Murre and al., 1994) (Murre et al., 1994).

Les protéines bHLH ont la capacité de se dimeriser *via* les hélices α -amphipatiques du domaine HLH afin d'exercer leur activité transcriptionnelle. La nature des complexes formés dépend des concentrations relatives des différents partenaires et de leurs affinités réciproques. Les protéines de la classe II ne sont pas stables sous forme homodimérique et forment préférentiellement des hétérodimères en s'associant avec des protéines de la classe I autrement nommées protéines E (Connerney et al., 2006; Vinals et al., 2004; Vinals & Ventura, 2004). Les protéines de la classe V

vont se dimeriser préférentiellement avec des protéines E afin de modifier l'équilibre des hétérodimères associant les protéines de la classe I et de la classe II (Connerney et al., 2006; Massari & Murre, 2000). Au-delà des protéines Id, des régulations post-traductionnelles semblent modifier la balance homodimère/hétérodimère et donc guider la formation de certains complexes, critique pour la fonction biologique (Figure 29A)(B. A. Firulli et al., 2005; Vichalkovski, Gresko, Hess, Restuccia, & Hemmings, 2010). Ainsi, selon l'environnement, certains dimères se formeront préférentiellement et auront, par conséquence, des fonctions spécifiques transcriptionnelles activatrices ou répressives au niveau de gènes cibles qui leur sont propres (Barnes & Firulli, 2009; Chen & Behringer, 1995).

La dimérisation des facteurs de transcription de la famille des bHLH est un prérequis pour la fixation de ces dimères à des éléments de réponse nommés E-box de l'ADN, *via* leur domaine basique (Figure 29B) (De Masi et al., 2011; B. A. Firulli, Redick, Conway, & Firulli, 2007; Massari & Murre, 2000). Les séquences E-box sont des séquences hexanucléotidiques palindromiques ou semi-palindromiques, de type 5'-CANNTG-3', constituées de 4 bases externes consensus et de deux bases centrales variables (Figure 29C et D). Les dimères permettront ainsi l'activation ou la répression de la transcription de gènes cibles spécifiques (Barnes & Firulli, 2009; Chen & Behringer, 1995). Cette caractéristique de liaison à l'ADN a permis la mise en place d'une classification des protéines bHLH en six groupes phylogénétiques de A à F selon les séquences E-box auxquelles ces protéines se fixent (Atchley & Fitch, 1997; S. Jones, 2004; Ledent, Paquet, & Vervoort, 2002). Lorsqu'un dimère lie une séquence E-box, chaque monomère lie un « demi-site 5'-CAN-3' » avec une certaine affinité (Figure 29D). Ainsi chaque dimère aura une affinité différente pour les E-box selon la nature des monomères qui les constituent.

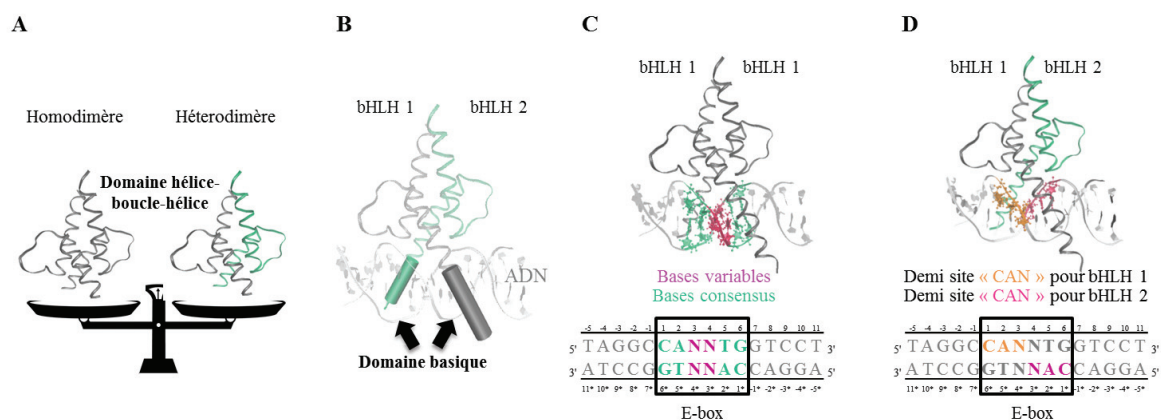


Figure 29 : deux domaines clés des facteurs de transcription bHLH.

A: Dimérisation *via* le domaine hélice-boucle-hélice (HLH) et balance homodimère (gris) / hétérodimère (gris et vert) selon le contexte cellulaire (niveau d'expression des facteurs bHLH et modifications post-traductionnelles). **B:** Liaison à l'ADN des dimères bHLH *via* leur domaine basique (cylindre gris et vert) suite à la dimérisation *via* les domaines HLH. **C:** Exemple de liaison d'un homodimère sur une séquence E-box avec deux bases centrales variables (magenta) et quatre bases externes variables (vert). **D:** Exemple de liaison d'un hétérodimère sur deux demi sites « CAN » respectivement orange et rose pour bHLH1 et bHLH2.

1.2 TWIST, facteur de transcription bHLH

1.2.1 Du gène à la protéine

En général tous les gènes *TWIST* ont une organisation génomique conservée (Gitelman, 2007). Les gènes *TWIST* sont constitués de deux exons. Le premier exon contient l'intégralité de la région codante flanquée par la région 5'UTR et une petite partie, de 35 paires de bases, de la région 3'UTR. Le reste de la région 3'UTR se situe dans le second exon (Gitelman, 2007). Entre ces deux exons, il y a un intron central de taille variable selon les espèces. La taille varie entre 400 paires et 600 paires de base pour le gène *TWIST1* alors qu'il est 5 à 150 fois plus long pour le gène *TWIST2* (Gitelman, 2007; M. S. Lee, Lowe, Flanagan, Kuchler, & Glackin, 2000).

La protéine TWIST1 humaine est codée par un gène *TWIST* humain localisé sur le bras court du chromosome 7 (7p21.1 pour ensembl et 7p21.2 pour NCBI gene ; références des séquences *via* NCBI : NM_000474.3 et NP_000465.1) (Bourgeois, Stoetzel, Bolcato-Bellemin, Mattei, & Perrin-Schmitt, 1996). Le premier exon de ce gène est composé de 609 paires de bases chez l'homme et contient l'intégralité de la séquence codante de la protéine TWIST1 de 202 acides aminés pour environ 21 kDa. Il s'en suit une partie non traduite de 45 paires de bases, un intron de 536 paires de bases, et un second exon non traduit.

La protéine TWIST2 humaine est codée par un gène localisé sur le bras long du chromosome 2 (2q37.3 ; références des séquences *via* NCBI : NM_001271893.3 et NP_001258822.1) (Perrin-Schmitt, Bolcato-Bellemin, Bourgeois, Stoetzel, & Danse, 1997). L'organisation de ce dernier est un peu différente puisque le gène *TWIST2* humain est une exception qui comporte 4 exons et 3 introns. Le premier exon de 650 paires de base contient l'intégralité de la séquence codante de la protéine TWIST2 de 160 acides aminés pour environ 18 kDa (M. S. Lee et al., 2000). Les 3 derniers exons, de 263, 474 et 699 paires de bases respectivement, peuvent chacun être utilisés comme deuxième exon. Ainsi, il existe trois transcrits de cette protéine qui sont différents au niveau de la région 3'UTR (Gitelman, 2007). Les 3 introns sont constitués respectivement de 18 021, 18 782 et 16 483 paires de bases.

1.2.2 TWIST1 au cours de l'évolution

Les protéines TWIST1 et TWIST2 sont des protéines structurellement conservées au cours de l'évolution (Germanguz, Lev, Waisman, Kim, & Gitelman, 2007). Nous allons nous focaliser sur TWIST1 dans cette partie puisque les séquences de cette protéine sont déterminées pour un plus grand nombre d'espèces que ne le sont celles de TWIST2. Les protéines TWIST1 présentent un taux d'identité de séquence supérieur à 40% si on considère la totalité des séquences des protéines TWIST1 de différentes espèces. Les structures des domaines bHLH sont extrêmement bien conservées au cours

de l'évolution puisque ces domaines caractéristiques des facteurs de transcription de la famille des bHLH présentent un taux d'identité de séquence entre 80 et 100% (Figure 30) (Castanon & Baylies, 2002; Thisse, el Messal, & Perrin-Schmitt, 1987).

Organisme/espèce	Code Uniprot	Taux d'identité de séquences		Références bibliographiques
		Séquence entière	Séquence HLH	
Humain <i>Homo sapiens</i>	Q15672	100%	100%	Wang et al., 1997
Souris <i>Mus musculus</i>	P26687	96,53%	100%	Wolf et al., 1991
Rat <i>Rattus norvegicus</i>	Q9EPJ1	96,53%	100%	Bloch-Zupan et al., 2001
Poulet <i>Gallus gallus</i>	Q9PVG4	94,18%	100%	Tavares et al., 2001
Grenouille <i>Xenopus laevis</i>	P13903	84,34%	92,31%	Hopwood et al., 1989
Poisson zèbre <i>Danio rerio</i>	Q9PTE3	80,12%	98,08%	Morin-Kensicki and Eisen, 1997
Lancelet <i>Branchiostoma belcheri</i>	O96642	67,65%	92,31%	Yasui et al., 1998
Méduse <i>Podocoryna carnea</i>	Q9GNV2	50,34%	80,77%	Spring et al., 2000
Nématode <i>Caenorhabditis elegans</i>	Q11094	43,64%	64,71%	Harfe and Fire., 1998
Drosophile <i>Drosophila melanogaster</i>	P10627	42,29%	86,54%	Thisse et al., 1987 ; Leptin et al., 1992
Sangsue <i>Helobdella robusta</i>	Q95VQ5	40,91%	88,46%	Soto et al., 1997

Figure 30 : alignement de protéines TWIST1 issues de différentes espèces avec Clustal Omega (Sievers & Higgins, 2014).

1.2.3 Structure des protéines TWIST1 et TWIST2

Elles sont constituées d'un domaine HLH de dimérisation strictement identique (en position 121-164 pour TWIST1 et en position 79-122 pour TWIST2) qui permet à ces protéines TWIST de se dimeriser avec un second facteur HLH. L'importance de ce domaine fonctionnel a été mise en évidence suite à la découverte de mutations ponctuelles localisées au niveau de ce domaine bHLH et associées au syndrome de Saethre-Chotzen qui se caractérise par une craniosténose et une malformation des membres (el Ghouzzi et al., 1997; Howard et al., 1997). Nous reviendrons sur les caractéristiques de cette maladie génétique ultérieurement. Ces protéines sont également constituées d'un domaine basique de liaison à l'ADN (en position 109-120 pour TWIST1 et en position 67-78 pour TWIST2). Le taux d'identité de séquence entre le domaine basique de TWIST1 et celui de TWIST2 est de 83,33%. La dimérisation de deux facteurs de transcription bHLH, *via* le domaine HLH, crée un domaine de liaison à l'ADN en rapprochant les domaines basiques des deux facteurs bHLH dimérisés permettant ainsi aux dimères de se lier spécifiquement à des séquences hexanucléotidiques de type 5'-CANNTG-3', dénommées E-box, présentes aux niveaux des promoteurs de gènes cibles des facteurs de transcription. La variabilité observée au niveau de ce domaine basique explique que l'activité transcriptionnelle induite par TWIST1 est différente de celle induite par TWIST2 (Figure 31C).

Outre ces deux domaines caractéristiques des bHLH, les protéines TWIST possèdent, au niveau de leur région N-terminale (Amino-terminale), deux séquences signal identiques de localisation

nucléaire (Singh & Gramolini, 2009). Ces deux séquences, riches en arginine et en lysine chargés très positivement, permettent de réguler le transport des protéines dans le noyau *via* la reconnaissance de ces séquences par des protéines réceptrices d'import nucléaire (Dingwall & Laskey, 1991). La première de ces deux séquences RKRR est située en position 37-40 pour TWIST1 et 31-34 pour TWIST2. La seconde KRGKK est située en position 73-77 pour TWIST1 et en position 49-53 pour TWIST2. De plus, elles sont constituées d'un domaine WR identique au niveau de leur extrémité C-terminale, domaine d'interaction protéique autrement nommé domaine Twist en position 183-202 pour TWIST1 et en position 141-160 pour TWIST2 (Puisieux et al., 2006; Singh & Gramolini, 2009). Une mutation faux sens F187L au niveau du domaine WR de la protéine TWIST1 humaine est suffisante pour induire le syndrome de Saethre-Chotzen ce qui montre l'importance fonctionnelle de ce domaine (Figure 28)(Gripp, Zackai, & Stolle, 2000; Kress et al., 2006). En revanche, contrairement à TWIST2, TWIST1 possède un domaine NSEEE en position 19-23 et un domaine riche en glycine en position 47-98, au niveau de son extrémité N-terminale (Puisieux et al., 2006; Singh & Gramolini, 2009). Afin d'étudier la fonction du domaine riche en glycine de TWIST1, différentes insertions ou délétions au niveau de ce domaine, associées au Syndrome de Saethre-Chotzen, ont été étudiées (Elanko et al., 2001; Funato et al., 2005). Ces études ont montré que ces mutations n'étaient pas les plus pathogènes : elles ne conduisent pas forcément au syndrome de Saethre-Chotzen puisqu'en effet elles sont été retrouvées chez les parents de sujets atteints, n'étant eux-mêmes pas touchés par ce syndrome (Elanko et al., 2001). D'autre part, les variations de la longueur de ce domaine riche en glycine ne semble pas avoir d'impact sur la fonction de la protéine TWIST1 (Funato et al., 2005). Ceci met en évidence la flexibilité de ce domaine qui, de part ces propriétés, est moins sujet aux contraintes imposées par les mutations (Figure 31)(Elanko et al., 2001; Funato et al., 2005).

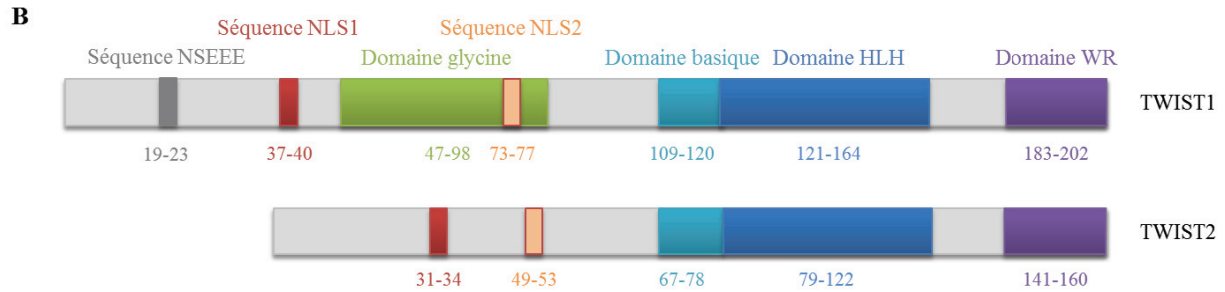
A

CLUSTAL O(1.2.1) multiple sequence alignment

```

Q15672|T1  MMQDVSSSPVSPADDSLSNSEEPDRQQPPSGKRGGRKRRSSRRSAGGGAAGPAGGAGGGVGGGDEPQS PAQGKRGKKSAGCGGGGAGGGGSSSSGGGSPQS
Q8WVJ9|T2  -MEEGSSSPVSPVDS-LGTSEEELERQP---KRFGRKRRYSKSS-----EDGSPTPGKRGK-----KGSPSAQSS
          *:: *****.*. *. *.* ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **
          *:: *****.*. *. *.* ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **

Q15672|T1  YEELQTRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQT LKLAARYIDFLYQVLQSDLEDSKMA SC SYVAHERLSYAF SVWRMEGAWSMSASH
Q8WVJ9|T2  FEELQSRILANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQT LKLAARYIDFLYQVLQSDLEMDNKMTSCSYVAHERLSYAF SVWRMEGAWSMSASH
          :***:***:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
  
```



C

	Séquence entière	NLS1	NLS2	basique	HLH	WR
Pourcentage d'identité de séquences entre TWIST1 et TWIST2	81,88%	100%	100%	83,33%	100%	100%

Figure 31: structure des protéines TWIST1 et TWIST2.

A: Alignement des séquences TWIST1 (Q15672) et TWIST2 (Q8WVJ9) avec Clustal Omega (Sievers & Higgins, 2014) B: Schématisation des domaines de structures de TWIST1 et TWIST2 (D'après Puisieux and al., 2006) (Puisieux, Valsesia-Wittmann, & Ansieau, 2006) C: Taux d'identité de séquence entre les domaines de structures de TWIST1 et de TWIST2 (Clustal Omega ; (Sievers & Higgins, 2014)).

1.2.3.1 TWIST1 selon la classification de Murre

Selon la classification la plus fréquente qui tient compte de la distribution tissulaire, des capacités de dimérisation et des capacités de liaison à l'ADN, TWIST1 appartient à la classe II (Massari & Murre, 2000; Murre et al., 1994). Cette classe regroupe des protéines qui ont une expression tissu spécifique. TWIST1 s'exprime dans les tissus dérivés du mésoderme et préférentiellement dans les cellules progénitrices telles que les ostéoblastes et les myocytes (Bialek et al., 2004; Hebrok, Wertz, & Fuchtbauer, 1994; Pettersson et al., 2011; Spicer, Rhee, Cheung, & Lassar, 1996). TWIST1 est également exprimé au niveau des tissus adipeux blancs et bruns (Pan, Fujimoto, Lopes, & Wang, 2009; Pettersson et al., 2010). D'un point de vue pathologique, TWIST1 est surexprimée dans de nombreux cancers humains et dans de nombreuses lignées cellulaires cancéreuses (Ansieau et al., 2008; J. Yang et al., 2004). D'autre part, les protéines TWIST1, sont capables de former des homodimères TWIST1/TWIST1 mais forment préférentiellement des hétérodimères avec des protéines E comme par exemple TWIST1/E12 ou TWIST1/E47 (Connerney et al., 2006; De Masi et al., 2011; Murre et al., 1994).

1.2.3.2 TWIST1 selon la classification de Jones

TWIST1, *via* son domaine basique, pourra lier sous forme dimérique, des séquences E-box hexanucléotidiques palindromiques ou semi-palindromiques de type 5'-CANNTG-3' (S. Jones, 2004). La classification de Jones est basée sur cette caractéristique de liaison à l'ADN. TWIST1 appartient ainsi à la classe A liant les E-box CAGCTG ou CACCTG. Certains acides aminés des domaines basiques et HLH semblent jouer un rôle clef dans la liaison de ces dimères à l'ADN. De Masi a montré que la Thréonine en position 121 est important dans la reconnaissance des demi-sites de l'E-box (De Masi et al., 2011). D'autre part, les résidus d'Arginine 118, de Sérine 144 et de Lysine 145 jouent un rôle clef dans la liaison des dimères de TWIST1 à l'ADN (Bouard et al., 2014). Ainsi le choix du partenaire de dimérisation, dépendant du contexte cellulaire, influence, de par sa composition en acides-aminés et ainsi sa structure, la réponse cellulaire puisqu'il impacte directement l'affinité de la liaison des dimères bHLH à l'ADN (Bouard et al., 2014; De Masi et al., 2011).

1.2.4 Partenaires de dimérisation de TWIST1 *via* son domaine HLH

Le choix du partenaire de dimérisation conduit à une diversité des réponses cellulaires et est un point de régulation cellulaire majeur (Castanon, Von Stetina, Kass, & Baylies, 2001; N. Jones, 1990; Kadesch, 1993). En effet, les homodimères ou les hétérodimères, selon la nature des monomères qui les constituent, n'auront pas la même affinité de liaison à l'ADN ni les mêmes cibles biologiques. Les deux types de complexes homodimériques ou hétérodimériques de TWIST1 possèdent des fonctions distinctes voire antagonistes au cours du développement notamment au cours de la myogenèse de la Drosophile (Castanon et al., 2001) ou lors du développement des membres et des structures crânio-faciales chez la souris (Connerney et al., 2006; Spicer et al., 1996).

Il a été montré chez la drosophile que l'expression de l'homodimère ne permettait pas de compenser les altérations phénotypiques induites par une délétion complète du gène *twi* (analogue du gène *TWIST1* humain) (Castanon et al., 2001; Spicer et al., 1996). Ceci implique donc l'existence d'autres complexes fonctionnels de TWIST1 de type hétérodimérique. Les protéines bHLH de la classe II, à laquelle TWIST1 appartient, forment préférentiellement des hétérodimères qui, selon leur nature, peuvent avoir des fonctions antagonistes (Castanon et al., 2001; Connerney et al., 2008; Connerney et al., 2006; De Masi et al., 2011; B. A. Firulli et al., 2005; Murre et al., 1994). TWIST1 est capable de former des hétérodimères avec les facteurs de transcription bHLH de la classe I que sont les protéines dites E2A : E12 et E47 (El Ghouzzi et al., 2000; Gajula et al., 2015; Spicer et al., 1996). Ces deux protéines sont issues de l'épissage alternative du même gène nommé TCF3 (Funato et al., 2005). Des expériences de FRET montrent en effet que TWIST1 peut former des homodimères TWIST1/TWIST1 et des hétérodimères en s'associant avec E47 mais également en s'associant avec HAND2, une autre protéine de classe II (Castanon et al., 2001; B. A. Firulli et al., 2005; Murre,

McCaw, & Baltimore, 1989). L'hétérodimère TWIST1/E12 semble être la forme pro-métastatique de TWIST1 dans le cancer de la prostate (Gajula et al., 2015).

L'état de phosphorylation de la protéine TWIST1 permettrait de réguler la balance homodimère/hétérodimère en modifiant son affinité pour son partenaire de dimérisation. Des mutations mimant un état phosphorylé de TWIST1 (T121D et S123D) favoriseraient la formation de l'hétérodimère TWIST1/HAND2 au profit de l'homodimère TWIST1/TWIST1. A l'inverse, des mutations mimant un état dé-phosphorylé de TWIST1 (T121A et S123A) favoriseraient la formation de l'homodimère de TWIST1 (B. A. Firulli et al., 2005). De plus, les mutations de TWIST1 à proximité des résidus T121 et S123, associées au syndrome de Saethre-Chotzen, impacteraient d'une part la phosphorylation des résidus 121 et 123 par la Protéine Kinase A (PKA), comme nous le verrons dans la partie du syndrome de Saethre-Chotzen, mais impacteraient également la dimérisation de TWIST1 avec MYOD1 (B. A. Firulli et al., 2005; Hamamori, Wu, Sartorelli, & Kedes, 1997). Ce qui confirme en effet que l'interaction de TWIST1, avec ces différents partenaires, peut être régulée par la phosphorylation notamment pour sa dimérisation avec MYOD1.

D'autre part des mutations faux sens au niveau des hélices du domaine bHLH rapportées dans le syndrome de Saethre-Chotzen abolissent la formation des dimères de TWIST1 contrairement aux insertions/délétions ou mutations faux sens localisées au niveau des boucles du domaine bHLH (El Ghouzzi et al., 2000).

Toutefois, l'équilibre entre ces deux formes dimériques, homo et hétérodimérique, semble dépendre également du niveau de la protéine HLH inhibitrice Id appartenant à la classe V qui va se dimeriser préférentiellement avec des protéines E (Connerney et al., 2006; Massari & Murre, 2000). Ces protéines Id sont exprimées au niveau des fronts ostéogéniques et sont donc en compétition avec TWIST1 pour la dimérisation avec les protéines E (Connerney et al., 2006). Au niveau des fronts ostéogéniques, la protéine Id s'associe avec E12, alors, TWIST1 est majoritairement présent sous forme homodimérique. A ce même niveau l'expression, des gènes FGFR2 (Fibroblast Growth Factor Receptor 2) et périostine, régulée par l'homodimère de TWIST1, est retrouvée de manière cohérente. A l'inverse, la forme hétérodimérique TWIST1/E12 est exprimée au milieu des sutures du crâne, puisque la protéine Id est absente, où se retrouve exprimé le gène codant la TSP-1 régulé par cet hétérodimère (Connerney et al., 2006).

Ainsi l'expression relative des partenaires, la modification post-traductionnelle de résidus clés de la protéine TWIST1 et la présence de protéines HLH Id permettent la régulation de la balance entre homodimères et hétérodimères (Figure 32) (Connerney et al., 2008; Connerney et al., 2006; B. A. Firulli et al., 2005).

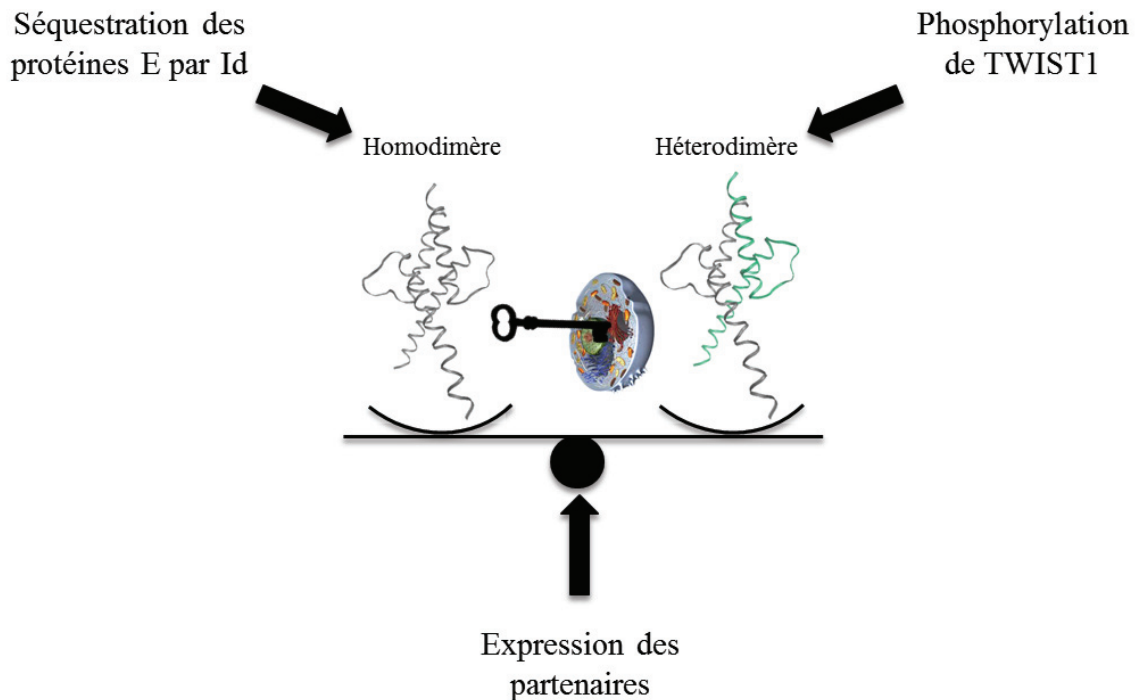


Figure 32: balance entre homodimère et hétérodimère de TWIST1.

1.2.5 Les E-box reconnues par TWIST1

Le domaine de liaison à l'ADN, formé suite au rapprochement des domaines basiques des deux monomères d'un dimère, permet au dimère bHLH de se lier à l'ADN en deux temps (Castanon & Baylies, 2002; Ephrussi, Church, Tonegawa, & Gilbert, 1985; Sauve, Naud, & Lavigne, 2007). Dans un premier temps, les hélices α amphipatiques des monomères ne seront que partiellement stabilisées de sorte à pouvoir scanner l'ADN. Dans un second temps, lorsque le dimère reconnaîtra une séquence E-box spécifique, alors les hélices α amphipatiques des monomères subiront un réarrangement conformationnel induit par l'ADN ce qui permettra donc la stabilisation du dimère sur la séquence E-box reconnue (Sauve et al., 2007). En revanche lorsque le dimère reconnaîtra une séquence E-box qui n'est pas spécifique, alors, les hélices α amphipatiques des monomères resteront dans un état conformationnel instable de sorte à ce que le dimère continue sa fonction de « scanning » (Sauve et al., 2007). De plus la dimérisation et la localisation nucléaire semblent être des prérequis à l'activité transcriptionnelle du facteur de transcription TWIST1 (Figure 33)(El Ghouzzi et al., 2000; Schwoebel & Moore, 2000).

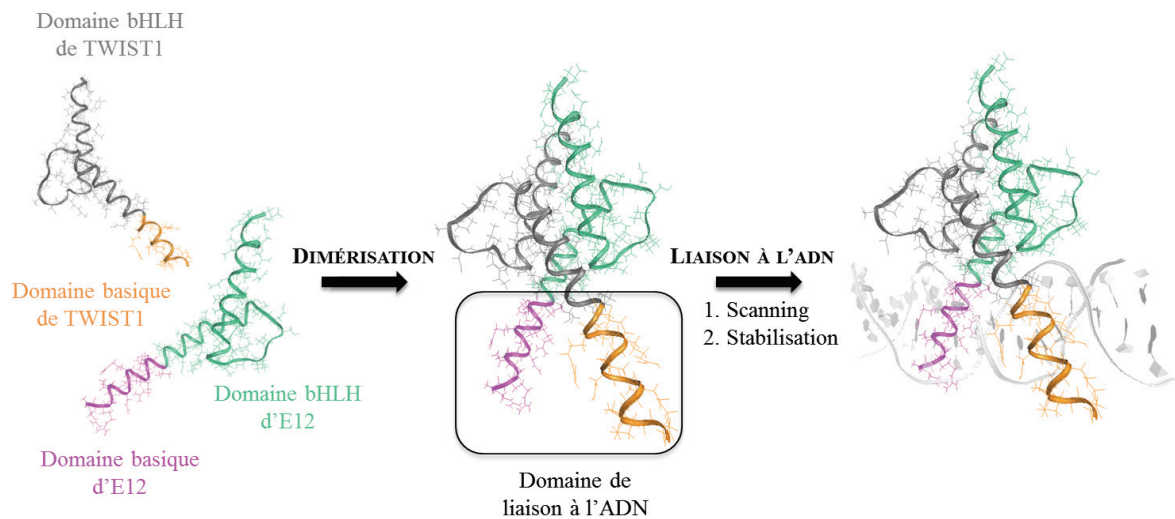


Figure 33 : dimérisation de TWIST1 est un prérequis à son activité transcriptionnelle.

De manière générale ces E-box, séquences hexanucléotidiques, sont constituées de quatre bases externes dites consensus et de deux bases centrales qui sont dites variables. L'affinité de liaison des dimères à l'ADN varie selon les acides aminés des domaines basiques des dimères et selon la nature des bases qui composent les E-box (Figure 34). Ces différences d'affinité permettent de moduler les fonctions de transactivation induites par ces facteurs de transcriptions (De Masi et al., 2011; Fong et al., 2012; Grove et al., 2009). Une autre approche montre que chacun des deux monomères constituant un dimère lie un demi site de l'E-box constitué de 3 nucléotides 5'-CAN-3' (De Masi et al., 2011). Dans ce cas, c'est la troisième base du demi-site qui régit la spécificité de liaison car elle est variable. Les travaux de De Masi ont montré que TWIST1 était une des deux protéines ayant un demi site unique de liaison à l'ADN défini 5'-CAT-3' (De Masi et al., 2011). Lors de la liaison d'un hétérodimère à une séquence non palindromique, la présence d'un demi-site de haute affinité pour un des deux monomères pourrait constituer une base solide de liaison du dimère à l'ADN et permettre ainsi à l'autre monomère bHLH de tolérer une liaison de plus faible affinité pour le second demi-site (De Masi et al., 2011). Ainsi il existe une grande variabilité de liaison des dimères de TWIST1 aux E-box selon le contexte cellulaire. En effet, la classification de Jones montre que TWIST1 lie les séquences E-box 5'-CAGCTG-3' ou 5'-CACCTG-3' (S. Jones, 2004). Les membres de la famille TWIST lie préférentiellement 5'-CATATG-3' aussi bien *in vitro* qu'*in vivo* (Castanon & Baylies, 2002; Ozdemir et al., 2011). Si on s'intéresse plus particulièrement à l'hétérodimère TWIST1/E12, les demi-sites de liaisons décrits pour TWIST1 et E12 sont respectivement 5'-CAT-3' et 5'-CAT-3' ou 5'-CAC-3' ou 5'-CAG-3' (De Masi et al., 2011). Ainsi TWIST1 lierait, sous forme hétérodimérique TWIST1/E12-E47, la séquence palindromique 5'-CATATG-3' ou les séquences non palindromiques 5'-CATCTG-3' and 5'-CATGTG-3' (Figure 34)(De Masi et al., 2011).

Selon Jones (S. Jones, 2004) (M. H. Yang et al., 2010)	Selon Castanon (Castanon & Baylies, 2002) (Ozdemir et al., 2011) (B. A. Firulli et al., 2007) (De Masi et al., 2011)	Selon De Masi sous forme hétérodimérique TWIST1/E2A (De Masi et al., 2011) (B. A. Firulli et al., 2007)
CAGCTG / CACCTG	CATATG	CATATG / CATCTG

Figure 34 : principales E-box décrites pour TWIST1.

Une étude récente montre que la régulation des fonctions de transactivation induites par l'hétérodimère TWIST1/E47 se ferait par l'intermédiaire d'une organisation tétramérique permise par le rapprochement des domaines WR de deux protéines TWIST1. Ainsi ces doubles complexes ne lieraient non plus une mais deux E-box consécutives. Ceci augmente les contraintes de reconnaissance et ainsi augmente la spécificité et la stabilité de liaison des tétramères à l'ADN (Chang et al., 2015).

1.2.6 Autres domaines de liaison de TWIST1

BioGRID référence plus de 60 interactions pour la protéine TWIST1 (BioGRID). Parmi ces interactions, il y a celles qui impliquent le domaine bHLH que nous avons vues précédemment. Mais la protéine TWIST1 contient également d'autres régions capables de se lier à des protéines. Les partenaires sont nombreux impliquant une grande variabilité de la fonction de TWIST1. Nous allons voir ensemble quelques-unes des interactions de TWIST1 décrites dans la littérature (Figure 35). La majorité de ces interactions ont été découvertes par des approches de type western blot ou par des méthodes de spectrométrie de masse, de double-hybrides ou encore *via* la reconstitution de complexes *in vitro*.

Des études ont montré que TWIST1 était capable d'inhiber la liaison de P53 à l'ADN *via* une liaison directe de la partie N-terminale ou de la partie C-terminale de TWIST1 (entre les résidus 1 et 50 ou entre les résidus 175 et 199 de TWIST1) avec P53 (Maestro et al., 1999; Piccinin et al., 2012; Shiota et al., 2008). Ainsi, l'interaction de la partie N-terminale de TWIST1 (1-50) avec le domaine de liaison à l'ADN de P53 (161-223) inhiberait l'apoptose induite par ce dernier en impactant l'induction des gènes cibles de P53 en réponse à des dommages de l'ADN et inhiberait également la répression du promoteur *YB-1* dépendante de TWIST1 (Shiota et al., 2008). D'autre part, l'interaction du domaine C-terminal TWIST-box de TWIST1 (175-199) avec le domaine de régulation de la protéine P53 impacterait les modifications post-traductionnelles de P53 et faciliterait ainsi sa dégradation médiée par la protéine MDM2 (Double Minute 2 protein) (Piccinin et al., 2012). Les études de Shiota et de Puccini se contredisent sur le domaine de TWIST1 impliqué dans l'inhibition de P53 (Piccinin et al., 2012; Shiota et al., 2008). Toutefois, les études de Stanopoulos confirment les données de Shiota. En effet, l'interaction de TWIST1 et P53 semble pouvoir être modulée partiellement par la protéine HOXA5 (Stasinopoulos et al., 2005). Les résidus de 1 à 50 de TWIST1 semblent permettre la liaison

directe à la protéine HOXA5, impliquée dans la tumorigenèse mammaire, *via* ses résidus de 81 à 105 (Figure 35) (Stasinopoulos et al., 2005).

D'autre part, TWIST1 est capable de moduler l'activité de deux Histones Acetyl-Transférases (HAT) indépendantes *via* une interaction directe. Les HAT sont des enzymes impliquées dans la régulation épigénétique de l'expression des gènes. La première HAT concernée est p300 qui joue un rôle à la fois au niveau de la chromatine et à la fois au niveau des HAT elles-mêmes (Delvecchio, Gaucher, Aguilar-Gurrieri, Ortega, & Panne, 2013). Hamamori et son équipe ont montré que p300 interagissait avec les résidus de TWIST1 compris entre les deux domaines NLS (Nuclear Localization Signal) localisés entre les acides aminés 30 et 80 (Hamamori et al., 1999). La seconde HAT concernée est la p300/CBP-Associated Factor (pCAF) qui appartient à la famille des HAT nucléaire GCN5/pCAF (GCN5 related N-acetyltransferases). En 1999, les travaux de Hamamori ont montré que cette HAT interagissait directement avec TWIST1 notamment avec les acides aminés 1 et 111 (Figure 35) (Hamamori et al., 1999).

TWIST1 interagit également avec plusieurs composants du complexe Mi2 β /NuRD (NUcleosome Remodeling and Deacetylase). Ce complexe joue de nombreuses fonctions notamment dans la répression des gènes de nombreux processus physiologiques comme l'initiation et la progression tumorale (Fu et al., 2011). TWIST1 semble lier MTA2 (MeTastasis Associated 1 family member 2), RbAp46 (Retinoblastoma-binding protein p46), Mi2 β (Mi-2 autoantigen 218 kDa protein) and HDAC2 (Histone DeACetylase 2). La région N-terminale de TWIST1, comprend les acides aminés de 1 à 112 lie Mi2 β alors que la région C-terminale de TWIST1, comprend les acides aminés de 110 à 202 lie RbAp46 (Fu et al., 2011). La protéine MTA2, *via* son domaine C-terminal (principalement les acides aminés compris entre 308 et 436 de ce domaine, puis, dans une moindre mesure les acides aminés de 434 à 561 et les acides aminés de 563 à 668) se lie à TWIST1 aussi bien au niveau de sa partie N et C-Terminale (Fu et al., 2011). De plus la partie C-terminale de MTA2 interagit également avec RbAp46 et HDAC2 (Fu et al., 2011). Contrairement aux interactions directes de TWIST1 et de MTA2, de TWIST1 et de RbAp46 ainsi que TWIST1 et de Mi2 β , l'interaction de HDAC2 avec TWIST1 semble toutefois indirecte et se fait par l'intermédiaire du domaine C-terminal de MTA2 (Figure 35) (Fu et al., 2011).

TWIST1 joue un rôle au niveau de la différenciation ostéoblastique *via* son interaction directe avec RUNX2 appartenant à la famille des facteurs de transcription RUNX. TWIST1 interagit avec RUNX2 *via* sa séquence d'acides aminés comprise entre 183 et 202, qui n'est autre que le domaine WR autrement nommé TWIST box (Figure 35) (Bialek et al., 2004). D'autre part, TWIST est capable de réguler la voie de signalisation des cytokines (Sosic, Richardson, Yu, Ornitz, & Olson, 2003). En effet, le domaine WR de TWIST1, permettrait à TWIST1 d'interagir directement avec NF- κ B (Nuclear Factor kappa B) et influencerait ainsi l'expression de l'interleukine 8 (Figure 35) (S. Li et al., 2012).

Le domaine N-terminal de TWIST1 comprenant les acides aminés situés entre 1 et 30, est également impliqué dans une autre interaction directe avec la protéine PGC-1 α (Peroxisome proliferator-activated receptor gamma coactivator 1-alpha PPARGC1A) (Pan et al., 2009). Cette dernière est un co-activateur transcriptionnel qui joue un rôle clef de régulateur dans la thermogenèse de la graisse brune (Figure 35) (Pan et al., 2009).

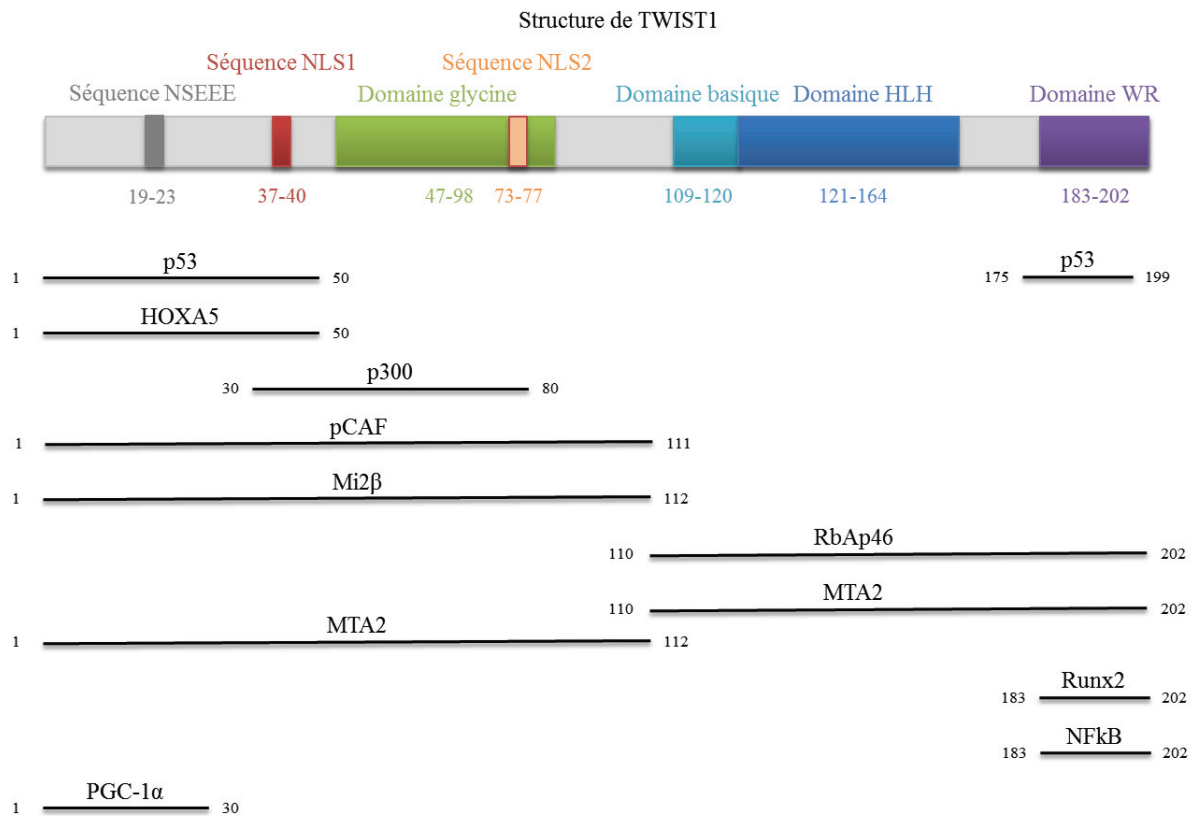


Figure 35 : domaines d'interactions de TWIST1 avec ces partenaires.

1.3 Pathologies associées aux mutations des protéines TWIST

Certaines mutations des gènes TWIST humains conduisent à une altération de l'expression et/ou de la fonction protéique. L'haplo-insuffisance de *Twist1* est associée à une perte fonctionnelle de la protéine TWIST1 caractéristique du syndrome de Saethre-Chotzen (SCS). Cette pathologie est caractérisée par une craniosténose et une malformation des membres, phénotype retrouvé chez des souris hétérozygotes *Twist1*^{+/-} (Ansieau, Morel, Hinkal, Bastid, & Puisieux, 2010; A. B. Firulli & Conway, 2008). Autrefois le syndrome de Robinow-Sorauf était un syndrome bien à part, mais les études de Cai en 2003 l'ont associé au syndrome de Saethre-Chotzen (Cai, Shoo, Sorauf, & Jabs, 2003; Robinow & Sorauf, 1975). Il a également été décrit qu'une trisomie au niveau du gène *Twist1* conduisait à de multiples anomalies dont une encéphalocèle (crâne bifidum) (Stankiewicz et al., 2001).

1.3.1 Pathologie associée aux mutations de TWIST1 : Le syndrome de Saethre-Chotzen

Les craniosténoses sont des anomalies très fréquentes de l'ossification du crâne dont l'incidence à la naissance concerne 1 nouveau-né sur 3000. Nous nous intéresserons dans cette partie, uniquement à une craniosténose autosomique dominante qui correspond au syndrome de Saethre-Chotzen mis en évidence dès 1931 par le neuropsychiatre norvégien H. Saethre et le pédiatre allemand F. Chotzen (Chotzen, 1932; Saethre, 1931). Ce syndrome est caractérisé par une fusion prématurée d'une ou plusieurs sutures crâniennes. Il entraîne une déformation du crâne et parfois une hypertension intracrânienne chronique, ainsi que des anomalies des membres (el Ghouzzi et al., 1997; Howard et al., 1997). Une haplo insuffisance du gène *TWIST1*, situé sur le chromosome 7p21.2 chez l'Homme, conduit à ce syndrome dont l'incidence à la naissance est estimée à 1/25 000 à 1/50 000 nouveau-nés.

1.3.1.1 Les signes cliniques et traitements

Les phénotypes associés à ce syndrome peuvent être plus ou moins marqués. L'intelligence des patients atteints est habituellement normale, bien que ceux ayant de grandes délétions génomiques soient plus susceptibles d'avoir des retards de développement. Le diagnostic clinique du SCS est basé sur les caractéristiques cliniques suivantes (Figure 36) :

- Une craniosténose qui correspond à une fusion prématurée d'une ou plusieurs sutures de la voûte crânienne (suture coronale). Cette caractéristique clinique est souvent associée à une brachycéphalie (morphologie du crâne plus large) et à une acrocéphalie (morphologie du crâne en pointe),
- Une implantation basse des cheveux sur le front, ptosis (chute de la paupière supérieure), strabisme, asymétrie faciale (en particulier chez les personnes atteintes synostose coronale unilatérale),
- Un aspect caractéristique de l'oreille (oreille petite et ronde avec une déformation caractéristique du pavillon),
- Des anomalies des membres, y compris une brachydactylie (doigts et orteils anormalement courts), une syndactylie partielle (accolement/fusion des deuxième et troisième doigts de la main) et une clinodactylie (déviation et chevauchements des doigts et orteils).

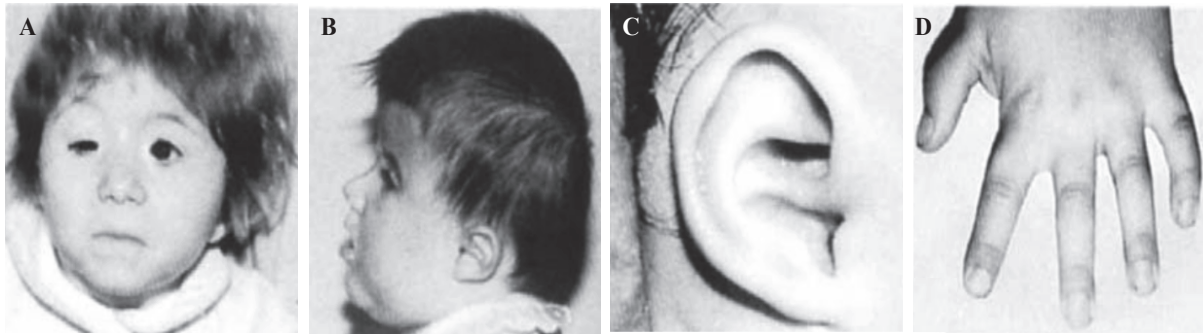


Figure 36 : signes cliniques associés au Syndrome de Saethre-Chotzen.

D'après El Ghouzzi and al., 1997(el Ghouzzi et al., 1997)

A: Dimorphisme facial avec ptosis asymétrique et brachycéphalie. B: brachycéphalie avec front plat, implantation des cheveux anormalement basse. C: Petite oreille ronde avec forme du pavillon caractéristique. D: Syndactylie partielle des 2^{ème} et 3^{ème} doigts.

En ce qui concerne la prise en charge des patients atteints par le syndrome de Saethre-Chotzen, il y a une prise en charge chirurgicale qui doit être réalisée dans les meilleurs délais après la naissance. Autrefois la prise en charge chirurgicale consistait à éliminer la compression cérébrale alors qu'actuellement la prise en charge est à la fois neurochirurgicale et plastique afin de corriger également les dysmorphies.

1.3.1.2 Tests diagnostics

Le diagnostic de ce SCS est principalement basé sur les résultats cliniques après avoir établi un arbre généalogique des deux parents afin de déterminer s'il y a des antécédents médicaux ou des maladies génétiques connus. C'est l'invalidation hétérozygote de *Twist1* dans un modèle murin, qui a mis en évidence un phénotype osseux similaire aux symptômes décrits dans le SCS, en parallèle de la localisation du gène *TWIST1* qui ont permis d'identifier le gène *TWIST1* comme responsable de cette maladie (el Ghouzzi et al., 1997; A. B. Firulli & Conway, 2008; Howard et al., 1997). Cependant il est important de noter que certains patients présentant des mutations du gène *FGFR3*, notamment porteurs de la mutation Pro250Arg impliquée dans le syndrome de Muenke, ont été assimilés pendant longtemps au Syndrome de Saethre-Chotzen, tout comme les patients présentant une mutation du gène *FGFR2* (Paznekas et al., 1998). Toutefois, Wolfram Kress et son équipe ont montré des caractéristiques fonctionnelles différentes de ces phénotypes issus des mutations des gènes *FGFR* permettant de les différencier du SCS (Kress et al., 2006). Parfois, les personnes touchées ont une translocation chromosomique impliquant le locus 7p21. Ainsi, les mutations du gène codant pour la protéine TWIST1 restent la principale cause du SCS (Kress et al., 2006).

1.3.1.3 Les mutations de TWIST1 associées au SCS

De nombreuses mutations ont été décrites au niveau de la séquence protéique de TWIST1 pour les patients atteints de ce syndrome. Nous pouvons retrouver des mutations décrites et associées au SCS tout le long du premier exon du gène codant pour la protéine TWIST1 (comprenant les acides nucléiques de 1 à 609). Toutefois, nous pouvons noter que plus de 75% des mutations concernent des acides nucléiques impliquant des acides aminés localisés au niveau des domaines bHLH. Les mutations peuvent être non-sens c'est à dire qui entraînent l'apparition d'un codon stop de manière prématurée. C'est le cas, par exemple, de la mutation C368A qui change le codon TCG en TAG et qui remplace donc le résidu de sérine en position 123 par un codon stop (el Ghouzzi et al., 1997; Kress et al., 2006). Les mutations peuvent également être des mutations faux sens. Dans ce cas la mutation ponctuelle du nucléotide au sein d'un codon conduit à un changement de l'acide aminé codé par le codon. La mutation G353A induit une mutation faux sens en substituant le résidu d'arginine 118 en histidine (Rose, Patel, Reardon, Malcolm, & Winter, 1997). Les mutations associées au SCS peuvent également être des insertions ou des duplications, avec ou sans décalage du cadre de lecture, ou également des délétions qui entraînent l'apparition prématurée d'un codon terminal, ou qui allongent au contraire la protéine (El Ghouzzi et al., 1999; Gripp et al., 2000; Howard et al., 1997). L'insertion de sept acides aminés au niveau de la boucle située entre les deux hélices du domaine bHLH, aussi bien en position 135 qu'en position 139, a été décrite dans le SCS et déstabilise les complexes de TWIST1 mutants sur l'ADN ce qui peut être une explication de la perte de fonction de cette protéine (Bouard et al., 2014; el Ghouzzi et al., 1997; Howard et al., 1997; Rose et al., 1997). D'autre part, certaines mutations associées à ce syndrome comme par exemple la mutation faux sens R39G conduit à un défaut d'import nucléaire de TWIST1 (Funato et al., 2005; Singh & Gramolini, 2009) (Vichalkovski et al., 2010). Contrairement au TWIST1 sauvage, la translocation nucléaire de ce mutant nécessite la présence de E12 (Funato et al., 2005). Néanmoins il n'existe pas de corrélation entre la nature des mutations et la sévérité des phénotypes associés au SCS qui est caractérisé par une perte fonctionnelle de TWIST1. En effet une même mutation, par exemple C309G, peut engendrer une expression phénotypique caractéristique du SCS dans une première famille ou engendrer une expression phénotypique non syndromique dans une seconde famille (Figure 37) (Kress et al., 2006).

Ces différentes mutations ont permis de mettre en évidence les zones fonctionnelles de la protéine TWIST1 puisque ce syndrome est associé à une perte fonctionnelle de la protéine TWIST1. De ce fait Vincent El Ghouzzi et son équipe ont mis en évidence une corrélation entre la nature de la mutation, sa position au niveau du domaine bHLH et la mécanistique de perte de fonction de TWIST1 associée (El Ghouzzi et al., 2000; El Ghouzzi et al., 2001). En effet, les mutations non-sens impactent la stabilité de la protéine en favorisant la dégradation des protéines tronquées. Lorsqu'une mutation faux sens est localisée au niveau d'une des deux hélices du domaine bHLH ceci impacte la

dimérisation de TWIST1 qui est un prérequis pour sa translocation nucléaire et pour sa liaison à l'ADN. Et enfin, les mutations concernant des acides aminés du domaine basique ou à la jonction entre la boucle et la seconde hélice n'ont pas d'impact sur la dimérisation de TWIST1 avec ces partenaires, mais altèrent la fixation des dimères à l'ADN.

D'autres part, au vu de la localisation de certaines mutations à proximité des sites de phosphorylation de TWIST1 au niveau de la première hélice α du domaine bHLH et compte tenu que l'état de phosphorylation d'une protéine semble influencer la balance homo/hétérodimère et donc l'activité de ces facteurs de transcription, l'équipe de Anthony B. Firulli a étudié l'impact de ces mutations sur la phosphorylation de TWIST1 (B. A. Firulli et al., 2005). L'étude individuelle de cinq mutations R116W, R118H, Q119P, R120A et E117K a mis en évidence une diminution de la phosphorylation des résidus T121 et S123 par la PKA en présence des mutations. Ceci démontre que les mutations adjacentes aux sites de phosphorylation peuvent influencer la phosphorylation de la protéine et donc modifier la balance homo/hétérodimère (B. A. Firulli et al., 2005). Ceci peut être pour Anthony B. Firulli, une explication de la perte fonctionnelle de TWIST1 associée au syndrome de Saethre-Chotzen. En accord avec ces hypothèses, l'altération de la dimérisation de TWIST1 avec la protéine HAND2 dans les bourgeons de membres de la souris conduit ainsi une polydactylie, un défaut typique du syndrome de Saethre-Chotzen, soulignant de nouveau l'importance des mécanismes de régulation post-traductionnelle dans le contrôle de l'activité de la protéine TWIST1 (B. A. Firulli et al., 2007).

Acides nucléiques mutés	Acides aminés modifiés	Domaine fonctionnel	Références bibliographiques
G61T	E21X	Domaine N-terminal	Kress et al., 2006
81-82delinsTT	Q28X	Domaine N-terminal	James and al., 2009
92C>T	Q28X	Domaine N-terminal	James and al., 2009
G106T	G36X	Domaine N-terminal	Kress et al., 2006
108delA	Hors du cadre de lecture	Domaine N-terminal	Kress et al., 2006
181G>T	G61X	Domaine N-terminal	Rose et al., 1997
190-200dup	S68T	Domaine N-terminal	James and al., 2009
193G>T	E65X	Domaine N-terminal	Kaspárová et al., 1998; Rose et al., 1997; el Ghouzzi et al., 1999; Kress et al., 2006
211C>T	Q71X	Domaine N-terminal	Kress et al., 2006
230delA; 232T>C	K77X ; K77S	Domaine N-terminal	Johnson et al., 1998; Rose et al., 1997; James and al., 2009
232T>C	S78P	Domaine N-terminal	Rose et al., 1997
263delG	G88A	Domaine N-terminal	Gripp et al., 2000
276-277 ins21bp	GAGGGGG92-92ins	Domaine N-terminal	Kaspárová et al., 1998
272-273 ins10bp	S93G	Domaine N-terminal	Kaspárová et al., 1998
308-309insA	Y103X	Domaine N-terminal	Howard et al., 1997; Paznekas et al., 1998
309C>A	Y103X	Domaine N-terminal	el Ghouzzi et al., 1997b; el Ghouzzi et al., 1999
309C>G	Y103X	Domaine N-terminal	Paznekas et al., 1998; el Ghouzzi et al., 1999; Kress et al., 2006; James and al., 2009
310G>T	E104X	Domaine N-terminal	el Ghouzzi et al., 1997a; Johnson et al., 1998; el Ghouzzi et al., 1999
336delG	Hors du cadre de lecture	Domaine de liaison à l'ADN	Kress et al., 2006
346C>T	R116W	Domaine de liaison à l'ADN	Paznekas et al., 1998
	R118C	Domaine de liaison à l'ADN	el Ghouzzi et al., 2001
352-354delCGC	R118del	Domaine de liaison à l'ADN	el Ghouzzi et al., 1997a
353G>A	R118H	Domaine de liaison à l'ADN	Rose et al., 1997
353G>C	R118P	Domaine de liaison à l'ADN	Kress et al., 2006
353del8bp	R118H	Domaine de liaison à l'ADN	el Ghouzzi et al., 1999
352del3	R118Q	Domaine de liaison à l'ADN	el Ghouzzi et al., 1997
355C>T	Q119X	Domaine de liaison à l'ADN	Kress et al., 2006
356A>C	Q119P	Domaine de liaison à l'ADN	Howard et al., 1997
359G>C	R120P	Domaine de liaison à l'ADN	Kaspárová et al., 1998
364C>T	Q122X	Hélice 1	Paznekas et al., 1998
368C>A	S123X	Hélice 1	el Ghouzzi et al., 1997b; Kress et al., 2006
368C>G	S123W	Hélice 1	Johnson et al., 1998; Kress et al., 2006
376G>T	E126X	Hélice 1	el Ghouzzi et al., 1997; Kaspárová et al., 1998; 1998; Rose et al., 1997; el Ghouzzi et al., 1999; James and al., 2009
379G>A and 398A>T	A127T et K133I	Hélice 1	Kress et al., 2006
C380A	A127GE	Hélice 1	Kress et al., 2006
384-385insC	A129R	Hélice 1	Paznekas et al., 1998
385-405dup	dans le cadre de lecture	Hélice 1	Kress et al., 2006
379G>C	A129P		Seifer and al., 2006
392T>C	L131P	Hélice / formation de la boucle	el Ghouzzi et al., 1997b
395G>T	R132L	Hélice 1	James and al., 2009
395G>C	R132P	Hélice 1	Paznekas et al., 1998
397-417dup	dans le cadre de lecture	Hélice 1 / boucle	Kress et al., 2006
402C>G Helix 1	I134M	Hélice 1	Rose et al., 1997
405-406 ins21bp	AALRKII135-136ins	Hélice 1	Howard et al., 1997
407C>T	P136L	Hélice 1	Johnson et al., 1998; Kaspárová et al., 1998
409A>C	T137P	Hélice 1	Kress et al., 2006
415C>T	P139S	Boucle	Paznekas et al., 1998
416-417 ins21bp	KIIPTLP139-140ins	Hélice 1 / boucle	el Ghouzzi et al., 1997; Howard et al., 1997; Johnson et al., 1998; Paznekas et al., 1998
417-418 ins21bp	KIIPTLP139-140ins	Hélice 1 / boucle	el Ghouzzi et al., 1997; Rose et al., 1997
418-419 ins21bp	S140X	Boucle	Rose et al., 1997
420-421 ins21bp	IIPTLP140-141ins	Boucle	Rose et al., 1997
421G>T	D141Y	Boucle	Johnson et al., 1998; Paznekas et al., 1998
421G>C	D141H	Boucle	James and al., 2009
422A>G	D141G	Boucle	Rose et al., 1997
423-424 ins25bp	DHPHAALG141-142ins	Boucle	el Ghouzzi et al., 1999
428delT	Hors du cadre de lecture	Boucle	Kress et al., 2006
430A>C	S144R	Boucle	el Ghouzzi et al., 1999
433A>G	K145E	Boucle	el Ghouzzi et al., 1997; el Ghouzzi et al., 1999
433del123bp	K145G	Boucle	Howard et al., 1997
435G>C	K145N	Boucle	Rose et al., 1997
438-439insT	Q147X	Boucle	James and al., 2009
442A>G	T148A	Boucle	Kaspárová et al., 1998
443C>A	T148N	Boucle	Ray et al., 1997
443C>G	T148S	Boucle	Kress et al., 2006
443C>T	T148I	Boucle	Kress et al., 2006
445C>T	L149F	Boucle	Paznekas et al., 1998
455C>T	A152V	Boucle	Paznekas et al., 1999
460A>G	R154G	Hélice 2	Rose et al., 1997
460-461insA	R154K	Hélice 2	Kunz et al., 1999
464del5bp	Y155X	Hélice 2	el Ghouzzi et al., 1997; el Ghouzzi et al., 1999
470A>T	N157V	Hélice 2	Kress et al., 2006
474C>G	F158L	Hélice 2	Kress et al., 2006
475C>T	L159F	Hélice 2	el Ghouzzi et al., 1997; el Ghouzzi et al., 1999
480C>G	Y160X	Hélice 2	Kaspárová et al., 1998
481C>T	Q161X	Hélice 2	el Ghouzzi et al., 1997; el Ghouzzi et al., 1999; Kress et al., 2006
487delC	L163F	Hélice 2	el Ghouzzi et al., 1999
490C>T	Q165X	Hélice 2	Gripp et al., 2000
541G>T	E181X	Domaine C-terminal	Gripp et al., 1999b
561C>G	F187L	TWIST Box	Kress et al., 2006
570G>A	W190X		Seifer and al., 2006
Deletion		Entire gene	Kress et al., 2006

Figure 37: mutations du gène codant pour TWIST1 associées au SCS.

1.3.2 Pathologie associée aux mutations de TWIST2 : Le syndrome de Setleis

Le syndrome de Setleis a été découvert pour la première fois par le pédiatre Américain Howard Setleis lors de l'étude de cinq enfants d'une famille Portoricaine consanguine (Setleis, Kramer, Valcarcel, & Einhorn, 1963). Le mode de transmission de ce syndrome n'est pas clairement établi puisqu'il est considéré comme maladie héréditaire autosomique récessive extrêmement rare d'une part et comme une maladie héréditaire autosomique dominante d'autre part (OMIM 227260) (Cervantes-Barragan et al., 2011) (Girisha et al., 2014; McGaughran & Aftimos, 2002; Ward & Moss, 1994).

1.3.2.1 Les signes cliniques, diagnostics et traitements

Cette maladie est une dysplasie ectodermique faciale de type III qui affecte généralement les cheveux, les dents, les ongles et / ou la peau. En effet les patients atteints du syndrome Setleis présentent une dysmorphie des traits du visage avec un faciès spécifique dit Léonin, des marques bilatérales au niveau des tempes, des traits supplémentaires au visage, des anomalies au niveau des cils, des sourcils (absence des glandes Meibonius), du nez (bout du nez en forme de bulbe c'est à dire très arrondi) et des lèvres (forme caractéristique en V inversé avec une lèvre supérieure anormalement importante, et / ou des lèvres anormalement épaisses ; Figure 38). De plus les patients atteints présentent également des manifestations non faciales comme par exemple une hyper- ou hypopigmentation de certaines zones du corps (Girisha et al., 2014; McGaughran & Aftimos, 2002; Tukul et al., 2010).



Figure 38 : signes cliniques associés au syndrome de Setleis.

D'après Girisha and al. 2014 (Girisha, Bidchol, Sarpangala, & Satyamoorthy, 2014) et Tukul and al., 2010 (Tukul et al., 2010).

A et C: Visage léonin avec lésions bitemporales vues de face. B et D: Lésions bitemporales vues de profil. E: Plis de peaux associés aux lésions bitemporales. F: Anomalies des cils supérieurs.

Ce syndrome de Setleis est basé sur une évaluation clinique approfondie et est diagnostiqué généralement peu après la naissance. En complément, des examens microscopiques de petits échantillons de peau ainsi que des tests génétiques peuvent être effectués. Des mutations homozygotes de *TWIST2* ont été récemment identifiées chez des patients atteints du syndrome de Setleis issus de deux familles consanguines indépendantes, alors que les patients de cinq autres familles atteints par ce syndrome de Setleis ne présentent pas de mutation de *TWIST2* (Tukel et al., 2010). De plus il existe des similitudes phénotypiques entre les souris *Twist2*^{-/-} et les patients atteints par la syndrome de Setleis (Tukel et al., 2010). Toutefois, au vu de la majorité des cas n'ayant pas de mutations du gène codant pour TWIST2 et compte tenu des doutes persistants sur le mode de transmission de cette maladie, les mécanismes moléculaires mis en jeu restent à élucider.

D'un point de vue thérapeutique, le traitement reste symptomatique. En effet, les anomalies visibles sur chacun des individus atteints (nasales, labiales et oculaires) peuvent être corrigées chirurgicalement. Toutefois, comme les anomalies faciales ont tendance à diminuer avec l'âge des patients dans certains cas, la chirurgie corrective peut éventuellement être retardée.

1.3.2.2 Les mutations de TWIST2 associées au Syndrome de Setleis

A ce jour, seules des mutations au niveau du facteur de transcription bHLH TWIST2 semblent pouvoir expliquer le syndrome de Setleis. En effet la découverte de mutations non-sens ou encore de délétion homozygote d'un seul nucléotide au niveau du gène codant pour la protéine TWIST2 chez des personnes atteintes conduit à l'expression d'une protéine instable et tronquée (Figure 39)(Girisha et al., 2014; Tukel et al., 2010). Les études de Tukel ont montré que deux mutations non-sens du gène codant la protéine TWIST2 ont été retrouvées chez deux patients atteints du syndrome de Setleis (Tukel et al., 2010). La première mutation non-sens concerne un membre d'une famille arabe. Il s'agit d'une substitution de l'acide nucléique en position 324 (C324T) qui entraîne l'apparition d'un codon stop au niveau du résidu 65 (Q65X)(Tukel et al., 2010). La seconde mutation non-sens concerne un membre d'une famille portoricaine. Il s'agit d'une substitution de l'acide nucléique en position 486 (C486T) causant l'apparition d'un codon stop au niveau du résidu 119 (Q119X) (Tukel et al., 2010). De plus, un enfant atteint par ce syndrome, né d'un couple consanguin indien, est porteur d'une délétion homozygote d'un seul nucléotide en position 91 (91delC) conduisant à une modification du cadre de lecture et donc à une modification de la séquence en acide aminé à partir du résidu 31, créant un codon stop en position 71 (p.R31GfsX71)(Girisha et al., 2014). Et enfin, dans une famille mexicaine, la simple délétion homozygote d'un nucléotide en position 168 (168delC) entraîne une modification du cadre de lecture et donc une modification de la séquence en acide aminé à partir du résidu 57 en créant un codon stop en position 100 et est responsable de ce syndrome (p.SS7AfsX45) (Cervantes-Barragan et al., 2011). Au regard des cas atteints, pour lesquels aucune mutation de

TWIST2 n'a été définie, des études complémentaires doivent être menées afin de mieux comprendre les mécanismes moléculaires associés à ce syndrome.

Acides nucléiques mutés	Acides aminés modifiés	Domaine fonctionnel	Références bibliographiques
C324T	Q65X	Domaine de liaison à l'ADN	Tukel and al., 2000
C486T	Q119X	Hélice 2	Tukel and al., 2000
91delC	p.R31GfsX71	proteine tronquée	Girisha and al., 2014
c168 del C	p.S57AfsX45	frame shift mutation	Cervantes-Barragan and al., 2011

Figure 39 : mutations de TWIST2 décrites dans le syndrome de Setleis.

1.4 Incidence des mutations de TWIST1 associées au syndrome de Saethre-Chotzen sur le développement de cancers.

Au-delà de ses propriétés embryonnaires, la protéine TWIST1 a des propriétés oncogéniques (nous reviendrons sur ce point ultérieurement). Une première étude des années 1983 menée par Elisabeth A. McKeen et son équipe a associé le syndrome de Saethre-Chotzen à la survenue de cancers en raison d'une diminution des capacités immunitaires des individus concernés (McKeen, Mulvihill, Levine, Dean, & Howley, 1984). L'étude concernait une famille de six frères et sœurs, les trois sœurs et le père présentaient un phénotype caractéristique du syndrome de Saethre-Chotzen associant craniosténose et faciès asymétrique alors que les trois frères ne présentaient pas de signes de dysmorphie (McKeen et al., 1984). Une des sœurs a eu un carcinome du nasopharynx à 32 ans, deux des frères ont développé la maladie de Hodgkin à 22 et 28 ans et le dernier des frères a eu un séminome avec un tératome des testicules à 18 ans. Pour la sœur atteinte du Syndrome de Saethre-Chotzen et les trois frères sans dysmorphie, l'apparition de malignomes a été associée à un dysfonctionnement immunitaire (McKeen et al., 1984). Ainsi le développement de tumeurs pour les frères sans dysmorphie serait lié au génotype et non pas au phénotype associé au syndrome de Saethre-Chotzen (McKeen et al., 1984; Seifert, Kress, Meisel, Henze, & Seeger, 2006). Par la suite, Georg Seifer semble avoir constaté, pour une patiente atteinte du syndrome de Saethre-Chotzen, une corrélation entre son syndrome et le développement d'un carcinome rénal (Seifert et al., 2006). Une autre équipe, l'équipe de Pelle Sahlin en 2007, a montré qu'il existait une corrélation entre le syndrome de Saethre-Chotzen et les cancers du sein (Sahlin et al., 2007). En effet, cette étude portant sur 15 familles atteintes par le syndrome de Saethre-Chotzen a permis de mettre en évidence que des femmes atteintes de ce syndrome, suite à des mutations de TWIST1, avaient un risque plus élevé de développer un cancer du sein (Sahlin et al., 2007).

A l'inverse, les études plus récentes de Paul A. James et son équipe ont montré que le risque de développer un cancer du sein n'est en aucun cas augmenté chez les patients avec des mutations de TWIST1 associées au SCS (James et al., 2009). L'étude de 89 individus atteints par le SCS a rapporté que, parmi ces 89 individus, seuls trois ont développé un cancer. Le premier individu a été

diagnostiqué pour un cancer de l'estomac à l'âge de 79 ans. Les deux autres ont été diagnostiqués pour un cancer de la prostate à l'âge de 65 ans (James et al., 2009). D'autre part, cette étude rapporte que les mutations de *TWIST1* non associées à un phénotype du SCS ne sont pas non plus corrélées à une augmentation du risque de développer un cancer, infirmant ainsi la relation entre le développement de tumeurs et le génotype/phénotype associés au SCS avancée par Elisabeth A. McKeen et Georg Seifer (James et al., 2009; McKeen et al., 1984; Seifert et al., 2006). De plus, l'étude de Bergman remet également en cause les études de Pelle Sahlin menées en 2007 qui montraient que les personnes porteuses de mutations de *TWIST1* associées au SCS avaient un risque plus élevé de développer un cancer du sein (Bergman et al., 2009; Sahlin et al., 2007). En effet, les travaux de Annika Bergman montrent que les mutations de *TWIST1* associées au SCS sont rares ou absentes dans les cancers du sein héréditaires négatifs pour *BRCA1/2* (Bergman et al., 2009).

A ce jour, aucune étude n'a été menée sur l'incidence des mutations de *TWIST2* sur le développement tumoral. Ceci peut s'expliquer par le très faible taux de mutations de *TWIST2* associé au syndrome de Setleis. En effet, les mécanismes moléculaires mis en jeu dans ce syndrome ne sont que très peu compris.

1.4.1 Les mutations de *TWIST1* et de *TWIST2* retrouvées dans les cancers.

1.4.1.1 Les mutations de *TWIST1* retrouvées dans les cancers

Cent vingt et une études concernant les altérations de *TWIST1* dans les cancers ont été rapportées sur la base de données cBioPortal (Cerami et al., 2012; J. Gao et al., 2013; Genomics). Les cancers présentant le plus d'altérations de *TWIST1* sont le carcinome invasif du sein (17,2% soit 5 cas sur 29), l'adénocarcinome de la prostate (9,8% soit 6 cas sur 61) et le carcinome urothélial (8,7% soit 11 cas sur 127). Cependant il existe différents types d'altérations (Figure 40).

L'amplification du gène est l'altération génomique la plus fréquente de *TWIST1*. Les cancers présentant le plus fréquemment une amplification de *TWIST1* sont le carcinome invasif du sein (17,2% soit 5 cas sur 29), l'adénocarcinome de la prostate (8,2% soit 5 cas sur 61) et le carcinome urothélial (7,9% soit 10 cas sur 127 ; Figure 40A et B).

Ensuite il y a des délétions au niveau du gène codant la protéine *TWIST1*. Les trois principaux cancers concernés par une délétion de *TWIST1* sont le lymphome diffus à grandes cellules B (2,1% soit 1 cas sur 48), le cancer urothélial (2,1% soit 2 cas sur 97) et le carcinome adénoïde kystique (1,7% soit 1 cas sur 60 ; Figure 40A et B).

Puis il y a les mutations, qui sont majoritairement des mutations faux sens et plus rarement des mutations non-sens. Celles-ci sont réparties uniformément le long de la séquence protéique de *TWIST1* et concernent différents domaines : le domaine riche en glycine et la séquence NLS2, le

domaine basique, le domaine HLH et enfin le domaine WR (Figure 40A et C). Les principaux types de cancers touchés par ce genre d'altération sont l'adénocarcinome pulmonaire (5,9% soit 2 cas sur 34), le cancer du poumon à petites cellules (2,7% soit 3 cas sur 110) et enfin l'adénocarcinome oesophagien (1,4% soit 2 cas sur 146 ; Figure 40A et B). D'autre part, si nous comparons les mutations retrouvées dans les cancers et celles associées au Syndrome de Saethre-Chotzen, nous pouvons voir que six d'entre elles concernent des acides aminés communs. Il s'agit des résidus R116, R120, P136, A152, E181 et W190. Parmi ces résidus, deux sont importants dans la liaison du domaine basique de TWIST1 à l'ADN (R116 et R120). Les deux dernières mutations concernent des acides aminés de la partie C-terminale de TWIST1 et pourraient être impliquées dans la conformation du facteur de transcription. Notons également que deux mutations R116W et A152V sont associées à la fois au syndrome de Saethre-Chotzen et aux cancers, respectivement, de l'estomac et du poumon. Ceci montre que certains cancers, en particulier ceux de l'estomac et des poumons, pourraient être associés à une perte fonctionnelle de TWIST1. Et enfin, nous terminerons cette partie par un cas particulier qui présente de multiples altérations au niveau du gène codant la protéine TWIST1. Il s'agit d'un patient atteint d'un carcinome épidermoïde du poumon (0,6% soit 1 cas sur 178). Ce patient est porteur d'une amplification du gène codant TWIST1 et d'une mutation faux sens P136H au niveau de la séquence protéique codante (Figure 40A).

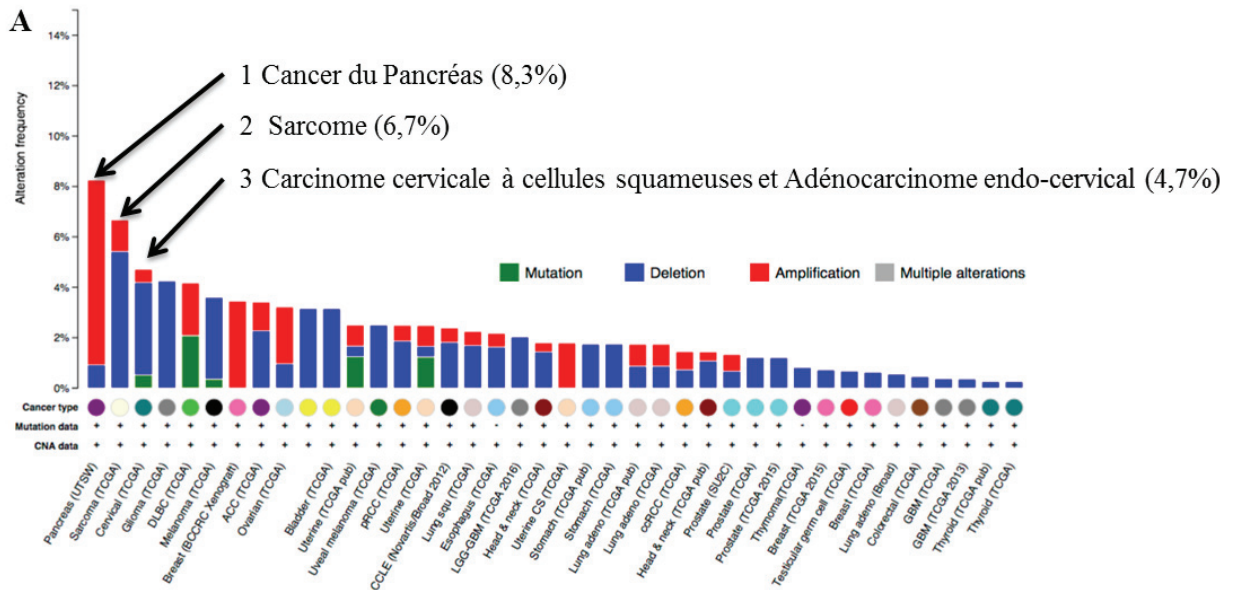
1.4.1.3 Les mutations de *TWIST2* retrouvées dans les cancers

Tout comme pour *TWIST1*, cent vingt et une études concernant les altérations de *TWIST2* dans les cancers ont été rapportées sur la base de données cBioPortal (Cerami et al., 2012; J. Gao et al., 2013; Genomics). Les cancers présentant le plus d'altérations de *TWIST2* sont le cancer du pancréas (8,3% soit 9 cas sur 109), le sarcome (6,7% soit 16 cas sur 240), le carcinome cervical à cellules squameuses et l'adénocarcinome endo-cervical (4,7% soit 9 cas sur 191). Cependant il existe plusieurs types d'altérations du gène codant pour *TWIST2* dans les cancers (Figure 41).

Il y a tout d'abord des délétions du gène *TWIST2*. En effet, contrairement à *TWIST1* pour lequel les altérations majoritaires sont des amplifications, dans ce cas ce sont les délétions du gène *TWIST2* qui sont majoritairement associées aux cancers. Ainsi les trois types de cancers les plus concernés par ces délétions sont le sarcome (5,4% soit 13 cas sur 240), les gliomes de bas grade (4,2% soit 12 cas sur 283), le carcinome cervical à cellules squameuses et l'adénocarcinome endo-cervical (3,7% soit 7 cas sur 191 ; Figure 41A et B).

Ensuite il y a des amplifications du gène codant pour *TWIST2*. Celles-ci se retrouvent principalement dans le cancer du pancréas (7,3% soit 8 cas sur 109), le carcinome invasif du sein (3,4% soit 1 cas sur 29) et enfin le cystadénocarcinome séreux ovarien (2,3% soit 7 cas sur 311 ; Figure 41A et B).

Enfin, il y a les mutations qui sont uniquement des mutations faux sens dans le cas des altérations du gène codant la protéine *TWIST2*. Dans les cancers, les mutations associées à *TWIST2* sont beaucoup moins nombreuses que celles associées à *TWIST1* (6 contre 23 mutations). Ces mutations sont retrouvées dans le lymphome diffus à grandes cellules B (2,1% soit 1 cas sur 48), le carcinome endométrioïde du corps utérin (1,3% soit 3 cas sur 240), le carcinome cervical à cellules squameuses et l'adénocarcinome endo-cervical (0,5% soit 1 cas sur 191 ; Figure 41A et C).



B

Types d'altérations retrouvées		
Délétion	Amplification	Mutation
1 Sarcome	Cancer du pancréas	Lymphome diffus à grandes cellules B
2 Gliomes de bas grade	Carcinome invasif du sein	Cancers du poumon à petites cellules
3 Carcinome cervicale à cellules squameuses et de l'adénocarcinome endo-cervical	Cystadénocarcinome séreux ovarien	Adénocarcinomes oesophagiens

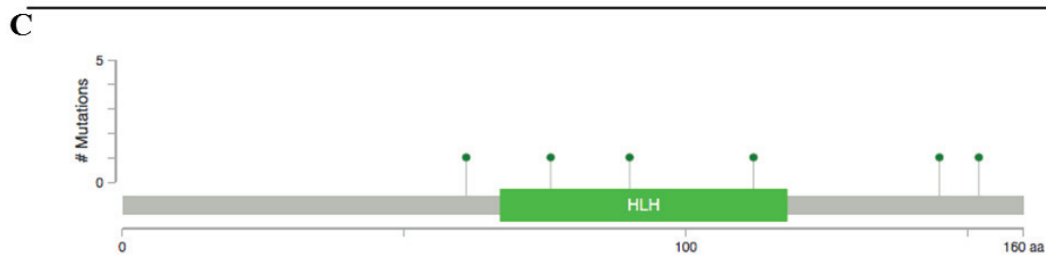


Figure 41: analyse des altérations somatiques de TWIST2 sur cBioPortal.

D'après les données de cBioPortal (J. Gao et al., 2013; Genomics).

A: Fréquences des altérations du gène *TWIST2* selon le type de cancer. B: Les trois premiers types de cancers concernés par les différentes altérations possibles. C: Répartition des mutations rapportées comme altérations du gène codant la protéine TWIST2.

2 TWIST d'un point de vue fonctionnel

2.1 Fonction embryonnaire de TWIST de la drosophile aux vertébrés

2.1.1 TWIST chez la drosophile

Le gène *twist* est requis pour de multiples étapes du développement embryonnaire, notamment la gastrulation, la formation du mésoderme et la migration des cellules de la crête neurale (Barnes & Firulli, 2009; Castanon & Baylies, 2002; Leptin, 1991; Thisse et al., 1987). L'invalidation du gène zygotique *twist* chez la Drosophile conduit à une gastrulation incomplète, une absence de formation du mésoderme, une torsion de l'abdomen (d'où la nomenclature de cette protéine TWIST) et à la mort à la fin de l'embryogenèse (Grau, Carteret, & Simpson, 1984; Simpson, 1983; Thisse et al., 1987) (Leptin, 1991). Le gène *twist* a été identifié chez la drosophile comme un gène nécessaire à la structuration dorso-ventrale au cours de l'embryogenèse (Nüsslein-Volhard, 1984; Simpson, 1983). En effet au stade blastula, *Twist* est initialement exprimée dans les cellules les plus ventrales de l'embryon de manière uniforme, sous le contrôle d'un gradient nucléaire de la protéine Dorsal et est nécessaire pour l'activation d'un certain nombre de régulateurs spécifiques du mésoderme (Barnes & Firulli, 2009; Ip, Park, Kosman, Yazdanbakhsh, & Levine, 1992; Leptin, 1991). Le gène *twist* de la Drosophile est également nécessaire à l'activation de *snail* qui est crucial pour la gastrulation et pour le maintien de l'expression de *twist* (Ip et al., 1992; Leptin, 1991). Toutefois l'implication de *twist* semble aujourd'hui bien plus vaste puisque des études de CHIP ont permis l'identification de plus de 500 gènes cibles de TWIST incluant des gènes nécessaires au développement du muscle, de la prolifération cellulaire, de la morphogenèse et de la migration cellulaire (Sandmann et al., 2007)(Figure 42).

2.1.2 TWIST chez les vertébrés

Chez les vertébrés, il existe deux gènes *Twist*, *Twist-1* et *Twist-2* (L. Li, Cserjesi, & Olson, 1995; Puisieux et al., 2006)(Figure 42). Alors que le gène *twist* de drosophile est nécessaire à la gastrulation et au développement du mésoderme, l'expression de *Twist1* chez la souris n'est pas détectée tant que la gastrulation n'est pas initiée (Barnes & Firulli, 2009). Au cours de l'embryogenèse chez la souris, l'expression de la protéine TWIST1 est détectée dans les somites, dans les bourgeons des membres, dans les cellules mésenchymateuses de la tête et dans les arcs branchiaux (Wolf et al., 1991).

TWIST1 inhibe le développement musculaire en séquestrant les protéines E qui forment des complexes myogéniques en s'associant avec les facteurs de transcription bHLH spécifiques du muscle squelettique MYOD1 et en inhibant la transactivation de MEF2 (Myocyte-specific Enhancer Factor 2) (Hebrok et al., 1994; Spicer et al., 1996) (Hamamori et al., 1997).

Les souris homozygotes *Twist1*^{-/-} ont une gastrulation normale mais meurent au stade embryonnaire E10,5. En effet chez la souris, la protéine TWIST1 joue un rôle dans la différenciation spatiale du mésoderme et régule à la fois la formation et le développement des dérivés de la crête neurale. Ces souris présentent également des anomalies au niveau des somites, des bourgeons des membres et de la suture du tube neural (Chen & Behringer, 1995). D'autre part ces souris montrent la présence de cellules apoptotiques au niveau des somites ce qui suggèrent que TWIST1 possède des propriétés anti-apoptotique (Chen & Behringer, 1995). Les effets anti-apoptotiques de TWIST1 ont été confirmés dans des fibroblastes embryonnaires murins (Maestro et al., 1999). En effet, TWIST1 bloque la réponse apoptotique dépendante de la voie ARF-p53 (Alternative Reading Frame) induite par c-MYC (Maestro et al., 1999). Après la naissance, l'expression de TWIST1 chute rapidement et reste confinée à quelques rares précurseurs de tissu d'origine mésodermique chez l'animal (Wolf et al., 1991).

Les souris hétérozygotes *Twist1*^{+/-} ont un phénotype plus modéré et présentent une craniosténose et une malformation des membres, un phénotype similaire à celui observé chez les patients humains atteints du syndrome de Saethre-Chotzen (Bourgeois et al., 1996; el Ghouzzi et al., 1997) (Franco, Casasnovas, Rodriguez-Medina, & Cadilla, 2011). D'autre part, la découverte de mutations ponctuelles localisées au niveau du domaine bHLH ou juste en amont du domaine WR de TWIST1 est également associée au syndrome de Saethre-Chotzen qui se caractérise par une suture prématurée des os du crâne et par une malformation des membres (Castanon & Baylies, 2002; el Ghouzzi et al., 1997; Howard et al., 1997). De plus *via* son domaine WR, TWIST1 interagit avec RUNX2 une protéine qui intervient dans la régulation du développement des ostéoblastes (Barnes & Firulli, 2009; Seto et al., 2007).

Au cours du développement, le gène *Twist2* présente un profil d'expression tissulaire proche mais distinct du gène *Twist1*, avec une accumulation au niveau du dermatome et des dérivés des arcs bronchiaux (d'où son ancien nom *Dermo-1*)(L. Li et al., 1995). Chez la souris, l'expression de TWIST2 augmente progressivement dans les somites et les bourgeons des membres au cours de l'embryogenèse. Par la suite elle se limite au perichondrium, c'est à dire au niveau du tissu conjonctif qui entoure le cartilage des os en développement. L'expression de TWIST2 augmente progressivement dans le derme et persiste jusqu'à la naissance. En revanche, à l'âge adulte, son expression chute rapidement puisque TWIST2 n'est détectable qu'en faible quantité dans le derme (L. Li et al., 1995).

Chez la souris, la délétion hétérozygote du gène *Twist2* est viable (Sosic et al., 2003). La perte de TWIST2 n'impacte pas le développement embryonnaire mais conduit à la mort des nouveau-nés au cours des premiers jours à la suite d'une augmentation de l'expression des cytokines pro-inflammatoires (Puisieux et al., 2006; Sosic et al., 2003). Cette augmentation conduit à une cachexie sévère induite par une production anormalement élevée de cytokines pro-inflammatoires due à l'incapacité de la protéine TWIST2 à réprimer l'activité du facteur de transcription NF-kB (L. Li et al.,

1995; Susic & Olson, 2003; Susic et al., 2003). Ces souris présentent également des lésions bitemporales et des anomalies oculaires, un phénotype similaire à celui observé chez les patients atteints du syndrome de Setleis caractérisé par des mutations non-sens de TWIST2 (Tukel et al., 2010). Les souris double hétérozygotes pour *Twist1* et *Twist2* montrent des phénotypes similaires à ceux obtenus avec les souris délétées de manière homozygote pour le gène *Twist2* (Puisieux et al., 2006).

Comme nous venons de la voir, les pertes hétérozygotes de *Twist1* et *Twist2* conduisent à des phénotypes bien différents avec un impact respectivement au niveau du développement des os et de la peau. Ceci confirme que, malgré leur homologie de séquence, les deux protéines possèdent des profils d'expression et des fonctions distinctes au cours du développement embryonnaire (Franco et al., 2011; Tukel et al., 2010).

	Espèce	Génotype	Phénotype	Références bibliographiques
TWIST1	Drosophile	<i>Tw1^{-/-}</i>	Gastrulation incomplète, absence de formation du mésoderme, torsion de l'abdomen, léthal en fin d'embryogenèse	(Grau et al., 1984; Simpson, 1983; Thisse et al., 1987) (Leptin, 1991)
	Souris	<i>Twist1^{-/-}</i>	Léthal à E10,5, gastrulation normale mais anomalies des somites, des bourgeons de membres et de la suture du tube neural, présence de cellules apoptotiques au niveau des somites	(Chen & Behringer, 1995).
	Souris	<i>Twist1^{+/-}</i>	Craniosténose, malformation des membres	(el Ghouzzi et al., 1997) (Bourgeois et al., 1996) (Franco et al., 2011)
	Homme	TWIST1+/-	Syndrome de Saethre-Chotzen : suture prématuré des os du crâne et par une malformation des membres	(el Ghouzzi et al., 1997) (Howard et al., 1997) (A. B. Firulli & Conway, 2008)
TWIST2	Souris	<i>Twist2^{-/-}</i>	Viable	(Susic et al., 2003) (Puisieux et al., 2006)
	Souris	<i>Twist2^{+/-}</i>	Pas d'impact sur le développement embryonnaire, lésions bitemporales et des anomalies oculaires, cachexie sévère dans les premiers jours de vie conduisant à la mort du nouveau-né	(Susic & Olson, 2003) (Susic et al., 2003) (Tukel et al., 2010) (L. Li et al., 1995)

Figure 42: Récapitulatif des phénotypes observés selon le génotype et les espèces.

2.2 Fonction de TWIST à l'âge adulte

De nombreuses études montrent chez l'animal que l'expression des protéines TWIST chute rapidement à la fin du développement embryonnaire pour être exprimée à l'âge adulte au niveau de rares précurseurs de tissus d'origine mésodermique (Isenmann et al., 2009; Wolf et al., 1991). Chez l'Homme, l'expression de TWIST1 à la naissance est limitée au placenta, au cœur et aux muscles

squelettiques qui sont dérivés du mésoderme (S. M. Wang et al., 1997). En effet à l'âge adulte, TWIST jouerait un rôle dans la croissance et la régénération musculaire en inhibant la différenciation des cellules musculaires squelettiques (Currie & Bate, 1991; Leshem, Spicer, Gal-Levi, & Halevy, 2000). De plus TWIST1 interviendrait également dans le maintien de l'homéostasie de la pulpe dentaire en régulant la différenciation des odontoblastes (Galler et al., 2007; Meng et al., 2015). TWIST1 est également exprimé dans le tissu adipeux comprenant la graisse brune et la graisse blanche impliquant donc une fonction de TWIST1 dans la balance énergétique. TWIST1 jouerait un rôle dans la régulation de la thermogenèse de la graisse brune (Pan et al., 2009). En outre, TWIST1 chez l'homme pourrait, contrairement aux observations faites chez la souris, être un régulateur positif de l'oxydation des acides gras dans les adipocytes blancs humains (Pettersson et al., 2010). TWIST1 a également été impliqué dans l'inflammation associée à l'obésité dans les tissus adipeux blancs (*via* peut être PPAR γ ; Peroxisome Proliferator-Activated Receptor gamma) avec une expression diminuée de TWIST1 chez les patients obèses (Ma et al., 2014; Pettersson et al., 2011). D'autre part, TWIST1 est également exprimé de manière dépendante de NF κ B, de l'Interleukine 12 et de STAT4 (Signal transducer and activator of transcription 4) dans les lymphocytes T « helpers 1 » LTH1 (Niesner et al., 2008). Dans ce contexte TWIST1 se comporte comme un facteur limitant la réponse pro-inflammatoire des LTH1 en réponse à leur activation (Niesner et al., 2008).

Comme nous l'avons vu précédemment, l'expression de TWIST2 à l'âge adulte est retrouvée en faible quantité au niveau du derme (L. Li et al., 1995). TWIST2 a été identifié dans les progéniteurs myéloïdes comme un inhibiteur de leur différenciation et leur prolifération (Sharabi et al., 2008). TWIST2 a également été trouvé dans les cellules myéloïdes matures en inhibant la production de cytokines pro-inflammatoires et en favorisant la production de cytokine anti-inflammatoire (Sharabi et al., 2008).

3 Impact des modifications post-traductionnelles sur les fonctions de TWIST1

Le niveau d'expression de la protéine TWIST1 est sous le contrôle de modifications post-traductionnelles que sont la phosphorylation et l'ubiquitination comme le montre la différence de concordance entre l'expression de l'ARNm (Acide RiboNucléique messenger) et de la protéine (Gitelman, 1997). En effet des analyses comparatives entre le taux d'ARNm et le taux de protéine TWIST1 ne sont pas comparables. En effet au niveau embryonnaire chez la souris, une forte expression de l'ARN (Acide RiboNucléique) dans le mésoderme pré-somitique, les somites épithéliaux, et le mésoderme antérieur a été détectée alors que la protéine restait indétectable (Gitelman, 1997). De plus le taux d'ARN est maximal à sept jours post-coïtum alors que celui de la protéine est maximal le neuvième jour au cours du développement embryonnaire de la souris. Au-delà de la phosphorylation et de l'ubiquitination sur lesquelles nous reviendrons, TWIST1 pourrait être acétylé au niveau du résidu de lysine 73 *via* d'après des études de prédiction (Hornbeck et al., 2012).

3.1 Phosphorylation

La phosphorylation est une modification post-traductionnelle consistant en l'addition d'un groupement phosphoryle PO_3^{2-} . Cette modification post-traductionnelle constitue un des mécanismes de régulation les plus importants. La phosphorylation se fait par l'intermédiaire de protéines kinases spécifiques alors que les protéines dites Phosphatases, quant à elles, ont une action de déphosphorylation. Ces protéines aux actions antagonistes sont régulées par des signaux extérieurs comme par exemple les hormones, les cytokines et les facteurs de croissances.

D'autre part, les mutations du syndrome de Saethre-Chotzen adjacentes aux sites de phosphorylation peuvent influencer la phosphorylation de la protéine TWIST1, modifier ainsi la balance homo/hétérodimère et modifier la fonction de la protéine (el Ghouzzi et al., 1997; B. A. Firulli et al., 2005; Howard et al., 1997). En effet, l'étude de mutations ponctuelles à proximité des résidus de thréonine 121 et de sérine 123 a montré un impact non négligeable de ces mutations au niveau de la phosphorylation par le PKA (B. A. Firulli et al., 2005; Lu et al., 2011). De plus, les mutations non-sens conduisant à la formation d'une protéine TWIST1 tronquée pourraient conduire à une dégradation de TWIST1 suite à une phospho-régulation du résidu 68 s'il est présent dans la protéine tronquée.

TWIST1 possède de nombreux sites de phosphorylation répartis tout au long de sa séquence. La sérine 45, la sérine 78, la sérine 99, la sérine 102, la tyrosine 103 et la thréonine 108 sont des sites de phosphorylations déterminés par spectrométrie de masse (Figure 43)(Klammer et al., 2012; Trinidad et al., 2012; Wei et al., 2015; Williams et al., 2016; H. Y. Wu et al., 2009).

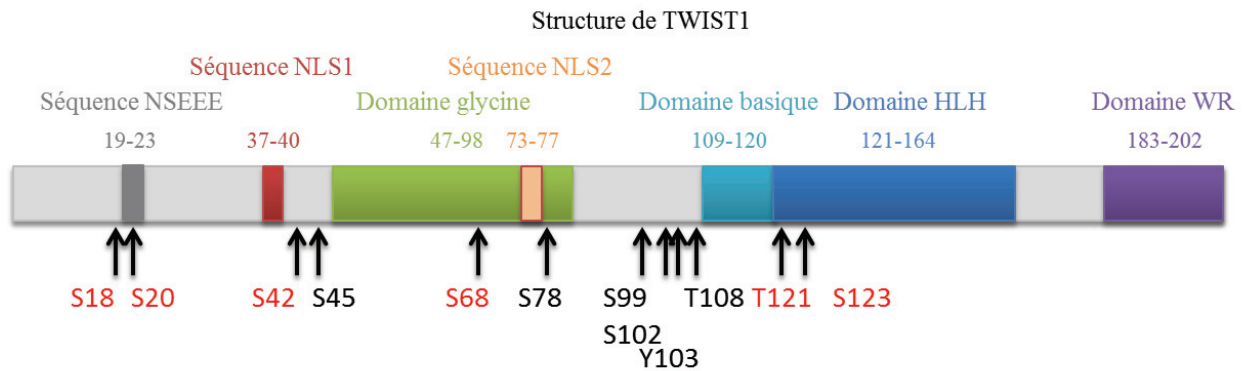


Figure 43 : sites de phosphorylation de TWIST1 (Hornbeck et al., 2015).

3.1.1 Phosphorylation des Sérines 18 et 20 de TWIST1

TWIST1 possède de nombreux sites de phosphorylation potentiels dont le premier se situe entre les résidus 18 et 21 et comprend deux résidus de sérine en position 18 et 20 (Figure 44). Ce site est un motif consensus de type « SNSE » conservé au sein de différentes espèces de TWIST1 et caractéristique des substrats de CDK2 (Caséine Kinase 2), une sérine/thréonine Kinase ubiquitaire (Su, Xie, Sano, & Myers, 2011). Des études de la stabilité protéique de TWIST1 sauvage ou mutant mimant, soit un état déphosphorylé de TWIST1 S18A/S20A, soit un état phosphorylé de TWIST1 S18D/S20D, ont été réalisées. Celles-ci ont montré que le temps de demi-vie de TWIST1 sauvage était de 4,69 heures alors que le mutant S18D/S20D avait un temps de demi-vie beaucoup plus long (9,62 heures) contrairement au mutant S18A/S20A qui avait un temps de demi-vie beaucoup plus court (2,45 heures). La phosphorylation de TWIST1 par CDK2 au niveau des résidus de sérine 18 et 20 permet la stabilisation de TWIST1 (Su et al., 2011). D'autre part la surexpression du mutant S18D/S20D semble augmenter la motilité cellulaire contrairement à TWIST1 sauvage ou au mutant S18A/S20A. La phosphorylation des deux résidus de sérine 18 et 20 par la CDK2 se fait en réponse à l'interleukine 6 (IL-6) ainsi l'IL-6 stabilise TWIST1 et favorise la motilité cellulaire (Su et al., 2011).

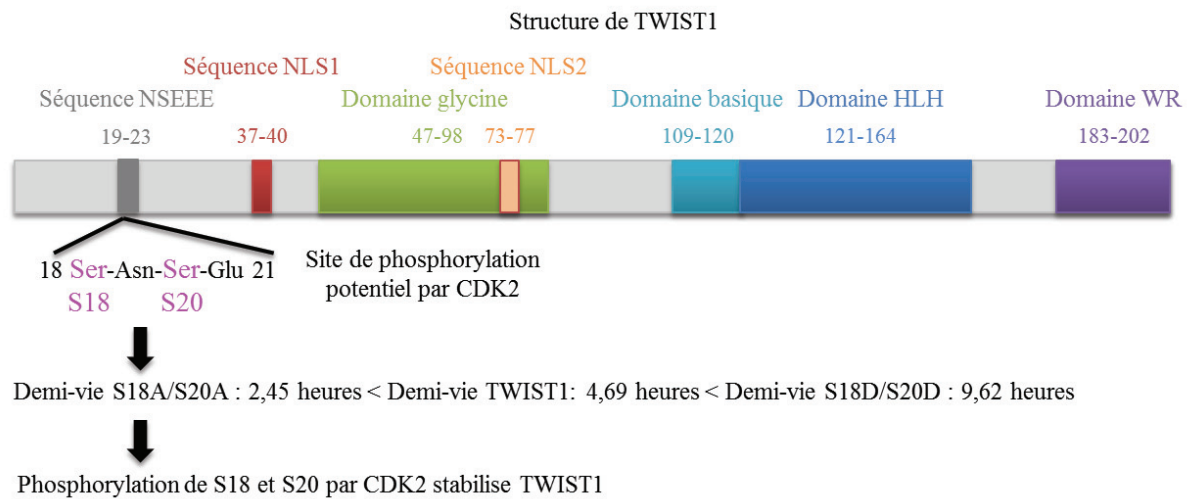


Figure 44 : phosphorylation de TWIST1 par CDK2 permet la stabilisation de la protéine TWIST1.

3.1.2 Phosphorylation de la Sérine 42 de TWIST1

TWIST1 est une protéine capable d'activer la transcription d'AKT2 en se liant à une E-box de son promoteur (RAC-alpha serine/threonine-protein kinase (PKB))(G. Z. Cheng et al., 2007). D'autre part, il semble exister une co-expression élevée de TWIST1 et d'AKT2 dans des échantillons de cancer du sein à des stades avancés (G. Z. Cheng et al., 2007). PKB/AKT correspond à la protéine kinase B dont l'expression est ubiquitaire. Cette kinase est impliquée dans la régulation de nombreux mécanismes cellulaires comme la croissance cellulaire, la survie, la prolifération, l'apoptose, le métabolisme et l'angiogenèse (Risso, Blaustein, Pozzi, Mammi, & Srebrow, 2015). Des mutations en alanine des résidus de sérine en position 42, 45, 123 et du résidu de thréonine en position 121 de la protéine TWIST1 ont permis de montrer que seules les mutations des résidus en positions 42 et 123 entraînaient une perte de la phosphorylation de la protéine TWIST1 par la PKB *in vitro* Figure 45)(Vichalkovski et al., 2010). En revanche, *in vivo* il semblerait que la PKB phosphoryle préférentiellement la S42 (Vichalkovski et al., 2010). De plus, cette phosphorylation de la S42 de TWIST1 par la PKB semble importante pour les effets anti-apoptotiques de TWIST1 sans toutefois impacter sur la localisation de la protéine. Ainsi TWIST1 phosphorylé favoriserait la survie cellulaire notamment dans le contexte tumoral. TWIST1 semble apparaître préférentiellement sous forme phosphorylé au niveau du résidu de S42 dans de nombreux cancers comme notamment les cancers du sein, du rectum et du colon (Vichalkovski et al., 2010).

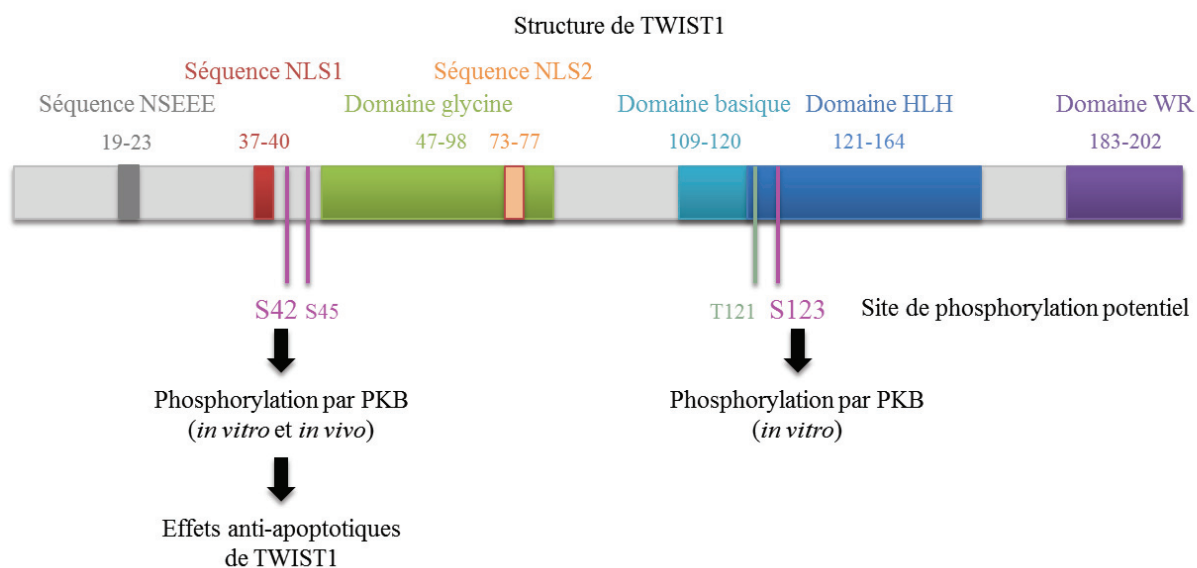


Figure 45: phosphorylation de la sérine 42 par PKB favorise les effets anti-apoptotiques de TWIST1.

3.1.3 Phosphorylation de la Sérine 68 de TWIST1

La sérine 68 de TWIST1 est un site majeur de phosphorylation de la protéine TWIST1. L'analyse de la séquence TWIST1 montre un site spécifique de phosphorylation par les MAPK de type Pro-X-Ser/Thr-Pro en position 66 : Pro-Gly-Ser-Pro (Figure 46)(Hong et al., 2011; Widmann, Gibson, Jarpe, & Johnson, 1999). La sérine 68 de TWIST1 inclue dans une séquence consensus de phosphorylation peut être phosphorylée par différents membres de la famille des MAPK : p38, JNK (c-Jun N-terminal Kinase), ERK1 and ERK2 (Extracellular signal Regulated Kinase 1 et 2) aussi bien *in vitro* qu'*in vivo* (Hong et al., 2011). Des études ont montré que la demi-vie du mutant mimant une hypo-phosphorylation S68A est de 1,8 heures alors que le temps de demi-vie de la protéine sauvage est de 6,2 heures. Ainsi, l'activation de la voie des MAPK ou le traitement au TGF β permettent la phosphorylation de TWIST1 et donc l'augmentation du temps de demi-vie de la protéine TWIST1 (Hong et al., 2011). Cette phosphorylation protège TWIST1 contre son ubiquitination et sa dégradation par le protéasome mais n'est toutefois pas nécessaire ni à la localisation nucléaire, ni à la dimérisation de TWIST1 (Figure 46)(Hong et al., 2011).

Toutefois la phosphorylation au niveau de la sérine 68 ne semble pas être le seul site impliqué dans la stabilisation de TWIST1. En effet, comme nous l'avons vu précédemment, les phosphorylations des résidus de sérine 18, 20 et 42 permettent également de stabiliser la protéine TWIST1 (Su et al., 2011; Vichalkovski et al., 2010).

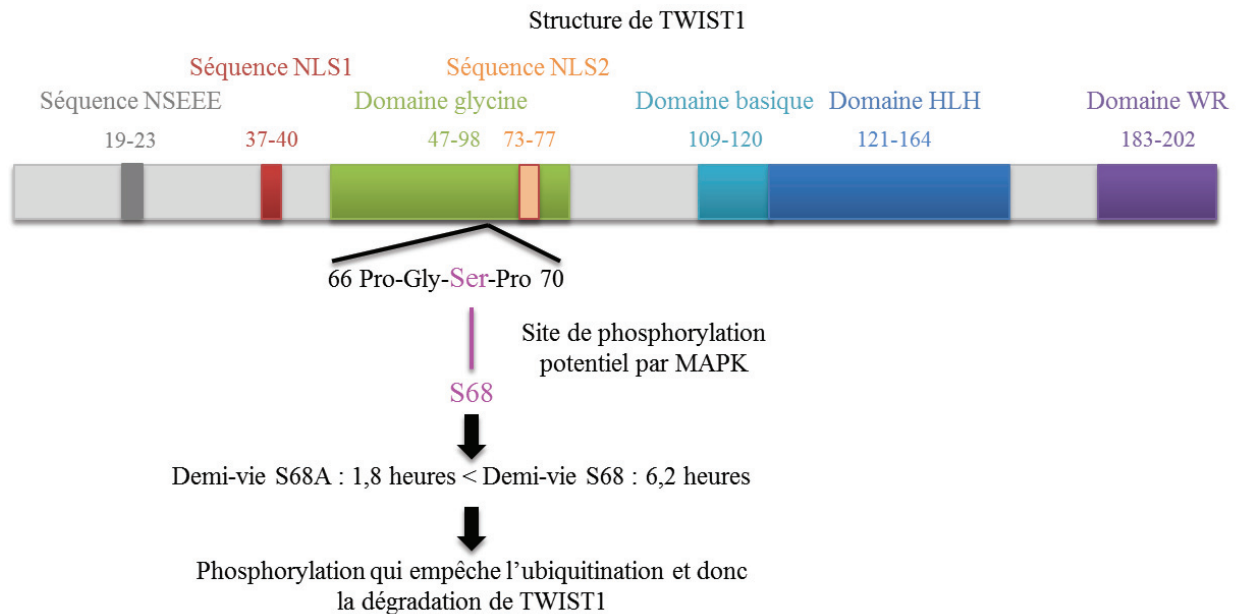


Figure 46 : phosphorylation de la sérine 68 stabilise la protéine TWIST1.

3.1.4 Phosphorylation des résidus de Thréonine 121 et Sérine 123 de TWIST1

Les résidus de T121 et de S123 situés au niveau de l'hélice α 1 de TWIST1 sont des résidus très conservés au niveau du domaine bHLH de facteurs de transcription de la même famille comme HAND1, HAND2 et TWIST2. Des alignements de séquences montrent que les protéines TWIST1, HAND1, TWIST2 et HAND2 possèdent dans plusieurs espèces un motif consensus de type $\frac{R}{K} X \frac{R}{K} X X \frac{S}{T}$ spécifiquement reconnu par la protéine kinase B (Lu et al., 2011). TWIST1 est, comme HAND1, phosphorylé aussi bien au niveau de la T121 que de la S123 par la PKB (Figure 47) (Lu et al., 2011). D'autre part la mutation de ces deux résidus de thréonine et de sérine en alanine ainsi que des expériences de co-immunoprécipitation ont permis de mettre en évidence leur phosphorylation par la protéine kinase A aussi bien au niveau de TWIST1 que de HAND2 (B. A. Firulli et al., 2003; B. A. Firulli et al., 2005). De plus ce double mutant T121A/S123A mimant un état déphosphorylé de TWIST1 semble favoriser l'homodimérisation de TWIST1 avec lui-même (B. A. Firulli et al., 2005). A l'inverse la double mutant T121D/S123D mimant un état phosphorylé semble favoriser l'hétérodimérisation de TWIST1 avec E47 (B. A. Firulli et al., 2005).

Il a également été montré que les mutations associées au syndrome de Saethre-Chotzen localisées à proximité de ces deux résidus de sérine et thréonine affectaient leur phosphorylation par la PKA (B. A. Firulli et al., 2005). De plus, l'état de phosphorylation de HAND1 impacterait le choix de son partenaire de dimérisation (Centonze, Firulli, & Firulli, 2004; B. A. Firulli, Hadzic, McDaid, & Firulli, 2000; B. A. Firulli et al., 2003). Les mutations de TWIST1 à proximité des résidus T121 et S123 associées au syndrome de Saethre-Chotzen impacteraient d'une part la phosphorylation des résidus 121 et 123 par la PKA comme nous l'avons vu précédemment mais impacteraient également la dimérisation de TWIST1 avec MYOD1 (B. A. Firulli et al., 2005; Hamamori et al., 1997). Ce qui laisse penser en effet que l'interaction de TWIST1 et MYOD1 peut être régulée par la phosphorylation. Ainsi l'état de phosphorylation de la protéine TWIST1 pourrait guider le choix de son partenaire de dimérisation tout comme cela a été précédemment montré pour HAND1, moduler ainsi sa liaison à l'ADN et influencer directement sur sa fonction (Figure 47)(B. A. Firulli et al., 2005; B.

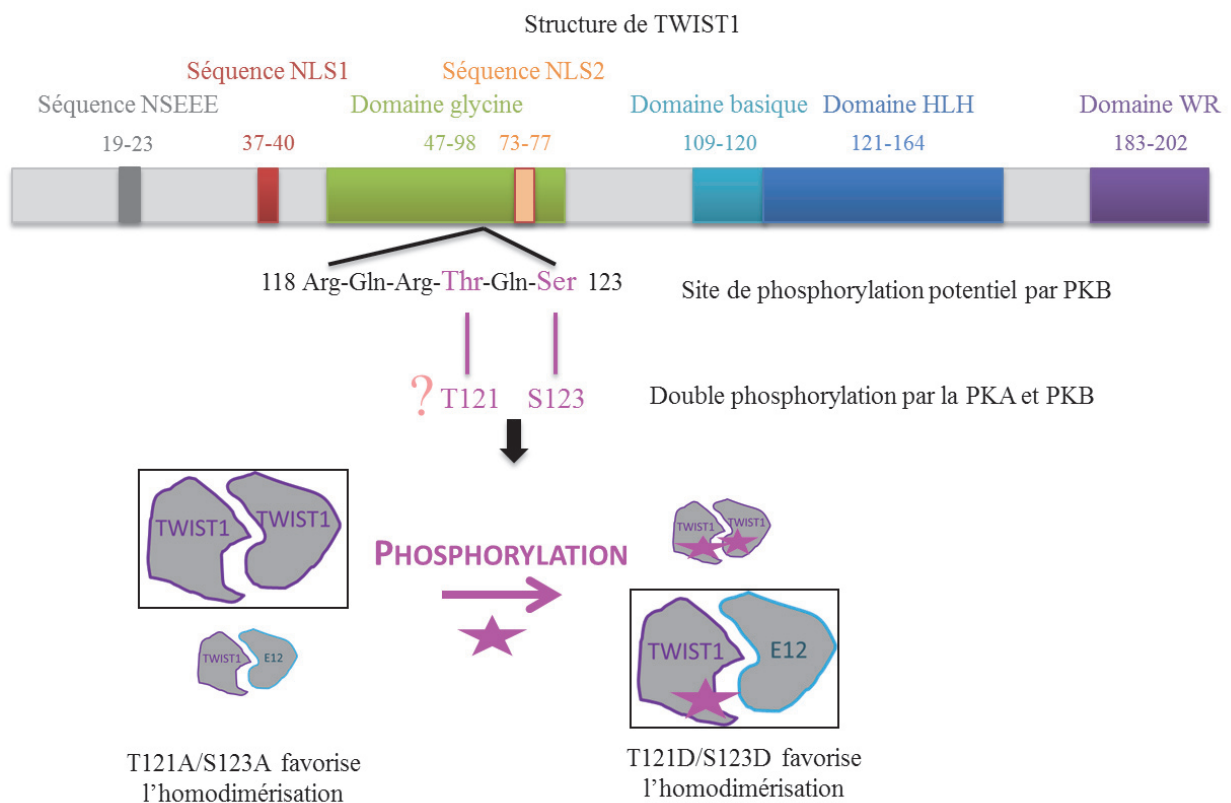


Figure 47: impact de la phosphorylation de la thréonine 121 et de la sérine 123 sur la balance homo-hétérodimère.

A. Firulli et al., 2007; Hamamori et al., 1997).

En accord avec ces données, il a été montré, dans des cellules cancéreuses de la prostate, que la phosphorylation de TWIST1 au niveau des résidus de thréonine et sérine respectivement en position 121 et 123 *via* la voie PI3K/AKT/mTOR favorisait la formation de l'hétérodimère TWIST1/E12 et conférait à ces cellules des capacités migratoires (Gajula et al., 2015). De manière intéressante, les

études de Vichalkovski semblent montrer *in vitro* que la phosphorylation *via* la PKB ne concerne non pas les résidus de thréonine 121 et de sérine 123 mais uniquement le résidu de sérine 123 (Vichalkovski et al., 2010).

3.2 Ubiquitination

L'ubiquitination est une modification post-traductionnelle importante pour la régulation de la stabilité de la protéine TWIST1 puisque l'utilisation d'un inhibiteur du protéasome entraîne l'inhibition de la dégradation de TWIST1 (Demontis et al., 2006; Lander, Nordin, & LaBonne, 2011). En effet, la TWIST-box au niveau de l'extrémité C-terminale de TWIST1 serait importante pour la dégradation de la protéine TWIST1 (Diaz, Vinas-Castells, & Garcia de Herreros, 2014; Lander et al., 2011). Dans un contexte cellulaire normal, TWIST1 est une protéine qui a une demi-vie plus courte avant d'être poly-ubiquitinylée et dégradé par le protéasome 26S. La poly-ubiquitination consiste en l'addition d'une ubiquitine, qui est une protéine de soixante-seize acides aminés, à la protéine cible. Ce mécanisme implique trois enzymes : l'Enzyme E1 (activation de l'ubiquitine *via* un ATP (Adénosine TriPhosphate)), l'Enzyme E2 (Transfert l'ubiquitine activée sur l'enzyme E2) et l'Enzyme E3 (transfert de l'ubiquitine activé vers la protéine cible) (Diaz et al., 2014). Lorsque la protéine cible est poly-ubiquitinylée par un branchement de l'ubiquitine *via* la lysine 48 alors elle peut être reconnue par le protéasome pour être dégradée (Diaz et al., 2014). Pour TWIST1, l'enzyme E3 impliquée serait de type RING-finger SCF (Skp1-Cullin1-F-box protein) avec Ppa (Partner of paired) comme protéine F (Lander et al., 2011). Par ailleurs, la protéine TWIST1 peut être dégradée par autophagie. Dans une cellule dépourvue d'autophagie, une accumulation de p62 (sequestosome 1(SQSTM1)) pourrait conduire à l'interaction de p62 avec TWIST1 poly-ubiquitinylée et ainsi à la stabilisation de TWIST1 en inhibant sa dégradation par le protéasome (Qiang et al., 2014).

4 Des fonctions embryonnaires aux fonctions oncogéniques de TWIST1

4.1 Forte expression de TWIST1 dans de nombreux cancers

Comme nous l'avons vu précédemment, TWIST1 est un facteur de transcription impliqué dans le développement embryonnaire (Chen & Behringer, 1995; Puisieux et al., 2006). De nombreuses études montrent qu'au-delà de son rôle embryonnaire, TWIST1 joue un rôle au cours du développement tumoral et métastatique (Kwok et al., 2005; Puisieux et al., 2006; J. Yang et al., 2004). A l'inverse des tissus normaux adultes où TWIST1 est peu détectable, ce facteur de transcription est ré-exprimé dans de nombreux cancers humains, comme le neuroblastome (Valsesia-Wittmann et al., 2004), le cancer du sein (J. Yang et al., 2004), le cancer de la prostate (Kwok et al., 2005; Raatikainen, Aaltomaa, Palvimo, Karja, & Soini, 2015; Yuen et al., 2007), le carcinome hépatocellulaire (T. K. Lee et al., 2006; M. H. Yang et al., 2009), le cancer du pancréas (Satoh et al., 2008), le carcinome à cellules squameuses de l'œsophage (Sasaki et al., 2009), le cancer de la vessie (Wallerand et al., 2010; Zhang et al., 2007), le cancer gastrique (Feng et al., 2009), et les cancers de la voie Oto-Rhino-Laryngologie (K. J. Wu & Yang, 2011; M. H. Yang et al., 2010).

cBioPortal est une banque de données de la génomique des cancers. Les études cBioPortal nous ont permis de voir que les mutations de TWIST1 dans les cancers étaient des mutations rares (Cerami et al., 2012; J. Gao et al., 2013; Genomics). De plus ce sont des mutations qui ne sont pas systématiquement associées au syndrome de Saethre-Chotzen et donc ne conduisent pas forcément à une perte fonctionnelle de la protéine TWIST1 surexprimée. Ces résultats montrent que TWIST1 joue un rôle clef dans le contexte cancéreux et qu'il est très rarement la cible de mutations conduisant à une inhibition fonctionnelle.

4.2 Les fonctions oncogéniques de TWIST1

TWIST1 est un facteur de transcription qui intervient à plusieurs étapes du développement tumoral. En effet il joue un rôle au cours de l'initiation et de la progression tumorale ainsi qu'au cours de la cascade métastatique. Une partie du potentiel oncogénique de TWIST1 consiste à neutraliser conjointement les deux voies onco-suppressives dépendantes des protéines Rb et P53 et à coopérer par ce biais avec différentes onco-protéines dans la transformation maligne (Figure 48) (Ansieau et al., 2008; Maestro et al., 1999; Valsesia-Wittmann et al., 2004; T. Wang et al., 2014). D'autres effets délétères sont associés à TWIST1, comme l'induction d'une instabilité chromosomique, en lien avec l'inactivation de la protéine P53, et un effet pro-angiogénique (Mironchik et al., 2005). En favorisant la conversion de cellules épithéliales en cellules mésenchymateuses *via* la transition épithélio-mésenchymateuse (TEM/EMT), la protéine TWIST1 facilite également la conversion maligne, accorde aux cellules un potentiel d'auto-renouvellement ainsi que des propriétés migratoires et

invasives, facilitant ainsi leur dissémination (Mani et al., 2008; Morel et al., 2012; Morel et al., 2008; J. Yang et al., 2004). Mais TWIST1 semble jouer un rôle également dans l'augmentation de la résistance aux traitements de chimiothérapie (G. Z. Cheng et al., 2007; Q. Q. Li et al., 2009).

4.2.1 TWIST1/E12 est une forme oncogénique de TWIST1

Des données récentes suggèrent que TWIST1/E12 est la forme pro-métastatique de TWIST1 dans le contexte tumoral prostatique (Gajula et al., 2015). Pour rappel, E12 est un facteur de transcription de la famille des bHLH, comme E47. E12 et E47 sont des protéines issues de l'épissage alternatif du gène E2A. En effet ce sont des protéines de Classe I qui ont la caractéristique d'être ubiquitaires et qui ont la capacité de former des hétérodimères en s'associant avec TWIST1 (El Ghouzzi et al., 2000; Gajula et al., 2015; Spicer et al., 1996).

4.2.2 TWIST1 et mécanismes de sauvegarde

Comme nous l'avons vu très succinctement au cours de l'introduction des fonctions oncogéniques de TWIST1, une partie du potentiel oncogénique de TWIST1 passe par l'inhibition des mécanismes de sauvegarde, que sont la sénescence et l'apoptose, dépendants des protéines Rb et P53. Cette inhibition permet aux cellules d'échapper à l'apoptose et à la sénescence, induit tous deux suite à une activation mitogénique (Ansieau et al., 2008; Maestro et al., 1999; Valsesia-Wittmann et al., 2004).

La sénescence et l'apoptose sont des barrières naturelles permettant de contenir l'émergence de cellules dotées d'un potentiel de prolifération exacerbé au sein de lésions précancéreuses. La neutralisation de ces systèmes de protection permet l'amplification de cellules à risque, favorisant l'initiation tumorale, et contribue à leur survie au cours des différentes étapes de la progression tumorale (Bartkova et al., 2005; Braig et al., 2005; Collado & Serrano, 2005; Gorgoulis et al., 2005). L'altération des voies dépendantes des protéines onco-suppressives Rb et P53 joue un rôle déterminant dans les mécanismes d'échappement à ces systèmes de sauvegarde, offrant ainsi une explication rationnelle à l'inactivation de ces protéines onco-suppressives et à la coopération oncogénique observée entre les onco-protéines MYC, RAS (RAr Sarcoma), B-RAF ou E2F1 (Larsen, 2004; Lowe et al., 2004).

TWIST1 est une protéine codée par le gène embryonnaire *TWIST1* réactivé de manière aberrante notamment dans le cadre du neuroblastome. TWIST1 neutralise l'activité pro-apoptotique de la protéine N-MYC et coopère avec la protéine mitogénique dans la neuroblastomagenèse (Ansieau et al., 2008; Maestro et al., 1999; Puisieux et al., 2006; Valsesia-Wittmann et al., 2004). En effet, au-delà de son potentiel oncogénique, le facteur de transcription c-MYC constitue également un puissant inducteur d'apoptose (Evan et al., 1992). La transformation cellulaire par c-MYC nécessite donc la

mise en place de mécanismes susceptibles de neutraliser son activité pro-apoptotique telle l'inactivation directe ou indirecte de gènes suppresseurs de tumeurs. L'analyse du mécanisme d'inactivation de l'apoptose induite en réponse à c-MYC par les gènes TWIST a montré qu'il repose sur une inactivation de la voie onco-suppressive ARF-P53, conférant aux protéines TWIST une activité pro-oncogénique (Maestro et al., 1999). Mais l'altération de l'activité des protéines P53 ne se limite pas à l'inactivation de la voie ARF. Nous avons vu précédemment que la protéine TWIST1 est capable d'interagir directement avec la protéine P53 ce qui *in fine* empêche la fixation de la protéine P53 sur ses éléments de réponse et ainsi annihile l'induction des gènes cibles en aval (Shiota et al., 2008; Stasinopoulos et al., 2005). TWIST1 inhibe les modifications post-traductionnelles stabilisatrices de P53, notamment la phosphorylation de la sérine 20 de P53 (Stasinopoulos et al., 2005). L'interaction directe entre la TWIST-box de TWIST1 et P53 inhibe la phosphorylation de la sérine 392 de P53 conduisant à une augmentation de l'affinité de P53 pour MDM2, ce qui facilite la dégradation de P53 médié par MDM2 (Piccinin et al., 2012). D'autre part l'interaction directe de TWIST1 et de HOX5 (Homeobox-leucine zipper protein HOX5) bloquerait également l'activité transcriptionnelle du gène TP53 puisque cela empêche la fixation activatrice de HOX5 au niveau du promoteur de P53 (Raman et al., 2000; Stasinopoulos et al., 2005). Et enfin, TWIST1 séquestre les protéines p300/CPB et pCAF *via* une interaction directe (Hamamori et al., 1999). Ces deux acétyltransférases sont nécessaires pour l'activité de transactivation induite par P53 (Hamamori et al., 1999).

Outre ces propriétés anti-apoptotiques, TWIST1 permet l'échappement à la sénescence en réponse à différents stress génotoxiques comme par exemple le peroxyde d'hydrogène et le cisplatine conduisant ainsi à la prolifération cellulaire avec une accumulation des dommages à l'ADN (Kwok et al., 2005). D'autre part, la surexpression de TWIST permet également l'échappement à la sénescence suite à une activation oncogénique (Burns et al., 2013). En effet, l'expression d'oncogène tel que RAS semble induire un mécanisme de sénescence contrairement à MYC qui induit plutôt l'apoptose (Serrano, Lin, McCurrach, Beach, & Lowe, 1997; Trost et al., 2005; Valsesia-Wittmann et al., 2004). Les protéines TWIST, en coopérant avec l'oncogène RAS, neutralisent l'induction de la sénescence et favorisent ainsi la transformation néoplasique des cellules (Ansieau et al., 2008). RAS est un proto-oncogène qui conduit à l'augmentation de p16^{INK4a} et de P53 (Serrano et al., 1997) (Burns et al., 2013). De plus, il a été montré la présence de séquences E-box au niveau du locus INK4A/ARF/INK4B contenant deux gènes avec chacun leur promoteur (P16^{INK4a} et ARF) (Hansson, Manetopoulos, Jonsson, & Axelson, 2003; Popov & Gil, 2010; Serrano et al., 1997). Celles-ci permettraient la régulation de ce locus par les facteurs de transcription tel que TWIST1 comme cela a été montré notamment pour TAL1 au niveau du promoteur p16^{INK4a} (Hansson et al., 2003). De plus il a été montré que TWIST1 activait deux protéines EZH2 (Enhancer of Zeste Homolog 2) et BMI1 (B Lymphoma

Mo-MLV insertion region 1 homolog (mouse)) qui sont des inhibiteurs des deux protéines codées par le locus INK4A/ARF/INK4B (Cakouros et al., 2012; Popov & Gil, 2010; M. H. Yang et al., 2010).

À l'appui de ces conclusions, la diminution des niveaux de TWIST1 d'expression dans différents modèles suite à un traitement par le 8-bromo-5-hydroxy-7-méthoxychrysin ou *via* l'utilisation de shRNA (small hairpin RNA or short hairpin RNA) ou de petits ARN interférents (small interfering RNA siRNA) permet de restaurer la sénescence et la mort cellulaire (Ansieau et al., 2008; Ren et al., 2013; Tran et al., 2012; Valsesia-Wittmann et al., 2004).

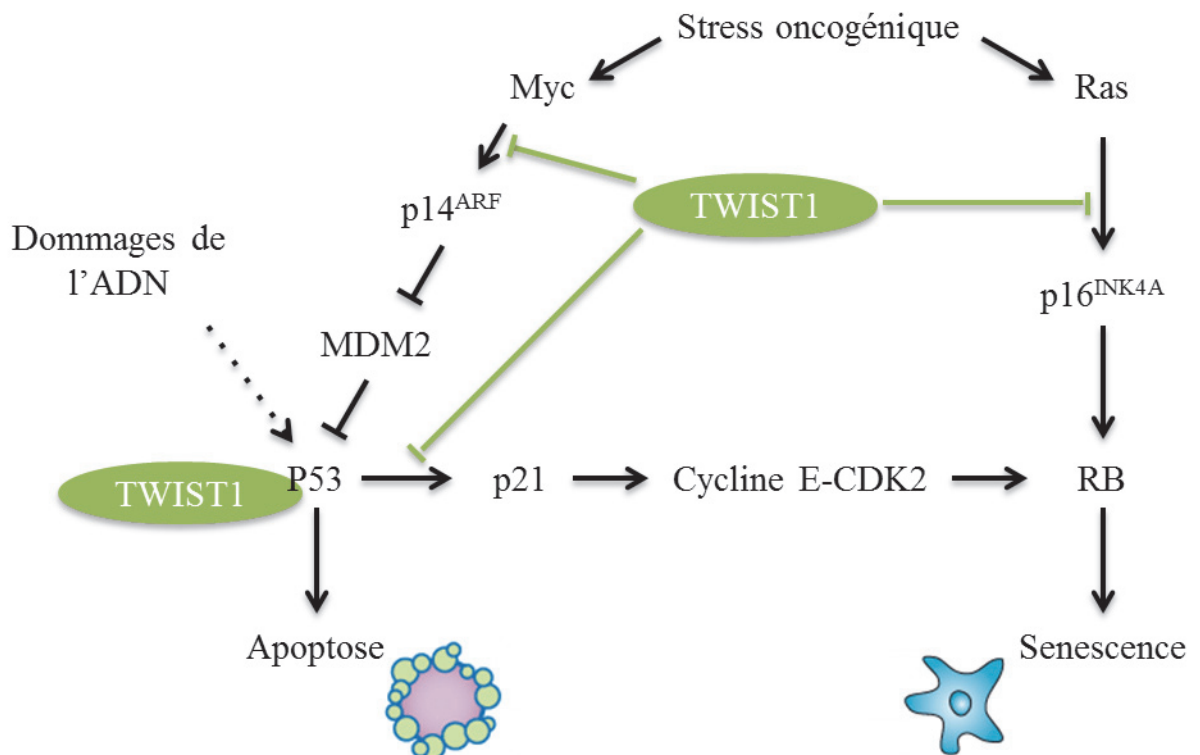


Figure 48: TWIST1 inhibe les mécanismes de sauvegardes cellulaires.
D'après Puisieux, Brabletz et Caramel 2014 (Puisieux, Brabletz, & Caramel, 2014)

4.2.3 TWIST1 est impliqué dans la TEM

La Transition Epithélio-Mésenchymateuse (TEM) est un processus embryonnaire fondamental au cours duquel les cellules épithéliales polarisées et jointives deviennent des cellules mésenchymateuses individuelles et motiles. Ce processus réversible favorise la migration des cellules au cours des processus embryonnaires notamment au cours de l'organogenèse (Nieto, 2013; Thiery et al., 2009). La TEM a tout d'abord été associé aux étapes d'invasions et de cascade métastatique, mais récemment il a été montré son rôle au cours de l'initiation et du développement des tumeurs primaires (Ansieau et al., 2008; Morel et al., 2012; Thiery et al., 2009; J. Yang & Weinberg, 2008). Ce programme de trans-différenciation est orchestré par des facteurs de transcription dits « EMT-TFs » et peut être réactivé à l'âge adulte notamment au cours de la tumorigenèse (Morel et al., 2012) (Puisieux

et al., 2014). TWIST1 est un facteur de transcription capable d'induire le processus de transdifférenciation qu'est la TEM. Ainsi, TWIST1 est un facteur de transcription qui joue un rôle important au cours de la tumorigenèse (notamment au cours de l'initiation et de la progression tumorale) et de la différenciation métastatique (Ansieau et al., 2008; Morel et al., 2012; Smit, Geiger, Song, Gitelman, & Peeper, 2009; J. Yang et al., 2004; J. Yang & Weinberg, 2008). En effet, l'expression ectopique de TWIST1 conduit à des caractéristiques morphologiques et moléculaires associées à la TEM (motilité cellulaire, perte de la E-cadhérine, activation des marqueurs mésenchymateux) (J. Yang et al., 2004).

4.3 TWIST1 : une cible thérapeutique

Comme nous l'avons vu précédemment, TWIST1 est une protéine indétectable dans la plupart des cellules normales adultes différenciées qui est surexprimée dans de nombreux cancers et est fréquemment associée à un mauvais pronostic dans les modèles humains et de souris (Kwok et al., 2005; Raatikainen et al., 2015; Wushou, Hou, Zhao, & Shao, 2014; J. Yang et al., 2004). De ce fait, TWIST1 constitue une cible attrayante pour élaborer de nouvelles stratégies thérapeutiques.

En accord avec cette hypothèse, l'inhibition de son expression dans différentes lignées tumorales induit effectivement la réactivation des mécanismes de sauvegarde cellulaire et entraîne un effet cytostatique sur des xénogreffes de carcinomes pulmonaires chez la souris (Ansieau et al., 2008; Tran et al., 2012; Valsesia-Wittmann et al., 2004). D'autre part la répression de TWIST1 par le 8-bromo-5-hydroxy-7-methoxychrysin, un analogue chimique de la chrysin, diminue la croissance de xénogreffes (Ren et al., 2013). Enfin, une diminution des niveaux d'expression de TWIST1 *via* l'utilisation de siRNA a montré une potentialisation de l'effet de la chimiothérapie du trioxyde d'arsenic (Seo et al., 2014).

CHAPITRE 4 : OBJECTIFS DE LA THESE

TWIST1 est une cible thérapeutique potentielle. Afin d'identifier des inhibiteurs spécifiques, il faut décrypter les mécanismes moléculaires structuraux. Les objectifs de mes travaux de thèse ont été (Figure 49) :

- de modéliser *in silico* les complexes TWIST1/TWIST1 et TWIST1/E12. Ces modèles *in silico* tridimensionnels ont été construits par homologie à partir de l'empreinte du complexe murin NeuroD1/E47. Ils ont permis d'identifier le rôle des boucles des domaines bHLH et le rôle des acides aminés clefs impliqués dans la fixation de TWIST1 à l'ADN (Travail 1 : (Bouard et al., 2014)).
- de participer à l'identification du complexe oncogénique : TWIST1/E12 afin de déterminer au mieux la cible moléculaire (Travail 2 : (Jacqueroud et al., 2016)).
- d'identifier les éléments clefs de TWIST1/E12 impliqués dans la reconnaissance des séquences E-box d'ADN cibles des promoteurs de gènes régulés par cet hétérodimère, d'estimer *in silico* l'affinité de TWIST1/E12 pour différentes séquences E-box et de valider *in vitro* ces prédictions par une approche stréptavidine/biotine (Travail 3 : (Bouard et al., 2016)).
- de prédire l'impact d'une mutation sur la fonction de dimérisation et de reconnaissance d'un E-box des complexes de TWIST1 en étudiant plus spécifiquement une mutation associée au syndrome de Saethre-Chotzen (Travail 4 : (Bouard et al., 2016b)).
- de mieux comprendre les impacts structuraux et les mécanismes moléculaires sous-jacents associés à la phosphorylation de la sérine 123 de TWIST1 dans sa forme hétérodimérique. Mais également d'identifier l'histidine 593 de E12 comme un résidu clef pour la formation de l'hétérodimère TWIST1/E12 (en cours).

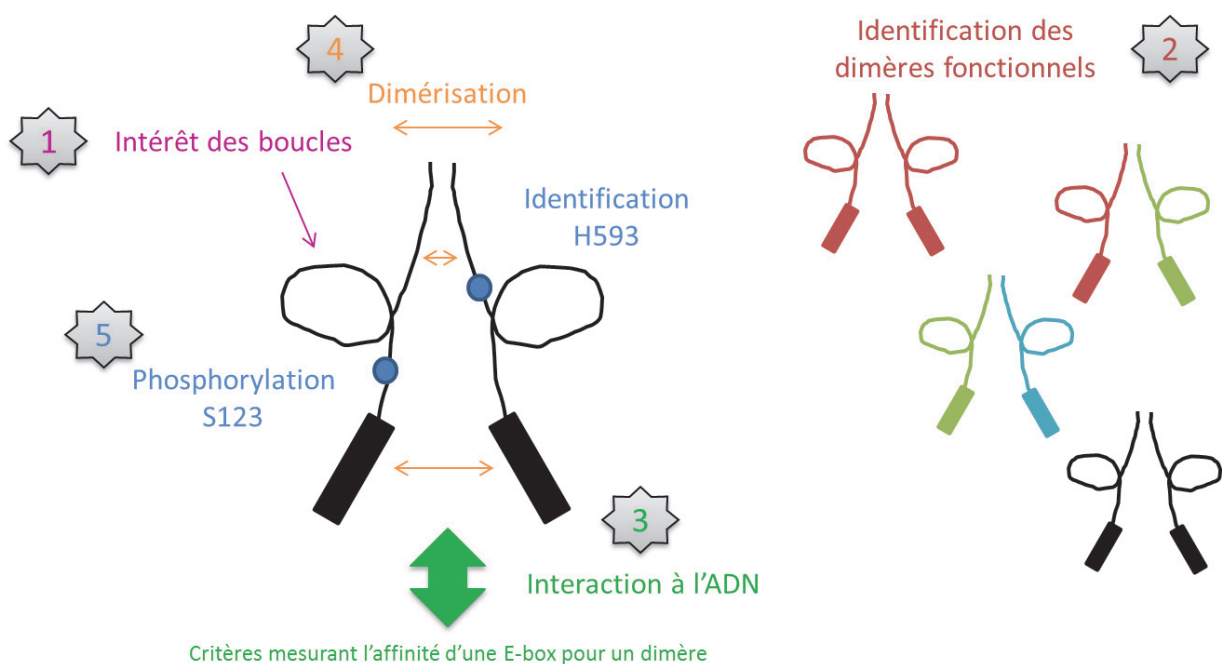


Figure 49: objectifs de la thèse

CHAPITRE 5 : TRAVAUX REALISES

Travail de synthèse

« Cancer cell plasticity and metastatic dissemination »

Moyret-Lalle C., Pommier R., Bouard C., Nouri E., Richard G., Puisieux A.

Médecine/Sciences, janvier 2016

1 Travail de synthèse : Plasticité des cellules cancéreuses et dissémination métastatique

La dissémination métastatique est une suite d'événements qui conduit certaines cellules cancéreuses à envahir des tissus distants de la tumeur primaire, on parle alors d'invasion. Chaque événement constitue un obstacle pour les cellules cancéreuses et seul un nombre limité de cellules cancéreuses sont capables de les franchir. Toutefois la diversité génétique des cellules cancéreuses au sein d'une tumeur et la plasticité cellulaire transmise suite aux signaux du microenvironnement facilitent le franchissement de ces étapes et la dissémination des cellules cancéreuses dans l'organisme et favorisent la formation de métastases en favorisant l'émergence de cellules tumorales ayant une résistance intrinsèque aux différentes pressions rencontrées. Ainsi, la progression tumorale et la dissémination métastatique font preuve d'une grande adaptabilité et font appel à des interactions étroites entre les cellules tumorales et le microenvironnement. Un processus de trans-différenciation embryonnaire, la transition épithélio-mésenchymateuse, est fréquemment réactivée au cours du développement de la tumeur. Ce processus donne aux cellules tumorales différentes propriétés de cellules souches et donc de plasticité permettant aux cellules tumorales de s'adapter aux nouveaux environnements et contraintes rencontrées lors de l'invasion en faveur également de l'hétérogénéité tumorale.

Dans le cadre de ma thèse, j'ai participé au travail de synthèse, à la rédaction de cette revue et à l'élaboration des figures 1 et 2

« Plasticité des cellules cancéreuses et dissémination métastatique »

English title « Cancer cell plasticity and metastatic dissemination »

Caroline Moyret-Lalle, Roxane Pommier Charlotte Bouard, Ebticem Nouri, Geoffrey Richard, et Alain Puisieux

Université de Lyon, F-69000 Lyon, France

Université Lyon I, Lyon, F-69622, France, F-69000 Lyon, France

INSERM U1052, Centre de Recherche en Cancérologie de Lyon, F-69000 Lyon, France

CRNS UMR 5286, Centre de Recherche en Cancérologie de Lyon, F-69000 Lyon, France

Centre Léon Bérard, Département de la Recherche, F-69008 Lyon, France

LabEx DEVweCAN, F-69000 Lyon, France

e-mail: alain.puisieux@lyon.unicancer.fr

Summary

Metastatic dissemination consists of a sequence of events resulting in the invasion by cancer cells of tissues located away from the primary tumour. This process is highly inefficient, since each event represents an obstacle that only a limited number of cells can overcome. However, two biological phenomena intrinsically linked with tumour development facilitate the dissemination of cancer cells throughout the body and promote the formation of metastases, namely the genetic diversity of cancer cells within a given tumour, which arises from their genetic instability and from successive clonal expansions, and cellular plasticity conveyed to the cells by micro-environmental signals. Genetic diversity increases the probability of selecting cells that are intrinsically resistant to biological and physical constraints encountered during metastatic dissemination, whereas cellular plasticity provides cells with the capacity to adapt to stressful conditions and to changes in the microenvironment. The epithelial-mesenchymal transition, an embryonic trans-differentiation process frequently reactivated during tumour development, plays an important role in that context by endowing tumor cells with a unique capacity of motility, survival and adaptability to the novel environments and stresses encountered during the invasion-metastasis cascade.

Résumé

La dissémination métastatique repose sur une séquence d'événements qui aboutit à la colonisation par les cellules cancéreuses d'un tissu à distance de la tumeur primaire. Il s'agit d'un processus inefficace, chaque étape représentant un obstacle que seul un petit nombre de cellules réussissent à franchir. Deux phénomènes intrinsèquement associés au développement tumoral favorisent pendant leur dissémination au sein de l'organisme et la formation de métastases : la diversité génétique des cellules cancéreuses au sein d'une tumeur et la plasticité conférée à ces cellules par des signaux provenant du microenvironnement. La transition épithélio-mésenchymateuse, processus de trans-différenciation cellulaire, joue un rôle important dans ce contexte en accordant aux cellules tumorales une capacité unique de motilité, de survie et d'adaptation aux stress et aux nouveaux environnements rencontrés au cours des différentes phases de la dissémination métastatique.

Introduction

Alors que 90% des cas de décès par cancer sont dus au développement de métastases, le processus de dissémination métastatique des cellules cancéreuses reste mal connu. Cette relative méconnaissance traduit la complexité biologique du phénomène. Il s'agit d'un processus séquentiel, impliquant plusieurs étapes successives, d'où la terminologie fréquemment employée de « cascade métastatique » [1]. Dans le contexte des tumeurs d'origine épithéliale, de loin les plus fréquentes chez l'adulte, le processus est initié par le détachement de certaines cellules cancéreuses du tissu épithélial par modification des interactions cellules-cellules. Les cellules qui survivent à ce stress pourront migrer, de manière individuelle ou collective, dans le tissu conjonctif intra- et péri-tumoral, entraînant une invasion locale. Cette étape précoce peut être dissociée des suivantes puisqu'une tumeur invasive localement ne devient pas forcément métastatique à distance. Le processus d'intravasation constitue une étape critique de la dissémination des cellules tumorales au sein de l'organisme puisqu'elle consiste en la traversée de la paroi des vaisseaux sanguins ou lymphatiques par les cellules cancéreuses pour entrer dans la circulation. Ce phénomène est favorisé par l'angiogenèse d'origine tumorale ou « néo-angiogenèse », c'est-à-dire la formation de nouveaux vaisseaux capables d'irriguer la tumeur en développement. Les cellules qui survivent aux pressions physiques et aux contraintes soumises par la circulation pourront procéder par la suite à l'étape d'extravasation, la sortie des vaisseaux leur permettant d'atteindre et de coloniser un organe cible. La sélection de ces organes est dépendante des caractéristiques biologiques de ces derniers, notamment de la microcirculation et du microenvironnement tissulaire, et des caractéristiques des cellules tumorales qui leur confèrent un tropisme tissulaire qui leur est propre. Les dernières étapes de la progression métastatique correspondent à l'implantation des cellules tumorales extravasées dans l'organe cible sous forme de cellules isolées ou d'amas paucicellulaires, puis à la formation de micrométastases et enfin à la formation de métastases cliniquement décelables (figure 1) [2,3]. De nombreuses observations et données expérimentales démontrent que le développement de métastases est un processus hautement inefficace. Ainsi, on estime que seuls 2% des cellules circulantes ou CTCs (pour *Circulating Tumor Cells*) qui survivent dans le flux sanguin sont à l'origine de micro-métastases et seulement 1% de ces micro-métastases progresse jusqu'à un stade macroscopique [4]. En effet, la majorité des cellules qui ont réussi à entrer dans la circulation, ou à arriver dans un tissu à distance, meurent du fait de leur incapacité à s'adapter à leur

nouvel environnement. Les cellules survivantes entrent pour la plupart dans un état de « dormance », caractérisé par une absence de prolifération. L'échappement à cet état de dormance est donc nécessaire au développement de métastases.

Le franchissement de chacune des étapes de la dissémination métastatique requiert des capacités uniques de motilité, d'adaptation métabolique et de survie des cellules cancéreuses. Nous proposons d'explorer le rôle de la plasticité cellulaire et du dialogue entre les cellules cancéreuses et leur microenvironnement dans ce processus complexe et d'évoquer l'importance de l'hétérogénéité intratumorale dans l'émergence de cellules à potentiel métastatique.

Plasticité cellulaire et dissémination métastatique

La dissémination métastatique a longtemps été présentée comme une phase tardive de la progression tumorale. En effet, sur la base d'une vision exclusivement génétique de la cancérogenèse, il était proposé que les étapes d'invasion et de dissémination étaient initiées par l'acquisition par quelques cellules tumorales d'anomalies génétiques leur conférant des propriétés de motilité et de survie. Cependant, trois observations majeures ont conduit à réviser cette notion : le faible nombre d'événements génétiques connus pour être directement associés au processus métastatique ; l'observation de processus précoces d'invasion et de dissémination dans certains cancers ; la mise en évidence de la complexité biologique du processus et du nombre de propriétés requises pour en franchir les différentes étapes. Aussi, une hypothèse mécanistique additionnelle a été proposée: la dissémination métastatique ferait appel à une reprogrammation épigénétique des cellules cancéreuses qui pourrait survenir dès les phases précoces de la progression tumorale. La réactivation de la transition épithélio-mésenchymateuse (TEM) se trouve au cœur de ce concept.

La TEM est un processus embryonnaire au cours duquel les cellules épithéliales perdent leur cohésion, modulent leur polarité et réorganisent la structure de leur cytosquelette. Elles s'individualisent, deviennent mobiles et acquièrent des caractéristiques morphologiques et fonctionnelles de cellules mésenchymateuses. La TEM est nécessaire à plusieurs étapes critiques de la morphogenèse aboutissant à la construction du plan de l'embryon et au développement des organes. Ce processus est contrôlé par des signaux

du microenvironnement et est orchestré par un réseau de facteurs de transcription (FT-TEM), dont les mieux caractérisés appartiennent aux familles SNAI, ZEB et TWIST [5,6]. Il consiste en une reprogrammation importante de l'épigénome, impliquant notamment des changements dans la méthylation de l'ADN, des modifications post-traductionnelles des histones, des changements d'expression de microARNs ou de lncARNs (ARNs longs non codants) [7,8]. Bien que l'implication de la TEM dans la dissémination métastatique reste encore un sujet de controverse [9,10], de nombreuses observations montrent, au moins dans certains types tumoraux, que sa réactivation aberrante peut constituer un événement déterminant des étapes précoces de celle-ci. En effet, des figures morphologiques de TEM associées à une surexpression de FT-TEM ont pu être mises en évidence sur le front invasif de certains carcinomes humains [11], au niveau des cellules tumorales circulantes chez les patients atteints de cancer, ainsi que dans des modèles expérimentaux de dissémination métastatique [12]. La TEM favorise ainsi l'étape d'invasion locale en conduisant à la perte des protéines d'adhésion intercellulaire, les FT-TEM se conduisant notamment comme des répresseurs de la E-cadhérine, et en initiant le réarrangement du cytosquelette. Ces événements confèrent aux cellules cancéreuses des capacités accrues de motilité et d'invasion. L'invasion est également facilitée par l'induction d'expression de métalloprotéases (MMPs), permettant aux cellules tumorales de cliver les composants de la matrice extracellulaire notamment au niveau des invadopodes. Ces structures spécialisées des cellules invasives sont des régions de réorganisation du cytosquelette d'actine dans lesquelles se concentre une MMP (MT1-MMP) essentielle à l'invasion et à la dégradation de la matrice [13]. De façon importante, l'expression des FT-TEM procure aux cellules tumorales une capacité de survie accrue, en particulier en inhibant les voies onco-suppressives dépendantes de p53 et de Rb [5,14,15]. Cette propriété favorise le développement de la tumeur primaire, et protège les cellules tumorales en cours d'invasion et de dissémination contre les stress associés aux processus d'anoïkie (mort cellulaire consécutive à la perte de contacts cellules-cellules et au détachement de la matrice extracellulaire sous-jacente), d'invasion et de dissémination. Enfin, la TEM permet aux cellules cancéreuses d'acquérir des propriétés caractéristiques de cellules souches, incluant des propriétés d'auto-renouvellement, favorisant leur adaptation dans un environnement différent de celui du tissu d'origine [16,17].

Si les étapes initiales de la dissémination métastatique sont favorisées, et peut-être parfois initiées, par une TEM, les étapes tardives telles que

l'extravasation et la colonisation du tissu hôte semblent nécessiter une transition inverse, dite mésenchymo-épithéliale (TME) [3,18], apportant une explication rationnelle à l'observation selon laquelle les métastases reproduisent généralement le phénotype de la tumeur primaire. Ainsi, expérimentalement, une cellule tumorale disséminée ou DTC (pour *Disseminated Tumor Cell*) ayant acquis un phénotype mésenchymateux doit subir une TME pour former un clone métastatique exprimant la E-cadhérine [18]. Cette nécessité est probablement liée aux capacités réduites de prolifération des cellules à l'état mésenchymateux. Comme la TEM, la TME est un processus dynamique contrôlé par des facteurs protéiques (GRHL2, ELF3/ELF5) et des micro-ARNs (miR-200, miR-3) [3,18,19] (figure 2). Ainsi, la cascade métastatique semble reposer sur une balance épithélio-mésenchymateuse, les cellules cancéreuses à même de former une métastase étant probablement celles qui présentent la plus forte plasticité phénotypique et fonctionnelle. Cette hypothèse donne une explication rationnelle à l'observation chez les patientes atteintes de cancer du sein de cellules circulantes présentant un phénotype mixte intégrant à la fois des caractéristiques épithéliales et mésenchymateuses [20]. Il est important de souligner, qu'au-delà des propriétés intrinsèques des cellules cancéreuses, la balance épithélio-mésenchymateuse, et de façon générale la plasticité cellulaire, sont fortement dépendantes des signaux provenant du microenvironnement. Le dialogue entre les cellules cancéreuses et leur microenvironnement constitue donc une composante essentielle de la cascade métastatique.

Rôle du microenvironnement dans la plasticité des cellules cancéreuses et la dissémination métastatique

Au sein de la tumeur primaire, les cellules tumorales sont intégrées dans un environnement complexe constitué des composants de la matrice extracellulaire (MEC) et de différents types cellulaires, incluant fibroblastes, cellules endothéliales, adipocytes et cellules du système immunitaire. La progression tumorale est le résultat d'un dialogue réciproque et dynamique entre les cellules tumorales et leur microenvironnement. Ce dernier va influencer le comportement cellulaire au travers de signaux diffusibles (facteurs de croissance, cytokines) et de contraintes structurelles (molécules d'adhésion) [21]. Initialement considérée comme une structure passive de soutien, la MEC joue un rôle actif dans l'acquisition de propriétés plastiques par les cellules tumorales en modulant au travers de récepteurs d'adhérence des

voies de signalisation de type IGF1 (*Insulin Growth Factor 1*)[25]. Une modification de la structure de la MEC telle qu'une rigidification peut induire un confinement de cellules tumorales conduisant à une reprogrammation épigénétique activant l'expression de FT-TEM. En retour, les cellules tumorales ayant subi une TEM vont modifier leur environnement. Elles libèrent en effet des facteurs solubles (facteurs de croissance, cytokines, chimiokines) qui servent de médiateurs de communication entre cellules cancéreuses et cellules stromales environnantes, et qui favorisent la création d'un environnement apte à la génération d'une niche pro-métastatique. Les fibroblastes constituent des cellules clefs de ce dialogue entre cellules cancéreuses et microenvironnement. Alors qu'ils sont relativement peu présents au sein de la matrice extracellulaire normale, ils représentent la composante cellulaire majoritaire du microenvironnement tumoral [22]. La sécrétion par les cellules tumorales de TGF β et de PDGF, ainsi que le stress oxydatif associé à une activation mitogénique aberrante favorisent, d'une part, l'initiation d'une TEM dans les cellules cancéreuses et conduisent, d'autre part, à l'activation des fibroblastes résidents [23]. Ces fibroblastes activés (CAFs pour *Cancer Associated Fibroblasts*) vont former des myofibroblastes, qui peuvent ainsi représenter jusqu'à 80% du compartiment fibroblastique. Les CAFs présentent des modifications de leur phénotype telles que l'expression de l' α -SMA (*α -Smooth Muscle Actin*), de la vimentine, de la desmine, de la FAP (*Fibroblast Activated Protein*) ainsi que l'acquisition de capacités migratoires [24]. Ils sont capables d'interagir avec la matrice extracellulaire qui les entoure, et son remodelage, qui fait appel à des protéases de types MMPs ou uPA (une sérine protéase), assure d'une part la modification des barrières physiques qui entourent les cellules tumorales et d'autre part la libération de molécules séquestrées telles que l'IGF, le TGF β , Wnt et les FGF [21,25] pouvant induire une TEM. L'uPA peut notamment générer une molécule active, la plasmine, ayant le potentiel de cliver des composants essentiels de la MEC que sont la laminine et la fibronectine. L'uPA induit la prolifération et la migration cellulaire et représente un facteur de mauvais pronostic dans les cancers du sein [26].

L'inflammation, fréquemment associée au processus de tumorigenèse, favorise également des interactions cellulaires qui facilitent le franchissement de certaines étapes de la dissémination métastatique. Le cancer est ainsi considéré comme une "blessure non réparée" où l'auto-renouvellement et la prolifération des cellules immunes induisent une inflammation permanente, par rétrocontrôle positif [27]. L'inflammation

conduit à la production d'espèces réactives de l'oxygène, de protéases, de cytokines et de chemokines par les cellules immunitaires. A titre d'exemple, les TAM (pour *Tumor Associated Macrophages*) sécrètent de l'IL-10 et du TGF β qui vont moduler la réponse immunitaire et activer la néo-angiogenèse [28], favorisant simultanément la plasticité cellulaire et la dissémination métastatique. La néo-angiogenèse, angiogenèse initiée par le développement tumoral, constitue en effet un élément clef du remodelage favorisant le développement métastatique [29]. Ce processus est initié par un déséquilibre entre facteurs pro- (ex. VEGF, FGF2) et anti-angiogéniques, (ex. TBSPD-1 et 2, IL-10) appelé switch angiogénique. Les néo-vaisseaux tumoraux présentent des anomalies structurales telles qu'une membrane basale fragmentée présentant une perméabilité accrue qui favorise l'intravasation des cellules tumorales. La néo-angiogenèse fait également intervenir le facteur de transcription HIF-1 α (*Hypoxia Inducible Factor-1*), qui stimule l'expression de gènes cibles comme le *VEGFA* ainsi que certains FT-TEM et le gène *SPARC* dont l'expression est souvent associée à une TEM. La protéine SPARC a été montrée récemment comme étant capable de favoriser l'extravasation des cellules de mélanome en provoquant l'ouverture des jonctions endothéliales vasculaires ce qui facilite la formation de métastases pulmonaires [30].

Au total, la progression tumorale et la dissémination métastatique font appel à des interactions étroites entre les cellules tumorales et les cellules et composants du microenvironnement (figure 3). Certains signaux provenant du microenvironnement tumoral favorisent ainsi l'initiation d'une TEM et le franchissement d'étapes clefs de la dissémination métastatique. En retour, les cellules cancéreuses peuvent modifier leur microenvironnement. Dans ce contexte, des travaux récents démontrent le rôle de la sécrétion par les cellules tumorales d'exosomes, structures à double-membrane hautement spécialisées, capables de contrôler la communication intracellulaire et le microenvironnement local. Les exosomes contiennent en effet des protéines, des lipides, de l'ADN, des ARNs et des microARNs qui peuvent être transférés après fusion avec des cellules et ainsi initier la TEM, promouvoir la néo-angiogenèse, favoriser l'échappement des cellules cancéreuses à la reconnaissance par les cellules immunitaires et participer à la formation d'une niche pré-métastatique [31,32].

Plasticité cellulaire, hétérogénéité tumorale et invasion métastatique

La plasticité cellulaire constitue une composante essentielle de l'hétérogénéité intra-tumorale. Cette hétérogénéité repose en effet sur deux phénomènes principaux. Le premier est un processus génétique basé sur la survenue de vagues successives d'expansion clonale au cours du développement tumoral. Chacune de celles-ci a pour origine l'acquisition d'une mutation conférant un avantage sélectif de croissance [33]. Ce processus est à l'origine d'une diversité génétique intra-tumorale accordant à la population cellulaire une forte capacité d'adaptation vis-à-vis d'un stress, au travers de la sélection de clones résistants. Le deuxième processus est de nature épigénétique, à l'origine de la plasticité phénotypique et fonctionnelle. Cette plasticité est associée à une capacité individuelle d'adaptation (figure 4). Comme illustré précédemment par le processus de TEM, cette notion de plasticité cellulaire est intrinsèquement associée aux capacités de différenciation cellulaire. Différents modèles théoriques évoquant le rôle de la différenciation et de la plasticité cellulaires au cours de la cancérogenèse ont été établis. Le modèle hiérarchique stipule qu'au sein d'une tumeur maligne, seule une minorité des cellules cancéreuses, appelées « cellules souches cancéreuses » (CSCs) présente des propriétés d'auto-renouveau et la capacité de régénérer une tumeur [34]. Le modèle stochastique d'inter-conversion repose quant à lui sur les propriétés intrinsèques de plasticité des cellules tumorales [35]. La TEM joue dans ce modèle un rôle important, en permettant aux cellules tumorales différenciées d'acquies des propriétés de cellules souches [16,17].

De nombreuses observations montrent que l'hétérogénéité intra-tumorale et la plasticité des cellules cancéreuses soutiennent le processus métastatique en permettant l'émergence des cellules tumorales présentant une résistance intrinsèque aux différentes pressions de sélection rencontrées, et des cellules aptes à s'y adapter [36]. Des modélisations mathématiques ont ainsi permis de mettre en évidence un processus darwinien de formation des métastases impliquant une évolution du phénotype cellulaire tout au long de l'invasion et de la dissémination [37]. Seules les cellules qui ont suffisamment évolué en s'affranchissant des contraintes du site primaire et pouvant se multiplier dans un environnement « étranger » peuvent générer la formation de métastases. Cette dynamique évolutive induit à la fois une divergence phénotypique entre différents clones métastatiques et une convergence phénotypique des cellules colonisant le même organe.

Conclusion et Perspectives

Le développement tumoral est fréquemment considéré comme un processus d'évolution génétique dont la formation des métastases constitue la phase ultime. L'importance de l'hétérogénéité génétique intra-tumorale est reflétée par l'observation selon laquelle une grande diversité clonale prédit la progression d'un état pré-malin vers la malignité [38]. Cependant, au-delà de la diversité génétique, il apparaît clairement que le franchissement des différentes étapes du développement métastatique fait également appel aux capacités d'adaptation des cellules cancéreuses, en relation directe avec leur microenvironnement. Ainsi, le processus de TEM, mécanisme de trans-différenciation cellulaire contrôlé par des signaux du microenvironnement, peut conférer aux cellules, dès les phases précoces de la tumorigenèse, une plasticité associée à une capacité unique de motilité cellulaire, de survie et d'adaptation aux stress et aux nouveaux environnements rencontrés au cours des différentes phases de la dissémination. Remettant en question le modèle d'évolution linéaire de la progression tumorale dont le développement métastatique constitue une phase tardive et indépendante, l'initiation d'une TEM favorise la croissance de la tumeur primaire, au travers de l'acquisition de propriétés d'auto-renouveau et de l'inhibition des voies onco-suppressives dépendantes de p53 et de Rb. Aussi, l'initiation de la dissémination métastatique pourrait être une conséquence secondaire d'événements nécessaires au développement de la tumeur primaire. La prévention des processus assurant une plasticité phénotypique et fonctionnelle aux cellules cancéreuses constitue une opportunité nouvelle pour le développement d'approches ciblant des étapes clés de la progression tumorale et capables d'empêcher l'adaptation des cellules malignes aux thérapies conventionnelles.

BIBLIOGRAPHIE

- 1 Mehlen P, Puisieux A. Metastasis: a question of life or death. *Nat Rev Cancer* 2006, 6: 449-58
- 2 Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell* 2009, 139: 871-90
- 3 De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* 2013, 13: 97-110
- 4 Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of

- early micrometastases. *Am J Pathol* 1998, 153: 865-73
- 5 Morel AP, Hinkal GW, Thomas C, Fauvet F, *et al.* EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS Genet* 2012, 8:e1002723
- 6 Puisieux A, Brabletz T, Caramel J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* 2014, 16: 488-94
- 7 Moyret-Lalle C, Ruiz E, Puisieux Epithelial-mesenchymal transition transcription factors and miRNAs: "Plastic surgeons" of breast cancer. *World J Clin Oncol* 2014, 5: 311-22
- 8 Richards EJ, Zhang G, Li ZP, Permut-Wey J, *et al.* Long non-coding RNAs (LncRNA) regulated by transforming growth factor (TGF) β : LncRNA-hit-mediated TGF β -induced epithelial to mesenchymal transition in mammary epithelia. *J Biol Chem* 2015, 290: 6857-67
- 9 Fischer KR, Durrans A, Lee S, Sheng J, *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* 2015, 527: 472-6
- 10 Zheng X, Carstens JL, Kim J, Scheible M, *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* 2015, 527: 525-30
- 11 Spaderna S, Schmalhofer O, Hlubek F, Berx G, *et al.* A transient, EMT-linked loss of basement membranes indicates metastasis and poor survival in colorectal cancer. *Gastroenterology*. 2006, 131: 830-40
- 12 Pasquier J, Abu-Kaoud N, Al Thani H, Rafii A. Epithelial to Mesenchymal Transition in a Clinical Perspective. *J Oncol* 2015, 2015: 792182
- 13 Eckert MA, Lwin TM, Chang AT, Kim J, *et al.* Twist1-induced invadopodia formation promotes tumor metastasis. *Cancer Cell* 2011, 19: 372-86
- 14 Valsesia-Wittmann S, Magdeleine M, Dupasquier S, Garin E *et al.* Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer Cell* 2004, 6: 625-30
- 15 Ansieau S, Bastid J, Doreau A, Morel AP, *et al.* Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell* 2008, 14: 79-89
- 16 Morel AP, Lièvre M, Thomas C, Hinkal G, *et al.* Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* 2008, 3: e2888
- 17 Mani SA, Guo W, Liao MJ, Eaton EN, *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008, 133: 704-15
- 18 Jolly MK, Boareto M, Huang B, Jia D, *et al.* Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Front Oncol* 2015, 5: 155
- 19 Ocaña OH, Córcoles R, Fabra A, Moreno-Bueno G, *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* 2012, 22: 709-24
- 20 Yu M, Bardia A, Wittner BS, Stott SL, *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* 2013, 339: 580-4
- 21 Mueller MM, Fusenig NE. Friends or foes - bipolar effects of the tumour stroma in cancer. *Nat Rev Cancer* 2004, 4: 839-49
- 22 Rønnov-Jessen L, Petersen OW, Bissell MJ. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol Rev* 1996, 76: 69-125
- 23 Scholer-Dahirel A., Costa A., Mechta-Grigoriou F. Control of cancer-associated fibroblasts function by oxidative stress: A new piece in the puzzle. *Cell Cycle* 2013, 12: 2169
- 24 Garin-Chesa P, Old LJ, Rettig WJ. Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody target in human epithelial cancers. *Proc Natl Acad Sci USA* 1990, 87: 7235-9
- 25 Lanigan F, O'Connor D, Martin F, Gallagher WM. Molecular links between mammary gland development and breast cancer. *Cell Mol Life Sci* 2007, 64: 3159-84
- 26 Duffy MJ. Urokinase plasminogen activator and its inhibitor, PAI-1, as prognostic markers in breast cancer: from pilot to level 1 evidence studies. *Clin Chem* 2002, 48: 1194-7
- 27 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011, 144: 646-74
- 28 Mantovani A, Sica A, Locati M. Macrophage polarization comes of age. *Immunity* 2005, 23: 344-6
- 29 Folkman J. New perspectives in clinical oncology from angiogenesis research. *Eur J Cancer* 1996, 32A: 2534-9
- 30 Tichet M, Prod'Homme V, Fenouille N, Ambrosetti D, *et al.* Tumour-derived SPARC drives vascular permeability and extravasation through endothelial VCAM1 signalling to promote metastasis. *Nat Commun* 2015, 6: 6993
- 31 Suchorska WM, Lach MS. The role of exosomes in tumor progression and metastasis. *Oncol Rep* 2015 Dec 22, doi: 10.3892
- 32 Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, *et al.* Tumour exosome integrins determine organotropic metastasis. *Nature* 2015, 527: 329-35
- 33 Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976, 194: 23-8
- 34 Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* 1997, 3: 730-7

- 35 Cobaleda C, Jochum W, Busslinger M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* 2007, 449: 473-7
- 36 Wagenblast E, Soto M, Gutiérrez-Ángel S, Hartl CA, *et al.* A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* 2015, 520: 358-62
- 37 Bethge A, Schumacher U, Wedemann G. Simulation of metastatic progression using a computer model including chemotherapy and radiation therapy. *J Biomed Inform* 2015, pii: S1532-0464(15)00151-3
- 38 Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genet* 2006, 38: 468-73

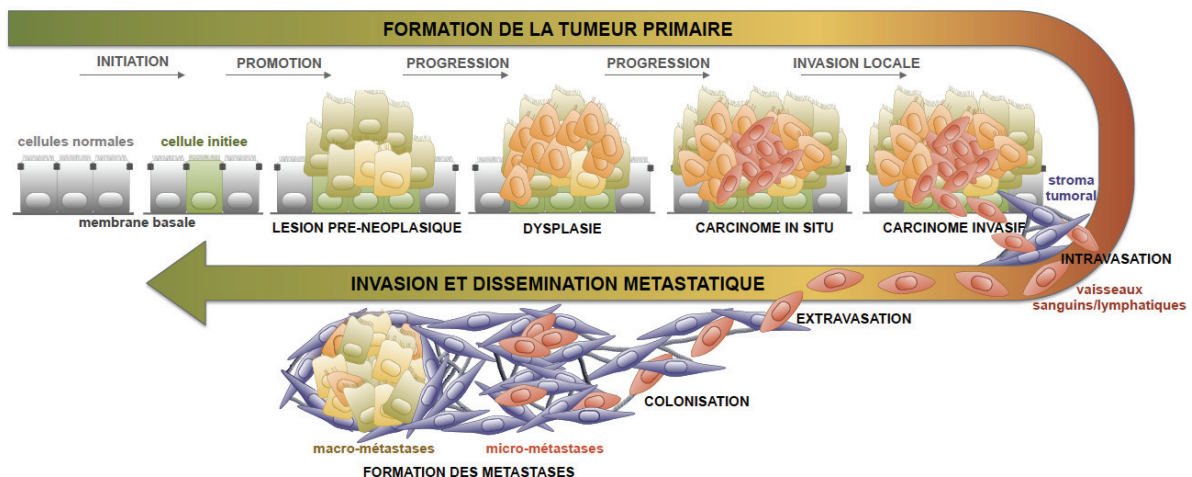


Figure 1 : Principales étapes du développement d'une tumeur d'origine épithéliale et de la cascade métastatique

La formation d'une tumeur bénigne (adénome) puis maligne (carcinome) repose sur l'acquisition progressive d'anomalies génétiques et épigénétiques conférant aux cellules des capacités accrues de prolifération et de survie. La cascade métastatique est initiée par une phase d'invasion tissulaire, menant à un carcinome invasif, suivie par une phase d'intravasation, les cellules cancéreuses pénétrant alors dans la circulation sanguine ou lymphatique. La survie dans la circulation de certaines cellules malignes et leur extravasation pourront conduire, après une phase de latence, à la colonisation d'un site secondaire pour former des micro-métastases puis des métastases détectables dans un site distant de celui de la tumeur primaire.

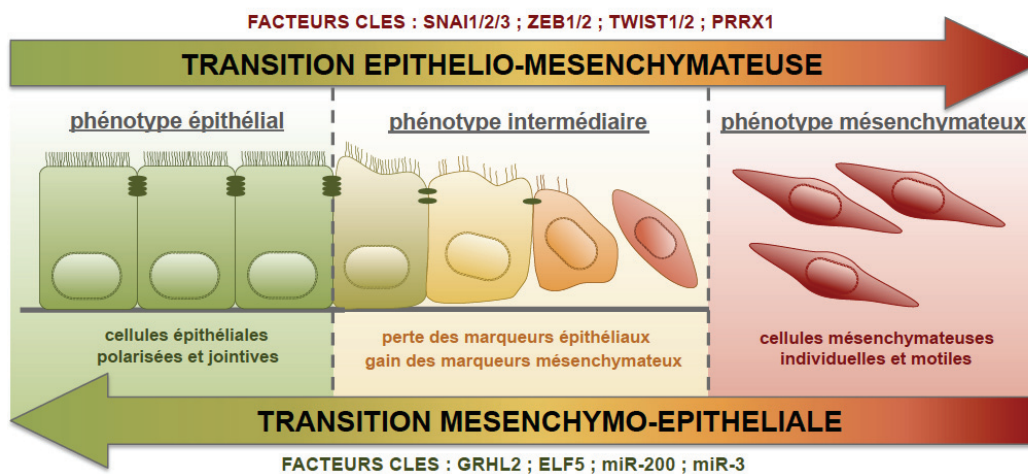


Figure 2 : Transition épithélio-mésenchymateuse et dissémination métastatique

La plasticité des cellules cancéreuses joue un rôle déterminant en leur permettant de s'adapter aux contraintes qui leur sont imposées au cours des différentes étapes de la cascade métastatique. La transition épithélio-mésenchymateuse (TEM) et son processus réverse (TME) sont impliqués respectivement dans les phases précoces et tardives de cette cascade. L'expression des FT-TEM des familles TWIST, ZEB et SNAI induit une TEM, conduisant à l'apparition de cellules mésenchymateuses motiles et invasives, alors que les facteurs GRLH2 ou ELF5 ainsi que certains micro-ARNs, tels que les membres de la famille miR-200, promeuvent un phénotype épithélial.

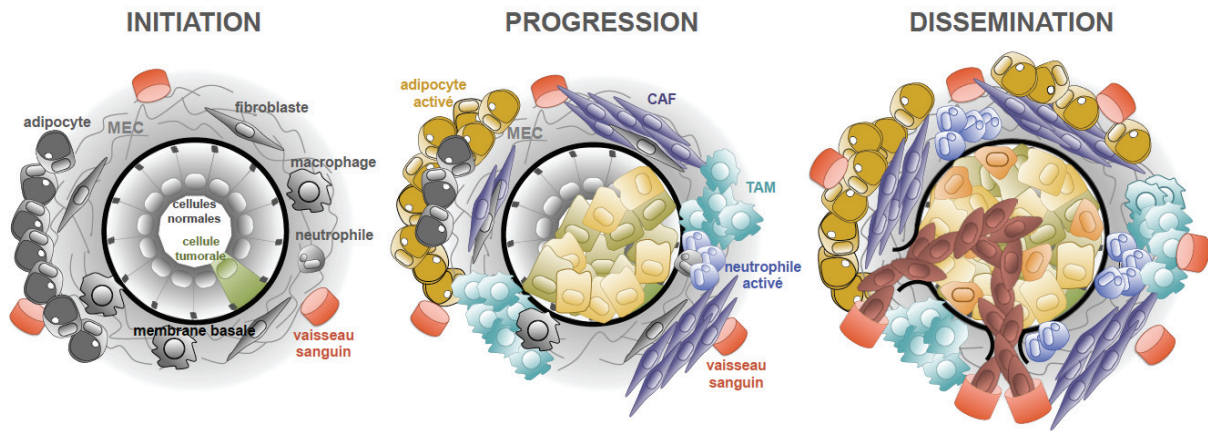


Figure 3 : Rôle du stroma au cours de la progression tumorale

Le stroma et la matrice extracellulaire (MEC) subissent d'importantes modifications au cours de la tumorigénèse et de la dissémination des cellules cancéreuses. Ainsi, les monocytes recrutés au niveau de la tumeur se différencient en macrophages activés (TAM) et participent à la création du microenvironnement tumoral, en favorisant, l'angiogénèse, la migration et la croissance des cellules tumorales. L'activation des adipocytes et des fibroblastes (CAFs) conduit à une augmentation de la sécrétion de facteurs de croissance, de composants de la matrice extracellulaire et de métalloprotéases facilitant le franchissement de la membrane basale, la néo-angiogénèse et le processus d'intravasation.

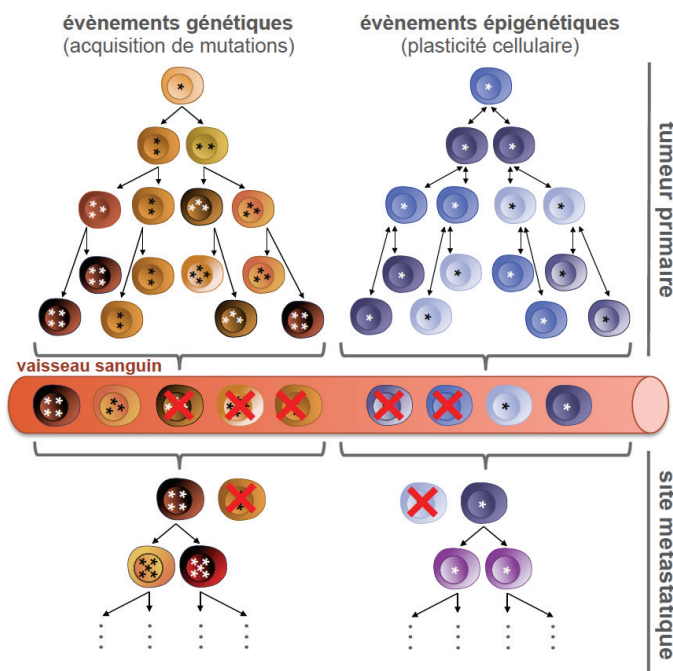


Figure 4 : Hétérogénéité intra-tumorale et métastases

L'hétérogénéité intra-tumorale, d'origine génétique et épigénétique, favorise le franchissement des différentes étapes de la cascade métastatique. L'instabilité génétique associée au développement tumoral promeut une diversité génétique intra-tumorale et augmente la probabilité de sélection de cellules intrinsèquement résistantes aux contraintes biologiques et physiques rencontrées au cours de la dissémination (exemples de l'intravasation, de la survie dans la circulation et de l'extravasation). Dans la figure, les anomalies génétiques acquises sont représentées par une étoile *. La plasticité qui caractérise certaines cellules cancéreuses en réponse à des signaux microenvironnementaux leur confère une capacité individuelle d'adaptation, leur permettant de franchir les différentes étapes de la cascade métastatique. L'importance relative des processus génétiques et épigénétiques dans ce phénomène est probablement dépendante du type tumoral.

Travail 1

« Interhelical loops within the bHLH domain are determinant in maintaining TWIST1–DNA complexes »

Bouard C., Terreux R., Hope J., Chemelle JA., Puisieux A., Ansieau S. et Payen L.

Journal of Biomolecular Structure & Dynamics, 2014

2 Travail 1 : élaboration de modèles *in silico* de TWIST1/TWIST1 et de TWIST1/E12

Afin de mieux décrypter l'organisation structurale de TWIST1, nous avons créé par modélisation moléculaire différents domaines bHLH des dimères de TWIST1 par homologie à la structure cristalline murine NeuroD1/E47 (PDB : 2QL2). Nous avons construit ainsi les dimères sauvages TWIST1/TWIST1 et TWIST1/E12 ainsi que des dimères mutants décrits dans le syndrome de Saethre-Chotzen TWIST1-ins135/E12, TWIST1-ins139/E12, TWIST1-ins135/TWIST1-ins135 et TWIST1-ins139/TWIST1-ins139. Nous avons soumis nos modèles à des dynamiques moléculaires de 10 ns afin de mimer au mieux le contexte physiologique. Nous avons étudié les éléments structuraux clés de ces complexes de TWIST1, notamment le rôle clef que jouent les boucles, situées entre les hélices alpha des domaines basiques, dans la stabilisation des complexes TWIST1-ADN.

Cette étude de modélisation *in silico* de mutants associés au syndrome de Saethre-Chotzen, caractérisé par une perte fonctionnelle de TWIST1, nous a permis de comprendre l'organisation structurale du domaine bHLH de la protéine TWIST1. Effectivement, nous avons identifié les résidus clés nécessaires à la dimérisation de TWIST1 mais également à la liaison à l'ADN de TWIST1 qui sont des prérequis aux fonctions de facteurs de transcription de cette protéine. En effet, la conformation des deux hélices α amphipatiques ainsi que la boucle du domaine bHLH diffèrent entre la forme monomérique et dimérique de TWIST1 d'une part mais aussi entre les différents dimères de TWIST1 formés montrant donc la rôle important de la flexibilité de ce domaine bHLH pour la dimérisation en impactant le choix du partenaire de dimérisation de TWIST1 mais également pour la reconnaissance des éléments de réponse E-box des promoteurs de gènes selon les dimères formés. Au-delà de la flexibilité induite par certaines mutations/insertions au niveau du domaine bHLH, nous avons montré que ces modifications impactaient également les liaisons hydrogènes formées par les résidus du dimère et impliquées dans la dimérisation et donc le maintien de la proximité des deux monomères ou dans la liaison des dimères à l'ADN et impliquées dans les fonctions de transactivation des complexes de TWIST1. D'autre part, l'impact structural est dépendant des modifications appliquées à TWIST1 puisque tous les mutants n'ont pas les mêmes variations structurales. De précédents travaux montrent que le résidu de sérine 144 joue un rôle déterminant dans la liaison des dimères de TWIST1 à l'ADN puisque la mutation ponctuelle associée au Syndrome de Saethre-Chotzen S144R abolit la liaison TWIST1-ADN (El Ghouzzi et al., 2001; Maia et al., 2012). Au cours de notre analyse nous avons pu confirmer l'importance de certains résidus de TWIST1 localisés au niveau des boucles. En effet, les deux insertions au niveau des boucles de TWIST1 (Ins-135 et Ins-139) que nous avons étudiées *in silico* impactaient fortement la capacité de la sérine 144 (résidu situé au niveau de la boucle inter-hélicale) à interagir avec la première base de l'E-box contenue dans l'ADN aussi bien au niveau des homodimères que des hétérodimères. Notre étude montre le rôle clef de stabilisateur que jouent les boucles *via* les liaisons hydrogènes directes établies entre les acides

aminées de ces boucles et l'ADN, pour stabiliser les complexes de TWIST1 sur l'ADN. De plus nous avons montré que la boucle de E12 dans le dimère TWIST1/E12 avait une forme caractéristique unique présentant une cavité créant ainsi une cible spécifique pour un modulateur pharmacologique potentiel de l'hétérodimère de TWIST1/E12. Cette approche nous a également permis d'étudier l'évolution de la stabilité de nos dimères sauvages ou mutants au cours du temps. Grâce à ces données, nous avons mis en évidence le rôle clef des boucles du domaine bHLH de TWIST1 et expliqué de manière structurale la perte fonctionnelle associée aux mutations/insertions retrouvées chez les patients atteints du syndrome de Saethre-Chotzen.

Compte tenu de la réactivation aberrante des facteurs de transcription embryonnaires au cours de la transformation maligne et de la dissémination métastatique, mieux comprendre leurs mécanismes moléculaires sous-jacents permettrait ouvrir la voie à de nouvelles approches thérapeutiques. En effet, la description et la caractérisation des complexes de TWIST1 permettrait de faire des complexes de TWIST1 une cible intéressante pour des études de criblages moléculaires *in silico* afin de trouver des inhibiteurs de TWIST1 efficaces. Ainsi des études complémentaires permettant de définir la forme dimérique et oncogénique de TWIST1 ont permis de compléter ces résultats *in silico* et de définir des zones structurales cibles et spécifiques de la forme oncogénique active de TWIST1.

Interhelical loops within the bHLH domain are determinant in maintaining TWIST1–DNA complexes

Charlotte Bouard^{a,b,c,d,e,f}, Raphael Terreur^{f,g,h,i}, Jennifer Hope^{a,h,c,d,e,f}, Julie Anne Chemelle^{f,g,h,i},
Alain Puisieux^{a,h,c,d,e,f,g,j}, Stéphane Ansieau^{a,b,c,d,e,f,g} and Léa Payen^{a,b,c,d,e,f,g,k*}

^aCentre de Recherche en Cancérologie de Lyon, Lyon F-69000, France; ^bCNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, Lyon; ^cINSERM U1052 Centre de Recherche en Cancérologie de Lyon, Lyon, France; ^dCentre Leon Berard, Lyon, France; ^eLabEx DEVweCAN, Lyon, France; ^fUniversité de Lyon, Lyon, France; ^gUniversité Lyon 1, ISPB, Lyon; ^hInstitut de Biologie et Chimie des Protéines (IBCP) Lyon, France; ⁱCNRS UMR 5086, Lyon, France; ^jInstitut Universitaire de France, Lyon, France; ^kHospices Civils of Lyon, Biochemistry Laboratory of Lyon Sud (CHLS), Lyon F-69495, France

Communicated by Ramaswamy H. Sarma

(Received 14 September 2012; final version received 23 December 2012)

The basic helix-loop-helix (bHLH) transcription factor TWIST1 is essential to embryonic development, and hijacking of its function contributes to the development of numerous cancer types. It forms either a homodimer or a heterodimeric complex with an E2A or HAND partner. These functionally distinct complexes display sometimes antagonistic functions during development, so that alterations in the balance between them lead to pronounced morphological alterations, as observed in mice and in Saethre–Chotzen syndrome patients. We, here, describe the structures of TWIST1 bHLH–DNA complexes produced *in silico* through molecular dynamics simulations. We highlight the determinant role of the interhelical loops in maintaining the TWIST1–DNA complex structures and provide a structural explanation for the loss of function associated with several TWIST1 mutations/insertions observed in Saethre–Chotzen syndrome patients.

An animated interactive 3D complement (I3DC) is available in Proteopedia at <http://proteopedia.org/w/Journal:JBSD:27>

Keywords: TWIST1; dimerization; DNA interaction; homology models; 3D-models; molecular dynamics; embryonic transcription factors; bHLH

Introduction

The basic helix-loop-helix (bHLH) nuclear transcription factors belong to an evolutionarily conserved superfamily of nearly 125 proteins playing pivotal roles in numerous essential cellular functions, including adaptation to micro-environmental changes, specification, differentiation, proliferation, and apoptosis. Although they were previously classified on the basis of a combination of functional features (Atchley & Fitch, 1997), unbiased sequence comparisons across proteins of numerous species have led to their classification into five clades (Stevens, Roalson, & Skinner, 2008). For information, the TWIST, HAND, and NEUROD proteins belong to clade A, the E2A proteins to clade B, and MYOD1 to clade C (Skinner, Rawls, Wilson-Rawls, & Roalson, 2010).

The bHLH structure comprises a motif of mainly basic residues (the b-domain), allowing protein binding to 5'-CANNTG-3' cis-regulatory elements called E-boxes (De

Masi et al., 2011; Murre, 1999), and a dimerization motif of 50 residues forming two amphipathic α -helices separated by an interhelical loop (the HLH domain) (Voronova & Baltimore, 1990). Post-translational regulations, including protein modifications and competition between partners for binding, are described to affect complex formation and the resulting bHLH transcription factor activities. The regulation of TWIST1 during embryonic development and its alteration in human syndromes are classical examples of this complexity. TWIST1 was originally identified in *Drosophila* (the *Tw* gene) as a zygotic developmental gene crucial to the establishment of the ventral furrow, a prerequisite to the development of all mesoderm-derived internal organs (Thisse, el Messal, & Perrin-Schmitt, 1987). In mice, *Twist1* haplo-insufficiency produces viable offspring but leads to abnormal craniofacial structures and polydactyly on the hind limb, a phenotype reminiscent of the dominantly inherited Saethre–Chotzen syndrome (SC) in

*Corresponding authors. Email: stephane.ansieau@lyon.unicancer.fr, lea.payen-gay@univ-lyon1.fr

humans, where *TWIST1* is mutated (Bourgeois et al., 1998). *TWIST1*-depleted mice die at E10.5–11 because of numerous defects reflecting a role of this transcription factor in neural crest cell migration and determination and in mesoderm differentiation (Hebrok, Wertz, & Fuchtbauer, 1994; Lee, Lowe, Strong, Wergedal & Glackin, 1999; Spicer, Rhee, Cheung, & Lassar, 1996). *TWIST1* has been found to dimerize, either with itself or with an E2A (E12 or E47) or HAND (HAND1 or HAND2) partner, and *TWIST1* dimer selection has been shown to determine both the regulation of limb development (Firulli et al., 2005; Firulli, Redick, Conway, & Firulli, 2007) and cranial suture patterning and fusion (Connemey et al., 2006, 2008). Mutations in *TWIST1*, as observed in SC patients, favor homodimerization, either by reducing the level of the protein (monoallelic nonsense mutations) or by disrupting the Ser123 phosphorylation that dictates partner selection (El Ghouzzi et al., 2000, 2001; Firulli et al., 2005; Vichalkovski, Gresko, Hess, Restuccia, & Hemmings, 2010).

While undetectable in most differentiated adult cells, *TWIST* genes (*TWIST1* and *TWIST2*) are aberrantly reactivated in numerous human cancer types, including carcinoma (breast, lung, and colon), sarcoma, melanoma, glioma, and retinoblastoma (Ansieau, Morel, Hinkal, Bastid & Puisieux, 2010). Because it promotes the epithelial-to-mesenchymal transition (EMT), *TWIST1* was originally depicted as a prometastatic factor (Yang, Mani, & Weinberg, 2006). *TWIST* proteins actually display additional oncogenic properties. They abolish the activation of failsafe programs (senescence and apoptosis) in response to an oncogenic insult by turning down both RB- and p53-dependent oncosuppressive pathways (Ansieau et al., 2008; Kwok, Ling, Yuen, Wong, & Wang, 2007; Maestro et al., 1999; Valsesia-Wittmann et al., 2004), by inducing the AKT2 serine/threonine kinase-dependent pathway (Cheng et al., 2007), and by promoting expression of anti-apoptotic BH3-only proteins (Kwok et al., 2005; Wang et al., 2004; Zhuo, Wang, Zhuo, Zhang, & Chen, 2008). They thereby cooperate with mitogenic proteins, such as RAS and ERBB2, in promoting cell transformation *in vitro* and tumor development *in vivo* (Morel et al., 2012; Tran et al., 2012). Both intrinsic properties (Pham et al., 2007) and associated EMT (Kajiyama et al., 2007; Li et al., 2009; Yang et al., 2006) additionally lead to cell chemoresistance. Lastly, cell commitment to EMT is associated with reacquisition of stem-cell-like properties, including self-renewal potential, an important prerequisite for tumor development (Mani et al., 2008; Morel et al., 2008; Vesuna, Lisok, Kimble, & Raman, 2009). Furthermore, EMT associated reprogramming facilitates cell transformation *in vitro* (Morel et al., 2012). *TWIST* proteins, thus, display pleiotropic oncogenic and prometastatic properties favoring tumor initiation, development, and dissemination.

When *TWIST* proteins are depleted *in vitro* or *in vivo* in cancer cells, senescence and apoptotic programs are restored (Ansieau et al., 2008; Kwok et al., 2007; Tran et al., 2012). This and the accelerated senescence and decreased lifespan of SC patient osteoblasts as compared to control cells (Cakouros et al., 2012) suggest that *TWIST* proteins may constitute appealing therapeutic targets. In the present study, we have established, on the basis of homology with the X-ray structure of the bHLH domains of the murine E47/NEUROD1 complex, homology models of the human *TWIST1* homodimer and heterodimers (formed with either the E12 or HAND1 dimerization partner) bound to their target DNA sequences. Although the protein backbone is similar for all of our *in silico*-generated models, the *TWIST* complexes differ substantially in their interhelical loops. Alterations of these loops, as observed in SC patients, significantly alter the stability of *TWIST* dimer–DNA complexes by inducing conformational changes. This suggests that these structures may constitute prime targets for pharmacological inhibitors.

Materials and methods

Selection of X-ray structure and model building

The primary sequences of human *TWIST1* (Q15672), *TWIST2* (Q8WVJ9), E12 (P15923-1), HAND1 (O906004) HAND2 (P61296), NEUROD1 (Q13562), E47 (P15923-2), and MYOD1 (P15172) were downloaded from the UniProtKB/Swiss-Prot website. The mouse NEUROD1 and E47 sequences were obtained from the 2ql2 PDB files. Primary sequences, restricted to the bHLH domains, were aligned with the ClustalW 2.0 software (Thompson, Higgins, & Gibson, 1994).

The X-ray structures of the human MYC/MAX dimer (PDB code lmkp) (Nair & Burley, 2003), mouse MYOD1/MYOD1 complex (PDB code lmdy) (Ma, Rould, Weintraub, & Pabo, 1994), and mouse NEUROD1/E47 complex (PDB code 2ql2) (Longo, Guanga, & Rose, 2008) were downloaded from the RCSB protein data bank. Two X-ray structures of mouse NEUROD1/E47 have been reported (2ql2A/2ql2B; 2ql2C/2ql2D), which slightly diverge (root-mean-square deviation, RMSD). The calculated RMSD (based on align seed residue and list distance parameters) on C-chains is weak and evaluated at 0.618 Å, suggesting that the two models can be considered similar.

Sequence alignments were submitted to the SWISS-MODEL Workspace (Arnold, Bordoli, Kopp, & Schwede, 2006) to generate models for homodimeric and heterodimeric *TWIST1* complexes by homology with the NEUROD1(2ql2D)/E47(2ql2C) or NEUROD1(2ql2B)/E47(2ql2A) template, respectively, as described in (Maia, da Silva, Mencialha, Caffarena, & Abdelhay, 2012). The RMSDs between the models generated by Swiss Model Workspace and the folds are, respectively, 1.065/0.069 Å

and 1.60/0.064 Å for TWIST1/E12 (modeled reference 2ql2B/A) and TWIST1/TWIST1 (modeled with the reference 2ql2D/C). The choice between the two homology models was finally made by calculating the RMSD of each dimer with respect to its reference. Homology modeling was performed with the same DNA sequence as used in the reported NEUROD1/E47 X-ray structure. The dynamics simulations represent series of possible conformational states of the DNA-bound complexes.

To gain further insight into TWIST1 dimer function and specificity, we generated additional models with two functionally active TWIST1 variants expressed in SC patients. These variants result from a 21-bp tandem repeat insertion into the *TWIST1* gene, leading to the aberrant presence of seven extra amino acids in the interhelical loop at position 135 or 139 (Ins-135 or Ins-139, Figure S5) (El Ghouzzi et al., 1997). We, thus, built homology models of the following complexes: TWIST1/TWIST1, TWIST1 Ins-135/TWIST1 Ins-135, TWIST1 Ins-139/TWIST1 Ins-139, TWIST1/E12, TWIST1 Ins-135/E12, and TWIST1 Ins-139/E12.

Minimizations (>10,000 steps with the conjugated gradient algorithm) were carried out with the Sybyl-X 1.1 software package, elaborated by the Tripos company. We applied the Tripos force field with the Gasteiger–Marsilli

partial charges and a dielectric constant of 80 to simulate an implicit water phase (the dielectric constant of water is 20.10 at 20 °C). No restraints were applied to our models. This step principally refines and corrects the positions of residue side chains. The effects of minimization on the homology models were estimated approximately by calculating their RMSDs before and after the minimization step. In the heterodimer, the RMSDs of the E12 and TWIST1 peptides were 1.58 and 1.97 Å, respectively. In the homodimer, the RMSDs of the two TWIST1 peptides were 3.21 and 1.69 Å. The E-loop, thus, seems more disordered than the T-loop in the homodimeric complex, in agreement with the NeuroD1/E47 X-ray structure (Longo et al., 2008). Nonetheless, the RMSD calculation during the minimization step showed that the E-loop structure converges to a stable conformation, which could be used in dynamic simulations.

Molecular dynamics analysis

The established homology models (including DNA sequences) were visualized with the VMD1.9.1 software (Humphrey, Dalke & Schulten, 1996; Jorgensen, Chandrasekhar, Madura, Impey & Klein, 1983). The resulting model was inserted into a parapipedic TIP3P solvent box

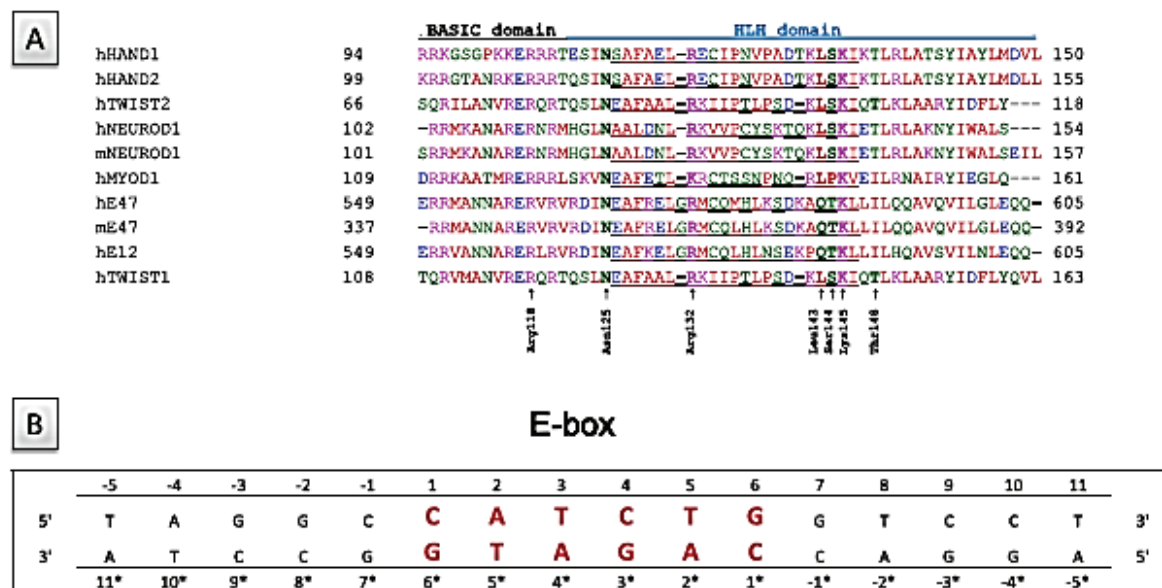


Figure 1. Primary sequences of the proteins and DNA sequence employed in the present study. (A) Primary sequence alignment of the bHLH domains of the NEUROD1, TWIST, HAND, and E2A proteins. TWIST1 residues identified as essential to the structure of the interhelical loops or to their binding to DNA are indicated. Residues within the interhelical loops are underlined, h and m stand for human and murine, respectively. (B) Sequence of the cis-regulatory element of the insulin promoter included in the NEUROD1/E47 X-ray structure and employed to establish the TWIST homology models.

To facilitate comparison of the DNA-binding properties of the dimers, numbering of the residues in each strand begins with the first base of the 5'-CATCTG-3' E-box core (uppercase bold). Consequently, the paired bases located on the opposite strand are differently numbered. The strands containing the CAT and CAG half-sites are bound to TWIST1 and E2, respectively.

by means of the add solvation box module of the VMD 1.9.1 software. A distance of 15 Å was set between the surfaces of the protein and the limit of the solvent box.

Conditions were computed to reach neutral charges before adding sodium and chloride to concentrations corresponding to physiological conditions. The model was minimized with the NaMD 2.8 bl software for 1000 steps before the molecular dynamics simulations (Phillips et al., 2005). It was computed on a 144 xeon core CPU cluster supercomputer (SGI Altix). Simulations were carried out at constant temperature (300 K) and pressure (1 atm) and by implementing the widely used CHARMM 27 force fields. The time step was set at 1 fs and Langevin dynamics was performed with a target piston pressure of 1.01325 bar and a damping coefficient of 5 ps^{-1} . There is no coupling of the Langevin temperature with hydrogen. The PME algorithms were applied with a grid extended by 10 Å from the periodic boundary condition size (Darden, Perera, Li, & Pedersen, 1999). The electrostatic cut-off was set at 14 Å. A conformation was sampled every 10 ps. As the solvent was described, the dielectric constant was set at 1. To identify steady conformations, 2D-RMSD calculations were carried out on 100 conformations selected, with a stride of 10, from the 1000 conformations produced during the 10-ns

simulation. The equilibrium state was reached around 30-ps for all studied models.

Results

Generation of 3D models of TWIST1 by homology with the NEUROD1/E47 X-ray structure

The X-ray structures of the NEUROD1/E47 and MYOD1/MYOD1 bHLH domains and MYC/MAX bHLH-LZ domains bound to the corresponding E-box sequences have been reported (Longo et al., 2008; Ma et al., 1994; Nair & Burley, 2003). To determine the suitable template structure for establishing by homology the structures homodimeric and heterodimeric TWIST1 complexes, primary sequence alignments were carried out. Use of the MYC/MAX structure was excluded, as the identity scores of TWIST1 and E12 vs. MYC and MAX are low (23 and 18% for TWIST1 and 17 and 9% for E12, respectively). Despite higher identity scores (37, 30, 35, 37, and 33% identity between MYOD1 and E47, E12, TWIST1, HAND2, and HAND1, respectively, Figure 1(A)), the MYOD/MYOD X-ray structure was also excluded, as the E-box sequence within it (5'-CAGCTG-3') is not recognized by TWIST complexes (De Masi et al., 2011) and as MYOD1 belongs to a different clade

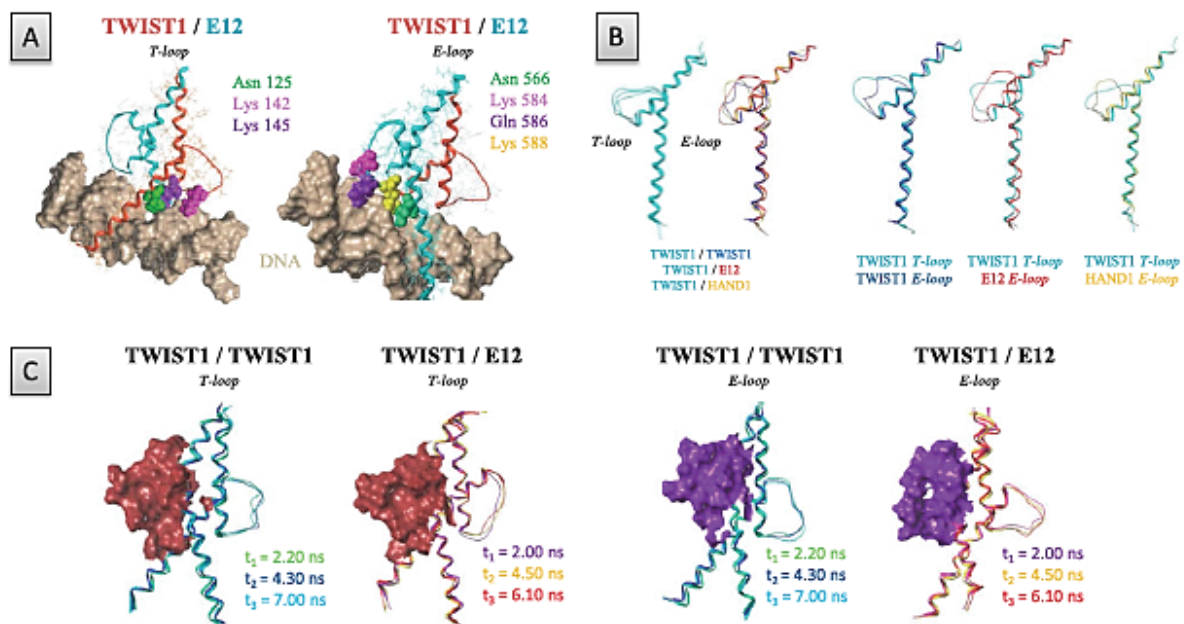


Figure 2. Comparison of the T- and E-loop structures in TWIST1 complexes. (A) Visualization of key residues within the T-loop and E-loop of the TWIST1/E12 heterodimeric complex, located at the vicinity of DNA. The protein structures are represented as ribbons. (B) Left panels: superposition of the T-loop and E-loop structures of the TWIST1/TWIST1, TWIST1/E12, and TWIST1/HAND1 dimers, as indicated at the bottom. Right panels: superposition of the TWIST1 T-loop with the E-loops of TWIST1, E12, and HAND1, as indicated at the bottom. Images correspond to the second steady states of the dynamics simulations. (C) Superposition of the E- and T-loop structures (represented as ribbons) of the different TWIST1 complexes as detected at the third steady state of the dynamics simulation. E- and T-loop surfaces are colored in violet and brown, respectively.

of bHLH proteins (Skinner et al., 2010). Strikingly, the mouse NEUROD1/E47 complex (2ql2 pdb file) meets all the criteria for constituting an appropriate template. In their bHLH domains, human E12 and E47 are 85 and 91% identical, respectively, to the murine E47 protein and TWIST1 is 48 and 29% identical, respectively, to the murine NEUROD1 and E47 proteins (Figure 1(A)). Moreover, the 5'-CATCTG-3' E-box sequence included in this X-ray structure (Figure 1(B)) has been reported as a TWIST-responsive element (Connerney et al., 2006; De Masi et al., 2011). Although the C-terminal sequences adjacent to the bHLH domain, known as the TWIST-box, constitute a protein interaction site essential for TWIST function (Bialek et al., 2004; Spicer et al., 1996), we reasoned that a structural model restricted to the bHLH domains should provide additional insights into the TWIST-DNA interface and highlight functional differences between TWIST1 complexes. Moreover, the extreme conservation (89% identity) between the two TWIST proteins (Figure 1) suggests that the results of the present work can most certainly be extrapolated to the TWIST2 protein.

Following the dynamics simulations, 2D-RMSD calculations taking the constraints of the template as a reference were carried out on all models. Analysis of the 2D-RMSD variations led to identifying stable conformations under physiological conditions. For each complex, three different steady states are achieved (Figure 1S, labeled 1–3). On the 2D-RMSD plots, these steady states are represented as the centers of yellow squares (Figure S1). As it took only 30-ps for each model to appear having achieved equilibrium, the 100-ps time point was excluded from the next analysis.

The resulting structural models highlight an interhelical loop between the two amphipathic α -helices, likely involved in protein-DNA complex stabilization and in specifying the target DNA sequence (Figure 2). This assumption is supported by the observation that in the TWIST1/E12 heterodimer, the highly conserved residues Asn125, Lys142, and Lys145 of the interhelical loop of TWIST1 and the corresponding residues Asn566, Lys584, and Lys588, as well as the residue Gln586 of the interhelical loop of E12, are all located in the vicinity of the DNA and likely contribute to DNA binding

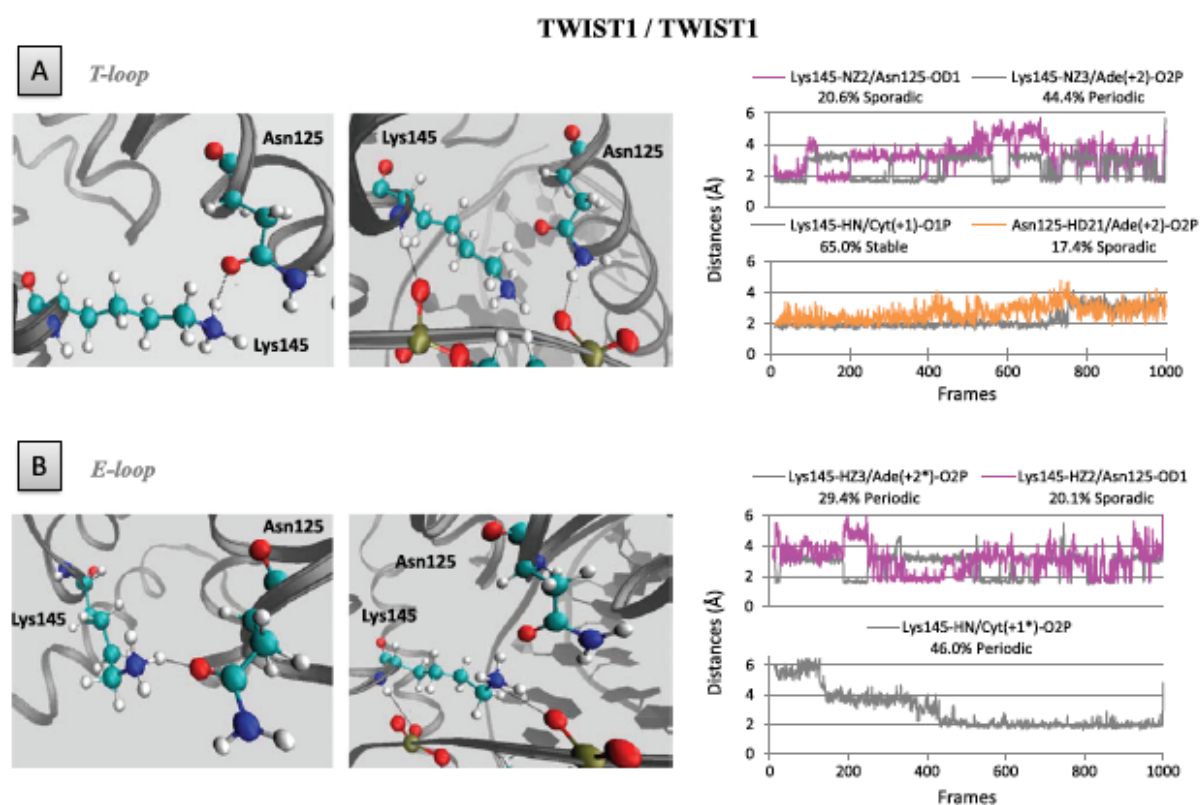


Figure 3. Hydrogen bonds involving the Asn125, Lys142, and Lys145 residues in the TWIST1 lateral loops. Left panels: positions of residues Asn125, Lys142, and Lys145 within the T-loop (A) and E-loop (B) of the TWIST1 homodimeric complex. Right panels: distances between residues and/or bases during the dynamics simulation are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames per ns). H-bonds are established when the two functional groups are less than 2 Å apart.

through hydrogen bonds (H-bonds) or allosteric effects (Figure 2(A)).

We next compared the dimer structures. For convenience, we arbitrarily called ‘T-loop’ the interhelical loop of the monomer modeled on the basis of the structure of murine NEUROD1 and ‘E-loops’ those of monomers modeled on the basis of the E47 structure. While the backbone of the T-loop showed similar conformations in the homodimer and different heterodimers, the backbone of the E-loop was found to diverge significantly between dimers (Figure 2(B)), because the amino acids composing the sequences of the different TWIST1 partners differ as regards length, polarity, charge, and hydrophobicity (Figure 1). Significant divergence was found between the T-loop and E-loop of the TWIST1 homodimer (Figure 2(B)). This divergence may be largely attributable to the use of a heterodimeric complex as a template. The E-loop of mE47 is quite disordered in the X-ray structure, but the 10-ns dynamics simulations showed its stability in space and time. The loops of the TWIST1/TWIST1 complex reach equilibrium at

5.5 ns (Figure S1(B)). This stabilization is likely due largely to interaction with DNA, as described below. We, then, investigated whether the conformation of the lateral loops might depend on the strand of DNA associated with the monomer, as E-boxes are asymmetrical. For this, we inverted the orientation of the DNA in the TWIST1/TWIST1 model, and carried out an additional 10-ns dynamics simulation. This inversion was found to affect the loop conformations only slightly, as symmetrical interactions between Arg118 and the DNA were found (Figure S2). We conclude that the divergence between T- and E-loops within the TWIST homodimeric complex is unlikely due to sequence difference between the two DNA strands.

Taken together, our data demonstrate that TWIST1 complexes differ significantly in the conformation of their interhelical loops. Remarkably, unlike the T- and E-loops of the TWIST1 homodimer, the E-loop of the heterodimer was found to form a deep cavity that might potentially be targeted in an effort to select specific inhibitors (Figure 2(C)).

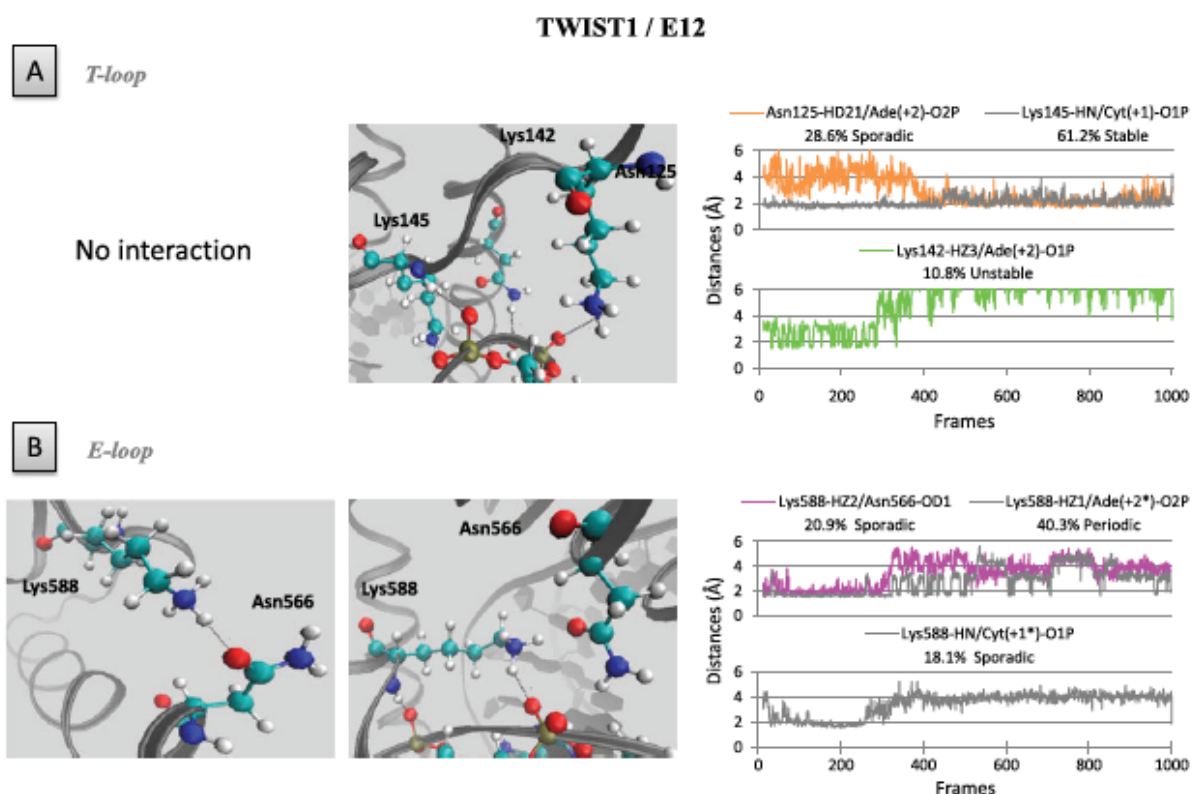


Figure 4. Hydrogen bonds involving the Asn125, Lys142, and Lys145 residues of TWIST1 and the Asn566 and Lys588 residues of E12 within the lateral loops of the TWIST1/E12 heterodimeric complex.

Left panels: positions of the Asn125, Lys142, Lys145, Asn566, and Lys588 residues within the T-loop (A) and E-loop (B) of the TWIST1/E12 heterodimeric complex. Right panels: distances between residues and/or bases during the dynamics simulation are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames per ns). H-bonds are established when the two functional groups are less than 2 Å apart.

Structural determinants of the interhelical loops

We next aimed to identify residues essential for interhelical loop conformation. Although these loops do not have a structural organization strictly speaking, H-bond formation can stiffen their structure. By limiting our study to H-bonds with an occupancy exceeding 5% of the total simulation and by excluding H-bonds between residues of different protein partners or involved in α -helix structures, we were led to focus on four TWIST1 residues: Lys145, Ser144, Arg132, and Arg118.

Our results show that in the T-loop and E-loop of the TWIST1 homodimer, the -NZ2 functional group of Lys145 interacts with the oxygen atom of the -OD1 group of Asn125 residue (Figure 3(A)), while it fails to do so in the T-loop of the heterodimer (Figure 4). Interestingly, the functional groups of Lys145 and/or Asn125 additionally establish H-bonds with oxygen groups of DNA bases (Figure 3). In both the T-loop and E-loop of the homodimer, the Lys145 residue interacts alternatively with the oxygen of the Asn125 residue and with the Ade(+2) nucleotide of the E-box during the 4 first ns of the simulation, and sporadically during the rest of the simulation (Figure 3). In parallel, a second H-bond is

established between the Lys145 residue and the Cyt(+1) base (Figure 3). In addition to the Lys145-Cyt(+1) interaction detectable in both the T-loop and the E-loop, the Asn125-Ade(+2) interaction is additionally established in the T-loop (Figure 3(A)).

Both the Lys145 and the Asn125 residues within the T-loop of the TWIST1/E12 heterodimer also establish H-bonds with DNA bases (Cyt(+1) and Ade(+2), respectively) (Figure 4). It is worth noting that the Lys142-Ade(+2) interaction established in the first part of the simulation is partially transferred to the Asn125 residue in the course of the dynamics simulation (Figure 4). Similar interactions between the corresponding Lys588 and Asn566 residues of E12 and the Ade(+2) and Cyt(+1) nucleotides also take place sporadically within the E-loop (Figure 4). In summary, these observations show that Asn125 and Lys145 residues play a determining role in stabilizing TWIST1 complexes on the DNA, by contributing to the establishment of the interhelical loop conformation and by directly interacting with oxygen atoms of DNA bases. In support of this view, the K145E mutation in TWIST1 abolishes the stable binding both the TWIST1 homodimer and the TWIST1/E12

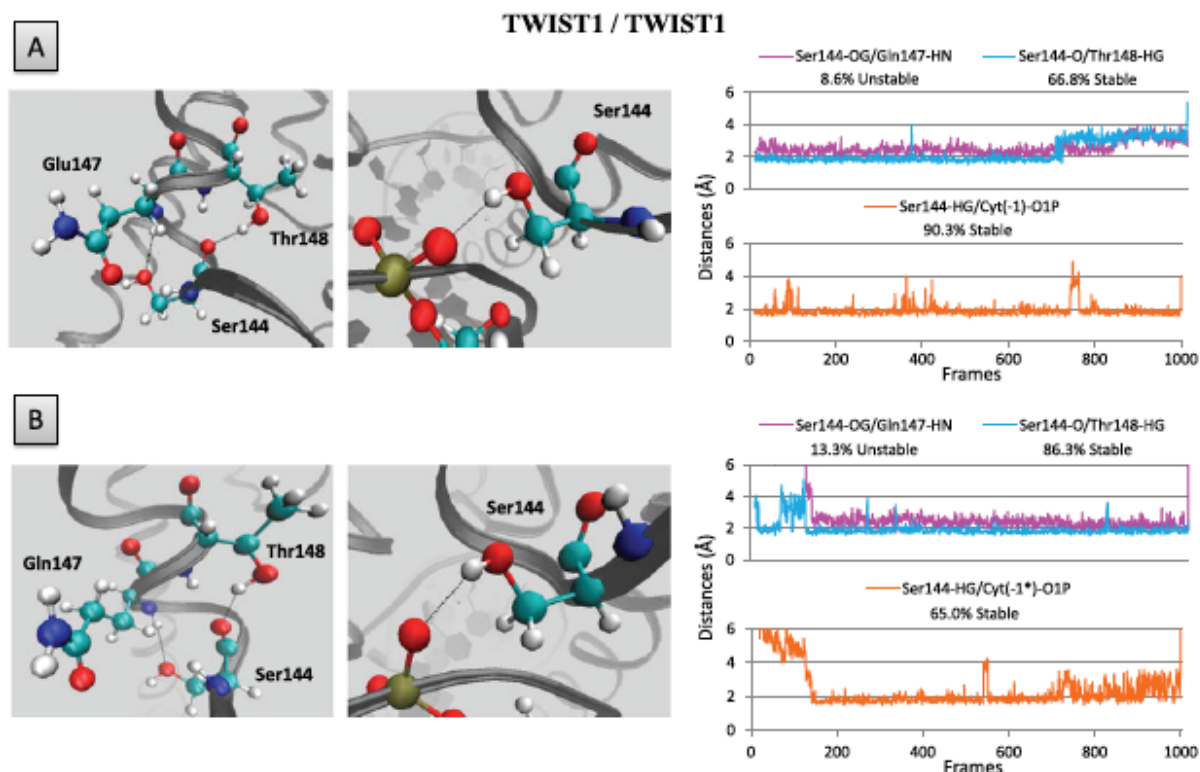


Figure 5. Hydrogen bonds involving the Ser144 residue of TWIST1.

Left panels: position of Ser144 residue within the T-loop (A) and E-loop (B) of the homodimeric TWIST1 complex. Right panels: distance between residues and/or bases during the dynamics simulation are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames per ns). H-bonds are established when the two functional groups are less than 2 Å apart.

heterodimer (Figure S3, (El Ghouzzi et al., 2001; Maia et al., 2012)).

In both the homodimeric and heterodimeric TWIST1 complexes, the Ser144 residue of TWIST1 and its counterpart Thr587 in E12 or Ser131 in HAND1 are invariably located along the bottom edge of a T- or E-loop (Figures 5 and 6 and S4). These residues form H-bonds with either Thr148 of TWIST1 (Figure 5), Leu590 (counterpart of Gln147) of E12 (Figure 6), or Lys134 and Thr135 (counterparts of Gln147 and Thr148) of HAND1 (Figure S4). While the Ser144–Thr148 interaction is fairly stable throughout the dynamics simulations in all homodimeric and heterodimeric complexes (Figures 5 and 6), the Ser144–Gln147 interaction appears to be less stable and even absent in the TWIST1/HAND1 heterodimer (Figure S4).

Interestingly, in both the T-loop and E-loop of the homodimer as well as in the E-loop of the TWIST1/E12 heterodimer, the HG group of Ser144 residue additionally establishes stable H-bonds with the O1P group of Cyt(–1) nucleotide (Figure 5). Surprisingly, the Ser144–Cyt(–1) and Ser131–Cyt(–1) H-bonds are not conserved in

the TWIST1/HAND 1 heterodimer (Figure S4). Taken together, these data demonstrate that H-bonds are determinant in maintaining interhelical loop conformation and contribute to determining the affinity of TWIST1 complexes for E-boxes. In support of this view, the S144R mutation in TWIST1 abolishes the stable binding to DNA of both the TWIST1 homodimer and the TWIST1/E12 heterodimer (Figure S3) (El Ghouzzi et al., 2001; Maia et al., 2012)).

Interestingly, in both the T- and E-loops of the TWIST1/TWIST1 and TWIST1/E12 dimers, the highly conserved residue Arg118 (Figure 1) additionally establishes stable H-bonds with both the O2P group of Cyt(+4) and the O2P group of Ade(+4^{*}) bases (Figure 7). The fourth base of the E-box is, thus, important in stabilizing TWIST1 complex binding to DNA. When the DNA is placed in the opposite orientation in the TWIST1 homodimeric model, the H-bonds between residue Arg118 and the fourth base of the E-box (Ade(+4^{*}) or Thy(+4)) are conserved (Figure S2(B)). We conclude that E-box asymmetry most likely does not determine either the sequence targeted or the stability of the

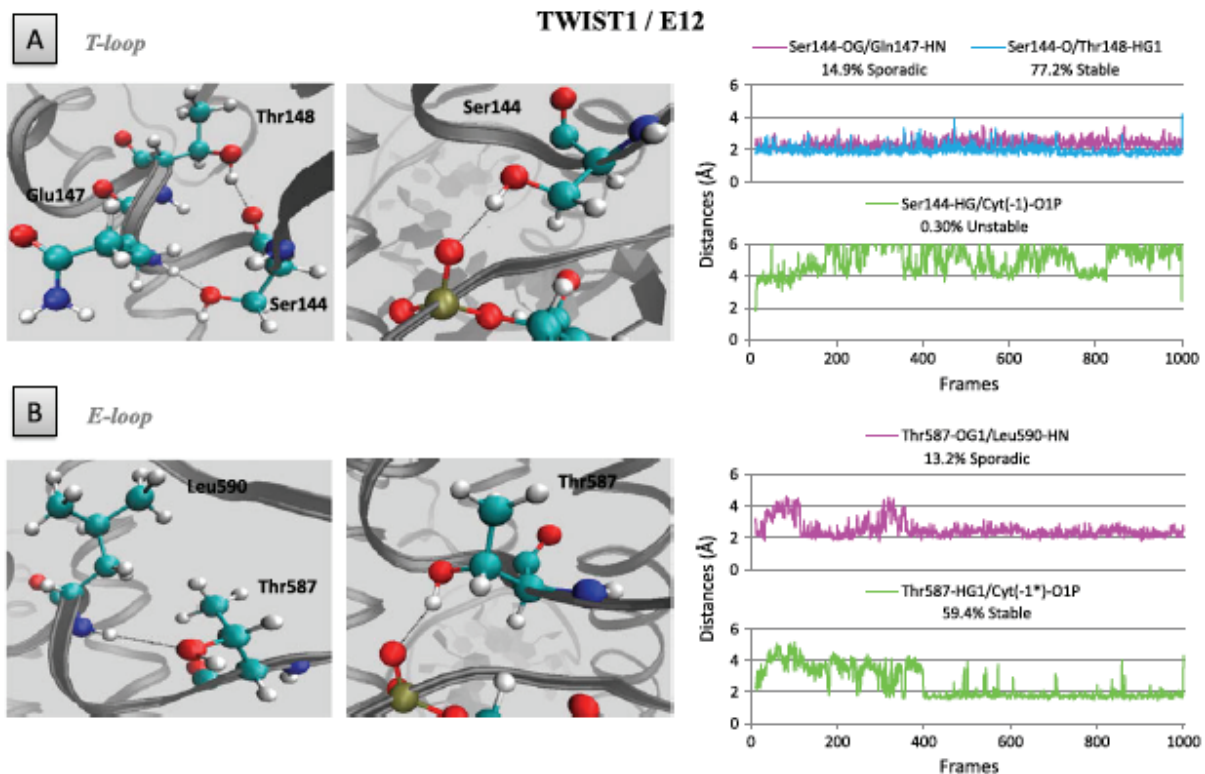


Figure 6. Hydrogen bonds involving residue Ser144 of and its counterpart Thr587 in E12. Left panels: position of the Ser144 residues within the T-loop (A) and E-loop (B) of the TWIST1/E12 heterodimeric complex. Right panels: distances between residues and/or bases during the dynamics simulation are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames per ns). H-bonds are established when the two functional groups are less than 2 Å apart.

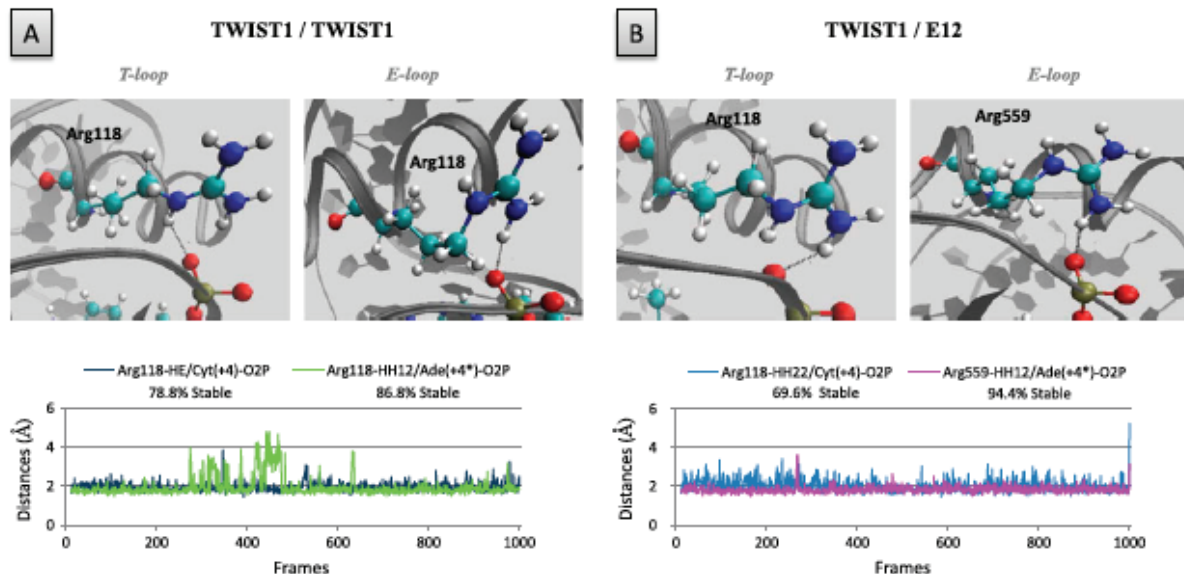


Figure 7. Hydrogen bonds involving the Arg118 residue of TWIST1 and its Arg559 counterpart in E12. The role of residues, Arg118 and Arg559, was examined in the context of both the TWIST1 homodimer (A) and the TWIST1/E12 heterodimer (B). Upper panels: positions of the Arg118 residues within the homodimeric TWIST1 complex (left panels) and of the Arg118 and Arg559 residues within the heterodimeric TWIST1/E12 complex (right panels). Lower panels: distances between residues and/or bases during the dynamics simulations are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames per ns). H-bonds are established when the two functional groups are less than 2 Å apart.

protein/DNA complex. Maintenance of these interactions largely explains the stability of T- and E- loop conformations in the TWIST1 homodimeric complex (Figure S2(A)).

In all TWIST1 complexes examined, the Arg132 residue is invariably oriented towards the inside of the T-loop (Figure 8). In both the homodimer and the TWIST1/E12 heterodimer, interestingly, this residue establishes H-bonds with the oxygen atoms of both Leu43 and Thr48, although the configuration adopted and the timing of bond formation differs according to the complex. In the homodimer, Arg132 interacts concomitantly with both residues during the first 2.88-ns of the dynamic simulation, i.e. in the first steady state identified, and both interactions are lost in the next two steady states. The T-loop, thus, evolves from a closed to an open conformation (Figure 8(A)). In the TWIST1/E12 heterodimer, Arg132 interacts alternatively with the two residues, the Arg132–Leu43 interaction being favored with time. The T-loop, thus, evolves from an open to a closed conformation (Figure 8(B)). In the TWIST1/HAND1 complex, Arg132 interacts concomitantly with both Thr48 and Leu43, with a preference for Thr48 over time (Figure S4). Despite these differences, Arg132 is likely to be determinant in establishing the T-loop conformation in all TWIST1 complexes. These H-bonds are not formed in the E-loop of the homodimer, or the bond with Ser140 forms only transiently (Figure 8); as a result,

the two helices come close together and reduce the space inside the E-loop. In support of the hypothesis that Arg132 plays a structural role in the T-loop only, the orientation of the side chain of its counterpart in E12 (Arg574) excludes an involvement in H-bonds. In contrast, in the E-loop of the TWIST1/HAND1 heterodimer, Arg118 (counterpart of Arg132) establishes an H-bond with Asn23 (Figure S5).

Insertions in the interhelical loops of TWIST1, as observed in SC patients, affect both their structure and interaction with DNA

To gain further insights into the role of the interhelical loops in TWIST1 dimer function and specificity, we next focused on additional Saerthe–Chotzen associated TWIST1 variants with aberrant insertion of seven amino acids into the interhelical loop at position 135 or 139 (Ins-135 or Ins-139, Figure S5) due to the presence of a 21-bp tandem repeat in the *TWIST1* gene (El Ghouzzi et al., 1997). Using a homology strategy similar to that described above, *in silico* models based on the murine NEUROD1/E47 X-ray structure were established and subjected to dynamics simulations. From the 2D-RMSD calculations, three steady states were selected as previously (Figure S6). The results show, as expected, that both insertions increase the length and surface of the interhelical loops and significantly move the T-loop towards the second helix of the HLH, this effect being

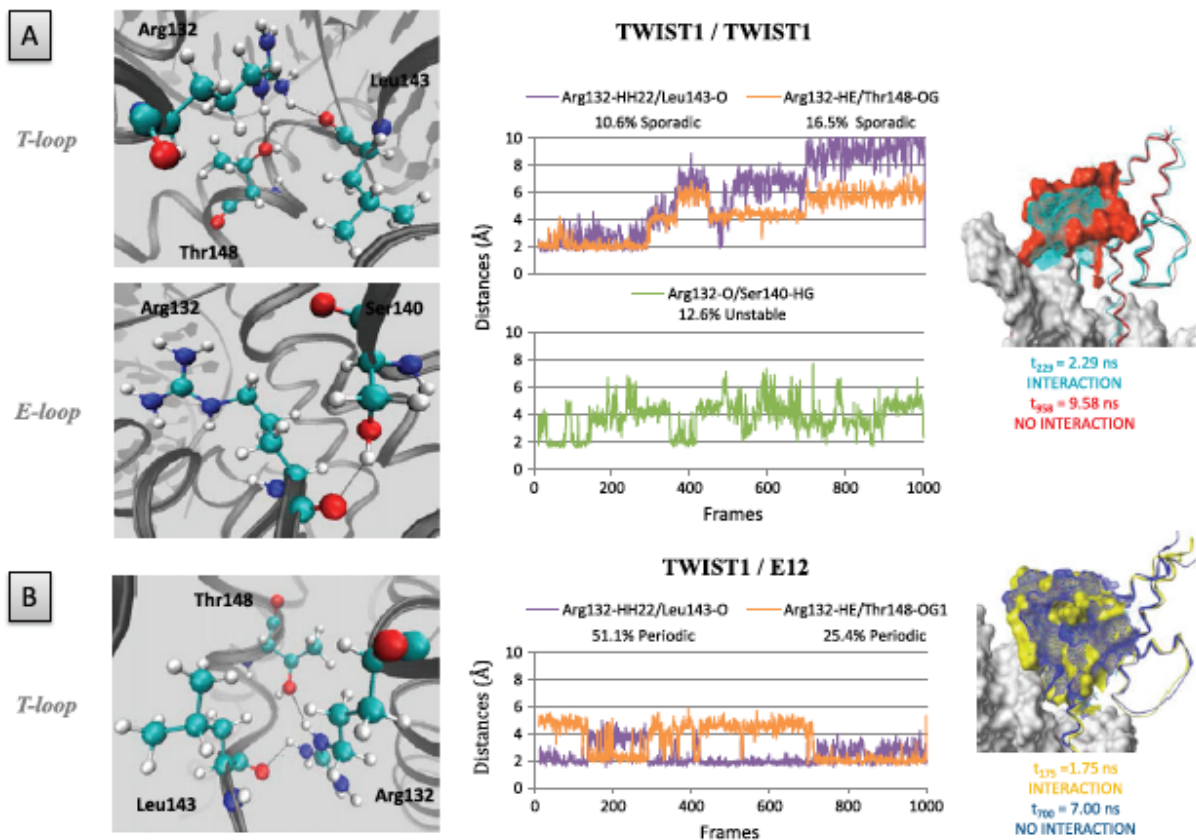


Figure 8. Hydrogen bonds involving residue Arg132 of TWIST1. The role of the residue was examined in the context of the TWIST1 homodimer (A) and TWIST1/E12 heterodimer (B). Left panels: positions of the Arg132 residues within the complexes. Central panels: distances between residues and/or bases during the dynamics simulations are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames = 1 ns). H-bonds are established when the two functional groups are 2 Å apart. Right panels: superposition of the complex structures as detected at different time points of the dynamics simulations.

more pronounced in the Ins-139 variant (Figure S5). Many interactions between loop residues are modified in complexes containing an Ins-variant of TWIST1 (e.g. the Arg132–Thr148 interaction is disrupted in the T-loops of homodimers containing either Ins-variant) (Table 1).

It is noteworthy that the two insertions have somewhat different effects (e.g. the Arg132–Leu143 interaction is disrupted in the T-loop of the Ins-135 homodimer and is fully stabilized in that of the Ins-139 homodimer) (Table 1). Conversely, the insertions create some novel concomitant H-bonds (Met575–His593, Gln577–Gln594, and Gln577–Glu583 in the E-loop of the TWIST1 Ins139/E12 heterodimer, Figure 9), Lys145–Thr148 in the T-loop and Arg139–Asp157 in the E-loop of the TWIST1 Ins-135 homodimer, (Figure 10), stiffening the structures of these loops. In the TWIST1 Ins-139 homodimer, some newly established H-bonds involve aberrantly inserted residues (Lys (Ins+1)–Asp141, Lys (Ins+1)–Pro (Ins+7)) (Figure 10).

The insertions additionally alter certain residue–DNA interactions (e.g. Ser144–Cyt(–1) in the T-loops of the homodimers containing either insertion variant) and create novel ones (e.g. Ser144–Cyt(–1) in the T-loops of the heterodimers containing either insertion variant) (Table 1). Taken together, these data suggest that both insertions distort the interhelical loops, thus, greatly affecting protein/DNA complex stability.

Discussion

As a regulator of cell migration and specification, the TWIST1 transcription factor is essential to embryonic development in multiple species (El Ghouzzi et al., 1997). As a reminiscence of its embryonic functions, its aberrant reactivation observed in numerous cancer types was originally associated with cancer cell dissemination (Yang et al., 2004). Further investigations have led to attributing to TWIST1 additional functions in malignant

Table 1. Main H-bonds established within the interhelical loops of the various TWIST1 complexes. Interactions are either established (% occupancy is indicated) or absent (-) over the time course of the dynamics simulations minus the first 100 ps. The percentage of occupancy represents the time during which the H-bond is established. Stable, periodic, and sporadic interactions are labeled in dark grey, light grey, white respectively.

A	DNA	TWIST1 corresponding residue	TWIST1					E12 corresponding residue	TWIST1		
			ins-135/ins-135	ins-139/ins-139	E12	ins-135/E12	ins-139/E12		ins-139/E12		
T-loop interaction		Ser144-Cyt(-1)	90.30%	-	-	-	-	89.50%	79.94%	-	
		Lys145-Asn(+2)	44.44%	37.78%	-	-	-	53.67%	47.97%	-	
		Asn125-Asp(+2)	17.37%	18.23%	-	-	28.59%	49.83%	3.49%	-	
		Lys145-Cyt(+1)	64.95%	53.63%	27.43%	-	61.21%	83.67%	-	-	
		Lys142-Asp(+2)	-	-	-	-	10.81%	16.17%	-	-	
Residue interaction		Lys145-Asn125	20.61%	24.83%	25.66%	-	-	-	19.77%	-	
		Ser144-Thr148	66.77%	74.24%	62.61%	-	77.17%	89.67%	48.99%	-	
		Ser144-Gln147	8.59%	-	35.62%	-	14.85%	-	42.90%	-	
		Arg132-Thr148	16.46%	-	-	-	25.35%	8.17%	-	-	
		Arg132-Leu143	10.61%	6.34%	60.62%	-	51.11%	48.17%	9.28%	-	
	Arg132-Ser140	-	-	-	-	-	-	-	-		
B	DNA	TWIST1 corresponding residue	TWIST1					E12 corresponding residue	TWIST1		
			ins-135/ins-135	ins-139/ins-139	E12	ins-135/E12	ins-139/E12		ins-139/E12		
E-loop interaction		Ser144-Cyt(-1*)	65.00%	36.72%	-	-	Thr587-Cyt(-1*)	59.39%	-	-	
		Lys145-Asp(+2*)	29.40%	57.73%	49.34%	-	Lys588-Asp(+2*)	40.30%	39.17%	55.65%	
		Asn125-Asp(+2*)	7.30%	17.97%	-	-	Asn566-Asp(+2*)	7.17%	3.00%	5.51%	
		Lys145-Cyt(+1*)	46.0%	37.91%	74.78%	-	Lys588-Cyt(+1*)	18.08%	71.33%	83.48%	
		Lys142-Asp(+2*)	-	-	-	-	Lys586-Asp(+2*)	-	-	-	
Residue interaction		Lys145-Asn125	20.10%	19.55%	21.68%	-	Lys588-Asn566	20.90%	21.0%	5.22%	
		Ser144-Thr148	86.26%	52.58%	42.48%	-	Iso591-Thr587	-	25.0%	13.33%	
		Ser144-Gln147	13.33%	-	31.86%	-	Leu590-Thr587	13.23%	-	5.51%	
		Arg132-Thr148	-	40.42%	-	-	Arg574-Leu590	-	-	-	
		Arg132-Leu143	-	8.19%	-	-	Arg574-Gln586	-	-	-	
	Arg132-Ser140	12.63%	-	-	-	Arg574-Ser582	-	-	-		

conversion and tumor progression, mainly involving fail-safe program inhibition and EMT induction (Ansieau et al., 2010). Interestingly, TWIST1 depletion in numerous cancer cell types is sufficient to reactivate senescence or apoptosis programs (Ansieau et al., 2008; Maestro et al., 1999; Tran et al., 2012; Valsesia-Wittmann et al., 2004), suggesting that TWIST1 inhibition may constitute a novel anti-cancer therapeutic route.

Beyond the difficulty of targeting nuclear factors, an additional complexity of TWIST1 activity arises from its ability to associate with different partners, i.e. with itself or with E12/E47 or HAND, to form functionally different homo- or heterodimers. The balance between these complexes determines TWIST1 function, and altered dimer selection has dramatic consequences, as demonstrated by the altered phenotype of mice and SC patients with TWIST1 haplo-insufficiency (Connerney et al., 2006; Firulli et al., 2005, 2007) and mice ectopically expressing TWIST1-tethered dimers (Connerney et al., 2008).

On the basis of this information, we have sought to compare the structures of different protein–DNA complexes to gain further insights into the functional consequences of the TWIST1 mutations/insertions encountered in SC patients, and potentially to identify structures that might be targeted to impair TWIST1 function.

Although the N- and C-terminal sequences of TWIST1 also play a role in TWIST1 functions, we have restricted our analysis to the bHLH domain, taking advantage of the available X-ray structure of the murine NEUROD1/E47 protein complex associated with its DNA target sequence (Longo et al., 2008). The high identity scores computed for the two clade-A proteins, TWIST1 and NEUROD1, and for the two E2A splice variants, E12 and E47, along with the fact that the E-box selected has been identified as a functional TWIST1-responsive cis-regulatory element in the promoting sequences of several of its target genes (Connerney et al., 2006; De Masi et al., 2011), argue strongly in

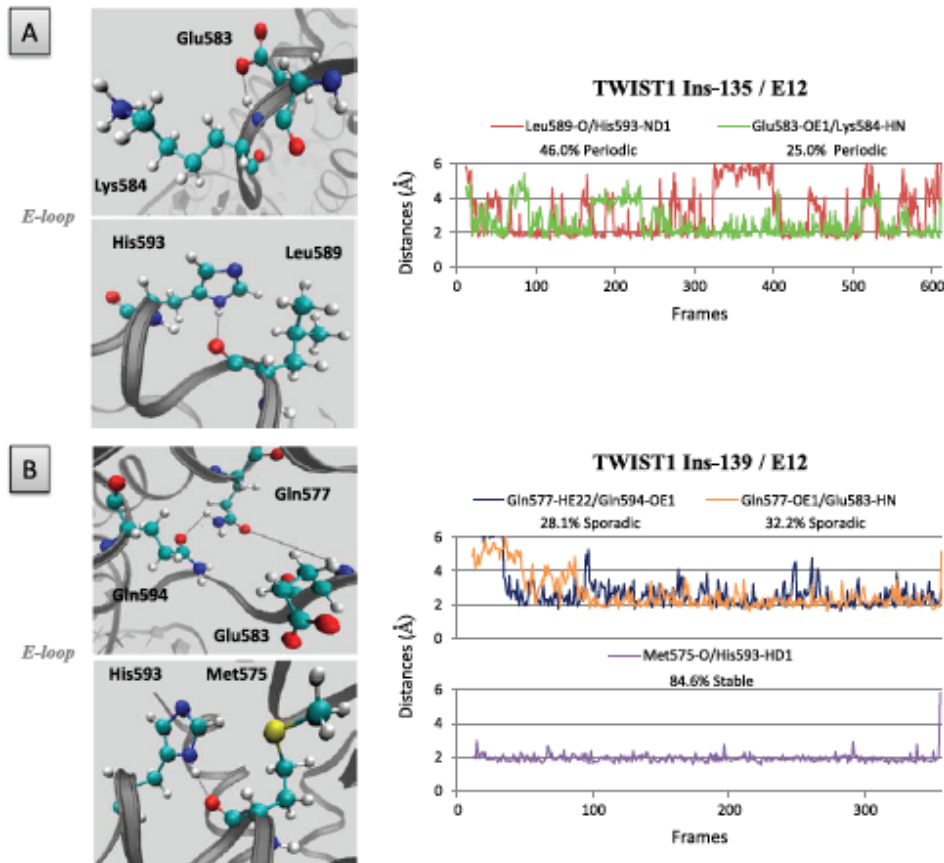


Figure 9. Amino acid insertions within the interhelical loops of the TWIST1/E12 heterodimer (Ins-135 and Ins-139), as observed in SC patients, generate novel H-bonds.

Left panels: positions of the aberrantly inserted residues involved in novel H-bonds within the interhelical loops in the Ins-135 TWIST1/E12 (upper panel) and Ins-139 TWIST1/E12 (lower panel) heterodimeric complexes. Right panels: distances between residues and/or bases during the dynamics simulations are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time 100 frames per 1 ns). H-bonds are established when the two functional groups are less than 2 Å apart.

favor of the reliability of the established TWIST1/E12 homology models. Although TWIST1 and HAND1 diverge more markedly from E47, the identity scores remain high enough to generate trustworthy homodimeric and TWIST1/HAND1 heterodimeric complexes.

Structural analysis highlights, between the two amphipathic α -helices, an interhelical loop whose conformation differs between the monomers forming a given TWIST1-containing dimer and according to the partner of TWIST1 in the dimer (Figure 2). We demonstrate that these structural divergences notably reflect differences in residue composition, as they affect the formation of H-bonds involving either residues or DNA bases. Interestingly, the E-loop of the TWIST1/E12 heterodimer forms a unique cavity that might reasonably be targeted in efforts to select specific intercalating molecules for pharmacological purposes. Preliminary docking analysis of the NCI library supports this hypothesis (data not shown).

Residues located in the loops of all of the homodimers and heterodimers studied interact with nucleotides -1 to $+4$ of the E-box sequence (Figures 3–8), known to be important in bHLH binding (De Masi et al., 2011; Ohno, Sadeh, Blatt, Brengman, & Engel 2003). In the light of the importance of these interactions, we have assessed the consequences of insertions within the interhelical loops (Ins-135 and Ins-139) (El Ghouzzi et al., 1997) observed in SC patients on TWIST1 complex structure. As expected, both insertions significantly increase the surface and strongly modify the conformations of the interhelical loops (Figures 9 and 10, S5 and S6). They notably abolish certain H-bonds between T-loop residues and promote the formation of novel ones (Table 1). The Ins-139 insertion additionally disrupts many interactions of the T-loop with DNA, especially within the homodimer (Figures 9 and 10 and Table 1). This observation corroborates the reported reduced

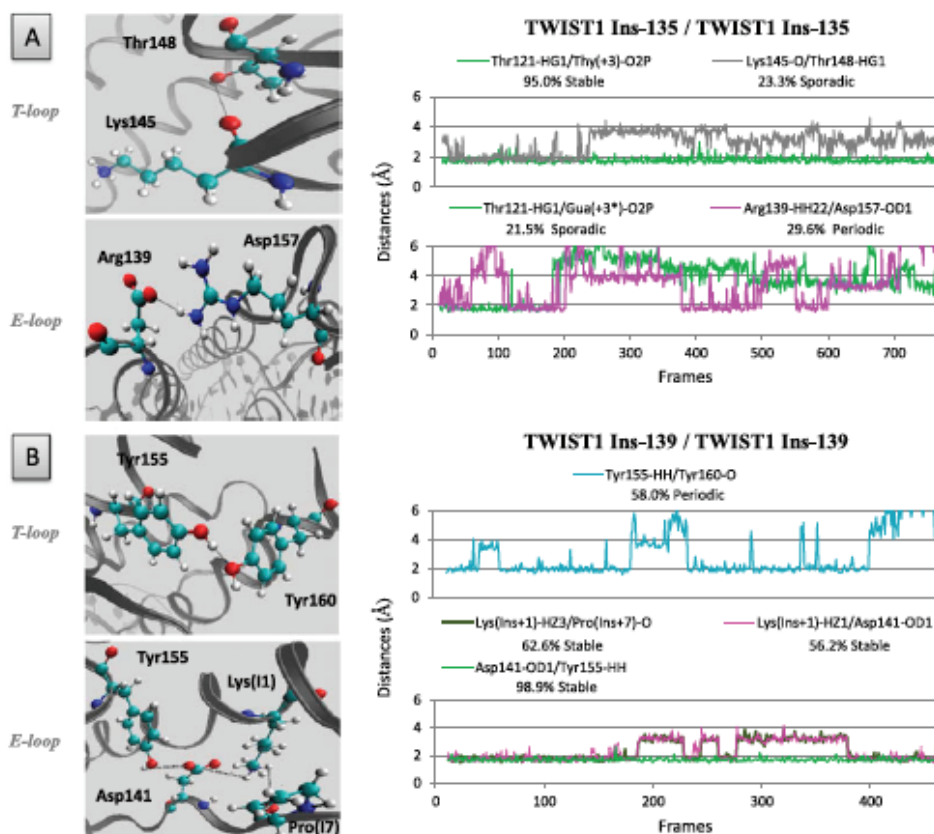


Figure 10. Amino acid insertions within the interhelical loops of the TWIST1/TWIST1 homodimer (Ins-135 and Ins-139), as observed in SC patients, generate novel H-bonds.

Left panels: positions of the aberrantly inserted residues involved in novel H-bonds within the interhelical loops of the Ins-135 (A) and Ins-139 (B) homodimeric TWIST complexes. Right panels: distances between residues and/or bases during the dynamics simulations are depicted as plots of distance from donor to acceptor atom (Å) versus frame number (total time, 100 frames per 1 ns). H-bonds are established when the two functional groups are less than 2 Å apart.

DNA-binding affinity of the TWIST1 Ins-139/E12 heterodimeric complex (El Ghouzzi et al., 2001).

Remarkably, the two SC-associated TWIST1 alterations are not equivalent, as the Ins-135 insertion disrupts much fewer contacts between interhelical loops and DNA (Table 1).

The relevance of our models is additionally supported by the analysis of two additional TWIST1 variants found in SC patients, characterized, respectively, by the point mutations K145E and S144R (El Ghouzzi et al., 2001). Residues 145 and 144 are located, respectively, at the center of the dimer and along the bottom edge of the interhelical loop, at the interface with the DNA (Figure S3). Both mutations appear to increase complex flexibility (Maia et al., 2012). The K145 substitution strongly impairs TWIST1 binding to DNA and only slightly reduces dimerization (El Ghouzzi et al., 2001). Lys145 actually interacts with the second, determinant nucleotide of the E-box. The K145E substitution probably abolishes this interaction through repulsion between the carbonyl group of Glu and the phosphate group on the DNA base. In addition, Ser144 in the T-loops of both the TWIST1 homodimer and the TWIST1/E12 heterodimer and its counterpart Thr587 in the E-loop of the latter are in close contact with Cyt(–1) (Figures 5 and 6). In support of the view that Ser144 plays a role in DNA binding, the SC-associated point mutation S144R was also found to abolish TWIST1–DNA binding, likely by destabilizing the DNA–protein complex (El Ghouzzi et al., 2001; Maia et al., 2012). Interestingly, the Ins-135 and Ins-139 insertions strongly alter the capacity of Ser144 (or Thr587) to interact with Cyt(–1) in both loops of both the homodimeric and the heterodimeric complex (Table 1). This provides an explanation for the reported reduced DNA affinity of the TWIST1 Ins-139/E12 heterodimer (El Ghouzzi et al., 2001). It is worth noting that in the TWIST1/HAND1 heterodimer, neither Ser144 in the TWIST1 T-loop nor its counterpart Ser131 in the HAND1 E-loop interacts with Cyt(–1) of the E-box. This highlights a functional divergence between TWIST1 complexes (Figure S4).

Taken together, our data support the view that interhelical loops within the bHLH play a determining role in maintaining TWIST1–DNA complex structures and provide a structural explanation for the loss of function associated with several TWIST1 mutations/insertions observed in SC patients.

Supporting Information

Supplementary material for this paper is available online at <http://dx.doi.org/10.1080/07391102.2012.762722>.

Supplementary material dealing with the “Inter-helical loops within the bHLH domain are determinant in maintaining TWIST1–DNA complexes” study is available from authors directly by e-mail transmission.

Author contributions

All authors have given approval to the final version of the manuscript. The manuscript was written through contributions of all authors. CB, RT, JH, JAC, and LP carried out experiments. RT, AP, SA, and LP designed the experimental strategy, and SA and LP wrote the manuscript.

Funding sources

Our team is labeled by the Ligue Nationale contre le Cancer. This work was additionally supported by institutional grants from the LabEX DEVweCAN (ANR-10-LABX-61) and from the LyRIC (Lyon Recherche Integree en Cancerologie, Institut National Contre le Cancer, INCa_4664).

Abbreviations

SC	Saethre–Chotzen syndrome
bHLH	basic helix-loop-helix
EMT	epithelial-to-mesenchymal transition

Acknowledgment

We thank all members of the laboratory for insightful discussions and Dr Kathleen Broman for critical reading of the manuscript.

References

- Ansieau, S., Bastid, J., Doreau, A., Morel, A. R., Bouchet, B. R., Thomas, C., ... Puisieux, A. (2008). Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell*, 14, 79–89.
- Ansieau, S., Morel, A. P., Hinkal, G., Bastid, J., & Puisieux, A. (2010). TWISTing an embryonic transcription factor into an oncoprotein. *Oncogene*, 29, 3173–3184.
- Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The Swiss-model workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195–201.
- Atchley, W. R., & Fitch, W. M. (1997). A natural classification of the basic helix-loop-helix class of transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 5172–5176.
- Bialek, P., Kern, B., Yang, X., Schrock, M., Sobic, D., Hong, N., ... Karsenty, G. (2004). A twist code determines the onset of osteoblast differentiation. *Developmental Cell*, 6, 423–435.
- Bourgeois, P., Bolcato-Bellemin, A. L., Danse, J. M., Bloch-Zupan, A., Yoshida, K., Stoetzel, C., & Perrin-Schmitt, F. (1998). The variable expressivity and incomplete penetrance of the twist-null heterozygous mouse phenotype resemble those of human Saethre–Chotzen syndrome. *Human Molecular Genetics*, 7, 945–957.
- Cakouros, D., Isenmann, S., Cooper, L., Zannettino, A., Anderson, P., Glackin, C., & Gronthos, S. (2012). Twist-1 induces recruitment regulating histone methylation along the Ink4A/Arf locus in mesenchymal stem cells. *Molecular and Cell Biology*, 32, 1433–1441.

- Cheng, G. Z., Chan, J., Wang, Q., Zhang, W., Sun, C. D., & Wang, L. H. (2007). Twist transcriptionally up-regulates AKT2 in breast cancer cells leading to increased migration, invasion, and resistance to paclitaxel. *Cancer Research*, *67*, 1979–1987.
- Connerney, J., Andreeva, V., Leshem, Y., Mercado, M. A., Dowell, K., Yang, X., Lindner, V., ... Spicer, D. B. (2008). Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure. *Developmental Biology*, *318*, 323–334.
- Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M. A., & Spicer, D. B. (2006). Twist1 dimer selection regulates cranial suture patterning and fusion. *Developmental Dynamics*, *235*, 1345–1357.
- Darden, T., Perera, L., Li, L., & Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*, *7*, R55–R60.
- De Masi, R., Grove, C. A., Vedenko, A., Alibes, A., Gisselbrecht, S. S., Serrano, L., Bulyk, ... Walhout, A. J. (2011). Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Research*, *39*, 4553–4563.
- El Ghouzzi, V., Le Merrer, M., Perrin-Schmitt, R., Lajeunie, E., Benit, R., Renier, D., ... Bonaventure, J. (1997). Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nature Genetics*, *15*, 42–46.
- El Ghouzzi, V., Legeai-Mallet, L., Aresta, S., Benoist, C., Munnich, A., de Gunzburg, J., & Bonaventure, J. (2000). Saethre-Chotzen mutations cause TWIST protein degradation or impaired nuclear location. *Human and Molecular Genetics*, *9*, 813–819.
- El Ghouzzi, V., Legeai-Mallet, L., Benoist-Lasselin, C., Lajeunie, E., Renier, D., Munnich, A., & Bonaventure, J. (2001). Mutations in the basic domain and the loop-helix II junction of TWIST abolish DNA binding in Saethre-Chotzen syndrome. *FEBS Letters*, *492*, 112–118.
- Firulli, B. A., Krawchuk, D., Centonze, V. E., Vargesson, N., Virshup, D. M., Conway, S. J., ... Firulli, A. B. (2005). Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nature Genetics*, *37*, 373–381.
- Firulli, B. A., Redick, B. A., Conway, S. J., & Firulli, A. B. (2007). Mutations within helix I of Twist1 result in distinct limb defects and variation of DNA binding affinities. *Journal of Biological Chemistry*, *282*, 27536–27546.
- Hebrok, M., Wertz, K., & Fuchtbauer, E. M. (1994). M-twist is an inhibitor of muscle differentiation. *Developmental Biology*, *165*, 537–544.
- Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38, 27–38.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, *79*, 926–935.
- Kajiyama, H., Shibata, K., Terauchi, M., Yamashita, M., Ino, K., Nawa, A. & Kikkawa, F. (2007). Chemoresistance to paclitaxel induces epithelial-mesenchymal transition and enhances metastatic potential for epithelial ovarian carcinoma cells. *International Journal of Oncology*, *31*, 277–283.
- Kwok, W. K., Ling, M. T., Lee, T. W., Lau, T. C., Zhou, C., Zhang, X., ... Wang, X. (2005). Up-regulation of TWIST in prostate cancer and its implications as a therapeutic target. *Cancer Research*, *65*, 5153–5162.
- Kwok, W. K., Ling, M. T., Yuen, H. F., Wong, Y. C., & Wang, X. (2007). Role of p14ARF in TWIST-mediated senescence in prostate epithelial cells. *Carcinogenesis*, *28*, 2467–2475.
- Lee, M. S., Lowe, G. N., Strong, D. D., Wergedal, J. E. & Glackin, C. A. (1999). TWIST, a basic helix-loop-helix transcription factor, can regulate the human osteogenic lineage. *Journal of Cellular Biochemistry*, *75*, 566–577.
- Li, Q. Q., Xu, J. D., Wang, W. J., Cao, X. X., Chen, Q., Tang, F., ... Xu, Z. D. (2009). Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multi-drug resistance and invasive potential in breast cancer cells. *Clinical Cancer Research*, *15*, 2657–2665.
- Longo, A., Guanga, G. P., & Rose, R. B. (2008). Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: Heterodimer selectivity and DNA recognition. *Biochemistry*, *47*, 218–229.
- Ma, P. C., Rould, M. A., Weintraub, H., & Pabo, C. O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell*, *77*, 451–459.
- Maestro, R., Dei Tos, A. P., Hamamori, Y., Krasnokutsky, S., Sartorelli, V., Kedes, L., ... Hannon, G. J. (1999). Twist is a potential oncogene that inhibits apoptosis. *Genes and Development*, *13*, 2207–2217.
- Maiá, A. M., da Silva, J. H., Mencialha, A. L., Caffarena, E. R., & Abdelhay, E. (2012). Computational modeling of the bHLH domain of the transcription factor TWIST1 and R118C, S144R and K145E mutants. *BMC Bioinformatics*, *13*, 184.
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., ... Weinberg, R.A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, *133*, 704–715.
- Morel, A. P., Hinkal, G. W., Thomas, C., Fauvet, F., Courtois-Cox, S., Wierinckx, A. ... Puisieux, A. (2012). EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS Genetics*, *8*, e1002723.
- Morel, A. P., Lievre, M., Thomas, C., Hinkal, G., Ansieau, S., & Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS ONE*, *3*, e2888.
- Murre, C. (1999). Role of helix-loop-helix proteins in lymphocyte development. *Cold Spring Harbor Symposia on Quantitative Biology*, *64*, 39–44.
- Nair, S. K., & Burley, S. K. (2003). X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, *112*, 193–205.
- Ohno, K., Sadeh, M., Blatt, I., Brengman, J. M., & Engel, A. G. (2003). E-box mutations in the RAPSN promoter region in eight cases with congenital myasthenic syndrome. *Human Molecular Genetics*, *12*, 739–748.
- Pham, C. G., Bubici, C., Zazzeroni, E., Knabb, J. R., Papa, S., Kuntzen, C., & Franzoso, G. (2007). Upregulation of Twist-1 by NF-kappaB blocks cytotoxicity induced by chemotherapeutic drugs. *Molecular and Cellular Biology*, *27*, 3920–3935.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*, 1781–1802.

- Skinner, M. K., Rawls, A., Wilson-Rawls, J., & Roalson, E. H. (2010). Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. *Differentiation*, *80*, 1–8.
- Spicer, D. B., Rhee, J., Cheung, W. L., & Lassar, A. B. (1996). Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science*, *272*, 1476–1480.
- Stevens, J. D., Roalson, E. H., & Skinner, M. K. (2008). Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: Genomic approach to cellular differentiation. *Differentiation*, *76*, 1006–1022.
- Thisse, B., el Messal, M., & Perrin-Schmitt, F. (1987). The twist gene: Isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic Acids Research*, *15*, 3439–3453.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*, 4673–4680.
- Tran, P. T., Shroff, E. H., Bums, T. F., Thiyagarajan, S., Das, S. T., Zabuawala, T ... Felsner, D. W. (2012). Twist1 suppresses senescence programs and thereby accelerates and maintains mutant Kras-induced lung tumorigenesis. *PLoS Genetics*, *8*, e1002650.
- Valsesia-Wittmann, S., Magdeleine, M., Dupasquier, S., Garin, E., Jallas, A. C., Combaret, V., ... Puisieux, A. (2004). Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer Cell*, *6*, 625–630.
- Vesuna, E., Lisok, A., Kimble, B., & Raman, V. (2009). Twist modulates breast cancer stem cells by transcriptional regulation of CD24 expression. *Neoplasia*, *11*, 1318–1328.
- Vichalkovski, A., Gresko, E., Hess, D., Restuccia, D. E., & Hemmings, B. A. (2010). PKB/AKT phosphorylation of the transcription factor Twist-1 at Ser42 inhibits p53 activity in response to DNA damage. *Oncogene*, *29*, 3554–3565.
- Voronova, A., & Baltimore, D. (1990). Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 4722–4726.
- Wang, X., Ling, M. T., Guan, X. Y., Tsao, S. W., Cheung, H. W., Lee, D. T., & Wong, Y. C. (2004). Identification of a novel function of TWIST, a bHLH protein, in the development of acquired taxol resistance in human cancer cells. *Oncogene*, *23*, 474–482.
- Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., ... Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*, *117*, 927–939.
- Yang, J., Mani, S. A., & Weinberg, R. A. (2006). Exploring a new twist on tumor metastasis. *Cancer Research*, *66*, 4549–4552.
- Zhuo, W. L., Wang, Y., Zhuo, X. L., Zhang, Y. S., & Chen, Z. T. (2008). Short interfering RNA directed against TWIST, a novel zinc finger transcription factor, increases A549 cell sensitivity to cisplatin via MAPK/mitochondrial pathway. *Biochemical and Biophysical Research Communications*, *369*, 1098–1102.

Travail 2

« The heterodimeric TWIST1-E12 complex drives the oncogenic potential of TWIST1 in human mammary epithelial cells »

Jacqueroud L., Bouard C., Richard G., Payen L., Devouassoux-Shisheboran M., Spicer DB., Caramel J., Collin G., Puisieux A., Tissier A., Ansieau S.

Neoplasia, mars 2016

3 Travail 2 : identification du complexe oncogénique de TWIST1 dans un contexte cancéreux mammaire humain

Le but de cette étude est de déterminer *via* l'utilisation de dimère forcé de TWIST1 la contribution de l'homodimère et des hétérodimères de TWIST1 dans la transformation maligne et ainsi de déterminer la forme oncogénique de TWIST1 pour pouvoir développer une nouvelle stratégie thérapeutique ciblée.

La construction du dimère TWIST1_{R154P}-E12 confirme que les interactions entre les deux partenaires d'un dimère forcé sont nécessaires pour générer un dimère forcé actif. Ce mutant n'impacte pas la localisation nucléaire de TWIST1 mais cette mutation diminue la dimérisation de TWIST1 mutée avec E12 et donc diminue l'activation des gènes cibles des dimères de TWIST1. La construction de l'homodimère TWIST1_{K145E}-TWIST1_{K145E} permet de générer une forme inactive de l'homodimère puisque K145 joue un rôle clef dans la stabilisation des complexes de TWIST1 à l'ADN (Bouard et al., 2014; El Ghouzzi et al., 2001). Le rôle de TWIST1 dans l'atténuation de la sénescence induite par un oncogène est lié à la forme hétérodimérique de TWIST1/E12 et non pas à TWIST1/TWIST1. Les formes mutantes sont également dépourvues de cette fonction. Ainsi, la dimérisation de TWIST1 et de E12 est nécessaire à la fonction de contournement de la sénescence induite par un oncogène.

Nous avons alors cherché à voir si la présence de l'homodimère TWIST1-TWIST1 ou de l'hétérodimère TWIST1-E12 forcées étaient suffisants pour favoriser la carcinogenèse du sein *in vivo*. Les résultats montrent que les souris produisant l'hétérodimère forcé TWIST1-E12 développent des cancers du sein avec une fréquence de 37,9% avec n=29 et une latence plus courte alors que les souris produisant l'homodimère TWIST1-TWIST1 développent des cancers du sein avec une fréquence de 18,1% avec n=14 contre 6,1% et n=16 chez les souris contrôles. Ainsi, l'expression du dimère forcé de TWIST1/E12 induit le développement spontané de cancer du sein chez la souris. Les complexes hétérodimériques TWIST1-E12 sont les formes oncogéniques de TWIST1 dans les cellules épithéliales mammaires. Le complexe hétérodimérique TWIST1-E12 est détecté dans des lésions pré-malignes du sein chez l'homme.

Ensuite, nous avons inhibé l'expression de TWIST1 endogène *via* l'utilisation d'un shRNA ciblant les séquences non traduites à l'extrémité 3' de l'ARN endogène de TWIST1, séquences absentes de l'ADNc codant pour les hétérodimères, afin de réaliser des essais de complémentation. Dans trois lignées cellulaires mammaires, l'utilisation de ce shRNA conduit à un arrêt de prolifération et au déclenchement du programme de sénescence. Après cela, nous avons infecté ces différentes lignées avec les dimères de TWIST1 homodimérique et hétérodimérique. Nous avons pu observer que le complexe hétérodimérique TWIST1-E12, à la différence du complexe homodimérique TWIST1-

TWIST1, était capable de compenser la perte de la protéine TWIST1 endogène dans le maintien de la prolifération des cellules cancéreuses épithéliales mammaires.

Cette étude a permis de confirmer les fonctions distinctes des différents dimères de TWIST1 d'une part et de montrer que TWIST1-E12 était la forme oncogénique dans le contexte cancéreux mammaire d'autre part. Ainsi en abaissant l'activité des voies onco-suppressives et le mécanisme de contrôle du cycle cellulaire, la protéine TWIST1, sous sa forme oncogénique TWIST1-E12, favorise la transformation néoplasique des cellules et protège également les cellules épithéliales mammaires de la sénescence induite par un oncogène et / ou de l'apoptose latente. Ainsi TWIST1-E12 pourrait être une cible thérapeutique intéressante dans le contexte tumoral. Ainsi TWIST1-E12 est la forme active de TWIST1 qu'il faut caractériser *in silico* afin de mieux comprendre les mécanismes mis en jeu par cette protéine dans le but de trouver un inhibiteur spécifique de cette TWIST1-E12.

Pour ce projet, j'ai contribué à la sélection des différents dimères inactifs de TWIST1. En effet j'ai participé au choix de l'hétérodimère TWIST1R154P-E12 comme étant un hétérodimère inactif de TWIST1-E12 puisque la présence de la proline conduit à un changement conformationnel de l'hélice du domaine bHLH de TWIST1 ce qui impacte directement la stabilité de ce domaine *in silico* et donc ces capacités de dimérisation avec d'autres protéines bHLH dans le contexte cellulaire. Au-delà de l'impact conformationnel cela impacte l'activité transcriptionnelle de ces facteurs puisque leur dimérisation est un prérequis à leur liaison à l'ADN et donc à leur fonction. De plus la caractérisation des résidus clefs impliqués dans la fixation à l'ADN des dimères de TWIST1 de l'article 1, a guidé le choix de l'homodimère mutant inactif TWIST1K145E-TWIST1K145E puisque le résidu de lysine 145 est capital dans la conformation du cœur des dimères de bHLH ainsi que dans la reconnaissance des motifs E-box de l'ADN. D'autre part, j'ai également contribué à l'interprétation et à l'analyse des résultats expérimentaux obtenus.

The heterodimeric TWIST1-E12 complex drives the oncogenic potential of TWIST1 in human mammary epithelial cells.

Laurent Jacqueroud^{1,2,3,4,5}, Charlotte Bouard^{1,2,3,4,5}, Geoffrey Richard^{1,2,3,4,5}, Léa Payen^{1,2,3,4,5,6,7}, Mojgan Devouassoux-Shisheboran^{1,2,3,4,5,6,7,8}, Douglas B. Spicer⁹, Julie Caramel^{1,2,3,4,5,6}, Guillaume Collin^{1,2,3,4,5}, Alain Puisieux^{1,2,3,4,5,6,10}, Agnès Tissier^{1,2,3,4,5}, Stéphane Ansieau^{1,2,3,4,5,*}.

Inserm UMR-S1052, Centre de Recherche en Cancérologie de Lyon, France; 2. CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, France; 3. LabEX DEVweCAN, Lyon, France; 4. UNIV UMR1052, Lyon, France; 5. Centre Léon Bérard, Lyon, France; 6. Université de Lyon 1, ISPB, Lyon, France; 7. Hospices Civils de Lyon, Lyon, France; 8. Hôpital de la Croix-Rousse, Lyon, France; 9. Center for Molecular Medicine, Main Medical Center Research Institute, Scarborough, Maine, USA; 10. Institut Universitaire de France, Paris, France.

*Corresponding author: Stéphane Ansieau, Inserm UMR1052, CNRS UMR 5286, Centre de Recherche en Cancérologie de Lyon, Centre Léon Bérard, 28 rue Laennec, 69008 Lyon. France. Phone: 0033-478785160, Fax: 0033-478782720; e-mail: stephane.ansieau@lyon.unicancer.fr

Abstract 241 words (limited to 250 words)

The TWIST1 embryonic transcription factor displays biphasic functions during the course of carcinogenesis. It facilitates the escape of cells from oncogene-induced fail-safe programs (senescence, apoptosis) and their consequent neoplastic transformation. Additionally, it promotes the epithelial-to-mesenchymal transition and the initiation of the metastatic spread of cancer cells. Interestingly, cancer cells recurrently remain dependent on TWIST1 for their survival and/or proliferation, making TWIST1 their Achilles' heel. TWIST1 has been reported to form either homodimeric or heterodimeric complexes mainly in association with the E bHLH class I proteins. These complexes display distinct, sometimes even antagonistic functions during development and unequal pro-metastatic functions in prostate cancer cells. Using a tethered dimer strategy, we successively assessed the ability of TWIST1 dimers to cooperate with an activated version of RAS in human mammary epithelial cell transformation, to provide mice with the ability to spontaneously develop breast tumors, and lastly to maintain a senescence program at a latent state in several breast cancer cell lines. We demonstrate that the TWIST1-E12 complex, unlike the homodimer, is an oncogenic form of TWIST1 in mammary epithelial cells, and that efficient binding of both partners is a prerequisite for its activity. The detection of the heterodimer in human premalignant lesions by a proximity ligation assay, at a stage preceding the initiation of the metastatic cascade, is coherent with such an oncogenic function. TWIST1-E protein heterodimeric complexes may thus constitute the main active forms of TWIST1 with regards to senescence inhibition over the time course of breast tumorigenesis.

Keywords: TWIST1 dimers, escape from fail-safe programs, tumor initiation, breast.

Introduction

The *TWIST1* gene is aberrantly reactivated in a large set of solid cancer types, including a variety of carcinomas, melanomas, sarcomas and neuroblastomas (Ansieau et al., 2010). As a reminiscence of its embryonic functions, this transcription factor was originally shown to promote the metastatic dissemination of cancer cells through its ability to trigger an epithelial-to-mesenchymal transition (Yang et al., 2004). We and others additionally highlighted its ability to alleviate the induction of fail-safe programs (senescence, apoptosis) in response to oncogenic activations, thereby cooperating with mitogenic oncoproteins in promoting cell neoplastic transformation *in vitro* and tumor initiation *in vivo* (Maestro et al., 1999; Valsesia-Wittmann et al., 2004; Kwok et al., 2007; Ansieau et al., 2008; Tran et al., 2012; Morel et al., 2012; Burns et al., 2013). Furthermore, cancer cells often remained dependent on *TWIST1* functions to protect them from latent senescence or apoptosis (Valsesia-Wittmann et al., 2004; Ansieau et al., 2008; Tran et al., 2012; Piccinin et al., 2012; Burns et al., 2013; Beck et al., 2015). Its oncogenic properties are mostly related to its ability to down-modulate the activity of p53 (Maestro et al., 1999; Valsesia-Wittmann et al., 2004; Stasinopoulos et al., 2005; Piccinin et al., 2012) and to dampen the expression of numerous cyclin-kinase inhibitors (p15^{INK4B}, p16^{INK4A}, p21^{CIP1}), the relative contribution of these activities varying according to the cellular context (Ansieau et al., 2008; Burns et al., 2013).

Basic-helix-loop-helix (bHLH) transcription factors were originally classified into different categories based on their expression pattern, partners and structural features (Massari and Murre, 2000). The *TWIST1* protein belongs to Class II, which encompasses tissue-specific bHLH, such as MyoD. These proteins heterodimerize with the ubiquitously expressed Class I bHLH proteins, termed E proteins (Massari and Murre, 2000). They include the two splice variants of the TCF3/E2A protein E12 and E47, the TCF4/ITF2, and the TCF12/HEB transcription factors. All *TWIST1*-E heterodimers have been shown to display similar functional properties (Connerney et al., 2006). The *TWIST1* protein also interacts with the Class II bHLH transcription factors HAND1 and HAND2, and additionally displays the property to constitute functional homodimers, as originally demonstrated in *Drosophila* (Castanon et al., 2001). The *TWIST1* protein preferentially heterodimerizes providing a rationale for the similar limb formation defects triggered by the enforced expression of the *TWIST1*-E12 heterodimer or of the *TWIST1* monomer (Villavicencio et al., 2002; Firulli et al., 2005; Laursen et al., 2007). Nonetheless, homodimerization is either privileged at low protein concentration (e.g. *TWIST1* haploinsufficiency in

Saethre-Chotzen patients), in the presence of ID HLH proteins which titrate E proteins, or following the phosphorylation of residues located in the helix I of *TWIST1* (i.e. Thr121 and Ser123) by protein kinase A (Connerney et al., 2006; Firulli et al., 2007; Connerney et al., 2008). Point mutations preventing the post-translational modification of these residues (TS121-123AA) or mimicking their constitutive phosphorylation (TS121-123ED) were demonstrated to functionally mimic the *TWIST1* homodimer and heterodimer, respectively (Firulli et al., 2005). The phospho-regulation of *TWIST1* was additionally shown to influence its affinity for E-boxes in a cis-element dependent manner (Firulli et al., 2007). Post-translational modifications of *TWIST1* thus have an impact on dimer choice and on the downstream activation of targeted genes (Firulli and Conway, 2008). The contribution of these complexes to the embryonic *TWIST1* functions has been largely explored using tethered dimers, the reliability of this approach being unquestionably established through successful *in vivo* complementation assays (Castanon et al., 2001; Firulli et al., 2005; Connerney et al., 2008). The validity of this strategy was also supported by the demonstration that the tethered *TWIST1*~E12 dimer binds DNA and trans-activates reporter genes similarly to when *TWIST1* and E12 are expressed as separated polypeptides (Castanon et al., 2001; Firulli et al., 2007). Exploiting this tethered dimer strategy unveiled differential and even antagonistic properties of the *TWIST1* complexes during embryonic development, demonstrating the key role of the *TWIST1* partner in determining and regulating *TWIST1* functions. Whether the balance between these dimers also modulates *TWIST1* functions during tumorigenesis remains relatively unexplored. Enforced expression of tethered *TWIST1*~*TWIST1* or *TWIST1*~E12 dimers in prostate cancer cell lines led to the conclusion that the pro-metastatic properties of *TWIST1* are allotted to the heterodimeric *TWIST1* complex (Gajula et al., 2015). Their contribution to the malignant transformation nonetheless still remains to be determined. Moreover, identifying the *TWIST1* complex implicated in the escape from fail-safe program and upon which cancer cells are dependent for their proliferation and survival, constitutes an essential step in the development of novel therapeutic strategies aiming at eradicating cancer cells through *TWIST1* inactivation. To tackle this question, we employed this tethered dimer strategy to assess the oncogenic potentials of both *TWIST1* dimers *in vitro* and *in vivo*.

Material and Methods

DNA constructs

The EcoR pBabe Zeo, ER-HRAS^{G12V} pLNCX2-Neo, and HRAS^{G12V} pbabe-Puro retroviral construct and the *VIMI*-luciferase

reporter have been previously described (Le et al., 2005; Boehm et al., 2005; Barradas et al., 2009; Mikaelian et al., 2013). The tethered dimers were generated by PCR using FLAG-human TWIST1 (Ansieau et al., 2008) and E12 expression constructs (Murre et al., 1991) as templates. The strategy employed to generate the tethered dimers is described in details in the Supplementary Information section. The shRNA *TWIST1* A pLKO.1 lentiviral construct was previously described (Ansieau et al., 2008). The shRNA *TWIST1* B pSIREN (5'-CTCTGGAGCTGATAACTAAA-3') retroviral construct was kindly provided by Patrice Lassus (IGMM, Montpellier). The shRNA *TCF3* pLKO.1 was supplied by Sigma.

Mouse strains

Animal maintenance and experiments were performed in a specific pathogen free animal facility "Anican" in accordance with the animal care guidelines of the European Union and French laws and were validated by the local Animal Ethic Committee (CECAAP). The CAG-LSL-*Twist1*-*Twist1* or CAG-LSL-*Twist1*-*E12* (FVB/NJ genetic background) (Connerney et al., 2008) were crossed with the WAP-Cre (Whey Acidic Protein promoter, Wagner et al., 1997) mouse strain (B6.Cg-Tg(Wap-Cre)11738 Mam, obtained from the mouse models of Human Cancers Consortium and back-crossed to the FVB/NJ genetic background). Cohorts of multiparous female WAP-Cre; *Twist1*-*Twist1* and WAP-Cre; *Twist1*-*E12* mice (with two rounds of lactation) were generated. Genotyping was performed on genomic DNA from tails using the 5'-AATGAAATGGAGAGCTTGGGCGAC-3' and 5'-CATCACTCGTTGCATCGACC-3' primer pair for the WAP-Cre transgene and 5'-GCAAGCGCGCAAGAAATCTG-3' and 5'-CCCGTTCAAGTCTCTTCAGAAATGAC-3' primer pair for the *Twist1* transgene as described previously (Wagner et al., 1997; Connerney et al., 2008). Mice were monitored twice a week for tumor incidence. End points were based on tumor diameter (up to 1.5 cm) unless natural death occurred. Mice were euthanized by CO₂ inhalation. Tumors were fixed in 10% phosphate-buffered formalin for 24 h, and embedded in paraffin. Mammary gland, lung, liver, kidney and spleen were collected either to check for malignant lesions or to assess potential metastatic dissemination.

Cell culture

HMEC-hTERT cells and derived cells were cultured as described previously (Morel et al., 2012). The Hs578T, MDA-MB-436, MDA-MB-231 and BT-474 breast cancer cell lines and HEK293T were provided by the ATCC, and cultured according to the supplier's instructions.

Protein analysis and immunohistochemistry

Cells extracts were performed in RIPA buffer (100 mM NaCl, 1% NP40, 0.1% SDS, 50 mM Tris pH8) supplemented with protease (Roche) and phosphatase inhibitor cocktails (Sigma) and clarified by centrifugation. Proteins were analyzed by Western-blot using the monoclonal anti-TWIST1 2C1a (Abcam), anti-p21^{CIP1} clone SX118 (Dako), and the polyclonal anti-TWIST2 (proteintech), anti-E2A V18, anti-HEB A20, anti-ITF2 K12, anti-H-RAS C20, anti-p15^{INK4B} C20, anti-p16^{INK4A} H156, p-histone H3 (Ser10)-R (Santa-Cruz Biotechnology) and anti-GAPDH Abs16 (Millipore) antibodies and horseradish-peroxidase-conjugated secondary antibodies (Dako). Antigen-antibody complexes were revealed with the Western-blotting Luminol reagent (Santa Cruz Biotechnology).

Analysis of TWIST1 by immunohistochemistry in mouse breast carcinomas was performed as described previously (Morel et al., 2012) using the monoclonal anti-TWIST1 2C1a (Abcam).

Gene expression analysis

RNA preparation and reverse transcription were performed as described previously (Gras et al., 2014). Real time PCR intron-spanning primers were designed with the primer3 software. The *HPRT1* housekeeping gene was used for normalization. The combinations of primers used are listed in the Supplementary Information section.

Retroviral infections, SA-β-galactosidase, growth curves, and soft-agar colony assays

Enforced expression of H-RAS^{G12V} and/or TWIST1 proteins was performed through retroviral infections as described previously (Gras et al., 2014). Briefly cells were "murinized" by expressing the ecotropic receptor (Baker et al., 1992; Boehm et al., 2005) before being infected with retroviral expression constructs 48 h later. Selection was initiated 24 h post-second infection with puromycin (0.5 μg/ml) or neomycin (100 μg/ml).

shRNA *TWIST1* lentiviral particles were generated through the co-transfection of 293T cells with pLKO.1, pCMV ΔR8.91 (gag-pop-Tat-Rev) (Zufferey et al., 1997) and pHCMVG-VSVG (env) (Burns et al., 1993) expression constructs and shRNA *TWIST1* retroviral particles through the transfection of HEK293GP cells with pSIREN and pHCMVG-VSVG (env) expression constructs using the calcium phosphate precipitation technique. *TWIST1* knock-down was achieved *via* double retroviral (shRNA *TWIST1* B pSIREN) or lentiviral (shRNA *TWIST1* A pLKO.1) infections at 48 h intervals. Over 90% of cells were infected. Consequences of TWIST depletion on cell survival

and proliferation was assessed in the absence of selection. Activation of ER-RAS^{G12V} with 4-OH-Tamoxifen (4-OHT, 625 nM, Sigma Aldrich) was specifically performed in medium supplemented with 1 mM D-Glucose. Three days later, cells were washed twice with PBS, either fixed in 3% formaldehyde and stained with crystal violet or tested with a SA- β -galactosidase assay (Dimri et al., 1995). Transformation assays were performed one week post-infection (*TWIST1* transduction) or post-second infection (*RAS* + *TWIST* transduction), as described previously (Gras et al., 2014).

Immunofluorescence

10⁴ cells were seeded on a 12-well IBIDI-chamber slide, fixed in 4% formaldehyde (Sigma) for 15 min, washed with 1% bovine serum albumin (BSA) in PBS buffer and permeabilized in 0.1% Triton 100X (Sigma), 1% BSA in PBS buffer at room temperature for 10 min. The cells were then washed 3 times with 1% BSA in PBS buffer and incubated with the anti-TWIST1 2C1a primary antibody (Abcam, diluted 1/100 in 0.5% BSA in PBS buffer) overnight at 4°C. Following extensive washing in 1% BSA in PBS buffer, cells were incubated with FITC-conjugated Alexa goat anti-mouse antibody (Life Technology, 1/600 in 0.5% BSA in PBS buffer) for 30 min at room temperature. Nuclei were stained with Hoechst 5 mg/ml in 0.5% BSA in PBS buffer for 10 min and mounted with Fluoromount-G (SouthernBiotech). Images were taken on an immunofluorescence microscope (Leica) at identical exposure times.

Immuno-precipitation

HEK293T cells were transfected with wild-type or mutant FLAG-tagged TWIST1 and/or Myc-tagged E12 expression constructs using the calcium-phosphate technique. Cells were lysed 36 h post-transfection in a 50 mM Tris-HCL pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton buffer supplemented with protease inhibitors (Roche), cleared by centrifugation and incubated with an anti-FLAG M2 resin (Sigma-Aldrich). After intensive washing to eliminate non-specific binding, the resin was resuspended in Laemmli buffer and boiled for 3 min. After elimination of the beads by centrifugation, samples were reduced by adding β -mercaptoethanol and separated on SDS-PAGE. Proteins were analyzed with the murine monoclonal anti-TWIST1 C2a antibody (Abcam) and the rabbit polyclonal V18 E2A antibody (Santa Cruz Biotechnology).

Immunoprecipitation of endogenous E12 protein by ectopically expressed FLAG-TWIST1 was performed in HMEC-hTERT cells. 2x10⁶ cells were resuspended in an extraction buffer (20 mM Tris pH8, 125 mM NaCl, 1mM EDTA, 0.5% NP40) supplemented with protease inhibitors (Roche), sonicated and cleared by centrifugation. Protein

extracts were diluted in a final 20 mM Tris pH8 buffer containing 125 mM NaCl, 1mM EDTA, NP40 0.125%, Glycerol 10%, and successively incubated in the presence of 4 μ g of M2 α -FLAG antibody (Sigma) for 4 h at 4°C and protein-A sepharose beads (protein A sepharose[®]4B, InVitrogen) for 1 h. After intensive washing in a 20 mM Tris pH8 buffer containing 125 mM NaCl, 1mM EDTA, NP40 0.1%, Glycerol 10% buffer, beads were resuspended in Laemmli buffer (in absence of β -mercaptoethanol), boiled at 95°C for 3 min. After elimination of the beads by centrifugation, samples were reduced by adding β -mercaptoethanol and separated on SDS-PAGE.

Proximity Ligation Assay

Human ductal carcinoma *in situ* (DCIS) samples were obtained through the Biological Resource Center of the Centre Léon Bérard with the agreement of the ethical review board of the Centre Léon Bérard. Samples were used with the patient's written informed consent. The present study was approved by the ethical review board of the Centre Léon Bérard. DCIS samples were selected on the basis of TWIST1 protein detection, as previously reported (Morel et al., 2012).

After deparaffinization and rehydration, tissue sections were boiled in a 10 mM pH6 citrate buffer for 40 min. After saturating unspecific binding sites for 20 min with a 1% BSA PBS buffer, sections were incubated overnight with the mouse monoclonal anti-TWIST1 Twist2C1a (Abcam) and the rabbit polyclonal anti-E2A N-649 (Santa Cruz) antibodies in a 0.5% BSA in PBS buffer. Incubation with PLA probes, ligation and amplification steps were performed according to the manufacturer instructions (Duolink[®] In Situ kit, Sigma). The specificity of the signal was confirmed using paraffin-embedded MDA-MB-436, Hs578T (TWIST1-positive, E12-positive) and HMEC-hTERT (TWIST1-negative, E12-positive) mammary cells, in which *TWIST1* or *TCF3* (the two E2A protein variants E12 and E47 encoding gene) were specifically knocked-down through RNA interference. Number of dots per cell was assessed using the ImageJ software.

Results

Design of wild-type and mutant tethered TWIST1 dimers.

Since both *HAND1* and *HAND2* genes were found to be transcriptionally inactive in breast cancer cells and since E proteins, are expressed as anticipated in all of the studied breast cancer cell lines (Figure S1A), the experiments were performed either with homodimeric (TWIST1-TWIST1, hereafter named T~T) or heterodimeric (TWIST1-E2A E12, hereafter named T~E) tethered dimers. Following previous studies (Castanon et al.,

2001; Firulli et al., 2007; Connerney et al., 2008; Gajula et al., 2015), the E2A splice variant E12 was employed as the model for E proteins, since all TWIST1-E protein complexes were reported to display similar properties (Connerney et al., 2006). Tethered dimers were generated by linking the two partners with a flexible glycine-serine polylinker, also used in previous studies (Castanon et al., 2001; Firulli et al., 2007; Connerney et al., 2008; Gajula et al., 2015), and the TWIST1 protein was tagged at the N-terminus with a single FLAG peptide, a position that we previously confirmed does not interfere with TWIST1 oncogenic functions (Ansieau et al., 2008) (Figure S1B). The production and correct nuclear localization of the resulting tethered dimers were validated (Figure S1C). Noticeably, C-terminal truncated sub-products (revealed with both anti-TWIST1 and anti-FLAG antibodies) were detected. Very similar sub-products were revealed with the functional murine MYC-tagged T~T and MYC-tagged T~E forced dimers generated in the laboratory of D. Spicer (Connerney et al., 2006) (Figure S1D). In particular, the detection with the T~T dimer of a predominant sub-product with a molecular size close to the monomer, likely reflects a difference in the accessibility of a consensus caspase 3/7 cleavage site present in the C-terminal of TWIST1 within the two tethered dimers. Cleavage of TWIST1 by caspase 3 was originally reported in apoptotic cells, and was quickly followed by its proteasome-mediated degradation (Demontis et al., 2006). It is, therefore, very likely that the cleavage occurs during protein extraction.

As the study progressed, additional mutants were generated including the T R154P~E heterodimer (hereafter named T RP~E) to confirm the need of the interaction between both partners to generate an active tethered heterodimer. The R154P mutation was previously reported to disrupt the TWIST1-E2A protein interaction (Spicer et al., 1996), and this was herein confirmed by performing a co-immunoprecipitation assay (Figure S2A). Insertion of this mutation into the T~E tethered dimer (T RP~E) did not have an impact on its nuclear sub-localization (Figure S2B) but, as anticipated, abrogated the ability of the fusion protein to activate the TWIST1-targeted vimentin (*VIM*) gene, in a reporter assay (Figure S2 C). A TWIST1 K145E-TWIST1 K145E (T KE~T KE) tethered dimer was also generated as an inactive version of the T~T tethered dimer. The Lys145 residue plays a determining role in stabilizing TWIST1 complexes on the DNA by contributing to the establishment of the inter-helical loops and by directly interacting with oxygen atoms of DNA bases (Maia et al., 2012; Bouard et al., 2014). The K145E mutation thus affects the global structure of TWIST1 complexes and abrogates TWIST1 DNA binding properties (El Ghouzzi et al., 2001).

The TWIST1 dimers unequally alleviate oncogene-induced senescence in human mammary epithelial cells.

To compare the ability of both TWIST1 dimers in preventing oncogene-induced senescence (OIS), immortalized and non-transformed human mammary epithelial cells (HMEC-hTERT) were sequentially infected with retroviral vectors encoding either for an inducible activated version of HRAS (ER-HRAS^{G12V}) or for TWIST1 (monomer T, T~T or T~E dimer) (Figure 1A). Activation of the ER-HRAS^{G12V} fusion protein by 4-hydroxyl-tamoxifen (+ 4-OHT) triggered a senescence program, as revealed by the permanent proliferation arrest and the detection of a SA- β -galactosidase activity (Figure 1B). While the *CDKN2A-ARF* locus is methylated in HMEC-hTERT cells (Huschtscha et al., 1998; Brenner et al., 1998; Kiyono et al., 1998), commitment into senescence was associated with an accumulation of p15^{INK4B} (p15) and p21^{CIP1/WAF1} (p21) (Figure 1B). In line with our previous observations (Ansieau et al., 2008), enforced expression of the TWIST1 monomer successfully prevented the activation of cyclin-kinase inhibitors and sustained cell proliferation, as assessed in a crystal violet coloration assay (Figure 1C-D). The heterodimer turned out to be functionally active in this assay and the homodimer inactive (Figure 1C), suggesting that the dimerization between TWIST1 and E protein partners is a prerequisite to carry out this function. To support this hypothesis, we observed that the disruption of the partner interaction through the insertion of the R154P point mutation in TWIST1 abolished T~E activity (Figure 1C). Collectively, this assay demonstrated that both monomeric and tethered heterodimers similarly prevent OIS, supporting the concept that ectopically expressed TWIST1 can form heterodimers with endogenous E proteins (Villavicencio et al., 2002; Laursen et al., 2007). Indeed, ectopic TWIST1 was confirmed to immuno-precipitate the endogenous E12 protein (Figure S3). Of note, in light of the absence of activity of the TWIST1 homodimeric complex, the functionality of the fusion protein was further controlled by assessing its ability to transactivate a set of *TWIST1* targeted genes, selected on the basis of previous studies (Taube et al., 2010; Morel et al., 2012). As shown in Figure S4, the homodimer was confirmed to be functional. As an internal control, the insertion of the K145E mutation (T KE~T KE tethered dimer), described to abolish the TWIST1 DNA binding capability (El Ghouzzi et al., 2001), annihilated its transactivation potential.

To strengthen the differences observed between the two TWIST1 complexes, we next assessed their ability to cooperate with the HRAS^{G12V} protein in promoting the neoplastic

transformation of HMEC-hTERT cells. To this end, cells were sequentially infected with *TWIST1* and constitutive *H-RAS*^{G12V} retroviral expression constructs. While the sole expression of the *TWIST1* tethered dimers failed to transform cells, concomitant production of H-RAS^{G12V} and the active heterodimeric complex successfully transformed cells, as assessed in a colony formation assay (Figure 2A-B). Neither the homodimeric T~T complex nor the inactive form of the T~E complex (T RP ~E) were found to be functional in this second assay.

The enforced expression of the T~E tethered dimer induces the spontaneous development of breast carcinoma in mice.

Next, we validated the different oncogenic activities observed between the two *TWIST1* complexes *in vivo*. We previously reported that the combined production of either the murine monomeric *TWIST1* or the T~E fusion protein with an activated version of K-RAS (K-RAS^{G12D}) in luminal committed epithelial cells induced breast tumor development in the first five months with a complete penetrance (Morel et al., 2012). However, not a single one of the three mice examined, co-expressing the homodimer T~T and K-RAS^{G12D}, developed breast tumors within the same period of time, suggesting that the homodimer failed to cooperate with RAS in promoting tumor initiation (G. W. Hinkal, unpublished data). To strengthen this result, we examined whether the ectopic expression of T~T or T~E in the absence of an oncogenic insult was sufficient to promote breast carcinogenesis, with an anticipated lower penetrance and extended latency to accumulate requested secondary events. T~T or T~E production was enforced in luminal committed cells through the generation of WAP: *Twist1~Twist1* and WAP: *Twist1~E12* mice (producing the murine T~T or T~E fusion proteins under the control of the whey acidic protein promoter) by crossing either the CAG-LSL-*Twist1~Twist1* or and CAG-LSL-*Twist1~E12* and the WAP-*Cre* mouse strains (Wagner et al., 1997; Connerney et al., 2008). Full activation of the WAP promoter was assured by two consecutive lactations. As shown in Figure 2 C-E, T~E-producing mice were found to develop breast carcinomas with a higher frequency (37.9%, n=29) and a shorter latency than T~T-producing mice (18.1%, n=14) and control littermates (6.2%, n=16). Tumor phenotypes were found to be heterogeneous and classified (Cardiff, 2001), as (i) adenosquamous, (ii) metaplastic spindle cell, (iii) papillary and cribriform, (iv) solid and cribriform, and (v) adenosquamous carcinomas (Table S1). No secondary tumor site was detected. Collectively, these observations support our conclusion that the heterodimeric complexes are the oncogenic forms of *TWIST1* in mammary epithelial cells.

The heterodimeric *TWIST1*-E12 complex is detected in human breast premalignant lesions.

If the assumption that the heterodimeric *TWIST1* complex plays a role in malignant transformation, breast cancer cell survival and proliferation is correct, we would expect that this complex would already be detectable in *in situ* ductal carcinoma (DCIS), before the initiation of the metastatic cascade. We thus assessed TE complex formation through a proximity ligation assay (PLA), selecting cases previously defined as *TWIST1*-positive or -negative by immunohistochemistry (Morel et al., 2012). Experimental conditions were established in mammary carcinoma cell lines and the specificity of the signal was confirmed by turning-down the expression of either the endogenous *TWIST1* or *TCF3* (the E12/E47 proteins-encoding gene) through RNA interference. Depletion in the *TWIST1* and E2A proteins significantly reduced the number of dots detectable in the *TWIST1*-positive MDA-MB436 (8.8 ± 1.1 dots/cell to 4.9 ± 1.0 and 1.6 ± 1.5 dots/cell, respectively) and Hs578T (9.8 ± 3.1 dots/cell to 3.1 ± 1.9 and 2.2 ± 1.7 dots/cell, respectively), while the number of dots/cell remained <1 in *TWIST1*-negative HMEC-hTERT cells (Figure S5). As shown in Figure 3, the TE complex was invariably detected in all *TWIST1*-positive DCIS, while no signal was detected in the control samples, namely *TWIST1*-negative DCIS and normal mammary glands. The detection of the TE complex in premalignant lesions is thus compatible its role in the inhibition of a fail-safe program.

The *TWIST1*-E12 heterodimeric complex, unlike the homodimeric *TWIST1* complex, complements the loss of the endogenous *TWIST1* protein in sustaining mammary epithelial cancer cell proliferation.

By lowering the activity of oncosuppressive pathways and of the cell cycle control machinery, the *TWIST1* protein not only favors the cell neoplastic transformation, but also protects cancer cells from latent OIS and/or apoptosis (Valsesia-Wittmann et al., 2004; Kwok et al., 2007; Ansieau et al., 2008; Tran et al., 2012; Burns et al., 2013; Beck et al., 2015). To assess the contribution of *TWIST1* dimers in these secondary activities, we selected a shRNA targeting the 3'-untranslated sequences of the endogenous *TWIST1* RNA, which are absent from the heterodimer cDNAs, in order to inhibit endogenous *TWIST1* expression and performed complementation assays with the tethered dimers. We confirmed that this shRNA (shRNA *TWIST1* B), similarly to the shRNA used in our previous studies (shRNA *TWIST1A*, (Ansieau et al., 2008)) induced a proliferation arrest

(as assessed by a crystal violet coloration assay and the amount of phospho-Ser10 histone H3 (H3-S10)) and triggered a senescence program, as revealed by the detection of a SA- β -galactosidase activity (Figure 4), in all three breast cell lines tested (MDA-MB-436, Hs578T and BT-474). The permanent growth-arrest was invariably associated with an accumulation of p15^{INK4B} (p15) and p21^{CIP1} (p21) cyclin-dependent kinase inhibitors, p16^{INK4A} (p16) remaining undetectable (Figure 4). No trace of cleaved caspases-3 or PARP fragments were detected, excluding a concomitant apoptosis induction (data not shown). To evaluate the contribution of both the homodimeric and heterodimeric TWIST1 complexes, MDA-MB-436 cells were next sequentially infected with TWIST1-expressing constructs (using a monomeric TWIST1 cDNA lacking 3' untranslated sequence as an internal positive control) and depleted in endogenous TWIST1 through RNA interference. As shown in Figure 5, similarly to the ectopically expressed TWIST1 monomer, the heterodimeric T~E complex avoided cyclin-kinase inhibitor accumulation and sustained cell proliferation. As anticipated, the inactive T~E variant (T RP ~E) failed to do so. The T~T homodimer also failed to complement the endogenous TWIST1 functions. Similar results were obtained in the two additional Hs578T and BT-474 cell lines (Figure S6). These data unambiguously demonstrated that the T~E heterodimeric complex is necessary to prevent OIS and to maintain the senescence program in a latent state in mammary epithelial cells.

Discussion

The embryonic transcription factor TWIST1 displays pleiotropic functions during carcinogenesis, promoting the neoplastic transformation of cells and tumor initiation, on the one hand, and the metastatic spread of cancer cells through EMT induction, on the other hand (Ansieau et al., 2010). Determining whether these functions rely on the activities on various TWIST1 complexes or whether the oncogenic and pro-metastatic properties are conferred upon a single complex is essential in order to further understand the TWIST1 functions, as well as to design strategies to inactivate the transcription factor. By using a tethered dimer approach, previously proven to delineate the TWIST1 complex functions *in vitro*, as well as *in vivo*, we demonstrated that the TWIST1-E12 heterodimer efficiently avoided OIS in mammary epithelial cells. Furthermore, its enforced expression complements the activity of the endogenous TWIST1 protein in keeping cyclin-kinase inhibitors under a threshold compatible with breast cancer cell proliferation. In comparison, the homodimeric complex failed to do so, highlighting the pivotal role of the TWIST1 partner in regulating its oncogenic functions. To support this conclusion,

the monomeric TWIST1 protein and the heterodimeric TWIST1 complex were found to efficiently cooperate with RAS in promoting breast carcinogenesis *in vivo* (Morel et al., 2012). Despite the limited number of mice examined, the absence of pathology observed when the homodimeric TWIST1 complex was combined with RAS, demonstrated that T~T was less efficient than T~E in promoting tumor initiation in this context (G. W. Hinkal, unpublished results). The tendency of mice, ectopically expressing T~E in mammary epithelial cells, to spontaneously develop breast carcinoma with a higher frequency than their T~T counterparts and control littermates, further supports this conclusion.

Several bHLH transcription factors are known to heterodimerize with TWIST1, including the TCF3/E2A splice variants E12 and E47, the TCF4/ITF2, the TCF12/HEB, and the HAND proteins. As the *HAND* genes were found to be transcriptionally inactive in the examined breast cancer cells (Figure S1), we focused on E proteins, using E12 as a prototype. Considering the degree of homology in their bHLH domains (81.8% identity with TCF3/E2A E47 splice variant, 89.1% identity with TCF4/E2-2, and 92.7% identity with TCF12/HEB), it is likely that all of these complexes display similar oncogenic properties. Furthermore E2A proteins are the predominant E proteins expressed in the breast cancer cell lines employed in the present study, as assessed by Western-blot (Figure S1A). Unfortunately, we cannot exclude that additional TWIST1 bHLH protein partners also contribute to its oncogenic activity. For example, the TWIST1-related TWIST2 protein is expressed in basal-B breast cancer cells (Figure S1A). Although co-expression of both transcription factors at the cellular level remains to be demonstrated, the two proteins are able to heterodimerize (Fu et al., 2011). It is worth noting that the heterodimeric TWIST1-E12 complex was additionally shown to drive the pro-metastatic activity of TWIST1 in prostate cancer cells (Gajula et al., 2015), suggesting that the TWIST1-E complexes are the active forms of TWIST1 with regards to tumorigenesis, promoting both tumor initiation and progression. Using a proximity ligation assay, we confirmed that the heterodimer is detectable in DCIS before the metastatic cascade is initiated (Figures 3 and S6).

Since the ectopic expression of the tethered dimer T~T did not affect breast cancer cell proliferation, it is possible that the homodimeric complex does not display transdominant negative properties, at least with regards to the functions and cellular models considered in this study. The balance between the two complexes rather spatially and temporally controls the amount of heterodimeric complex, and, thereby, the oncogenic and pro-metastatic properties of TWIST1. To

support this hypothesis, ID HLH proteins which tilt the balance in favor of the TWIST1 homodimer, were shown to suppress the TWIST1-driven invasive properties and to favor secondary site colonization through the induction of a mesenchymal-to-epithelial transition (Stankic et al., 2013; Rahme and Israel, 2015). Interestingly, in both *C. elegans* and mouse development, respectively, the HLH8 (the *C. elegans* TWIST ortholog) and HAND1 homodimeric complexes were found to play a key role and the heterodimers functions were restricted to that of counterbalancing their activity (Philogene et al., 2012; Hu et al., 2013). The relative contribution of bHLH complexes during embryonic development and senescence inhibition over the time course of breast carcinogenesis may thus be inverted. Nonetheless, the ability of the homodimer to modulate downstream targeted genes suggests that it might also display unrelated and yet-uncharacterized pro-tumoral properties.

In numerous tumor models, cancer cells were shown to remain dependent on TWIST1 to sustain proliferation (Valesia-Wittmann et al., 2004; Ansieau et al., 2008; Tran et al., 2012; Piccinin et al., 2012; Burns et al., 2013; Beck et al., 2015). We now extend this observation to EMT-committed basal-B breast cancer cells (such as MDA-MB-436 and Hs578T), which express a large spectrum of EMT-inducers including ZEB1. ZEB1, similarly to TWIST1, has been demonstrated to down-regulate the expression of *INK4B* and *CIP1*, and to prevent EGFR-driven OIS in lung cancer cells (Liu et al., 2008; Ohashi et al., 2010). The observed lack of complementation between the two proteins in these breast cancer cells either suggests cell type specific regulations (e.g. post-translational regulations) or unveils yet-unidentified TWIST1-dependent properties. Among the various possibilities, the embryonic transcription factor has been reported to interact directly with p65^{cRelA} and to alleviate the transactivation of cytokine expressing genes, such as *IL6* (Sosic et al., 2003), known to take part in the senescence-associated secretory phenotype (SASP) and, thereby, to contribute to the maintenance of a senescent phenotype. The key regulators leading to the specificity of the genetic program driven by TWIST1 complexes remain to be defined. Of note, a tandem E-box separated by a 5 nucleotide spacer was recently shown to be specifically recognized by the TWIST1-E2A complexes (similarly with both E2A splice variants), a tetrameric complex formed and stabilized through the interaction between the α -helixes of the WR domain of TWIST1 proteins (Chang et al., 2015). Enforced expression of the T~E heterodimer in HMEC-hTERT cells affected the expression of a limited number of genes (data not shown), suggesting that the tethered dimer strategy avoids titration of bHLH factors. This

experimental approach may lead the way to defining the TWIST1-E complex-downstream targeted genes and to exploring their contribution to the oncogenic activity of TWIST1.

Acknowledgments

We are grateful to the Biological Resource Center of the Centre Léon Bérard for providing the cohort of human breast samples. We are thankful to Pr Cornelis Murre (University of California, San-Diego, USA) for providing the human E12 expression construct, Pr William C. Hahn for providing the EcoR retroviral expressing construct (Harvard University, USA), Dr Patrice Lassus (IGMM, Montpellier, France) for the shRNA *TWIST1* directed against its 3'UTR (shRNA *TWIST1* B), Pr Jesus Gil (MRC, London, UK) for the ER-RAS construct, Dr Ivan Mikaelian (CRCL, Lyon, France) for the *VIMI*-luciferase reporter, and Dr Didier Negre (INSERMU111/CIRI, Ecole Normale Supérieure de Lyon, France) for providing the pCMV Δ R8.91 and pHCMVG-VSVG vectors. Authors are thankful to Dr Isabelle Treilleux, Audrey Pierrot, Sophie Léon, Nicolas Gado, for their technical help and to Dr Louise Hill and Dr Brigitte Manship for critical reading of the manuscript.

Funding

Our team is labeled by the Ligue Nationale contre le Cancer. This work was additionally supported by institutional grants from the LabEX DEVweCAN (ANR-10-LABX-61) and from the LyRIC (Lyon Recherche Intégrée en Cancérologie, Institut National contre le Cancer, INCa_4664). LJ, GR and CB were recipients of fellowships from the Ligue Nationale contre le Cancer and the Foundation ARC pour la Recherche sur le Cancer.

Author contributions

Conceptualization and design: LJ, LP, AT, SA; Development of methodology: LJ, GR, MD, JC, AT, SA; Material support: DBS; Analysis and interpretation of data: LJ, CB, LP, AP, AT, SA; Writing: SA; Revision: GC and SA.

References

- Ansieau,S., Bastid,J., Doreau,A., Morel,A.P., Bouchet,B.P., Thomas,C., Fauvet,F., Puisieux,I., Doglioni,C., Piccinin,S., Maestro,R., Voeltzel,T., Selmi,A., Valesia-Wittmann,S., Caron de,F.C., and Puisieux,A. (2008). Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell* 14, 79-89.
- Ansieau,S., Morel,A.P., Hinkal,G., Bastid,J., and Puisieux,A. (2010). TWISTing an embryonic transcription factor into an oncoprotein. *Oncogene* 29, 3173-3184.

- Baker, B.W., Boettiger, D., Spooncer, E., and Norton, J.D. (1992). Efficient retroviral-mediated gene transfer into human B lymphoblastoid cells expressing mouse ecotropic viral receptor. *Nucleic Acids Res.* *20*, 5234.
- Barradas, M., Anderton, E., Acosta, J.C., Li, S., Banito, A., Rodriguez-Niedenfuhr, M., Maertens, G., Banck, M., Zhou, M.M., Walsh, M.J., Peters, G., and Gil, J. (2009). Histone demethylase JMJD3 contributes to epigenetic control of INK4a/ARF by oncogenic RAS. *Genes Dev.* *23*, 1177-1182.
- Beck, B., Lapouge, G., Rorive, S., Drogat, B., Desaedelaere, K., Delafaille, S., Dubois, C., Salmon, I., Willekens, K., Marine, J.C., and Blanpain, C. (2015). Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. *Cell Stem Cell* *16*, 67-79.
- Boehm, J.S., Hession, M.T., Bulmer, S.E., and Hahn, W.C. (2005). Transformation of human and murine fibroblasts without viral oncoproteins. *Mol. Cell Biol.* *25*, 6464-6474.
- Bouard, C., Terreux, R., Hope, J., Chemelle, J.A., Puisieux, A., Ansieau, S., and Payen, L. (2014). Interhelical loops within the bHLH domain are determinant in maintaining TWIST1-DNA complexes. *J. Biomol. Struct. Dyn.* *32*, 226-241.
- Brenner, A.J., Stampfer, M.R., and Aldaz, C.M. (1998). Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation. *Oncogene* *17*, 199-205.
- Burns, J.C., Friedmann, T., Driever, W., Burrascano, M., and Yee, J.K. (1993). Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc. Natl. Acad. Sci. U. S. A* *90*, 8033-8037.
- Burns, T.F., Dobromilskaya, I., Murphy, S.C., Gajula, R.P., Thiagarajan, S., Chatley, S.N., Aziz, K., Cho, Y.J., Tran, P.T., and Rudin, C.M. (2013). Inhibition of TWIST1 leads to activation of oncogene-induced senescence in oncogene-driven non-small cell lung cancer. *Mol. Cancer Res.* *11*, 329-338.
- Cardiff, R.D. (2001). Validity of mouse mammary tumour models for human breast cancer: comparative pathology. *Microsc. Res. Tech.* *52*, 224-230.
- Castanon, I., Von, S.S., Kass, J., and Baylies, M.K. (2001). Dimerization partners determine the activity of the Twist bHLH protein during *Drosophila* mesoderm development. *Development* *128*, 3145-3159.
- Chang, A.T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J.W., Paz, H., Wang, D., Li, H.R., Fu, X.D., Rauscher, F.J., III, and Yang, J. (2015). An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev.* *29*, 603-616.
- Connerney, J., Andreeva, V., Leshem, Y., Mercado, M.A., Dowell, K., Yang, X., Lindner, V., Friesel, R.E., and Spicer, D.B. (2008). Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure. *Dev. Biol.* *318*, 323-334.
- Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M.A., and Spicer, D.B. (2006). Twist1 dimer selection regulates cranial suture patterning and fusion. *Dev. Dyn.* *235*, 1345-1357.
- Demontis, S., Rigo, C., Piccinin, S., Mizzau, M., Sonogo, M., Fabris, M., Brancolini, C., and Maestro, R. (2006). Twist is substrate for caspase cleavage and proteasome-mediated degradation. *Cell Death. Differ.* *13*, 335-345.
- Dimri, G.P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E.E., Linskens, M., Rubelj, I., Pereira-Smith, O., and . (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc. Natl. Acad. Sci. U. S. A* *92*, 9363-9367.
- El Ghouzzi, V., Legeai-Mallet, L., Benoit-Lasselin, C., Lajeunie, E., Renier, D., Munnich, A., and Bonaventure, J. (2001). Mutations in the basic domain and the loop-helix II junction of TWIST abolish DNA binding in Saethre-Chotzen syndrome. *FEBS Lett.* *492*, 112-118.
- Firulli, A.B. and Conway, S.J. (2008). Phosphoregulation of Twist1 provides a mechanism of cell fate control. *Curr. Med. Chem.* *15*, 2641-2647.
- Firulli, B.A., Krawchuk, D., Centonze, V.E., Vargesson, N., Virshup, D.M., Conway, S.J., Cserjesi, P., Laufer, E., and Firulli, A.B. (2005). Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nat. Genet.* *37*, 373-381.
- Firulli, B.A., Redick, B.A., Conway, S.J., and Firulli, A.B. (2007). Mutations within helix I of Twist1 result in distinct limb defects and variation of DNA binding affinities. *J. Biol. Chem.* *282*, 27536-27546.
- Fu, J., Qin, L., He, T., Qin, J., Hong, J., Wong, J., Liao, L., and Xu, J. (2011). The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Res.* *21*, 275-289.
- Gajula, R.P., Chettiar, S.T., Williams, R.D., Nugent, K., Kato, Y., Wang, H., Malek, R., Taparra, K., Cades, J., Annadanam, A., Yoon, A.R., Fertig, E., Firulli, B.A., Mazzacurati, L., Burns, T.F., Firulli, A.B., An, S.S., and Tran, P.T. (2015). Structure-Function Studies of the bHLH Phosphorylation Domain of TWIST1 in Prostate Cancer Cells. *Neoplasia* *17*, 16-31.

- Gras,B., Jacqueroud,L., Wierinckx,A., Lamblot,C., Fauvet,F., Lachuer,J., Puisieux,A., and Ansieau,S. (2014). Snail family members unequally trigger EMT and thereby differ in their ability to promote the neoplastic transformation of mammary epithelial cells. *PLoS. One.* 9, e92254.
- Hu,D., Scott,I.C., Snider,F., Geary-Joo,C., Zhao,X., Simmons,D.G., and Cross,J.C. (2013). The basic helix-loop-helix transcription factor Hand1 regulates mouse development as a homodimer. *Dev. Biol.* 382, 470-481.
- Huschtscha,L.I., Noble,J.R., Neumann,A.A., Moy,E.L., Barry,P., Melki,J.R., Clark,S.J., and Reddel,R.R. (1998). Loss of p16INK4 expression by methylation is associated with lifespan extension of human mammary epithelial cells. *Cancer Res.* 58, 3508-3512.
- Kiyono,T., Foster,S.A., Koop,J.I., McDougall,J.K., Galloway,D.A., and Klingelutz,A.J. (1998). Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* 396, 84-88.
- Kwok,W.K., Ling,M.T., Yuen,H.F., Wong,Y.C., and Wang,X. (2007). Role of p14ARF in TWIST-mediated senescence in prostate epithelial cells. *Carcinogenesis* 28, 2467-2475.
- Laursen,K.B., Mielke,E., Iannaccone,P., and Fuchtbauer,E.M. (2007). Mechanism of transcriptional activation by the proto-oncogene Twist1. *J. Biol. Chem.* 282, 34623-34633.
- Le,M.S., Fromigue,O., and Marie,P.J. (2005). Sp1/Sp3 and the myeloid zinc finger gene MZF1 regulate the human N-cadherin promoter in osteoblasts. *Exp. Cell Res.* 302, 129-142.
- Liu,Y., El-Naggar,S., Darling,D.S., Higashi,Y., and Dean,D.C. (2008). Zeb1 links epithelial-mesenchymal transition and cellular senescence
1. *Development* 135, 579-588.
- Maestro,R., Dei Tos,A.P., Hamamori,Y., Krasnokutsky,S., Sartorelli,V., Kedes,L., Doglioni,C., Beach,D.H., and Hannon,G.J. (1999). Twist is a potential oncogene that inhibits apoptosis. *Genes Dev.* 13, 2207-2217.
- Maia,A.M., da Silva,J.H., Mencialha,A.L., Caffarena,E.R., and Abdelhay,E. (2012). Computational modeling of the bHLH domain of the transcription factor TWIST1 and R118C, S144R and K145E mutants. *BMC. Bioinformatics.* 13, 184.
- Massari,M.E. and Murre,C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell Biol.* 20, 429-440.
- Mikaelian,I., Malek,M., Gadet,R., Viallet,J., Garcia,A., Girard-Gagnepain,A., Hesling,C., Gillet,G., Gonzalo,P., Rimokh,R., and Billaud,M. (2013). Genetic and pharmacologic inhibition of mTORC1 promotes EMT by a TGF-beta-independent mechanism. *Cancer Res.* 73, 6621-6631.
- Morel,A.P., Hinkal,G.W., Thomas,C., Fauvet,F., Courtois-Cox,S., Wierinckx,A., Devouassoux-Shisheboran,M., Treilleux,I., Tissier,A., Gras,B., Pourchet,J., Puisieux,I., Browne,G.J., Spicer,D.B., Lachuer,J., Ansieau,S., and Puisieux,A. (2012). EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS. Genet.* 8, e1002723.
- Murre,C., Voronova,A., and Baltimore,D. (1991). B-cell- and myocyte-specific E2-box-binding factors contain E12/E47-like subunits. *Mol. Cell Biol.* 11, 1156-1160.
- Ohashi,S., Natsuzaka,M., Wong,G.S., Michaylira,C.Z., Grugan,K.D., Stairs,D.B., Kalabis,J., Vega,M.E., Kalman,R.A., Nakagawa,M., Klein-Szanto,A.J., Herlyn,M., Diehl,J.A., Rustgi,A.K., and Nakagawa,H. (2010). Epidermal growth factor receptor and mutant p53 expand an esophageal cellular subpopulation capable of epithelial-to-mesenchymal transition through ZEB transcription factors. *Cancer Res.* 70, 4174-4184.
- Philogene,M.C., Small,S.G., Wang,P., and Corsi,A.K. (2012). Distinct Caenorhabditis elegans HLH-8/twist-containing dimers function in the mesoderm. *Dev. Dyn.* 241, 481-492.
- Piccinin,S., Tonin,E., Sessa,S., Demontis,S., Rossi,S., Pecciarini,L., Zanatta,L., Pivetta,F., Grizzo,A., Sonogo,M., Rosano,C., Dei Tos,A.P., Doglioni,C., and Maestro,R. (2012). A "twist box" code of p53 inactivation: twist box: p53 interaction promotes p53 degradation. *Cancer Cell* 22, 404-415.
- Rahme,G.J. and Israel,M.A. (2015). Id4 suppresses MMP2-mediated invasion of glioblastoma-derived cells by direct inactivation of Twist1 function. *Oncogene* 34, 53-62.
- Sosic,D., Richardson,J.A., Yu,K., Ornitz,D.M., and Olson,E.N. (2003). Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity
3. *Cell* 112, 169-180.
- Spicer,D.B., Rhee,J., Cheung,W.L., and Lassar,A.B. (1996). Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science* 272, 1476-1480.
- Stankic,M., Pavlovic,S., Chin,Y., Brogi,E., Padua,D., Norton,L., Massague,J., and Benezra,R. (2013). TGF-beta-Id1 signaling opposes Twist1 and promotes metastatic colonization via a mesenchymal-to-epithelial transition. *Cell Rep.* 5, 1228-1242.
- Stasinopoulos,I.A., Mironchik,Y., Raman,A., Wildes,F., Winnard,P., Jr., and Raman,V. (2005). HOXA5-twist interaction alters

p53 homeostasis in breast cancer cells. *J. Biol. Chem.* 280, 2294-2299.

Taube, J.H., Herschkowitz, J.I., Komurov, K., Zhou, A.Y., Gupta, S., Yang, J., Hartwell, K., Onder, T.T., Gupta, P.B., Evans, K.W., Hollier, B.G., Ram, P.T., Lander, E.S., Rosen, J.M., Weinberg, R.A., and Mani, S.A. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. U. S. A* 107, 15449-15454.

Tran, P.T., Shroff, E.H., Burns, T.F., Thiagarajan, S., Das, S.T., Zabuawala, T., Chen, J., Cho, Y.J., Luong, R., Tamayo, P., Salih, T., Aziz, K., Adam, S.J., Vicent, S., Nielsen, C.H., Withofs, N., Sweet-Cordero, A., Gambhir, S.S., Rudin, C.M., and Felsner, D.W. (2012). Twist1 suppresses senescence programs and thereby accelerates and maintains mutant Kras-induced lung tumorigenesis. *PLoS. Genet.* 8, e1002650.

Valsesia-Wittmann, S., Magdeleine, M., Dupasquier, S., Garin, E., Jallas, A.C., Combaret, V., Krause, A., Leissner, P., and Puisieux, A. (2004). Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer Cell* 6, 625-630.

Villavicencio, E.H., Yoon, J.W., Frank, D.J., Fuchtbauer, E.M., Walterhouse, D.O., and Iannaccone, P.M. (2002). Cooperative E-box regulation of human GLI1 by TWIST and USF. *Genesis.* 32, 247-258.

Wagner, K.U., Wall, R.J., St-Onge, L., Gruss, P., Wynshaw-Boris, A., Garrett, L., Li, M., Furth, P.A., and Hennighausen, L. (1997). Cre-mediated gene deletion in the mammary gland. *Nucleic Acids Res.* 25, 4323-4330.

Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 117, 927-939.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat. Biotechnol.* 15, 871-875.

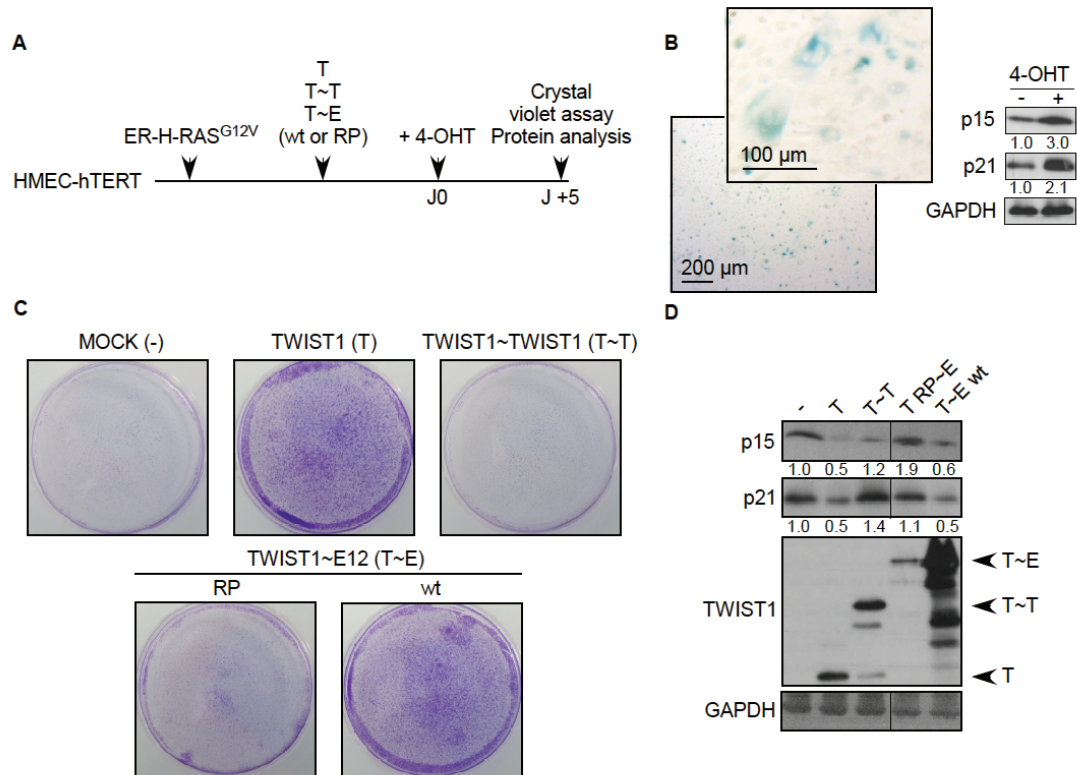


Figure 1

Figure 1. Homo- and heterodimeric TWIST1 complexes unequally avoid RAS-induced senescence in immortalized human mammary epithelial cells.

A, HMEC-hTERT cells were sequentially infected with the inducible ER-H-RAS^{G12V} fusion protein and with either monomeric TWIST1, the homodimeric TWIST1~TWIST1 tethered dimer, or the wild-type (wt) or mutant (RP) TWIST1~E12 heterodimeric tethered dimer. **B**, In the absence of TWIST1, the activation of the ER-H-RAS^{G12V} protein, using 4-OHT, was associated with the induction of senescence as shown in the SA-β-galactosidase assay, and the accumulation of the cyclin-kinase inhibitors p15^{INK4B} (p15) and p21^{CIP1} (p21), as assessed by Western-blot. **C-D**, Five days post-activation of ER-H-RAS^{G12V} using 4-OHT, cell proliferation was assessed through a crystal violet staining (**C**) and p21^{CIP1} (p21), p15^{INK4B} and TWIST1 proteins were analyzed by Western-blot (**D**). Relative signal quantification with respect to MOCK cells is indicated.

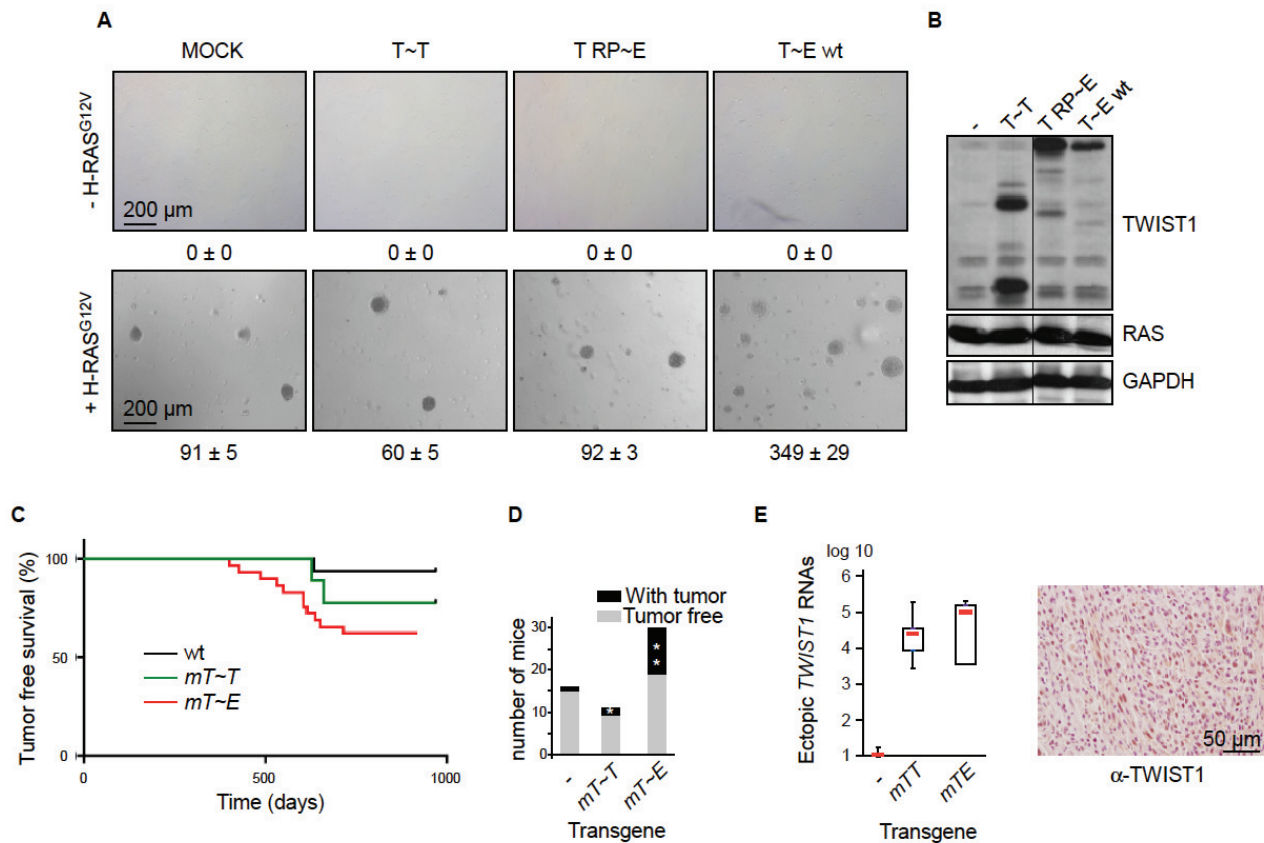


Figure 2

Figure 2. The heterodimeric TWIST1 complex displays oncogenic properties *in vitro* and *in vivo*.

A, HMEC-hTERT cells were sequentially infected with *Twist1* (T~T, wt or mutant T~E) and *H-RAS*^{G12V} retroviral expression vectors. The transformation potential of the established cell lines was assessed in a soft-agar colony formation assay. Mean of numbered colonies and SD of one experiment performed in triplicates are shown. **B**, Analysis of TWIST1 and RAS proteins by Western-blot. **C-E**, Assessment of the TWIST oncogenic potential *in vivo*. The enforced expression of *Twist1* in luminal committed mammary epithelial cells was achieved by crossing either *CAG-LSL-Twist~Twist1* or *CAG-LSL-Twist1~E12* with WAP-Cre mice. **C**, Kaplan-Meier tumor-free survival. **D**, Histogram indicating the number of mice spontaneously developing breast carcinoma during their lifespan. Stars mark mice with two breast tumors. **E**, Left panel: plot boxes indicating the expression of ectopic *Twist1* transcripts in tumors, as assessed by qRT-PCR. Levels were expressed relative to the housekeeping *HPRT1* gene transcript, and were normalized against wild-type mice-derived breast carcinoma. Right panel: example of TWIST1 staining in a metaplastic spindle cell carcinoma.

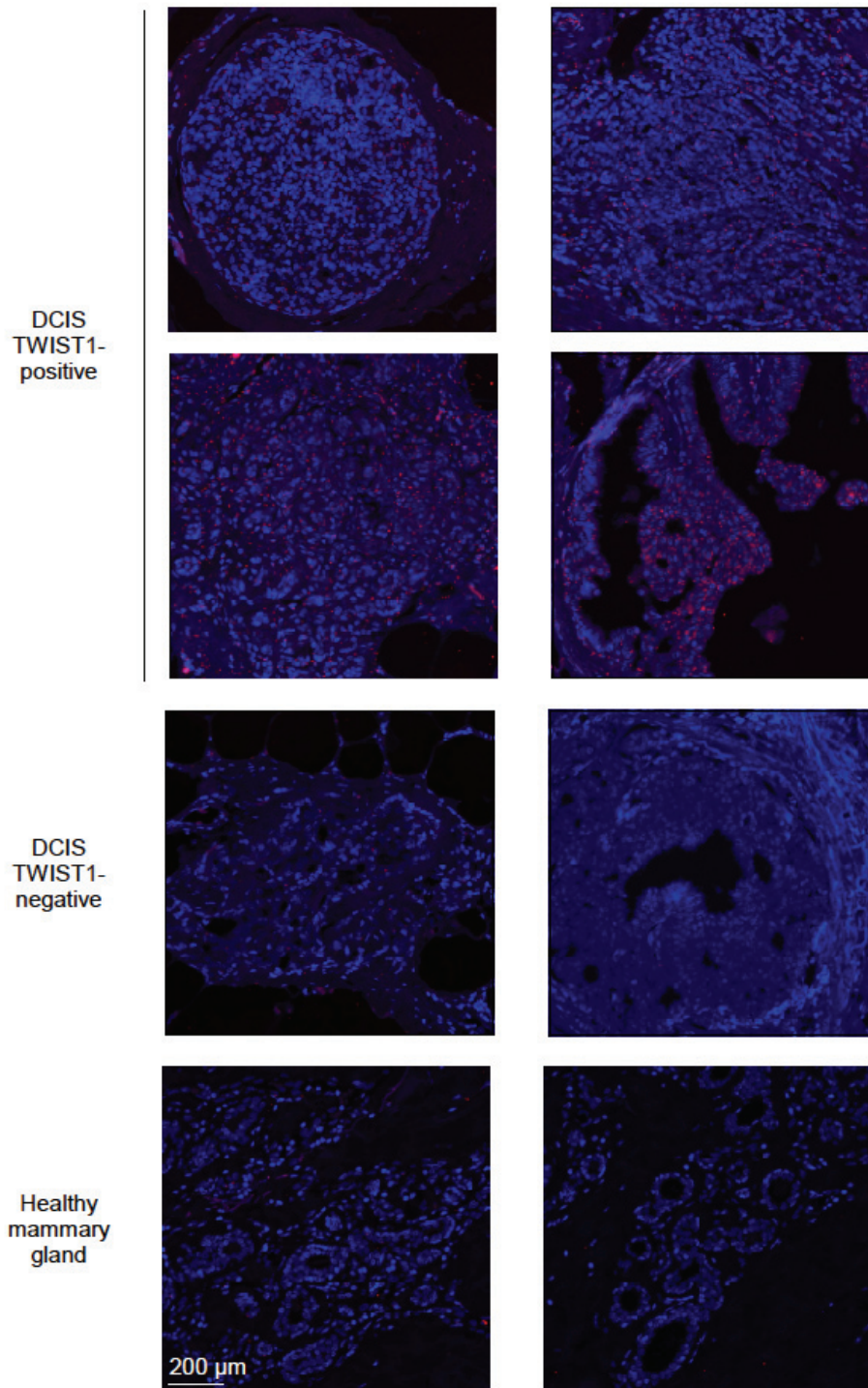


Figure 3

Figure 3. The heterodimeric complex is detectable in breast premalignant lesions, as assessed in a Proximity Ligation Assay.

Assessment of the TWIST1-E12 heterodimeric complex in a proximity ligation assay (PLA) in TWIST1-positive and TWIST1-negative *in situ* ductal carcinoma (DCIS) (Morel et al., 2012). Healthy mammary gland was used as a control. Each single dot corresponds to the detection of a TE complex.

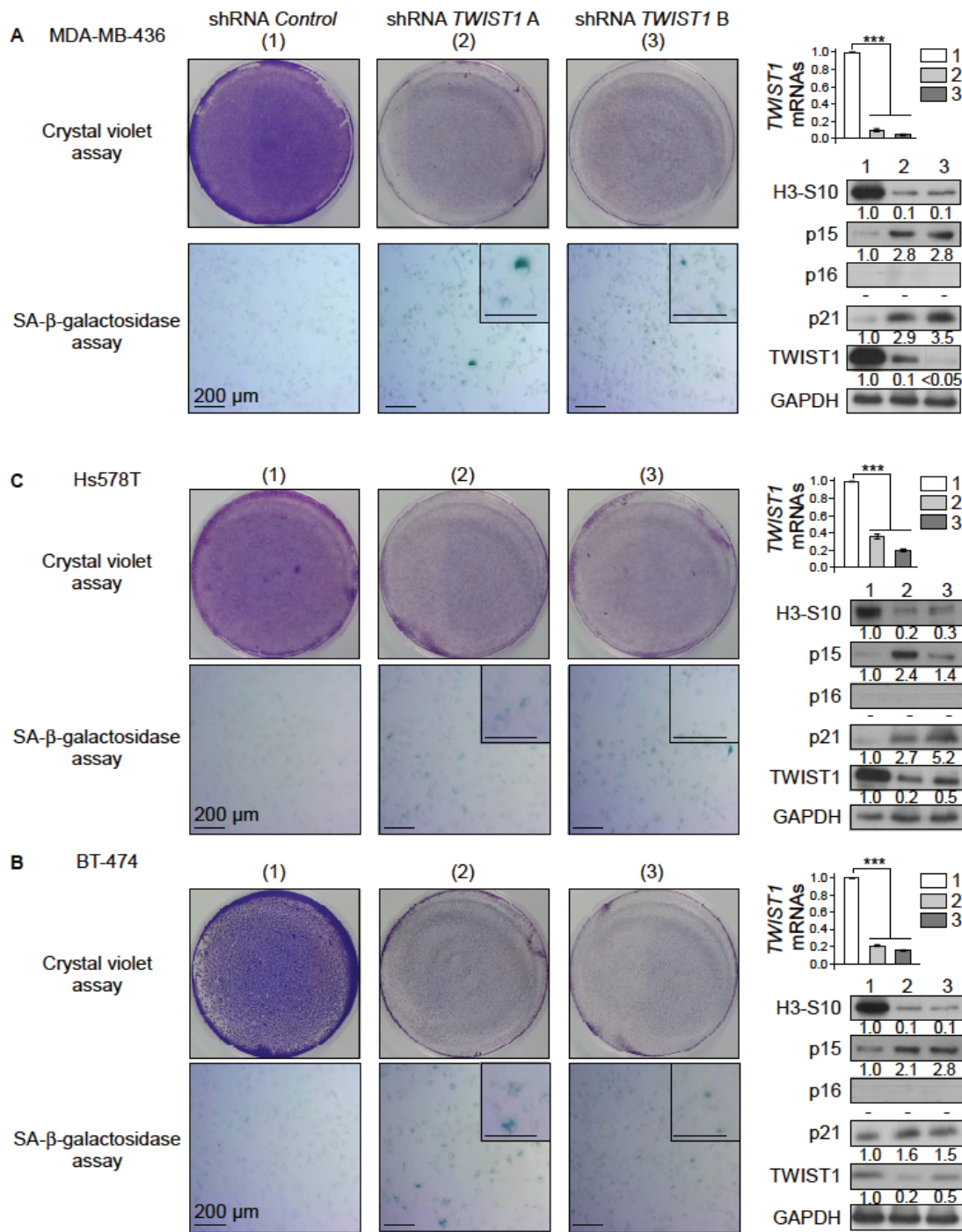


Figure 4

Figure 4. Depletion in TWIST1 reactivates latent OIS in breast cancer cells.

Breast cancer cells (A, MDA-MB-436; B, Hs578T; C, BT-474) were depleted in TWIST1 through RNA interference (shRNA *TWIST1* A or B). Left panels: Two days post-infection, cells were seeded at low density and stained after 3 days with crystal violet or assessed for their SA-β-galactosidase activity. Right upper panels: analysis of the level of the endogenous *TWIST1* transcripts by qRT-PCR. Levels were expressed relative to the housekeeping *HPRT1* gene transcript and were normalized against uninfected cells. Mean, SD and Student's t-test of one experiment performed in triplicates are shown. Right lower panels: analysis of Ser10-histone H3 (H3-S10), p15^{INK4B} (p15), p16^{INK4A} (p16), p21^{CIP1} (p21), and TWIST1 by Western-blotting (relative signal quantification with respect to shRNA *Control* infected cells is indicated). - means below the threshold.

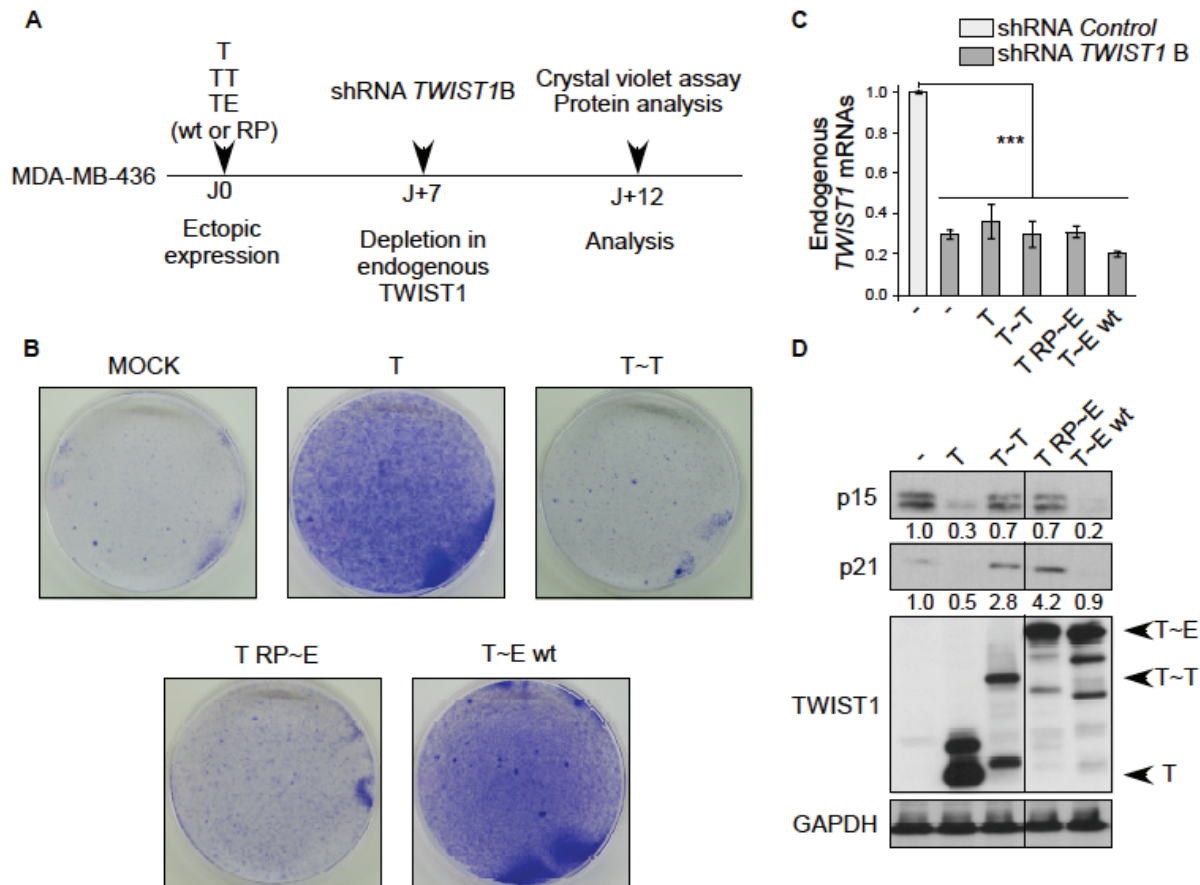


Figure 5

Figure 5. A fully active TWIST1-E12 heterodimeric complex is required to complement the endogenous TWIST1 function in sustaining MDA-MB-436 breast cancer cell proliferation.

A, MDA-MB-436 breast cancer cells were sequentially infected with T, T~T or T~E (either wild-type or RP mutant) retroviral expression vectors and depleted in endogenous TWIST1 through RNA interference (shRNA *TWIST1B*). Of note, the shRNA targets 3' untranslated sequences absent from ectopically expressed TWIST1 constructs and, thus, specifically turns-down the endogenous *TWIST1* expression. **B**, Cells were seeded at low density 48 h post-second infection and stained with crystal violet 3 days later. **C**, Analysis of the level of expression of endogenous *TWIST1* by qRT-PCR. Levels were expressed relative to the housekeeping *HPRT1* gene transcript, and were normalized against uninfected MDA-MB-436 cells. Mean, SD and Student's t-test of one experiment performed in triplicates are shown. **D**, Analysis of p15^{INK4B} (p15), p21^{CIP1} (p21), and TWIST1 proteins by Western-blot (relative signal quantification in respect to MOCK cells is indicated).

Travail 3

« Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences »

Bouard C., Terreux R., Honorat M., Manship B., Ansieau S., Vigneron A., Puisieux A., Payen L.

Nucleic Acid Research, avril 2016

4 Travail 3 : étude de l'interaction de TWIST1/E12 avec différentes séquences E-box

La dimérisation est un prérequis à la fixation des différents dimères à l'ADN au niveau de séquence dites E-box. Ce sont des séquences hexanucléotidiques composées de deux bases centrales variables et de quatre bases externes conservées. Lorsqu'un dimère bHLH se lie à une séquence E-box, le premier monomère se lie à la première moitié de l'E-box qui correspond au premier demi-site, tandis que le second monomère se lie au second demi-site qui se situe à l'extrémité inverse de l'E-box et qui est sur le brin complémentaire. Ainsi chaque monomère contribue à la spécificité d'interaction des dimères aux E-box. Afin d'appréhender les mécanismes moléculaires mis en jeu lors de la liaison à l'ADN de ces dimères, nous avons analysé la dynamique moléculaire du modèle tridimensionnel de l'hétérodimère TWIST1/E12 lié à différentes E-box. Nous avons ainsi comparé la fixation de l'hétérodimère TWIST1/E12 sur une E-box active (de haute affinité du dimère TWIST1/E12) et sur différentes E-box modifiées actives ou inactives. Les modifications des E-box concernent les bases variables ou consensus de l'E-box ou les bases adjacentes à l'E-box. Nous avons validé toutes nos analyses *in silico* par des analyses *in vitro* mesurant l'affinité de TWIST1/E12 aux différentes séquences E-box.

Nous avons tenté d'évaluer l'affinité de liaison de l'hétérodimère TWIST1/E12 à différentes séquences E-box. Pour cela, nous avons analysé les liaisons hydrogènes et hydrophobes établies au sein du dimère protéique lui-même ou entre le dimère et les différentes E-box étudiées. Dans un premier temps, nous avons décrit les interactions protéiques au cœur du dimère. La stabilité des interactions hydrophobes situées au cœur du dimère impliquant à la fois des résidus de Leucine et de phénylalanine, sont de bons indicateurs de l'orientation des domaines bHLH des deux partenaires dans le sillon de l'ADN. Au-delà des interactions hydrophobes, il existe des liaisons hydrogènes clefs au cœur du dimère. Deux couples de résidus clefs localisés, N125-K145 au niveau de TWIST1 et N566-K588 au niveau de E12, jouent un rôle clef dans le maintien de la conformation du cœur du dimère et donc dans la stabilisation de la liaison des dimères de TWIST1 à l'ADN. En effet, il existe des liaisons hydrogènes entre les résidus de lysine et d'asparagine au sein de ces deux couples de résidus qui impactent directement les liaisons spécifiques de résidus clefs comme la R118 de TWIST1 et le E558 de E12 avec les bases variables centrales de l'E-box. Plus les liaisons spécifiques de R118 de TWIST1 et de E558 de E12 sont préservées au cours du temps plus l'affinité de liaison des dimères de TWIST1 à l'ADN est importante. Nous avons noté que l'interaction de la S144 (résidu clef de la boucle de TWIST1), stable tout au long de la dynamique moléculaire contenant la TE-box et donc stabilisatrice de la liaison du dimère TWIST1 à la TE-box, était perdue lorsque les bases adjacentes sont modifiées. Ainsi les résidus retrouvés mutés dans le SCS sont importants de par leur interaction pour la stabilité de la fixation de TWIST1 à l'ADN et donc pour les fonctions de TWIST1.

Dans un second temps, nous avons étudié les interactions établies entre les résidus d'acides aminés des protéines et les acides nucléiques de l'ADN. Notre approche *in silico* nous a permis de différencier l'affinité de TWIST1/E12 pour les E-box actives ou inactives et de montrer le rôle important des premières bases de chacun des deux demi-sites de l'E-box. Ce sont les bases consensus (1) et (1*) qui jouent un rôle clef dans la reconnaissance des E-box par les dimères b-HLH. En effet, les interactions spécifiques impliquant ces bases ont été perdues lorsque la dynamique moléculaire a été réalisée avec le modèle contenant une E-box inactive (BMID-box où toutes les bases sont dégénérées). Ensuite, nous avons étudié l'impact de la substitution au niveau des bases conservées et variables de la TE-box sur l'affinité de liaisons établies au sein du dimère et entre le dimère et l'ADN. Cette approche nous a permis de mettre en évidence d'une part que les différences d'affinité de TE pour les différentes E-box étudiées s'expliquent par une redistribution des liaisons hydrogènes spécifiques au cours des simulations et d'autre part que les bases variables de l'E-box, qui sont également la troisième base de chacun des deux demi-sites de ces E-box, contribuent également à la spécificité de reconnaissance des différentes E-box par les dimères.

Pour résumer nos résultats, nous avons décrit trois états conformationnels selon la séquence ADN présente entre les deux domaines basiques du dimère. La première conformation est représentée par la liaison du dimère à une E-box dite de haute affinité avec le complexe (TE-box). Dans cette configuration, il existe des liaisons hydrogènes stables établies entre les bases variables de l'E-box et les chaînes latérales des résidus d'acides aminés. Cette conformation permettant une activité transcriptionnelle complète du dimère TWIST1/E12. La seconde conformation est dite intermédiaire et comprend toutes les E-box actives établies à partir de la TE-box où une base conservée ou variable a été substituée. L'affinité de TWIST1/E12 pour ces différentes E-box actives est diminuée ce qui rend ces E-box moins fonctionnelles pour TWIST1. Et enfin le troisième état conformationnel est non spécifique (fonction de « scanning »). Il correspond à une conformation protéique instable qui permet au dimère de TWIST1 de conserver toute sa flexibilité pour continuer à rechercher des E-box plus spécifiques. Dans ce cas, nous observons une augmentation des interactions des chaînes latérales des résidus de TWIST1/E12 avec les bases adjacentes à l'E-box. En accord avec ces résultats *in silico*, nous n'avons pas observé dans ce cas de liaison à l'ADN *in vitro* confirmant donc l'absence d'activité transcriptionnelle induite lorsque l'hétérodimère TWIST1/E12 rencontre une telle séquence.

Au cours de cette étude, nous avons mis en évidence le rôle prépondérant de la nature des chaînes latérales des résidus de protéines localisés à proximité du cœur du complexe et des séquences d'ADN. Une diminution de l'affinité de liaison se traduit par un glissement des interactions du dimère protéique des bases variables centrales de l'E-box vers les bases conservées externes de l'E-box ou vers les bases adjacentes à l'E-box (-1; -1*). Toutes nos analyses *in silico* ont été validées par des études *in vitro* via une approche stréptavidine/biotine. Ainsi cette approche nous a permis de montrer

la valeur prédictive de nos simulations *in silico* et pourrait être utilisée pour différentes approches. En effet ces modèles tridimensionnels pourraient être utilisés pour étudier l'implication structurale de certains acides aminés mutés et associés notamment au syndrome de Saethre-Chotzen et pour le criblage d'inhibiteurs spécifiques de l'hétérodimère oncogénique de TWIST1, TWIST1/E12. Cette stratégie *in silico* peut également être utilisée pour la caractérisation d'autres protéines.

Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences

Charlotte Bouard^{1,2,3,4,5,6}, Raphael Terreux^{6,7,8}, Mylène Honorat⁷, Brigitte Manship⁵, Stéphane Ansieau^{2,3,4,5}, Arnaud M. Vigneron^{1,2,3,4,5}, Alain Pulsieux^{1,2,3,4,5,6,9} and Léa Payen^{1,2,3,4,5,6,10,*}

¹Inserm UMR-S1052, Centre de Recherche en Cancérologie de Lyon, Lyon 69373, France, ²CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, Lyon 69373, France, ³LabEX DEVweCAN, Lyon, France, ⁴UNIV UMR1052, Lyon 69008, France, ⁵Centre Léon Bérard, Lyon 69373, France, ⁶Université de Lyon1, ISPB, Lyon 69008, France, ⁷Institut de Biochimie des protéines IBCP, Lyon 69007, France, ⁸CNRS UMR 5305, Lyon 69007, France, ⁹Institut Universitaire de France, Paris 75231, France and ¹⁰Hospices Civils de Lyon, Laboratoire de Biochimie et Biologie Moléculaire du CHLS, Lyon 69003, France

Received January 15, 2016; Revised April 11, 2016; Accepted April 13, 2016

ABSTRACT

The TWIST1 bHLH transcription factor controls embryonic development and cancer processes. Although molecular and genetic analyses have provided a wealth of data on the role of bHLH transcription factors, very little is known on the molecular mechanisms underlying their binding affinity to the E-box sequence of the promoter. Here, we used an *in silico* model of the TWIST1/E12 (TE) heterocomplex and performed molecular dynamics (MD) simulations of its binding to specific (TE-box) and modified E-box sequences. We focused on (i) active E-box and inactive E-box sequences, on (ii) modified active E-box sequences, as well as on (iii) two box sequences with modified adjacent bases the AT- and TA-boxes. Our *in silico* models were supported by functional *in vitro* binding assays. This exploration highlighted the predominant role of protein side-chain residues, close to the heart of the complex, at anchoring the dimer to DNA sequences, and unveiled a shift towards adjacent ((-1) and (-1*)) bases and conserved bases of modified E-box sequences. In conclusion, our study provides proof of the predictive value of these MD simulations, which may contribute to the characterization of specific inhibitors by docking approaches, and their use in pharmacological therapies by blocking the tumoral TWIST1/E12 function in cancers.

INTRODUCTION

The epithelial-mesenchymal transition (EMT) is a highly conserved cellular process that allows the transient conversion of epithelial cells into mesenchymal cells and promotes cell migration during embryonic processes. The TWIST1 protein has been associated with this cellular mechanism and was originally described in the developmental process, as an initiator of the formation of the mesoderm. Indeed, this protein is involved in several biological activities *via* its different functional domains, including the TWIST-box domain (1), which is highly homologous between different eukaryotic species (2), and the basic Helix Loop Helix (bHLH) domains, which enable its dimerization with other bHLH partners (3,4). The TWIST-box domain binds with the runt domain of the RUNX2 protein, and inhibits its transactivation function (2). In parallel, TWIST1 is able to directly repress the expression of *Runtx2*, thus preventing RUNX2 from inducing premature osteoblast differentiation (5,6).

However, the aberrant activation of the EMT in adult cancer cells contributes to the invasion of carcinoma cells and metastasis (7,8), and concomitantly, TWIST1 has also been implicated in tumorigenesis (8,9). In this context, the molecular events triggered by TWIST1 in cancer cells, may be induced through both the TWIST-box domain and the bHLH domain. For example, the TWIST-box domain mediates the formation of a protein complex comprised of TWIST1 and the nuclear factor-kappaB (NF- κ B) subunit RELA (p65/NF- κ B3), which activates the transcription of the *IL8* gene (10). The TWIST-box domain also interacts

*To whom correspondence should be addressed. Tel: +33 478863208; Fax: +33 478782720; Email: Lea.payen-gay@univ-lyon1.fr
Present address: Pr. Lea Payen, Inserm UMR-S1052, Centre de Recherche en Cancérologie de Lyon, Lyon 69008, France.

with p53, which antagonizes p53-dependent activity, especially its pro-apoptotic and transcription functions (11).

Clinically, while the re-expression of TWIST1 in tumors is frequently associated with a poor prognosis in humans, on the one hand (12–14), its repression on the other hand decreases the growth of xenografts (15). Additionally, a decrease in the TWIST1 protein, both in a transgenic mutant Kras lung cancer mouse model and in human lung tumors, restores a senescence program inducing the loss of a neoplastic phenotype (16). Similarly, such a phenomenon is observed using a 8-bromo-5-hydroxy-7-methoxychrysin treatment (15). Finally, the decrease in the level of expression of TWIST1 by small interfering RNA (siRNA) was shown to potentiate the effect of arsenic trioxide chemotherapy (17).

In the present work, we focused on the bHLH domain of a TWIST1 complex, and on its function in the binding of targeted gene promoters. It is well known that TWIST1 is a transcription factor belonging to the bHLH superfamily, which includes TWIST1, TWIST2, E12, E47, HAND1 and HAND2 (3,4). All of them are highly conserved in organisms from yeasts to humans and are composed of two characteristic domains: a HLH domain and a basic domain (4,18). The dimerization carried out *via* the HLH domains is a prerequisite for the interaction with E-box sequences occurring *via* basic domains and key residues (3) (this model is depicted in Figure 2C to facilitate its understanding). According to the formed homo- or heterodimer, their transcriptional functions differ. For instance the TWIST1 functions have been largely explored using tethered dimers in several eukaryotic models (19–21). A previous study focusing on the characterization of different prometastatic forms of TWIST1 demonstrated the predominant role of the TWIST1/E2A (E12 or E47, two alternative splicing products of the TCF3 gene) heterodimer in prostate cancer (22). Furthermore, the phosphorylation of TWIST1 at the level of the threonine T121 and the serine S123 residues may enhance the metastatic potential of the TWIST1/E12 (TE) complex (22).

A comprehensive classification of these transcription factors has been established according to their tissue distribution, dimerization capabilities (homodimer or heterodimer), and binding specificities to various E-box sequences (23). Since bHLH proteins can form both homo- and heterodimers, they can be further classified into two groups based on their binding preferences with numerous canonical core recognition E-box sequences (CANNTG), namely CACCTG or CAGCTG (group A) and CACGTG or CATGTG (group B) (23). When a bHLH dimer binds to an E-box sequence, the first monomer binds to the forward half-site, while the second monomer binds to the reverse complementary half-site (24). Each monomer, thereby, contributes to making its binding to a particular E-box sequence highly specific, presumably by preferring particular half-sites. For the TE complex, the half-site preferences are the CAT sequence for TWIST1, and the CAT, CAC and CAG sequences for E12 (24). These differences in the composition of the E-box sequences largely modulate the *in vivo* function of cellular transactivation by closely related transcription factors (24–26) (for brief review see Figure 1A). Recently, Chang demonstrated that a secondary organization might greatly modulate the transactivation function

of the bHLH domain (27). The TE complex can dimerize to form a tetramer able to recognize specific double E-box motifs and, in doing so, is able to align two TE complexes on the consecutive DNA grooves, providing a highly stable, bound tetramer (27). Nevertheless, little is known about the molecular characterization of the canonical core recognition of E-box sequences (CANNTG) by the TE complex (23).

The use of molecular dynamics (MD) simulations previously enabled us to study the *in silico* impact of TWIST1 alterations (insertion or single base mutations) on the binding of the TE complex to DNA (28,29). This approach provided a structural explanation for the loss of function associated with haplo-insufficiency of TWIST1 observed in patients suffering from Saethre-Chotzen syndrome (SCS) (30–34). In the present study, the molecular mechanisms underlying the recognition of the E-box by the TE heterodimer and their stabilization on DNA were investigated using the previously published dimer, initially generated by homology with the murine NEUROD1/E47 crystal structure (28,35). We generated various *in silico* models of the TE complex bound to putative functional E-box sequences of targeted gene promoters in order to verify the specificity of the binding domains during MD simulations. Consequently we assumed that a positive outcome of our approach, combining *in silico* models with *in vitro* assays, may result in the dynamic 3D visualization of a novel predictive model for the identification of specific and functional E-box sequences by the TE complex. This could be used in the future for developing and testing pharmacological TWIST1/E2A blocking molecules by *in silico* docking prior to *in vitro* or *in vivo* studies. Such therapeutic strategies may prove to be invaluable for restoring the pro-senescence and apoptosis responses (9,36–39). Furthermore, identification of a specific interaction between the TWIST1/E2A complex and a distinct E-box sequence, could lead to the determination of a putative pro-oncogenic signature.

MATERIALS AND METHODS

In silico study: generation of putative functional models of the TE complex binding to E-box sequences

Selection of X-ray structure and *in silico* model building. The conserved TE dimer model used throughout this study was generated previously (28). Briefly, sequences of human TWIST1 (Q15672), E12 (P15923-1), NEUROD1 (Q13562) and E47 (P15923-2) were downloaded from the UniProtKB/Swiss-Prot website. Primary sequences, restricted to the bHLH domains, were aligned using the ClustalW 2.0 software (40). In their bHLH domains, human E12 and TWIST1 share 85% and 48% homology with the murine E47 and NEUROD1 sequences, respectively. The X-ray structures of the mouse NEUROD1/E47 complex (PDB code 2ql2) were downloaded from the RCSB protein data bank (35). Two X-ray structures of the mouse NEUROD1/E47 were reported, namely the 2ql2A/2ql2B and the 2ql2C/2ql2D (35), which diverge slightly (root-mean-square deviation evaluated at 0.618 Å) and contain a well-known E-box sequence named the TE-box (5'-TAGGCCATCTGGTCCT-3'). Taking into account the degree of homology between residues in the sequences, *in sil-*

ico models of the conserved TE complex (TE-model) were generated by homology to the 2ql2A/2ql2B structure (28).

The molecular interactions of this TE dimer with several different E-box sequences were then assessed. E-box sequence modifications (or flanking base substitutions) were based on studies evaluating the specificity of the interactions between the canonical 5'-CANNTG-3' E-box sequences and bHLH transcription factors (20,24,41,42) (Figure 1A). All the DNA base substitutions were carried out in the initial TE-box using the mutation function of the Sybyl-X 1.1 software package. The resulting *in silico* models studied herein were, therefore, the TE complex bound either to an inactive BMID-box (degenerate sequence with low affinity for TWIST1 complexes), or to various active sequences, namely the initial TE-box, a BMI-box, an E12-box, a TWIST-box and a D-box, or to sequences displaying modifications in their flanking bases, such as the TA-box and the AT-box (Figure 1A).

Molecular dynamics simulations. Prior to conducting molecular dynamics (MD) simulations, which have been described previously (28), minimization of all of the *in silico* models was carried out using the Sybyl-X 1.1 software package, elaborated by the Tripos Company. The software parameters included a force field using Gasteiger–Marsilli partial charges with a cut off set at 12 Å, and a dielectric constant of 80 to simulate an implicit water phase (the dielectric constant of water being 20.10 at 20°C). No restraints were applied to the *in silico* models. The established homology model (including DNA sequences) was visualized *via* the VMD1.9.1 software (43). The resulting model was inserted into a parapepetic TIP3P solvent box using the 'add solvation box' module of the VMD 1.9.1 software. A distance of 15 Å was set between the surface of the proteins and the limit of the solvent box (44). For each structure a 10 ns MD simulation was computed using the NaMD software and CHARMM 27 as a force field. The state of equilibrium was reached after approximately 30 ps for all of the models studied.

Analysis of the stability of interactions during MD simulations. Among the interactions occurring between residue-residue and residue-base, we focused our investigation on hydrophobic and hydrogen bond (H-bond) interactions established between donor and acceptor groups. In the first place, hydrophobic interactions are mainly due to the tendency of nonpolar residues to aggregate in aqueous solutions and repulse water molecules. These interactions contribute to the folding of the protein and to the stabilization of the macromolecule. Residues such as alanine, valine, leucine, isoleucine, phenylalanine, tryptophan and methionine can cluster and form hydrophobic interactions. Here, we observed the presence of numerous leucine and phenylalanine residues in the vicinity of the heart of the TE complex. During the MD simulations, the distances in angstrom (Å) between pairs of residues were measured and plotted as a function of time (represented by a frame number: 1 frame per 10 ps). The distances between hydrophobic interactions were represented *via* GraphPad Prism 5: 'smooth, differentiate or integrate curve, with eight neighbours'. In the second place, the H-bond interaction is an intermolecular force in-

volving a hydrogen atom and an electronegative atoms such as oxygen or nitrogen. H-bonds can be established between donor atoms (arginine, lysine and tryptophan) and acceptor atoms (aspartic acid, glutamic acid) of side chain residues. We studied the residue-residue or residue-base interactions, which are established between atoms (O–H...:N; O–H...:O; N–H...:N; N–H...:O). Again, the lengths of H-bond interactions were represented *via* GraphPad Prism 5: 'smooth, differentiate or integrate curve, with eight neighbours'. These lengths have previously been estimated around 2–2.8 Å depending on the nature of the atoms (45). Using a computer software, we set an arbitrary threshold at 2.1 Å and made the following hypothesis: any interaction occurring under this threshold was deemed 'true' and assigned a score of 1, whereas any interaction occurring above this threshold was deemed 'false' and given a score of 0 (SI function: SI(test_logic; value_if_true;value_if_false) with logic test: '<2.10', value of 1 attributed if true and 0 if false; NB.SI function: NB.SI(range;criterion)). The scores were reckoned and the value obtained was designated as the rate of occupancy of H-bond interactions. Thus, for a given H-bond, the higher the rate of occupancy, the greater the number of interactions occurring under 2.10 Å, the higher the level of stability of that H-bond during the MD simulation. All calculations were carried using Excel.

In vitro study: binding of the TE complex to the E-box sequences

Vector constructs. The pBABE-neo expression vector was donated by H. Land & J. Morgenstern & B. Weinberg (Addgene plasmid # 1767). The cDNA of the tethered TE complex composed of 896 amino acids, including a FLAG-tag at the N-terminal position, TWIST1 cDNA, an amino acid linker and E12 cDNA, was cloned into the pBABE-neo expression vector (21). This construct of the ectopic TE protein is approximately 99 kDa in size. The final construct was named the FLAG-TWIST1/E12 pBABE-neo expression vector (TE pBABE; Figure 1Bi).

The cDNA of the tethered TT complex consists of 445 amino acids, including a FLAG-tag at the N-terminal position, TWIST1 cDNA, an amino acid linker and TWIST1 cDNA, was cloned into the pBABE-neo expression vector (21). The final construct (~49 kDa in size) was named the FLAG-TWIST1/TWIST1 pBABE-neo expression vector (TT pBABE; Figure 1Bii). Finally, the cDNA of the E12 protein (18) was cloned into the pBABE-neo expression vector. The ectopic E12 protein includes a MYC-tag at the N-terminal position, and the final construct is composed of 666 amino acids with a size of 73 kDa. This construct was named the MYC-E12 pBABE-neo expression vector (E12 pBABE; Figure 1Biii). The ectopic E12 homodimer has a size of 147 kDa (E12 pBABE).

Cell culture, transfection and protein extraction. HeLa cells were transfected in 10 cm in diameter culture dishes at a density of 80–90% using the FuGENE HD Transfection Reagent for 48 h (Promega #E2311). The transfection reagent contained 19 µg of the empty pBABE-neo expression vector or of the pBABE-neo expression vector (for the FLAG-TWIST1/E12 pBABE vector), as well as 56 µl of

FuGENE. Following their transfection, cells were washed with cold PBS (2.7 mM KCl, 1.47 mM KH₂PO₄, 0.14 M NaCl, 3.4 mM Na₂HPO₄) and lysed in EBC solution (20 mM Tris pH 8.0, 385 mM NaCl, 2 mM EDTA, 3.5% NP40, PIC 1×, PMSF 1×). After incubation on ice for 20 min, lysates were collected by scrubbing the plates, sonicated (6 cycles of 10 s of ultrasound sonication with a 20 s gap between each cycle) and centrifuged for 20 min at 15 000 rpm at 4°C. The supernatant was collected and the total protein content was quantified using the Bio-Rad Protein Assay (#500-0001).

Streptavidin/Biotin binding assays. The Streptavidin/Biotin binding assay was conducted to isolate proteins (the TE complex) bound to specific DNA sequences (E-box sequences). To do so, total cellular proteins (extracted from cells transfected with the TE pBABE construct or with the empty pBABE control) were incubated with biotin-coupled DNA sequences, and the DNA bound TE complex was then purified by immunoprecipitation (IP) via the recognition of biotin by Streptavidin beads. Hence, this approach decreased the risk of isolating non-specifically bound proteins. Briefly, both strands of DNA probes corresponding to various E-box sequences were labelled in their 3' extremity with biotin. The probes used were: the inactive BMID-box (5'-CGTAGGCTGGTGTGTCCTCG-3'), the active TE-box (5'-CGTAGGCCATCTGGTCCTCG-3'), the BMI-box (5'-CGTAGGCCAAGTGGTCCTCG-3'), the E12-box (5'-CGTAGGCCACCTGGTCCTCG-3'), the TWIST-box (5'-CGTAGGCCATATGGTCCTCG-3'), the D-box (5'-CGTAGGCCGTCTGGTCCTCG-3'), the TA-box (5'-CGTAGGTCATCTGATCCTCG-3') and the AT-box (5'-CGTAGGACATCTGTCCTCG-3'). For probe hybridization, an initial denaturing step at 95°C for 5 min was followed by a hybridization step for 10 min at the specific hybridizing temperature of each probe. Probe hybridizations were then confirmed on 3% agarose gels.

The binding between the DNA hybridized probes (3 µg) and the TE complex (80 µg) was carried out in binding buffer (100 mM KCl, 2% glycerol, 20 mM Hepes, 1 mM dithiothreitol and 0.1 mg/ml BSA) for 1 h at 30°C. Biotinylated DNA probes bound to the TE complex were purified using DynaBeads® M-280 Streptavidin (Invitrogen #11205D). Coupled Streptavidin DynaBeads® bound to the DNA/TE complex were run on a 12% SDS-polyacrylamide gel using a migration buffer (Tris glycine-SDS 1X) at 80 V for 10 min in the stacking gel, followed by the migration phase at 100 V for 120 min. The liquid transfer onto a PVDF membrane was conducted using a transfer buffer (Tris glycine 1× and 20% ethanol) at 200 mA for 60 min. The membrane was then saturated with 5% Tris-buffered saline (TBS) containing 0.05% Tween and 0.5% milk for 45 min, and incubated with the mouse monoclonal TWIST1 antibody (abcam #ab50887) as a primary antibody, and a polyclonal rabbit anti-mouse immunoglobulin conjugated with horseradish peroxidase as the secondary antibody (Dako #P0260). Figures 11B and 11C, the membrane was incubated with the rabbit polyclonal antibody E2A (V-18) as a primary antibody (SantaCruz #sc-349), and a polyclonal goat anti-rabbit immunoglobulin conju-

gated with horseradish peroxidase as the secondary antibody (Dako #P0448). The detection was performed by Western blot analysis using the Luminol Reagent (Santa Cruz #sc-2048). The major steps of this *in vitro* transfection and Streptavidin/Biotin procedure are summarized in a schematic diagram (Figure 1C).

RESULTS

To characterize the molecular mechanisms underlying the binding and recognition of the E-box sequences (Figure 1A) by the previously generated TWIST1/E12 (TE) complex (28), we submitted all *in silico* models to 10 ns molecular dynamics (MD) simulations. The resulting binding configuration clearly showed the localization of the TE-box sequence (5'-CATCTG-3') in the DNA groove, between the basic domains of the TWIST1 and E12 proteins (Figure 2A and 2C). The positions of the consensus and variable bases of the E-box in the DNA groove were also represented (Figure 2Bi). The proximal flanking bases (called adjacent bases) of the E-box are represented in yellow and blue (Figure 2Bii). The basic domains of TWIST1 and E12 proteins are represented in solid grey and green, respectively. To explore the interactions between the TE-complex and the various E-box sequences, we differentiated: (i) the 'non-specific' H-bonds (Figure 2Di), established between oxygen elements on the phosphate groups of the DNA bases and an atom of the side chain of the TWIST1 or E12 residues, from (ii) the 'specific' H-bonds established between atoms on the purine and pyrimidine nucleobases of the DNA and an atom of the side chain of TWIST1 or E12 residues (Figure 2Dii).

Comparison of the TE complex bound to both active (TE-box, BMI-box) and inactive (BMID-box) E-box sequences

To determine whether our *in silico* approach was able to predict modifications in the binding of the TE complex to DNA, we studied the binding of the dimer to two active E-box sequences, namely the TE- and BMI-boxes, and to the inactive BMID-box. Our investigations focused on two major interactions occurring in the heart of the TE complex, which may alter its structure, namely the hydrophobic interactions involved in the dimerization of the protein via HLH domains, and the H-bonds established between critical residues. Several leucine (L) and phenylalanine (F) residues close to the heart of the dimer were of particular interest (Figure 3A and 3B). We studied whether they were involved in dimer cohesion by hydrophobic interactions contributing to the dimerization function. In the first instance, the distances between hydrophobic residues localized at the interface of the TE complex (L124-L592, L149-L592, L131-L592, L124-L589, L149-L572, F128-F569) were determined (Figure 3Ci-Cvi). During the MD simulations, L124-L592 and L149-L592 distances were around 2.5 Å, while L131-L592, L149-L572 distances were under 2.3 Å. This strongly indicated that these hydrophobic interactions strongly structured the TE dimer (Figure 3Ci-Civ) irrespective of the E-box sequence included in the model. During the MD simulations the residues (L124-L589) moved closer together in the active BMI and TE models compared to the inactive BMID model, with distances between residues (for

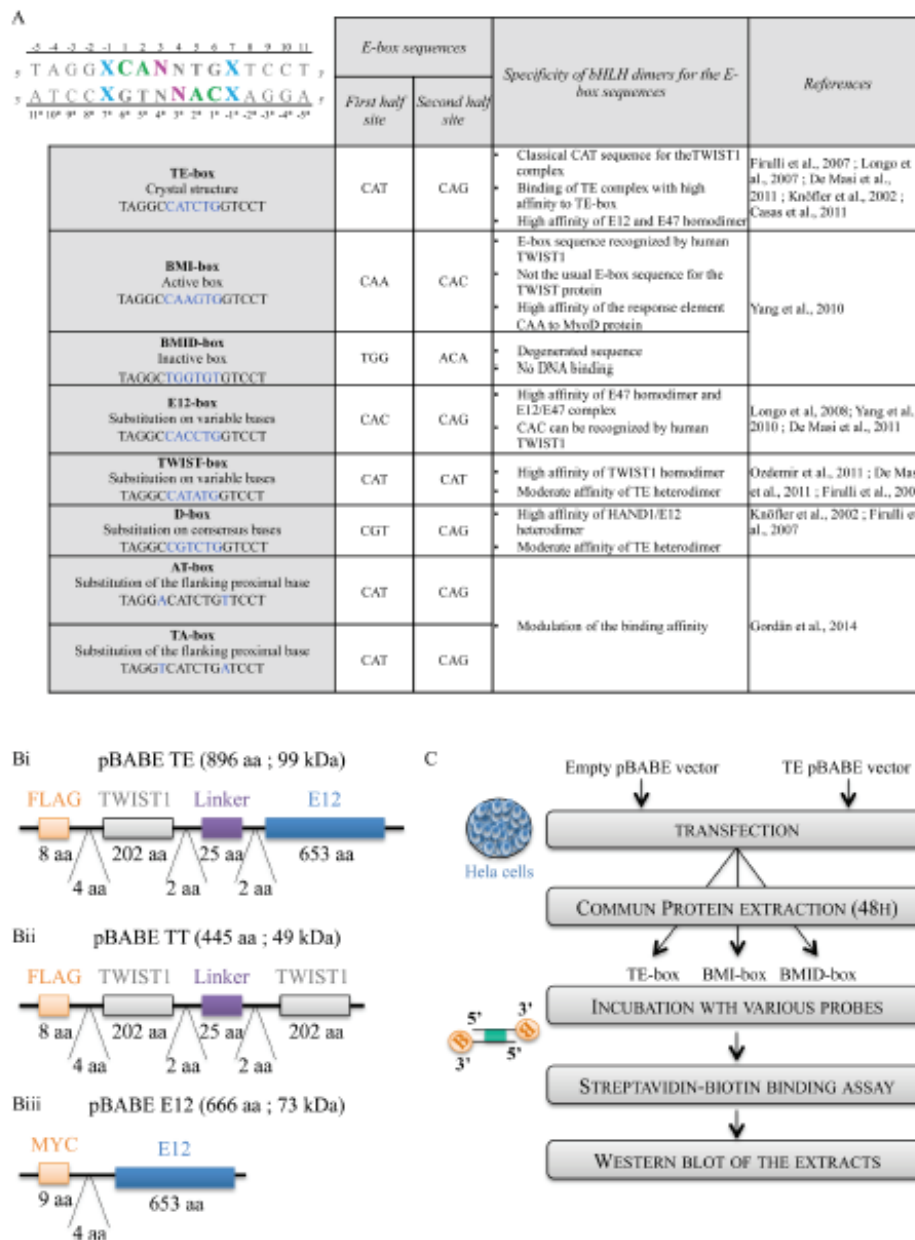


Figure 1. Description of various E-box sequences and the *in vitro* experiments. (A) Presentation of the different E-box sequences selected in the present study, with the localization of base substitutions and the sequences of the two half-sites. Top, representation of the E-box sequence, with the adjacent bases (-1) and (-1*) in blue, conserved bases in green the E-box and the variable bases in pink. The origin of these boxes along with the previous *in vitro* biochemical findings that prompted the choice of these specific E-box sequences are also reported in the table. The E-box sequences are highlighted in blue in the table. (B) Schematic representation of the different protein constructs used in our study. (i) The cDNA of the tethered TE complex is composed of 896 amino acids and its size is ~99 kDa. It contains a FLAG-tag at the N-terminal position, TWIST1 cDNA, an amino acid linker and E12 cDNA. (ii) The ~49 kDa cDNA of the tethered TT complex is composed of 443 amino acids, and includes a FLAG-tag at the N-terminal position, TWIST1 cDNA, an amino acid linker and TWIST1 cDNA. (iii) The cDNA of the E12 protein includes a MYC-tag at the N-terminal position, and the final ~73 kDa construct is composed of 666 amino acids. (C) Schematic diagram briefly summarizing the Streptavidin/Biotin assay using the TWIST1 antibody. Briefly, in step 1, cells were transfected with a plasmid to produce large amounts of the TWIST1/E12 complex. In step 2, total proteins were extracted. In step 3, proteins were incubated with E-box probes. In step 4, TWIST1/E12 complexes bound to E-box probes were isolated by immunoprecipitation in a Streptavidin/Biotin assay using biotin-labelled E-box probes. In step 5, the protein complexes bound to E-box sequences were visualized by Western blot analysis and the TWIST1/E12 signal was detected using the TWIST1 antibody. A detailed description of these steps is reported in the Materials and Methods section.

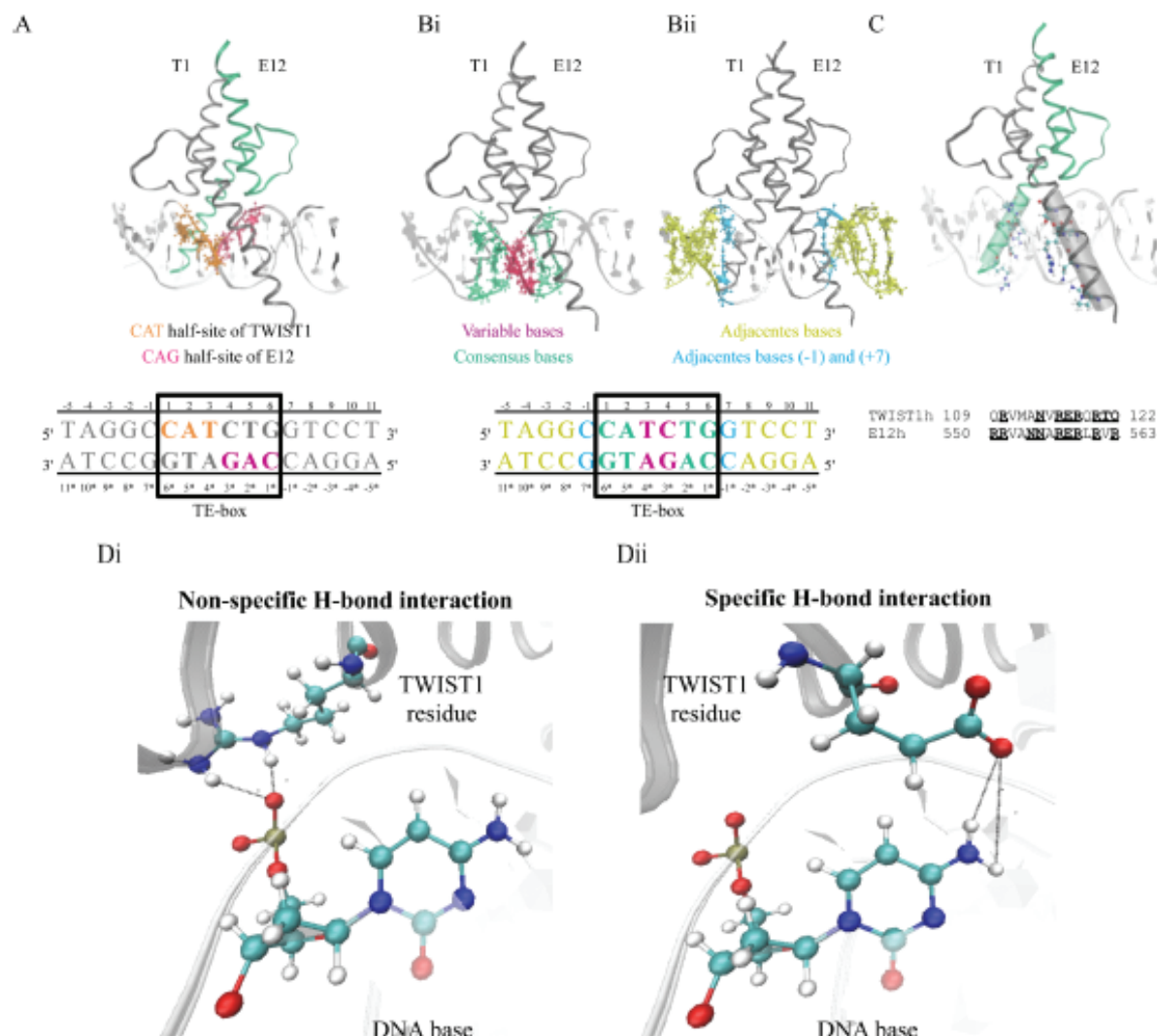


Figure 2. Description of several canonical recognition cores of E-box sequences (CANNTG) and basic domains of the TE complex. (A-D) The position of important TWIST1/E12 amino acid residues and E-box bases were visualized with the VMD 1.9.1 software. Different views are shown for a better understanding and interpretation of the findings. (A-C) 3D representation of the conserved TWIST1 (T1; grey ribbon)/E12 (green ribbon) complex bound to DNA (light white ribbon) along with the TE-box sequence (CATCTG). The residues are numbered in each strand of the 5'-CATCTG-3' E-box core. The complementary bases of the opposite strand are highlighted with a star. (A) Representation of the TWIST1 half-site (5'-CAT-3') highlighted in orange and the E12 half-site (5'-CAG-3') highlighted in magenta. (B) (i) Representation of the variable bases and consensus bases highlighted in pink and green, respectively. (ii) Representation of the flanking bases ((\geq -2) or (\geq -2*)) and proximal flanking bases ((-1) or (-1*)) of E-box shown in yellow and in cyan, respectively. (C) 3D *in silico* visualization and ClustalW alignment of the basic domains of the TE complex. The underlined residues are able to establish H-bond interactions with DNA and/or other TE residues. The 3D visualization shows the basic TWIST1 and E12 domains in solid grey and green, respectively. The CPK representation highlights the TE residues able to establish H-bonds with DNA. (D) *In silico* model of 'specific' or 'non-specific' H-bonds established between TWIST1 and the DNA sequence. (i) The H-bonds are considered as 'non-specific' when the lateral chain residues interact with atoms on the phosphate group. In contrast, (ii) the H-bonds are considered as 'specific', when the lateral chain residues interact with atoms on the purine or pyrimidine nucleobases. The composition of the atoms varies according to the nucleobase and the number of rings.

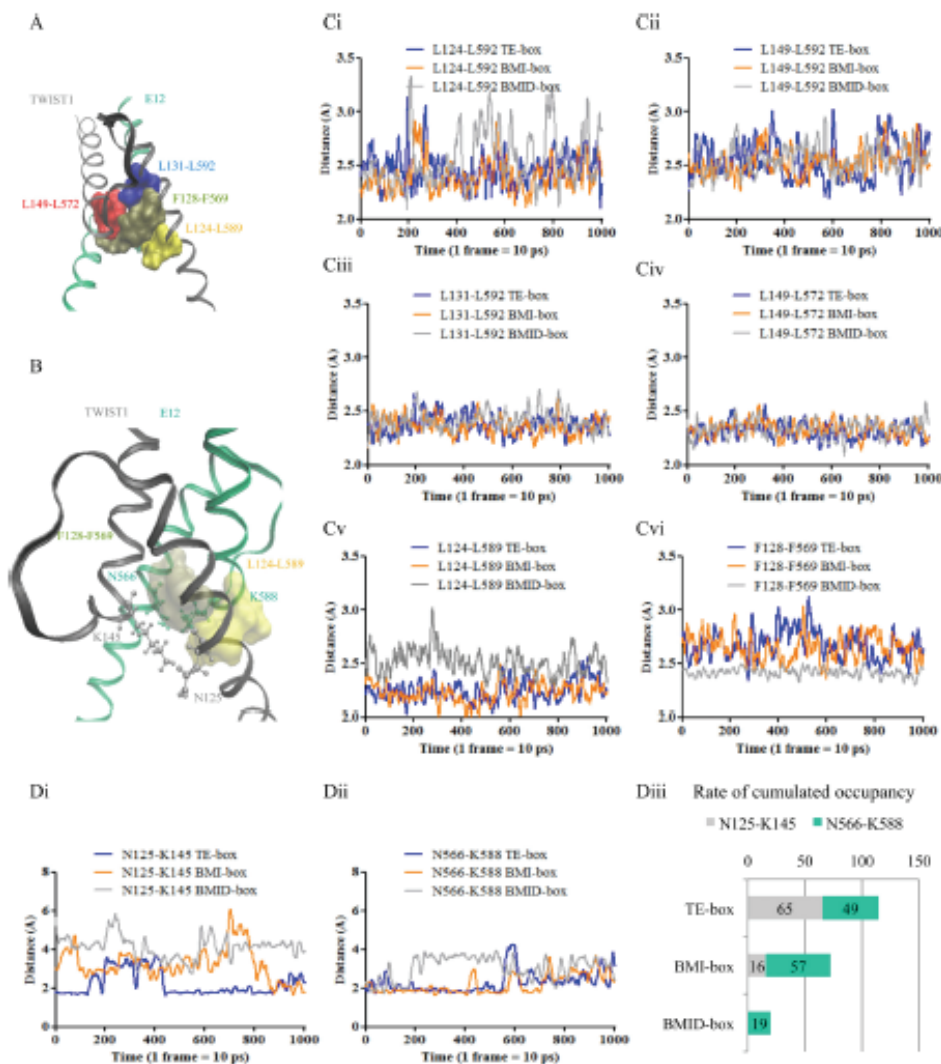


Figure 3. Binding of the TE complex to the active TE-box and BMI-box sequences, as well as to the inactive BMID-box sequences. (A and B) 3D visualization of the localization of the putative hydrophobic bonds in the heart of the TE dimer. TWIST1 and E12 are represented in grey and green ribbons, respectively. The position of important TWIST1/E12 amino acid residues and E-box bases were visualized with the VMD 1.9.1 software. Different views are shown for a better understanding and interpretation of the findings. (A) Two phenylalanine residues (F128 of TWIST1 and F569 of E12), represented in solid green and six leucine residues may strongly contribute to the structure of the heart of the TE complex *via* hydrophobic interactions. Pairs of leucine residues named L124-L589, L131-L592 and L149-L572 are shown in solid yellow, blue and red colours, respectively. (B) Pairs of leucine L124-L589 and phenylalanine residues F128-F569 are represented in solid yellow and green, respectively. The H-bond interactions between asparagine and lysine residues (N125 and K145 of TWIST1 and N566 and K588 of E12) are shown as grey and green CPK representations, respectively. Since the side chain of the lysine has three donors atom (NZ1, NZ2, NZ3), it can make alternatively three H-bonds with the same hydrogen acceptor atom (OD1) of asparagine. This case occurred when we studied the H-bond stability established between N125-K145 and N566-K588. (C) Plots representing the distances (Å) between pairs of residues close to the heart of the dimer as a function of time (total time = 10 ns, 1 frame per 10 ps), during the 10 ns MD simulations of the three TE (blue line), BMI (orange line) and BMID (grey line) *in silico* models. These distances were estimated during the MD simulations using the VMD 1.9.1 software (Graph of Labels Bonds). The lengths of H-bond interactions were represented *via* GraphPad Prism 5: 'smooth, differentiate or integrate curve, with eight neighbours'. The pairs of leucine residues studied were (i) L124-L592, (ii) L149-L592, (iii) L131-L592, (iv) L149-L572 and (v) L124-L589, while the pair of phenylalanine residues was (vi) F128-F569. (D) Plots representing the shortest distances (Å) between the OD1 atom of asparagine N125 and the NZ1, NZ2 or NZ3 atoms of the lysine K145 residues (i), and between N566 and K588 residues (ii) as a function of time (total time = 10 ns, 1 frame per 10 ps), during the TE (blue line), BMI (orange line) and BMID (grey line) simulations. The distances were estimated as described in Figure 3C. (iii) The horizontal bar chart shows the cumulated occupancy values for the H-bond interactions between the N125-K145 residues of TWIST1 (grey) and the N566-K588 residues of E12 (green) during the TE, BMI and BMID MD simulations. The cumulated occupancy values were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors (interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test.Logic; value.if.true;value.if.false) with logic test: <math>< 2.10^{\circ}</math>, value of 1 if true and 0 if false; NB.SI function: NB.SI(range;criterion)). Calculations were carried using Excel.

example the L124 (T1) and L589 (E12) residues) often under 2.2 Å compared to over 2.6 Å, respectively (Figure 3Cv). The proximity of these residues may largely contribute to the dimerization and to the stability of the TE dimer bound to these active E-boxes. In the second instance, two phenylalanine residues, namely F128 (T1) and F589 (E12) moved closer together during the BMID MD simulation compared to the TE and BMI MD simulations (Figure 3Cvi), inducing a decrease in the level of contact with the E-box bases localized at the heart of the TE complex. The F128 (T1) and F589 (E12) also moved towards adjacent bases of E-box sequences. Among the bHLH proteins, pairs of highly conserved asparagine (N) and lysine (K), namely N125-K145 (T1) and N566-K588 (E12) residues were suspected to contribute directly to the state of conformation of the heart of the TE complex bound to E-box sequences (Figure 3B and 3Di-iii). The H-bonds established between the K145 and N125 residues of TWIST1 and their K588 and N566 counterparts of E12 were largely disrupted in the BMID simulation compared to BMI and TE simulations (Figure 3Di-iii). Taken together, the hydrophobic interactions and H-bond establishments that drive alterations in the structure of the heart of the TE complex, are highly dependent on the E-box sequences used in our *in silico* models. Furthermore, although modifications arising from hydrophobic interactions seem to impede the binding of the dimer to the E-box sequences, H-bonds appear to play an essential role in the anchoring of the dimer to these boxes. We, therefore, investigated the effect of these bonds in the various models more thoroughly.

The cumulated rates of occupancy were calculated (as described in the Materials and Methods section), focusing on the 'non-specific' and 'specific' H-bond interactions (Figure 2D), in order to evaluate the overall stability of H-bond interactions between the TE complex and DNA during the *in silico* MD simulations. H-bond interactions were predominantly established between the dimer and E-box bases compared to adjacent bases, and these essentially occurred with conserved bases rather than with variable bases of the half-sites of the E-boxes (Figure 4A). However, while the cumulated occupancy values of H-bonds established with the different E-box bases were similar, this value was markedly higher between residues and adjacent bases in the BMI (+114%) and BMID (+83%) simulations compared to the TE MD simulations (Figure 4A). These findings indicate that there is an increase in H-bond interactions between the dimer and adjacent bases in the modified E-box sequences. Next, since the specificity of E-box recognition is likely due to these specific H-bond interactions, we focused on the specific H-bond binding affinities of the lateral chain residues either of TWIST1 or of E12. Of the H-bonds established between the dimer and E-box bases in the TE-model (Figure 4A), the specific H-bonds were principally attributed to TWIST1 residues (Figure 4Bi). Interestingly, few specific H-bond interactions were established with the adjacent bases in the TE simulation, indicating that they had a lower impact on the E-box recognition specificity. In contrast, of the elevated specific H-bond interactions observed between the dimer and adjacent bases in the BMI and BMID simulations (Figure 4A), the specific H-bond interactions were predominantly related to the E12 residues (Figure 4Bi).

Furthermore, these specific H-bond interactions mainly occurred with the flanking ((-1) and (-1*)) bases (Figure 4Bi and ii), representing 50% of the total specific H-bond interactions. Specific H-bond interactions were, overall, much lower in the non-functional BMID model than in the BMI MD simulation (Figure 4Bi). This suggests that adjacent E-box bases, play an important role in the binding affinity of the TE dimer to less functional E-box sequences, such as the BMI-box. Taken together, our data on the stability of specific H-bond interactions, suggest that the E-box composition largely participates in the selectivity of the TE binding, this may influence the panel of regulatory targeted genes for the various TWIST1 complexes.

Having addressed the importance of the E-box composition in the binding of the dimer, we focused our investigations on the exact mechanisms underlying this binding process. According to the literature (42), the first conserved Cytosine bases ((+1) and (+1*)) on each half-site of the E-box are important for the recognition of the E-box by a bHLH complex. Consequently, we studied the distribution of the specific H-bonds established between particular residues of the TE complex (two glutamic acids: E558 and E117, an arginine: R118, and an asparagine: N554) and bases of the E-box sequences (Figure 4Ci and Cii for better visualization). Specific H-bond occupancies with the first Cytosine bases ((+1) and (+1*)) of the half-sites of the E-box sequences represented 65%, 75% and 87% of the total occupancies in the TE, BMI, and BMID MD simulations, respectively (Figure 4Ci, Cii, Ciii). Interestingly, in both of the functional TE and BMI MD simulations, specific H-bonds were established between Cytosine (+1) and E558, while this was completely disrupted in the BMID simulation, where the Cytosine (+1) base was replaced by a Thymine. However, the specific H-bond established between base (+1*) and E117 was conserved in all of the simulations (Figure 4Ci-Ciii). Of note, in the TE simulation, numerous additional specific H-bond interactions were established between TE residues and the ((+3) and (+4*)) variable bases of the E-box sequence (represented in pink, Figure 4Ci and Di). In contrast, the H-bond interactions established with central variable bases were completely disrupted in BMI and BMID MD simulations (Figure 4Cii-iii and 4Dii-iii). Nevertheless, there was a clear shift towards the conserved base (+6*) in the BMI simulation (Figure 4Cii), and a slight shift towards base (+5*) during the BMID simulation (Figure 4Ciii and 4Diii). The 3D representations of the position of these H-bond interactions showed the binding of the TE-complex to different areas of the DNA groove (Figure 4Di-Diii). During the TE model simulation, these H-bond interactions occurred in four areas and acted as anchors allowing the specific recognition of the E-box sequence (Figure 4Di). Specific H-bond interactions with conserved central bases were partially lost, during the BMI and BMID model simulations, with only three (Figure 3Dii) and two (Figure 3Diii) areas of interactions remaining, respectively, resulting in a less stable anchoring of the dimer. To corroborate these *in silico* findings, we clearly observed *in vitro* that the TE complex was able to bind to the active TE-box and BMI-box sequences (Figure 5). Having demonstrated an impaired *in silico* binding during the BMID MD simulations, we expected a similar non-functional binding *in vitro*. As expected, a

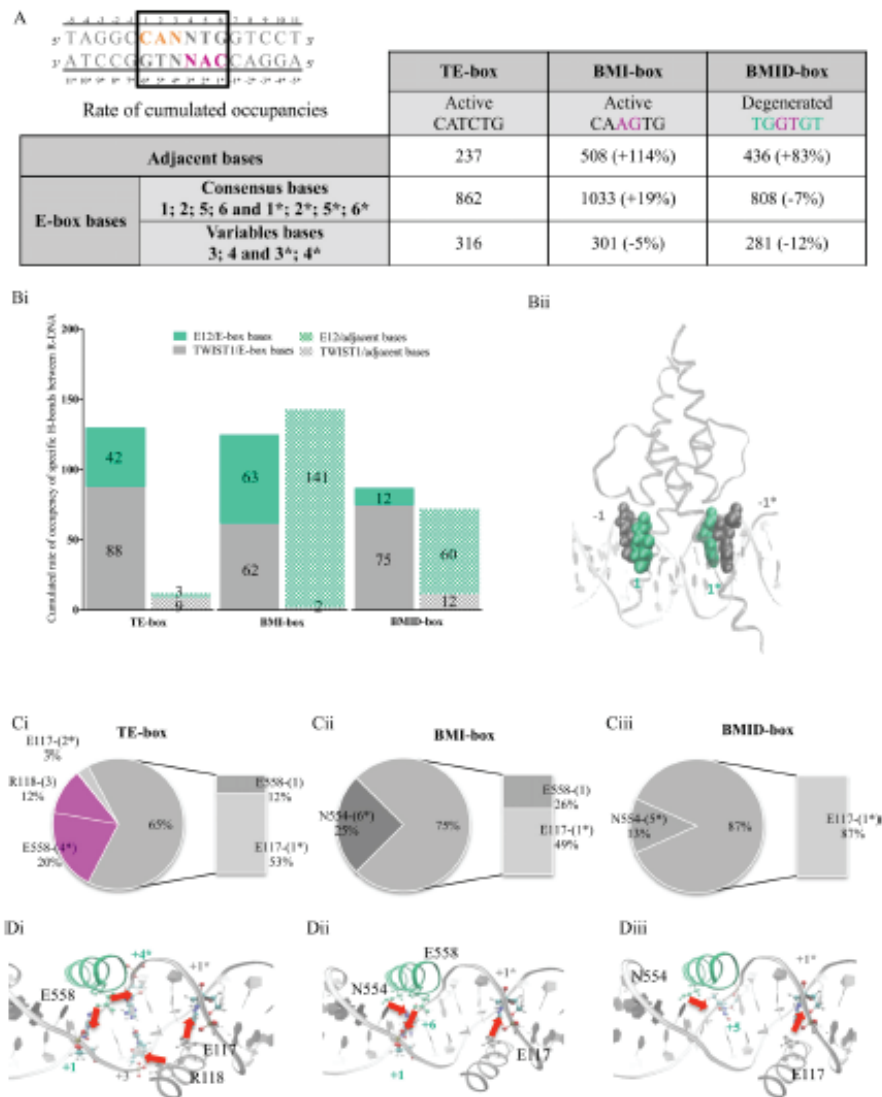


Figure 4. H-bond interactions between the TWIST1/E12 complex and active or inactive E-box sequences. (A–C) All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the time (the 10 ns of the MD) (interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test: "<2.10", value of 1 if true and 0 if false; NB.SI function: NB.SI(range; criterion)). The calculations were carried out using Excel. (A) Top, representation of the E-box sequence, with the forward half-site and the reverse complementary half-site in orange and pink, respectively. Table showing the rate of cumulated occupancy of H-bonds established between residues and DNA, during MD simulations of the TE complex binding to the active TE and BMI or inactive BMID E-box sequences. We focused on H-bonds established between the TE complex and either adjacent bases or E-box sequences (consensus and variable bases). The percentage of variation of the BMI- and BMID-boxes compared to the TE-box is reported in brackets. (B) (i) Bar chart representing the cumulated rate of occupancy of 'specific' H-bonds (as defined in Figure 2D) established during MD simulations either between E12 (green stippling) and TWIST1 (grey stippling) and adjacent bases of E-box sequences, or between E12 (solid green) and TWIST1 (solid grey) and E-box bases, in the case of the TE-, BMI- and BMID-box models. (ii) 3D *in silico* model showing the binding of the TE complex to DNA, carried out *via* the VMD 1.9.1 software. The proximal flanking ((-1) and (-1*)) bases and the first base of each half-site of the E-box are represented in grey and green VDW, respectively. (C) Distribution of 'specific' H-bonds established between residues and E-box bases in the (i) TE, (ii) BMI and (iii) BMID molecular dynamics simulations. The pie charts show percentages of the total rate of occupancy of specific H-bonds established during the MD simulation. Consensus bases and variable bases of the E-box sequences are represented in grey and pink, respectively. The bar chart highlights the specific residues binding to Cytosine ((+1) and (+1*)) bases, and shows their implication in the DNA binding affinity. (D) 3D representation of the TE complex within the DNA groove of the TE-, BMI- and BMID-boxes sequences. CPK representation of the glutamic acid E558 and asparagine N554 residues of E12 (shown in green), and the E117 and arginine R118 residues of TWIST1 (shown in grey). The red arrow points to the interactions of the TE complex with E-box sequences according to the DNA substitutions. (i) CPK representation of the Cytosine ((+1), (+1*)) and (+4*)) and Thymine (+3) bases of the TE-box, (ii) Cytosine ((+1) and (+1*)) and Guanine (+6*) of the BMI-box, and (iii) Adenine (+1*) and Cytosine (+5*) of the BMID-box.

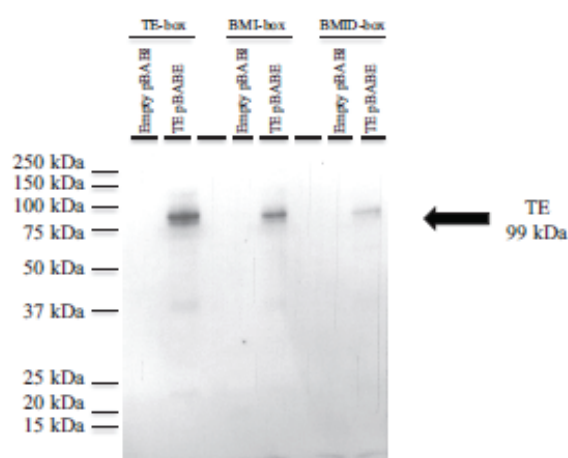


Figure 5. Streptavidin/Biotin assay of DNA bound to the TWIST1/E12 complex, in the case of active TE and BMI or inactive BMID E-box sequences. Streptavidin/Biotin assay using the TWIST1 antibody. Western blot showing the TE complex bound to various E-boxes, namely the TE-box, the BMI-box, and the BMID-box. The TE dimer was extracted from cells transfected with the TE pBABE vector. The empty pBABE vector was used as an assay control. The ectopic TE protein is visualized around 99 kDa, while a degraded peptide is observed at ~37 kDa. A detailed description of the assay is reported in the Materials and Methods section.

residual signal was observed with the inactive BMID-box sequence (Figure 5), which can likely be explained by the fact that bHLH complexes scan gene promoters to look for E-box sequences in the absence of stabilization of amphipathic helix conformations (46). No binding was observed in cells transfected with the empty pBABE-neo vector.

Deciphering of the key elements involved in the binding specificity of the TE complex to various functional E-boxes

Next, we determined whether the MD simulations accurately predicted the binding specificity of the TE dimer to various functional E-box sequences. We, thus, explored how the previously published TE complex (28) (Figure 1A) interacts with modified E-box sequences during the MD simulations. We substituted several bases in the DNA sequence of the TE-box (CATCTG), which was previously described by Firulli to bind with high affinity to heterodimers including to the TE complex (20), and compared the stabilities of the TE/modified E-box interactions (Figure 1A). The modified E-box sequences were as follows: the CACCTG sequence (named E12-box), described to bind to the E12 homodimer (24,35), the CATATG sequence (named TWIST-box) reported to bind to the TWIST1 homodimer (20,24,47) and finally the CGTCTG sequence (named D-box) known to bind to a lesser extent to the TE complex (20,48). We considered interactions with adjacent, conserved and variable bases of the E-box sequences separately.

Overall, we did not observe any major modifications in the positions of the leucine and phenylalanine residues during these MD simulations, suggesting that the establishment of hydrophobic interactions was not modified in the presence of modified E-box sequences (data not shown). In con-

trast, the stability of the H-bonds created between the asparagine and lysine N125-K145 residues (Figure 6Ai-Aiv; Bi-Bii) was completely disrupted in TE simulations bound to the E12- (Figure 6Aii and Bii), the TWIST- (Figure 6Aiii and Bii) and the D-box sequences (Figure 6Aiv and Bii). The H-bond interactions established between the N566-K588 of the E12 protein were conserved, irrespective of the E-box sequence chosen (Figure 6Ai-iv and Bii). These findings indicate that the heart of the dimer had a slightly altered stability according to the modified E-box sequence. To further explore the molecular mechanisms underlying this binding specificity, the overall stability of the H-bond interactions was then investigated during *in silico* MD simulations, focusing particularly on interactions occurring with all DNA bases. Total occupancy rates were calculated and revealed that the dimer preferentially bound to the conserved bases of the E-box sequences (Figure 6C). However, in the case of the modified E-box sequences, a shift was observed towards adjacent bases, in particular in the D-box with an increase in 90% of H-bond interactions with adjacent bases, confirming a decrease in the DNA binding specificity and the importance of E-box bases in the sequence recognition (Figure 6C) (20). We then identified the residues that created H-bonds with the E-box or with the adjacent ((-1), (+1), (-1*) and (+1*)) bases (Figure 6Dii-v). We observed that these four bases were close to the heart of the TE complex (Figure 6Di), and that H-bond interactions with these bases were evenly distributed between the TWIST1 and E12 proteins during the MD simulations, irrespective of the E-box model studied (Figure 6Dii-v). While the H-bond interactions between the adjacent (-1) base and the dimer represented only 2% of the total rate of occupancy during the MD simulations of the TE-box (Figure 6Dii), this percentage increased to 16%, 19% and 30% in the case of the E12- (Figure 6Diii), TWIST- (Figure 6Div) and D-box sequences (Figure 6Dv), respectively. The TWIST1 protein being the only one to interact with the adjacent (-1*) base of the TWIST-box, albeit at a very low rate (1%) (Figure 6Div). The shift of these H-bond establishments towards adjacent bases was concomitant with the experimental binding affinities of these E-box sequences (TE-box \ll E12-box \approx TWIST-box \ll D-box) (Figure 6Di-v) (20,24,47,48). The pivotal residues were the serine S144 on TWIST1 and the arginine R561 on E12 for the DNA binding function, since they are located at the vicinity of the proximal flanking ((-1) and (-1*)) bases (Figure 6Di). Indeed, S144 is localized at the bottom of the TWIST1 loops, while R561 is present in the C-alpha HLH domain of E12 (Figure 6Di). Overall, the stability of the TE dimer binding to modified E-box sequences is dependant of H-bond interactions with ((-1) and (-1*)) bases.

The analysis of the distribution of the specific H-bond interactions between all of the conserved or adjacent bases with the dimer, confirmed the predominant binding to the active E-box sequences compared to adjacent bases in all four models (Figure 7A). Surprisingly, the cumulated specific H-bond interactions in the D-box model were higher than the others (Figure 7A). In order to unravel this phenomenon, we studied the distribution of specific H-bonds with all of the bases. This binding was principally observed between the dimer and the conserved E-box bases with rates

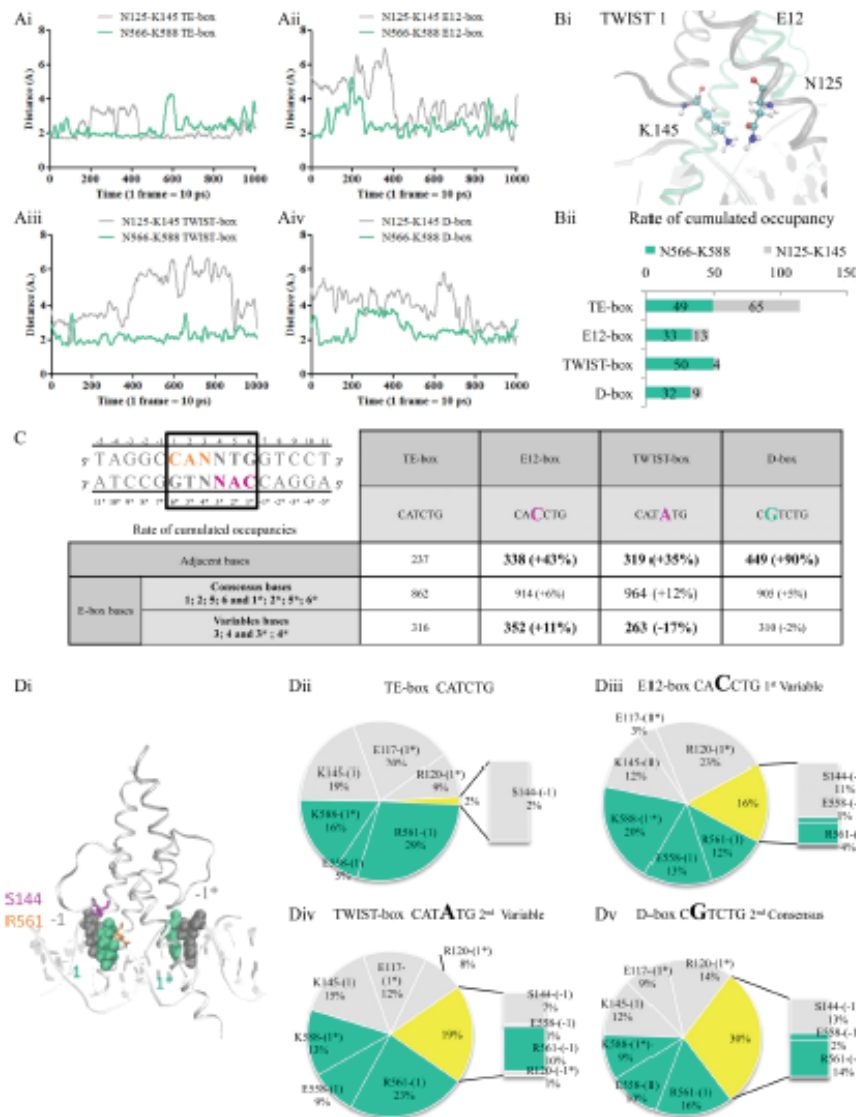


Figure 6. Binding of the TE complex to various active and modified E-box sequences, namely the E12-, TWIST- and D-boxes. (A) Plots represent the distances (Å) between the asparagine and lysine N125-K145 residues of TWIST1 and the N566-K588 residues of E12 as a function of time (total time = 10 ns, 1 frame per 10 ps) for the four (i) TE, (ii) E12, (iii) TWIST and (iv) D models during the MD simulations. These distances were estimated during the MD simulations using the VMD 1.9.1 software (Graph of Labels Bonds). The lengths of H-bond interactions were represented via GraphPad Prism 5: 'smooth, differentiate or integrate curve, with 8 neighbours'. (B and C) All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the time (the 10 ns of the MD)(interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test: '<2.10', value of 1 if true and 0 if false; NB.SI function: NB.SI(range;criterion)). The calculations were carried out using Excel. (B) (i) 3D representation of the TE dimer highlighting the position of the interaction between the asparagine and lysine residues N125-K145. (ii) The horizontal bar chart represents the rate of cumulated occupancy of the N125-K145 of TWIST1 and the N566-K588 of E12, in grey and green, respectively, for the four E-box sequences. (C) Top, representation of the E-box sequence, with the forward half-site and the reverse complementary half-site in orange and pink, respectively. Table showing the rate of cumulated occupancy of H-bonds established between the TE complex and either adjacent bases, or consensus and variable bases of TE-, E12-, TWIST-, and D-box sequences. The percentage of variation compared to the TE-box is reported in brackets. (D) (i) 3D *in silico* representation of the TE complex bound to DNA. The proximal flanking ((-1) and (-1*)) and first ((+1*) and (+1*)) bases of the E-box are represented in orange and grey VDW, respectively. The serine S144 and arginine R561 residues are highlighted in magenta and orange, respectively. (ii-v) The pie charts show the percentages of cumulated occupancy of H-bonds established between residues of TWIST1 (grey) and E12 (green) and DNA bases during the (ii) TE-, (iii) E12-, (iv) TWIST- and (v) D-box MD simulations. Occupancies of H-bonds established with the proximal flanking ((-1) and (-1*)) bases are represented in yellow, while the bar charts highlight the exact residues binding to those bases. The H-bonds established with the first consensus ((+1) and (+1*)) bases of the E-box are represented in grey and green for the TWIST1 and the E12 residues, respectively.

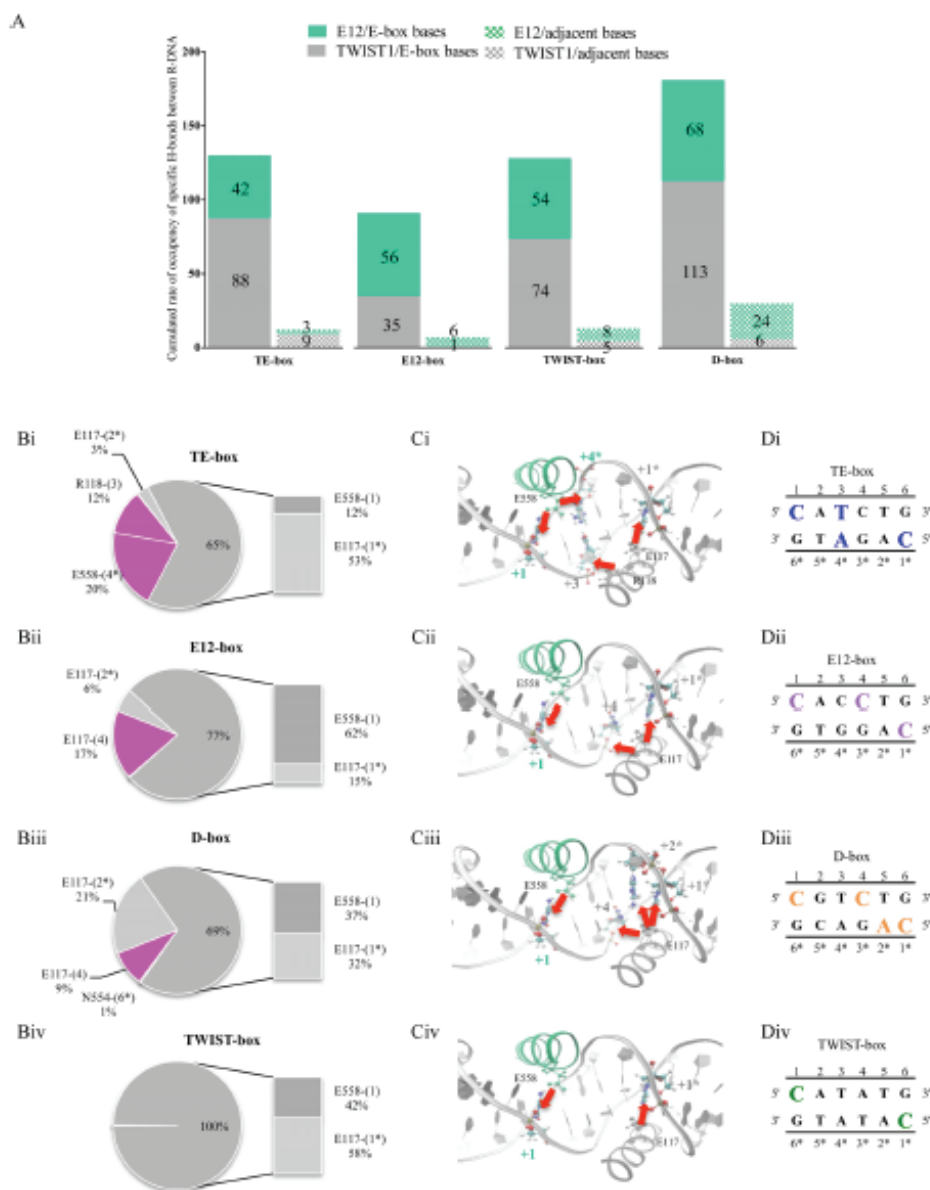


Figure 7. Predictive functional effect of base substitutions of E-box sequences on the H-bond establishment between the TE complex and DNA. **(A and B)** All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the time (the 10 ns of the MD)(interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test: '<2.10', value of 1 if true and 0 if false; NB.SI(range;criterion)). The calculations were carried out using Excel. **(A)** Bar chart showing the rate of cumulated occupancy of the 'specific' H-bonds (see Figure 2D) established between E12 (green) and TWIST1 (grey) and adjacent bases (stippling) or E-box sequences (solid colouring). The values are given for the four E-box models (TE-, E12-, TWIST-, and D-box sequences). **(B-D)** Pie-charts depicting the percentages of cumulated occupancy of 'specific' H-bonds established between specific residues of the TE complex and variable (pink) or conserved (grey) bases of the E-box sequences, during MD simulations of the (Bi) TE-, (Bii) E12-, (Biii) D- and (Biv) TWIST-models. The bar charts highlight the specific residue interactions with conserved Cytosine ((+1) and (+1*)) bases. The 3D representations show the positions of the TE complex in the DNA groove of the various (Ci) TE-, (Cii) E12-, (Ciii) D- and (Civ) TWIST-box sequences. The glutamic acid residue E558 of the E12 protein and the arginine R118 and E117 of the TWIST1 protein are represented in green and grey CPK shapes, respectively. The red arrows point to the major H-bond interactions of the TE complex with bases of the DNA groove, according to the DNA base substitutions listed opposite. Active bases in the (Di) TE-box: Cytosine ((+1) and (+1*)), Thymine (+3) and Adenine (+4*) bases (purple), (Dii) E12-box: Cytosine ((+1), (+4) and (+1*)) bases (pink), (Diii) D-box Cytosine ((+1), (+4) and (+1*)) and Adenine (+2*) bases (orange), and (Div) TWIST-box: Cytosine ((+1) and (+1*)) bases (green).

of cumulated occupancies of 65%, 77%, 69% and 100% in the case of the TE- (Figure 7Bi), E12- (Figure 7Bii), D- (Figure 7Biii) and TWIST-boxes (Figure 7Biv), respectively. The TE complex created close contacts and specific interactions with the four ((+1), (+1*), (+3) and (+4*)) bases in the TE-box (Figure 7Bi, Ci and Di). These interactions mainly relied on the binding of the glutamic acid residue E117 of TWIST1 and the glutamic acid residue E558 of E12 to the conserved and functional Cytosine ((+1*) and (+1)), driving the E-box recognition (Figure 7Bi-Biv), as well as the binding of the arginine residue R118 of TWIST1 and the glutamic acid residue E558 of E12 to the conserved ((+3) and (+4*)) bases (Figure 7Bi). These residues, therefore, contributed to the binding specificity of the dimer to the E-box sequence. In the modified models, the H-bond interactions established between the glutamic acid residue E117 of TWIST1 and the (+1*) base were partially preserved, while those between the (+4*) base and the glutamic acid residue E558 of E12 were completely lost (Figure 7B-D). Furthermore, in these models a clear shift towards the conserved (+1) base was observed, particularly in the TWIST-box simulation (Figure 7Biv, Civ, Div), which may explain the low specificity of the TE heterodimer for this box.

Interestingly, in the functional TE simulations, specific H-bonds established between Cytosine (+1) and E558, and between Cytosine (+1*) and E117, were conserved in all of the simulations (Figure 7Bi). Of note, in the TE-box and E12-box simulations, numerous additional specific H-bond interactions were established between TE residues and the (+3) and (+4*) variable bases of E-box sequences (represented in pink, Figure 7Bi and 7Biii). In contrast, the H-bond interactions established with central variable bases were partially disrupted in TWIST-box simulations (Figure 7Biv and Civ). There was, however, a clear shift towards the conserved base in the D-box and TWIST-box simulations (Figure 7Ciii). The 3D representations of the position of these H-bond interactions showed the binding of the TE-complex to different areas of the DNA groove (Figure 7Di-Div). During the TE MD simulations, these H-bond interactions occurred in four areas and acted as anchors allowing the specific recognition of the E-box sequence (Figure 7Di). Specific H-bond interactions with conserved central bases were partially lost, during the modified E12- and D-box simulations, with only three (Figure 7Dii) and two (Figure 7Div) areas of interactions remaining, respectively, resulting in a less stable anchoring of the dimer. The *in vitro* experiments conducted in parallel using the Streptavidin/Biotin binding assay, corroborated these *in silico* findings (Figure 8). Indeed, the binding of the TE complex to the TE-box probes was stronger than that obtained with E12-box, TWIST-box and D-box probes. Moreover, the residual signal observed when using the degenerate BMID-box probe confirmed that the modified boxes were truly active (Figure 8). This result was further confirmed by running a blot using the E12 antibody, instead of the TWIST1 antibody (Figure 11C). These findings highlight the functional importance of the variable bases of the E-boxes. The correlation obtained between the *in silico* prediction and the *in vitro* findings could lead to the screening of gene promoters in order to cluster targeted genes according to the E-box sequences they contain.

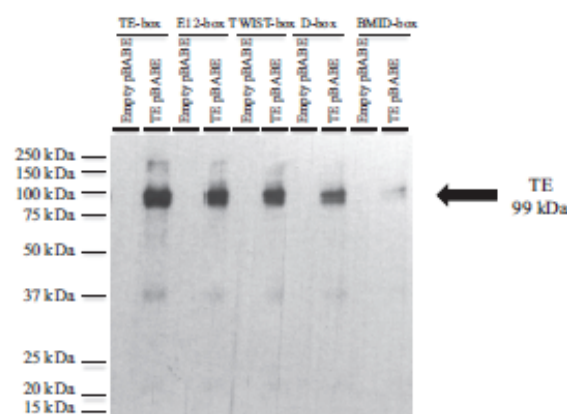


Figure 8. Streptavidin/Biotin assay of DNA bound to TE complex, in the case of modified E-box sequences. Streptavidin/Biotin assay using the TWIST1 antibody. Western blot showing the TE complex bound to various E-boxes, namely the TE-box, the E12-box, the D-box, the TWIST-box, as well as to the degenerate BMID-box. The TE dimer was extracted from cells transfected with the TE pBABE vector. The empty pBABE vector was used as an assay control. The ectopic TE protein is visualized around 99 kDa, while a degraded peptide is observed at ~37 kDa. A detailed description of the assay is reported in the Materials and Methods section.

Prediction of the contribution of the flanking proximal bases to the recognition of the E-box sequence

Having studied both the impact of using active *versus* inactive box-sequences and non-modified *versus* modified active box sequences on the binding of the TE complex, we next focused on the contribution of adjacent bases to its binding affinity. Indeed, following a study on the influence of proximal flanking sequences on E-box recognition by bHLH dimers (49), and in light of our own findings (see sections above), we investigated the contribution of the ((-1) and (-1*)) adjacent bases to the binding affinity of the TE-complex during MD simulations. We then exchanged the proximal flanking Cytosine (-1), Guanine (+7), Cytosine (-1*) and Guanine (+7*) bases by Adenine (-1), Thymine (+7), Adenine (-1*) and Thymine (+7*) in the AT-box, respectively, and proceeded to the opposite substitutions for the TA-box (Figure 1A, Figure 2Bii and Figure 9Aii).

The critical S144 (TWIST1) and R561 (E12) residues (described in the previous section) were localized close to the ((-1) and (-1*)) substituted bases (Figure 6Di and 9Aiii). In the modified TE-box sequences, and particularly in the case of the AT-box, base substitutions led to the presence of two aromatic rings, namely an Adenine (-1) and a Guanine (-1*), in these critical base locations (Figure 9Ai and Aii). These aromatic double rings have a higher molecular volume compared to the Thymine and Cytosine bases, which likely modifies the H-bond interactions at the interface of the DNA and the TE complex. Indeed, the aromatic rings on the Adenine ((-1) and (-1*)) bases strongly contributed to the stability of H-bond interactions with neighbouring S144 and R561 (Figure 9Bi and Biii) residues during the AT-box simulation. Furthermore, in the case of this model, of the 28% H-bond interactions occurring with adjacent bases, 3% were associated with the Adenine (-1*) base (Figure 9Ciii),

whereas these interactions were non-existent in the 10% and in the 2% of adjacent H-bond interactions established in the TA-box (Figure 9Cii) and TE-box simulations (Figure 9Ci), respectively. These interactions with the flanking Adenine ((-1) and (-1*)) bases likely contributed to stabilizing the H-bond interactions with adjacent sequences. They had a detrimental effect on the H-bond interaction between the N125-K145 residue pair of the TWIST1 protein that are at heart of the dimer, which may have resulted in a decrease in the stability and affinity of the TE complex for the TE-box (Figure 9Di and Diii). The H-bonds created between the asparagine and lysine N125-K145 residues were partially and completely disrupted in TA- and AT-box simulations, respectively, while the H-bond interactions occurring in the N566-K588 residue pair remained stable in all three models (Figure 9Di-iii). These findings indicate that the heart of the dimer had an altered stability according to the proximal flanking base included in the models. Overall our findings demonstrate that the stability of the H-bonds established between residues of the TE complex and adjacent ((-1) and (-1*)) bases is higher during the molecular dynamics of the TA- and AT-box simulations compared to the TE simulation, and that it strongly impacts the establishment of a stable H-bond interaction between the N125-K145 residue pair of TWIST1. In conclusion, to screen gene-targets of the TE complex, the flanking proximal bases should be taken into account, since they modulate the function of key residues, such as K145. The latter is reported as a mutation of TWIST1, linked to the SCS syndrome.

When examining the rate of cumulated occupancy of H-bond interactions between the dimer and E-box sequences, taking into account all of the adjacent bases and not only the ((-1) and (-1*)) bases, the AT- and TA-box simulations displayed higher rates of binding with adjacent bases, with 82% and 32%, respectively, compared to the TE-box (Figure 10A). Moreover, a decrease (by 21%) in the rate of cumulated occupancies of H-bond interactions with central bases of the TA-box was observed (Figure 10A). To identify key structural modifications in altered MD simulations, we then studied the specific H-bond interactions (as defined in Figure 2D) focusing on the variable and conserved bases, in particular ((+1) and (+1*)) bases (Figure 10Bi). While specific H-bond interactions with variable ((+3), (+4), (+3*)) and ((+4*)) bases represented 32% of all of the interactions with the TE-box sequence (Figure 10Bii), these were largely redistributed to the conserved ((+1) and (+1*)) bases in the AT- (Figure 9Biii) and TA-box simulations (Figure 10Biv). These data denote a decrease in the interactions with central bases in the TA-box simulation, which are highly involved in the specific recognition of the E-box by the TE complex. These findings were confirmed using the Streptavidin/Biotin assay, since the binding of TE complex to the TE-box probe was stronger than the binding obtained with AT-box and TA-box probes (Figure 10C). The altered TE binding to these modified E-boxes is in agreement with Chip experiments, in which authors observed a clear decrease in the binding of the yeast Tye7 and Cbfl bHLH to a functional E-box, by modifying the flanking proximal sequence (49). In conclusion, these findings confirmed the critical function of the flanking proximal bases in the binding of the TE complex to DNA. Furthermore, the

MD simulations were able to predict and identify the molecular mechanism of TE complex binding on these modified DNA sequences.

DISCUSSION

Our study provides an original proof that the *in silico* approach we used to determine the impact of base mutations on the binding of the oncogenic TE complex to E-box sequences, during MD simulations, could accurately predict the functional binding affinity of the dimer. We first addressed the molecular role of the arginine R118, serine S144 and lysine K145 residues of TWIST1. Several variants of the TWIST1 protein have been found in patients suffering from TWIST1 haploinsufficiency. This is the case for example in patients with SCS syndrome, where TWIST1 mutations occur in the R118H/Q/C (arginine118histidine/glutamine/cysteine), R120P (arginine120proline), S144R, and K145E/Q residues (30,32–34,50–52). These variations may modify the binding affinity of TWIST1 with the E-box sequences during the developmental process leading to SCS syndrome (32,53,54). A previous study using similar MD simulations focusing on the effect of structural changes between wild-type *versus* three alterations in the R118C, S144R and K145 residues, showed an increase in the instability of the mutated dimers (29). In the present work, we completed this study by showing that the H-bond interactions between the N125-K145 residue pair is an important element involved in stabilizing the structure of the heart of the TE dimer, enabling the establishment of specific H-bonds of R118 of TWIST1 and E558 of E12 with central variable bases of the E-box sequence. We also highlighted the crucial function of R118 and E558 in the recognition and binding specificity, since we observed that the higher the affinity between the TE complex and the E-box, the greater the specific H-bond interactions between R118 and E558 residues and central variable bases of the E-box (Figure 7). The third S144 residue, close to the hydroxyl group, is also involved in the anchoring of the TE dimer (i) to E-box sequences with low binding affinity or (ii) when the flanking proximal bases are modified. By doing so, it appears to compensate for the low binding specificity to such sequences. Overall, we showed that these residues are differentially involved in H-bond interactions and likely participate in the diverse functions of the TWIST1 complexes.

Our work clearly demonstrated that the MD simulations allowed us to differentiate active and inactive E-box sequences, and revealed the critical role of the consensus ((+1) and (+1*)) bases in the E-box recognition process. The affinity of TE complex for an E-box is partially defined by the specific H-bond interactions between residues and DNA bases. We determined their localization and uncovered their role in stabilizing the binding of TE complex to various active and modified E-box sequences during MD simulations. We found that specific H-bond interactions are principally established with the ((+1) or (+1*)) bases of the half-sites of the active E-box sequences. This was clearly reduced in the MD simulation of the inactive BMID model, confirming the pivotal function of these consensus Cytosine ((+1) and (+1*)) in E-box recognition by bHLH complexes. In-

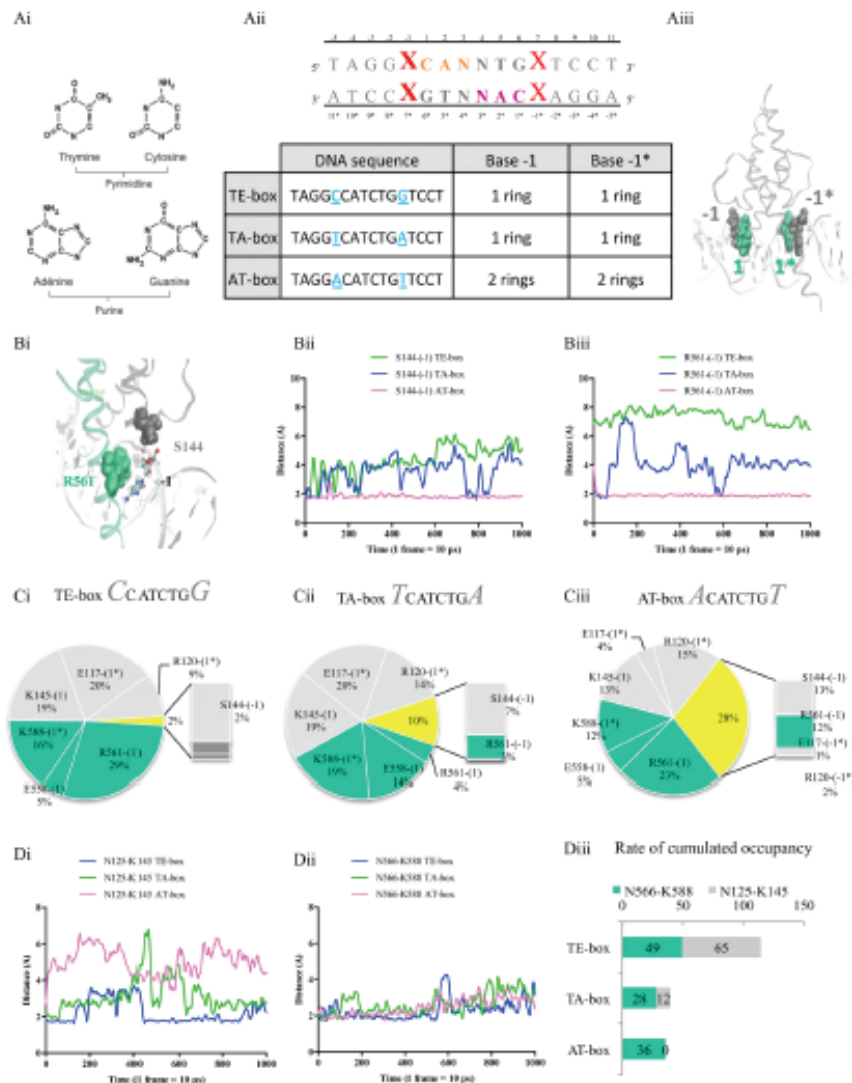


Figure 9. Predictive functional effect of the substitution of the proximal adjacent bases on the binding of the TE complex to DNA. (A) (i) Chemical structure of the purine and pyrimidine bases, showing the aromatic rings. (ii) The table depicts the base substitution generated in the TE-box, resulting in the TA- and AT-boxes. The latter substitutions led to the presence of two aromatic rings in the crucial ((-1) and (-1*)) base locations. (iii) The 3D representation of the TE complex (grey ribbon) in the DNA groove highlighting the location of the ((-1) and (-1*)) and ((+1) and (+1*)) base pairs in solid grey and green, respectively. (B) (i) *in silico* 3D model of the TWIST1 (grey ribbon)/E12 (green ribbon) dimer bound to DNA (white ribbon). Serine S144 and arginine R561 residues are shown in solid grey and green colouring, respectively, while CPK representation is used to highlight the first adjacent base of the E-box. (ii-iii) Plots representing the distances (Å) between residues and the (-1) base of the E-box sequences according to time (total time = 10 ns, 1 frame per 10 ps), focusing on (ii) S144 and (-1) base interactions, and (iii) R561 and (-1) base interactions, for all of the TE-box (in blue), TA-box (in green) and AT-box models (in purple). These distances were estimated during the MD simulations using the VMD 1.9.1 software (Graph of Labels Bonds). The lengths of H-bond interactions were represented *via* GraphPad Prism 5: 'smooth, differentiate or integrate curve, with 8 neighbours'. (C and D) All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the time (the 10 ns of the MD) (interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test: '< 2.10', value of 1 if true and 0 if false; NBSI function: NBSI(range;criterion)). The calculations were carried out using Excel. (C) Pie-charts representing the percentages of cumulated occupancy of H-bonds established between residues of the TWIST1 (grey) and E12 (green) proteins and ((-1), (-1*), (+1) and (+1*)) bases. Percentage occupancies of H-bonds established with ((-1) and (-1*)) bases are represented in yellow, the bar charts highlight the exact residues interacting with these bases, in the (i) TE-, (ii) TA-, and (iii) AT-models. (D) Plots representing the distances (Å) between (i) N125-K145 residues of TWIST1, and (ii) N566-K588 residues of E12 as a function of time (total time = 10 ns, 1 frame per 10 ps), during TE- (in blue), TA- (in green) and AT-box (in purple) simulations. These distances were estimated during the MD simulations using the VMD 1.9.1 software (Graph of Labels Bonds). The lengths of H-bond interactions were represented *via* GraphPad Prism 5: 'smooth, differentiate or integrate curve, with eight neighbours'. (iii) The horizontal bar chart shows the rate of cumulated occupancy of the interactions between the N125-K145 residue pairs of TWIST1 (in grey) and the N566-K588 residue pairs of E12 (in green) for each of the TE-, TA- and AT-box models.

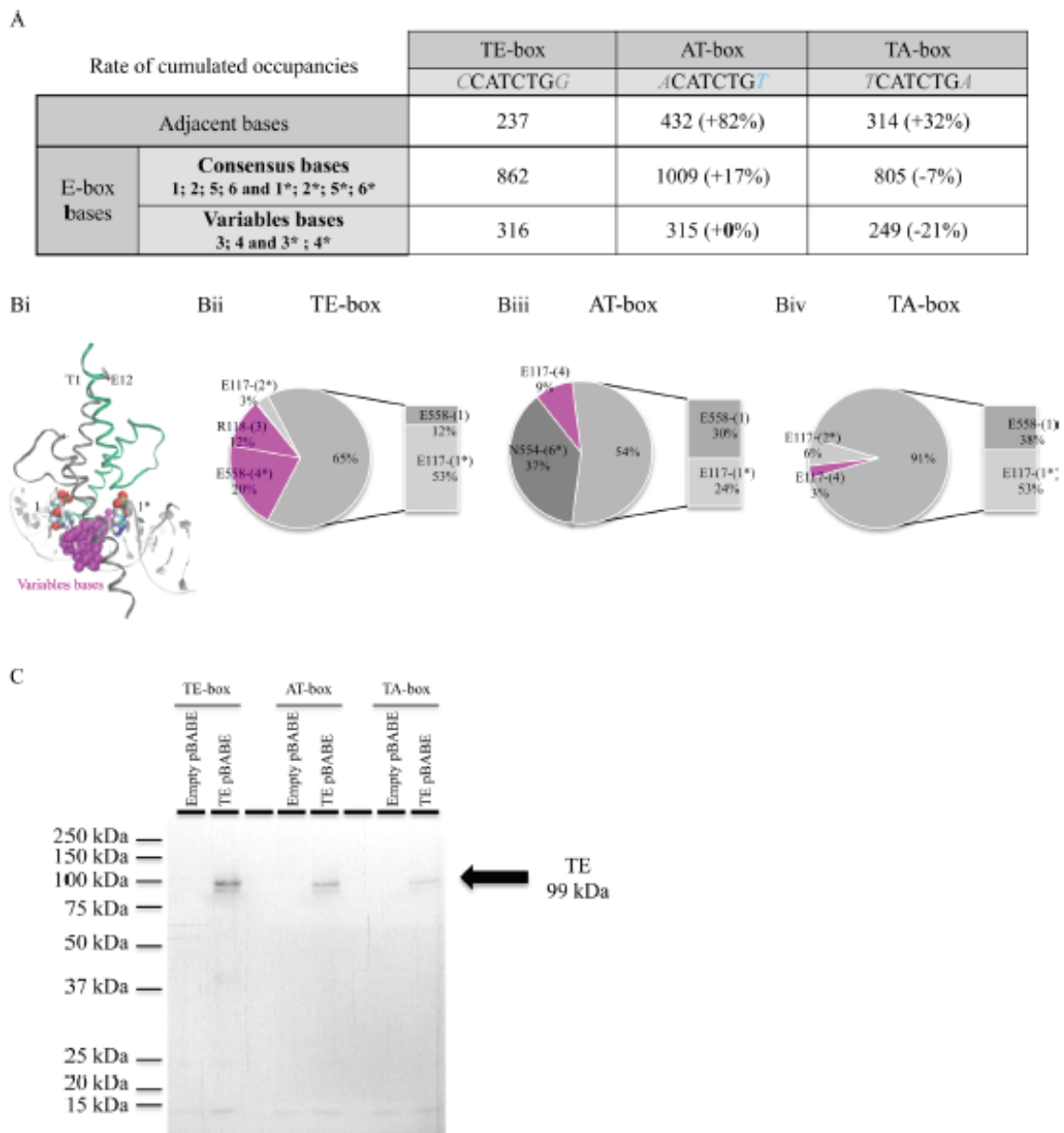


Figure 10. Prediction of the impact of proximal adjacent base substitutions on the specific interactions of the TE complex with E-box sequences. **(A and B)** All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the time (the 10 ns of the MD)(interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test: '<2.10', value of 1 if true and 0 if false; NB.SI function: NB.SI(range; criterion)). The calculations were carried out using Excel. **(A)** Table showing the rate of cumulated occupancy of H-bonds established between the TE complex and either adjacent bases, or consensus and variable bases of TE-, AT-, and TA-box sequences. The percentage of variation compared to TE-box is reported in brackets. **(B)** (i) 3D representation showing the position of the TE complex in the DNA groove. The TWIST1 and E12 dimers are represented in grey and green ribbons, respectively, while the variable base interactions, close to the ((+1) and (+1*)) bases, are shown in magenta CPK representations. (ii-iv) Pie-charts representing the percentage of cumulated occupancies of specific H-bonds established between the TE complex and either conserved (grey) or variable (magenta) bases. The bar charts highlight the exact residues interacting with the conserved ((+1) and (+1*)) bases, in the (ii) TE-, (iii) TA-, and (iv) AT-models. **(C)** Streptavidin/Biotin assay using the TWIST1 antibody. Western blot showing the TE complex bound to various E-boxes, namely the TE-box, the AT-box, and the TA-box. The TE dimer was extracted from cells transfected with the TE pBABE vector. The empty pBABE vector was used as an assay control. The weaker intensity of the signal was due to an unintentional shorter exposure-time of the radiographic film. The ectopic TE protein is visualized around 99 kDa, while a degraded peptide is observed at ~37 kDa. A detailed description of the assay is reported in the Materials and Methods section.

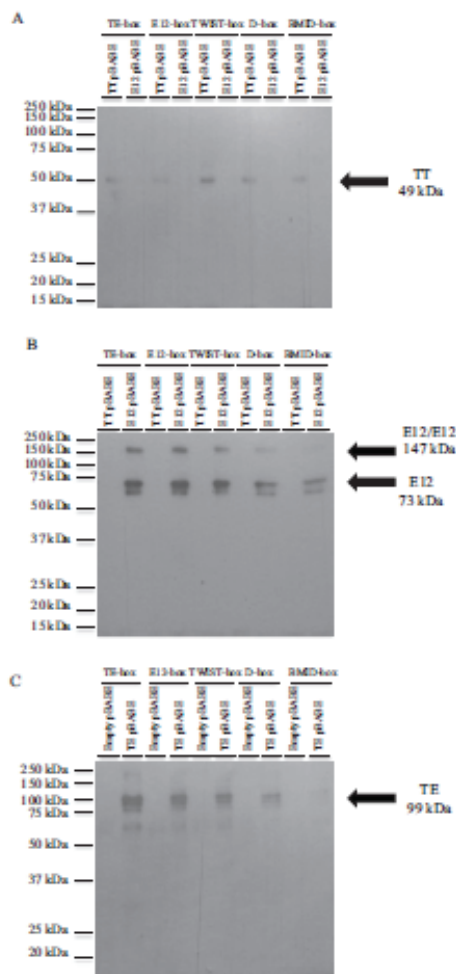


Figure 11. Streptavidin/Biotin assay of the tethered TWIST1/TWIST1 dimer or of the E12 protein. (A) Streptavidin/Biotin assay using the TWIST1 antibody. Western blot showing the TT and E12 complexes bound to various E-boxes, namely the TE-box, the E12-box, the TWIST-box, the D-box and the BMID-box. The proteins were extracted from cells transfected either with the TT pBABE vector or E12 pBABE vector. The ectopic TT protein is visualized around 49 kDa. The E12 protein was not detected using this antibody. A detailed description of the assay is reported in the Materials and Methods section. (B) Streptavidin/Biotin assay using the E12 antibody. Western blot showing the TT and E12 complexes bound to various E-boxes, namely the TE-box, the E12-box, the TWIST-box, the D-box and the BMID-box. The TT and E12/E12 dimers, as well as the E12 proteins were extracted from cells transfected either with the TT pBABE vector or E12 pBABE vector. The ectopic E12 protein is visualized around 73 kDa and the E12-E12 dimer appears at ~147 kDa. For the ectopic E12 monomer, we can observe a double band (E12 corresponds to the highest one). The TT dimer was not detected using this antibody. A detailed description of the assay is reported in the Materials and Methods section. (C) Streptavidin/Biotin assay using the E12 antibody. Western blot showing the TE complex bound to various E-box sequences, namely the TE-box, the E12-box, the TWIST-box, the D-box and the BMID-box. The TE dimer was extracted from cells transfected with the TE pBABE vector. The empty pBABE vector served as an assay control. The ectopic TE protein is visualized around 99 kDa as a double-band (TE correspond to the highest one), while a degraded peptide is observed at ~37 kDa. A detailed description of the assay is reported in the Materials and Methods section.

deed, the substitution of Cytosine to Thymine in BMID sequence was sufficient to impair the interaction between the E-box sequence and the bHLH complex (42,55) as shown by the loss of E558-(+1) specific H-bonds during the dynamic simulation of BMID simulation. These *in silico* findings were supported by *in vitro* data, in which the binding of the TE complex was higher in the TE- and BMI-boxes, while only a residual signal was observed in the negative control and in the BMID-box (Figure 5). Our predictions are in agreement with previously published biochemical findings (42), which demonstrated the ability of different TWIST1 complexes to bind various E-box sequences using EMSA assays, and measured the transactivation function of these complexes using reporter gene assays. These studies found that the higher the affinity between the TE-box and the TE complex, the stronger the binding (20,24,42,47,48,55). In the present study, this observation was verified for different complexes, since we also obtained a stronger binding between the TWIST1/TWIST1 homodimer and the TWIST-box using the TWIST1 antibody (Figure 11A), and between the E12/E12 homodimer and the E12-box using the E12 antibody (Figure 11B).

We also evaluated whether our MD simulations were able to predict the degree of affinity of E-box sequences, by studying the distribution of specific H-bonds established inside various E-box sequences with modified central or variable bases. We observed a shift in the specific interactions to conserved bases of the E-box and a decrease in the binding affinity of the TE-complex in Streptavidin/Biotin assays (TE-box \ll E12-box \approx TWIST-box \ll D-box). The pivotal function of the variable third base of the half-site of the E-box was highlighted, since the substitution of the complementary half-site CAG by the CAT sequence of TWIST-box abolished the specific H-bond interactions between the TE complex and the TWIST-box sequence. The half-site CGT of the D-box and CAC of the E12-box were, however, still recognized by TWIST1 (20,24). This is supported by the fact that the CAC half-site is specific to the E12 homodimers and its specificity is decreased in the TE heterodimers, while CAT and CAG half-sites are normally recognized with high affinity by the TE complex (24). These findings are in agreement with the *in vitro* Streptavidin/Biotin assays conducted with the TE heterodimer, and the TWIST1 and E12 homodimers (Figure 11). In the present study, we clearly highlighted a redistribution of the specific H-bonds during the simulations according to the biological binding affinity of the TE complex to a given E-box sequence, with central variable bases of the E-box largely contributing to the TE complex binding specificity. Overall, this *in silico* approach may provide a valuable predictive tool to select E-box sequences with flanking proximal sequences with a higher probability of binding affinity to the TE complex. Lastly, it will enable the screening of gene promoters for highly specific E-box sequences in order to establish a clustering of targeted gene sets. Such an approach may also be applicable to other proteins (or complexes) bound to their ligands. Furthermore, the identification of a specific interaction between the oncogenic TE complex and a distinct E-box sequence could lead to the determination of a putative pro-oncogenic signature.

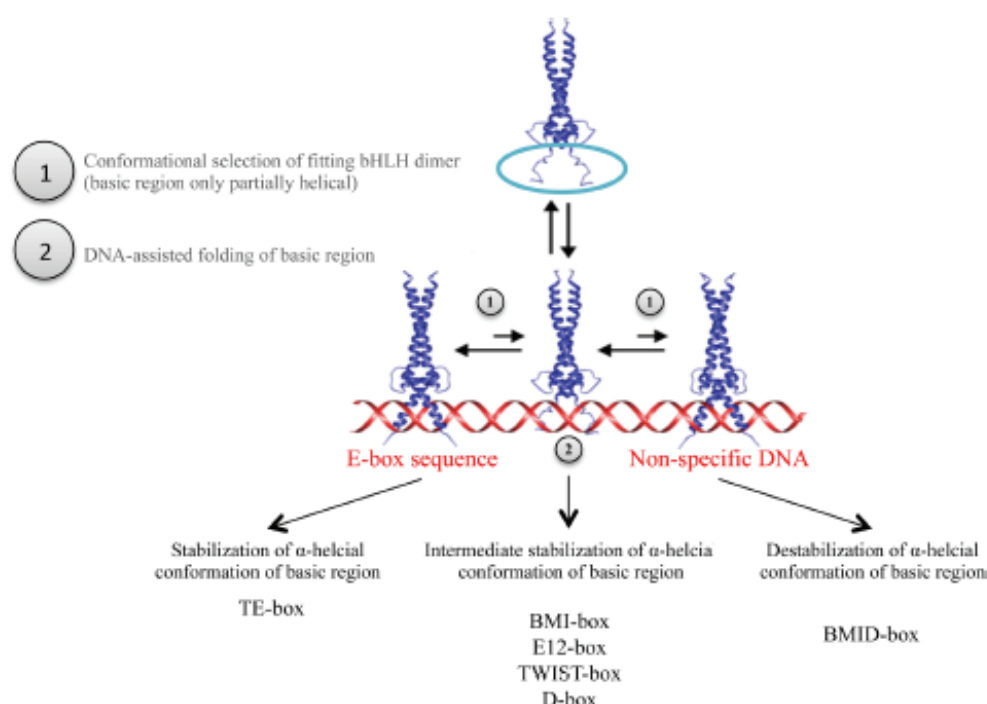


Figure 12. Synopsis of the three conformational states of the TE complex bound to the E-box sequences based on Sauve's model. Three protein structures are depicted on this diagram. The first state, represented by TE-box simulation, demonstrates the complete stabilization of the alpha-helical conformation of the basic region. The second state, represented by BMI-box, E12-box, TWIST-box and D-box simulations, shows the partial stabilization of the alpha-helical conformation of the basic region. The third state, represented by BMID-box simulation, reveals that there is no stabilization of the alpha-helical conformation of the basic region.

To summarize our findings and based on Sauve's model (46), we described three different states of binding of the bHLH complex to functional and degenerate E-box sequences based on molecular dynamics simulations. The first conformation is represented by the TE-box dynamics simulation, in which specific H-bond interactions are established with the variable bases of the E-box (Figure 12). This 'ideal' conformational state favours stability and complete transactivation of the TE complex. The second or 'intermediate' conformation includes all of the active, modified E-box sequences, which are not highly specific for the TE complex. Although, these sequences are recognized (Figure 7B) by the TE complex, the binding specificity and efficiency to these BMI-, E12-, D- and TWIST-boxes is not ideal (Figure 12). The third state is a non-specific conformation that enables the TE complex to scan DNA to look for E-box sequences. This state is, represented by the BMID simulation. In this context, only the specific E117-(+1*) interaction is preserved and a shift of the interactions towards adjacent bases of the E-box sequence occurs. The *in vitro* binding also decreases, since the binding affinity is partially impaired. This strongly suggests that there is an increase in the binding flexibility of the TE complex to DNA, allowing the proper promoter scanning function (Figure 12).

In conclusion, our study is a proof of concept of the predictive value of these molecular modeling approaches on deciphering molecular mechanisms underlying the binding of

the putative oncogenic TE form to E-box sequences. In the future, it may be possible to identify specific transcriptional signatures according to the TWIST1 heterodimers generated, for example TWIST1/E2A *versus* TWIST1/TWIST1 or TWIST1/HANDs, and to their E-box binding specificities. Furthermore, these validated *in silico* models of the TE complex may contribute to the characterization of specific inhibitors by docking approaches of the putative oncogenic form of TE complex targeting the heart of the dimer. This could be used in the future for developing and testing pharmacological TWIST1/E12 blocking molecules by *in silico* docking prior to *in vitro* or *in vivo* studies.

ACKNOWLEDGEMENTS

The pBABE-neo expression vector was donated by H. Land & J. Morgenstern & B. Weinberg (Addgene plasmid # 1767). Our team is accredited by the Ligue Nationale contre le Cancer.

FUNDING

LabEX DEVweCAN [ANR-10-LABX-61]; LyRIC (Lyon Recherche Intégrée en Cancérologie, Institut National contre le Cancer [INCa_4664]; Ligue Nationale contre le Cancer (to C.B.). Funding for open access charge: LabEX DEVweCAN [ANR-10-LABX-61]; LyRIC (Lyon Recherche

Intégrée en Cancérologie, Institut National contre le Cancer [INCa_4664]; Ligue Nationale contre le Cancer (to C.B.).
Conflict of interest statement. None declared.

REFERENCES

- Bialek, P., Kern, B., Yang, X., Schrock, M., Sasic, D., Hong, N., Wu, H., Yu, K., Ornitz, D.M., Olson, E.N. *et al.* (2004) A twist code determines the onset of osteoblast differentiation. *Dev. Cell*, **6**, 423–435.
- Seto, M.L., Hing, A.V., Chang, J., Hu, M., Kapp-Simon, K.A., Patel, P.K., Burton, B.K., Kane, A.A., Smyth, M.D., Hopper, R. *et al.* (2007) Isolated sagittal and coronal craniosynostosis associated with TWIST box mutations. *Am. J. Med. Genet. A*, **143**, 678–686.
- Massari, M.E. and Murru, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell Biol.*, **20**, 429–440.
- Teng, Y. and Li, X. (2014) The roles of HLH transcription factors in epithelial mesenchymal transition and multiple molecular mechanisms. *Cell. Exp. Metast.*, **31**, 367–377.
- Liu, T.M. and Lee, E.H. (2013) Transcriptional regulatory cascades in Runx2-dependent bone development. *Tissue Eng. B, Rev.*, **19**, 254–263.
- Kosty, J. and Vogel, T.W. (2015) Insights into the development of molecular therapies for craniosynostosis. *Neurosurg. Focus*, **38**, E2.
- Thierry, J.P., Aclouque, H., Huang, R.Y. and Nieto, M.A. (2009) Epithelial-mesenchymal transitions in development and disease. *Cell*, **139**, 871–890.
- Yang, J. and Weinberg, R.A. (2008) Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. Cell*, **14**, 818–829.
- Ansieau, S., Bastid, J., Doreau, A., Morel, A.P., Bouchet, B.P., Thomas, C., Fauvet, F., Puisieux, I., Doglioni, C., Piccinin, S. *et al.* (2008) Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell*, **14**, 79–89.
- Li, S., Kendall, S.E., Raices, R., Finlay, J., Covarrubias, M., Liu, Z., Lowe, G., Lin, Y.H., Teh, Y.H., Leigh, V. *et al.* (2012) TWIST1 associates with NF-kappaB subunit RELA via carboxyl-terminal WR domain to promote cell autonomous invasion through IL8 production. *BMC Biol.*, **10**, 73.
- Piccinin, S., Tonin, E., Sessa, S., Demontis, S., Rossi, S., Pecciarini, L., Zanatta, L., Pivetta, F., Grizzo, A., Sonogo, M. *et al.* (2012) A "twist box" code of p53 inactivation: twist box: p53 interaction promotes p53 degradation. *Cancer Cell*, **22**, 404–415.
- Kwok, W.K., Ling, M.T., Lee, T.W., Lau, T.C., Zhou, C., Zhang, X., Chua, C.W., Chan, K.W., Chan, F.L., Glackin, C. *et al.* (2005) Up-regulation of TWIST in prostate cancer and its implication as a therapeutic target. *Cancer Res.*, **65**, 5153–5162.
- Wushou, A., Hou, J., Zhao, Y.J. and Shao, Z.M. (2014) Twist-1 up-regulation in carcinoma correlates to poor survival. *Int. J. Mol. Sci.*, **15**, 21621–21630.
- Raatikainen, S., Aaltomaa, S., Palvimo, J.J., Karja, V. and Soini, Y. (2015) TWIST overexpression predicts biochemical recurrence-free survival in prostate cancer patients treated with radical prostatectomy. *Scand. J. Urol.*, **49**, 51–57.
- Ren, K.Q., Cao, X.Z., Liu, Z.H., Guo, H., Quan, M.F., Liu, F., Jiang, L., Xiang, H.L., Deng, X.Y. and Cao, J.G. (2013) 8-bromo-5-hydroxy-7-methoxychrysin targeting for inhibition of the properties of liver cancer stem cells by modulation of Twist signaling. *Int. J. Oncol.*, **43**, 1719–1729.
- Tran, P.T., Shroff, E.H., Burns, T.F., Thiyagarajan, S., Das, S.T., Zabuawala, T., Chen, J., Cho, Y.J., Luong, R., Tamayo, P. *et al.* (2012) Twist1 suppresses senescence programs and thereby accelerates and maintains mutant Kras-induced lung tumorigenesis. *PLoS Genet.*, **8**, e1002650.
- Seo, S.K., Kim, J.H., Choi, H.N., Choe, T.B., Hong, S.I., Yi, J.Y., Hwang, S.G., Lee, H.G., Lee, Y.H. and Park, J.C. (2014) Knockdown of TWIST1 enhances arsenite trioxide- and ionizing radiation-induced cell death in lung cancer cells by promoting mitochondrial dysfunction. *Biochem. Biophys. Res. Commun.*, **449**, 490–495.
- Murre, C., Bain, G., van Dijk, M.A., Engel, I., Furnari, B.A., Massari, M.E., Matthews, J.R., Quong, M.W., Rivera, R.R. and
- Stuiver, M.H. (1994) Structure and function of helix-loop-helix proteins. *Biochim. Biophys. Acta*, **1218**, 129–135.
- Castanon, I., Von Stetina, S., Kass, J. and Baylies, M.K. (2001) Dimerization partners determine the activity of the Twist bHLH protein during Drosophila mesoderm development. *Development*, **128**, 3145–3159.
- Firulli, B.A., Redick, B.A., Conway, S.J. and Firulli, A.B. (2007) Mutations within helix I of Twist1 result in distinct limb defects and variation of DNA binding affinities. *J. Biol. Chem.*, **282**, 27536–27546.
- Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M.A. and Spicer, D.B. (2006) Twist1 dimer selection regulates cranial suture patterning and fusion. *Dev. Dyn.*, **235**, 1345–1357.
- Gajula, R.P., Chettiari, S.T., Williams, R.D., Nugent, K., Kato, Y., Wang, H., Malek, R., Taparra, K., Cades, J., Annadanam, A. *et al.* (2015) Structure-function studies of the bHLH phosphorylation domain of TWIST1 in prostate cancer cells. *Neoplasia*, **17**, 16–31.
- Jones, S. (2004) An overview of the basic helix-loop-helix proteins. *Genome Biol.*, **5**, 226.
- De Masi, F., Grove, C.A., Vedenko, A., Alibes, A., Gisselbrecht, S.S., Serrano, L., Bulyk, M.L. and Walhout, A.J. (2011) Using a structural and logic systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res.*, **39**, 4553–4563.
- Fong, A.P., Yao, Z., Zhong, J.W., Cao, Y., Ruzzo, W.L., Gentleman, R.C. and Tapscott, S.J. (2012) Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev. Cell*, **22**, 721–735.
- Grove, C.A., De Masi, F., Barrasa, M.I., Newburger, D.E., Alkema, M.J., Bulyk, M.L. and Walhout, A.J. (2009) A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. *Cell*, **138**, 314–327.
- Chang, A.T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J.W., Paz, H., Wang, D., Li, H.R., Fu, X.D. *et al.* (2015) An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev.*, **29**, 603–616.
- Bouard, C., Terreux, R., Hope, J., Chemelle, J.A., Puisieux, A., Ansieau, S. and Payen, L. (2014) Interhelical loops within the bHLH domain are determinant in maintaining TWIST1-DNA complexes. *J. Biomol. Struct. Dyn.*, **32**, 226–241.
- Maia, A.M., da Silva, J.H., Mencalha, A.L., Caffarena, E.R. and Abdelhay, E. (2012) Computational modeling of the bHLH domain of the transcription factor TWIST1 and R118C, S144R and K145E mutants. *BMC Bioinformatics*, **13**, 184.
- Gripp, K.W., Kasparcova, V., McDonald-McGinn, D.M., Bhatt, S., Bartlett, S.P., Storm, A.L., Drumheller, T.C., Emanuel, B.S., Zackai, E.H. and Stolke, C.A. (2001) A diagnostic approach to identifying submicroscopic 7p21 deletions in Saethre-Chotzen syndrome: fluorescence in situ hybridization and dosage-sensitive Southern blot analysis. *Genet. Med.*, **3**, 102–108.
- Corsi, A.K., Brodigan, T.M., Jorgensen, E.M. and Krause, M. (2002) Characterization of a dominant negative C. elegans Twist mutant protein with implications for human Saethre-Chotzen syndrome. *Development*, **129**, 2761–2772.
- el Ghouzzi, V., Le Merrer, M., Perrin-Schmitt, F., Lajeunie, E., Benit, P., Renier, D., Bourgeois, P., Bolcato-Bellemin, A.L., Munnich, A. and Bonaventure, J. (1997) Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat. Genet.*, **15**, 42–46.
- el Ghouzzi, V., Lajeunie, E., Le Merrer, M., Cormier-Daire, V., Renier, D., Munnich, A. and Bonaventure, J. (1999) Mutations within or upstream of the basic helix-loop-helix domain of the TWIST gene are specific to Saethre-Chotzen syndrome. *Eur. J. Hum. Genet.: EJHG*, **7**, 27–33.
- El Ghouzzi, V., Legeai-Mallet, L., Benoist-Lasselin, C., Lajeunie, E., Renier, D., Munnich, A. and Bonaventure, J. (2001) Mutations in the basic domain and the loop-helix II junction of TWIST abolish DNA binding in Saethre-Chotzen syndrome. *FEBS Lett.*, **492**, 112–118.
- Longo, A., Guanga, G.P. and Rose, R.B. (2008) Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry*, **47**, 218–229.
- Nardella, C., Clohessy, J.G., Alimonti, A. and Pandolfi, P.P. (2011) Pro-senescence therapy for cancer treatment. *Nat. Rev. Cancer*, **11**, 503–511.
- Acosta, J.C. and Gil, J. (2012) Senescence: a new weapon for cancer therapy. *Trends Cell Biol.*, **22**, 211–219.

38. Valsesia-Wittmann, S., Magdeleine, M., Dupasquier, S., Garin, E., Jallas, A.C., Combaret, V., Krause, A., Leisner, P. and Puisieux, A. (2004) Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer Cell*, **6**, 625–630.
39. Morel, A.P., Hinkal, G.W., Thomas, C., Fauvet, F., Courtois-Cox, S., Wierinckx, A., Devouassoux-Shisheboran, M., Treilleux, J., Tissier, A., Gras, B. *et al.* (2012) EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS Genet.*, **8**, e1002723.
40. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
41. Zhang, Y., Hassan, M.Q., Li, Z.Y., Stein, J.L., Lian, J.B., van Wijnen, A.J. and Stein, G.S. (2008) Intricate gene regulatory networks of helix-loop-helix (HLH) proteins support regulation of bone-tissue related genes during osteoblast differentiation. *J. Cell. Biochem.*, **105**, 487–496.
42. Yang, M.H., Hsu, D.S., Wang, H.W., Wang, H.J., Lan, H.Y., Yang, W.H., Huang, C.H., Kao, S.Y., Tzeng, C.H., Tai, S.K. *et al.* (2010) Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat. Cell Biol.*, **12**, 982–992.
43. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
44. Slovak, J. and Tanaka, H. (2005) Computer simulation study of metastable ice VII and amorphous phases obtained by its melting. *J. Chem. Phys.*, **122**, 204512.
45. Song, X.-J., Rienstra, C.M. and McDermott, A.E. (2001) N—H bond stretching in histidine complexes: a solid-state NMR study. *Magn. Reson. Chem.*, **39**, S30–S36.
46. Sauve, S., Naud, J.F. and Lavigne, P. (2007) The mechanism of discrimination between cognate and non-specific DNA by dimeric b/HLH/LZ transcription factors. *J. Mol. Biol.*, **365**, 1163–1175.
47. Ozdemir, A., Fisher-Aylor, K.L., Pepke, S., Samanta, M., Dunipace, L., McCue, K., Zeng, L., Ogawa, N., Wold, B.J. and Stathopoulos, A. (2011) High resolution mapping of Twist to DNA in *Drosophila* embryos: Efficient functional analysis and evolutionary conservation. *Genome Res.*, **21**, 566–577.
48. Knöfler, M., Meinhardt, G., Bauer, S., Loregger, T., Vasicek, R., Bloor, D.J., Kimber, S.J. and Husslein, P. (2002) Human Hand1 basic helix-loop-helix (bHLH) protein: extra-embryonic expression pattern, interaction partners and identification of its transcriptional repressor domains. *Biochem. J.*, **361**, 641–651.
49. Gordan, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R. and Bulyk, M.L. (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.*, **3**, 1093–1104.
50. Johnson, D., Horsley, S.W., Moloney, D.M., Oldridge, M., Twigg, S.R., Walsh, S., Barrow, M., Njolstad, P.R., Kunz, J., Ashworth, G.J. *et al.* (1998) A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1. *Am. J. Hum. Genet.*, **63**, 1282–1293.
51. Rose, C.S., Patel, P., Reardon, W., Malcolm, S. and Winter, R.M. (1997) The TWIST gene, although not disrupted in Saethre-Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases. *Hum. Mol. Genet.*, **6**, 1369–1373.
52. Paznekas, W.A., Cunningham, M.L., Howard, T.D., Korf, B.R., Lipson, M.H., Grix, A.W., Feingold, M., Goldberg, R., Borochoy, Z., Aleck, K. *et al.* (1998) Genetic heterogeneity of Saethre-Chotzen syndrome, due to TWIST and FGFR mutations. *Am. J. Hum. Genet.*, **62**, 1370–1380.
53. Howard, T.D., Paznekas, W.A., Green, E.D., Chiang, L.C., Ma, N., Ortiz de Luna, R.I., Garcia Delgado, C., Gonzalez-Ramos, M., Kline, A.D. and Jabs, E.W. (1997) Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nat. Genet.*, **15**, 36–41.
54. Firulli, B.A., Krawchuk, D., Centonze, V.E., Vargesson, N., Virshup, D.M., Conway, S.J., Cserjesi, P., Laufer, E. and Firulli, A.B. (2005) Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nat. Genet.*, **37**, 373–381.
55. Ohno, K., Sadeh, M., Blatt, I., Brengman, J.M. and Engel, A.G. (2003) E-box mutations in the RAPSN promoter region in eight cases with congenital myasthenic syndrome. *Hum. Mol. Genet.*, **12**, 739–748.

Travail 4

« Destabilization of the TWIST1/E12 complex dimerization following the R154P point-mutation of TWIST1: an *in silico* approach »

Bouard C., Terreux R., Tissier A., Jacqueroud L., Vigneron A., Ansieau S., Puisieux A., Payen L.

Soumission Journal of Biomolecular Structure & Dynamics, mai 2016

5 Travail 4 : impact de la mutation R154P sur la dimérisation du complexe TWIST1/E12

Comme nous l'avons déjà vu précédemment, TWIST1 appartient à la famille des facteurs de transcription bHLH et joue un rôle clef au cours du développement embryonnaire et de la tumorigenèse. En effet certaines mutations, associées à une perte fonctionnelle de la protéine, du gène codant cette protéine au niveau du chromosome 7p21 sont responsables du syndrome de Saethre-Chotzen. Au cours de cette étude, nous nous sommes focalisés sur l'étude d'une mutation rare associée à ce syndrome. Il s'agit de la substitution rare du résidu d'arginine 154 par un résidu de proline au niveau de la séquence protéique de TWIST1, localisée dans le domaine HLH de TWIST1 au niveau de la seconde hélice α amphipatique de ce domaine. Pour cette étude, nous avons utilisé nos outils de dynamique moléculaire en construisant un modèle TWIST1_{R154P}/E12 (TWIST1 muté au niveau du résidu d'arginine 154 en proline et E12 sauvage) afin de mieux comprendre les mécanismes moléculaires sous-jacents à la perte fonctionnelle de TWIST1 associée à ce syndrome.

Au cours de notre approche *in silico*, nous avons étudié les capacités de dimérisation et de liaison à l'ADN en comparant notre nouveau modèle TWIST1_{R154P}/E12 à notre modèle hétérodimérique sauvage. De manière générale, on observe une distorsion de la structure de l'hétérodimère en présence de la mutation R154P avec une augmentation de l'écartement des deux hélices α amphipatiques inférieures de TWIST1 et de E12 contenant le domaine basique de liaisons à l'ADN de ces deux protéines, impliquant donc une modification majeure de l'affinité de liaison du dimère à l'ADN. Toutefois, comme la dimérisation précède la liaison du complexe aux séquences E-box de l'ADN, nous avons d'abord étudié la stabilité de dimérisation de notre modèle TE lors d'une simulation de MD sans tenir compte des séquences ADN.

Nous avons montré *in silico*, via l'étude de la stabilité des liaisons hydrogènes formées entre les résidus et entre les résidus et l'ADN au cours des différentes dynamiques moléculaires, que la mutation R154P induisait une diminution de la stabilité des hélices α amphipatiques de TWIST1 et de son partenaire E12. L'altération de la liaison hydrogène établie entre la paire de résidus N566-K588 de E12 montre une déstabilisation du cœur du dimère. Cette diminution de stabilité impacte notamment la stabilité des domaines de dimérisation localisés au niveau des points de contact entre TWIST1 et E12. Ainsi, cette mutation R154P conduit à une altération de la dimérisation de la protéine TWIST1 et entraîne donc une diminution de l'affinité de la liaison à une E-box de l'ADN. Ainsi, nos résultats *in silico* nous permettent de conclure que la mutation R154P induit une diminution de la dimérisation de TWIST1 mutée avec E12. Ces résultats ont été validés par des approches *in vitro* d'immunoprécipitation montrant donc une diminution de la dimérisation de TWIST1 avec E12 en présence de la mutation R154P de TWIST1. Comme la dimérisation est un prérequis à la liaison à l'ADN des complexes de TWIST1, une diminution de celle-ci impacte l'activité de transactivation induite par ces

complexes. Toutefois nous avons pu observer une activité de transactivation non nulle pour TWIST1_{R154P}/E12 (Jacqueroud et al., 2016). Cette diminution de transactivation montre probablement que la stabilité du complexe est diminuée (temps de demi-vie des complexes diminué).

Par conséquent, nous nous sommes donc intéressés à la capacité de liaison à l'ADN de ces différentes dynamiques moléculaires mutante ou sauvage. Ainsi, nous avons étudié l'affinité de liaison de nos dimères TWIST1/E12 R154P ou sauvages à une séquence E-box décrite comme étant de haute affinité pour l'hétérodimère TWIST1/E12 sauvage. Nous avons montré une forte augmentation de l'implication des bases adjacentes à l'E-box associées à une baisse des interactions spécifiques avec les bases variables (et donc responsable de la spécificité de liaison) de l'E-box en présence de la mutation R154P de TWIST1. Au-delà de l'altération de la dimérisation par la mutation R154P de TWIST1, la présence de la mutation affecte la spécificité de liaison de l'hétérodimère à l'ADN.

Ainsi les effets de la mutation R154P modifient, non seulement la fonction de dimérisation, mais aussi la liaison du complexe à l'ADN. Nos résultats *in silico* permettent donc de donner une explication structurelle à la perte fonctionnelle des mutations de TWIST1 associées au Syndrome de Saethre-Chotzen. De plus nos résultats sont en parfait accord avec les données bibliographiques et les résultats obtenus *in vitro* ce qui confirme une nouvelle fois la valeur prédictive des simulations de dynamiques moléculaires *in silico*. Ces approches de dynamiques moléculaires pourraient ainsi être utilisées pour prédire l'impact de mutations présentes au niveau du domaine bHLH de TWIST1 ou d'autres facteurs de transcription de la famille des bHLH compte tenu du haut degré d'homologie entre les différentes séquences bHLH de ces facteurs de transcription. De plus, elles pourraient également permettre d'étudier l'impact des mutations sur les capacités de dimérisation et de liaison à l'ADN de ces protéines. D'autre part, au vu de l'implication de TWIST1 dans les cancers, ces approches pourraient permettre de faciliter l'identification d'inhibiteurs spécifiques du complexe TWIST1/E12 afin de bloquer la fonction oncogénique de TWIST1.



Destabilization of the TWIST1/E12 complex dimerization following the R154P point-mutation of TWIST1: an in silico approach

Journal:	<i>PROTEINS: Structure, Function, and Bioinformatics</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Short Communication
Date Submitted by the Author:	n/a
Complete List of Authors:	Bouard, Charlotte; Centre de Recherche en Cancérologie de Lyon, équipe Alain Puisieux Terreux, Raphael; Institut de Biologie et Chimie des Proteines, Bio-informatique Tissier, Aqnes; Centre de Recherche en Cancérologie de Lyon, équipe Alain Puisieux Jacqueroud, Laurent; Centre de Recherche en Cancérologie de Lyon, équipe Alain Puisieux Vigneron, Arnaud; Centre de Recherche en Cancérologie de Lyon, équipe Alain Puisieux Ansieau, Stephane; Centre de Recherche en Cancérologie de Lyon, équipe Platteroti Puisieux, Alain; Centre de Recherche en Cancérologie de Lyon, équipe Alain Puisieux Payen, Ilea; Centre de Recherche en Cancérologie de Lyon, équipe Alain Puisieux
Key Words:	TWIST1/E12, Dimerization, bHLH Transcription Factor, molecular dynamics, in silico analysis

SCHOLARONE™
Manuscripts

Destabilization of the TWIST1/E12 complex dimerization following the R154P point-mutation of TWIST1: an *in silico* approach

Charlotte Bouard^{1,2,3,4,5}, Raphael Terreux^{6,8,9}, Agnès Tissier^{1,2,3,4,5}, Laurent Jacqueroūd^{1,2,3,4,5}, Arnaud Vigneron^{1,2,3,4,5}, Stéphane Ansieau^{1,2,3,4,5}, Alain Puisieux^{1,2,3,4,5,6,10}, Léa Payen^{1,2,3,4,5,6,7*}

¹Inserm UMR-S1052, Centre de Recherche en Cancérologie de Lyon, 69373, France,

²CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, 69373, France,

³LabEX DEVweCAN, Lyon, France,

⁴UNIV UMR1052, Lyon, 69008, France,

⁵Centre Léon Bérard, Lyon, 69373, France,

⁶Université de Lyon1, ISPB, Lyon, 69008, France,

⁷Laboratoire de Biochimie et Biologie Moléculaire (CHLS), Hospices Civils de Lyon, Lyon, 69003, France,

⁸Institut de Biochimie des protéines IBCP, Lyon, 69007, France,

⁹CNRS UMR 5086, Lyon, France,

¹⁰Institut Universitaire de France, Paris, 75231, France.

To whom correspondence should be addressed. Pr. Lea PAYEN Tel: 0 +33 478863208; Fax: 0 + 33 478782720; Email: Lea.payen-gay@univ-lyon1.fr

ABSTRACT

The bHLH transcription factor TWIST1 plays a key role in the embryonic development and in tumorigenesis. Some loss-of-function mutations of the *TWIST1* gene on chromosome 7p21 have been shown to cause an autosomal dominant craniosynostosis, known as Saethre-Chotzen syndrome (SCS). Although the biochemical impacts of many *TWIST1* mutations have been reported, little is known on the molecular mechanisms underlying their loss-of-function. In a previous study, we highlighted the predictive value of *in silico* molecular dynamics (MD) simulations in deciphering the molecular function of several TWIST1 residues in ensuring its dimerization and subsequent DNA binding activities. Here, based on these studies, we demonstrate the impact of a rare SCS mutation, namely the substitution of the arginine 154 amino acid by a proline residue (R154P) localized in the HLH domain of TWIST1, on the protein/protein interactions of TWIST1 and E12. Indeed, MD simulations highlighted a clear decrease of the α -helix stability during the dimerization of the mutated R154P TWIST1/E12 (TE) dimer compared to the wild-type TE complex. In conclusion, MD simulations provide a structural explanation for the loss-of-function associated with the SCS TWIST1 mutation and provide a proof of concept of the predictive value of these MD simulations. This *in silico* methodology could be used to determine reliable pharmacophore sites leading to the application of docking approaches in order to identify specific inhibitors of TWIST1 complexes. Ultimately, these putative inhibitors may be used to block the function of the oncogenic TWIST1/E12 complex in cancer.

KEY WORDS

TWIST1/E12; dimerization; bHLH transcription factor; molecular dynamics; *in silico* analysis.

INTRODUCTION

TWIST1 is a transcription factor belonging to the basic helix–loop–helix (bHLH) superfamily. During embryogenesis, it functions as an inducer of the epithelial-to-mesenchymal transition (EMT), a transdifferentiation process promoting the transient conversion of epithelial cells into

mesenchymal cells (Thiery, Acloque, Huang, & Nieto, 2009; Yang & Weinberg, 2008). The TWIST1 protein is highly conserved from *Drosophila* to humans, where it plays a key role during embryonic morphogenesis, mesoderm patterning and development. It is involved in cell-type determination and differentiation during myogenesis, cardiogenesis, neurogenesis, hematopoiesis

and osteogenesis (Anderson et al., 1997; Lister & Baron, 1998; Murray et al., 1992). The aberrant expression of TWIST1 has recently been implicated in cancer development, by fostering tumorigenesis and promoting the invasion-metastasis cascade (Ansieau et al., 2008; Gajula et al., 2015; Yang & Weinberg, 2008).

One of the major functions of proteins belonging to the bHLH superfamily, which includes TWIST1, TWIST2, E12, E47, HAND1 and HAND2 (Murre et al., 1994; Teng & Li, 2014), is their interaction, once dimerized, with E-box sequences (CANNTG). TWIST1 proteins can form either homodimeric (TWIST1/TWIST1) or heterodimeric complexes, mainly in association with E2A proteins. This dimerization is a prerequisite for the recognition of the E-box sequences, and involves basic domains and key residues of the bHLH domains (Bouard et al., 2016; Massari & Murre, 2000). These complexes can bind to numerous variable canonical core recognition E-box sequences with different levels of affinity, and display distinct and sometimes even antagonistic functions during the embryonic development (Connerney et al., 2008; Connerney et al., 2006; Firulli, Redick, Conway, & Firulli, 2007; Jones, 2004; Simionato et al., 2007). The differences in the composition of the E-box sequences, in effect, largely modulate the *in vivo* function of cellular transactivation by closely related transcription factors (Bouard et al., 2016; Chang et al., 2015; De Masi et al., 2011; Fong et al., 2012; Grove et al., 2009). Very recently, TWIST1 was shown to recognize a unique double E-box motif, constituted of two sequentially positioned E-boxes interspaced by five nucleotides, providing a highly specific DNA binding site for the TWIST1/E47 complex, thus resulting in a highly stable intramolecular tetramer. This specific tetramerization was primarily due to the formation of molecular interactions arising from the TWIST-box domains (WR domain) of the TWIST1 proteins (Chang et al., 2015).

The Saethre-Chotzen syndrome (SCS) is an autosomal dominant craniosynostosis characterized by facial and limb deformities caused by loss-of-function mutations of the *TWIST1* gene on chromosome 7p21 (Kress et al., 2006). Over 160 mutations of the *TWIST1*

gene have been described in SCS patients, a majority of which are present in the bHLH domains mediating protein dimerization, and involve a single base pair substitution (54% of mutations). They either create premature termination codons, which lead to truncated proteins, or substitute highly conserved residues in the bHLH region. These mutations affect the dimerization ability and the DNA-binding properties of TWIST1. Deletions (25%) and insertions (15%) of nucleotides have also been reported in SCS patients. One such mutation, termed R154P, is the substitution of an arginine 154 amino acid by a proline residue on the TWIST1 protein, and is very rare in SCS patients. It was previously suggested that this substitution disrupted the dimerization of TWIST1 and E2A (Spicer, Rhee, Cheung, & Lassar, 1996), although the molecular mechanisms underlying this disruption remain unclear.

In the present study, these mechanisms were investigated by comparing the dimerizing ability of the wild-type TWIST1/E12 (TE) complex, with that of the mutated R154P TE complex. This study was based both on the use of the recently published structural model of the TE dimer (Bouard et al., 2016; Bouard et al., 2014), initially generated using the murine NEUROD1/E47 crystal structure (Bouard et al., 2014; Longo, Guanga, & Rose, 2008), and on the previously described molecular dynamics (MD) simulation approach (Bouard et al., 2016; Bouard et al., 2014; Maia, da Silva, Mencialha, Caffarena, & Abdelhay, 2012). Indeed, this approach enabled the study of the effects of molecular modifications (insertions or single base mutations) in TWIST1 on the binding of the TE complex to DNA (Bouard et al., 2016; Bouard et al., 2014; Maia et al., 2012). It also provided a structural explanation for the loss-of-function associated with TWIST1-haploinsufficiency observed in SCS patients (Corsi, Brodigan, Jorgensen, & Krause, 2002; Gripp et al., 2001). Here, we generated the *in silico* R154P mutated TE complex and verified whether it bound to functional E-box sequence of targeted gene promoters using MD simulations, as previously described (Bouard et al., 2016; Bouard et al., 2014). A positive outcome of our *in silico* approach, combined with the functional *in vitro* assays conducted on cancer cells (Jacqueroud et al., 2016), would provide a

proof of concept of the predictive value of MD simulations in deciphering molecular mechanisms involved in protein dimerization and E-box-binding. These MD strategies may be used for the design of inhibitors of pro-oncogenic dimers via docking strategies.

MATERIALS AND METHODS

Selection of X-ray structure and in silico model building

The conserved TWIST1/E12 (TE) dimer model used throughout this study was previously generated by homology with the NeuroD1/E47 crystal structure, using a DNA sequence containing an E-box 5'-TAGCCATCTGGTCCT-3' (Bouard et al., 2016; Bouard et al., 2014; Longo et al., 2008). The R154P mutated TE complex was constructed by substituting the arginine 154 residue by a proline residue, using the mutation function of the Sybyl-X 1.1 software package. The resulting *in silico* models studied herein were, therefore, the TE complex (original R154 residue) bound or not to the E-box, sequence as well as the R154P mutated (P154 residue instead of R154) TE complex bound or not to the E-box sequence.

Molecular dynamics simulations

Prior to conducting molecular dynamics (MD) simulations, exhaustively described previously (Bouard et al., 2016; Bouard et al., 2014), minimization of all of the *in silico* models was carried out using the Sybyl-X 1.1 software package, elaborated by the Tripos Company. In the first instance, the effect of the R154P mutation of TWIST1 on the molecular interactions of the TE dimer with the E-box sequence was investigated in comparison with the effect of the WT TWIST1 protein, over a 10 ns experimental time-course. In the second instance, the impact of the anchor provided by the DNA sequence (E-box) on the binding efficacy of the WT or R154P mutated TE dimer was studied by conducting MD simulations over a time course of 10 ns.

Analysis of the stability of the interactions arising during MD simulations

Among the interactions occurring between residue-residue and residue-base, we focused our current investigation on hydrogen

bond (H-bond) interactions established between donor and acceptor groups. The H-bond interaction is an intermolecular force involving a hydrogen atom and an electronegative atom, such as oxygen or nitrogen. H-bonds can be established between donor atoms (arginine, lysine and tryptophan) and acceptor atoms (aspartic acid, glutamic acid) of side chain residues. We studied the residue-residue or residue-base interactions, which are established between atoms (O-H...N; O-H...O; N-H...N; N-H...O). The lengths of the H-bond interactions were represented via GraphPad Prism 5: “smooth, differentiate or integrate curve, with 8 neighbours”. These lengths were previously estimated around 2-2.8 Å depending on the nature of the atoms (Song, Rienstra, & McDermott, 2001). Using a computer software, we set an arbitrary threshold at 2.1 Å and made the following hypothesis: any interaction occurring under this threshold was deemed “true” and assigned a score of 1, whereas any interaction occurring above this threshold was deemed “false” and given a score of 0 (SI function: SI(test_logic; value_if_true;value_if_false) with logic test:“<2.1”, value of 1 attributed if true and 0 if false; NB.SI function: NB.SI(range;criterion)). The scores were reckoned and the value obtained was designated as the rate of occupancy of H-bond interactions. Thus, for a given H-bond, the higher the rate of occupancy, the greater the number of interactions occurring under 2.1 Å, the higher the level of stability of that H-bond during the MD simulation.

RESULTS

To characterize the molecular mechanisms involved in the dimerization of the TE complex (Bouard et al., 2014), we submitted the four *in silico* models (TE model with and without DNA, R154P model with and without DNA) to 10 ns MD simulations. Since protein dimerization precedes the binding of the complex to DNA sequences (Ellenberger, Fass, Arnaud, & Harrison, 1994), we initially studied the dimerizing stability of our TE model during a DNA sequence-free MD simulation. We clearly observed that the presence of DNA stabilizes the dimerization of the TE complex, possibly by limiting the

movement of the basic domains within the DNA groove, during the MD simulation (Figure 1Ai-Aii). Indeed, the evolution of the distances between the backbone methionine (M112) and valine (V552) residues of TWIST1 and E12, respectively, revealed that the interactions were more stable in the presence, rather than in the absence of DNA (Figure 1Ai-Aii). Since such a molecular instability in the DNA-free model may have overweighed the structural impact of the R154P punctual mutation, or led to a biological artefact, all of the following MD simulations will include a specific E-box sequence (Bouard et al., 2016).

We next evaluated the predictive value of our *in silico* strategy in assessing the functional disruption generated by the R154P mutation of TWIST1 on protein dimerization and DNA binding. Indeed, we integrated the R154P mutation into the recently established *in silico* structural model of the bHLH domains of the heterodimeric TE complex bound to its DNA target sequence (E-box) (Bouard et al., 2016; Bouard et al., 2014). We then conducted MD simulations focusing on alterations occurring in the top, middle and bottom regions of TE complex (Figure 1Bi). A clear distortion of the TE structure with the R154P mutation was observed, as indicated by the variation in distances between the TWIST1 and E12 backbone residues. These distances either increased, namely between the M112 and V552 residues situated at the bottom of the dimer (Figure 1Bii), and between the leucine (L124) and the aspartic acid (D564) residues of TWIST1 and E12 located in the middle of the dimer (Figure 1Biii), or decreased, namely between the isoleucine (I134) and L578 residues of TWIST1 and E12, found at the top of the dimer (Figure 1Biv). These results strongly suggest that the TE basic domains are further apart from the E-box sequence (positioned at the lower half of the dimer) during the mutated MD simulation compared to the TE simulation, thus, impeding proper binding (Figure 1Bii-Biii).

Next, we examined the consequences of the R154P mutation on a group of residues found on the α -helix domains of TWIST1 and E12, which are directly involved in dimerization and DNA binding, by assessing the stability of H-bonds established between residues, and between residues and DNA

during MD simulations (Figure 1Ciii). We studied four dimerizing domains along the mutated complex, named A and B boxes for TWIST1, and C and D boxes for E12 (Figure 1Ci-Cii), with A and D boxes and the B and C boxes facing each other, respectively (Figure 1Ci-Ciii). As expected, the R154P mutation impacted the most the stability of H-bonds established by amino-acid residues situated within the B box, where the R154 residue resides, with a 41% decrease in H-bonds establishment, regardless of the DNA interactions (Figure 1Ciii). When evaluating the stability of H-bonds exclusively established between residues (excluding interactions with DNA bases), we observed that these were also impaired in the A box of TWIST1 (-22%) and D box of E12 (-11%). This clear decrease in the stability of these boxes (excluding interactions with DNA bases), when comparing the TE and R154P variant of the TE MD model, confirms that DNA acts as an anchor for the C- α chain of HLH domains of TWIST1 (Figure 1Ci-Ciii, near here).

The destabilizing effect of the R154P substitution on TWIST1 was further explored, by studying the difference between the cumulated interactions of each residue within the A and B boxes of the wild-type and mutated TWIST1, during MD simulations. These simulations focused exclusively on the A and B boxes, since these domains were reported to experience major alterations (Figure 1 Ciii). This strategy revealed that the impact of the R154P mutation is distributed throughout the α -helix. Moreover, the residues that presented the highest rate of variation in their H-bond interactions, such as threonine (T121), glutamine (Q122), and asparagine (N125) within the A box (Figure 1Di), and leucine (L149), lysine (K150), and alanine (A152) within the B box (Figure 1Dii), were considered to be directly implicated in the stabilization of the complex. Furthermore, critical H-bonds localized in the heart of the dimer were disrupted, namely between the N125-K145 residue pair of TWIST1 and the N566-K588 residue pair of E12 in the R154P TWIST1 variant (Figure 2Ai-Aiii). This disruption occurs between the OD1 atom of the asparagine (N125 or N566) residues and either one of the three NZ1, NZ2 or NZ3 atoms of the lysine (K145 or K588) residues. Indeed, the interaction between the N566-K588 residue

pair of E12 was largely impaired, revealing the clear impact of the TWIST1 mutation on the structure of heart of the TE complex, which is at the interface of the functional E-box recognition (respectively Figure 1Di-Dii and Figure 2Ai-Aiii). Taken together, these results corroborate findings of Spicer et al. (Spicer et al., 1996), who suggested that the SCS mutation R154P (helix I) led to a decrease in dimerization. This, in turn, may affect the quantity of functional TE complexes present in cells, explaining the intermediate loss of luciferase activity, as well as the lack of detectable TE dimers in immunoprecipitation assays, following the R154P mutation compared to the TE complex (Jacqueroud et al., 2016). Consequently, we explored the binding affinity of the wild-type or mutated TE complexes, to investigate how such a mutation may affect the canonical E-box recognition. Based on recent findings (Bouard et al., 2016), we studied the particular involvement of the (1), (1*), (-1) and (-1*) DNA bases in the establishment of H-bond interactions between the wild-type or R154P mutated TE complex and the E-box sequence (Figure 2Bi). We noted an important increase (+30%) in the implication of (-1) and (-1*) adjacent bases in the R154P variant (Figure 2Bii-Biii). Moreover, this increase involved residues, namely serine (S144), R561 and R120, which were shown to play a pivotal role in the specific DNA binding affinity of the TE complex (Bouard et al., 2016; Bouard et al., 2014). In conclusion, these data strongly indicate that the R154P mutant, similarly to the TE complex binding to modified E-boxes (Bouard et al., 2016), establishes alternative H-bond interactions with adjacent bases in our *in silico* models in an attempt to compensate for the decrease in structuration of the heterodimer. Our previous findings demonstrated that the specific affinity of the TE complex to functional E-box sequences was associated with specific H-bond interactions established with consensus bases (+1), (+2), (+2*) and (+1*), as well as variable central bases (+3), (+4), (+3*) and (+4*) of the E-box (Bouard et al., 2016). Here, we studied their distribution in the current MD simulations (Figure 2C). We clearly noted that specific H-bond interactions were lost in the case of the mutated R154P TE complex, especially with variable central bases (Figure 2Ci-Civ, near here). This decrease was largely compensated

by an increase (+26%) in specific H-bond interactions with consensus bases (+1) and (+1*) of the E-box sequences in both cases (Figure 2Ci-Civ). The R154P mutation not only affects the dimerization function but also the DNA binding function of the TE complex. Collectively, our results demonstrate that the mutated R154P TE complex is not able to recognize or to bind to active E-box sequences with high affinity. Thus, the present *in silico* strategy is able to predict the impact of mutations occurring within the HLH domains of TWIST1, and possibly of E2A proteins, on the dimerization of the proteins and on DNA binding.

DISCUSSION

Our study provides a yet another rationale for using molecular dynamics (MD) simulations as a predictive tool in the context of protein interactions, dimerization and DNA-binding, and in determining the impact of protein point-mutations on the stability of the HLH domains of the oncogenic TE complex (Jacqueroud et al., 2016). Using a TWIST1 protein displaying the R154P mutation, which is one of the variants found in patients suffering from TWIST1 haploinsufficiency, we accurately predicted both an impairment in the protein dimerizing function and a subsequent decrease in the E-box binding affinity. We highlighted the determinant role of the interhelical loops in maintaining the TWIST1-DNA complex structure, by studying two variants with a 21-bp tandem repeat insertion in the TWIST1 gene, leading to the aberrant presence of seven extra amino-acids on positions 135 and 139 of the interhelical loop in SCS patients (el Ghouzzi et al., 1997). These insertions strongly modified the structure of the interhelical loops, and led to fewer contacts between interhelical loops and DNA. The present study thus provides a molecular explanation for the loss-of-function associated with several TWIST1 mutations/insertions observed in SCS patients (Bouard et al., 2014).

Previously, using MD simulations and biochemical assays, we and others highlighted the pivotal function of the R118H/Q/C, R120P, S144R, and K145E/Q residues in the molecular binding of the TE complex (Bouard

et al., 2016; El Ghouzzi et al., 1999; el Ghouzzi et al., 1997; El Ghouzzi et al., 2001; Gripp et al., 2001; Johnson et al., 1998; Maia et al., 2012; Paznekas et al., 1998; Rose, Patel, Reardon, Malcolm, & Winter, 1997). Mutations affecting these residues, as observed in SCS, lead to a decrease in the binding affinity of the dimer with the regulatory E-box sequences, and likely modify the transactivation functions of TWIST1, (el Ghouzzi et al., 1997; Firulli et al., 2005; Howard et al., 1997) resulting in an imbalance in the TWIST1/TWIST1 homodimer and in the TWIST1/E12 heterodimer during the developmental process (Connerney et al., 2006). In the present work, we completed this study by showing that the H-bond integrity between residues of bHLH was partially impaired, which appears to be the major cause of the decrease in protein dimerization. To compensate for this instability, the R154P mutant alternatively establishes H-bond interactions with both non-specific and consensus bases of the E-box, thus modifying the conformation of the heart of the dimer by losing the crucial N125-K145 and N566-K588 H-bonds interactions. It is worth noting that K145 mutations are also reported in patients suffering from SCS syndrome (El Ghouzzi et al., 2001).

Several bHLH proteins are known to dimerize with TWIST1, including the HAND proteins and proteins derived from the following genes: the TCF2/HEB, the TCF3/E2A (splice variants E12 and E47) and the TCF4/E2-2. Considering the degree of homology in their bHLH domains (92.7% identity with TCF2/HEB, 81.8% identity with TCF3/E2A E47 splice variant, and 89.1% identity with TCF4/E2-2), applying the current *in silico* approach to the evaluation of the dimerizing properties of all of these complexes would be highly relevant. In conclusion, these MD approaches allowed us to determine the key residues of TWIST1 and its partners involved in HLH dimerization, which may completely modify the balance of homo and heterodimer of TWIST1 complexes. Overall, MD simulations may constitute a powerful tool to predict the biological impact of these alterations in a cancer cell, in terms of dimerization, but also of binding affinity to active E-box sequences.

ACKNOWLEDGEMENTS

Authors are thankful to Dr Brigitte Manship for critical reading of the manuscript.

FUNDING

Our team is accredited by the Ligue Nationale contre le Cancer. This work was also supported by institutional grants from the LabEX DEVweCAN (ANR-10-LABX-61) and from the LyRIC (Lyon Recherche Intégrée en Cancérologie, Institut National contre le Cancer, INCa_4664). CB and LJ were recipients of fellowships from the Ligue Nationale contre le Cancer.

REFERENCES

- Anderson, D. J., Groves, A., Lo, L., Ma, Q., Rao, M., Shah, N. M., & Sommer, L. (1997). Cell lineage determination and the control of neuronal identity in the neural crest. *Cold Spring Harb Symp Quant Biol*, *62*, 493-504.
- Ansieau, S., Bastid, J., Doreau, A., Morel, A. P., Bouchet, B. P., Thomas, C., . . . Puisieux, A. (2008). Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell*, *14*(1), 79-89. doi: 10.1016/j.ccr.2008.06.005
- Bouard, C., Terreux, R., Honorat, M., Manship, B., Ansieau, S., Vigneron, A., . . . Payen, L. (2016). Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences. *Nucleic Acid Research (NAR-00118-N-2016.R2)*.
- Bouard, C., Terreux, R., Hope, J., Chemelle, J. A., Puisieux, A., Ansieau, S., & Payen, L. (2014). Interhelical loops within the bHLH domain are determinant in maintaining TWIST1-DNA complexes. *J Biomol Struct Dyn*, *32*(2), 226-241. doi: 10.1080/07391102.2012.762722
- Chang, A. T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J. W., . . . Yang, J. (2015). An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev*, *29*(6), 603-616. doi: 10.1101/gad.242842.114
- Connerney, J., Andreeva, V., Leshem, Y., Mercado, M. A., Dowell, K., Yang, X., . . . Spicer, D. B. (2008). Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure. *Dev Biol*, *318*(2), 323-334. doi: 10.1016/j.ydbio.2008.03.037
- Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M. A., & Spicer,

- D. B. (2006). Twist1 dimer selection regulates cranial suture patterning and fusion. *Dev Dyn*, 235(5), 1345-1357. doi: 10.1002/dvdy.20717
- Corsi, A. K., Brodigan, T. M., Jorgensen, E. M., & Krause, M. (2002). Characterization of a dominant negative *C. elegans* Twist mutant protein with implications for human Saethre-Chotzen syndrome. *Development*, 129(11), 2761-2772.
- De Masi, F., Grove, C. A., Vedenko, A., Alibes, A., Gisselbrecht, S. S., Serrano, L., . . . Walhout, A. J. (2011). Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res*, 39(11), 4553-4563. doi: 10.1093/nar/gkr070
- El Ghouzzi, V., Lajeunie, E., Le Merrer, M., Cormier-Daire, V., Renier, D., Munnich, A., & Bonaventure, J. (1999). Mutations within or upstream of the basic helix-loop-helix domain of the TWIST gene are specific to Saethre-Chotzen syndrome. *Eur J Hum Genet*, 7(1), 27-33. doi: 10.1038/sj.ejhg.5200240
- el Ghouzzi, V., Le Merrer, M., Perrin-Schmitt, F., Lajeunie, E., Benit, P., Renier, D., . . . Bonaventure, J. (1997). Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat Genet*, 15(1), 42-46. doi: 10.1038/ng0197-42
- El Ghouzzi, V., Legeai-Mallet, L., Benoist-Lasselain, C., Lajeunie, E., Renier, D., Munnich, A., & Bonaventure, J. (2001). Mutations in the basic domain and the loop-helix II junction of TWIST abolish DNA binding in Saethre-Chotzen syndrome. *FEBS Lett*, 492(1-2), 112-118.
- Ellenberger, T., Fass, D., Arnaud, M., & Harrison, S. C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev*, 8(8), 970-980.
- Firulli, B. A., Krawchuk, D., Centonze, V. E., Vargesson, N., Virshup, D. M., Conway, S. J., . . . Firulli, A. B. (2005). Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nat Genet*, 37(4), 373-381. doi: 10.1038/ng1525
- Firulli, B. A., Redick, B. A., Conway, S. J., & Firulli, A. B. (2007). Mutations within helix I of Twist1 result in distinct limb defects and variation of DNA binding affinities. *J Biol Chem*, 282(37), 27536-27546. doi: 10.1074/jbc.M702613200
- Fong, A. P., Yao, Z., Zhong, J. W., Cao, Y., Ruzzo, W. L., Gentleman, R. C., & Tapscott, S. J. (2012). Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev Cell*, 22(4), 721-735. doi: 10.1016/j.devcel.2012.01.015
- Gajula, R. P., Chettiar, S. T., Williams, R. D., Nugent, K., Kato, Y., Wang, H., . . . Tran, P. T. (2015). Structure-function studies of the bHLH phosphorylation domain of TWIST1 in prostate cancer cells. *Neoplasia*, 17(1), 16-31. doi: 10.1016/j.neo.2014.10.009
- Gripp, K. W., Kasparcova, V., McDonald-McGinn, D. M., Bhatt, S., Bartlett, S. P., Storm, A. L., . . . Stolle, C. A. (2001). A diagnostic approach to identifying submicroscopic 7p21 deletions in Saethre-Chotzen syndrome: fluorescence in situ hybridization and dosage-sensitive Southern blot analysis. *Genet Med*, 3(2), 102-108. doi: 10.1097/00125817-200103000-00003
- Grove, C. A., De Masi, F., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. (2009). A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell*, 138(2), 314-327. doi: 10.1016/j.cell.2009.04.058
- Howard, T. D., Paznekas, W. A., Green, E. D., Chiang, L. C., Ma, N., Ortiz de Luna, R. I., . . . Jabs, E. W. (1997). Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nat Genet*, 15(1), 36-41. doi: 10.1038/ng0197-36
- Jacqueroud, L., Bouard, C., Richard, G., Payen, L., Devouassoux-Shisheboran, M., Spicer, D., . . . Ansieau, S. (2016). The heterodimeric TWIST1-E12 complex drives the oncogenic potential of TWIST1 in human mammary epithelial cells. *Neoplasia (NEO-D-15-00360R1)*.
- Johnson, D., Horsley, S. W., Moloney, D. M., Oldridge, M., Twigg, S. R., Walsh, S., . . . Wilkie, A. O. (1998). A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1. *Am J Hum Genet*, 63(5), 1282-1293. doi: 10.1086/302122
- Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol*, 5(6), 226. doi: 10.1186/gb-2004-5-6-226
- Kress, W., Schropp, C., Lieb, G., Petersen, B., Busse-Ratzka, M., Kunz, J., . . . Collmann, H. (2006). Saethre-Chotzen syndrome caused by TWIST 1 gene mutations: functional differentiation from Muenke coronal synostosis syndrome. *Eur J Hum Genet*, 14(1), 39-48. doi: 10.1038/sj.ejhg.5201507
- Lister, J. A., & Baron, M. H. (1998). Induction of basic helix-loop-helix protein-containing

- complexes during erythroid differentiation. *Gene Expr*, 7(1), 25-38.
- Longo, A., Guanga, G. P., & Rose, R. B. (2008). Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry*, 47(1), 218-229. doi: 10.1021/bi701527r
- Maia, A. M., da Silva, J. H., Mencialha, A. L., Caffarena, E. R., & Abdelhay, E. (2012). Computational modeling of the bHLH domain of the transcription factor TWIST1 and R118C, S144R and K145E mutants. *BMC Bioinformatics*, 13, 184. doi: 10.1186/1471-2105-13-184
- Massari, M. E., & Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol*, 20(2), 429-440.
- Murray, S. S., Glackin, C. A., Winters, K. A., Gazit, D., Kahn, A. J., & Murray, E. J. (1992). Expression of helix-loop-helix regulatory genes during differentiation of mouse osteoblastic cells. *J Bone Miner Res*, 7(10), 1131-1138. doi: 10.1002/jbmr.5650071004
- Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., . . . Stuver, M. H. (1994). Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta*, 1218(2), 129-135.
- Paznekas, W. A., Cunningham, M. L., Howard, T. D., Korf, B. R., Lipson, M. H., Grix, A. W., . . . Jabs, E. W. (1998). Genetic heterogeneity of Saethre-Chotzen syndrome, due to TWIST and FGFR mutations. *Am J Hum Genet*, 62(6), 1370-1380. doi: 10.1086/301855
- Rose, C. S., Patel, P., Reardon, W., Malcolm, S., & Winter, R. M. (1997). The TWIST gene, although not disrupted in Saethre-Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases. *Hum Mol Genet*, 6(8), 1369-1373.
- Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., . . . Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol Biol*, 7, 33. doi: 10.1186/1471-2148-7-33
- Song, X.-j., Rienstra, C. M., & McDermott, A. E. (2001). N—H bond stretching in histidine complexes: a solid-state NMR study. *Magnetic Resonance in Chemistry*, 39(S1), S30-S36. doi: 10.1002/mrc.956
- Spicer, D. B., Rhee, J., Cheung, W. L., & Lassar, A. B. (1996). Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science*, 272(5267), 1476-1480.
- Teng, Y., & Li, X. (2014). The roles of HLH transcription factors in epithelial mesenchymal transition and multiple molecular mechanisms. *Clin Exp Metastasis*, 31(3), 367-377. doi: 10.1007/s10585-013-9621-6
- Thiery, J. P., Acloque, H., Huang, R. Y., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, 139(5), 871-890. doi: 10.1016/j.cell.2009.11.007
- Yang, J., & Weinberg, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell*, 14(6), 818-829. doi: 10.1016/j.devcel.2008.05.009

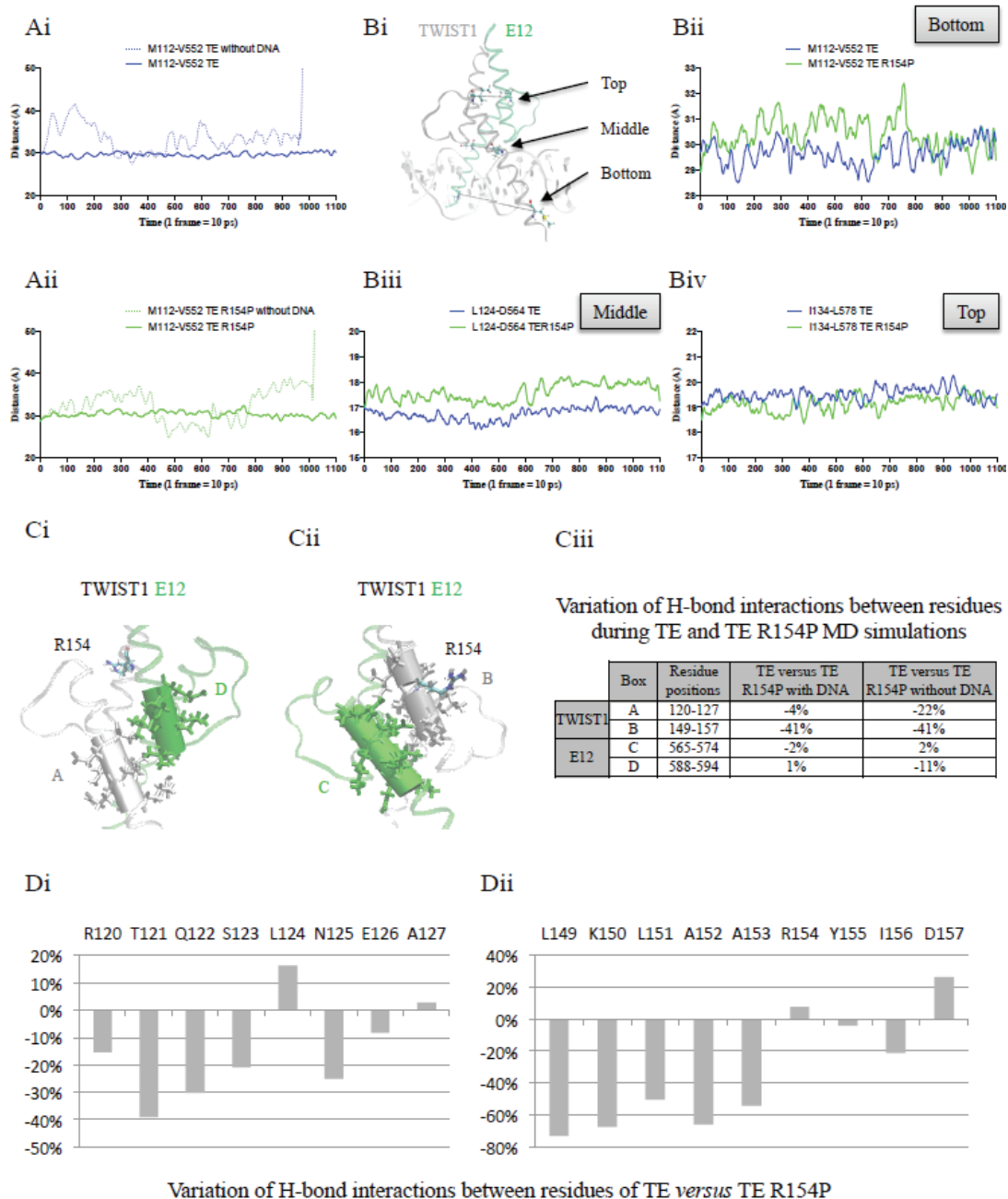


Figure 1 : Comparison of the stability of the wild-type TWIST1/E12 (TE) dimer and mutated R154P TE dimer.

Figure 1: Comparison of the stability of the wild-type TWIST1/E12 (TE) dimer and mutated R154P TE dimer.

A: Plots representing the distances (Å) between pairs of residues close to the bottom of the dimer as a function of time (total time = 10 ns, 1 frame per 10 ps), during 10 ns *in silico* molecular dynamics (MD) simulations. These distances were estimated during the MD simulations using the VMD 1.9.1 software (Graph of Labels Bonds). The lengths of H-bond interactions were represented via GraphPad Prism 5: “smooth, differentiate or integrate curve, with 8 neighbours”. The dynamic interactions between the methionine (M112) residue of TWIST1 and valine (V552) residue of E12 was studied in the (i) wild-type TWIST1/E12 (TE)-model, with (blue line) or without a DNA anchor (dashed blue line), and (ii) mutated R154P TE-model with (green line) or without (dashed green line) DNA anchor.

B: (i) 3D representation of the conserved TWIST1 (grey ribbon)/E12 (green ribbon) complex bound to DNA (light white ribbon) along with the E-box sequence (CATCTG), carried out using the VMD 1.9.1 software. The residues are numbered in each strand of the 5'-CATCTG-3' E-box core.

The three pairs of residues shown, namely M112-V552, L124-D564, I134-L578, were located at the bottom, in the middle and at the top of the dimer, respectively. (ii-iv) Plots representing the distances (Å) between pairs of residues of the TE (blue line) and TE R154P (green line) dimers as a function of time (total time = 10 ns, 1 frame per 10 ps), during 10 ns *in silico* MD simulations. Interactions between the (ii) M112 of TWIST1 and V552 of E12 residue pair, close to the bottom of the dimer, (iii) L124 of TWIST1 and D564 of E12 residue pair, close to the middle of the dimer, and (iv) I134 of TWIST1 and L578 of E12 residue pair, close to the top of the dimer.

C: (i-ii) 3D representation of the conserved TWIST1 (grey ribbon)/E12 (green ribbon) complex. Description of the position of the two series of residues (i) boxes A and D, and (ii) boxes B and C, in the TE dimer is represented as a cartoon and CPK. The first dimerization blocs composed of boxes A and D (i), B and C (ii) are represented by cylindrical grey and green solid surfaces, on the TWIST1 and E12 proteins, respectively. The localization of R154 on the TWIST1 ribbon is highlighted in CPK. (iii) Evaluation of the impact of the R154P mutation on the number of H-bonds, in each individual box A to D, established between residues and DNA (with DNA) or exclusively between residues (without DNA).

D: Bar chart representing the variation of H-bond interactions between residues of the TE versus R154P TE complex. The sum of the interactions occurring between residues within the box A (i) and box B (ii) of the wild type or mutated TWIST1 protein is highlighted. The percentage of cumulated occupancies of H-bond interactions occurring in the mutated R145P TE model is normalized against the TE model (100%). All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the 10 ns time-course of the MD simulation (interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test:“<2,10”, value of 1 if true and 0 if false; NB.SI function: NB.SI(range;criterion)).

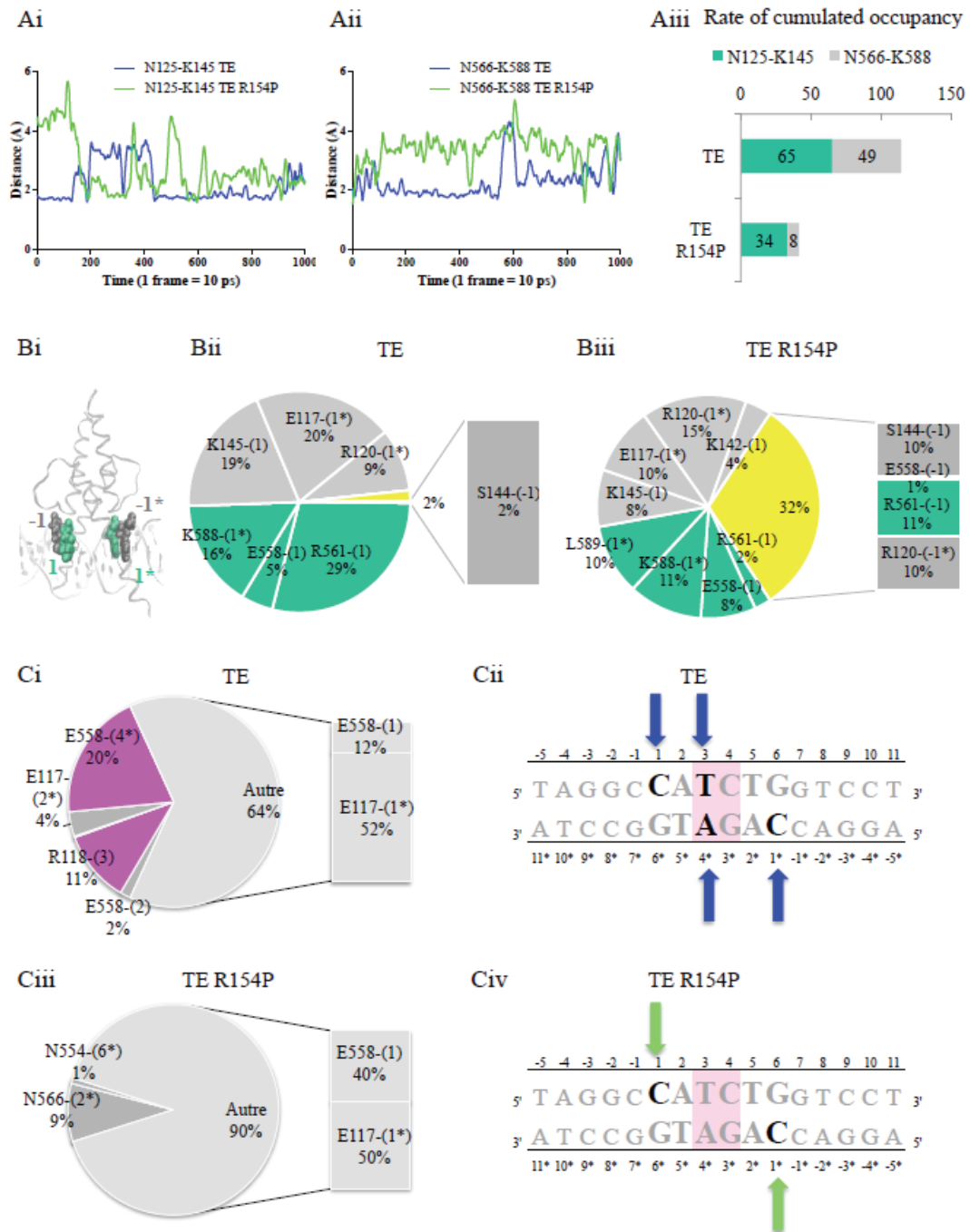


Figure 2 : Consequence of impaired TWIST1/E12 dimerization on DNA binding.

Figure 2: Consequence of impaired TWIST1/E12 dimerization on DNA binding.

A: Plots representing the shortest distances (Å) between the OD1 atom of asparagine (N125) and the NZ1, NZ2 or NZ3 atoms of the lysine (K145) residues (i), and between N566 and K588 residues (ii) as a function of time (total time = 10 ns, 1 frame per 10 ps), during the wild-type TWIST1/E12 (TE) (blue line) and mutated TE R154P (green line) simulations. (iii) The horizontal bar chart shows the cumulated occupancy values for the H-bond interactions established between the N125-K145 residues of TWIST1 (grey) and the N566-K588 residues of E12 (green) during the TE and TE R154P MD simulations. All cumulated occupancy values of the H-bonds were calculated as described in the Materials and Methods section. Briefly, H-bond interactions are assigned a value according to the distance between their atomic donors/acceptors during the time (the 10 ns of the

MD)(interactions score 1 if their distance is under 2.10 Å, and 0 if above). Higher occupancy values being obtained for shorter and, therefore, more stable interactions. (SI function: SI(test_logic; value_if_true;value_if_false) with logic test:“<2,10”, value of 1 if true and 0 if false; NB.SI function: NB.SI(range;riterion)).

B: (i) 3D *in silico* representation of the TE complex bound to DNA, carried out using the VMD 1.9.1 software. The proximal flanking (-1; -1*) and first (1*; 1*) bases of the E-box are represented in grey and green VDW, respectively. (ii-iii) The pie charts show the percentages of cumulated occupancy of H-bonds established between residues of TWIST1 (grey) and E12 (green) and DNA bases during the (ii) TE and (iii) R154P TE MD simulations. Occupancies of H-bonds established with the proximal flanking (-1 and -1*) bases are represented in yellow, while the bar charts highlight the exact residues binding to those bases. The H-bonds established with the first consensus bases of the E-box (1 and 1*) are represented in grey and green for the TWIST1 and the E12 residues, respectively.

C: Distribution of “specific” H-bonds established between residues and E-box bases in the (i) TE and (iii) mutated R154P TE molecular dynamics simulations. The pie charts show percentages of the total rate of cumulated occupancy of specific H-bonds established during the MD simulations. Consensus bases and variable bases of the E-box sequences are represented in grey and pink, respectively. The bar charts highlight the specific residues binding to cytosine (+1) and (+1*) bases, and show their implication in the DNA binding affinity. The DNA sequences are represented for (ii) the TE and (iv) mutated R154P TE dimers. The arrows highlight the major base interactions for the TE complex (blue arrow) and mutated R154P TE complex (green arrow).

CHAPITRE 6 : PARTICIPATION COLLABORATIVE AUX TRAVAUX DE L'EQUIPE

Mon implication dans ces travaux de l'équipe, notamment ceux du MBL-141 et du TMTZ m'ont permis de participer à des études précliniques. En effet dans ces deux études, j'ai participé au suivi et à la réalisation des protocoles ainsi qu'à l'interprétation des résultats. Ceci m'a permis de consolider mes connaissances et de valider mon niveau 1 en expérimentation animale.

Travail collaboratif 1

**« MBL-II-141, a chromone derivative, enhances irinotecan (CPT-11) anticancer efficiency
in ABCG2-positive xenografts »**

Honorat M., Guitton J., Gauthier C., Bouard C., Lecerf-Schmidt F., Peres B., Terreux R.,
Gervot H., Rioufol C., Boumendjel A., Puisieux A., Di Pietro A. Payen L..

Oncotarget, 2014

1 Travail collaboratif 1 : Etude de l'activité inhibitrice du MBLI-141 sur les fonctions de transport d'ABCG2 (études précliniques)

Les protéines ABC, pour protéines « ATP-binding cassette », sont des protéines membranaires capables de transporter leurs substrats à l'encontre du gradient de concentration en se servant de l'énergie fournie par l'hydrolyse de l'ATP. Ces protéines permettent de protéger les cellules des altérations physiologiques ou pathologiques induites par des molécules endogènes et exogènes (xénobiotiques ou des toxines). La protéine ABCG2 (ATP-binding cassette sub-family G member 2), membre de la sous famille G, qui est une protéine de résistance des cancers du sein impliquée dans les mécanismes de résistances multiples aux médicaments (phénotype MDR). Le transporteur ABCG2 est responsable de la sécrétion de nombreuses molécules comme les agents anti-cancéreux que sont la métoxantrone, le topotecan, l'irinotecan et son métabolite actif le SN-38. Ainsi en diminuant la rétention intracellulaire de ces agents anti-cancéreux, ABCG2 réduit considérablement la cytotoxicité de ces drogues. ABCG2 joue également un rôle physiologique. Elle est fortement exprimée au niveau des barrières physiologiques au niveau du pôle apicale des cellules polarisées (comme par exemple la barrière hémato-encéphalique) et limite ainsi la distribution cérébrale des médicaments. La forte l'expression de ce transporteur est associée à une plus faible survie des patients porteurs de tumeurs exprimant ABCG2. L'utilisation de siRNA contre ABCG2 bloque les propriétés malignes de cellules de carcinome hépatocellulaire exprimant ABCG2.

La protéine ABCG2 est donc une cible thérapeutique intéressante pour vaincre la chimiorésistance. L'équipe d'Attilio Di Pietro a travaillé sur le criblage d'une bibliothèque de molécules chimiques dans le but d'identifier des inhibiteurs spécifiques de ABCG2 dépourvue de toxicité. Ils ont identifié un inhibiteur de haute affinité et de faible cytotoxicité : le MBL-II-141 ou 5-(4-bromobenzyloxy)-2-(2-(5-methoxyindolyl) ethyl-1-carbonyl)-4H-chromen-4-one. Il est décrit comme le modulateur le plus puissant de la protéine ABCG2. Cette molécule est capable d'inhiber l'efflux de tous les substrats testés. Le but de cette étude a donc été d'évaluer *in vivo* la puissance de cet inhibiteur au vu de son haut index thérapeutique (>900) dans la but de sensibiliser les tumeurs positives à ABCG2. Pour cela, notre équipe a mis en place un protocole de traitement pharmacologique basé sur la dose et la voie d'administration pour évaluer l'efficacité du MBL-II-141 sur la croissance tumorale de xénogreffes ABCG2 positives et négatives.

Nous avons étudié la sensibilité du MBL-II-141 sur des souris ayant des xénogreffes ABCG2 positives et négatives. Le MBL-II-141 seul a un effet négligeable sur la croissance tumorale. L'Irinotecan ralenti la croissance des tumeurs ABCG2 mais moins efficacement que sur les tumeurs ABCG2 négatives. Nous avons pu observer que la combinaison du MBL-II-141 et de CPT-11 diminuait considérablement la croissance tumorale des xénogreffes (ABCG2 positives) et favorise ainsi la survie des souris. Ainsi le traitement au MBL-II-141 permet de sensibiliser les cellules de

xénogreffes ABCG2 positives aux traitements CPT-11. En effet, la combinaison de 10 mg / kg de MBL-II-141 avec l'agent anticancéreux CPT-11 permet de bloquer complètement la croissance de 90% des tumeurs fraîchement implantées positives pour ABCG2. Par ailleurs, la même combinaison permet de ralentir la croissance des tumeurs déjà établies. Nous avons également étudié l'impact du MBL-II-141 sur la pharmacocinétique de CPT-11 et de SN-38 au niveau du sang, du foie et du cerveau puisque cela est requis pour le développement préclinique. En effet, le MBL-II-141 modifie les pharmacocinétiques du CPT11 et du SN38. Le MBL-II-141 s'accumule très peu lorsqu'il est absorbé par voie intra-péritonéale et est bien toléré. De plus le MBL-II-141 augmente fortement les niveaux de CPT-11 dans le cerveau, et serait un agent précieux pour améliorer la distribution des médicaments dans le cerveau lors du traitement de gliomes agressifs. Toutes ces données sont supportées par les prédictions *in silico* du logiciel Percepta. En effet, le MBL-II-141 est une molécule prédite comme bien tolérée. Ce logiciel permet de prédire son absorption par voie orale, sa distribution tissulaire, sa génotoxicité et sa cancérogénicité *via* la comparaison de données expérimentales de molécules structurellement liée au MBL-II-141.

Cette étude a permis de mettre en évidence que le rôle important que joue le MBL-II-141 sur les xénogreffes ABCG2 positives en synergie avec CPT-11 pour une activité anticancéreuse. Ainsi au vu de la haute tolérance des souris pour le MBL-II-141 et des effets potentialisateurs observés du MBL-II-141, cette molécule serait adaptée pour le développement clinique pour la mise en place d'un traitement des cancers résistants ou le transporteur ABG2 joue un rôle central.

MBL-II-141, a hormone derivative, enhances irinotecan (CPT-11) anticancer efficiency in ABCG2-positive xenografts

Mylène Honorat^{1,2,3}, Jérôme Guitton^{1,4,5}, Charlotte Gauthier³, Charlotte Bouard², Florine Lecerf-Schmidt⁶, Basile Peres⁶, Raphaël Terreux^{1,3}, Héloïse Gervot², Catherine Rioufol⁵, Ahcène Boumendjel⁶, Alain Puisieux², Attilio Di Pietro³, Léa Payen^{1,2,5}

¹Institut des Sciences Pharmaceutiques et Biologiques (ISPB), Université Lyon 1, Lyon 69373, France

²INSERM UMR-S1052, CNRS UMR 5286, Université Lyon 1, Centre de Recherche en Cancérologie de Lyon (CRCL), LabEx DEVweCAN, Centre Léon Bérard, Lyon 69373, France

³Equipe labellisée Ligue 2014, BMSSI UMR 5086 CNRS-Université Lyon 1, IBCP, Lyon 69373, France

⁴EA 3738, Laboratoire de Ciblage Thérapeutique en Cancérologie, Centre Hospitalier Lyon-Sud, Pierre Bénite, France

⁵Hospices Civils of Lyon, Laboratoire de Biochimie de Lyon Sud (CBS), Lyon, France

⁶Univ. Grenoble Alpes/CNRS, DPM UMR 5063, F-38041 Grenoble, France

Correspondence to:

Léa Payen, e-mail: lea.payen-gay@univ-lyon1.fr

Keywords: ABCG2, MDR phenotype, Inhibitors, Chemosensitization

Received: June 17, 2014

Accepted: October 02, 2014

Published: January 20, 2015

ABSTRACT

ABCG2 is responsible for the multidrug resistance (MDR) phenotype, and strongly modulates cancer outcomes. Its high expression at a number of physiological barriers, including blood-brain and intestinal barriers, impacts on drug pharmacokinetics parameters. We characterized MBL-II-141, a specific and potent ABCG2 inhibitor. Combination of 10 mg/kg MBL-II-141 with the anticancer agent CPT-11 completely blocked the growth of 90% freshly implanted ABCG2-positive tumors. Moreover, the same combination slowed the growth of already established tumors. As required for preclinical development, we defined the main pharmacokinetics parameters of MBL-II-141 and its influence on the kinetics of CPT-11 and its active metabolite SN-38 in mice. MBL-II-141 distribution into the brain occurred at a low, but detectable, level. Interestingly, preliminary data suggested that MBL-II-141 is well tolerated (at 50 mg/kg) and absorbed upon force-feeding. MBL-II-141 induced a potent sensitization of ABCG2-positive xenografts to CPT-11 through *in vivo* ABCG2 inhibition. MBL-II-141 strongly increased CPT-11 levels in the brain, and therefore would be a valuable agent to improve drug distribution into the brain to efficiently treat aggressive gliomas. Safety and other pharmacological data strongly support the reglementary preclinical development of MBL-II-141.

INTRODUCTION

ATP-binding cassette (ABC) proteins secrete various endogenous and exogenous molecules to protect cells from xenobiotic- and toxin-induced alterations under physiological and pathological conditions. Three major ABC transporters, P-glycoprotein (P-gp/MDR1/ABCB1), multidrug resistance protein 1 (MRP1/ABCC1), and breast cancer resistance protein

(BCRP/MXR/ABCG2), are overexpressed and widely contribute to the multidrug resistance (MDR) phenotype. The latter is also associated with a lower survival of cancer patients [1–4]. ABCG2 expression provides an early prediction of the clinical outcome of cancers, which can be used to improve the clinical management of gastric cancer patients [5]. Interestingly, expression levels of ABCG2 have been described as an important determinant of the “side-population” (SP)

phenotype, characteristic of stem cells (as a Hoechst 33342-negatively stained population compared to the majority of cells) [6]. ABCG2-positive cells from hepatocellular carcinoma (HCC), displayed enhancements in tumorigenicity, proliferation capacity, doxorubicin resistance, cell migration, and invasion potential, whereas ABCG2 down-expression by siRNA significantly blocked these malignant properties [7]. Consequently, anticancer therapy could eliminate the majority of tumor cells, generate clonal selection of aggressive tumor cells overexpressing ABCG2 and cause their relapse. In conclusion, this suggests that the ABCG2 protein is an appealing therapeutic target, and inhibiting its mediated secretion or decreasing its expression level in cancer cells constitute potential strategies to overcome chemoresistance.

ABCG2 is responsible for the secretion of a large range of molecules, including anticancer agents such as mitoxantrone [8], topotecan, irinotecan and its SN-38 active metabolite [9, 10]. *In vitro* experiments showed that, by decreasing the intracellular retention of anticancer agents, ABCG2 drastically reduces drug cytotoxicity. In addition, since it is strongly expressed in polarized cells within physiological barriers, including the blood-brain barrier (BBB) [11], hepatocytes [12] and enterocytes [13], it also limits drug distribution within the body. ABCG2 is also expressed in the breast [14] and the kidney [12]. It is well recognized to influence the pharmacokinetics parameters of substrates such as topotecan [15], camptothecin [16], methotrexate (in combination with ABCC2 and ABCC3) [17], tandutinib [18], and erlotinib [19]. In consequence, inhibition of ABCG2 may strongly improve distribution of its substrates into the brain [20] and thus modulate brain tumor treatment. Recently, Kawamura *et al.* showed the role of ABCG2 and ABCB1 on CPT-11 (irinotecan) distribution inside mouse brain [21] which could be harnessed to improve efficiency of aggressive gliomas treatment. Strongly expressed at the physiological barriers, ABCG2 could also decrease plasma levels of anticancer drugs that are rapidly metabolized and excreted without reaching pharmacological concentration levels.

Whilst initial clinical assays were often discordant and not fully convincing, targeting ABC proteins with modulators remains a possible approach to improve chemotherapy efficiency. Chemosensitization strategies targeting ABCB1 have been so far unsuccessful due to low numbers of candidate inhibitors and unsuitable trials [22]. Inhibitors still need to be improved by increasing substrate specificity and lowering toxicity. Moreover, they should not show any adverse effect on drug pharmacokinetics that could lead to toxic plasma levels of anticancer drugs. The neurotoxic fumitremorgin C (FTC) was the first selective ABCG2 inhibitor identified [23].

Later, synthetic analogs of FTC were studied, including Ko143 which was devoid of toxicity and increased the oral availability of topotecan, giving plasma levels 4 to 6-fold higher in ABCB1(-/-) mice [24]. *In vitro*, other ABCG2-specific modulators were developed from the structure of the ABCB1 inhibitor tariquidar [25], although their *in vivo* efficiency over ABCG2-mediated chemoresistance has not been yet reported. Very recently, telatinib (15 mg/kg) in combination with doxorubicin (1.8 mg/kg) was shown to significantly decrease the growth rate and tumor size of ABCG2-overexpressing tumors in a xenografted nude mouse model [26]. In order to specifically study the impact of ABCG2 modulation on substrate pharmacokinetics and reversion of MDR phenotype, we designed an ABCG2-selective inhibitor (MBL-I-87) using the structure of the ABCB1 inhibitor elacridar as a template. It proved to be an efficient inhibitor of ABCG2 excretion [27] which improved *in vivo* CPT-11 anticancer activity on the growth of ABCG2-positive xenografted tumors [28, 29]. We also developed an inhibition model of tumor growth [30] predicting tumor growth dynamics, to be used as a template to design studies for producing more efficient ABCG2 inhibitors. Indeed, our team recently screened an original chemical library to identify more powerful inhibitors, some being devoid of toxicity. Using a chromone core instead of the acridone moiety, we identified a highly promising candidate for *in vivo* studies: MBL-II-141 or 5-(4-bromobenzyloxy)-2-(2-(5-methoxyindolyl) ethyl-1-carbonyl)-4H-chromen-4-one, also called chromone 6g [31]. MBL-II-141 displayed a high-affinity inhibition activity (IC₅₀ of 0.11 μM) and a very low cytotoxicity (IG₅₀ > 100 μM). It was shown to inhibit ABCG2 in various cell lines, either. Transfected with expression vector of ABCG2 or selected with mitoxantrone [32] and was found to be the most powerful modulator yet described. Similarly to other ABCG2-selective inhibitors such as acridones [27] and methoxy trans-stilbenes [33], MBL-II-141 inhibition is non-competitive towards mitoxantrone [31]. It is able to strongly inhibit the efflux of all tested substrates including mitoxantrone and SN-38 [31], pheophorbide A and BODIPY-prazosin [32]. Its remarkable high therapeutic index (> 900) completely justified its *in vivo* evaluation aimed at chemosensitizing ABCG2-positive tumors. In this study, we established a protocol of pharmacological treatment (*i.e.* dose and route of administration) based on the prediction model developed with MBL-I-87 [29] to evaluate the efficacy of MBL-II-141 against tumor growth. Furthermore, we looked at its tolerance in mice and we found that its pharmacokinetic parameters met the preliminary requirements for further clinical drug development.

RESULTS

MBL-II-141 *in vivo* sensitizes ABCG2-positive xenografts to CPT-11 treatment

In vitro studies previously demonstrated the valuable inhibitory effect of MBL-II-141 on ABCG2 drug-efflux activity. This compound was found to be a powerful and highly-specific inhibitor of ABCG2, blocking mitoxantrone efflux with an IC_{50} of 0.11 μ M, without any effect on ABCG1-mediated transport using either mitoxantrone or calcein as substrates [31]. Furthermore, little effect on viability was observed *in vitro*, with less than 25% cell death of HEK293 cells at 100 μ M; therefore, MBL-II-141 displayed a low cytotoxicity and appeared to be a promising molecule for *in vivo* assays of ABCG2 inhibition.

Firstly, we evaluated MBL-II-141 efficacy to chemosensitize ABCG2-positive, freshly established, tumors to CPT-11 treatment. Following the onset of tumors growing from HEK293-ABCG2 cells or control HEK293-pcDNA3.1 cells (implanted into the flank of SCID mice), the animals were treated by CPT-11, either alone or in combination with MBL-II-141. All xenografts grew rapidly in the presence of the vehicles (corn oil/5% DMSO and PBS) (Figure 1), while MBL-II-141 displayed a slight inhibitory effect on the growth of ABCG2-negative tumors (Figure 1A and Supplementary table 1). This effect was not observed for ABCG2-positive tumors (Figure 1B), and was not related to an enhanced cytotoxicity of MBL-II-141 on HEK293-pcDNA3.1 control cells. On the contrary, it was found to be slightly more cytotoxic *in vitro* on HEK293-ABCG2 cells than on HEK293-pcDNA3.1 cells (leading to 60% and 75% cell survival, respectively, at 100 μ M) [31].

The CPT-11 treatment completely blocked the growth of ABCG2-negative cells, whereas it only delayed, by about two months, the growth of ABCG2-positive tumors. Interestingly, co-treatment of CPT-11 with MBL-II-141 was able to completely block the growth of 90% ABCG2-positive tumors. In terms of animal survival (Figure 1A), CPT-11 alone was sufficient to ensure the survival of 100% of mice bearing ABCG2-negative tumors. It also increased the survival of mice bearing ABCG2-positive tumors, although these mice had to be sacrificed at day 110 (Figure 1B) as the tumor size reached the maximal size authorized by the ethical code. The combination of MBL-II-141 and CPT-11 remarkably increased the survival rate of mice, up to 90%, even after 40 days following the end of the treatment (Figure 1B).

Since the combination efficiently prevented the growth of early ABCG2-positive tumors, we then tested

the effects of the same combination on established tumors (Figure 2). We delayed the pharmacological treatment until tumors reached 8 mm in diameter (mean volume of 238 mm^3); this represented well-established tumors, fully developed and undergoing exponential growth. Treatments with MBL-II-141 alone had a negligible effect on tumor growth (Figure 2A), with a slope of 68 ± 3 mm^3/day versus 61 ± 3 mm^3/day for tumors treated with solvents. CPT-11 alone was able to delay the tumor growth, with a slope of 50 ± 2 mm^3/day , but did not markedly improve mice survival (Figure 2B). Only the drug combination was able to greatly lower tumor growth, with a slope of 41 ± 1 mm^3/day , and the tumors reached their maximal size about two weeks later than those treated with CPT-11 alone (Figure 2A). Consequently, the combination markedly increased mice survival (Figure 2B and Supplementary table 2).

MBL-II-141 modifies CPT-11 and SN-38 pharmacokinetics

From the perspective of putative pharmaceutical development, we fully analyzed the kinetics parameters of MBL-II-141 in SCID mice (Figure 3 and Supplementary table 3). The highest level of MBL-II-141 in total blood was observed at 0.5 h (Figure 3A). The levels then rapidly decreased to below the detection limit at 6 h. The elimination profile was similar in either the presence or absence of CPT-11 after 1 h. MBL-II-141 distribution into the brain was at a low, but still detectable, level (Figure 3C). CPT-11 appeared to decrease the MBL-II-141 distribution into the brain until 4 h, then MBL-II-141 levels became similar in either absence or presence of CPT-11.

While the CPT-11 injected dose was 2-fold higher than MBL-II-141 (20 versus 10 mg/kg), the CPT-11 blood levels were relatively much higher (Figure 3A), and the levels were globally 10-fold higher than those of MBL-II-141 at the same time points. 90% of CPT-11 was lost from the blood at 4 h. CPT-11 levels also rapidly decreased in the liver (Figure 3B). A very weak distribution was observed into the brain, with a value 200-fold lower than in the liver. However, the CPT-11 brain levels were significantly and durably increased in the presence of MBL-II-141 (Figure 3C), by approximately 2-fold (Supplementary table 4), whereas no impact was observed on the liver levels.

In parallel, we also quantified the levels of SN-38 in blood and tissues, which acts as both the CPT-11 active metabolite and an ABCG2 substrate. It was detected in blood until 6 h (Figure 3A), and the median level greatly decreased after 2 h. In the presence of MBL-II-141, SN-38 kinetics displayed a similar profile (with an

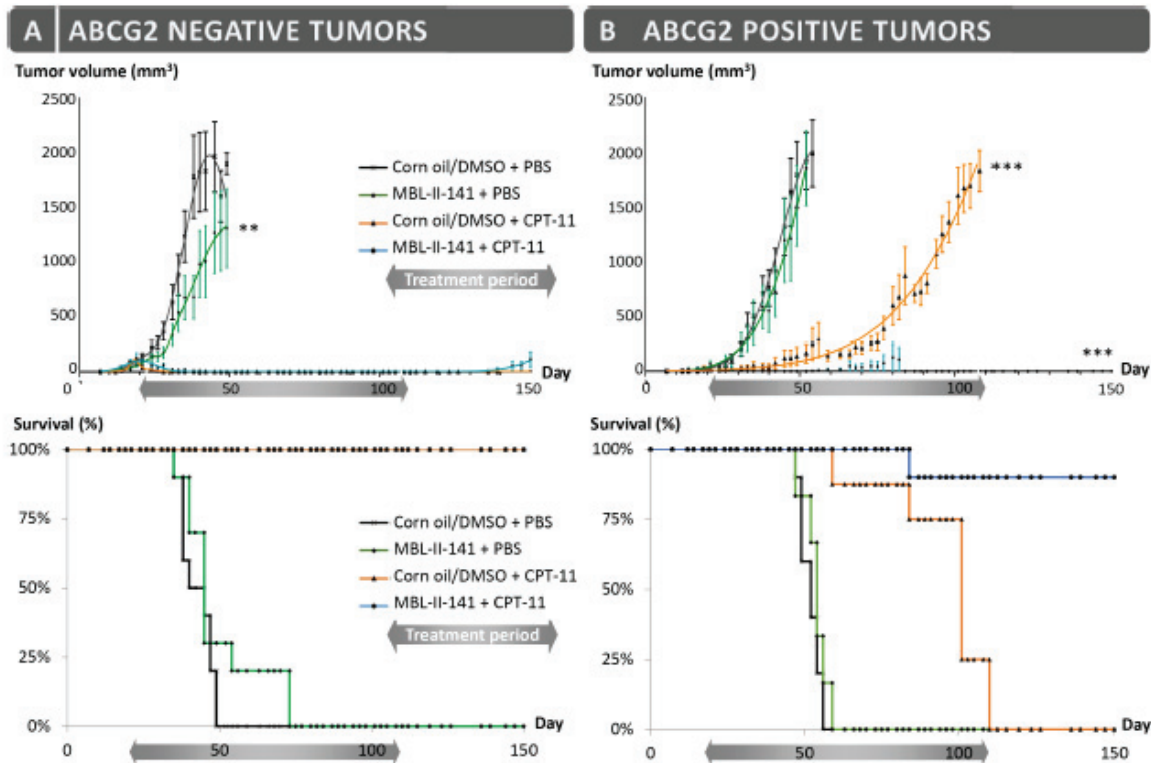


Figure 1: Effects of CPT-11 treatment in combination with MBL-II-141 on tumor growth and mice survival. (A) HEK-pcDNA3 and (B) HEK-ABCG2 xenografts were implanted sub-cutaneously into mice. After onset of the tumors, mice were treated three times a week by: corn oil/0.5% DMSO + PBS (black crosses), MBL-II-141 at 10 mg/kg + PBS (green squares), corn oil/5% DMSO + CPT-11 at 20 mg/kg (orange triangles) and MBL-II-141 at 10 mg/kg + CPT-11 at 20 mg/kg (blue circles). The treatments started at day 21 and lasted until day 108. $**P < 0.001$

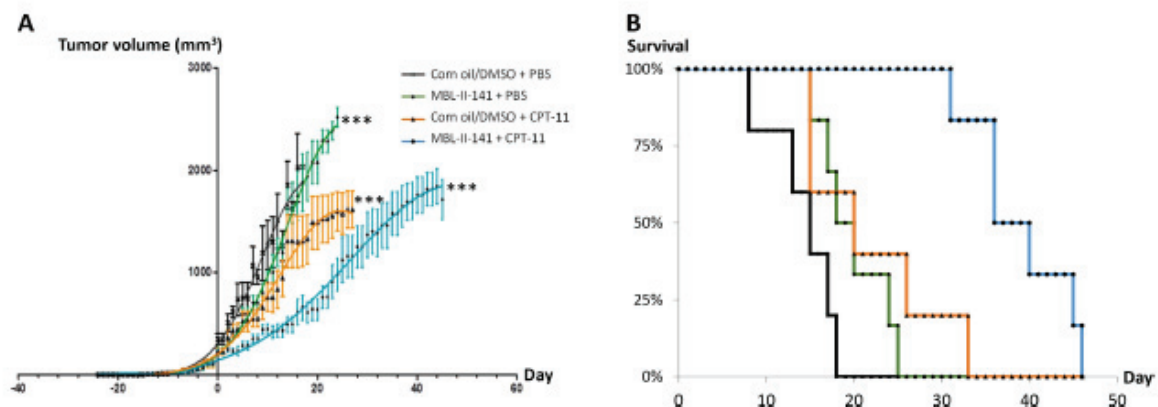


Figure 2: Effects of CPT-11 treatment in combination with MBL-II-141 on ABCG2-positive tumors of 8 mm diameter. (A) Tumor growth curves and (B) Mice survival. HEK293 ABCG2 xenografts were implanted sub-cutaneously into mice. When tumors reached 8 mm in diameter, mice were treated three times a week by: corn oil/0.5% DMSO + PBS (black crosses), MBL-II-141 at 10 mg/kg + PBS (green squares), corn oil/5% DMSO + CPT-11 at 20 mg/kg (orange triangles) and MBL-II-141 at 10 mg/kg + CPT-11 at 20 mg/kg (blue circles). Treatment started at day 0. $***P < 0.001$

increase of blood level at 1 h) but, interestingly, MBL-II-141 increased the SN-38 blood levels. This effect disappeared after 8 h, when levels became almost undetectable. In the liver, SN-38 followed a similar profile as in the blood, but the increase observed in the presence of MBL-II-141 was barely significant (Figure 3B and Supplementary Table 3), while SN-38 levels in the brain were below the limits of quantification (Figure 3C).

MBL-II-141 is accumulated at low levels and is orally absorbed

As previously observed, MBL-II-141 was rapidly eliminated from blood after a single injection. To study a possible accumulation effect, we quantified blood levels of MBL-II-141 at 1 or 4 h after the last injection in a multiple-administration protocol. MBL-II-141 blood levels were similar after one or two injection(s) (Figure 4A), while the third injection led to a slight increase of blood level, regardless of the time of blood collection (Figure 4A). In addition, we evaluated whether this molecule was orally absorbed upon

force-feeding. The blood level was approximately 2-fold lower at 4 h under force-feeding in comparison to an intraperitoneal (IP) route (Figure 4C). Nevertheless, the observed difference was not significant due to the high variability of blood levels.

MBL-II-141 is predicted to be a highly tolerated molecule

In view of the requirements of this new drug for pharmaceutical development, we investigated *in silico* safety data by using the Percepta software. By comparing *in vitro* experimental data from structurally-related molecules, it was possible to predict its oral absorption, tissue distribution, genotoxicity and carcinogenicity (Table 1). Percepta predictions indicated only a weak acute toxicity of MBL-II-141. In mice, the predicted LD50 was 290, 1200, 77 or 320 mg/kg, respectively, by IP, oral, intravenous or sub-cutaneous route. Similar data were predicted for rats, which are frequently used in preclinical studies (Table 1). In our pilot study, no acute toxicity was experimentally observed at 10 mg/kg per mouse. Furthermore, reiterated administrations confirmed

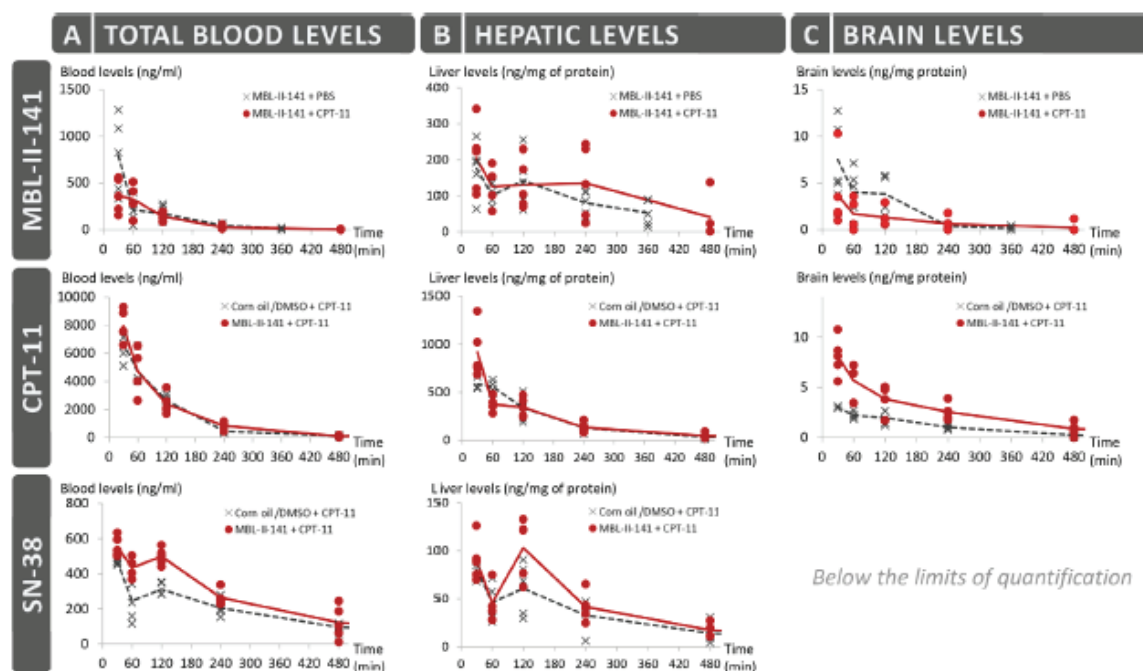


Figure 3: Pharmacokinetics studies of MBL-II-141 levels in total blood (A) liver (B) and brain (C). MBL-II-141 alone (at 10 mg/kg) or in combination with 20 mg/kg of CPT-11 was intraperitoneally administered to SCID mice. Blood, brain and liver were taken for LC-MS/MS analysis after 30, 60, 120, 240 or 360 min for treatments with MBL-II-141 alone, and after 30, 60, 120, 240 or 480 min for combination treatments ($N = 5$ mice/group).

Table 1: Predictions of MBL-II-141 behavior by the ACD Percepta software. Predictions were obtained according to data available for structurally related molecules. LD50: lethal dose 50; RBA: relative binding affinity (Log RBA) to a receptor, compared to the reference ligand estradiol; HSA: human serum albumin; PPB: percentage of a compound bound to human plasma proteins; V_D : volume of distribution; BB: brain/blood distribution ratio; PS: rate of passive diffusion/permeability (logPS).

Physicochemistry		Oral bioavailability	
Molecular mass	547.40 g/mol	Bioavailability	9%
Index of refraction	1.68	Active transport	Not transported by PepT1 et ASBT
Density	1.48	Passive absorption	
logP	4.87		100% (transcellular)
Acute toxicity		ABCB1 specificity	
Mouse LD50		Inhibitor	Weak probability: 0.15 (Ki < 1µM) (RI = 0.18)
Intraperitoneal	290 mg/kg	Substrate	Weak probability: 0.03 (RI = 0.26)
Oral	1200 mg/kg	Distribution	
Intravenous	77 mg/kg	Protein Binding	Acid compound
Sub-cutaneous	320 mg/kg		Bound fraction to HSA; %PPB 99.68% (RI = 0.31)
Rat LD50			Unbound in plasma = 0.0032
Intraperitoneal	570 mg/kg	V_D	3.5 L/kg
Oral	210 mg/kg	Blood-Brain Barrier (BBB)	
Genotoxicity		BBB	Low brain penetration
Ames test positivity	Weak probability: 0, 17	log BB	-0.41
Irritation by local exposure		log PS	-1.3
Eye irritation	Probability: 0.21		
Skin irritation	Probability: 0.25		
Endocrine system: estrogen receptor (ER)			
Binding to ER α	Weak probability (-3 < log RBA < 0)		

MBL-II-141 safety without any associated animal death or loss of weight. Upon IP administration at 10, 30 or 50 mg/kg MBL-II-141 in combination with CPT-11 at 20 mg/kg, mice did not show any modification in behavior (physical activity, appearance). Complete blood count proportion was unchanged regardless of the administered dose of MBL-II-141 (data not shown), and macroscopic analysis of various organs (brain, heart, lungs, kidneys, liver and spleen) showed no apparent lesions. These results were confirmed with the Percepta software where the Ames test was negative, with and without metabolism systems.

In the absence of structurally-related molecules in the Percepta software database, a low index of confidence was not achieved for prediction validation of complementary genotoxicity tests (micronucleus *in vitro* and *in vivo*). Only a weak probability was predicted for MBL-II-141 on its skin irritating nature and putative effect on the endocrine system via interaction with estrogen receptor alpha (ER). To complete the predicted safety of MBL-II-141, we observed that its structure would not permit transport either by PepT1 and ASBT, or by ABCB1. Moreover, it must bind with high affinity to human serum albumin

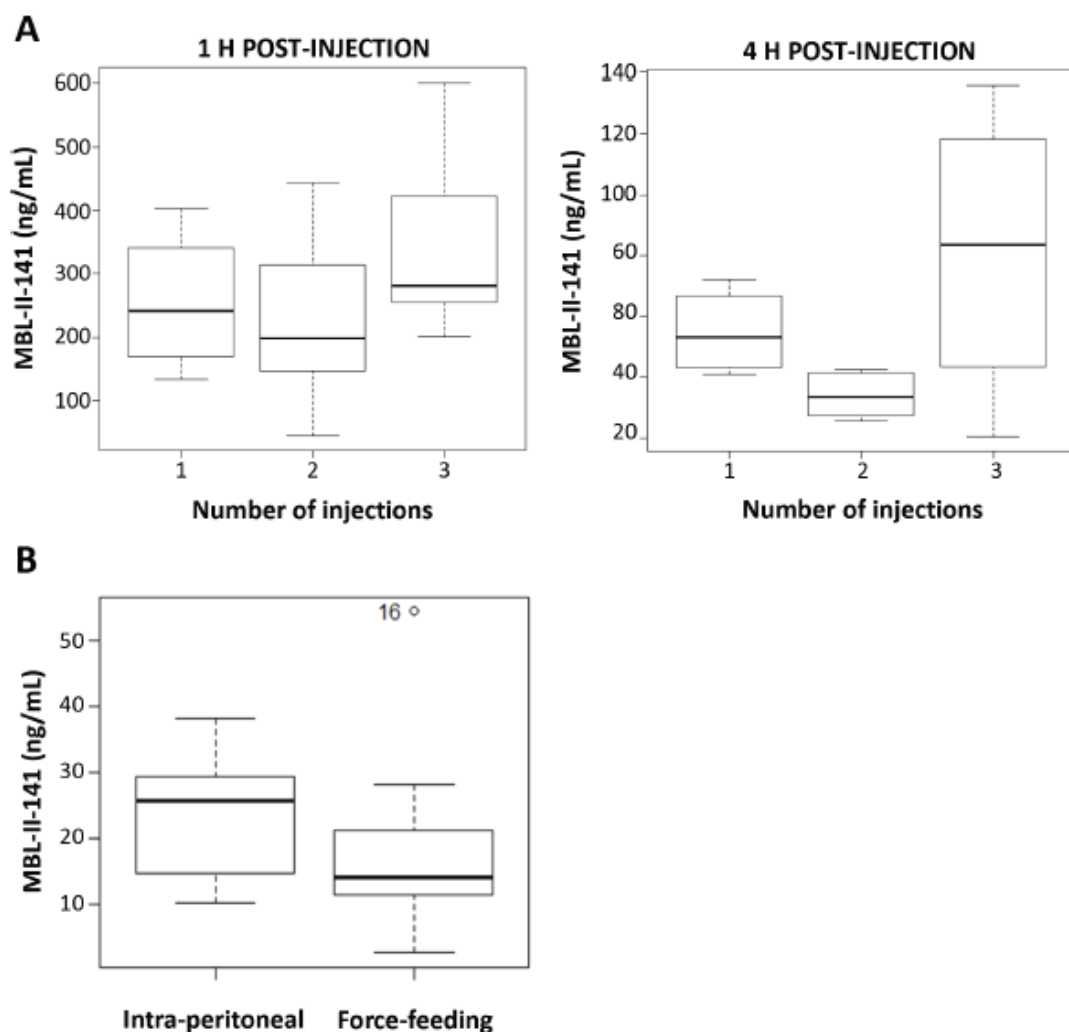


Figure 4: (A) Impact of multiple-injection treatments on MBL-II-141 plasma levels at 1 h or 4 h after the last injection. MBL-II-141 was administered at 10 mg/kg to SCID mice by IP injection one, two or three times separated by 4 h intervals. Blood was withdrawn for HPLC-MS/MS analysis at 1 h or 4 h after the last injection ($N = 6$ mice/group). (B) Comparison of the route administration. MBL-II-141 at 10 mg/kg was administered by either IP injection or force-feeding. The results correspond to plasma levels observed 4 h after the administration ($N = 7$ mice/group). One value from the force-feeding treatment group was considered as an outlier by the R software, and so not considered in the analysis.

(HSA) in blood, suggesting a low unbound fraction in plasma that represents the active fraction on the target. All these data were in full agreement with our preliminary data indicating that MBL-II-141 was orally absorbed (Figure 4) and not transported (35).

DISCUSSION

Chemosensitization targeting ABCG2 has been widely explored using inhibitors, and other strategies such as the use of shRNA [34]. Its combination with

the anticancer agent 5-FluoroUracil reduced the size of tumors by 24% after a unique direct intratumoral injection. However, safety requirements of such a silencing system in clinical applications are complicated and under development. Thus, research of chemical molecules is still a valuable strategy. Our present study concerns the inhibitory effect of MBL-II-141 on ABCG2-mediated drug resistance to CPT-11. As expected, CPT-11 alone blocked the growth of ABCG2-negative xenografts, but only delayed the growth of ABCG2-positive xenografts. In contrast, its combination with MBL-II-141 is shown not

only to completely block the growth of freshly implanted ABCG2-positive tumors, but also to strongly slow down the growth of already established tumors. ABCG2 inhibition by MBL-II-141 in ABCG2-positive xenografts therefore plays a significant and synergistic role with CPT-11 anticancer activity. With an IC_{50} of 0.11 μ M, MBL-II-141 is known as one of the most powerful inhibitors ever described for ABCG2 (Table 2) [31, 32]. Its elevated IG_{50} (50% of cell growth inhibition), > 100 μ M, leads to a high *in vitro* therapeutic ratio [31], and is consistent with low toxicity for mice in a repeated-administration protocol (with 10 mg/kg MBL-II-141 in combination with 20 mg/kg CPT-11 for more than 100 days). Several ABCG2 inhibitors have been tested *in vivo*

(Table 2): some of them, such as GF120918 [35], Telatinib [26], Sorafenib [36] and YHO-13177 [37], have been described to delay the growth of small tumors in combination with doxorubicin or CPT-11. However, MBL-II-141 is clearly shown here to be the most effective, by completely inhibiting the growth of 90% tumors. In addition, its *in vivo* efficiency was validated at a relatively low dose (10 mg/kg), which also suggests a high inhibition activity in animal models. Moreover, MBL-II-141 is the first ABCG2 inhibitor ever described to significantly delay the growth of well-established tumors. Together with YHO-13177 [37] and Ko143 [24], MBL-II-141 is one of the very few ABCG2-specific inhibitors assayed in animal models.

Table 2: Comparison of ABCG2 inhibitors described in the literature for *in vivo* experiments. IC_{50} : Inhibitory concentration of 50% of the ABCG2 activity; IG_{50} : Inhibitory growth of 50% of the cell proliferation; I.V.: Intravenous; I.P.: Intraperitoneal; P.O.: Per Os.

Name	<i>In vitro</i>			<i>In vivo</i> Toxicity		<i>In vivo</i> Impact on		
	Inhibition (IC_{50})		ABCG2 Specificity	Cytotoxicity (IG_{50})	Acute	Chronic	Tumor growth	Pharmacokinetics
GF120918 [13, 21, 35, 45-48]	0.31 μ M	Mitoxantrone	No	1-40 mM	> 5 mg/kg I.V.	> 400 mg oral (human)	Yes	Doxorubicin : No
					> 10 mg/kg I.P.			Topotecan: Yes
					> 50 mg/kg oral			CPT-11: Yes
Ko143 [6, 24, 49]	0.22 μ M	Hoechst 3342	Yes	18-34 μ M	> 50 mg/kg oral			Topotecan: Yes
	0.35 μ M	Pheophorbide A			> 10 mg/kg I.P.			
Telatinib [26, 50, 51]		[3H]-Mitoxantrone [3H]-E217BG	No	> 10 μ M	>15 mg/kg oral	>1 500 mg (human)	Yes	
Triclabendazole [7]	\approx 10 μ M	Mitoxantrone			> 100 mg/kg I.P.			Nitrofurantoin: Yes
Sorafenib [25]			No	2-6 μ M	> 25 mg/kg oral	> 25 mg/kg P.O.	Yes	
YHO-13177 [40]	\approx 0.3 μ M	Hoechst 3342	Yes	> 10 μ M	> 200 mg/kg oral		Yes	
					> 30 mg/kg I.V.			
MBL-II-141	0.11 μ M	Mitoxantrone	Yes	> 100 μ M	> 10 mg/kg oral	> 10 mg/kg I.P.	Yes	CPT-11: Yes
					> 50 mg/kg I.P.	> 50 mg/kg I.P.		

It is indeed selective for ABCG2, and not known to interact with other regulatory functions. Compared to the above inhibitors, MBL-II-141 still remains the most powerful as suggested by its inhibition activity (Table 2) and the complete disappearance of 90% tumors observed in the present study.

In light of ABCG2-positive tumor sensitization and as required by preclinical development, we decided to define the pharmacokinetics parameters of MBL-II-141 and its possible interactions on CPT11 and SN-38 pharmacokinetics. The ABCG2 transporter plays a key role in protecting the brain parenchyma by catalyzing the efflux of xenobiotics from capillary endothelial cells at the BBB. It largely prevents the entry of its anticancer substrates at the BBB, thereby limiting their efficacy [20, 38]. In addition, Kawamura *et al.* (2013) recently demonstrated that ABCG2 strongly decreased CPT-11 distribution into the brain [21] by using labeled [¹⁴C]-CPT-11 to specifically track CPT-11, even in the presence of its SN-38 active metabolite. In complete agreement with these data, and despite high blood levels, we also found that CPT-11 displayed an apparently low distribution into the brain. However, similarly to GF120918 (5 mg/kg) [21], MBL-II-141 significantly increased CPT-11 levels into the brain. The SN-38 blood levels were strongly increased by MBL-II-141, but remained under the limit of detection into the brain. In comparison with Lin's work with normal FVB mice, intravenous injection of CPT-11 led to lower plasma levels than those observed here for our immunodeficient SCID mice. In contrast, the brain levels were higher in these normal strain mice, and SN-38 was consequently detectable [16]. In ICR mice, Yamazaki *et al.* described SN-38 plasma levels 27-fold lower than CPT-11, 1 h after the injection, [37] which is consistent with our difference of 19-fold between CPT-11 and SN-38 levels. Taken altogether, the results show a lower brain distribution in our mouse strain; nevertheless, CPT-11 distribution is increased and can be locally metabolized to SN-38 [39] by brain carboxylesterases [40]. The results suggested that MBL-II-141 positively influences SN-38 distribution into the brain and tumors, increasing its antitumor effect. Nevertheless, from the perspective of clinical use, the distribution of SN-38 has to be completed by evaluating the possible SN-38 accumulation in blood after reiterated administration of CPT-11 in combination with MBL-II-141. Our preliminary data did not show any significant change in SN-38 blood level after a 2 h treatment with CPT-11 and MBL-II-141 whether the animals were submitted to an acute or a chronic treatment (Supplementary figure 2). Furthermore, no side effects (loss of weight, modification of blood count) were observed during the several months of combination treatments.

Consequently, MBL-II-141 would be useful to increase the distribution of CPT-11 into the brain and, potentially, to locally treat gliomas since CPT-11 has proved its antitumor activity against MDR glioblastoma

cells [41]. This potential clinical development is also well supported by the fact that ABCG2 function at the BBE is relatively higher in human than in mouse. CPT-11 distribution may clearly be improved in the presence of MBL-II-141 in aggressive gliomas treatments. This is further supported by recent data showing that irinotecan in combination with temozolomide and bevacizumab produced promising results [41].

From the perspective of a putative pharmaceutical development, our preliminary data indicate that MBL-II-141 is absorbed after force-feeding to a lesser extent than through the IP route. MBL-II-141 bioavailability was predicted to be around 9% (Table 1), consistent with our data after force-feeding. As likely due to its high hydrophobicity, MBL-II-141 was predicted to strongly bind to human serum albumin (HSA) in blood, suggesting a low unbound active fraction in plasma. Even though reglementary guidelines generally do not require such knowledge in cancer use (for treating aggressive tumors with a low survival rate), the Percepta software predicted that MBL-II-141 is not genotoxic and not carcinogenic. In agreement with data from our pilot experiments, MBL-II-141 displays a low toxicity level during both acute and repeated-administration protocols. Nevertheless, several injections in a short period of time increased CPT-11 blood level. Indeed, two injections separated by 4 h did not change the blood level, while the third injection led to a slight increase. This may be due to an accumulative effect in the peritoneal cavity, due to the high hydrophobicity, and possible precipitation, of MBL-II-141. This last hypothesis was supported by the fact that, at 50 mg/kg for 3 times a week for 1 month, a slight deposit in the peritoneal cavity was observed (data not shown). Furthermore, no sign of toxicity (such as decrease of weight and modification of behavior) has ever been observed.

To conclude, by inhibiting the *in vivo* transport activity of ABCG2, MBL-II-141 has proved its efficiency of chemosensitization in xenograft models, and is therefore of interest for improving drug distribution into the brain. All observations and prediction data suggest a high mouse tolerance to MBL-II-141, and indicate that it would indeed be suitable for clinical development towards resistant cancer therapy or treatment of gliomas, which are limited by physiological barriers where the ABCG2 transporter plays a central role.

METHODS

Products

MBL-II-141 was synthesized by Prof. A. Boumendjel (UMR 5063, Grenoble, France) as previously described [31]. Corn oil (#C8267-500ML) and DMSO (#D2650-100ML) were purchased from Sigma Aldrich (France). Matrigel (#734-1100) was obtained from VWR. DMEM (#E15-883), penicillin/streptomycin (#P11-010) and G418

(#P11-012) were supplied by GE Healthcare Life Sciences (France). FBS (#ES98PS) was obtained from Abcys. CPT-11 and heparin were kindly provided by Dr. C. Rioufol (Hospices Civils of Lyon, France).

Cell lines

The human embryonic kidney (HEK) 293 cells were transfected as previously described [42]. HEK293 pcDNA3.1 (empty vector) and HEK ABCG2 (pcDNA3 vector coding for ABCG2) cell lines were cultured in DMEM (4.5 g/L glucose) supplemented with 10% FBS, 1% penicillin/streptomycin and 0.75 mg/mL G418. These cell lines are weakly tumorigenic [28].

Animal experiments

All *in vivo* experiments were conducted by accredited researchers (Drs. L. Payen and M. Honorat) in accordance with the animal care guidelines of the European Union and French laws, and were validated by the local Animal Ethic Evaluation Committee (CECCAPP). The investigation was conducted in accordance with the ethical standards and according to the Declaration of Helsinki and to national and international guidelines and has been approved by the authors' institutional review board. Four week-old female mice (Charles River Laboratories) were maintained in a specific pathogen-free animal facility AniCan at the Centre Léon Bérard (agreement n° B 69 388 0202, Lyon, France). Mice were randomly distributed into groups.

The experimental formulation of MBL-II-141 was determined according to its solubility properties. MBL-II-141 [31] was dissolved in DMSO (maximal solubility of 250 mg/mL), and 20-fold diluted in sterile corn oil [43, 44].

At the end of experiments, total cardiac blood was collected either from animals intraperitoneally treated with 50 U.I./kg heparin (50 μ L/mouse) for 1.5 h before withdrawal of blood samples, or using dried heparinized syringes to completely block clotting. Plasma was separated from blood cells by centrifugation (for 30 s at 2000 g). Blood and plasma were stored at -80 °C until drug quantification by LC-MS/MS.

MBL-II-141 tolerance assay

As recommended in the guidelines of The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), the pilot tolerance study was conducted by using three doses of MBL-II-141 in combination with the anticancer agent CPT-11. SCID mice were randomly distributed into three groups ($n = 10$) according to the MBL-II-141 doses: 10 mg/kg (low concentration), 30 mg/kg (intermediate concentration) or 50 mg/kg (maximal solubility). They were intraperitoneally treated three

times a week with MBL-II-141 and CPT-11 (20 mg/kg) for one month.

MBL-II-141 efficacy of chemosensitization on just-established tumors

4.10⁶ HEK-pcDNA3.1 or HEK-ABCG2 cells were sub-cutaneously injected into one flank of each SCID mouse, in solution with 50% matrigel. After the apparition of most tumors, mice were randomly distributed into 4 groups of 20 animals (10 ABCG2 negative and 10 ABCG2-positive tumors). In detail, group 1 received only vehicles (PBS and corn oil/5% DMSO); group 2 received only MBL-II-141 (10 mg/kg and PBS); group 3 received only CPT-11 (20 mg/kg and corn oil/5% DMSO) and group 4 received both molecules (MBL-II-141 at 10 mg/kg and CPT-11 at 20 mg/kg). Two independent injections were carried out per mouse.

Treatments were intraperitoneally administered three times a week until mice sacrifice. To evaluate the toxicity impacts of treatments, three times a week, mice were weighed and their behaviour was observed; tumor size was measured with a caliper in two perpendicular dimensions. Finally, mice were sacrificed when the tumor diameter reached 17 mm.

Efficacy of chemosensitization on well-established tumors

1.10⁶ HEK ABCG2 cells were sub-cutaneously injected into one flank of each SCID mouse, in solution with 30% matrigel. Once the tumor reached 8 mm in diameter (268 mm³ total volume), as previously described, mice were treated either by vehicles (PBS and corn oil/5% DMSO), MBL-II-141 (10 mg/kg and PBS), CPT-11 (20 mg/kg and corn oil/5% DMSO) or with the combination (MBL-II-141 at 10 mg/kg and CPT-11 at 20 mg/kg). Each group contained 10 mice. The same protocol of treatment and follow-up were applied.

MBL-II-141 pharmacokinetic study

SCID mice ($n = 6$ per group) were intraperitoneally treated by MBL-II-141 either alone (10 mg/kg) or in combination with CPT-11 (20 mg/kg). MBL-II-141 was administered for 0.5, 1, 2, 4 or 6 h. CPT-11 was administered either alone or in combination with MBL-II-141 for 0.5, 1, 2, 4, 8 and 24 h. In order to achieve possible cumulative effects, reiterated administrations (two or three) were carried out with a 4 h time interval between administrations. Blood samples were collected at either 1 h or 4 h after the last injection.

Total blood samples were collected using dried heparinized syringes, avoiding modification of samples. Blood and plasma were stored at -80°C. Brain and liver were collected and stored at -80°C for LC MS/MS quantification of MBL-II-141, CPT-11 and SN-38. Tissues were crushed using a dried cold system.

MBL-II-141 oral absorption study

The availability of MBL-II-141 was evaluated by force-feeding ($n = 4$), and compared to intraperitoneal injection ($n = 4$). CD1 mice were treated with 10 mg/kg of MBL-II-141. After 4 h of treatment, total blood was collected from animals treated with heparin. Samples were stored at -80°C until quantification.

HPLC-MS/MS analysis

CPT-11, SN-38 and MBL-II-141 were quantified in whole blood and in tissues by HPLC MS/MS (ThermoElectron, San Jose, USA). One assay was used for CPT-11 and SN-38 using gradient of elution with water-formic acid 0.1% and acetonitrile-formic acid 0.1%. Another gradient of elution was used for MBL-II-141 using ammonium acetate buffer (pH 6.5, 50 mM), acetonitrile and propanol-2. Both assays were performed on a C18 column HypersilGold, (ThermoElectron), 3.0 μm , (100 mm \times 2.1 mm i.d.). The mass spectrometry analysis was performed in positive ion mode with electrospray source. Labeled CPT-11, SN-38 and MBL-II-141 were used as internal standards. CPT-11 and $^{13}\text{C}_6$ -CPT-11 were quantitated in selected reaction monitoring (SRM) mode using the transition m/z 587 \rightarrow 124 and 593 \rightarrow 124, respectively. SN-38 and $^{13}\text{C}_6$ -SN-38 were quantitated in SRM mode using the transition m/z 393 \rightarrow 349 and 399 \rightarrow 355, respectively. MBL-II-141 and $^2\text{H}_3$ -MBL-II-141 were quantitated in SRM mode using the transition m/z 547 \rightarrow 342 and 550 \rightarrow 345, respectively. A volume of 50 μL of whole blood or tissue solution was introduced into a conic centrifuge and 50 μL of solution of the internal standard and 300 μL of cold acetonitrile were added. The sample was mixed and centrifuged for 10 min at 13000 g at 5°C . For MBL-II-141, 250 μL of the clear supernatant was removed and evaporated to dryness under a stream of nitrogen. The residues were resuspended in 100 μL of mobile phase and 10 μL were injected into the HPLC device. For CPT-11 and SN-38, 50 μL of clear supernatant were diluted (1/5) in mobile phase and 10 μL were injected into the chromatographic device. For MBL-II-141, calibration curves were from 5 to 1000 ng/mL for whole blood and tissues analysis. For CPT-11 and SN-38, calibration curves were from 50 to 5000 ng/mL for whole blood and liver and from 50 to 500 ng/mL for brain. When the concentration was over the upper standard concentration, a dilution of the sample was performed using blank whole blood or appropriate tissues. Peak area ratios of compound to the respective internal standard measured at each nominal concentration were used to construct weighted (1/concentration) least-square linear regression curves. For the three compounds, the within-run precision and the between-run precision of the assay was less than 10%. Assay accuracy was in the range of 91–108%. The lower limit of quantification of compounds in medium was set at the lowest standard concentration.

Results in tissues were normalized to cellular protein content. Results in tissues were normalized to cellular protein content (C1600 Abbott technology).

Prediction by ACD Percepta software

Predictions of some ADMET (absorption, distribution, metabolism, and excretion - toxicity) properties were computed using the ACD Percepta software, version 14.0.0 running on an intel i7 CPU laptop computer under windows 7. The default set and the ACD database was used without any modifications. The software uses proprietary set of QSAR (quantitative structure-activity relationship) equations to predict ADMET properties. The software is able to propose a reliable index (RI) made on structurally-related compounds found by a Tanimoto indice. The ACD standard library was used as database.

Statistical analysis

Growth models. Tumor growth curves were modeled with GraphPad[®] Software. Gaussian distribution models were obtained according to the equation $Y = \text{Amplitude} \cdot \exp(-0.5 \cdot ((X - \text{Mean}) / \text{SD})^2)$. The amplitude was the height of the center of the distribution in Y units. Mean was the X value at the center of the distribution. SD was a measure of the width of the distribution, in the same units as X. Another model was applied for sum of two Gaussian distributions with $Y = \text{one} + \text{two}$ ($\text{One} = \text{Amplitude} 1 \cdot \exp(-0.5 \cdot ((X - \text{Mean}1) / \text{SD}1)^2)$ and $\text{Two} = \text{Amplitude} 2 \cdot \exp(-0.5 \cdot ((X - \text{Mean}2) / \text{SD}2)^2)$). Amplitudes were the heights of the center of the distribution in Y units. Means were the X values at the center of the two distributions. SDs were measures of the widths of the distributions, in the same units as X.

Paired t-test. The statistical test used was the paired t-test. It was applied for all data of growth tumor experiments and pharmacokinetics data.

ACKNOWLEDGMENTS

We are grateful to the ethical committee for advice on *in vivo* protocols, and to Dr. Pierre Falson for critical discussion. We also thank Sarah Kabani for her help in reviewing the English of the manuscript.

This project was supported by the French National League against Cancer and an international grant from French ANR and Hungarian NIH (2010-INT-1101-01). C. G. was the recipient of a doctoral fellowship from the French National League against Cancer. The authors are grateful to the Agence Nationale pour la Recherche for financial support (Labex Arcanen ANR-11-LABX-0003-01).

Conflict of interest

The authors declare that there is no conflict of interest.

The funders had no role in study design, data collection and analysis, preparation of the manuscript.

REFERENCES

1. Gottesman, M.M., T. Fojo, S.E. Bates. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat Rev Cancer*, 2002; 2:48–58.
2. Amiri-Kordestani, L., A. Basseville, K. Kurdziel, A.T. Fojo, S.E. Bates. Targeting MDR in breast and lung cancer: discriminating its potential importance from the failure of drug resistance reversal studies. *Drug Resist Updat*, 2012; 15:50–61.
3. de Moraes, A.C., C.K. Maranhão, G.S. Rauber, M.C. Santos-Silva. Importance of detecting multidrug resistance proteins in acute leukemia prognosis and therapy. *J Clin Lab Anal*, 2013; 27:62–71.
4. Stacy, A.E., P.J. Jansson, D.R. Richardson. Molecular pharmacology of ABCG2 and its role in chemoresistance. *Mol Pharmacol*, 2013; 84:655–69.
5. Zhang, Q., K. Li, J.H. Xu, C.G. Zhao, Q. Gao, B. Wu, X.Y. Liu. Role of ABCG2 expression driven by cisplatin in platinum-containing chemotherapy for gastric cancer. *World J Gastroenterol*, 2013; 19:6630–6.
6. Zhou, S., J.D. Schuetz, K.D. Bunting, A.M. Colapietro, J. Sampath, J.J. Morris, I. Lagutina, G.C. Grosveld, M. Osawa, H. Nakauchi, B.P. Sorrentino. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat Med*, 2001; 7:1028–34.
7. Zhang, G., Z. Wang, W. Luo, H. Jiao, J. Wu, C. Jiang. Expression of Potential Cancer Stem Cell Marker ABCG2 is Associated with Malignant Behaviors of Hepatocellular Carcinoma. *Gastroenterol Res Pract*, 2013; 2013:782581.
8. Miyake, K., L. Mickley, T. Litman, Z. Zhan, R. Robey, B. Cristensen, M. Brangi, L. Greenberger, M. Dean, T. Fojo, S.E. Bates. Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone-resistant cells: demonstration of homology to ABC transport genes. *Cancer Res*, 1999; 59:8–13.
9. Brangi M, Litman T, Ciotti M, Nishiyama K, Kohlhagen G, Takimoto C, Robey R, Pommier Y, Fojo T, Bates S.E. Camptothecin resistance: role of the ATP-binding cassette (ABC), mitoxantrone-resistance half-transporter (MXR), and potential for glucuronidation in MXR-expressing cells. *Cancer Res*, 1999; 59:5938–46.
10. Kawabata, S., M. Oka, K. Shiozawa, K. Tsukamoto, K. Nakatomi, H. Soda, M. Fukuda, Y. Ikegami, K. Sugahara, Y. Yamada, S. Kamihira, L.A. Doyle, D.D. Ross, et al. Breast cancer resistance protein directly confers SN-38 resistance of lung cancer cells. *Biochem Biophys Res Commun*, 2001; 280:1216–23.
11. Tang, S.C., N. de Vries, R.W. Sparidans, E. Wagenaar, J.H. Beijnen, A.H. Schinkel. Impact of P-glycoprotein (ABCB1) and breast cancer resistance protein (ABCG2) gene dosage on plasma pharmacokinetics and brain accumulation of dasatinib, sorafenib, sunitinib. *J Pharmacol Exp Ther*, 2013; 346:486–94.
12. Takashima, T., C. Wu, M. Takashima-Hirano, Y. Katayama, Y. Wada, M. Suzuki, H. Kusuhabara, Y. Sugiyama, Y. Watanabe. Evaluation of breast cancer resistance protein function in hepatobiliary and renal excretion using PET with ¹¹C-SC-62807. *J Nucl Med*, 2013; 54:267–76.
13. Jonker, J.W., J.W. Smit, R.F. Brinkhuis, M. Maliepaard, J.H. Beijnen, J.H. Schellens, A.H. Schinkel. Role of breast cancer resistance protein in the bioavailability and fetal penetration of topotecan. *J Natl Cancer Inst*, 2000; 92:1651–6.
14. Jonker, J.W., M. Buitelaar, E. Wagenaar, M.A. Van Der Valk, G.L. Scheffer, R.J. Scheper, T. Plosch, F. Kuipers, R.P. Elferink, H. Rosing, J.H. Beijnen, A.H. Schinkel. The breast cancer resistance protein protects against a major chlorophyll-derived dietary phototoxin and protoporphyria. *Proc Natl Acad Sci U S A*, 2002; 99:15649–54.
15. de Vries, N.A., J. Zhao, E. Kroon, T. Buckle, J.H. Beijnen, O. van Tellingen. P-glycoprotein and breast cancer resistance protein: two dominant transporters working together in limiting the brain penetration of topotecan. *Clin Cancer Res*, 2007; 13:6440–9.
16. Lin, F., S. Marchetti, D. Plum, D. Iusuf, R. Mazzanti, J.H. Schellens, J.H. Beijnen, O. van Tellingen. Abcc4 together with abcb1 and abcg2 form a robust cooperative drug efflux system that restricts the brain entry of camptothecin analogues. *Clin Cancer Res*, 2013; 19:2084–95.
17. Vlaming, M.L., A. van Esch, Z. Pala, E. Wagenaar, K. van de Wetering, O. van Tellingen, A.H. Schinkel. Abcc2 (Mrp2), Abcc3 (Mrp3), and Abcg2 (Bcrp1) are the main determinants for rapid elimination of methotrexate and its toxic metabolite 7-hydroxymethotrexate *in vivo*. *Mol Cancer Ther*, 2009; 8:3350–9.
18. Yang J.J, Milton M.N, Yu S, Liao M, Liu N, Wu J.T, Gan L, Balani S.K, Lee F.W, Prakash S, Xia C.Q. P-glycoprotein and breast cancer resistance protein affect disposition of tandutinib, a tyrosine kinase inhibitor. *Drug Metab Lett*, 2010; 4:201–12.
19. Marchetti, S., N.A. de Vries, T. Buckle, M.J. Bolijn, M.A. van Eijndhoven, J.H. Beijnen, R. Mazzanti, O. van Tellingen, J.H. Schellens. Effect of the ATP-binding cassette drug transporters ABCB1, ABCG2, and ABCC2 on erlotinib hydrochloride (Tarceva) disposition in *in vitro* and *in vivo* pharmacokinetic studies employing Bcrp1^{-/-}/Mdr1a/1b^{-/-} (triple-knockout) and wild-type mice. *Mol Cancer Ther*, 2008; 7:2280–7.
20. Bakhsheshian, J., B.R. Wei, K.E. Chang, S. Shukla, S.V. Ambudkar, R.M. Simpson, M.M. Gottesman, M.D. Hall. Bioluminescent imaging of drug efflux at the blood-brain barrier mediated by the transporter ABCG2. *Proc Natl Acad Sci U S A*, 2013; 110:20801–6.

21. Kawamura, K., H. Hashimoto, M. Ogawa, J. Yui, H. Wakizaka, T. Yamasaki, A. Hatori, L. Xie, K. Kumata, M. Fujinaga, M.R. Zhang. Synthesis, metabolite analysis, and *in vivo* evaluation of [(11)C]irinotecan as a novel positron emission tomography (PET) probe. *Nucl Med Biol*, 2013; 40:651–7.
22. Szakacs, G., J.K. Paterson, J.A. Ludwig, C. Booth-Genthe, M.M. Gottesman. Targeting multidrug resistance in cancer. *Nat Rev Drug Discov*, 2006; 5:219–34.
23. Rabindran, S.K., H. He, M. Singh, E. Brown, K.I. Collins, T. Annable, L.M. Greenberger. Reversal of a novel multidrug resistance mechanism in human colon carcinoma cells by fumitremorgin C. *Cancer Res*, 1998; 58:5850–8.
24. Allen J.D, van Loevezijn A, Lakhai J.M, van der Valk M, van Tellingen O, Reid G, Schellens J.H, Koomen G.J, Schinkel A.H. Potent and specific inhibition of the breast cancer resistance protein multidrug transporter *in vitro* and in mouse intestine by a novel analogue of fumitremorgin C. *Mol Cancer Ther*, 2002; 1:417–25.
25. Marighetti, F., K. Steggemann, M. Hanl, M. Wiese. Synthesis and quantitative structure-activity relationships of selective BCRP inhibitors. *ChemMedChem*, 2013; 8:125–35.
26. Sodani, K., A. Patel, N. Anreddy, S. Singh, D.H. Yang, R.J. Kathawala, P. Kumar, T.T. Talele, Z.S. Chen. Telatinib reverses chemotherapeutic multidrug resistance mediated by ABCG2 efflux transporter *in vitro* and *in vivo*. *Biochem Pharmacol*, 2014; 89:52–61.
27. Boumendjel, A., S. Macalou, A. Ahmed-Belkacem, M. Blanc, A. Di Pietro. Acridone derivatives: design, synthesis, and inhibition of breast cancer resistance protein ABCG2. *Bioorg Med Chem*, 2007; 15:2892–7.
28. Arnaud, O., A. Boumendjel, A. Geze, M. Honorat, E.L. Matera, J. Guitton, W.D. Stein, S.E. Bates, P. Falson, C. Dumontet, A. Di Pietro, L. Payen. The acridone derivative MBLL-87 sensitizes breast cancer resistance protein-expressing xenografts to irinotecan. *Eur J Cancer*, 2011; 47:640–8.
29. Sostelly A, Payen L, Guitton J, Di Pietro A, Falson P, Honorat M, Boumendjel A, Geze A, Freyer G, Tod M. Quantitative evaluation of the combination between cytotoxic drug and efflux transporter inhibitors based on a tumour growth inhibition model. *Fundam Clin Pharmacol*, 2014; 28:161–9.
30. Sostelly, A., L. Payen, J. Guitton, A. Di Pietro, P. Falson, M. Honorat, G. Valdameri, A. Geze, A. Boumendjel, G. Freyer, M. Tod. A template model for studying anticancer drug efflux transporter inhibitors *in vitro*. *Fundam Clin Pharmacol*, 2013; 27:544–56.
31. Valdameri, G., E. Genoux-Bastide, B. Peres, C. Gauthier, J. Guitton, R. Terreux, S.M. Winnischofer, M.E. Rocha, A. Boumendjel, A. Di Pietro. Substituted chromones as highly potent nontoxic inhibitors, specific for the breast cancer resistance protein. *J Med Chem*, 2012; 55:966–70.
32. Winter, E., F. Lecerf-Schmidt, G. Gozzi, B. Peres, M. Lightbody, C. Gauthier, C. Ozvegy-Laczka, G. Szakacs, B. Sarkadi, T.B. Crezynski-Pasa, A. Boumendjel, A. Di Pietro. Structure-activity relationships of chromone derivatives toward the mechanism of interaction with and inhibition of breast cancer resistance protein ABCG2. *J Med Chem*, 2013; 56:9849–60.
33. Valdameri, G., L. Pereira Rangel, C. Spatafora, J. Guitton, C. Gauthier, O. Arnaud, A. Ferreira-Pereira, P. Falson, S.M. Winnischofer, M.E. Rocha, C. Tringali, A. Di Pietro. Methoxy stilbenes as potent, specific, untransported, and noncytotoxic inhibitors of breast cancer resistance protein. *ACS Chem Biol*, 2012; 7:322–30.
34. Xie, N., L. Mou, J. Yuan, W. Liu, T. Deng, Z. Li, Y. Jin, Z. Hu. Modulating Drug Resistance by Targeting BCRP/ABCG2 Using Retrovirus-Mediated RNA Interference. *PLoS One*, 2014; 9:e103463.
35. Hyafil, F., C. Vergely, P. Du Vignaud, T. Grand-Perret. *In vitro* and *in vivo* reversal of multidrug resistance by GF120918, an acridonecarboxamide derivative. *Cancer Res*, 1993; 53:4595–602.
36. Mazard, T., A. Causse, J. Simony, W. Leconet, N. Vezzio-Vie, A. Torro, M. Jarlier, A. Evrard, M. Del Rio, E. Assenat, P. Martineau, M. Ychou, B. Robert, et al. Sorafenib overcomes irinotecan resistance in colorectal cancer by inhibiting the ABCG2 drug-efflux pump. *Mol Cancer Ther*, 2013; 12:2121–34.
37. Yamazaki, R., Y. Nishiyama, T. Furuta, H. Hatano, Y. Igarashi, N. Asakawa, H. Kodaira, H. Takahashi, R. Aiyama, T. Matsuzaki, N. Yagi, Y. Sugimoto. Novel acrylonitrile derivatives, YHO-13177 and YHO-13351, reverse BCRP/ABCG2-mediated drug resistance *in vitro* and *in vivo*. *Mol Cancer Ther*, 2011; 10:1252–63.
38. Bihorel, S., G. Camenisch, M. Lemaire, J.M. Schermann. Influence of breast cancer resistance protein (Abcg2) and p-glycoprotein (Abcb1a) on the transport of imatinib mesylate (Gleevec) across the mouse blood-brain barrier. *J Neurochem*, 2007; 102:1749–57.
39. Sparreboom, A., M.J. de Jonge, P. de Bruijn, E. Brouwer, K. Nooter, W.J. Loos, R.J. van Alphen, R.H. Mathijssen, G. Stoter, J. Verweij. Irinotecan (CPT-11) metabolism and disposition in cancer patients. *Clin Cancer Res*, 1998; 4:2747–54.
40. Hosokawa, M. Structure and catalytic properties of carboxylesterase isozymes involved in metabolic activation of prodrugs. *Molecules*, 2008; 13:412–31.
41. Vredenburg, J.J., A. Desjardins, D.A. Reardon, H.S. Friedman. Experience with irinotecan for the treatment of malignant glioma. *Neuro Oncol*, 2009; 11:80–91.
42. Robey, R.W., Y. Honjo, K. Morisaki, T.A. Nadjem, S. Runge, M. Risbood, M.S. Poruchynsky, S.E. Bates. Mutations at amino-acid 482 in the ABCG2 gene affect substrate and antagonist specificity. *Br J Cancer*, 2003; 89:1971–8.

43. Badham, H.J., D.P. LeBrun, A. Rutter, L.M. Winn. Transplacental benzene exposure increases tumor incidence in mouse offspring: possible role of fetal benzene metabolism. *Carcinogenesis*, 2010; 31:1142–8.
44. Gin, D.W., H.R. Kim, D.M. Lee, S.H. Lee, A.R. Seong, J.A. Hwang, J.H. Park. 2,3,7,8-Tetrachlorodibenzo-p-dioxin up-regulates stathmin 1 in the eye: suggestive evidence for the involvement of stathmin 1 in accelerated eye opening in mice induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Lett*, 2010; 196:87–94.
45. de Bruin, M., K. Miyake, T. Litman, R. Robey, S.E. Bates. Reversal of resistance by GF120918 in cell lines expressing the ABC half-transporter, MXR. *Cancer Lett*, 1999; 146:117–26.
46. Kruijtzter, C.M., J.H. Beijnen, H. Rosing, W.W. ten Bokkel Huinink, M. Schot, R.C. Jewell, E.M. Paul, J.H. Schellens. Increased oral bioavailability of topotecan in combination with the breast cancer resistance protein and P-glycoprotein inhibitor GF120918. *J Clin Oncol*, 2002; 20:2943–50.
47. Ahmed-Belkacem, A., A. Pozza, F. Munoz-Martinez, S.E. Bates, S. Castanys, F. Gamarro, A. Di Pietro, J.M. Perez-Victoria. Flavonoid structure-activity studies identify 6-prenylchrysin and tectochrysin as potent and specific inhibitors of breast cancer resistance protein ABCG2. *Cancer Res*, 2005; 65:4852–60.
48. Kuppens, I.E., E.O. Witteveen, R.C. Jewell, S.A. Radema, E.M. Paul, S.G. Mangun, J.H. Beijnen, E.E. Voest, J.H. Schellens. A phase I, randomized, open-label, parallel-cohort, dose-finding study of elacridar (GF120918) and oral topotecan in cancer patients. *Clin Cancer Res*, 2007; 13:3276–85.
49. Zander, S.A., W. Sol, L. Greenberger, Y. Zhang, O. van Tellingen, J. Jonkers, P. Borst, S. Rottenberg. EZN-2208 (PEG-SN38) overcomes ABCG2-mediated topotecan resistance in BRCA1-deficient mouse mammary tumors. *PLoS One*, 2012; 7:e45248.
50. Eskens, F.A., N. Steeghs, J. Verweij, J.L. Bloem, O. Christensen, L. van Doorn, J. Ouwkerk, M.J. de Jonge, J.W. Nortier, J. Kraetzschmar, P. Rajagopalan, H. Gelderblom. Phase I dose escalation study of telatinib, a tyrosine kinase inhibitor of vascular endothelial growth factor receptor 2 and 3, platelet-derived growth factor receptor beta, and c-Kit, in patients with advanced or metastatic solid tumors. *J Clin Oncol*, 2009; 27:4169–76.
51. Mross, K., A. Frost, M.E. Scheulen, J. Krauss, D. Strumberg, B. Schultheiss, U. Fasol, M. Buchert, J. Kratzschmer, H. Delesen, P. Rajagopalan, O. Christensen. Phase I study of telatinib (BAY 57-9352): analysis of safety, pharmacokinetics, tumor efficacy, and biomarkers in patients with colorectal cancer. *Vasc Cell*, 2011; 3:16.

Travail collaboratif 2

« Quantification and kinetic study in plasma and tissues of (E)-1,1,4,4-tétraméthyl-2-tétrazène, a liquid propellant and a transformation product of 1,1-dimethyl hydrazine »

Payen L., Honorat M., Bouard C., Jacob G., Terreux R., Delalu H., Labarthe E., Guitton J.

Analytical and Bioanalytical Chemistry, juin 2015

2 Travail collaboratif 2 : Etude de la tolérance et de la toxico-cinétique d'un produit à usage industriel (TMTZ)

Les propergols utilisant les dérivés d'hydrazine sont très utilisés dans la défense et l'aérospatiale. Aujourd'hui le monométhylhydrazine (MMH) et le 1,1-diméthylhydrazine (UDMH) sont les principales molécules utilisées à grande échelle pour l'industrielle spatiale. Cependant le MMH est génotoxique et reste longtemps dans l'environnement et l'UDMH, classé groupe 2B par le CIRC, augmenterait la fréquence des souris et des rats à développer un cancer. Afin d'éviter l'utilisation de ces produits dangereux, le (E)-1,1,4,4-tétraméthyl-2-tétrazène (TMTZ) est envisagé comme un nouveau propulseur pertinent pour l'espace. Le TMTZ a également été détecté parmi les nombreux composés formés par la décomposition de l'UDMH.

Dans cette présente étude, nous avons évalué le niveau d'exposition et la toxicité *in vivo* du TMTZ chez la souris. Pour cela nous avons développé une méthode analytique de quantification du TMTZ dans différents milieux biologiques. La quantification du TMTZ extrait de la matrice biologique (50 µl de plasma ou de solution tissulaire) par précipitation protéique avec de l'acétonitrile froid a été réalisée par chromatographie liquide couplée en tandem à la spectrométrie de masse (LC-MS/MS).

Des études de la toxicité du TMTZ sont importantes pour déterminer les mesures de protection nécessaires pour empêcher les travailleurs d'être exposés à des doses nocives lors de la préparation des lanceurs. Une étude *in silico* a simulé la concentration plasmatique et les paramètres pharmacocinétiques du TMTZ. Afin d'approfondir les connaissances sur la toxicité du TMTZ, la pharmacocinétique et la distribution tissulaire de ce composé ont été étudiées de manière expérimentale chez la souris suite à des injections intra péritonéales. Ainsi le but de cette étude est de décrire une méthode permettant la quantification du TMTZ dans les milieux biologiques.

L'étude de tolérance a été menée pendant 14 jours sur 12 souris traitées avec des doses de TMTZ de 0 à 300 mg/kg. Il n'a pas été observé de toxicité aiguë (Pas de modifications de leur comportement). Lors de l'administration de 300 mg/kg de TMTZ, après 24h d'exposition, le TMTZ était sous la limite de détection ce qui suggère une demi vie courte du TMTZ. Les études de la distribution tissulaire montrent que le TMTZ diffuse rapidement dans certains tissus et organes. Des concentrations élevées ont été retrouvées dans le foie et dans les reins montrant donc l'implication de ces deux organes dans l'élimination du TMTZ. D'autre part des concentrations élevées de TMTZ ont été retrouvés dans le cerveau montrant donc que ce dernier passe la barrière hémato-épithéliale facilement suggérant une certaine neurotoxicité.

Certains résultats expérimentaux ne sont pas en accord avec les prédictions *in silico* des paramètres pharmacocinétiques (vitesse élimination et toxicité du TMTZ). Ainsi ces premières

données expérimentales permettent d'enrichir les connaissances sur la toxicité du TMTZ. Cependant d'autres expériences devront être réalisées afin de compléter et d'optimiser les règles de sécurité et d'utilisation de ce composé qui pourrait être utilisé comme un nouveau propulseur dans le secteur de l'aérospatial.

Quantification and kinetic study in plasma and tissues of (*E*)-1,1,4,4-tetramethyl-2-tetrazene, a liquid propellant and a transformation product of 1,1-dimethyl hydrazine

Léa Payen^{1,2,3} · Mylène Honorat^{2,3,4} · Charlotte Bouard² · Guy Jacob⁵ · Raphael Terreur^{3,4} · Henri Delalu⁵ · Emilie Labarthe⁶ · Jérôme Guitton^{1,3}

Received: 21 March 2015 / Revised: 26 May 2015 / Accepted: 5 June 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract (*E*)-1,1,4,4-tetramethyl-2-tetrazene (TMTZ) is formed from the oxidation of the unsymmetrical 1,1-dimethylhydrazine (UDMH) and is used as a storable liquid fuel which can be considered as a new potential propellant for space rocket propulsion. To better understand the toxicological behavior of the compound, an intraperitoneal administration of TMTZ was performed in mice to define its toxicokinetics and tissue distribution. A fully validated liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) assay was developed to determine TMTZ levels in biological samples. Determination of TMTZ was achieved using 50 μ L of plasma or tissue solution. Precipitation with ammonium sulfate and acetonitrile was used for sample preparation. Liquid chromatography was performed on an Atlantis HILIC Silica column (Waters; 3 μ m, 150 mm \times 2.1 mm i.d.). Isocratic elution with a mixture of ammonium acetate buffer (pH 5, 100 mM)/water/acetonitrile (3:2:95, *w/v/v*) was used.

The detection was conducted using an electrospray source in positive ion mode. TMTZ and ¹⁵N₂-TMTZ (internal standard) were quantitated in selected reaction monitoring mode using the transition *m/z* 117 \rightarrow 72 and 119 \rightarrow 74, respectively. Standard curves exhibited excellent linearity in the range of 10–500 ng/mL for plasma and 50–2000 ng/mL for all tissues (heart, liver, brain, kidney, and lung) analyzed, and acceptable precision and accuracy (<10 %) were obtained. The elimination rate constant strongly suggests that TMTZ was very quickly eliminated from the body. The results of tissue distribution experiments indicated that TMTZ underwent a rapid distribution into limited organs such as the liver, kidney, and brain.

Keywords TMTZ · Alkyltetrazene · Liquid propellant · Tissues distribution · Liquid chromatography–mass spectrometry

✉ Jérôme Guitton
jerome.guitton@univ-lyon1.fr

¹ Centre Hospitalier Lyon-Sud, Laboratoire de Biochimie-Toxicologie, Hospices Civils de Lyon, 69495 Pierre-Bénite, France

² Centre de Recherche en Cancérologie de Lyon, Centre Léon Bérard, UMR INSERM U1052/CNRS 5286, 69373 Lyon, France

³ ISPBL, Faculté de Pharmacie, Laboratoire de Toxicologie, Université Lyon 1, 8 Avenue Rockefeller, 69373 Lyon, France

⁴ IBCP, BMSSI UMR CNRS 5086, Université Lyon 1, 7 Passage du Vercors, 69373 Lyon, France

⁵ Faculté des Sciences et Technologies, UMR CNRS 5278 Hydrazines et Composés Energetiques Polyazotés, Université Lyon 1, 69622 Lyon, France

⁶ CNES Launcher Directorate, 52 Rue Jacques Hillairet, 75612 Paris, France

Introduction

Propellants using hydrazine derivatives are widely used in the aerospace and defense industries. As a bi-propellant with nitrogen tetroxide, monomethylhydrazine (MMH) is used in storable propellant engines for launchers and in apogee motors of satcoms. It has also been used in the Automated Transfer Vehicle (ATV), the European cargo that supplied the International Space Station, and also raised its orbit by 30 km per year. As a mono-propellant, hydrazine (N₂H₄) is used for launcher trajectory correction and for the propulsion of small satellites. Hydrazine compounds are perfectly adapted for applications which need to be both reliable and accurate, such as approach maneuvers in space. At present, MMH and the unsymmetrical 1,1-dimethylhydrazine (UDMH) are the principal molecules used on a large industrial scale for spatial

applications. Unfortunately, MMH is toxic and persists long term in the environment. It displays genotoxic properties (methyl adducts formation, Ames test is positive in several strains of bacteria, and micronuclei increases) [1–3]. In addition, exposure to 0.05–0.5 ppm UDMH for acute- or intermediate-duration exposure is known to increase frequency of mice and rat cancers [4]. In consequence, it was classified as a carcinogenic compound by IARC (group 2B) [1] and appropriated manufacturing and handling processes are applied to prevent any issues. For several years, great effort has been devoted to develop a less hazardous molecule allowing to alleviate these security measures. These investigations result in the identification of (*E*)-1,1,4,4-tetramethyl-2-tetrazene (TMTZ) as a storable liquid fuel which could be considered as a new potential relevant propellant for space propulsion [5]. TMTZ was also detected among the numerous compounds formed from the decomposition of UDMH, a widely used rocket propellant [4, 6, 7]. Possible atmospheric dispersion of UDMH and various decomposition products was studied using an *in silico* approach and also after analysis of soil samples [4, 7]. TMTZ was clearly identified in laboratory samples [7]. However, due to its volatility, TMTZ could only be detected in ground samples if sampling was performed immediately after spillage of UDMH [7].

If TMTZ were to be exploited as a useful compound for space propulsion, workers would potentially be exposed during the various stages of synthesis, including transport and rocket re-fuelling. Thus, data on TMTZ toxicity should be available to determine the necessary protection measures to prevent workers from being exposed to harmful doses. Moreover, as transformation products of UDMH, TMTZ also has to be taken into account in the risk assessment.

Toxicity response is often directly related to the target tissue concentration of the chemical. The risk usually decreases with a short-time exposure. Thus, the pharmacokinetic and the tissue distribution largely contribute to enhance the knowledge on toxicological behavior of a compound. Plasma concentration and pharmacokinetic parameters were available for TMTZ from an *in silico* study [6]. They addressed absorption, distribution, metabolism, and excretion features as derived by ADME Boxes (Pharma Algorithms 1) as well as toxicological effects derived by Tox Boxes (Pharma Algorithms 1, 2008) and by PASS (PASS1, 2008) [6]. The goal of the present work was to experimentally study for the first time the pharmacokinetic and the tissue distribution of TMTZ in mice. For this purpose, intraperitoneal administration of TMTZ was performed in mice and plasma and tissue concentrations were measured. Previous assays were used to analyze TMTZ from water samples using ion-pair chromatography coupled with amperometric or spectrophotometric detectors [8, 9]. Kenessov et al. developed a gas chromatography-mass spectrometry (GC-MS) method using headspace solid-phase microextraction (SPME) for detection of TMTZ and other

transformation products in soil and water [10]. To enhance performance and to create a simple, short, and sensitive assay, we developed a liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) method to determine for the first time TMTZ levels in biological samples. The method was validated in terms of selectivity, sensitivity, accuracy, precision, and recovery.

Material and method

Reagents and chemicals

TMTZ and $^{15}\text{N}_2$ -TMTZ, used as internal standard, were synthesized by the Hydrazine and Polynitrogen Energetic Compounds Laboratory (UMR CNRS 5278, Lyon, France). TMTZ was stored at +4 °C in the dark with a slight nitrogen stream for its conservation. For the internal standard, the substitution $^{14}\text{N}/^{15}\text{N}$ occurs on the nitrogen in positions 2 and 3 and the isotopic enrichment is over to 98.5 %. Dimethyl sulfide, ammonium sulfate, and acetonitrile (HPLC gradient grade) were from Sigma-Aldrich (Saint-Quentin-Fallavier, France). Acetic acid (Ultra, >99.5 % pure) and ammoniac (30 %) were provided by Carlo Erba (Milan, Italy). Milli-Q deionized water was used throughout the study. Drug-free normal plasma was provided by the regional blood bank (EFS Rhône-Alpes, France).

Standard solution and calibration

Stock solutions of TMTZ (1 mg/mL) were prepared in acetonitrile and stored at +4 °C. Standard solutions were obtained daily by successive dilutions in acetonitrile. Drug-free normal human plasma was spiked with the appropriate diluted solutions to give final concentrations of calibration standards at 10, 25, 50, 100, 200, and 500 ng/mL. In addition, quality control (QC) samples were made at 20, 75, and 300 ng/mL from these standard solutions. To validate the dilution procedure performed from mouse samples when the concentration was over 500 ng/mL, solutions of TMTZ at 4000 and 20,000 ng/mL in plasma were diluted by 1/10 and 1/50, respectively, to obtain a final concentration of 400 ng/mL. For the quantification of TMTZ in tissues (see preparation below), TMTZ-free control mouse tissue (brain, kidney, liver, lung, and heart) was spiked with the appropriate diluted solutions to give final concentrations of calibration standards at 50, 100, 200, 1000, and 2000 ng/mL. Plasma mouse samples with TMTZ concentrations greater than the upper calibration standard were diluted with blank human plasma prior to extraction. The blank plasma for the dilution was the same lot used to prepare calibration standards. For tissue analysis, the respective tissues from control group were used for the dilution.

Chromatographic and mass spectrometry conditions

The quantification of TMTZ was performed on a liquid chromatography coupled with triple–quadrupole mass spectrometer (Quantum Ultra; Thermo Electron, San Jose, USA). TMTZ was analyzed on an Atlantis HILIC Silica column (Waters, Milford, USA; 3 μm , 150 mm \times 2.1 mm i.d.). Isocratic elution with a mixture of ammonium acetate buffer (pH 5, 100 mM)/water/acetonitrile (3:2:95, v/v/v) was used. The total run time was fixed at 5 min. The column oven temperature was maintained at 25 $^{\circ}\text{C}$, and the mobile phase flow rate was 200 $\mu\text{L}/\text{min}$. The mass spectrometer operated in positive ion mode with electrospray ionization (ESI) source. Argon was used as collision gas at 1.5 mTorr. Spray voltage and capillary temperature were set at 3.5 kV and 300 $^{\circ}\text{C}$, respectively. TMTZ and $^{15}\text{N}_2$ -TMTZ were quantitated in selected reaction monitoring mode using the transition m/z 117 \rightarrow 72 and 119 \rightarrow 74, respectively.

Calibration curve and validation procedure

Calibration curves were constructed by plotting the ion abundance peak area ratio as a function of plasma TMTZ or tissue concentration. Data were fitted by weighted (one/concentration) for least squares regression, and standard curves were determined by linear regression analysis. A total of six calibration curves (prepared as a single replicate and analyzed on six different days) were generated throughout the entire validation process using freshly prepared samples from plasma. Calibration curves were also generated from tissue preparations. To determine the inter-day precision and the accuracy, three runs, including a calibration curve and QC at three concentrations in six replicates, were performed. The limit of quantitation (LOQ) was chosen as the concentration with a limit of variance within 20 % for both precision and accuracy. The freeze-thaw stability of TMTZ in plasma was tested in triplicate on the three QC following three freeze-thaw cycles at -80°C . Extraction recovery was measured on the three QC analyzed on two different days.

Sample preparation

All the samples were kept on ice during the sample preparation procedure. A volume of 50 μL of plasma or tissue solution was introduced in a conic centrifuge tube containing 50 mg of ammonium sulfate. The sample was mixed, and 50 μL of the solution of the internal standard and 400 μL of cold acetonitrile were added. After protein precipitation, the sample was mixed and centrifuged for 10 min at 13,000g at 5 $^{\circ}\text{C}$. Then, the clear supernatant was transferred to a glass vial. Finally, the glass vial was kept at 5 $^{\circ}\text{C}$ in an autosampler and 10 μL was injected into a chromatographic device.

Safety and kinetic studies

All in vivo experiments were conducted by accredited researchers (Dr. L. Payen and Dr. M. Honorat) in accordance with the animal care guidelines of the European Union and French laws and were validated by the local Animal Ethic Evaluation Committee (CECCAPP). Four-week-old female mice (Charles River Laboratories) were maintained in a specific pathogen-free animal facility, AniCan, at the Centre Léon Bérard (agreement no. B 69 388 0202; Lyon, France). Mice were randomly distributed into groups. The assay was applied to safety and kinetic studies of TMTZ in CD1 mice (Charles River Laboratories). The mice were housed and maintained under controlled environmental conditions with free access to food and drinking water until 4 h prior to experiments.

TMTZ was dissolved in a mixture of DMSO/NaCl (0.9 %) (30:70, v/v) solution. Three hundred microliters of the TMTZ mixture was intraperitoneally administered. Table 1 presents the design of safety and kinetic studies. Protocols included a control group (dose 0 or time 0), which was only treated with vehicles [DMSO/NaCl (0.9 %) (30:70, v/v)]. In the safety 1 study, the mice were examined for symptoms of toxicity every hour for the first 12 h. In the safety 2 study, once daily, the mice were examined for pains or symptoms of toxicity, which would require ethical sacrifice, as set out in the ethical procedure of animal health care (local Animal Ethic Evaluation Committee (CECCAPP)). All findings were recorded. All animals were weighed using an electronic balance on the date of acquisition and on the first day of the experimental period, and then the mice were weighed once every 3 days.

At the end of the three independent experiments, total cardiac blood was collected from animals intraperitoneally treated with 50 UI/kg heparin (50 $\mu\text{L}/\text{mouse}$) 1.5 h before the blood samples to completely block clotting. Plasma was separated from blood cells by centrifugation (30 s at 2000g). Blood and plasma were stored at -80°C until drug quantification by LC-MS/MS. Several tissues including brain, kidney, liver, lung, and heart were collected and stored at -80°C . The preparation of tissues was performed as follows: tissue was mixed in a volume of deionized water from 0.5 to 1.5 mL according to the weight of the organ then freeze-ground using a Cryonext freeze grinder (Saint-Gély-du-Fesc, France). Tissues were dried cold grinding before TMTZ quantification, and total protein was quantified using the Bradford method and expressed in milligrams per milliliter.

Data for the plasma kinetic study was treated by the non-linear least square analysis which was performed using “nls” function in the R statistical software (2.14.1 version, <http://cran.r-project.org/>).

Table 1 Design of safety and kinetic studies

Study	Doses of TMTZ (mg/kg)	Exposure	Animals
Safety 1	0, 75, 150 and 300	24 h	<i>n</i> =6 per dose
Safety 2	0–300	14 days	<i>n</i> =6 per dose
Kinetic	300	0, 30, 60, 120, 180, and 300 min	<i>n</i> =5 per time

Results

Quantification by LC-MS/MS

TMTZ was extracted from the biological matrix by simple protein precipitation with cold acetonitrile. A concentration step, with evaporation of the liquid phase under nitrogen and reconstitution of dry residue by low volume of mobile phase, was not an option due to TMTZ volatility. Chromatographic separation performed with HILIC column produces a good peak shape for TMTZ. Matrix components from plasma or tissues did not interfere with the analysis (Fig. 1). A retention time for TMTZ of 2.3 min was found to be sufficient to analyze numerous samples within a single run. Protonated molecule ($[M+H]^+$) was predominantly formed with TMTZ in the mobile phase and was selected as the precursor ion. Mass spectrum exhibited two daughter ions at m/z 72 and 44 (Fig. 2). For $^{15}N_2$ -TMTZ (internal standard), two daughter ions were observed at m/z 74 and 44, in agreement with the fragmentation pathway and the labeling of the compound. The MS/MS parameters were optimized to maximize the response for TMTZ parent/daughter ion of m/z 117→72 and $^{15}N_2$ -TMTZ of m/z 119→74. In the present condition of mobile phase, no signal for TMTZ was obtained in negative mode. The calibration curves were obtained by weighted one/[concentration] least squares linear regression analysis of TMTZ concentration versus peak area ratio of TMTZ/ $^{15}N_2$ -TMTZ area ratios. Standard curves exhibited excellent linearity in the range of 10–500 ng/mL for plasma and 50–2000 ng/mL for all tissues analyzed. Assay accuracy and precision are determined by validating the QC samples. The resulting assay precision and accuracy data for plasma are presented in Tables 2 and 3. The lower limit of quantification of TMTZ in all media was set at the lowest standard concentration. To study the acute doses (30 min, 1 h, or 2 h at high concentration) of the pharmacokinetic assay, the samples were diluted as described in the “Material and method” section to concentrations falling within the standard curve. To validate the dilution procedure, the assessment of accuracy was carried out with a quality control at 4000 and 20,000 ng/mL during the assay validation process. Results of analysis from plasma revealed a very satisfactory precision and accuracy (Table 4). Using the three QC samples analyzed on two different days, an extraction yield was obtained from the peak area ratio of TMTZ extracted/unextracted. A satisfactory mean extraction yield from 86 to 90 % was found for

TMTZ. Various experiments were performed to study the stability of TMTZ in the present assay. After extraction from spiked plasma, the stability of TMTZ was demonstrated for at least 24 h when vials were maintained in the autosampler at +5 °C. No significant degradation (<10 %) was observed for TMTZ after three freeze-thaw cycles in plasma, allowing multiple rounds of re-analysis (data not shown).

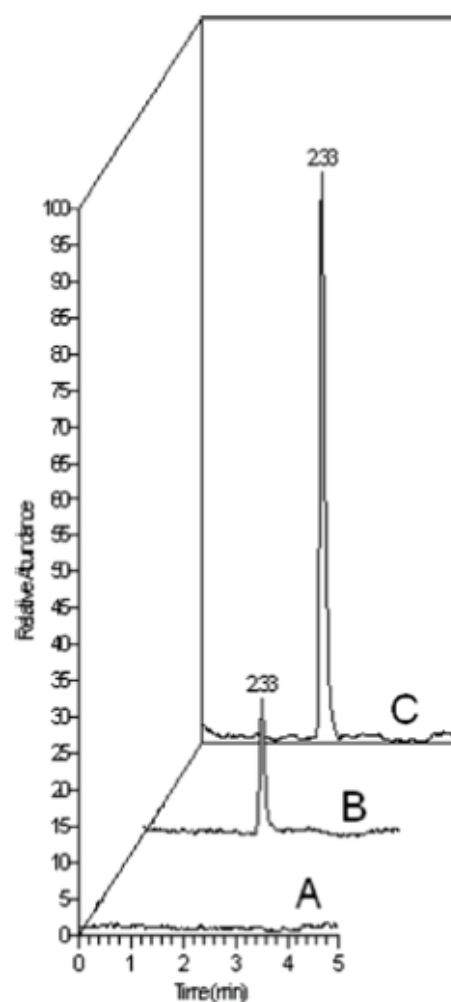


Fig. 1 MRM chromatograms for TMTZ (m/z 117→72) in the human blank plasma (A), in the human blank plasma spiked with TMTZ at 50 ng/mL (B), and in the plasma from mice used for pharmacokinetic study (TMTZ 131 ng/mL) (C)

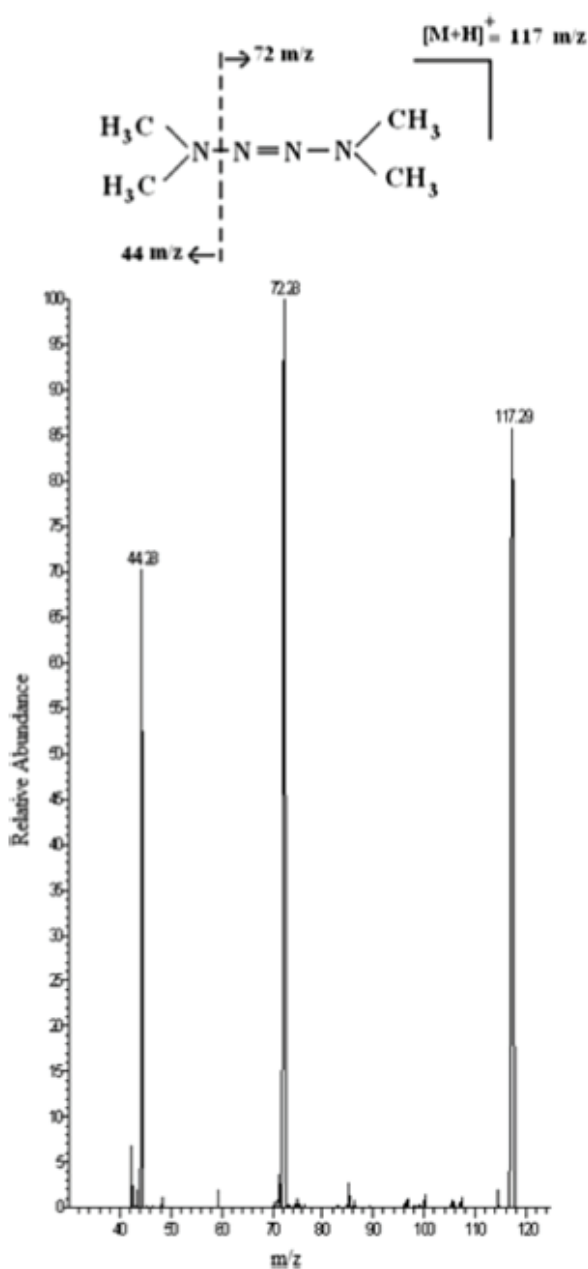


Fig. 2 Product ion mass spectrum (partial fragmentation) of TMTZ

Safety studies

In both safety studies, regardless of the exposure period (24 h and 14 days), 300 mg/kg TMTZ administrations did not cause death or any loss of weight. Under these protocols, mice did not show any modification in behavior (physical activity, appearance). There were no significant differences observed between the overall mean intakes of food or water between control and treated mice. For the further kinetic experiment, 300 mg/kg dose was selected to treat mice.

Kinetic of TMTZ in plasma

The plot of the plasma concentration–time profile of TMTZ in mice is shown in Fig. 3. The concentrations ranged from 84 to 98,289 ng/mL. A bi-compartmental model was built using the R statistical software, with a model equation for plasmatic concentration (C_p) corresponding to $C_p = A \cdot e^{(-\alpha \cdot t)} + B \cdot e^{(-\beta \cdot t)}$. For TMTZ model, the equation was (concentration expressed in ng/mL and time in min) $C_p = 217,970 \cdot e^{(-0.06597923 \cdot t)} + 3700 \cdot e^{(-0.01128446 \cdot t)}$.

The pharmacokinetic parameters calculated from the present model were 36 mL (1.8 L/kg) for the volume of distribution at steady state ($V_{d_{ss}}$) with central and peripheral compartment equal to 27 and 9 mL, respectively. The *micro* rate constants of k_{10} (elimination rate constant; $k_{10} = (\alpha \cdot \beta) / k_{21}$), k_{12} (distribution rate constant from compartment 1 to compartment 2; $k_{12} = \alpha + \beta - k_{21} - k_{el}$), and k_{21} (distribution rate constant from compartment 2 to compartment 1; $k_{21} = (A \cdot \beta + B \cdot \alpha) / (A + B)$) were 0.061, 0.0040, and 0.012 min^{-1} , respectively.

Kinetic of TMTZ in tissues

Table 5 shows the set of values of the tissue kinetics from 30 to 300 min for the five tissues studied. The body clearance of TMTZ was elevated since at 300 min, with the exception of liver of one mouse, the compound was not detected in any tissues. We also noticed that TMTZ half-times were strongly reduced in the heart and lung since the compound was no longer detected at 120 min in both tissues. Concentrations observed in the liver, kidney, and brain from 30 to 120 min are presented in Fig. 4. It is within the liver that the distribution of TMTZ was the most important; kidney and brain were fairly similar concentrations.

We also investigated the metabolism of TMTZ. The identification of the potential metabolites was performed using liquid chromatography coupled with two different mass spectrometry technologies: tandem quadrupole mass spectrometry and high-resolution mass spectrometry. Metabolizer (Metabolizer 15.1.5.0, 2014 ChemAxon (<http://www.chemaxon.com>)) is capable of predicting the major and minor metabolites and estimating their metabolic stability. The enumeration process of the metabolites is based on a manually curated knowledge base. Based on Metabolizer *in silico* predictions, several metabolic products were considered including *N*-demethylation, *N*-dealkylation, *N*-hydroxylation, and *C*-hydroxylation. We considered either one or two reactions of the same type (two *N*-demethylation, for example) or two types of reaction (*N*-demethylation followed by *C*-hydroxylation, for example). Thus, more than ten different structures were investigated following *in vitro* experiments, by using mouse liver microsomes, or from blood and tissue samples used for pharmacokinetic experiment. Using microsomes, a time-dependent decrease

Table 2 Inter-day validation of determination of tetramethyl-tetrazene (TMTZ) in the plasma. Data are from calibration curves prepared as a single replicate and analyzed on six different days

TMTZ (ng/mL) Spiked	Found (mean±s.d.)	Precision (%) (between run)	Accuracy (%)
10	9.9±0.6	5.9	99.9
25	25.3±1.1	4.5	101.1
50	10.4±0.3	4.1	99.0
100	100.1±4.2	4.2	100.1
200	200.4±4.7	2.3	100.2
500	499.9±1.2	0.2	100.0

of TMTZ was observed, which was inhibited using the non-selective CYP450 inhibitor aminobenzotriazole (data not shown). This strongly suggested a CYP450-dependent metabolism for TMTZ. However, no metabolite was detected in microsomes or in blood or tissues.

Discussion

Two assays based on liquid chromatography were previously published for the simultaneous determination of TMTZ and hydrazine, methylhydrazine, and asymmetric dimethylhydrazine [8, 9]. These authors have thoroughly studied the influence of eluent composition and tested several ion-pair reagents. For TMTZ, the limit of quantification in the water sample was 9 and 90 ng/mL using an amperometric and spectrophotometric detector, respectively [8]. Another study based on headspace SPME-GC-MS apparatus was developed for the screening of transformation products from soil and water samples [7, 11]. The present assay is the first one describing the quantification of TMTZ in biological media. Satisfactory limits of quantification allowing plasma and tissue pharmacokinetic studies have been obtained. The validation of TMTZ quantification by LC-MS/MS was fully performed in plasma. The recommended precision (15 %) and accuracy (100±15 %) requirements for the three QC plasma samples were obtained since the within-run and between-run precision of the assay was less than 6 %. Assay accuracy was in the range of 98–102.3 %. Therefore, this assay was accurate, reproducible, and sensitive for the quantification of TMTZ throughout a wide dynamic range. No interference was

observed from plasma samples as well as tissues. Only a partial validation was performed for tissue samples. For the quantification of TMTZ in mice tissue samples, standard curves were built using the appropriate matrix (brain, kidney, liver, lung, or heart) and QC samples were added in each run.

To explore *in vivo* toxicity, we have to determine the route of administration, the frequency, and administrated dose, modeled on the pharmaceutical development of new drugs. The choice of the intraperitoneal (I.P.) route was driven by several arguments including technical ease, distribution efficiency, the authorized administrated doses, and pharmacokinetic characteristics. As the peritoneal cavity is strongly vascularized by lymphatic and blood vessels, the distribution levels of administrated compounds into the body are high within a short time of exposure (less than 15 min). Compared to intravenous administration (I.V.), the I.P. administration is technically convenient and the distribution is very similar to that with I.V. In addition, as TMTZ solubility into vehicles is a limiting factor, the I.P. route is beneficial as the authorized volumes by I.P. (300 µL/day) are three times the authorized injected volumes by I.V. (100 µL/day), thus allowing an increased injected dose per mouse. For oral administration, the forced feeding route makes it possible to evaluate the exact administrated quantity, but it is limited by the intestinal absorption. As it is a micelle mixture, the absorbed dose in drinking water per mouse would be highly variable and it would be impossible to indicate the moment of administration, a critical point for pharmacokinetic analysis.

This safety study was carried out to determine, for pharmacokinetic study, a dose usable without toxic or animal suffering symptoms. However, to our knowledge, no experimental information was available on toxicity of

Table 3 Assessment of accuracy and precision in the plasma. Data are from six replicates for each concentration and analyzed on three different days

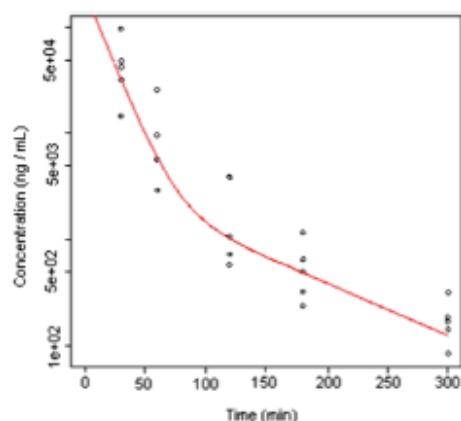
TMTZ (ng/mL) Spiked	Found (mean±s.d.)	Precision (%)		Accuracy (%)
		Within run	Between run	
20	20.5±1.0	4.2	2.6	102.3
75	75.3±5.3	5.5	1.9	100.4
300	294.0±17.6	5.9	0.4	98.0

Table 4 Assessment for extended concentration range and validation of dilution procedure (^a and ^b: samples diluted prior to extraction in blank plasma at 1/10 and 1/50, respectively)

TMTZ (ng/mL) Spiked	Found (mean±s.d.)	Precision (%) (within run)	Accuracy (%)
4000 ^a	3747±131	3.5	93.7
20,000 ^b	18,495±824	4.5	92.5

TMTZ and only conflicting data were available about the lethal dose 50 (LD50). Carlsen et al., using an in silico approach (ADME Box version 3.5 software), proposed a LD50 (mg/kg) in mouse at 170, 230, 68, and 110 for I.P., oral, I.V., and subcutaneous administration, respectively [6]. Likewise, in mouse, the ACD/Percepta software (ACD/Labs, version 12.01; Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2014.) assessed the LD50 at 230, 480, 130, and 160 for I.P., oral, I.V., and subcutaneous administration, respectively. Data from Toxnet (toxicology data network) reported data from experimental studies indicating a LD50 in mouse at 1080 and 1500 mg/kg for intramuscular and oral administration, respectively [12]. However, original studies from which these values are derived are not available. Thus, on the basis of this potential wide range of toxic doses, a short (24 h) safety study was conducted in 24 mice ($n=6$ of each dosing group) to determine the safety of 0, 75, 150, and 300 mg/kg of TMTZ. A 2-week safety study was conducted in 12 mice ($n=6$ of each dosing group) to determine the safety of 0 and 300 mg/kg of TMTZ. All mice grew normally and survived until the end of the study. In both studies, no pains, distress, or change of their behavior was observed in animals.

Since no experimental data are available, we determined the pharmacokinetics of TMTZ in mice after I.P. administration of 300 mg/kg administration to explore

**Fig. 3** Plasma concentration of TMTZ in mice after intraperitoneal administration (300 mg/kg)

how TMTZ is distributed and the kinetics of elimination from the organism. Since TMTZ fell under the limit of detection in tissues and blood samples at 24 h of exposure, it strongly supported previous evidence that TMTZ has a short half-life. Thus, we used shorter time ranges for tissues and blood collections after administration in our experimental assay. After single I.P. administration, the plasma concentration of TMTZ versus time was well fitted by bi-exponential equation corresponding at a two-compartment model. A two-compartment model assumes that, following drug administration into the central compartment (comprising tissues that are highly perfused such as heart, lungs, kidneys, liver, and brain), the compound distributes between that compartment and the peripheral compartment (comprising less well-perfused tissues such as muscle, fat, and skin). Thus, the model splits the body into these two compartments, which have no physiological or anatomical meaning. To our knowledge, only in silico data are available for kinetics of TMTZ [6]. The results of tissue distribution indicated that the TMTZ underwent a rapid distribution into limited tissues and organs within the time course examined. High concentrations were detected in the liver and kidney, suggesting that both organs mainly eliminated the TMTZ. It was also found at elevated concentrations in the brain, clearly indicating that TMTZ efficiently crosses the blood-brain barrier, suggesting a potential risk of neurotoxicity. This observation is supported by in silico prediction performed with the ACD/Percepta software (LogBB is -2.6 leading to the report "brain penetration sufficient for CNS activity"). Although the heart and lung are highly vascularized, they are exposed to the TMTZ for shorter amounts of time, supported by the fact that half-lives are short and concentrations are only barely detectable at early time points. Carlsen et al. reported the pharmacokinetic parameters of TMTZ using in silico expert systems (ADME Boxes software) [6]. This approach generated a $V_{d_{ss}}$ and k_e value equal to 1.13 L/kg and 0.002 min^{-1} , respectively. In comparison to our experimental data, values for $V_{d_{ss}}$ were very similar (1.8 vs 1.13 L/kg). However, a very large difference was observed concerning the elimination rate constant (0.012 vs 0.002 min^{-1}), meaning that, based on the experimental data, TMTZ was very quickly eliminated from the body and was clearly faster than assessed by in silico calculation. Both metabolism and

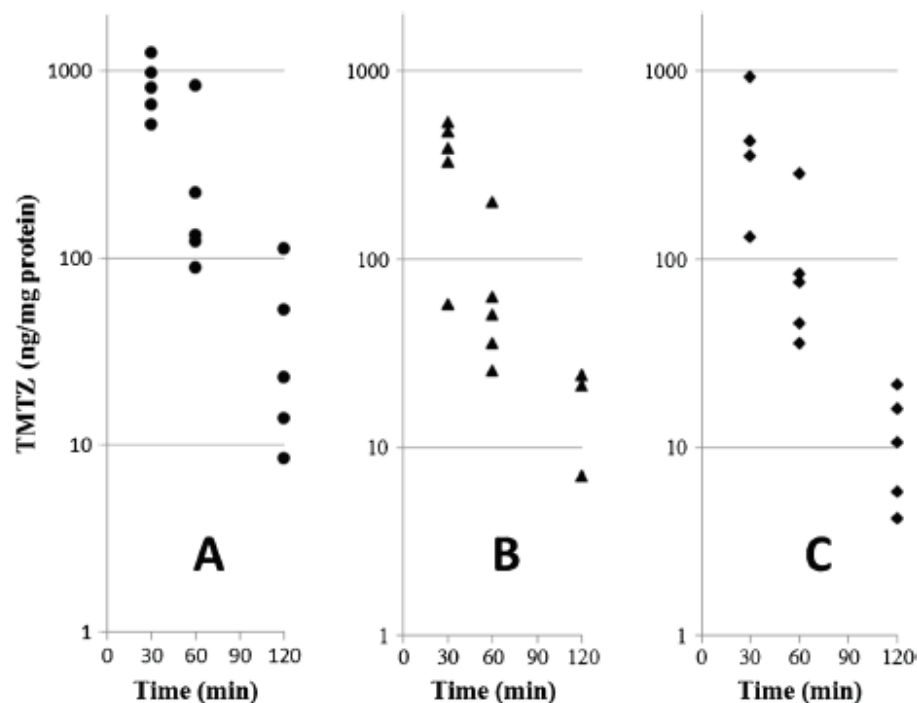
Table 5 Range of tissue concentration of TMTZ in mice after intraperitoneal administration (300 mg/kg)

Time (min)	Tissue concentration (ng/mg protein)				
	Liver	Brain	Kidney	Heart	Lung
30	[514–1245]	[132–933]	[57–532]	[12.5–378]	[8.1–398]
60	[89–873]	[36–283]	[25–200]	[8.5–55]	73 and 77
120	[9–112]	[4–22]	[7–24]	<LOQ	<LOQ
180	[4.8–25]	[2.7–6.3]	[2.5–2.6]	<LOQ	<LOQ
300	9	<LOQ	<LOQ	<LOQ	<LOQ

A range of tissue concentration is indicated in brackets when at least three out of the five mice had a higher concentration than the limit of quantification (LOQ = 2.5 ng/mg protein). Otherwise the tissue concentration of each mouse is indicated

excretion are involved in the body clearance of compounds, and the detection of TMTZ in the kidney and liver is suggestive of their implication in the excretion of the compound. In comparison, MMH showed first-order kinetics in the blood following I.V. administration and a wide and rapid distribution in the murine body. The elimination phase was also rapid and measured at 37–52 min after I.V. dosing [13]. Thus, the MMH and TMTZ have a close enough kinetic behavior. Since for TMTZ no metabolite was detected in vitro or in vivo, one hypothesis might be the formation of reactive metabolite leading to protein adduct. Adduct formation can have many toxicological consequences, and further experiments should be carried out to confirm this assumption.

From a toxicological point of view, these preliminary data suggest that the TMTZ could be considered as a relevant challenger to replace the MMH. LD50 for TMTZ was estimated at 270 mg/kg by in silico analysis, while we did not observe any modification in mice for a dose equal to 300 mg/kg. The LD50 of MMH is clearly lower since in mice the value was assessed between 15 and 33 mg/kg according to the route of administration [12, 14]. Moreover, the experimental vapor tensions (at 20 °C) are 37.35 and 6.57 mmHg for MMH and TMTZ, respectively. This is in favor of a less atmospheric exposure to TMTZ. Based on the present data, further experiments on TMTZ toxicity mechanisms should be performed to better investigate the potential interest of TMTZ as a new potential relevant propellant for space propulsion.

Fig. 4 Concentration of TMTZ in the liver (A), kidney (B), and brain (C) at 30, 60, and 120 min after intraperitoneal administration of 300 mg/kg TMTZ

Conclusion

This work led to the development of a powerful and selective LC-MS/MS assay on biological samples with an excellent sensitivity and selectivity for the analysis of TMTZ. This assay was performed to determine the pharmacokinetic parameters and the main organ distributions of TMTZ in mouse. These results will be especially useful in determining the toxicology of TMTZ and, more generally, will increase knowledge about the potential impacts of alkyltetrazene propellants in biological media. Some pharmacokinetic parameters previously assessed using an *in silico* approach were similar in value to those determined from experimental data. Inversely, the *in silico* data suggest a relatively slow elimination of TMTZ which is not shown experimentally. Similarly, we can suggest that the *in silico* data overestimates the toxicity of TMTZ. Finally, the experimental data will improve *in silico* models, providing a powerful and indispensable approach to investigate the toxicity of thousands of chemicals.

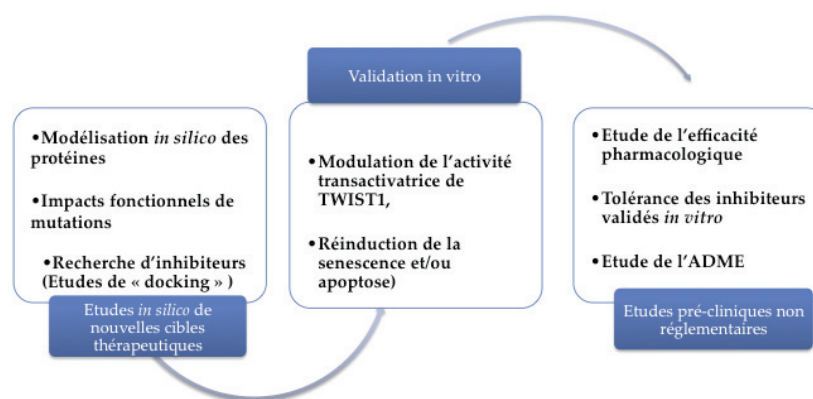
Acknowledgments We thank Emilie Henin (EMR 3738, Laboratoire de Ciblage Thérapeutique en Cancérologie, Centre Hospitalier Lyon-Sud, Pierre-Bénite, France) for her assistance in the determination of the pharmacokinetic parameters. We also thank Sarah Kabani for her help in reviewing the English of the manuscript.

References

1. Choudhary G, Hansen H (1998) Human health perspective on environmental exposure to hydrazines: a review. *Chemosphere* 37: 801–843
2. Carlsen L, Kenessov BN, Batyrbekova SY, Kolumbaeva SZ, Shalakhmetova TM (2009) Assessment of the mutagenic effect of 1,1-dimethyl hydrazine. *Environ Toxicol Pharmacol* 28:448–452
3. Reddy G, Song J, Mecchi MS, Johnson MS (2010) Genotoxicity assessment of two hypergolic energetic propellant compounds. *Mutat Res* 700:26–31
4. Carlsen L, Kenessova OA, Batyrbekova SE (2007) A preliminary assessment of the potential environmental and human health impact of unsymmetrical dimethylhydrazine as a result of space activities. *Chemosphere* 67:1108–1116
5. Miro Sabate C, Delalu H, Guelou Y, Dhenain A, Perrut C, (2011) Préparation d'alkyltetrazènes. Patent FR1153221
6. Carlsen L, Kenessov BN, Batyrbekova SY (2009) A QSAR/QSTR study on the human health impact of the rocket fuel 1,1-dimethyl hydrazine and its transformation products multicriteria hazard ranking based on partial order methodologies. *Environ Toxicol Pharmacol* 27:415–423
7. Kenessov B, Alimzhanova M, Sailaukhanuly Y, Baimatova N, Abilev M, Batyrbekova S, Carlsen L, Tulegenov A, Nauryzbayev M (2012) Transformation products of 1,1-dimethylhydrazine and their distribution in soils of fall places of rocket carriers in Central Kazakhstan. *Sci Total Environ* 427–428:78–85
8. Ponomarenko SA, Smolenkov AD, Shpigun OA (2009) Determination of 1,1-dimethylhydrazine and its decomposition products using ion-pair chromatography. *Mosc Univ Chem Bull* 64:147–153
9. Smolenkov AD, Ponomarenko SA, Shpigun OA, (2009) The retention of 1,1-dimethylhydrazine and its decomposition products on alkyl-grafted silica gels under the ion pair chromatography conditions. *Russ J Phys Chem* 83:478–487
10. Kenessov B, Sailaukhanuly Y, Koziel JA, Carlsen L, Nauryzbayev M (2011) GC-MS and GC-NPD determination of formaldehyde dimethylhydrazone in water using SPME. *Chromatographia* 73: 123–128
11. Kenessov BN, Koziel JA, Grotenhuis T, Carlsen L (2010) Screening of transformation products in soils contaminated with unsymmetrical dimethylhydrazine using headspace SPME and GC-MS. *Anal Chim Acta* 674:32–39
12. TOXNET, <http://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+1172>. TOXNET data network
13. Bingham E, Cohrssen B, Powell CH, (2001) Patty's toxicology, volumes 1-9. V4 1293
14. Hawks A, Hicks RM, Holsman JW, Magee PN (1974) Morphological and biochemical effects of 1,2-dimethylhydrazine and 1-methylhydrazine in rats and mice. *Br J Cancer* 30:429–439

CHAPITRE 7 : RESUME DE MES TRAVAUX DE RECHERCHE

Mes activités de recherche sont réalisées au Centre de Recherche en Cancérologie de LYON UMR Inserm 1052-CNRS 5286 (CRCL, Pr A. Puisieux). Mon projet de recherche de thèse s'inscrit dans la thématique de l'EMT et de la plasticité de la cellule cancéreuse de l'équipe du Pr A. Puisieux. Je contribue à la caractérisation de nouvelles cibles thérapeutiques impliquées dans l'initiation et la progression tumorale avec Pr Léa Payen-Gay. Nous recherchons des molécules bioactives bloquant ces cibles. Mes recherches vont de la modélisation *in silico* de cibles thérapeutiques potentielles, jusqu'au développement de tests pharmacologiques et toxicologiques *in vivo* incluant la validation préalable de l'efficacité *in vitro* de médicaments candidats. Ces travaux sont décrits dans différents articles publiés ou en cours de publication.



- Modélisation *in silico* de cibles thérapeutiques: nous mettons en place des stratégies expérimentales *in silico* : modélisation par homologie, dynamique moléculaire, amarrage moléculaire («docking»). Ces études font appel à des collaborations avec le Pr. R. Terreux (Lyon UMR CNRS 5305 Institut de Biologie et de Chimie des Protéines (IBCP)).
- Validation *in vitro* des cibles thérapeutiques ou des médicaments candidats : initialement dans le domaine des transporteurs ABC, et maintenant dans le domaine des facteurs de transcription embryonnaires, nous mettons en place des outils cellulaires (lignées transfectées ou sélectionnées) exprimant des protéines d'intérêt (TWIST1/ E12 ou ABCG2). Nous étudions l'activité de modulation de médicaments candidats vis-à-vis soit de la régulation d'expression de gènes cibles (TWIST1) impliqués dans la progression tumorale, soit de la diminution de la chimiorésistance aux anticancéreux (transporteurs ABCG2).
- Validation *in vivo* des cibles thérapeutiques potentielles ou des médicaments candidats: nous menons des expériences mesurant l'efficacité thérapeutique de médicaments candidats sur différents types de tumeurs issues de modèles murins xénotransplantés. L'efficacité thérapeutique est mesurée par le ralentissement de la croissance tumorale. L'étude des paramètres pharmacocinétiques est réalisée en parallèle. Tous ces protocoles sont accompagnés d'études de tolérance (toxicité aiguë, toxicité chronique). Ces données sont nécessaires au dossier toxicologique non clinique non réglementaire.

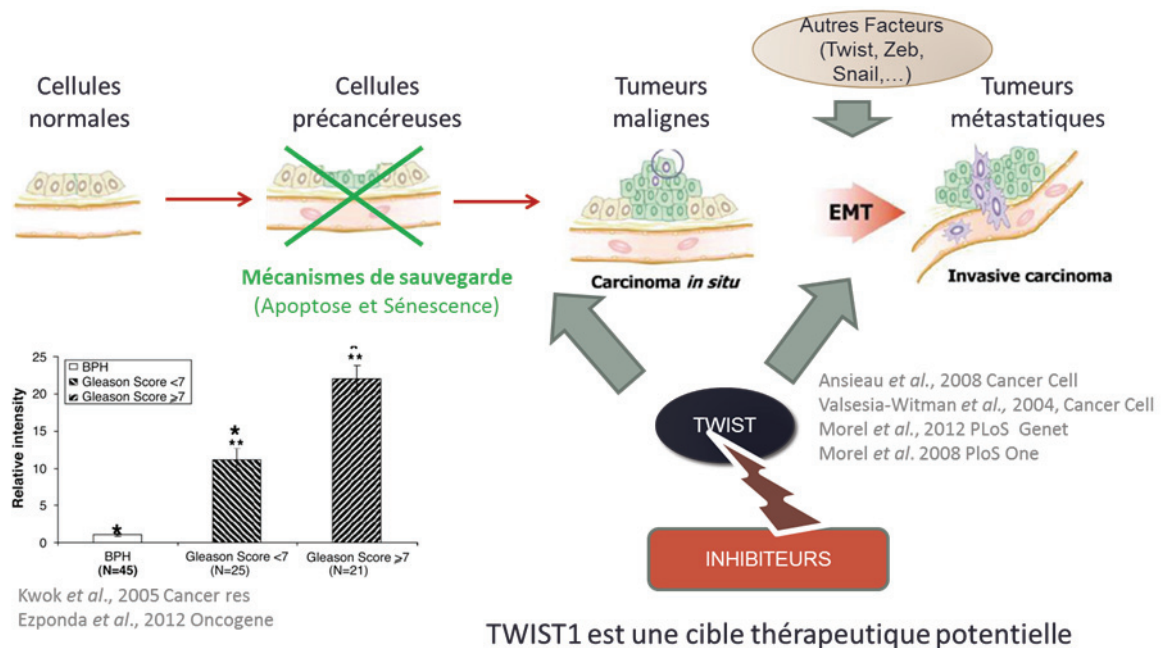
1 Revue bibliographique : « Plasticité des cellules cancéreuses et dissémination métastatique »

La dissémination métastatique est une suite d'événements qui conduit certaines cellules cancéreuses à envahir des tissus distants de la tumeur primaire, on parle alors d'invasion. Chaque événement constitue un obstacle pour les cellules cancéreuses et seul un nombre limité de cellules cancéreuses sont capables de les franchir. Toutefois la diversité génétique des cellules cancéreuses au sein d'une tumeur et la plasticité cellulaire transmise suite aux signaux du microenvironnement facilitent le franchissement de ces étapes et la dissémination des cellules cancéreuses dans l'organisme et favorisent la formation de métastases en favorisant l'émergence de cellules tumorales ayant une résistance intrinsèque aux différentes pressions rencontrées. Ainsi la progression tumorale et la dissémination métastatique font preuve d'une grande adaptabilité et font appel à des interactions étroites entre les cellules tumorales et le microenvironnement. Un processus de trans-différenciation embryonnaire, la transition épithéliale-mésenchymateuse, fréquemment réactivée au cours du développement de la tumeur, donne aux cellules tumorales différentes propriétés de cellules souches et donc de plasticité permettant donc aux cellules tumorales de s'adapter aux nouveaux environnements et contraintes rencontrées lors de l'invasion en faveur également de l'hétérogénéité tumorale.

Ces travaux ont été valorisés dans une publication internationale : « Cancer cell plasticity and metastatic dissemination » Moyret-Lalle C., Pommier R., Bouard C., Nouri E., Richard G., Puisieux A. (Médecine/Sciences, janvier 2016 ; IF 2016: 0.671).

2 Etude *in silico* des dimères TWIST1/TWIST1 et TWIST1/E12

La sénescence et l'apoptose constituent des barrières naturelles érigées contre l'apparition de lésions précancéreuses. Ces mécanismes diminuent et protègent l'organisme de l'émergence de cellules hyper-prolifératives. L'échappement à ces programmes de sauvegarde favorise la conversion maligne et constitue un préalable à la survie des cellules cancéreuses aux différents stades de progression tumorale. La réactivation aberrante des gènes *TWIST* embryonnaires (TWIST1 ou TWIST2) a été identifiée dans le laboratoire comme un mécanisme d'inactivation des voies P53 et RB dans de nombreux types de tumeurs {Ansieau, 2008 #6}. Ces protéines appartiennent à la grande famille des facteurs de transcription de la famille des b-HLH. En diminuant la sénescence et l'induction de l'apoptose, les facteurs de transcription TWIST coopèrent avec les protéines mitogéniques dans la transformation cellulaire *in vitro* et interviennent dans l'initiation et la progression tumorale *in vivo* {Ansieau, 2008 #6; Morel, 2008 #100; Ansieau, 2010 #101; Morel, 2012 #26}. Inversement, la diminution de TWIST dans des cellules tumorales TWIST-positives est invariablement associée au programme de restauration des mécanismes de sauvegarde aussi bien *in vitro* qu'*in vivo* {Morel, 2012 #26; Tran, 2012 #86}.



2.1 Construction des modèles *in silico* des homo et hétérodimères de TWIST1

Dans un premier travail, *in silico*, nous avons élaboré par homologie des modèles homodimériques et hétérodimériques de TWIST1. L’empreinte utilisée est NeuroD1/E47 de souris (pdb 2QL2) {Longo, 2008 #23}. Les homodimères et hétérodimères activent différents pools de gènes cibles, menant à des régulations différentes.

Ceci nous a permis de mettre en évidence les résidus d’acides aminés critiques susceptibles d’intervenir dans les interactions avec les éléments de réponse (E-box) des promoteurs des gènes cibles {Bouard, 2014 #7}. Nous avons aussi caractérisé l’impact des insertions Ins-139 et Ins-135 (retrouvées dans la maladie génétique appelée Saethre-Chotzen). Elles conduisent à un défaut de fonction des dimères TWIST, lié à une modification des boucles latérales.

Ces travaux ont été valorisés :

Publication internationale : « Interhelical loops within the bHLH domain are determinant in maintaining TWIST1–DNA complexes » Bouard C., Terreux R., Hope J., Chemelle JA., Puisieux A., Ansieau S. et Payen L. (Journal of Biomolecular Structure & Dynamics, 2014. ; IF 2014: 2.919).

Communication écrite : « Interhelical loops within the bHLH domain are determinant in maintaining TWIST1-DNA complexes » Payen L., Bouard C., Terreux R., Hope J., Chemelle JA., Puisieux A. et Ansieau S. (Symposium du CRCL, février 2013).

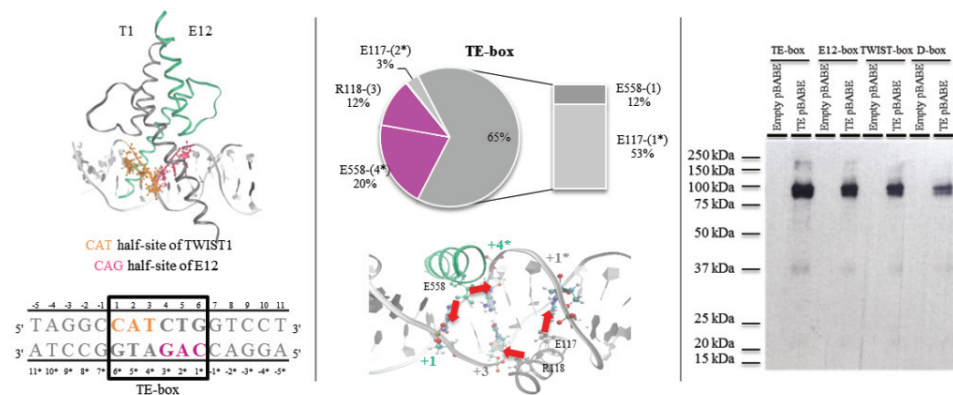
2.2 Identification des éléments clefs de TWIST1/E12 qui influencent la reconnaissance des motifs ADN cibles des promoteurs de gènes

Afin d'appréhender les mécanismes moléculaires mis en jeu lors de la liaison à l'ADN du dimère TWIST1/E12, nous avons analysé la dynamique moléculaire du modèle tridimensionnel de l'hétérodimère TWIST1/E12 lié à différentes E-box. Nous avons ainsi comparé la fixation de l'hétérodimère TWIST1/E12 sur une E-box active (de haute affinité du dimère TWIST1/E12) et sur différentes E-box modifiées actives ou inactives. Nous avons validé toutes nos analyses *in silico* par des analyses *in vitro* mesurant l'affinité de TWIST1/E12 aux différentes séquences E-box.

Au cours de cette étude, nous avons mis en évidence le rôle prépondérant de la nature des chaînes latérales des résidus de protéines localisés à proximité du cœur du complexe et des séquences d'ADN. Une diminution de l'affinité de liaison se traduit par un glissement des interactions du dimère protéique des bases variables centrales de l'E-box vers les bases conservées externes de l'E-box ou vers les bases adjacentes à l'E-box (-1; -1*). Toutes nos analyses *in silico* ont été validées par des études *in vitro* via une approche streptavidine/biotine. Ainsi nous avons montré la valeur prédictive de nos simulations *in silico*. En effet ces modèles tridimensionnels pourraient être utilisés pour étudier l'implication structurale de certains acides aminés mutés et associés notamment au syndrome de Saethre-Chotzen et pour le criblage d'inhibiteurs spécifiques de l'hétérodimère oncogénique de TWIST1,

TWIST1/E12.

Cette stratégie *in silico* peut également être utilisée pour la caractérisation d'autres protéines.



Ces travaux ont été valorisés :

Publication internationale : « Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences » [Bouard C.](#), Terreux R., Honorat M., Manship B., Ansieau S., Vigneron A., Puisieux A., Payen L. (Nucleic Acid Research, avril 2016 ; IF 2016: 9.112).

Communications écrites :

1. « Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences » [Bouard C.](#), Terreux R., Honorat M., Manship B., Ansieau S., Vigneron A., Puisieux A., Payen L. (Journée du CRCL, mai 2016).

2. « Deciphering the molecular mechanism of the TWIST1/E12 complex DNA binding on E-box regulatory sequences » Bouard C., Terreux R., Honorat M., Ansieau S., Vigneron A., Puisieux A., Payen L. (Symposium du CRCL, septembre 2015).

Communications orales :

1. « Deciphering of key mechanism of DNA binding function of the TWIST1/E12 complex (*in silico* and *in vitro* approaches)» Bouard C., Payen L., Bouchelouche R., Puisieux A. (Journée de retraite de département, juin 2015).

2. « Deciphering of TWIST1/E12 binding to regulatory sequences of the gene promoters (*in silico* and *in vitro* approaches) » Bouard C., Payen L., Bouchelouche R., Puisieux A. (Journée de retraite d'équipe, janvier 2015).

2.3 Recherche d'inhibiteurs spécifiques de TWIST1/E12

Après la caractérisation fine et la validation de nos modèles de prédiction, nous allons utiliser ces modèles afin de lancer des screening *in silico* de molécules chimiques bloquant de façon spécifique l'hétérodimère cible, fixé à différents types de E-box. Nous avons effectué une première de faisabilité en réalisant un amarrage moléculaire (« Docking » réalisée sous SYBYL X 1.1) d'une librairie chimique de 12 000 molécules chimiques du NCI en appliquant les règles de sélections de Lipinski. En parallèle, nous menons également un criblage à partir d'une banque chimique originale sous MOE (collaboration avec A. Boumendjel, DPM Grenoble). Ainsi, nous pouvons évaluer si notre outil bio-informatique est en mesure de prédire la relation structure activité de molécules chimiques proche, comme décrit dans la figure ci-après. Actuellement des tests *in vitro*, sont en cours afin de mesurer l'efficacité de ces molécules et explorer leur site d'interaction sur le complexe TWIST1/E12.

Ces travaux ont été valorisés :

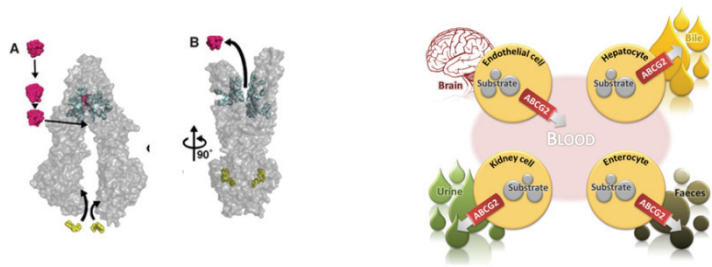
Communication écrite : « *In silico* screening of xenobiotics interacting with TWIST1 dimers » Bouard C., Payen L., Terreux R., Puisieux A. et Ansieau S. (Journée médico-scientifique du CRCL, décembre 2011).

Communication orale : « Identification de modulateurs spécifiques de la protéine TWIST1 sur la base d'analyses structurales et fonctionnelles » Bouard C., Payen L., Puisieux A. (Journée JPS de l'ISPB, septembre 2014).

3 Implications dans des travaux de pharmaco-toxicologie

3.1 Etudes précliniques non réglementaires: tolérance et efficacité d'inhibiteurs de nouvelles cibles thérapeutiques (*in vivo*)

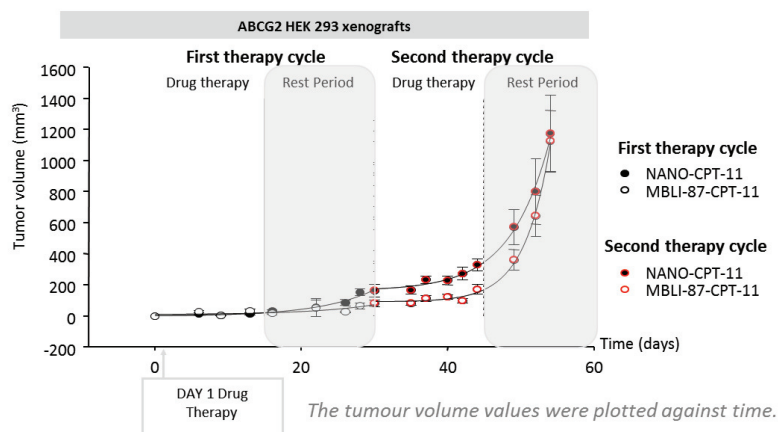
Dans le cadre du développement de tests précliniques sur des molécules actives, j'ai participé à des protocoles expérimentaux utilisant des modèles murins porteurs de xénogreffes. Par exemple, la mise en place d'un modèle de xénogreffes exprimant fortement la protéine ABCG2 nous a permis :



- d'étudier la voie d'administration (orale ou intrapéritonéale) la plus appropriée,
- d'évaluer la toxicité aiguë et chronique d'inhibiteurs de transporteurs ABCG2 (MBL-II-141),
- de déterminer les paramètres pharmacocinétiques de ces molécules d'intérêt (taux plasmatique et distribution tissulaire),
- d'évaluer l'efficacité pharmacologique *in vivo* de ces molécules bioactives.

Ces travaux nous ont permis de mettre en place une structure réalisant des tests précliniques mesurant l'efficacité pharmacologique de médicaments candidats, et générant les premières données de toxicité indispensables avant la réalisation des tests précliniques réglementaires. Les activités réalisées se font dans le respect des Bonnes Pratiques de Laboratoire françaises et européennes (BPL).

ABCG2-expressing xenografts.



Ces travaux ont été valorisés dans une publication internationale : « **MBL-II-141, a chromone derivative, enhances irinotecan (CPT-11) anticancer efficiency in ABCG2-positive xenografts** » Honorat M., Guittou J., Gauthier C., Bouard C., Lecerf-Schmidt F., Peres B., Terreux R., Gervot H., Rioufol C., Boumendjel A., Puisieux A., Di Pietro A. Payen L. (Oncotarget, 2014 ; IF 2014: 6.359).

3.2 Mesure de la toxicité de molécules d'intérêt industriel.

Grâce aux activités de valorisation menées par le Pr. L. Payen-Gay, j'ai eu l'occasion de réaliser des tests de cytotoxicités (MTT) de molécules chimiques ayant un intérêt industriel au niveau *in vitro* et *in vivo*. Ceci m'a permis d'élargir le panel de mes compétences. Ceci a été rendu possible grâce à notre maîtrise de différents modèles cellulaires (fibroblastes dermiques, cellules épithéliales

dermiques, cellules hépatiques) ainsi que des tests mesurant la cytotoxicité (tests *in vitro* de cytotoxicité, MTT, mesure de l'apoptose, de la sénescence ...).

Ces travaux ont été valorisés dans une publication internationale : « Quantification and kinetic study in plasma and tissues of (E)-1,1,4,4-tétraméthyl-2-tétrazène, a liquid propellant and a transformation product of 1,1-dimethyl hydrazine » Payen L., Honorat M., Bouard C., Jacob G., Terreux R., Delalu H., Labarthe E., Guitton J. (Analytical and Bioanalytical Chemistry, juin 2015 ; IF 2014: 3.436).

CHAPITRE 8 : PERSPECTIVES

A ce jour la structure tridimensionnelle cristalline de la protéine TWIST1 n'a pas été déterminée. Ainsi la modélisation par homologie de TWIST1 permet d'approximer sa structure et ainsi d'identifier les éléments structuraux clefs et nécessaires à ses fonctions. L'étude *in silico* des altérations génétiques de TWIST1 associées au syndrome de Saethre-Chotzen nous a permis de mettre en évidence les résidus d'acides aminés critiques susceptibles d'intervenir dans les interactions avec les éléments de réponse (E-box) des promoteurs des gènes cibles, notamment de déterminer le rôle de stabilisateur des boucles interhélicales des domaines b-HLH (Bouard et al., 2014). De plus nous avons définie l'hétérodimère TWIST1/E12 comme la forme oncogénique de TWIST1 dans des cellules épithéliales mammaires humaines (Jacqueroud et al., 2016). Afin d'étudier la fixation de la forme oncogénique de TWIST1 à des séquences E-box contenues dans des promoteurs de gènes, nous avons sélectionné différentes E-box (séquence consensus « CANNTG ») définies avec plus ou moins d'affinité pour TWIST1 ou E12 à partir des données bibliographiques. Selon la composition des E-box et en accord avec les données bibliographiques, nous avons montré différentes affinités de reconnaissances des E-box par le complexe TWIST1/E12 aussi bien *in silico* qu'*in vitro* par des essais de stréptavidine/biotine (Bouard et al., 2016; B. A. Firulli et al., 2007; Gordan et al., 2013). Ainsi nous avons défini trois états de stabilité du dimère protéine/ADN corrélés à la stabilité et à la localisation des interactions entre TE et l'ADN :

- un premier état avec une absence de stabilisation de l'hélice α -amphipatique associée à une absence de reconnaissance de la séquence E-box de l'ADN par le complexe ;
- un second état correspond à une stabilisation partielle de l'hélice α -amphipatique associé à une reconnaissance partielle de la séquence E-box de l'ADN par TE ;
- un troisième état correspondant à une stabilisation complète de l'hélice α -amphipatique associé à une reconnaissance spécifique de la séquence E-box de l'ADN par TE (Sauve et al., 2007).

Nous avons également montré le rôle essentiel que jouent les bases adjacentes à l'E-box (hautement variable entre les différents promoteurs de gènes) sur la stabilité de fixation de TWIST1/E12 à l'ADN. Et enfin comme la phosphorylation de TWIST1 impact sa dimérisation, nous avons étudié *in silico* l'impact de la phosphorylation de ce facteur de transcription sur sa fixation à l'ADN (A. B. Firulli & Conway, 2008). Dans cette étude nous avons vu que l'état de phosphorylation des résidus de thréonine 121 et/ou de serine 123 influençait fortement l'affinité de reconnaissance de l'hétérodimère TWIST1/E12 à E-box (Bouard et al., 2016). De plus cette approche nous a permis d'identifier un résidu clef, dont le rôle, est influencé par l'état de phosphorylation du résidu de serine 123, il s'agit de l'histidine 593. En effet la mutation de ce résidu d'histidine en alanine suffit pour déstabiliser l'ensemble du complexe protéine-ADN montrant donc son implication majeure dans la liaison à l'ADN des complexes de TWIST1.

Les nouveaux objectifs de ce projet pourraient être :

- identifier la forme oncogénique de TWIST1 dans d'autres contextes tumoraux que le contexte tumoral mammaire. En effet, TWIST1 agit au cours du développement embryonnaire sous forme homodimérique et sous différentes formes hétérodimériques. Il serait intéressant de voir ce qu'il en est dans différents contextes cancéreux afin d'élargir les connaissances des formes actives de TWIST1 d'une part mais également sur les mécanismes d'actions sous-jacents mis en jeu par ces complexes d'autre part.
- déterminer d'autres formes oncogéniques de TWIST1 et la caractérisation structurale de ces complexes qui permettraient de voir si la présence de la cavité au niveau de la boucle de E12 est spécifique de l'hétérodimère TWIST1/E12 ou si cette cavité est une caractéristique structurale de E12 indépendante de son partenaire de dimérisation. Cette information est importante pour la spécificité de reconnaissance de molécules chimiques.
- déterminer si les boucles ont une conformation adaptable selon le partenaire de dimérisation puisque les boucles interhélicales jouent un rôle de stabilisateur dans la liaison des complexes de TWIST1 à l'ADN.
- déterminer l'état de phosphorylation de TWIST1 afin d'étudier l'influence de cette phosphorylation sur la dimérisation de TWIST1 avec son partenaire. Ainsi, définir l'état de phosphorylation de la forme oncogénique TWIST1/E12 dans le contexte tumoral mammaire dans un premier temps, puis dans d'autres contextes tumoraux dans un second temps, permettrait de voir les phosphorylations activatrices des fonctions oncogéniques de TWIST1. Les études de phosphorylation de la protéine TWIST1 menées jusqu'à-là ont leurs limites. L'équipe de Firulli considère qu'il y a une double phosphorylation des résidus de T121 et de S123 ce qui semble peu probable dans la réalité. Au vu de l'encombrement stérique d'un groupement phosphate, il semble compliqué que deux résidus séparés d'un seul résidu puissent être phosphorylés en même temps compte tenu de l'encombrement stérique induit et de la force de répulsion entre deux charges de même signe. De plus, le résidu de T121 semble impliqué directement dans les interactions à l'ADN ce qui pose un problème lors de la liaison à l'ADN du dimère doublement phosphorylé. Et enfin, l'état de phosphorylation est mimé par le remplacement des résidus de thréonine et de sérine par un résidu d'Acide Aspartique. Ce dernier ayant une charge et un encombrement stérique différent d'un groupement phosphate. La caractérisation de l'état de phosphorylation de la forme oncogénique de TWIST1 est indispensable avant des études de criblages moléculaires. Au-delà de l'impact de la phosphorylation sur la dimérisation de TWIST1, la phosphorylation impacte également la liaison à l'ADN des dimères de TWIST1. Ainsi, il semble évident que l'état de phosphorylation puisse influencer la fixation et donc l'activité d'un modulateur potentiel de TWIST1.

La caractérisation fine et la validation de ces modèles tridimensionnels de TWIST1 pourraient être utilisées à des fins de prédiction dans l'identification de modulateurs putatifs de ces complexes afin d'en limiter les fonctions oncogéniques. Ces connaissances structurales permettent d'affiner et de définir des sites spécifiques stratégiques à cibler par les molécules chimiques testées. Les zones cibles clés ne seront pas les mêmes selon les fonctions ciblées de TWIST1. En effet les zones ciblées ne seront pas les mêmes si l'objectif est de moduler les capacités de dimérisation de TWIST1 avec ces partenaires ou si l'objectif est de moduler les capacités de liaisons à l'ADN des dimères de TWIST1. Si l'objectif est de bloquer la fixation à l'ADN de la protéine TWIST1, il est important de ne pas oublier les éventuels cofacteurs qui peuvent lier TWIST1 sur d'autres domaines que son domaine bHLH et qui peuvent avoir des effets secondaires non négligeables. Si l'objectif est de bloquer la dimérisation de la protéine TWIST1 avec ces partenaires, il ne faut pas, non plus, négliger l'impact que cela peut avoir sur les autres dimères du bHLH. En effet nous avons vu que certains dimères se formaient préférentiellement suite à la titration d'un des leurs. Une fois les zones à cibler déterminées, des études d'amarrages moléculaires (« Docking ») *in silico*, peuvent être réalisées. Après avoir défini les boucles de l'hétérodimère TWIST1/E12 comme pharmacophore, nous avons réalisé des études d'amarrages moléculaires à partir des 12000 molécules du NCI en appliquant les règles de Lepinski et à partir de chimiothèques originales du Professeur Achène Boumendjel de Grenoble sous SYBYL X1.1 puis sous MOE. Nous avons pu observer des différences de spécificités pour TWIST1 ou pour E12 selon la composition des molécules chimiques et notamment selon la nature des substituants des cycles aromatiques de celles-ci (Figure 50). De plus, au vu de l'impact structural de la phosphorylation de TWIST1, nous avons réalisé nos études de criblage moléculaire sur différents états de phosphorylation de TWIST1. La pré-sélection *in silico* de molécules spécifiques de TWIST1 ou de E12 permet d'avoir une explication structurale et ainsi de mieux comprendre le mécanisme moléculaire éventuellement induit *in vitro* et *in vivo* sous-jacent à la modulation de TWIST1 par les molécules chimiques sélectionnées *in silico*. Cette approche permet également de limiter les expériences *in vitro* et *in vivo* qui seront réalisées.

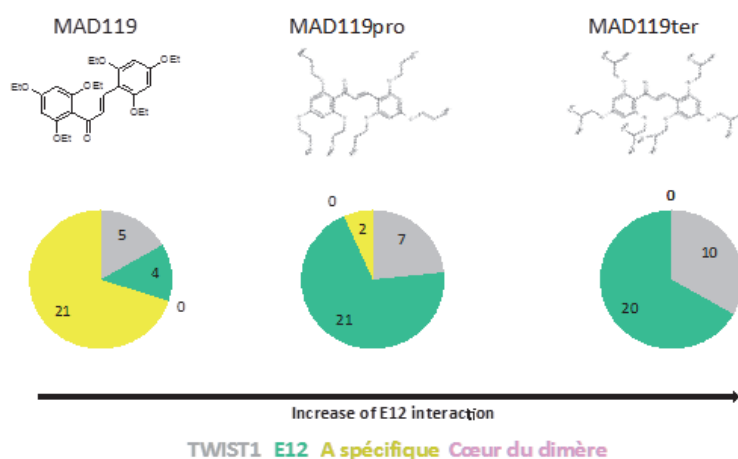


Figure 50: spécificité d'interaction de trois molécules chimiques.

Une fois les molécules sélectionnées *in silico*, leur efficacité pourrait être estimées *in vitro* par différentes approches afin de faire une seconde sélection des molécules les plus affines avec un bon rapport efficacité/spécificité. En effet, l'efficacité de ces molécules pourrait être testée *in vitro*. Des immuno-précipitations en présence et en absence des molécules chimiques pourraient dire si celles-ci bloquent la dimérisation de TWIST1 avec ces partenaires. De plus, des essais stréptavidine/biotine permettraient de dire si la molécule chimique testée impacte ou non la capacité de fixation du complexe de TWIST1 à l'ADN. Ensuite, des expériences de transactivation en présence et en absence de ces molécules chimiques permettraient de tester si celles-ci bloquent l'activité transcriptionnelle induite par les complexes de TWIST1. Et enfin, la modulation des molécules chimiques pourra être testée en étudiant l'impact d'une molécule chimique sur le complexe de TWIST1 en se focalisant sur l'étude de la restauration des mécanismes de sauvegardes cellulaires que sont la sénescence et l'apoptose. Ainsi, l'évaluation de l'efficacité *in vitro* des molécules spécifiques sélectionnées *in silico*, nous permettra d'affiner la sélection des molécules chimiques et d'évaluer la puissance prédictive de notre outil bio-informatique en termes de relation « structure-activité » de molécules chimiques.

La validation de modulateurs chimiques potentiels *in vitro* pourrait conduire à des expériences *in vivo*. En effet l'efficacité des molécules les plus affines pourrait être évaluée *in vivo* en jugeant leur efficacité sur le potentiel tumorigène de modèles xenogreffés avec des lignées sur-exprimant TWIST1/E12 après que la prise tumorale ait été initiée. Ces différentes approches alternatives permettront de valider le choix des molécules opéré à partir des criblages *in silico* par amarrages moléculaires et pourraient ouvrir la voie à des études précliniques (Figure 51).

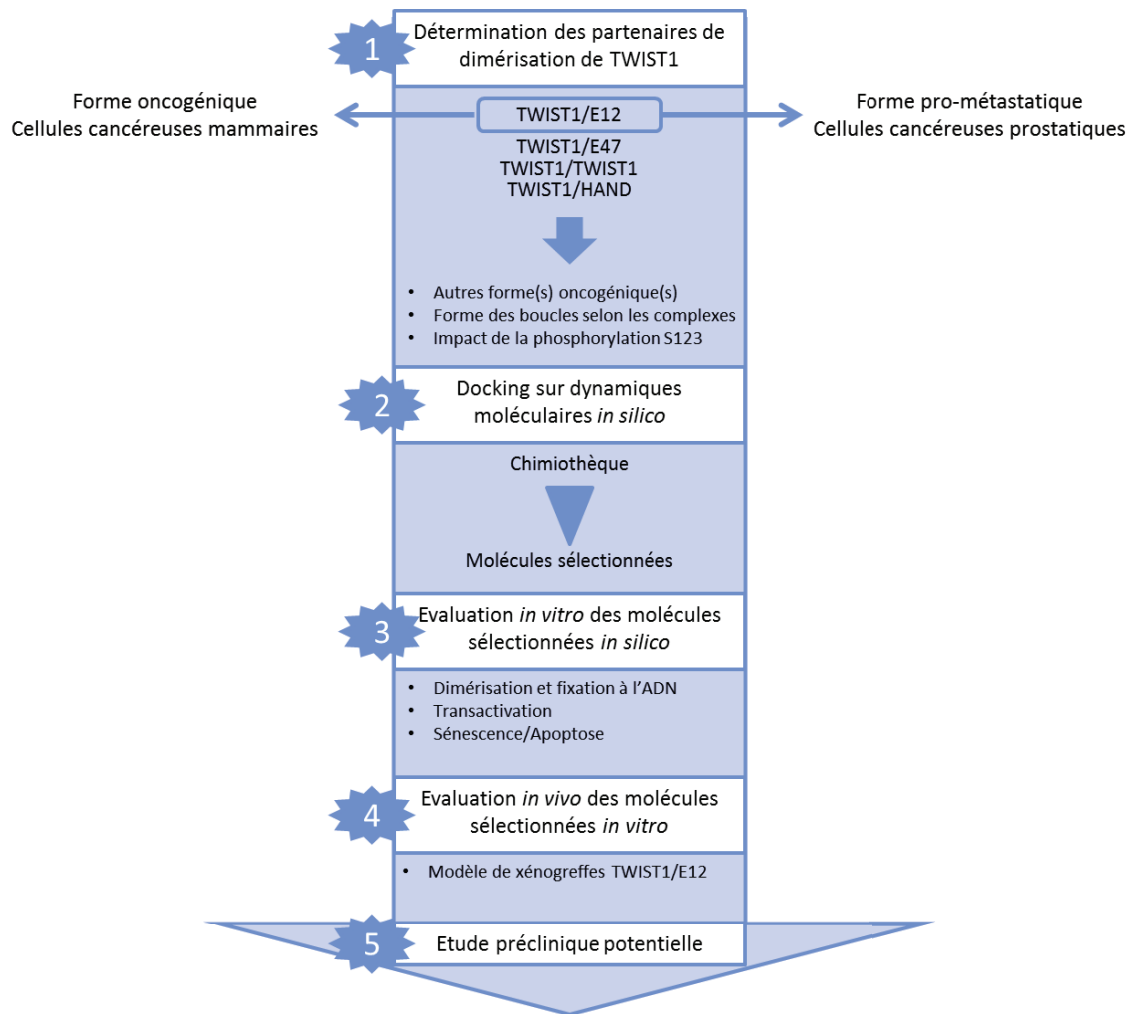


Figure 51 : perspectives de mes travaux de thèse.

REFERENCES BIBLIOGRAPHIQUES

- Adams, J. M. (2003). Ways of dying: multiple pathways to apoptosis. *Genes Dev*, *17*(20), 2481-2495. doi: 10.1101/gad.1126903
- Adams, J. M., & Cory, S. (2007). The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*, *26*(9), 1324-1337. doi: 10.1038/sj.onc.1210220
- Adams, R. H., & Alitalo, K. (2007). Molecular regulation of angiogenesis and lymphangiogenesis. *Nat Rev Mol Cell Biol*, *8*(6), 464-478. doi: 10.1038/nrm2183
- Ahn, C. H., Jeong, E. G., Lee, J. W., Kim, M. S., Kim, S. H., Kim, S. S., . . . Lee, S. H. (2007). Expression of beclin-1, an autophagy-related protein, in gastric and colorectal cancers. *APMIS*, *115*(12), 1344-1349. doi: 10.1111/j.1600-0463.2007.00858.x
- Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., . . . Eger, A. (2007). The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene*, *26*(49), 6979-6988. doi: 10.1038/sj.onc.1210508
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A*, *100*(7), 3983-3988. doi: 10.1073/pnas.0530291100
- Allinen, M., Beroukhi, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., . . . Polyak, K. (2004). Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell*, *6*(1), 17-32. doi: 10.1016/j.ccr.2004.06.010
- Alt, F. W., DePinho, R., Zimmerman, K., Legouy, E., Hatton, K., Ferrier, P., . . . Nisen, P. (1986). The human myc gene family. *Cold Spring Harb Symp Quant Biol*, *51 Pt 2*, 931-941.
- Ansieau, S., Bastid, J., Doreau, A., Morel, A. P., Bouchet, B. P., Thomas, C., . . . Puisieux, A. (2008). Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell*, *14*(1), 79-89. doi: 10.1016/j.ccr.2008.06.005
- Ansieau, S., Morel, A. P., Hinkal, G., Bastid, J., & Puisieux, A. (2010). TWISTing an embryonic transcription factor into an oncoprotein. *Oncogene*, *29*(22), 3173-3184. doi: 10.1038/onc.2010.92
- Artandi, S. E., Chang, S., Lee, S. L., Alson, S., Gottlieb, G. J., Chin, L., & DePinho, R. A. (2000). Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature*, *406*(6796), 641-645. doi: 10.1038/35020592
- Artandi, S. E., & DePinho, R. A. (2010). Telomeres and telomerase in cancer. *Carcinogenesis*, *31*(1), 9-18. doi: 10.1093/carcin/bgp268
- Atchley, W. R., & Fitch, W. M. (1997). A natural classification of the basic helix-loop-helix class of transcription factors. *Proc Natl Acad Sci U S A*, *94*(10), 5172-5176.
- Ayer, D. E., Kretzner, L., & Eisenman, R. N. (1993). Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell*, *72*(2), 211-222.
- Baeriswyl, V., & Christofori, G. (2009). The angiogenic switch in carcinogenesis. *Semin Cancer Biol*, *19*(5), 329-337. doi: 10.1016/j.semcancer.2009.05.003
- Baluk, P., Hashizume, H., & McDonald, D. M. (2005). Cellular abnormalities of blood vessels as targets in cancer. *Curr Opin Genet Dev*, *15*(1), 102-111. doi: 10.1016/j.gde.2004.12.005
- Barnes, R. M., & Firulli, A. B. (2009). A twist of insight - the role of Twist-family bHLH factors in development. *Int J Dev Biol*, *53*(7), 909-924. doi: 10.1387/ijdb.082747rb
- Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., . . . Bartek, J. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*, *434*(7035), 864-870. doi: 10.1038/nature03482
- Behrens, J., Mareel, M. M., Van Roy, F. M., & Birchmeier, W. (1989). Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. *J Cell Biol*, *108*(6), 2435-2447.
- Benezra, R., Davis, R. L., Lassar, A., Tapscott, S., Thayer, M., Lockshon, D., & Weintraub, H. (1990). Id: a negative regulator of helix-loop-helix DNA binding proteins. Control of terminal myogenic differentiation. *Ann N Y Acad Sci*, *599*, 1-11.
- Bergers, G., & Benjamin, L. E. (2003). Tumorigenesis and the angiogenic switch. *Nat Rev Cancer*, *3*(6), 401-410. doi: 10.1038/nrc1093
- Bergman, A., Sahlin, P., Emanuelsson, M., Caren, H., Tarnow, P., Martinsson, T., . . . Stenman, G. (2009). Germline mutation screening of the Saethre-Chotzen-associated genes TWIST1 and

- FGFR3 in families with BRCA1/2-negative breast cancer. *Scand J Plast Reconstr Surg Hand Surg*, 43(5), 251-255. doi: 10.3109/02844310903247228
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*, 28(1), 235-242.
- Bernards, R., & Weinberg, R. A. (2002). A progression puzzle. *Nature*, 418(6900), 823. doi: 10.1038/418823a
- Berx, G., & van Roy, F. (2009). Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol*, 1(6), a003129. doi: 10.1101/cshperspect.a003129
- Bhowmick, N. A., Neilson, E. G., & Moses, H. L. (2004). Stromal fibroblasts in cancer initiation and progression. *Nature*, 432(7015), 332-337. doi: 10.1038/nature03096
- Bialek, P., Kern, B., Yang, X., Schrock, M., Sobic, D., Hong, N., . . . Karsenty, G. (2004). A twist code determines the onset of osteoblast differentiation. *Dev Cell*, 6(3), 423-435.
- Biegging, K. T., & Attardi, L. D. (2012). Deconstructing p53 transcriptional networks in tumor suppression. *Trends Cell Biol*, 22(2), 97-106. doi: 10.1016/j.tcb.2011.10.006
- BioGRID. the Biological General Repository for Interaction Datasets.
- Bissell, M. J., & Hines, W. C. (2011). Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat Med*, 17(3), 320-329. doi: 10.1038/nm.2328
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., . . . Sondak, V. (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, 406(6795), 536-540. doi: 10.1038/35020115
- Blackburn, E. H., & Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J Mol Biol*, 120(1), 33-53.
- Blackwood, E. M., & Eisenman, R. N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science*, 251(4998), 1211-1217.
- Blasco, M. A. (2005). Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet*, 6(8), 611-622. doi: 10.1038/nrg1656
- Bodnar, A. G., Ouellette, M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., . . . Wright, W. E. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science*, 279(5349), 349-352.
- Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*, 3(7), 730-737.
- Bordet, I. J. Historique. *Origine du cancer*. Retrieved from Institut Jules Bordet institut website: <http://www.bordet.be/fr/presentation/historic/cancer/cancer5.htm>
- Bouard, C., Terreux, R., Honorat, M., Manship, B., Ansieau, S., Vigneron, A. M., . . . Payen, L. (2016). Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences. *Nucleic Acids Res*. doi: 10.1093/nar/gkw334
- Bouard C., Terreux R., Tissier A., Jacqueroud L., Vigneron A., Ansieau S., Puisieux A., Payen L. (2016b). Destabilization of the TWIST1/E12 complex dimerization following the R154P point-mutation of TWIST1: an in silico approach. *Soumission Journal of Biomolecular Structure & Dynamics*.
- Bouard, C., Terreux, R., Hope, J., Chemelle, J. A., Puisieux, A., Ansieau, S., & Payen, L. (2014). Interhelical loops within the bHLH domain are determinant in maintaining TWIST1-DNA complexes. *J Biomol Struct Dyn*, 32(2), 226-241. doi: 10.1080/07391102.2012.762722
- Bourgeois, P., Stoetzel, C., Bolcato-Bellemin, A. L., Mattei, M. G., & Perrin-Schmitt, F. (1996). The human H-twist gene is located at 7p21 and encodes a B-HLH protein that is 96% similar to its murine M-twist counterpart. *Mamm Genome*, 7(12), 915-917.
- Braig, M., Lee, S., Loddenkemper, C., Rudolph, C., Peters, A. H., Schlegelberger, B., . . . Schmitt, C. A. (2005). Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature*, 436(7051), 660-665. doi: 10.1038/nature03841
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. et Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimisation, and dynamics calculations. *J. Comp. Chem*, 4(2), 187-217.
- Burkhardt, D. L., & Sage, J. (2008). Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer*, 8(9), 671-682. doi: 10.1038/nrc2399

- Burns, T. F., Dobromilskaya, I., Murphy, S. C., Gajula, R. P., Thiyagarajan, S., Chatley, S. N., . . . Rudin, C. M. (2013). Inhibition of TWIST1 leads to activation of oncogene-induced senescence in oncogene-driven non-small cell lung cancer. *Mol Cancer Res*, *11*(4), 329-338. doi: 10.1158/1541-7786.MCR-12-0456
- Cai, J., Shoo, B. A., Sorauf, T., & Jabs, E. W. (2003). A novel mutation in the TWIST gene, implicated in Saethre-Chotzen syndrome, is found in the original case of Robinow-Sorauf syndrome. *Clin Genet*, *64*(1), 79-82.
- Cakouros, D., Isenmann, S., Cooper, L., Zannettino, A., Anderson, P., Glackin, C., & Gronthos, S. (2012). Twist-1 induces Ezh2 recruitment regulating histone methylation along the Ink4A/Arf locus in mesenchymal stem cells. *Mol Cell Biol*, *32*(8), 1433-1441. doi: 10.1128/MCB.06315-11
- (Canada), S. d. r. s. l. c. Faits saillants sur le cancer. *Historique du cancer*. Retrieved from Société de recherche sur le cancer website: <http://www.societederecherchesurlecancer.ca/page.aspx?pid=1968>
- cancer, Les cancers en France (Février 2015). <http://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Les-cancers-en-France-Edition-2014>
- Carmeliet, P. (2005). VEGF as a key mediator of angiogenesis in cancer. *Oncology*, *69 Suppl 3*, 4-10. doi: 10.1159/000088478
- Carmeliet, P., & Jain, R. K. (2000). Angiogenesis in cancer and other diseases. *Nature*, *407*(6801), 249-257. doi: 10.1038/35025220
- Carnero, A., Blanco-Aparicio, C., Renner, O., Link, W., & Leal, J. F. (2008). The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. *Curr Cancer Drug Targets*, *8*(3), 187-198.
- Castanon, I., & Baylies, M. K. (2002). A Twist in fate: evolutionary comparison of Twist structure and function. *Gene*, *287*(1-2), 11-22.
- Castanon, I., Von Stetina, S., Kass, J., & Baylies, M. K. (2001). Dimerization partners determine the activity of the Twist bHLH protein during Drosophila mesoderm development. *Development*, *128*(16), 3145-3159.
- Cavallaro, U. G. O., & Christofori, G. (2004). Multitasking in Tumor Progression: Signaling Functions of Cell Adhesion Molecules. *Annals of the New York Academy of Sciences*, *1014*(1), 58-66. doi: 10.1196/annals.1294.006
- Centonze, V. E., Firulli, B. A., & Firulli, A. B. (2004). Fluorescence Resonance Energy Transfer (FRET) as a method to calculate the dimerization strength of basic Helix-Loop-Helix (bHLH) proteins. *Biol Proced Online*, *6*, 78-82. doi: 10.1251/bpo75
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, *2*(5), 401-404. doi: 10.1158/2159-8290.CD-12-0095
- Cervantes-Barragan, D. E., Villarroel, C. E., Medrano-Hernandez, A., Duran-McKinster, C., Bosch-Canto, V., Del-Castillo, V., . . . Desnick, R. J. (2011). Setleis syndrome in Mexican-Nahua sibs due to a homozygous TWIST2 frameshift mutation and partial expression in heterozygotes: review of the focal facial dermal dysplasias and subtype reclassification. *J Med Genet*, *48*(10), 716-720. doi: 10.1136/jmedgenet-2011-100251
- Chang, A. T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J. W., . . . Yang, J. (2015). An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev*, *29*(6), 603-616. doi: 10.1101/gad.242842.114
- Che, Q., Liu, B. Y., Wang, F. Y., He, Y. Y., Lu, W., Liao, Y., . . . Wan, X. P. (2014). Interleukin 6 promotes endometrial cancer growth through an autocrine feedback loop involving ERK-NF-kappaB signaling pathway. *Biochem Biophys Res Commun*, *446*(1), 167-172. doi: 10.1016/j.bbrc.2014.02.080
- Chen, Z. F., & Behringer, R. R. (1995). twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev*, *9*(6), 686-699.
- Cheng, G. Z., Chan, J., Wang, Q., Zhang, W., Sun, C. D., & Wang, L. H. (2007). Twist transcriptionally up-regulates AKT2 in breast cancer cells leading to increased migration, invasion, and resistance to paclitaxel. *Cancer Res*, *67*(5), 1979-1987. doi: 10.1158/0008-5472.CAN-06-1479

- Cheng, N., Chytil, A., Shyr, Y., Joly, A., & Moses, H. L. (2008). Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion. *Mol Cancer Res*, *6*(10), 1521-1533. doi: 10.1158/1541-7786.MCR-07-2203
- Chin, K., de Solorzano, C. O., Knowles, D., Jones, A., Chou, W., Rodriguez, E. G., . . . Lockett, S. J. (2004). In situ analyses of genome instability in breast cancer. *Nat Genet*, *36*(9), 984-988. doi: 10.1038/ng1409
- Chin, L., Artandi, S. E., Shen, Q., Tam, A., Lee, S. L., Gottlieb, G. J., . . . DePinho, R. A. (1999). p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell*, *97*(4), 527-538.
- Chotzen, F. (1932). Eine eigenartige familiaere Entwicklungsstoerung (Akrocephalosyndaktylie, Dysostosis craniofacialis und Hypertelorismus).
- Clark, A. G., & Vignjevic, D. M. (2015). Modes of cancer cell invasion and the role of the microenvironment. *Curr Opin Cell Biol*, *36*, 13-22. doi: 10.1016/j.ceb.2015.06.004
- Collado, M., & Serrano, M. (2005). The senescent side of tumor suppression. *Cell Cycle*, *4*(12), 1722-1724.
- Connerney, J., Andreeva, V., Leshem, Y., Mercado, M. A., Dowell, K., Yang, X., . . . Spicer, D. B. (2008). Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure. *Dev Biol*, *318*(2), 323-334. doi: 10.1016/j.ydbio.2008.03.037
- Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M. A., & Spicer, D. B. (2006). Twist1 dimer selection regulates cranial suture patterning and fusion. *Dev Dyn*, *235*(5), 1345-1357. doi: 10.1002/dvdy.20717
- Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., . . . Piccolo, S. (2011). The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell*, *147*(4), 759-772. doi: 10.1016/j.cell.2011.09.048
- Cory, S., & Adams, J. M. (2002). The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer*, *2*(9), 647-656. doi: 10.1038/nrc883
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., & Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J*, *11*(5), 1921-1929.
- Counter, C. M., Hahn, W. C., Wei, W., Caddle, S. D., Beijersbergen, R. L., Lansdorp, P. M., . . . Weinberg, R. A. (1998). Dissociation among in vitro telomerase activity, telomere maintenance, and cellular immortalization. *Proc Natl Acad Sci U S A*, *95*(25), 14723-14728.
- Crews, S. T. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes Dev*, *12*(5), 607-620.
- Currie, D. A., & Bate, M. (1991). The development of adult abdominal muscles in *Drosophila*: myoblasts express twist and are associated with nerves. *Development*, *113*(1), 91-102.
- d'Adda di Fagagna, F., Reaper, P. M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., . . . Jackson, S. P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, *426*(6963), 194-198. doi: 10.1038/nature02118
- Davies, M. A., & Samuels, Y. (2010). Analysis of the genome to personalize therapy for melanoma. *Oncogene*, *29*(41), 5545-5555. doi: 10.1038/onc.2010.323
- De Craene, B., & Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer*, *13*(2), 97-110. doi: 10.1038/nrc3447
- de Lange, T. (2006). Lasker Laurels for telomerase. *Cell*, *126*(6), 1017-1020. doi: 10.1016/j.cell.2006.09.004
- De Masi, F., Grove, C. A., Vedenko, A., Alibes, A., Gisselbrecht, S. S., Serrano, L., . . . Walhout, A. J. (2011). Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res*, *39*(11), 4553-4563. doi: 10.1093/nar/gkr070
- Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., & Panne, D. (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat Struct Mol Biol*, *20*(9), 1040-1046. doi: 10.1038/nsmb.2642
- Demontis, S., Rigo, C., Piccinin, S., Mizzau, M., Sonogo, M., Fabris, M., . . . Maestro, R. (2006). Twist is substrate for caspase cleavage and proteasome-mediated degradation. *Cell Death Differ*, *13*(2), 335-345. doi: 10.1038/sj.cdd.4401744

- Deng, Y., Chan, S. S., & Chang, S. (2008). Telomere dysfunction and tumour suppression: the senescence connection. *Nat Rev Cancer*, 8(6), 450-458. doi: 10.1038/nrc2393
- Deshpande, A., Sicinski, P., & Hinds, P. W. (2005). Cyclins and cdks in development and cancer: a perspective. *Oncogene*, 24(17), 2909-2915. doi: 10.1038/sj.onc.1208618
- Diaz, V. M., Vinas-Castells, R., & Garcia de Herreros, A. (2014). Regulation of the protein stability of EMT transcription factors. *Cell Adh Migr*, 8(4), 418-428. doi: 10.4161/19336918.2014.969998
- Dingwall, C., & Laskey, R. A. (1991). Nuclear targeting sequences--a consensus? *Trends Biochem Sci*, 16(12), 478-481.
- Dreyer, C., Sablin, M. P., Faivre, S., & Raymond, E. (2009). [Topics in mTOR pathway and its inhibitors]. *Bull Cancer*, 96(1), 87-94. doi: 10.1684/bdc.2009.0807
- Ducray, C., Pommier, J. P., Martins, L., Boussin, F. D., & Sabatier, L. (1999). Telomere dynamics, end-to-end fusions and telomerase activation during the human fibroblast immortalization process. *Oncogene*, 18(29), 4211-4223. doi: 10.1038/sj.onc.1202797
- El Ghouzzi, V., Lajeunie, E., Le Merrer, M., Cormier-Daire, V., Renier, D., Munnich, A., & Bonaventure, J. (1999). Mutations within or upstream of the basic helix-loop-helix domain of the TWIST gene are specific to Saethre-Chotzen syndrome. *Eur J Hum Genet*, 7(1), 27-33. doi: 10.1038/sj.ejhg.5200240
- el Ghouzzi, V., Le Merrer, M., Perrin-Schmitt, F., Lajeunie, E., Benit, P., Renier, D., . . . Bonaventure, J. (1997). Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat Genet*, 15(1), 42-46. doi: 10.1038/ng0197-42
- El Ghouzzi, V., Legeai-Mallet, L., Aresta, S., Benoist, C., Munnich, A., de Gunzburg, J., & Bonaventure, J. (2000). Saethre-Chotzen mutations cause TWIST protein degradation or impaired nuclear location. *Hum Mol Genet*, 9(5), 813-819.
- El Ghouzzi, V., Legeai-Mallet, L., Benoist-Lasselain, C., Lajeunie, E., Renier, D., Munnich, A., & Bonaventure, J. (2001). Mutations in the basic domain and the loop-helix II junction of TWIST abolish DNA binding in Saethre-Chotzen syndrome. *FEBS Lett*, 492(1-2), 112-118.
- Elanko, N., Sibbring, J. S., Metcalfe, K. A., Clayton-Smith, J., Donnai, D., Temple, I. K., . . . Wilkie, A. O. (2001). A survey of TWIST for mutations in craniosynostosis reveals a variable length polyglycine tract in asymptomatic individuals. *Hum Mutat*, 18(6), 535-541. doi: 10.1002/humu.1230
- Ellenberger, T., Fass, D., Arnaud, M., & Harrison, S. C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev*, 8(8), 970-980.
- Ephrussi, A., Church, G. M., Tonegawa, S., & Gilbert, W. (1985). B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. *Science*, 227(4683), 134-140.
- Evan, G. I., & Vousden, K. H. (2001). Proliferation, cell cycle and apoptosis in cancer. *Nature*, 411(6835), 342-348. doi: 10.1038/35077213
- Evan, G. I., Wyllie, A. H., Gilbert, C. S., Littlewood, T. D., Land, H., Brooks, M., . . . Hancock, D. C. (1992). Induction of apoptosis in fibroblasts by c-myc protein. *Cell*, 69(1), 119-128.
- Feng, M. Y., Wang, K., Song, H. T., Yu, H. W., Qin, Y., Shi, Q. T., & Geng, J. S. (2009). Metastasis-induction and apoptosis-protection by TWIST in gastric cancer cells. *Clin Exp Metastasis*, 26(8), 1013-1023. doi: 10.1007/s10585-009-9291-6
- Ferrara, N. (2009). Vascular endothelial growth factor. *Arterioscler Thromb Vasc Biol*, 29(6), 789-791. doi: 10.1161/ATVBAHA.108.179663
- Ferrara, N., Chen, H., Davis-Smyth, T., Gerber, H. P., Nguyen, T. N., Peers, D., . . . Schwall, R. H. (1998). Vascular endothelial growth factor is essential for corpus luteum angiogenesis. *Nat Med*, 4(3), 336-340.
- Ferrara, N., & Henzel, W. J. (1989). Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. *Biochem Biophys Res Commun*, 161(2), 851-858.
- Firulli, A. B., & Conway, S. J. (2008). Phosphoregulation of Twist1 provides a mechanism of cell fate control. *Curr Med Chem*, 15(25), 2641-2647.
- Firulli, B. A., Hadzic, D. B., McDaid, J. R., & Firulli, A. B. (2000). The basic helix-loop-helix transcription factors dHAND and eHAND exhibit dimerization characteristics that suggest

- complex regulation of function. *J Biol Chem*, 275(43), 33567-33573. doi: 10.1074/jbc.M005888200
- Firulli, B. A., Howard, M. J., McDavid, J. R., McIlreavey, L., Dionne, K. M., Centonze, V. E., . . . Firulli, A. B. (2003). PKA, PKC, and the protein phosphatase 2A influence HAND factor function: a mechanism for tissue-specific transcriptional regulation. *Mol Cell*, 12(5), 1225-1237.
- Firulli, B. A., Krawchuk, D., Centonze, V. E., Vargesson, N., Virshup, D. M., Conway, S. J., . . . Firulli, A. B. (2005). Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nat Genet*, 37(4), 373-381. doi: 10.1038/ng1525
- Firulli, B. A., Redick, B. A., Conway, S. J., & Firulli, A. B. (2007). Mutations within helix I of Twist1 result in distinct limb defects and variation of DNA binding affinities. *J Biol Chem*, 282(37), 27536-27546. doi: 10.1074/jbc.M702613200
- Folkman, J. (2002). Role of angiogenesis in tumor growth and metastasis. *Semin Oncol*, 29(6 Suppl 16), 15-18. doi: 10.1053/sonc.2002.37263
- Folkman, J. (2006). Angiogenesis. *Annu Rev Med*, 57, 1-18. doi: 10.1146/annurev.med.57.121304.131306
- Folkman, J., & Shing, Y. (1992). Angiogenesis. *J Biol Chem*, 267(16), 10931-10934.
- Fong, A. P., Yao, Z., Zhong, J. W., Cao, Y., Ruzzo, W. L., Gentleman, R. C., & Tapscott, S. J. (2012). Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev Cell*, 22(4), 721-735. doi: 10.1016/j.devcel.2012.01.015
- Franco, H. L., Casanovas, J., Rodriguez-Medina, J. R., & Cadilla, C. L. (2011). Redundant or separate entities?--roles of Twist1 and Twist2 as molecular switches during gene transcription. *Nucleic Acids Res*, 39(4), 1177-1186. doi: 10.1093/nar/gkq890
- Friedl, P., & Wolf, K. (2010). Plasticity of cell migration: a multiscale tuning model. *J Cell Biol*, 188(1), 11-19. doi: 10.1083/jcb.200909003
- Fu, J., Qin, L., He, T., Qin, J., Hong, J., Wong, J., . . . Xu, J. (2011). The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Res*, 21(2), 275-289. doi: 10.1038/cr.2010.118
- Funato, N., Twigg, S. R., Higashihori, N., Ohyama, K., Wall, S. A., Wilkie, A. O., & Nakamura, M. (2005). Functional analysis of natural mutations in two TWIST protein motifs. *Hum Mutat*, 25(6), 550-556. doi: 10.1002/humu.20176
- Gajula, R. P., Chettiar, S. T., Williams, R. D., Nugent, K., Kato, Y., Wang, H., . . . Tran, P. T. (2015). Structure-function studies of the bHLH phosphorylation domain of TWIST1 in prostate cancer cells. *Neoplasia*, 17(1), 16-31. doi: 10.1016/j.neo.2014.10.009
- Galler, K. M., Yasue, A., Cavender, A. C., Bialek, P., Karsenty, G., & D'Souza, R. N. (2007). A novel role for Twist-1 in pulp homeostasis. *J Dent Res*, 86(10), 951-955.
- Galluzzi, L., & Kroemer, G. (2008). Necroptosis: a specialized pathway of programmed necrosis. *Cell*, 135(7), 1161-1163. doi: 10.1016/j.cell.2008.12.004
- Galluzzi, L., Pietrocola, F., Bravo-San Pedro, J. M., Amaravadi, R. K., Baehrecke, E. H., Cecconi, F., . . . Kroemer, G. (2015). Autophagy in malignant transformation and cancer progression. *EMBO J*, 34(7), 856-880. doi: 10.15252/embj.201490784
- Gao, D., Vahdat, L. T., Wong, S., Chang, J. C., & Mittal, V. (2012). Microenvironmental regulation of epithelial-mesenchymal transitions in cancer. *Cancer Res*, 72(19), 4883-4889. doi: 10.1158/0008-5472.CAN-12-1223
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, 6(269), p11. doi: 10.1126/scisignal.2004088
- Gavard, J., Marthiens, V., Monnet, C., Lambert, M., & Mege, R. M. (2004). N-cadherin activation substitutes for the cell contact control in cell cycle arrest and myogenic differentiation: involvement of p120 and beta-catenin. *J Biol Chem*, 279(35), 36795-36802. doi: 10.1074/jbc.M401705200
- Genomics, cBioPortal for Cancer Genomics <http://www.cbioportal.org>
- Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., . . . Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*, 366(10), 883-892. doi: 10.1056/NEJMoa1113205

- Germanguz, I., Lev, D., Waisman, T., Kim, C. H., & Gitelman, I. (2007). Four twist genes in zebrafish, four expression patterns. *Dev Dyn*, 236(9), 2615-2626. doi: 10.1002/dvdy.21267
- Gire, V. (2005). [Senescence: a telomeric limit to immortality or a cellular response to physiologic stresses?]. *Med Sci (Paris)*, 21(5), 491-497. doi: 10.1051/medsci/2005215491
- Girisha, K. M., Bidchol, A. M., Sarpangala, M. K., & Satyamoorthy, K. (2014). A novel frameshift mutation in TWIST2 gene causing Setleis syndrome. *Indian J Pediatr*, 81(3), 302-304. doi: 10.1007/s12098-013-1253-y
- Gitelman, I. (1997). Twist protein in mouse embryogenesis. *Dev Biol*, 189(2), 205-214. doi: 10.1006/dbio.1997.8614
- Gitelman, I. (2007). Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. *Mol Biol Evol*, 24(9), 1912-1925. doi: 10.1093/molbev/msm120
- Goede, V., Schmidt, T., Kimmuna, S., Kozian, D., & Augustin, H. G. (1998). Analysis of blood vessel maturation processes during cyclic ovarian angiogenesis. *Lab Invest*, 78(11), 1385-1394.
- Gordan, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., & Bulyk, M. L. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep*, 3(4), 1093-1104. doi: 10.1016/j.celrep.2013.03.014
- Gorgoulis, V. G., Vassiliou, L. V., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., . . . Halazonetis, T. D. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature*, 434(7035), 907-913. doi: 10.1038/nature03485
- Grabiner, B. C., Nardi, V., Birsoy, K., Possemato, R., Shen, K., Sinha, S., . . . Sabatini, D. M. (2014). A diverse array of cancer-associated MTOR mutations are hyperactivating and can predict rapamycin sensitivity. *Cancer Discov*, 4(5), 554-563. doi: 10.1158/2159-8290.CD-13-0929
- Grau, Y., Carteret, C., & Simpson, P. (1984). Mutations and Chromosomal Rearrangements Affecting the Expression of Snail, a Gene Involved in Embryonic Patterning in DROSOPHILA MELANOGASTER. *Genetics*, 108(2), 347-360.
- Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, 43(2 Pt 1), 405-413.
- Gripp, K. W., Stolle, C. A., Celle, L., McDonald-McGinn, D. M., Whitaker, L. A., Zackai, E. H. (1999). TWIST gene mutation in a patient with radial aplasia and craniosynostosis: further evidence for heterogeneity of Baller-Gerold syndrome. *Am J Med Genet*, 82(2), 170-176.
- Gripp, K. W., Zackai, E. H., & Stolle, C. A. (2000). Mutations in the human TWIST gene. *Hum Mutat*, 15(2), 150-155. doi: 10.1002/(SICI)1098-1004(200002)15:2<150::AID-HUMU3>3.0.CO;2-D
- Grivennikov, S. I., Greten, F. R., & Karin, M. (2010). Immunity, inflammation, and cancer. *Cell*, 140(6), 883-899. doi: 10.1016/j.cell.2010.01.025
- Grove, C. A., De Masi, F., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. (2009). A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell*, 138(2), 314-327. doi: 10.1016/j.cell.2009.04.058
- Haapala, K., Kuukasjarvi, T., Hyytinen, E., Rantala, I., Helin, H. J., & Koivisto, P. A. (2007). Androgen receptor amplification is associated with increased cell proliferation in prostate cancer. *Hum Pathol*, 38(3), 474-478. doi: 10.1016/j.humpath.2006.09.008
- Halgren, T. A., & BNachbar, R. B. (1996). Merck molecular force field. IV. conformational energies and geometries for MMFF94. *Journal of Computational Chemistry*
- Hamamori, Y., Sartorelli, V., Ogryzko, V., Puri, P. L., Wu, H. Y., Wang, J. Y., . . . Kedes, L. (1999). Regulation of histone acetyltransferases p300 and PCAF by the bHLH protein twist and adenoviral oncoprotein E1A. *Cell*, 96(3), 405-413.
- Hamamori, Y., Wu, H. Y., Sartorelli, V., & Kedes, L. (1997). The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist. *Mol Cell Biol*, 17(11), 6563-6573.
- Hanahan, D., & Folkman, J. (1996). Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*, 86(3), 353-364.
- Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1), 57-70.
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646-674. doi: 10.1016/j.cell.2011.02.013

- Hansson, A., Manetopoulos, C., Jonsson, J. I., & Axelson, H. (2003). The basic helix-loop-helix transcription factor TAL1/SCL inhibits the expression of the p16INK4A and pTalpha genes. *Biochem Biophys Res Commun*, *312*(4), 1073-1081.
- Harley, C. B., Futcher, A. B., & Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature*, *345*(6274), 458-460. doi: 10.1038/345458a0
- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D. K., & Allshire, R. C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature*, *346*(6287), 866-868. doi: 10.1038/346866a0
- Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res*, *25*, 585-621.
- Hebrok, M., Wertz, K., & Fuchtbauer, E. M. (1994). M-twist is an inhibitor of muscle differentiation. *Dev Biol*, *165*(2), 537-544. doi: 10.1006/dbio.1994.1273
- Hitomi, J., Christofferson, D. E., Ng, A., Yao, J., Degtarev, A., Xavier, R. J., & Yuan, J. (2008). Identification of a molecular signaling network that regulates a cellular necrotic cell death pathway. *Cell*, *135*(7), 1311-1323. doi: 10.1016/j.cell.2008.10.044
- Hoglund, M., Gisselsson, D., Sall, T., & Mitelman, F. (2002). Coping with complexity. multivariate analysis of tumor karyotypes. *Cancer Genet Cytogenet*, *135*(2), 103-109.
- Holash, J., Maisonpierre, P. C., Compton, D., Boland, P., Alexander, C. R., Zagzag, D., . . . Wiegand, S. J. (1999). Vessel cooption, regression, and growth in tumors mediated by angiopoietins and VEGF. *Science*, *284*(5422), 1994-1998.
- Hong, J., Zhou, J., Fu, J., He, T., Qin, J., Wang, L., . . . Xu, J. (2011). Phosphorylation of serine 68 of Twist1 by MAPKs stabilizes Twist1 protein and promotes breast cancer cell invasiveness. *Cancer Res*, *71*(11), 3980-3990. doi: 10.1158/0008-5472.CAN-10-2914
- Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., . . . Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*, *40*(Database issue), D261-270. doi: 10.1093/nar/gkr1122
- Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*, *43*(Database issue), D512-520. doi: 10.1093/nar/gku1267
- Howard, T. D., Paznekas, W. A., Green, E. D., Chiang, L. C., Ma, N., Ortiz de Luna, R. I., . . . Jabs, E. W. (1997). Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nat Genet*, *15*(1), 36-41. doi: 10.1038/ng0197-36
- Hugo, H., Ackland, M. L., Blick, T., Lawrence, M. G., Clements, J. A., Williams, E. D., & Thompson, E. W. (2007). Epithelial--mesenchymal and mesenchymal--epithelial transitions in carcinoma progression. *J Cell Physiol*, *213*(2), 374-383. doi: 10.1002/jcp.21223
- Iliopoulos, D., Hirsch, H. A., & Struhl, K. (2009). An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell*, *139*(4), 693-706. doi: 10.1016/j.cell.2009.10.014
- Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K., & Levine, M. (1992). dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. *Genes Dev*, *6*(8), 1518-1530.
- Isenmann, S., Arthur, A., Zannettino, A. C., Turner, J. L., Shi, S., Glackin, C. A., & Gronthos, S. (2009). TWIST family of basic helix-loop-helix transcription factors mediate human mesenchymal stem cell growth and commitment. *Stem Cells*, *27*(10), 2457-2468. doi: 10.1002/stem.181
- Jacqueroud, L., Bouard, C., Richard, G., Payen, L., Devouassoux-Shisheboran, M., Spicer, D., . . . Ansieau, S. (2016). The heterodimeric TWIST1-E12 complex drives the oncogenic potential of TWIST1 in human mammary epithelial cells. *Neoplasia (NEO-D-15-00360R1)*.
- Jain, R. K., & Carmeliet, P. (2012). SnapShot: Tumor angiogenesis. *Cell*, *149*(6), 1408-1408 e1401. doi: 10.1016/j.cell.2012.05.025
- James, P. A., Culling, B., Mullan, G., Jenkins, M., Elakis, G., Turner, A. M., . . . Roscioli, T. (2009). Breast cancer risk is not increased in individuals with TWIST1 mutation confirmed Saethre-Chotzen syndrome: an Australian multicenter study. *Genes Chromosomes Cancer*, *48*(7), 533-538. doi: 10.1002/gcc.20661

- Jan, Y. N., & Jan, L. Y. (1993). HLH proteins, fly neurogenesis, and vertebrate myogenesis. *Cell*, 75(5), 827-830.
- Jiang, B. H., & Liu, L. Z. (2009). Chapter 2 PI3K/PTEN Signaling in Angiogenesis and Tumorigenesis. *102*, 19-65. doi: 10.1016/s0065-230x(09)02002-8
- Jolly, M. K., Boareto, M., Huang, B., Jia, D., Lu, M., Ben-Jacob, E., . . . Levine, H. (2015). Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Front Oncol*, 5, 155. doi: 10.3389/fonc.2015.00155
- Jones, N. (1990). Transcriptional regulation by dimerization: two sides to an incestuous relationship. *Cell*, 61(1), 9-11.
- Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol*, 5(6), 226. doi: 10.1186/gb-2004-5-6-226
- Junttila, M. R., & Evan, G. I. (2009). p53--a Jack of all trades but master of none. *Nat Rev Cancer*, 9(11), 821-829. doi: 10.1038/nrc2728
- Kadesch, T. (1993). Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth Differ*, 4(1), 49-55.
- Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J Clin Invest*, 119(6), 1420-1428. doi: 10.1172/JCI39104
- Kazerounian, S., Yee, K. O., & Lawler, J. (2008). Thrombospondins in cancer. *Cell Mol Life Sci*, 65(5), 700-712. doi: 10.1007/s00018-007-7486-z
- Kenific, C. M., & Debnath, J. (2015). Cellular and metabolic functions for autophagy in cancer cells. *Trends Cell Biol*, 25(1), 37-45. doi: 10.1016/j.tcb.2014.09.001
- Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., . . . Shay, J. W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science*, 266(5193), 2011-2015.
- Kimmelman, A. C. (2011). The dynamic nature of autophagy in cancer. *Genes Dev*, 25(19), 1999-2010. doi: 10.1101/gad.17558811
- Klammer, M., Kaminski, M., Zedler, A., Oppermann, F., Blencke, S., Marx, S., . . . Schaab, C. (2012). Phosphosignature predicts dasatinib response in non-small cell lung cancer. *Mol Cell Proteomics*, 11(9), 651-668. doi: 10.1074/mcp.M111.016410
- Klingelutz, A. J., Barber, S. A., Smith, P. P., Dyer, K., & McDougall, J. K. (1994). Restoration of telomeres in human papillomavirus-immortalized human anogenital epithelial cells. *Mol Cell Biol*, 14(2), 961-969.
- Knudsen, E. S., & Knudsen, K. E. (2008). Tailoring to RB: tumour suppressor status and therapeutic response. *Nat Rev Cancer*, 8(9), 714-724. doi: 10.1038/nrc2401
- Kress, W., Schropp, C., Lieb, G., Petersen, B., Busse-Ratzka, M., Kunz, J., . . . Collmann, H. (2006). Saethre-Chotzen syndrome caused by TWIST 1 gene mutations: functional differentiation from Muenke coronal synostosis syndrome. *Eur J Hum Genet*, 14(1), 39-48. doi: 10.1038/sj.ejhg.5201507
- Kwok, W. K., Ling, M. T., Lee, T. W., Lau, T. C., Zhou, C., Zhang, X., . . . Wang, X. (2005). Up-regulation of TWIST in prostate cancer and its implication as a therapeutic target. *Cancer Res*, 65(12), 5153-5162. doi: 10.1158/0008-5472.CAN-04-3785
- Lander, R., Nordin, K., & LaBonne, C. (2011). The F-box protein Ppa is a common regulator of core EMT factors Twist, Snail, Slug, and Sip1. *J Cell Biol*, 194(1), 17-25. doi: 10.1083/jcb.201012085
- Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., . . . Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464), 645-648. doi: 10.1038/367645a0
- Larsen, C. J. (2004). [pRB, p53, p16INK4a, senescence and malignant transformation]. *Bull Cancer*, 91(5), 399-402.
- Ledent, V., Paquet, O., & Vervoort, M. (2002). Phylogenetic analysis of the human basic helix-loop-helix proteins. *Genome Biol*, 3(6), RESEARCH0030.
- Lee, M. S., Lowe, G., Flanagan, S., Kuchler, K., & Glackin, C. A. (2000). Human Dermo-1 has attributes similar to twist in early bone development. *Bone*, 27(5), 591-602.
- Lee, T. K., Poon, R. T., Yuen, A. P., Ling, M. T., Kwok, W. K., Wang, X. H., . . . Fan, S. T. (2006). Twist overexpression correlates with hepatocellular carcinoma metastasis through induction of

- epithelial-mesenchymal transition. *Clin Cancer Res*, 12(18), 5369-5376. doi: 10.1158/1078-0432.CCR-05-2722
- Leptin, M. (1991). twist and snail as positive and negative regulators during Drosophila mesoderm development. *Genes Dev*, 5(9), 1568-1576.
- Leshem, Y., Spicer, D. B., Gal-Levi, R., & Halevy, O. (2000). Hepatocyte growth factor (HGF) inhibits skeletal muscle cell differentiation: a role for the bHLH protein twist and the cdk inhibitor p27. *J Cell Physiol*, 184(1), 101-109. doi: 10.1002/(SICI)1097-4652(200007)184:1<101::AID-JCP11>3.0.CO;2-D
- Leslie, N. R., & Longy, M. (2016). Inherited PTEN mutations and the prediction of phenotype. *Semin Cell Dev Biol*, 52, 30-38. doi: 10.1016/j.semcdb.2016.01.030
- Levine, B., & Kroemer, G. (2008). Autophagy in the pathogenesis of disease. *Cell*, 132(1), 27-42. doi: 10.1016/j.cell.2007.12.018
- Li, L., Cserjesi, P., & Olson, E. N. (1995). Dermo-1: a novel twist-related bHLH protein expressed in the developing dermis. *Dev Biol*, 172(1), 280-292. doi: 10.1006/dbio.1995.0023
- Li, Q. Q., Xu, J. D., Wang, W. J., Cao, X. X., Chen, Q., Tang, F., . . . Xu, Z. D. (2009). Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells. *Clin Cancer Res*, 15(8), 2657-2665. doi: 10.1158/1078-0432.CCR-08-2372
- Li, S., Kendall, S. E., Raices, R., Finlay, J., Covarrubias, M., Liu, Z., . . . Glackin, C. A. (2012). TWIST1 associates with NF-kappaB subunit RELA via carboxyl-terminal WR domain to promote cell autonomous invasion through IL8 production. *BMC Biol*, 10, 73. doi: 10.1186/1741-7007-10-73
- Longo, A., Guanga, G. P., & Rose, R. B. (2008). Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry*, 47(1), 218-229. doi: 10.1021/bi701527r
- Lowe, S. W., Cepero, E., & Evan, G. (2004). Intrinsic tumour suppression. *Nature*, 432(7015), 307-315. doi: 10.1038/nature03098
- Lu, S., Nie, J., Luan, Q., Feng, Q., Xiao, Q., Chang, Z., . . . Yang, Z. (2011). Phosphorylation of the Twist1-family basic helix-loop-helix transcription factors is involved in pathological cardiac remodeling. *PLoS One*, 6(4), e19251. doi: 10.1371/journal.pone.0019251
- Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F., & Groom, A. C. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol*, 153(3), 865-873. doi: 10.1016/S0002-9440(10)65628-3
- Ma, W., Lu, S., Sun, T., Wang, X., Ma, Y., Zhang, X., . . . Wang, Y. (2014). Twist 1 regulates the expression of PPARgamma during hormone-induced 3T3-L1 preadipocyte differentiation: a possible role in obesity and associated diseases. *Lipids Health Dis*, 13, 132. doi: 10.1186/1476-511X-13-132
- Mac Gabhann, F., & Popel, A. S. (2008). Systems biology of vascular endothelial growth factors. *Microcirculation*, 15(8), 715-738. doi: 10.1080/10739680802095964
- Maestro, R., Dei Tos, A. P., Hamamori, Y., Krasnokutsky, S., Sartorelli, V., Kedes, L., . . . Hannon, G. J. (1999). Twist is a potential oncogene that inhibits apoptosis. *Genes Dev*, 13(17), 2207-2217.
- Magee, J. A., Piskounova, E., & Morrison, S. J. (2012). Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell*, 21(3), 283-296. doi: 10.1016/j.ccr.2012.03.003
- Maiuri, M. C., Tasdemir, E., Criollo, A., Morselli, E., Vicencio, J. M., Carnuccio, R., & Kroemer, G. (2009). Control of autophagy by oncogenes and tumor suppressor genes. *Cell Death Differ*, 16(1), 87-93. doi: 10.1038/cdd.2008.131
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., . . . Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133(4), 704-715. doi: 10.1016/j.cell.2008.03.027
- Maniotis, A. J., Folberg, R., Hess, A., Sefror, E. A., Gardner, L. M. G., Pe'er, J., . . . Hendrix, M. J. C. (1999). Vascular Channel Formation by Human Melanoma Cells in Vivo and in Vitro: Vasculogenic Mimicry. *The American Journal of Pathology*, 155(3), 739-752. doi: 10.1016/s0002-9440(10)65173-5

- Martin, S. J., & Green, D. R. (1995). Apoptosis and cancer: the failure of controls on cell death and cell survival. *Crit Rev Oncol Hematol*, *18*(2), 137-153.
- Massari, M. E., & Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol*, *20*(2), 429-440.
- Masutomi, K., & Hahn, W. C. (2003). Telomerase and tumorigenesis. *Cancer Letters*, *194*(2), 163-172. doi: 10.1016/s0304-3835(02)00703-6
- Masutomi, K., Yu, E. Y., Khurts, S., Ben-Porath, I., Currier, J. L., Metz, G. B., . . . Hahn, W. C. (2003). Telomerase maintains telomere structure in normal human cells. *Cell*, *114*(2), 241-253.
- Mathew, R., Karantza-Wadsworth, V., & White, E. (2007). Role of autophagy in cancer. *Nat Rev Cancer*, *7*(12), 961-967. doi: 10.1038/nrc2254
- McGaughran, J., & Aftimos, S. (2002). Setleis syndrome: three new cases and a review of the literature. *Am J Med Genet*, *111*(4), 376-380. doi: 10.1002/ajmg.10632
- McKeen, E. A., Mulvihill, J. J., Levine, P. H., Dean, J. H., & Howley, P. M. (1984). The concurrence of Saethre-Chotzen syndrome and malignancy in a family with in vitro immune dysfunction. *Cancer*, *54*(12), 2946-2951.
- Meek, D. W. (2015). Regulation of the p53 response and its relationship to cancer. *Biochem J*, *469*(3), 325-346. doi: 10.1042/BJ20150517
- Meng, T., Huang, Y., Wang, S., Zhang, H., Dechow, P. C., Wang, X., . . . Lu, Y. (2015). Twist1 Is Essential for Tooth Morphogenesis and Odontoblast Differentiation. *J Biol Chem*, *290*(49), 29593-29602. doi: 10.1074/jbc.M115.680546
- Mironchik, Y., Winnard, P. T., Jr., Vesuna, F., Kato, Y., Wildes, F., Pathak, A. P., . . . Raman, V. (2005). Twist overexpression induces in vivo angiogenesis and correlates with chromosomal instability in breast cancer. *Cancer Res*, *65*(23), 10801-10809. doi: 10.1158/0008-5472.CAN-05-0712
- Mizushima, N. (2007). Autophagy: process and function. *Genes Dev*, *21*(22), 2861-2873. doi: 10.1101/gad.1599207
- Morel, A. P., Hinkal, G. W., Thomas, C., Fauvet, F., Courtois-Cox, S., Wierinckx, A., . . . Puisieux, A. (2012). EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS Genet*, *8*(5), e1002723. doi: 10.1371/journal.pgen.1002723
- Morel, A. P., Lievre, M., Thomas, C., Hinkal, G., Ansieau, S., & Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One*, *3*(8), e2888. doi: 10.1371/journal.pone.0002888
- Muller, H. (1938). The remaking of chromosomes. *The Collecting Net. Woods Hole*, *13*: 181-198.
- Muller, P. A., & Vousden, K. H. (2014). Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell*, *25*(3), 304-317. doi: 10.1016/j.ccr.2014.01.021
- Muntoni, A., & Reddel, R. R. (2005). The first molecular details of ALT in human tumor cells. *Hum Mol Genet*, *14 Spec No. 2*, R191-196. doi: 10.1093/hmg/ddi266
- Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., . . . Stuiver, M. H. (1994). Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta*, *1218*(2), 129-135.
- Murre, C., McCaw, P. S., & Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell*, *56*(5), 777-783.
- Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., . . . et al. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell*, *58*(3), 537-544.
- N., A. (1977). Conformational Analysis. 130. MM2. A hydrocarbone force field utilizing V1 and V2 torsional terms. *J. Amer. Chem. Soc.*, *99*(25), 8127-8134.
- Nagy, J. A., Chang, S. H., Shih, S. C., Dvorak, A. M., & Dvorak, H. F. (2010). Heterogeneity of the tumor vasculature. *Semin Thromb Hemost*, *36*(3), 321-331. doi: 10.1055/s-0030-1253454
- Nendza, M., Gabbert, S., Kuhne, R., Lombardo, A., Roncaglioni, A., Benfenati, E., . . . Schuurmann, G. (2013). A comparative survey of chemistry-driven in silico methods to identify hazardous

- substances under REACH. *Regul Toxicol Pharmacol*, 66(3), 301-314. doi: 10.1016/j.yrtph.2013.05.007
- Niesner, U., Albrecht, I., Janke, M., Doebis, C., Loddenkemper, C., Lexberg, M. H., . . . Radbruch, A. (2008). Autoregulation of Th1-mediated inflammation by twist1. *J Exp Med*, 205(8), 1889-1901. doi: 10.1084/jem.20072468
- Nieto, M. A. (2013). Epithelial plasticity: a common theme in embryonic and cancer cells. *Science*, 342(6159), 1234850. doi: 10.1126/science.1234850
- Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., . . . Breast Cancer Working Group of the International Cancer Genome, C. (2012). The life history of 21 breast cancers. *Cell*, 149(5), 994-1007. doi: 10.1016/j.cell.2012.04.023
- Norman L. Allinger, Fanbing Li, Liqun Yan, & Tai, J. C. (1990). Molecular mechanics (MM3) calculations on conjugated hydrocarbons. *Journal of Computational Chemistry*.
- Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260), 23-28.
- Nüsslein-Volhard. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. *Wilhelm Roux's archives of developmental biology*, 193(5), pp 267-282. doi: 10.1007/BF00848156
- Nyberg, P., Xie, L., & Kalluri, R. (2005). Endogenous inhibitors of angiogenesis. *Cancer Res*, 65(10), 3967-3979. doi: 10.1158/0008-5472.CAN-04-2427
- Olive, K. P., Jacobetz, M. A., Davidson, C. J., Gopinathan, A., McIntyre, D., Honess, D., . . . Tuveson, D. A. (2009). Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science*, 324(5933), 1457-1461. doi: 10.1126/science.1171362
- Ozdemir, A., Fisher-Aylor, K. I., Pepke, S., Samanta, M., Dunipace, L., McCue, K., . . . Stathopoulos, A. (2011). High resolution mapping of Twist to DNA in *Drosophila* embryos: Efficient functional analysis and evolutionary conservation. *Genome Res*, 21(4), 566-577. doi: 10.1101/gr.104018.109
- Pan, D., Fujimoto, M., Lopes, A., & Wang, Y. X. (2009). Twist-1 is a PPARdelta-inducible, negative-feedback regulator of PGC-1alpha in brown fat metabolism. *Cell*, 137(1), 73-86. doi: 10.1016/j.cell.2009.01.051
- Paznekas, W. A., Cunningham, M. L., Howard, T. D., Korf, B. R., Lipson, M. H., Grix, A. W., . . . Jabs, E. W. (1998). Genetic heterogeneity of Saethre-Chotzen syndrome, due to TWIST and FGFR mutations. *Am J Hum Genet*, 62(6), 1370-1380. doi: 10.1086/301855
- PDB. Protein Data Bank. <http://www.rcsb.org/>.
- Perrin-Schmitt, F., Bolcato-Bellemin, A. L., Bourgeois, P., Stoetzel, C., & Danse, J. M. (1997). The locations of the H-twist and H-dermo-1 genes are distinct on the human genome. *Biochim Biophys Acta*, 1360(1), 1-2.
- Pettersson, A. T., Laurencikiene, J., Mejhert, N., Naslund, E., Bouloumie, A., Dahlman, I., . . . Ryden, M. (2010). A possible inflammatory role of twist1 in human white adipocytes. *Diabetes*, 59(3), 564-571. doi: 10.2337/db09-0997
- Pettersson, A. T., Mejhert, N., Jernas, M., Carlsson, L. M., Dahlman, I., Laurencikiene, J., . . . Ryden, M. (2011). Twist1 in human white adipose tissue and obesity. *J Clin Endocrinol Metab*, 96(1), 133-141. doi: 10.1210/jc.2010-0929
- Piccinin, S., Tonin, E., Sessa, S., Demontis, S., Rossi, S., Pecciarini, L., . . . Maestro, R. (2012). A "twist box" code of p53 inactivation: twist box: p53 interaction promotes p53 degradation. *Cancer Cell*, 22(3), 404-415. doi: 10.1016/j.ccr.2012.08.003
- Polyak, K., Haviv, I., & Campbell, I. G. (2009). Co-evolution of tumor cells and their microenvironment. *Trends Genet*, 25(1), 30-38. doi: 10.1016/j.tig.2008.10.012
- Popov, N., & Gil, J. (2010). Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics*, 5(8), 685-690.
- Poulin, L. (2005). Étude de la relation structure-fonction de la protéine BI-1 chez *Saccharomyces cerevisiae*
- Puisieux, A., Brabletz, T., & Caramel, J. (2014). Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol*, 16(6), 488-494. doi: 10.1038/ncb2976
- Puisieux, A., Valsesia-Wittmann, S., & Ansieau, S. (2006). A twist for survival and cancer progression. *Br J Cancer*, 94(1), 13-17. doi: 10.1038/sj.bjc.6602876

- Qiang, L., Zhao, B., Ming, M., Wang, N., He, T. C., Hwang, S., . . . He, Y. Y. (2014). Regulation of cell proliferation and migration by p62 through stabilization of Twist1. *Proc Natl Acad Sci U S A*, *111*(25), 9241-9246. doi: 10.1073/pnas.1322913111
- Qiu, H., Fang, X., Luo, Q., & Ouyang, G. (2015). Cancer stem cells: a potential target for cancer therapy. *Cell Mol Life Sci*, *72*(18), 3411-3424. doi: 10.1007/s00018-015-1920-4
- Qu, X., Yu, J., Bhagat, G., Furuya, N., Hibshoosh, H., Troxel, A., . . . Levine, B. (2003). Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. *J Clin Invest*, *112*(12), 1809-1820. doi: 10.1172/JCI20039
- Raatikainen, S., Aaltomaa, S., Palvimo, J. J., Karja, V., & Soini, Y. (2015). TWIST overexpression predicts biochemical recurrence-free survival in prostate cancer patients treated with radical prostatectomy. *Scand J Urol*, *49*(1), 51-57. doi: 10.3109/21681805.2014.909529
- Radisky, D. C., & Bissell, M. J. (2006). Matrix metalloproteinase-induced genomic instability. *Curr Opin Genet Dev*, *16*(1), 45-50. doi: 10.1016/j.gde.2005.12.011
- Raica, M., Cimpian, A. M., & Ribatti, D. (2009). Angiogenesis in pre-malignant conditions. *Eur J Cancer*, *45*(11), 1924-1934. doi: 10.1016/j.ejca.2009.04.007
- Raman, V., Martensen, S. A., Reisman, D., Evron, E., Odenwald, W. F., Jaffee, E., . . . Sukumar, S. (2000). Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature*, *405*(6789), 974-978. doi: 10.1038/35016125
- Ramasamy, T., & Selvam, C. (2015). Performance evaluation of structure based and ligand based virtual screening methods on ten selected anti-cancer targets. *Bioorg Med Chem Lett*, *25*(20), 4632-4636. doi: 10.1016/j.bmcl.2015.08.040
- Raynaud, C. M., Hernandez, J., Llorca, F. P., Nuciforo, P., Mathieu, M. C., Commo, F., . . . Soria, J. C. (2010). DNA damage repair and telomere length in normal breast, preneoplastic lesions, and invasive cancer. *Am J Clin Oncol*, *33*(4), 341-345. doi: 10.1097/COC.0b013e3181b0c4c2
- Ren, K. Q., Cao, X. Z., Liu, Z. H., Guo, H., Quan, M. F., Liu, F., . . . Cao, J. G. (2013). 8-bromo-5-hydroxy-7-methoxychrysin targeting for inhibition of the properties of liver cancer stem cells by modulation of Twist signaling. *Int J Oncol*, *43*(5), 1719-1729. doi: 10.3892/ijo.2013.2071
- Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, *414*(6859), 105-111. doi: 10.1038/35102167
- Ribatti, D. (2009). Endogenous inhibitors of angiogenesis: a historical review. *Leuk Res*, *33*(5), 638-644. doi: 10.1016/j.leukres.2008.11.019
- Risau, W. (1997). Mechanisms of angiogenesis. *Nature*, *386*(6626), 671-674. doi: 10.1038/386671a0
- Risso, G., Blaustein, M., Pozzi, B., Mammi, P., & Srebrow, A. (2015). Akt/PKB: one kinase, many modifications. *Biochem J*, *468*(2), 203-214. doi: 10.1042/BJ20150041
- Robinow, M., & Sorauf, T. J. (1975). Acrocephalopolysyndactyly, type Noack, in a large kindred. *Birth Defects Orig Artic Ser*, *11*(5), 99-106.
- Rose, C. S., Patel, P., Reardon, W., Malcolm, S., & Winter, R. M. (1997). The TWIST gene, although not disrupted in Saethre-Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases. *Hum Mol Genet*, *6*(8), 1369-1373.
- Rushlow, C. A., Hogan, A., Pinchin, S. M., Howe, K. M., Lardelli, M., & Ish-Horowicz, D. (1989). The Drosophila hairy protein acts in both segmentation and bristle patterning and shows homology to N-myc. *EMBO J*, *8*(10), 3095-3103.
- Saethre, M. (1931). Ein Beitrag zum Turmschaedelproblem (Pathogenese, Erblichkeit und Symptomatologie).
- Sahlin, P., Windh, P., Lauritzen, C., Emanuelsson, M., Gronberg, H., & Stenman, G. (2007). Women with Saethre-Chotzen syndrome are at increased risk of breast cancer. *Genes Chromosomes Cancer*, *46*(7), 656-660. doi: 10.1002/gcc.20449
- Salmena, L., Carracedo, A., & Pandolfi, P. P. (2008). Tenets of PTEN tumor suppression. *Cell*, *133*(3), 403-414. doi: 10.1016/j.cell.2008.04.013
- Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., & Furlong, E. E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev*, *21*(4), 436-449. doi: 10.1101/gad.1509007
- Sasaki, K., Natsugoe, S., Ishigami, S., Matsumoto, M., Okumura, H., Setoyama, T., . . . Aikou, T. (2009). Significance of Twist expression and its association with E-cadherin in esophageal squamous cell carcinoma. *J Exp Clin Cancer Res*, *28*, 158. doi: 10.1186/1756-9966-28-158

- Sato, T., Nakashima, A., Guo, L., Coffman, K., & Tamanoi, F. (2010). Single amino-acid changes that confer constitutive activation of mTOR are discovered in human cancer. *Oncogene*, *29*(18), 2746-2752. doi: 10.1038/onc.2010.28
- Satoh, K., Hamada, S., Kimura, K., Kanno, A., Hirota, M., Umino, J., . . . Shimosegawa, T. (2008). Up-regulation of MSX2 enhances the malignant phenotype and is associated with twist 1 expression in human pancreatic cancer cells. *Am J Pathol*, *172*(4), 926-939. doi: 10.2353/ajpath.2008.070346
- Sauve, S., Naud, J. F., & Lavigne, P. (2007). The mechanism of discrimination between cognate and non-specific DNA by dimeric b/HLH/LZ transcription factors. *J Mol Biol*, *365*(4), 1163-1175. doi: 10.1016/j.jmb.2006.10.044
- Schwoebel, E. D., & Moore, M. S. (2000). The control of gene expression by regulated nuclear transport. *Essays Biochem*, *36*, 105-113.
- Scoazec, J. Y. (2000). [Perspectives on anti-angiogenesis treatments and tumor pathology]. *Ann Pathol*, *20 Suppl*, S96-97.
- Seifert, G., Kress, W., Meisel, C., Henze, G., & Seeger, K. (2006). Genetic investigations of Saethre-Chotzen syndrome presenting with renal cell carcinoma. *Cancer Genet Cytogenet*, *171*(1), 76-78. doi: 10.1016/j.cancergencyto.2006.06.012
- Seo, S. K., Kim, J. H., Choi, H. N., Choe, T. B., Hong, S. I., Yi, J. Y., . . . Park, I. C. (2014). Knockdown of TWIST1 enhances arsenic trioxide- and ionizing radiation-induced cell death in lung cancer cells by promoting mitochondrial dysfunction. *Biochem Biophys Res Commun*, *449*(4), 490-495. doi: 10.1016/j.bbrc.2014.05.030
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D., & Lowe, S. W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*, *88*(5), 593-602.
- Setleis, H., Kramer, B., Valcarcel, M., & Einhorn, A. H. (1963). Congenital Ectodermal Dysplasia of the Face. *Pediatrics*, *32*, 540-548.
- Seto, M. L., Hing, A. V., Chang, J., Hu, M., Kapp-Simon, K. A., Patel, P. K., . . . Cunningham, M. L. (2007). Isolated sagittal and coronal craniosynostosis associated with TWIST box mutations. *Am J Med Genet A*, *143A*(7), 678-686. doi: 10.1002/ajmg.a.31630
- Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell*, *138*(5), 822-829. doi: 10.1016/j.cell.2009.08.017
- Sharabi, A. B., Aldrich, M., Sosic, D., Olson, E. N., Friedman, A. D., Lee, S. H., & Chen, S. Y. (2008). Twist-2 controls myeloid lineage development and function. *PLoS Biol*, *6*(12), e316. doi: 10.1371/journal.pbio.0060316
- Shay, J. W., Van Der Haegen, B. A., Ying, Y., & Wright, W. E. (1993). The frequency of immortalization of human fibroblasts and mammary epithelial cells transfected with SV40 large T-antigen. *Exp Cell Res*, *209*(1), 45-52. doi: 10.1006/excr.1993.1283
- Shay, J. W., & Wright, W. E. (2000). Hayflick, his limit, and cellular ageing. *Nat Rev Mol Cell Biol*, *1*(1), 72-76. doi: 10.1038/35036093
- Sherr, C. J., & McCormick, F. (2002). The RB and p53 pathways in cancer. *Cancer Cell*, *2*(2), 103-112.
- Shibata, D., Schaeffer, J., Li, Z. H., Capella, G., & Perucho, M. (1993). Genetic heterogeneity of the c-K-ras locus in colorectal adenomas but not in adenocarcinomas. *J Natl Cancer Inst*, *85*(13), 1058-1063.
- Shiota, M., Izumi, H., Onitsuka, T., Miyamoto, N., Kashiwagi, E., Kidani, A., . . . Kohno, K. (2008). Twist and p53 reciprocally regulate target genes via direct interaction. *Oncogene*, *27*(42), 5543-5553. doi: 10.1038/onc.2008.176
- Sievers, F., & Higgins, D. G. (2014). Clustal omega. *Curr Protoc Bioinformatics*, *48*, 3 13 11-16. doi: 10.1002/0471250953.bi0313s48
- Simpson, P. (1983). Maternal-Zygotic Gene Interactions during Formation of the Dorsal-Ventral Pattern in Drosophila Embryos. *Genetics*, *105*(3), 615-632.
- Singh, S., & Gramolini, A. O. (2009). Characterization of sequences in human TWIST required for nuclear localization. *BMC Cell Biol*, *10*, 47. doi: 10.1186/1471-2121-10-47
- Sinha, S., & Levine, B. (2008). The autophagy effector Beclin 1: a novel BH3-only protein. *Oncogene*, *27 Suppl 1*, S137-148. doi: 10.1038/onc.2009.51

- Smit, M. A., Geiger, T. R., Song, J. Y., Gitelman, I., & Peeper, D. S. (2009). A Twist-Snail axis critical for TrkB-induced epithelial-mesenchymal transition-like transformation, anoikis resistance, and metastasis. *Mol Cell Biol*, *29*(13), 3722-3737. doi: 10.1128/MCB.01164-08
- Society, A. C. The History of Cancer.
- Sosic, D., & Olson, E. N. (2003). A new twist on twist--modulation of the NF-kappa B pathway. *Cell Cycle*, *2*(2), 76-78.
- Sosic, D., Richardson, J. A., Yu, K., Ornitz, D. M., & Olson, E. N. (2003). Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity. *Cell*, *112*(2), 169-180.
- Spaderna, S., Schmalhofer, O., Wahlbuhl, M., Dimmler, A., Bauer, K., Sultan, A., . . . Brabletz, T. (2008). The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. *Cancer Res*, *68*(2), 537-544. doi: 10.1158/0008-5472.CAN-07-5682
- Spicer, D. B., Rhee, J., Cheung, W. L., & Lassar, A. B. (1996). Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science*, *272*(5267), 1476-1480.
- Stankiewicz, P., Thiele, H., Baldermann, C., Kruger, A., Giannakudis, I., Dorr, S., . . . Hansmann, I. (2001). Phenotypic findings due to trisomy 7p15.3-pter including the TWIST locus. *Am J Med Genet*, *103*(1), 56-62.
- Stasinopoulos, I. A., Mironchik, Y., Raman, A., Wildes, F., Winnard, P., Jr., & Raman, V. (2005). HOXA5-twist interaction alters p53 homeostasis in breast cancer cells. *J Biol Chem*, *280*(3), 2294-2299. doi: 10.1074/jbc.M411018200
- Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., . . . Stratton, M. R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, *486*(7403), 400-404. doi: 10.1038/nature11017
- Stoica, G. E., Franke, T. F., Wellstein, A., Morgan, E., Czubayko, F., List, H. J., . . . Stoica, A. (2003). Heregulin-beta1 regulates the estrogen receptor-alpha gene expression and activity via the ErbB2/PI 3-K/Akt pathway. *Oncogene*, *22*(14), 2073-2087. doi: 10.1038/sj.onc.1206311
- Stommel, J. M., & Wahl, G. M. (2005). A new twist in the feedback loop: stress-activated MDM2 destabilization is required for p53 activation. *Cell Cycle*, *4*(3), 411-417.
- Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719-724. doi: 10.1038/nature07943
- Su, Y. W., Xie, T. X., Sano, D., & Myers, J. N. (2011). IL-6 stabilizes Twist and enhances tumor cell motility in head and neck cancer cells through activation of casein kinase 2. *PLoS One*, *6*(4), e19412. doi: 10.1371/journal.pone.0019412
- Sudarsanam, S., & Johnson, D. E. (2010). Functional consequences of mTOR inhibition. *Curr Opin Drug Discov Devel*, *13*(1), 31-40.
- Swift, M. R., & Weinstein, B. M. (2009). Arterial-venous specification during development. *Circ Res*, *104*(5), 576-588. doi: 10.1161/CIRCRESAHA.108.188805
- Talmadge, J. E., & Fidler, I. J. (2010). AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer Res*, *70*(14), 5649-5669. doi: 10.1158/0008-5472.CAN-10-1040
- Tao, L., Zhang, P., Qin, C., Chen, S. Y., Zhang, C., Chen, Z., . . . Chen, Y. Z. (2015). Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. *Adv Drug Deliv Rev*, *86*, 83-100. doi: 10.1016/j.addr.2015.03.014
- Teng, Y., & Li, X. (2014). The roles of HLH transcription factors in epithelial mesenchymal transition and multiple molecular mechanisms. *Clin Exp Metastasis*, *31*(3), 367-377. doi: 10.1007/s10585-013-9621-6
- Thiery, J. P., Acloque, H., Huang, R. Y., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, *139*(5), 871-890. doi: 10.1016/j.cell.2009.11.007
- Thisse, B., el Messal, M., & Perrin-Schmitt, F. (1987). The twist gene: isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic Acids Res*, *15*(8), 3439-3453.
- Tour, U. d.). [Cours de microbiologie: virus et cancer].
- Tran, P. T., Shroff, E. H., Burns, T. F., Thiyagarajan, S., Das, S. T., Zabuawala, T., . . . Felsher, D. W. (2012). Twist1 suppresses senescence programs and thereby accelerates and maintains mutant Kras-induced lung tumorigenesis. *PLoS Genet*, *8*(5), e1002650. doi: 10.1371/journal.pgen.1002650

- Trinidad, J. C., Barkan, D. T., Gullede, B. F., Thalhammer, A., Sali, A., Schoepfer, R., & Burlingame, A. L. (2012). Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics*, *11*(8), 215-229. doi: 10.1074/mcp.O112.018366
- Trost, T. M., Lausch, E. U., Fees, S. A., Schmitt, S., Enklaar, T., Reutzel, D., . . . Spangenberg, C. (2005). Premature senescence is a primary fail-safe mechanism of ERBB2-driven tumorigenesis in breast carcinoma cells. *Cancer Res*, *65*(3), 840-849.
- Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S., & Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell*, *22*(6), 725-736. doi: 10.1016/j.ccr.2012.09.022
- Tukel, T., Sobic, D., Al-Gazali, L. I., Erazo, M., Casasnovas, J., Franco, H. L., . . . Desnick, R. J. (2010). Homozygous nonsense mutations in TWIST2 cause Setleis syndrome. *Am J Hum Genet*, *87*(2), 289-296. doi: 10.1016/j.ajhg.2010.07.009
- Turner, H. E., Harris, A. L., Melmed, S., & Wass, J. A. (2003). Angiogenesis in endocrine tumors. *Endocr Rev*, *24*(5), 600-632. doi: 10.1210/er.2002-0008
- Tursz, T. (Septembre 2013). *Nouvelle Médecine du cancer (La)*.
- Valsesia-Wittmann, S., Magdeleine, M., Dupasquier, S., Garin, E., Jallas, A. C., Combaret, V., . . . Puisieux, A. (2004). Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer Cell*, *6*(6), 625-630. doi: 10.1016/j.ccr.2004.09.033
- Vassilev, L. T. (2005). p53 Activation by small molecules: application in oncology. *J Med Chem*, *48*(14), 4491-4499. doi: 10.1021/jm058174k
- Vermeulen, K., Van Bockstaele, D. R., & Berneman, Z. N. (2005). Apoptosis: mechanisms and relevance in cancer. *Ann Hematol*, *84*(10), 627-639. doi: 10.1007/s00277-005-1065-x
- Vichalkovski, A., Gresko, E., Hess, D., Restuccia, D. F., & Hemmings, B. A. (2010). PKB/AKT phosphorylation of the transcription factor Twist-1 at Ser42 inhibits p53 activity in response to DNA damage. *Oncogene*, *29*(24), 3554-3565. doi: 10.1038/onc.2010.115
- Vinals, F., Reiriz, J., Ambrosio, S., Bartrons, R., Rosa, J. L., & Ventura, F. (2004). BMP-2 decreases Mash1 stability by increasing Id1 expression. *EMBO J*, *23*(17), 3527-3537. doi: 10.1038/sj.emboj.7600360
- Vinals, F., & Ventura, F. (2004). Myogenin protein stability is decreased by BMP-2 through a mechanism implicating Id1. *J Biol Chem*, *279*(44), 45766-45772. doi: 10.1074/jbc.M408059200
- Visvader, J. E., & Lindeman, G. J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer*, *8*(10), 755-768. doi: 10.1038/nrc2499
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, *339*(6127), 1546-1558. doi: 10.1126/science.1235122
- Wallerand, H., Robert, G., Pasticier, G., Ravaud, A., Ballanger, P., Reiter, R. E., & Ferriere, J. M. (2010). The epithelial-mesenchymal transition-inducing factor TWIST is an attractive target in advanced and/or metastatic bladder and prostate cancers. *Urol Oncol*, *28*(5), 473-479. doi: 10.1016/j.urolonc.2008.12.018
- Wang, S. M., Coljee, V. W., Pignolo, R. J., Rotenberg, M. O., Cristofalo, V. J., & Sierra, F. (1997). Cloning of the human twist gene: its expression is retained in adult mesodermally-derived tissues. *Gene*, *187*(1), 83-92.
- Wang, T., Li, Y., Tuerhanjiang, A., Wang, W., Wu, Z., Yuan, M., & Wang, S. (2014). Correlation of Twist upregulation and senescence bypass during the progression and metastasis of cervical cancer. *Front Med*, *8*(1), 106-112. doi: 10.1007/s11684-014-0307-5
- Wang, Y., Xing, J., Xu, Y., Zhou, N., Peng, J., Xiong, Z., . . . Jiang, H. (2015). In silico ADME/T modelling for rational drug design. *Q Rev Biophys*, *48*(4), 488-515. doi: 10.1017/S0033583515000190
- Ward, K. A., & Moss, C. (1994). Evidence for genetic homogeneity of Setleis' syndrome and focal facial dermal dysplasia. *Br J Dermatol*, *130*(5), 645-649.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737-738.

- Wei, S. C., Fattet, L., Tsai, J. H., Guo, Y., Pai, V. H., Majeski, H. E., . . . Yang, J. (2015). Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nat Cell Biol*, *17*(5), 678-688. doi: 10.1038/ncb3157
- Weiner, P. K. e. K., P. A. (1981). AMBER: Assisted Model Building with Energy Refinement. A general program for modeling molecules and their interactions. *J. Comp.Chem.*, *2*(3), 287-303.
- White, E. (2012). Deconvoluting the context-dependent role for autophagy in cancer. *Nat Rev Cancer*, *12*(6), 401-410. doi: 10.1038/nrc3262
- White, E., & DiPaola, R. S. (2009). The double-edged sword of autophagy modulation in cancer. *Clin Cancer Res*, *15*(17), 5308-5316. doi: 10.1158/1078-0432.CCR-07-5023
- Widmann, C., Gibson, S., Jarpe, M. B., & Johnson, G. L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev*, *79*(1), 143-180.
- Williams, G. R., Bethard, J. R., Berkaw, M. N., Nagel, A. K., Luttrell, L. M., & Ball, L. E. (2016). Exploring G protein-coupled receptor signaling networks using SILAC-based phosphoproteomics. *Methods*, *92*, 36-50. doi: 10.1016/j.ymeth.2015.06.022
- Wolf, C., Thisse, C., Stoetzel, C., Thisse, B., Gerlinger, P., & Perrin-Schmitt, F. (1991). The M-twist gene of *Mus* is expressed in subsets of mesodermal cells and is closely related to the *Xenopus X-twi* and the *Drosophila twist* genes. *Dev Biol*, *143*(2), 363-373.
- Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., . . . Vogelstein, B. (2007). The genomic landscapes of human breast and colorectal cancers. *Science*, *318*(5853), 1108-1113. doi: 10.1126/science.1145720
- Wright, W. E., & Shay, J. W. (1992). The two-stage mechanism controlling cellular senescence and immortalization. *Exp Gerontol*, *27*(4), 383-389.
- Wu, H. Y., Tseng, V. S., Chen, L. C., Chang, Y. C., Ping, P., Liao, C. C., . . . Liao, P. C. (2009). Combining alkaline phosphatase treatment and hybrid linear ion trap/Orbitrap high mass accuracy liquid chromatography-mass spectrometry data for the efficient and confident identification of protein phosphorylation. *Anal Chem*, *81*(18), 7778-7787. doi: 10.1021/ac9013435
- Wu, K. J., & Yang, M. H. (2011). Epithelial-mesenchymal transition and cancer stemness: the Twist1-Bmi1 connection. *Biosci Rep*, *31*(6), 449-455. doi: 10.1042/BSR20100114
- Wushou, A., Hou, J., Zhao, Y. J., & Shao, Z. M. (2014). Twist-1 up-regulation in carcinoma correlates to poor survival. *Int J Mol Sci*, *15*(12), 21621-21630. doi: 10.3390/ijms151221621
- Xie, Z., & Klionsky, D. J. (2007). Autophagosome formation: core machinery and adaptations. *Nat Cell Biol*, *9*(10), 1102-1109. doi: 10.1038/ncb1007-1102
- Xu, D.-W., Zhang, G.-Q., Wang, Z.-W., Xu, X.-Y., & Liu, T.-X. (2015). Autophagy in Tumorigenesis and Cancer Treatment. *Asian Pacific Journal of Cancer Prevention*, *16*(6), 2167-2175. doi: 10.7314/apjcp.2015.16.6.2167
- Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., . . . Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*, *117*(7), 927-939. doi: 10.1016/j.cell.2004.06.006
- Yang, J., & Weinberg, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell*, *14*(6), 818-829. doi: 10.1016/j.devcel.2008.05.009
- Yang, M. H., Chen, C. L., Chau, G. Y., Chiou, S. H., Su, C. W., Chou, T. Y., . . . Wu, J. C. (2009). Comprehensive analysis of the independent effect of twist and snail in promoting metastasis of hepatocellular carcinoma. *Hepatology*, *50*(5), 1464-1474. doi: 10.1002/hep.23221
- Yang, M. H., Hsu, D. S., Wang, H. W., Wang, H. J., Lan, H. Y., Yang, W. H., . . . Wu, K. J. (2010). Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat Cell Biol*, *12*(10), 982-992. doi: 10.1038/ncb2099
- Yu, D., & Hung, M. C. (2000). Overexpression of ErbB2 in cancer and ErbB2-targeting strategies. *Oncogene*, *19*(53), 6115-6121. doi: 10.1038/sj.onc.1203972
- Yu, J., & Zhang, L. (2005). The transcriptional targets of p53 in apoptosis control. *Biochem Biophys Res Commun*, *331*(3), 851-858. doi: 10.1016/j.bbrc.2005.03.189
- Yuan, T. L., & Cantley, L. C. (2008). PI3K pathway alterations in cancer: variations on a theme. *Oncogene*, *27*(41), 5497-5510. doi: 10.1038/onc.2008.245

- Yuen, H. F., Chua, C. W., Chan, Y. P., Wong, Y. C., Wang, X., & Chan, K. W. (2007). Significance of TWIST and E-cadherin expression in the metastatic progression of prostatic cancer. *Histopathology*, *50*(5), 648-658. doi: 10.1111/j.1365-2559.2007.02665.x
- Zakian, V. A. (1995). Telomeres: beginning to understand the end. *Science*, *270*(5242), 1601-1607.
- Zee, Y. K., O'Connor, J. P., Parker, G. J., Jackson, A., Clamp, A. R., Taylor, M. B., . . . Jayson, G. C. (2010). Imaging angiogenesis of genitourinary tumors. *Nat Rev Urol*, *7*(2), 69-82. doi: 10.1038/nrurol.2009.262
- Zhang, Z., Xie, D., Li, X., Wong, Y. C., Xin, D., Guan, X. Y., . . . Wang, X. (2007). Significance of TWIST expression and its association with E-cadherin in bladder cancer. *Hum Pathol*, *38*(4), 598-606. doi: 10.1016/j.humpath.2006.10.004
- Zhao, G. Q., Zhao, Q., Zhou, X., Mattei, M. G., & de Crombrughe, B. (1993). TFEC, a basic helix-loop-helix protein, forms heterodimers with TFE3 and inhibits TFE3-dependent transcription activation. *Mol Cell Biol*, *13*(8), 4505-4512.
- Zhou, B. P., Hu, M. C., Miller, S. A., Yu, Z., Xia, W., Lin, S. Y., & Hung, M. C. (2000). HER-2/neu blocks tumor necrosis factor-induced apoptosis via the Akt/NF-kappaB pathway. *J Biol Chem*, *275*(11), 8027-8031.
- Zhou, X., Tan, M., Stone Hawthorne, V., Klos, K. S., Lan, K. H., Yang, Y., . . . Yu, D. (2004). Activation of the Akt/mammalian target of rapamycin/4E-BP1 pathway by ErbB2 overexpression predicts tumor progression in breast cancers. *Clin Cancer Res*, *10*(20), 6779-6788. doi: 10.1158/1078-0432.CCR-04-0112
- Zhu, H., Tannous, P., Johnstone, J. L., Kong, Y., Shelton, J. M., Richardson, J. A., . . . Hill, J. A. (2007). Cardiac autophagy is a maladaptive response to hemodynamic stress. *J Clin Invest*, *117*(7), 1782-1793. doi: 10.1172/JCI27523
- Zhu, J., Wang, H., Bishop, J. M., & Blackburn, E. H. (1999). Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. *Proc Natl Acad Sci U S A*, *96*(7), 3723-3728.
- Zong, W. X., & Thompson, C. B. (2006). Necrotic death as a cell fate. *Genes Dev*, *20*(1), 1-15. doi: 10.1101/gad.1376506