HAL
open science

Caractérisation des domaines N-terminal et de liaison à l'ADN du récepteur des androgènes par des approches biophysiques

Sandra Meyer

▶ **To cite this version:**

Sandra Meyer. Caractérisation des domaines N-terminal et de liaison à l'ADN du récepteur des androgènes par des approches biophysiques. Biophysique. Université de Strasbourg, 2015. Français. NNT : 2015STRAJ091 . tel-01451349

HAL Id: tel-01451349

https://theses.hal.science/tel-01451349

Submitted on 1 Feb 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.

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

UMR7104 – U964 Institut de Génétique et Biologie Moléculaire et Cellulaire

THÈSE présentée par :

Sandra MEYER

soutenue le : **14 Décembre 2015**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences de la Vie et de la Santé / Aspects moléculaires et cellulaires de la biologie

Caractérisation des domaines N-terminal et de liaison à l'ADN du récepteur des androgènes par des approches biophysiques.

THÈSE dirigée par :

M. KIEFFER Bruno	PR, Université de Strasbourg

RAPPORTEURS :

Mme LANDRIEU Isabelle	DR, Université de Lille
M. BOURGUET William	DR, Centre de biochimie structurale de Montpellier

EXAMINATEUR :

M. METZGER Daniel	DR, IGBMC, Strasbourg

MEMBRES INVITES :

M. CERALINE Jocelyn	Co-Directeur de thèse, MCU-PH, Université de Strasbourg
M. DELSUC Marc-André	DR, IGBMC, Strasbourg

A mes grands-parents,

A mes filles,

A Romain,

Remerciements

Cette thèse représente quatre années de science mais également quatre années d'apprentissage sur soi-même ainsi que quatre années parsemées de jolies rencontres.

Tout d'abord, je tiens à remercier Mme Isabelle Landrieu, M. William Bourguet et M. Daniel Metzger, Directeurs de Recherche, pour l'honneur qu'ils me font de juger ce travail.

Je remercie Jocelyn Céraline, mon co-directeur de thèse. Merci, de m'avoir accueilli dans ton laboratoire il y a de çà cinq ans pour réaliser mon stage de Master 2, puis de m'avoir dirigé vers le laboratoire de RMN de l'IGBMC pour débuter ma thèse.

C'est ainsi que j'en viens à remercier mon directeur de thèse, Bruno Kieffer. Merci de m'avoir permis de réaliser ma thèse dans un environnement scientifique de qualité. Le passage du monde de la biologie au monde de la biophysique ne s'est pas fait sans mal ! Je te remercie très chaleureusement pour ton aide devant la console du spectro' à mes débuts, ta disponibilité et ta patience...pour ton exigence, ta rigueur et l'autonomie que tu m'as accordée et qui m'ont permises d'apprendre à tes côtés...pour la confiance que tu m'as accordée lors des enseignements en première année de l'ESBS et lors de l'encadrement des stagiaires au labo notamment le Master 2 de Justine...pour toutes les opportunités d'ouvertures et de participation à des congrès ou écoles de formation qui ont permis d'augmenter mon bagage scientifique...pour ton soutien et ton aide précieuse au moment de la rédaction de ce manuscrit ! Tu m'as transmis ta passion pour la science et fait découvrir le monde fascinant de la RMN, qui j'espère me réserve encore de belles surprises !

Un énorme MERCI à Marc-André Delsuc. A mon arrivée dans l'équipe, tu me parlais de choses étranges comme MeDor, HCA, SAXS... et Python ! ! Je me demandais bien où est-ce que j'avais mis les pieds, moi qui aie la phobie des serpents ! Finalement, j'y ai pris goût petit à petit pour ne jamais m'en lasser ! Pour cela et pour ton aide, je te remercie très sincèrement pour ta grande disponibilité, pour avoir toujours su répondre à mes questions de débutante en Python, entre autres, même à 19h passé ! Merci d'avoir été présent pour tous nos road trips à SOLEIL où nous n'avons jamais travaillé de nuit ! Merci pour tes connaissances encyclopédiques ! Finalement, tu as été un petit peu mon directeur de thèse aussi !

Que de monde à remercier ! Allez, c'est parti ! !

D'abord les collègues...

Un merci à tous les membres actuels et passés de l'équipe de RMN biomoléculaire. Claude, le grand manitou des spectros et du réseau ! Merci pour ta gentillesse, ta disponibilité et ton aide ! Isabelle, merci de m'avoir transmis ton savoir-faire concernant les expériences de titration RMN ! Lionel, je t'ai mis à contribution ces dernières semaines pour python et le débruitage des spectres et je te remercie pour ton aide ! Merci également pour toutes nos discussions scientifiques ou non et nos rires partagés lors des repas du midi, pauses café ou fin de journée quand on était les derniers au labo ! Raphaël, tu te dis timide mais pourtant tu es le plus avenant de l'équipe ! Tu es toujours prêt à aider dans n'importe quelle circonstance. Merci pour ta bonne humeur quotidienne ! Justine V., merci pour les repas à l'extérieur qui nous ont permis de nous aérer l'esprit avec Julia. Et enfin Justine W., ma super stagiaire de M2 ! Notre cobaye et ma référente CcpNmr ! ! Merci pour tous les fous-rires que l'on a eu tout au long de ton stage, merci d'avoir soniqué les protéines et mis mes tubes RMN dans le spectro pendant ma grossesse ! Merci pour ton aide en cette fin de thèse !

Un merci à tous les membres de l'équipe oncoprotéine du 2ème étage et plus particulièrement à :

Yves, pour ton implication à 200% dans l'enseignement et ton aide au dichro', pour les discussions enrichissantes du midi.

Denise, pour avoir partagé ta paillasse et tes protocoles à mon arrivée, pour être toujours disponible et souriante lorsqu'on vient te voir, pour m'avoir fait découvrir les sauterelles et autres spécialités de chez toi ! J'attends le voyage du labo au Mexique !

Anne, pour les discussions science, enfants et bredele que nous avons pu avoir. D'ailleurs, c'est bientôt le moment de se mettre aux fourneaux ! Merci également pour les astuces techniques où la découverte des pièces de l'ESBS comme la chambre chaude du 1er ainsi que les nombreuses fois où tu as soniqué mes protéines pendant ma grossesse.

Un merci également à tous les membres de l'équipe de modélisation avec qui nous partageons également le 2ème étage !

Un merci également à un expert de biologie moléculaire, Gunter. Thank you for having taught me the tricks of all your exotic clonings.

Un merci également à nos supers collaborateurs de Marseille.

Véronique, merci pour ta disponibilité et ton accueil qui m'ont permis d'en savoir plus sur

le SAXS. Hélène, merci pour la semaine intense de purification que nous avons passée ensemble à Strasbourg. Mais surtout un grand merci pour ta relecture des chapitres d'introduction et tes suggestions de corrections pertinentes.

Merci aux personnes des services commun de l'IGBMC : Jean-Marie Garnier de la plateforme de clonage, Pierre et Alastair de la plateforme de cristallisation, Annick du service des missions et Maïté, toujours souriante et disponible pour les impressions.

Puis, il y a les collègues devenues amies...

Marie-Elodie, notre rencontre marque le début d'une aventure remplie de péripéties allant de la morsure du chat à l'escorte de la police pour aller à ton entretien INSERM jusqu'aux soirées mémorables à Langatte ! Merci pour les fous rires partagés, pour ta présence dans les bons moments comme dans les moins bons, pour ta patience, ton aide et tes conseils lors de mes nombreuses répétitions d'oraux qui ont duré jusqu'au bout de la nuit ! Merci à ton cher et tendre, Thom, pour sa contribution à me déstresser avec ses incontournables Bonguouuuuur ! ! ! .

Adeline, notre globe-trotteuse ! Que de parcours ces quatre dernières années ! On s'est trouvé pas mal de points communs avec Marie-Elodie pour ensuite former un trio de choc. Merci pour ton soutien, pour les pauses déjeuners ou les petits verres de fin de journée réconfortants ! Merci d'être la personne que tu es tout simplement ! J'espère avoir l'occasion de venir te voir dans ton nouveau pays un de ces jours !

Julia, notre espagnole-suédoise et ma colocataire de bureau ! Merci pour ton amitié, pour toutes les pauses café, déjeuners et rires partagés, pour ta générosité, pour les souvenirs de nos passages prolongés en chambre froide, devant la console du spectro' pour paramétrer la DOSY, supprimer l'eau...pour ta participation à mettre la sous-couche sur les murs de notre maison (merci à Luis qui peint plus vite que son ombre !)...pour le babysitting de cet été au labo...pour toute l'aide que tu m'as apportée en cette fin de thèse.

Puis la seule, l'unique et véritable... Baie ! !

Aline, ces quelques lignes ne résumeront pas toutes nos années d'amitié ! Tout simplement merci pour la simplicité que tu incarnes et ta passion pour la science même si nous avons étudié des choses complètement différentes. Les centaines de kilomètres qui nous séparent n'auront pas eu raison de nos dizaines de textos quotidiens ni de nos sessions tel' jusqu'au bout de la nuit ! Merci pour ton soutien infaillible, ton assistance CV, Illustrator, mais aussi ton assistance déco de ma salle de mariage, ta planification de journées détente au Spa et bien

d'autres ! Enfin un énorme merci pour le temps et tout le sérieux que tu as mis à relire cette thèse malgré qu'elle soit trop complexe et qu'il te faudrait une bible de bio' mol' pour la comprendre en détail !

Et enfin la famille...

Merci à tous les membres de ma famille pour leur aide dans n'importe quelle circonstance et particulièrement en ce moment où nous construisons notre maison !

Un énorme MERCI à mes grands-parents sans qui rien de tout cela n'aurait été possible ! Merci pour votre soutien sans faille. Les mots ne sont rien pour résumer ce que vous représentez pour moi. Merci pour tout.

Merci à mes filles, Mathilde et Louise, d'être les rayons de soleil de mes journées.

Et le dernier, mais pas le moindre, merci à Romain pour ta présence quotidienne, ton soutien, pour avoir fait tourner la maison et la construction ces derniers mois. Merci d'avoir fait partie de cette aventure scientifique !

Enfin, je terminerai par une pensée pour mes trois poilus , Téquila, Pépito et Cannelle, à mes côtés depuis mes débuts à la fac et dont la présence est source de réconfort.

En essayant continuellement
On nit par réussir.
Donc : plus ça rate,
Plus on a de chances que ça marche.
Proverbe Shadock

Table des matières

1 Objectifs **1**

2 Les récepteurs nucléaires : une machine allostérique complexe **5**

2.1 Organisation des récepteurs nucléaires 6

2.1.1 Le domaine NTD 6

2.1.2 Le domaine DBD 7

2.1.3 La région charnière 8

2.1.4 Le domaine LBD 9

2.2 Les RNs intégrateurs de signaux 10

2.2.1 Régulation allostérique par des effecteurs 11

2.3 Description moléculaire du récepteur des androgènes (AR) 13

2.3.1 Mode d'action de AR 20

2.3.2 Régulation post-traductionnelle de AR 22

2.3.3 AR et pathologies 24

2.3.4 Nouvelles stratégies thérapeutiques ciblant l'AR 25

Bibliographie 27

3 Le motif protéique doigt de zinc **35**

3.1 Introduction 35

3.2 Publication 38

Bibliographie 47

4 Description des protéines intrinsèquement désordonnées (IDPs) : concept et méthodes **49**

4.1 Découverte et définition des IDPs 50

4.2 Propriétés des IDPs 52

4.2.1 Caractéristiques physico-chimiques des IDPs 52

4.2.2 Flexibilité et plasticité structurale des IDPs 53

4.2.3 Transition désordre-ordre 57

4.3 Fonction des IDPs 60

4.4 Méthodes de caractérisation des IDPs 64

4.4.1 Outils bio-informatiques de prédiction du désordre 64

4.4.2 SDS-PAGE 66
4.4.3 Dichroïsme circulaire (CD) 66
4.4.4 Spectroscopie de fluorescence 68
4.4.5 Diffusion des rayons X aux petits angles (SAXS) 69
4.4.6 Résonance Magnétique Nucléaire (RMN) 75
4.5 Conclusion 86
Bibliographie 88

5 Mécanismes moléculaires de reconnaissance de l'ADN par AR-DBD 101
5.1 Introduction 101
5.2 Publication 101

6 Analyse bio-informatique de la séquence protéique humaine de AR-NTD 131
6.1 Prédiction ordre/désordre dans le NTD du récepteur des androgènes 132
6.1.1 metaPrDOS 132
6.1.2 Analyse des clusters hydrophobes (HCA) 133
6.2 Conservation de séquence 135
6.3 Analyse des motifs d'interaction dans AR-NTD 137
6.4 Production et purification 139
6.5 Conclusion 143
Bibliographie 144

Annexe A 147
A.1 Analyse bioinformatique 147
A.2 Production et purification des sous-domaines de AR-NTD 147
Bibliographie 150

7 Caractérisation biophysique de la région C-terminale de AR-NTD 153
7.1 Résonance magnétique nucléaire (RMN) 153
7.1.1 Analyse des spectres protons 153
7.1.2 Mesure de la diffusion translationnelle 155
7.1.3 Marquage isotopique des fragments 157
7.2 Spectroscopie de fluorescence 159
7.3 Dichroïsme circulaire (CD) 161
7.4 Diffusion des rayons X aux petits angles (SAXS) 164
7.4.1 Structures résiduelles 165

7.4.2 Dimensions et forme des protéines 166
7.4.3 Estimation du désordre par la physique des polymères 167
7.5 Conclusion 170
Bibliographie 171

Annexe B **173**
B.1 Expériences RMN 173
B.2 Expériences SAXS 173
B.3 Dichroïsme circulaire 174
B.4 Spectroscopie de fluorescence 175
Bibliographie 176

8 Etude du fragment W1-DBD par RMN **177**
8.1 Production et purification du fragment 178
8.2 Stratégie d'attribution des fréquences de résonances 179
8.3 Publication : Attribution et étude de la structure secondaire résiduelle 182
8.4 Etude de l'interaction du fragment W1-DBD avec un élément de réponse IR3 193
8.4.1 Mise en place des expériences de titrage 193
8.4.2 Résultats 193
8.5 Conclusion 198
Bibliographie 201

Annexe C **203**
C.1 Acquisition, interprétation des spectres RMN du fragment W1-DBD 203
Bibliographie 205

9 Conclusion générale **207**

Liste des Abréviations

G	Variation d'énergie libre
H	Variation d'enthalpie
S	Variation d'entropie
	Déplacement chimique (en ppm)
	Longueur d'onde
D_T	Coefficient de diffusion translationnelle
J	Constante de couplage
R_g	Rayon de giration
R_h	Rayon hydrodynamique
3D	Tri-dimensionnel
ADN	Acide Désoxyribonulcéique
AF-1	Activation Function 1, *Fonction d activation 1*
AF-2	Activation Function 2, *Fonction d activation 2*
ANS	1-anilino-naphtalène-8-sulfonate
ANTS	AR-NTD Signature sequence, *Séquence signature du domaine N-terminal de AR*
ARE	Androgen Response Element, *Elément de réponse aux androgènes*
ARN	Acide Ribonucléique
AR	Androgen Receptor, *Récepteur des androgènes*
ASTEROIDS	A Selection Tool for Ensemble Representation Of Intrinsically Disordered States
BF-3	Binding Function 3, *Fonction de liaison 3*
BMRB	Biological Magnetic Resonance Bank
CARM-1	Coactivator Associated Arginine Methyltransferase 1
CBP	CREB Binding Protein
CD	Circular Dichroïsm, *Dichroïsme circulaire*
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CLEANEX	CLEAN chemical EXchange
CPa	Cancer de la prostate

CREB	C-AMP Response Element Binding Protein
CSP	Chemical Shift Perturbation, *Perturbation des déplacements chimiques*
CTE	C-terminal Extension, *Extension C-terminale*
DBD	DNA Binding Domain, *Domaine de liaison à l ADN*
DHT	Dihydrotestosterone
DOSY	Diffusion Ordered SpectrocopY
DO	Densité Optique
DR	Direct Repeat, *Répétition directe*
DTT	Dithiothréitol
EDTA	Éthylène Diamine Tétra-Acétique
ELM	Eukaryotic Linear Motif, *Motif linéaire eucaryote*
ER	Estrogen Receptor, *Récepteur des oestrogènes*
EXSY	EXchange SpectrocopY
FID	Free Induction Decay
FM	Flexible Meccano
GR	Glucocorticoid Receptor, *Récepteur des glucocorticoïdes*
GST	Glutathione-S-transferase
HAT	Histone Acétyltransférase
HCA	Hydrophobic Cluster Analysis, *Analyse de cluster hydrophobe*
HDAC	Histone Déacétylase
HMQC	Heteronuclear Multiple Quantum Correlation
HMT	Histone Methyltransferase
HNF-4	Hepatocyte Nuclear Factor 4, *Facteur nucléaire hépatocytaire 4*
HSP	Heat Shock Protein, *Protéine de choc thermique*
HSQC	Heteronuclear Single Quantum Correlation
IDP	Intrinsically Disordered Protein, *Protéine intrinsèquement désordonnée*
IPTG	Isopropyl -D-1-Thiogalactopyranoside
IR	Inverted Repeat, *Répétition inversée*
IUPRED	Prediction of Intrinsically Unstructured Proteins
JR	Jump & Return
KID	Kinase Inductible Domain
LBD	Ligand Binding Domain, *Domaine de liaison du ligand*

LBP	Ligand Binding Pocket, *Poche de liaison du ligand*
LB	Luria-Bertani
MAP	Microtubule Associated Protein, *protéine associée aux microtubules*
MBP	Maltose Binding Protein
mDUB	DeUbiquitination module of SAGA
MD	Molecular Dynamic
MnM	Minimotif Miner
MoRF	Molecular Recognition Feature
MR	Mineralocorticoid Receptor, *Récepteur des Minéralocorticoïdes*
NCOR	Nuclear receptor CoRepressor
NES	Nuclear Export Signal, *Signal d export nucléaire*
NF- B	Nuclear Factor B
NLS	Nuclear Localisation Signal, *Signal de localisation nucléaire*
NOESY	Nuclear Overhauser Effect SpectroscopY
NTD	N-terminal Domain, *Domaine N-terminal*
NUS	Non Uniform Sampling
PDB	Protein Data Bank, *Banque de donnée protéique*
PEST	Proline-Glutamate-Serine-Threonine rich sequence
PFG-SE	Pulse Field Gradient- Spin Echo
PFG	Pulse Field Gradient
PMG	*Pre-molten globule*
PONDR	Predictor Of Naturally Disordered Regions
PPAR	Preoxisome Proliferator-Activated Receptor, *Récepteur activé par les prolifé-rateurs de péroxisomes*
ppm	partie par million
PR	Progesterone Receptor, *Récepteur à la progestérone*
PTM	Post-Translational Modification, *Modi cations post-translationnelles*
PULCON	PULse Length CONcentration determination
RAP74	RNA Polymerase II Associated Protein
RAR	Retinoic Acid Receptor, *Récepteur à l acide rétinoïque*
RC	Random Coil
RDC	Residual Dipolar Coupling, *Couplage résiduel dipolaire*

RE	Response Element, *Elément de réponse*
RMN	Résonance Magnétique Nucléaire
RMSD	Root Mean Square Deviation
RN	Récepteur Nucléaire
RXR	Retionid X Receptor, *Récepteur de l acide rétinoïque X*
SAXS	Small Angle X-Ray Scattering, *Di usion des rayons X aux petits angles*
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis, *Gel d électrophorèse de polyacrylamide contenant du SDS*
SLiM	Small Linear Motif, *Petit motif linéaire*
SLP	Sex-limited Protein
SMRT	Silencing Mediator of Retinoid and Thyroid Receptors
SOFAST	band-Selective Optimized-Flip-Angle Short-Transient spectra
SRC	Steroid Receptor Coactivator
SSP	Secondary Structure Propensity, *Propensité de structure secondaire*
TAU1/5	Transactivation Unit 1/5, *Unité de transactivation 1/5*
TEV	Tobacco Etch Virus
TFE	Trifluoroethanol
TFIIF	Transcription Factor II F
TFIIIA	Transcription Factor III A
TMAO	Trimethylamine-N-oxide
TR	Thyroid Receptor, *Récepteur thyroïdien*
UV	Ultra Violet
VDR	Vitamin D Receptor, *Récepteur à la vitamine D*
ZnF	Zinc Finger, *Motif protéique en doigt de zinc*

Chapitre 1

Objectifs

Les récepteurs nucléaires (RNs) sont des facteurs de transcription qui interviennent dans la régulation de l'expression des gènes chez les eucaryotes supérieurs. Ils interviennent dans de nombreux processus biologiques comme la croissance cellulaire, le développement, l'homéostasie, l'embryogenèse, la reproduction, le métabolisme ou la différenciation. Les RNs possèdent une architecture commune organisée en une structure modulaire de quatre domaines fonctionnels. Un domaine N-terminal (NTD) qui contient la fonction d'activation indépendante du ligand (AF-1), suivi du domaine de liaison à l'ADN (DBD) ou domaine C puis du domaine de liaison au ligand (LBD) ou domaine E qui dans certains cas est suivi d'une extrémité C-terminale nommée domaine F. Une région charnière fait le lien entre le DBD et LBD, aussi nommée domaine D. De nombreuses pathologies humaines ont pour origine des dysfonctionnements de RN, pour diverses raisons dont l'apparition de mutations. Ce sont des cibles privilégiées pour la conception de nouvelles molécules thérapeutiques.

Mon projet de thèse porte sur l'étude des mécanismes moléculaires de régulation de l'activité du récepteur nucléaire des androgènes (AR). De nombreuses mutations ont été décrites et ont pour conséquences l'implication de AR dans des pathologies variées telles que le cancer de la prostate ou le syndrome de Kennedy (une maladie neurodégénérative). Ce travail a été motivé par la découverte, par l'équipe du Dr Céraline, de formes tronquées du récepteur chez des patients diagnostiqués avec un cancer de la prostate et traités par chimiothérapie. Les récepteurs tronqués sont dépourvus de leur domaine LBD et ne conservent que les domaines NTD et DBD. Ces récepteurs tronqués sont capables d'activer de façon constitutive l'expression de gènes qui ne sont pas sous le contrôle de l'hormone androgène, dans une situation normale. L'activité transcriptionnelle pathologique de ces récepteurs tronqués met en évidence le rôle joué par le domaine NTD dans la régulation de la transcription par les récepteurs nucléaires.

Le caractère désordonné des domaines NTD rend difficile l'établissement de relations entre leurs compositions en acide-aminés et leurs fonctions. Cependant, depuis plusieurs années, de nombreux travaux mettent en lumière l'importance des régions désordonnées dans les protéines et proposent des méthodes expérimentales et conceptuelles permettant d'établir des relations

entre leur rôle biologique et leurs propriétés dynamiques et structurales. Le domaine NTD de AR comporte 554 acides aminés et représente plus de la moitié de la protéine totale.

A mon arrivée dans le groupe de RMN biomoléculaire, la biologie structurale et les méthodes associées m'étaient totalement inconnues. Mes travaux de recherche se sont attachés à (i) comprendre le mécanisme de reconnaissance de l'ADN par un récepteur nucléaire, (ii) caractériser les propriétés structurales du domaine AR-NTD intrinsèquement désordonné en combinant les méthodes biophysiques comme la RMN, le SAXS, le CD et la spectroscopie de fluorescence.

Etude des mécanismes de reconnaissance de l ADN par AR-DBD

Une première partie de ma thèse est consacrée à la caractérisation du mécanisme moléculaire de reconnaissance de l'ADN par le DBD de AR, en s'appuyant sur une comparaison systématique des propriétés d'un DBD sauvage et portant la mutation T575A. Cette mutation conduit à une altération de la reconnaissance de l'ADN, et des études préliminaires menées au laboratoire dans le cadre d'une thèse (Ying-Hui Wang 2010) et d'un stage de Master 2 (Wissem Medjaoui 2011) ont montré que la structure du DBD mutant, identique à celle du sauvage, n'explique pas les différences observées entre les deux DBD. Pour comprendre ce phénomène de dérégulation nous avons entrepris d'étudier la formation du complexe DBD/ADN en solution par RMN (chapitre 5).

Etude structurale de AR-NTD intrinsèquement désordonné

La deuxième partie de ma thèse consiste en la caractérisation du domaine NTD de AR. Nous avons réalisé une analyse bio-informatique à l'aide des outils ELM (Eukaryotic Linear Motif) et HCA (Hydrophobic Cluster Analysis) permettant de prédire des sites d'interaction ou de repliement (chapitre 6). Cette étude a mis en évidence plusieurs régions du NTD, qui ne présentent pas les caractéristiques attendues pour des IDPs (faible complexité de composition, répartition des résidus hydrophobes/hydrophiles, séquences répétées). En particulier, la partie C-terminale du NTD, localisée à proximité immédiate du DBD (la région TAU5) présente des caractéristiques particulières qui permettent de penser que cette région pourrait adopter un repliement tri-dimensionnel dans certaines conditions. En vue d'étudier de possibles interactions entre ces régions de la protéine et l'activité de reconnaissance à l'ADN, nous avons entrepris le clonage, l'expression de plusieurs fragments comprenant des segments du NTD de AR de plus

en plus longs. Ces fragments ont ensuite été caractérisés par plusieurs méthodes biophysiques telles que la RMN, le SAXS, le CD et la fluorescence (chapitre 7).

La caractérisation par RMN du fragment le plus court, comprenant le domaine DBD, a été entièrement réalisée et a permis l'analyse de structures secondaires à partir des déplacements chimiques expérimentaux (chapitre 8).

Chapitre 2

Les récepteurs nucléaires : une machine allostérique complexe

2.1 Organisation des récepteurs nucléaires **6**

2.1.1 Le domaine NTD 6

2.1.2 Le domaine DBD 7

2.1.3 La région charnière 8

2.1.4 Le domaine LBD 9

2.2 Les RNs intégrateurs de signaux **10**

2.2.1 Régulation allostérique par des effecteurs 11

2.3 Description moléculaire du récepteur des androgènes (AR) **13**

2.3.1 Mode d'action de AR 20

2.3.2 Régulation post-traductionnelle de AR 22

2.3.3 AR et pathologies 24

2.3.4 Nouvelles stratégies thérapeutiques ciblant l'AR 25

Bibliographie **27**

Les récepteurs nucléaires (RNs) représentent une des voies principales de transduction du signal permettant la régulation de la transcription des gènes. Il s'agit de facteurs de transcription répondant à des stimuli et, comme leur nom l'indique, dont l'action se situe dans le noyau. Les RNs forment une superfamille de 48 membres identifiés dans le génome humain [Robinson-Rechavi 2001]. Chacun d'entre eux possède un rôle non redondant dans des processus biologiques variés comme la croissance cellulaire, le développement, l'homéostasie, l'embryogenèse, la reproduction, le métabolisme ou la différenciation [Ribeiro 1995]. Diverses fonctions sont incluses dans l'entité protéique que forme un RN, notamment la liaison d'un ligand, la fixation à l'ADN et le contrôle de l'expression de gènes cibles. Ces ligands sont des petites molécules lipophiles qui agissent comme des signaux permettant de modifier l'activité du récepteur. Alors que les ligands de plusieurs RNs ont été identifiés (hormones stéroïdienne

ou thyroïdienne, vitamine D3, acide rétinoïque, prostaglandines et acides gras), il existe encore un ensemble de RNs définis comme orphelins pour lesquels aucun ligand n'a été défini à ce jour. Ces derniers ne forment pas un groupe distinct des autres RNs, ce qui a mené plusieurs auteurs à postuler que les RNs orphelins aient été les premiers membres de cette superfamille, puis la capacité de fixation d'un ligand serait apparue pour permettre de réguler l'activité du récepteur [Escriva 1997, Laudet 1997, Escriva 2000].

2.1 Organisation des récepteurs nucléaires

Les RNs possèdent une architecture commune organisée en une structure modulaire de quatre domaines fonctionnels (Figure 2.1). Le domaine N-terminal (NTD) ou A/B qui contient la fonction d'activation indépendante du ligand (AF-1), suivi du domaine de liaison à l'ADN (DBD) ou domaine C puis du domaine de liaison au ligand (LBD) ou domaine E qui dans certains cas est suivi d'une extrémité C-terminale nommée domaine F. Une région charnière fait le lien entre le DBD et LBD, aussi nommée domaine D.

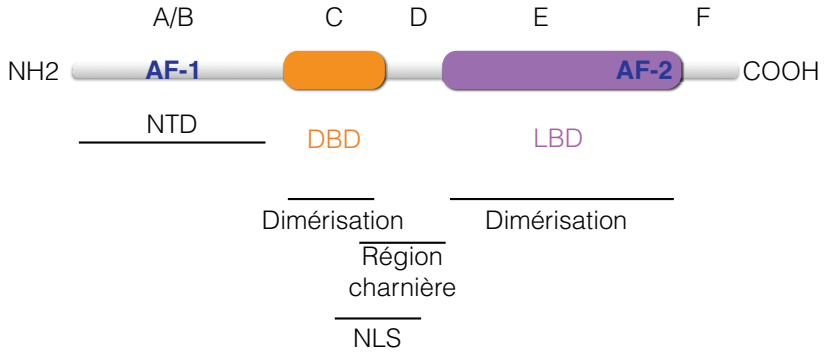


Figure 2.1 – Organisation modulaire des récepteurs nucléaires. Adapté de [Gronemeyer 2004].

2.1.1 Le domaine NTD

Le domaine NTD est la région la plus variable, à la fois en taille et en séquence, parmi les RNs. Cela va de 23 résidus pour le récepteur à la vitamine D (VDR) à 602 résidus pour le récepteurs aux minéralocorticoïdes (MR). Le domaine NTD des RNs présente les caractéristiques d'une protéine intrinsèquement désordonnée (IDP, voir chapitre 4) et aucune donnée cristallographique n'est donc disponible à ce jour [Kumar 2003, McEwan 2007]. Il est alors impossible d'établir un lien structure/fonction. Cependant la présence de désordre s'avère importante pour

la fonction puisque ce domaine est la cible de nombreuses modifications post-traductionnelles (PTMs) telles que la phosphorylation ou sumoylation et d'interactions avec des protéines co-régulatrices des RNs. De plus, la présence de désordre semble être une propriété commune des facteurs de transcription eucaryotes comme par exemple le co-activateur CREB, $NF_{\kappa B}$ et p53 [Wright 2015].

2.1.2 Le domaine DBD

Le domaine de liaison à l'ADN (DBD) est le plus conservé parmi les RNs avec 66 résidus hautement conservés sur un total d'environ 70 résidus. La structure de ce domaine est stabilisée par deux atomes de zinc qui coordonnent les cystéines conservées du DBD formant ainsi deux motifs en doigt de zinc. Ces motifs sont similaires à la structure en doigt de zinc classique identifiée dans le facteur de transcription TFIIIA [Miller 1985].

Eléments structuraux impliqués dans la spéci cité d'interaction avec l'ADN

Les nombreuses structures de DBD, complexés ou non à l'ADN, obtenues par cristallographie ou RMN ont permis de mettre en évidence les déterminants de la spécificité de reconnaissance des RNs vis à vis des séquences d'ADN [Luisi 1991, Schwabe 1993]. Ces différents éléments structuraux nommés boîtes P, D et une extension C-terminale (CTE) [Germain 2006] sont présentés de façon schématique en Figure 2.2.

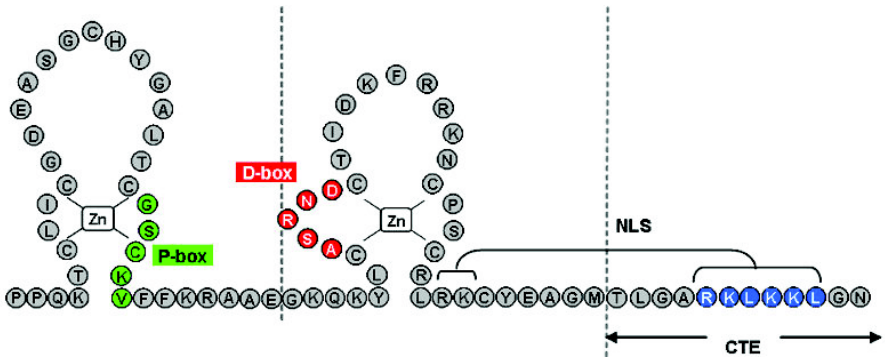


Figure 2.2 – Représentation schématique générale du DBD des récepteurs nucléaires, notamment du récepteur des androgènes ici. La boîte P est colorée en verte, la boîte D en rouge et l'extension C-terminale (CTE) en bleu. Figure extraite de [Claessens 2008]

Le premier doigt de zinc contient la boite P (boîte proximale) qui est constituée d'une

suite de 5 acides aminés interagissant directement avec le grand sillon de l'ADN. Cette boîte P est située au niveau de la première hélice du motif en doigt de zinc et module la spécificité d'interaction avec l'ADN. La boîte D (boîte distale), située au niveau du second motif en doigt de zinc, est impliquée dans l'interface de dimérisation entre deux RNs dont le rôle est également de dicter l'espacement entre les deux demi-sites (détaillé dans le paragraphe ci-dessous). Selon le RN, comme par exemple pour le récepteur thyroïdien (TR) et le récepteur de l'acide rétinoïque X (RXR), la partie variable CTE joue un rôle dans la reconnaissance de l'ADN et la dimérisation des monomères au niveau du petit sillon de l'ADN [Rastinejad 1995].

Types d'éléments de réponses reconnus

Les RNs fonctionnent sous forme d'homo-dimères comme par exemple les récepteurs stéroïdiens, ou d'hétéro-dimères comme le VDR ou les récepteurs à l'acide rétinoïque (RAR) qui s'hétérodimérisent avec RXR. Ces dimères lient des séquences d'ADN, nommées élément de réponse (RE), organisées en deux demi-sites hexamériques espacés d'un nombre variable de nucléotides et arrangés de manière répétée directe (DR, *Direct Repeat*) ou inversée (IR, *Inverted Repeat*). Ces REs dériveraient tous d'une séquence hexamérique canonique 5'-PuGGTCA-3' (Pu= A ou G) et des mutations ou duplications auraient permis de générer ce répertoire de REs variés spécifiques à chaque type de RN [Gronemeyer 1995]. Le mode de fixation à l'ADN combiné à la reconnaissance d'un certain type de RE a permis de classer les RNs en quatre catégories [Mangelsdorf 1995]. La classe I comprend les récepteurs stéroïdiens qui interagissent avec des REs de type IR sous forme d'homo-dimères. Les RNs appartenant à la classe II fonctionnent en hétéro-dimère avec RXR et lient des REs de type DR. Puis les deux autres classes incluent les RNs orphelins qui lient un demi-site de RE sous forme d'homo-dimère (classe III) ou sous forme de monomères (classe IV).

2.1.3 La région charnière

Tout comme le domaine NTD, la région charnière ou domaine D est très peu conservée à la fois en séquence et en taille.

Initialement la communauté scientifique s'accordait pour dire que cette région jouait un rôle de charnière flexible entre les domaines DBD et LBD permettant le positionnement du DBD au niveau des REs au sein du complexe dimère/ADN. En effet, l'interface de dimérisation des RNs est différente selon le type de RE reconnu.

Par la suite, la mise en évidence d'une séquence de localisation nucléaire (NLS) bipartite entre le DBD et cette région charnière a alors permis d'impliquer cette région dans la régula-

tion de la localisation des RNs [Zhou 1994]. Enfin cette région est également impliquée dans l'interaction des RNs avec l'ADN et la régulation de leur activité [Haelens 2007].

2.1.4 Le domaine LBD

Le LBD est le second domaine le plus conservé parmi les RNs. Il s'agit d'un domaine multifonctionnel permettant : (i) la fixation du ligand au niveau de la poche de liaison du ligand (LBP, *Ligand Binding Pocket*), (ii) la transactivation de la transcription dépendante du ligand relayée par la fonction d'activation 2 (AF-2) sur l'hélice H12, (iii) la dimérisation des RNs. (Gronemeyer et laudet, 1995)

Les structures cristallographiques révèlent la présence de 12 hélices (H1 à H12) anti-parallèles disposées en trois couches pour la plupart des RNs [Wurtz 1996, Moras 1998]. Cependant, des variations de cette organisation ont été recensées notamment pour (RAR) et le récepteur des androgènes (AR) ou l'hélice H2 est absente [Renaud 1995, EP 2002].

Fixation du ligand et modèle du piège à souris

La fixation du ligand a lieu dans une poche (LBP) formée par les hélices H3, H7 et H10. Un modèle élégant a été proposé quant au mécanisme moléculaire de la fixation du ligand : le mécanisme du piège à souris (Figure 2.3) [Wurtz 1996, Moras 1998]. En effet, la fixation du ligand entraîne un réarrangement structural majeur menant à un changement conformationnel du LBD où la flexibilité de l'hélice H12 est l'élément crucial. Ce modèle a été proposé à partir de la comparaison des structures cristallographiques du LBD du sous forme *apo* (en absence de ligand) [Bourguet 1995] et *holo* (en présence de ligand) (Figure 2.3) [Egea 2000, Renaud 1995]. Les résultats ont montré que le passage de la forme *apo* à *holo* du LBD s'accompagne d'une compaction du domaine. Les changements conformationnels majeurs mettent en jeu principalement les hélices H3, H11 et H12. En absence de ligand, l'hélice H11 est logée dans la poche de liaison du ligand permettant sa stabilisation. Lors de la fixation du ligand, l'hélice H3 est déplacée de quelques Angström ce qui entraîne un repositionnement de l'hélice H11 dans la continuité de H10. Ce mouvement induit le basculement et le repliement de l'hélice H12, initialement exposée au solvant, sur le ligand.

Ce repliement permet de stabiliser le ligand dans la poche de liaison et également d'exposer une surface d'interaction formée par H3, H4 et H12 comprenant la fonction d'activation 2 (AF-2) impliquée dans les interactions avec des co-régulateurs des RNs contenant le motifs LxxLL (où L correspond à la leucine et x à n'importe quel acides aminé) [Gronemeyer 2004].

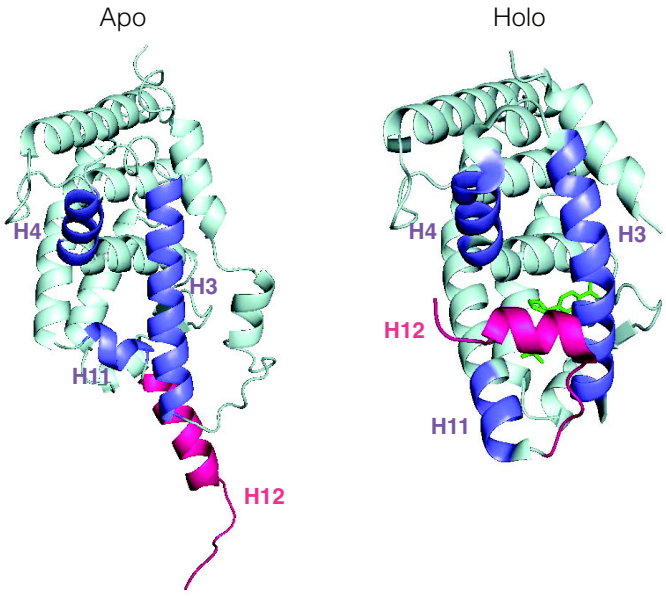


Figure 2.3 – Changement de conformation du LBD de RXRα suite à la fixation du ligand. Les éléments de structure secondaire subissant les réarrangements structuraux majeurs lors de la fixation du ligand (H3, H4 et H11 en violet et H12 en rose) entre la forme *apo* (absence de ligand) (PDB : 1LBD) et la forme *holo* (présence du ligand) sont indiqués (PDB : 1MVC).

Interface de dimérisation

La dimérisation au niveau de ce dernier domaine permet de stabiliser l'homo-/hétéro-dimère sur l'ADN mais n'intervient en aucun cas dans la sélection du type de RE.

La comparaison de structures 3D des LBDs de l'hétéro-dimère RARα/RXRα et de l'homo-dimère formé par le récepteur aux oestrogènes α (ERα) indique que les éléments structuraux impliqués dans l'interface de dimérisation sont identiques [Germain 2013]. Les résidus des hélices H7, H9, H10 et des boucles L8-9 et L9-10 forment l'interface de dimérisation mais H9 et H10 y contribuent à plus de 75% [Bourguet 1995, Bourguet 2000b, Gampe 2000].

2.2 Les RNs intégrateurs de signaux

Jusqu'à il y a peu de temps, les RNs étaient étudiés uniquement par domaine individuel. La vision que l'on a eu pendant longtemps était celle d'une architecture globale organisée en quatre domaines structurellement et fonctionnellement distincts. Cependant, ces domaines appartiennent à une même entité protéique et il convient donc de considérer les RNs dans leur ensemble comme des intégrateurs de signaux.

La cellule, unité fondamentale du vivant, a mis au point des outils permettant de détecter des signaux extérieurs ou intérieurs, réguler les voies métaboliques et la communication intercellulaire. La manière dont tout cela est orchestré débute par (i) la détection du signal, (ii) la transmission de l'information à des effecteurs qui permettent (iii) la propagation et l'intégration du signal aboutissant à (iv) la régulation de la transcription de gènes cibles en réponse au signal.

Les RNs font partie des outils développés par la cellule pour faire face à la multitude de signaux perçus et sont un exemple d'allostérie permettant l'adaptation du comportement cellulaire. L'allostérie peut être vue comme une transition d'un ensemble d'états vers un état permettant la transduction du signal après stimulation par des effecteurs [del Sol 2009]. La transduction du signal dans le contexte des RNs provient de l'action commune de plusieurs effecteurs qui sont le ligand, l'ADN, les modifications post-traductionnelles et les co-régulateurs.

2.2.1 Régulation allostérique par des e ecteurs

Le ligand et les protéines co-régulatrices

L'effet allostérique qui fait suite à la liaison d'un ligand agoniste au niveau de la LBP des RNs est décrit dans le paragraphe précédent. Dans le cas des RNs complexés à l'ADN en absence de ligand, le changement conformationnel suite à la fixation du ligand induit la dissociation des co-represseurs pour permettre l'initiation de la transcription.

A l'inverse, pour les récepteurs situés dans le cytoplasme en absence de ligand comme AR, le changement conformationnel induit par la fixation du ligand déstabilise l'interaction avec les chaperonnes et permet d'exposer un signal de localisation nucléaire nécessaire à la translocation dans le noyau pour activer la transcription de gènes cibles.

De plus, une seconde surface, nommée BF-3, dont les résidus sont extrêmement conservés parmi les récepteurs stéroïdiens a été mise en évidence. Celle-ci est adjacente à AF-2 et des études structurales suggèrent une relation allostérique entre ces deux surfaces où l'occupation de BF-3 par des inhibiteurs affecte le recrutement de co-régulateurs au niveau de AF-2 [Estébanez-Perpiñá 2007, Buzón 2012].

Ainsi, la régulation de l'expression des gènes par les RNs est le résultat d'un échange très bien coordonné avec leurs co-régulateurs. Comme nous le verrons en section 2.3.1, ces co-régulateurs possèdent des activités histone acétylase ou méthylase permettant la décompaction de la chromatine et donc la transcription de gène cible.

L'ADN

La vision actuelle que nous avons du mécanisme de régulation médié par les RNs considère l'ADN comme un effecteur allostérique.

Ce concept est mis en avant par les travaux de Yamamoto et coll., où la séquence d'ADN correspondant à un RE joue un rôle majeur dans la modulation de l'activité transcriptionnelle du récepteur des glucocorticoïdes (GR) [Meijsing 2009], bien que cette étude aie été réalisée par domaine isolé. Le GR reconnait différents REs de type DR3 et dont l'affinité de liaison n'est pas forcément corrélées avec son activité. Les structures de dimères de GR sur des séquences d'ADN différentes révèlent une organisation similaire du dimère de manière tête-à-tête. Mais, un segment de 6 résidus présente une orientation différente selon la séquence d'ADN utilisée. Le même groupe à ensuite montré qu'une mutation dans l'interface de dimérisation abolissait la formation du dimère de GR sur l'ADN. Ces résultats mettent en avant la régulation allostérique de la conformation du DBD de GR au contact de l'ADN qui se propage au niveau de l'interface de dimérisation facilitant la formation du dimère [Watson 2013].

L'analyse des séquences d'ADN liées aux récepteurs RAR à partir d'expériences de Chip-seq (séquençage après immunoprécipitation de la chromatine) qui a révélé un répertoire d'élément de réponse aux récepteurs plus large que prévu fournit également les preuves de concept [Moutier 2012]. Alors que les éléments de réponse de l'hétéro-dimère RAR/RXR de type DR1 DR2 et DR5 étaient décrits comme canoniques, cette étude a montré que l'hétéro-dimère occupait également des REs de type DR0, DR8 et IR0. L'incapacité de l'hétéro-dimère à activer la transcription lors de l'interaction avec un DR0 contrairement à un DR2 ou DR8 suggère que l'espacement entre les demi-sites régule la fonction du récepteur de manière allostérique.

De plus, les premières structures 3D de complexes ternaires entres RNs entiers et leur REs respectifs comme PPAR /RXR, obtenue par cristallographie, et VDR/RXR, obtenue par cryo-microscopie électronique, indique que d'autres régions comme le LBD ou la région charnière [Chandra 2008] ou encore le NTD [Orlov 2012] contribuent à l'interaction avec l'ADN.

Les modi cations post-traductionnelles

Les RNs et notamment leur domaines NTD et LBD sont la cible de nombreuses modifications post-traductionnelles (PTMs) comme la phosphorylaion, acétylation, sumoylation, méthylation qui agissent en tant que modulateur de la transcription. Les structures cristallographiques obtenues pour les RNs entiers montrent que des PTMs à distance de l'interface avec l'ADN peuvent impacter les propriétés d'interaction avec l'ADN.

En effet, la structure du récepteur HNF4- complexé à l'ADN et à un co-activateur, indique

que ces signaux traverses de longues distances à travers la communication interdomaine pour moduler l'interaction du récepteur avec l'ADN [Chandra 2013].

Par ailleurs, des études de simulation de dynamique moléculaire (MD) ont également étudié cette communication interdomaine. Un exemple est la cascade de phosphorylation du récepteur RAR initiée sur une sérine du domaine LBD (S369) et qui se termine sur une sérine du NTD (S77) permettant la transcription de gènes cibles [Bour 2005, Gaillard 2006, Samarut 2011]. Les résultats expérimentaux montrent que la première phosphorylation augmente l'affinité de l'interaction du récepteur avec la cycline H dont le site d'interaction est situé à plus de 40 Å. Chebaro et al. ont mis en évidence le mécanisme allostérique engendré par la phosphorylation de la sérine du LBD par MD [Chebaro 2013]. La phosphorylation entraîne une réorientation des hélices H9 et H10 qui affecte le site de liaison de la cycline H.

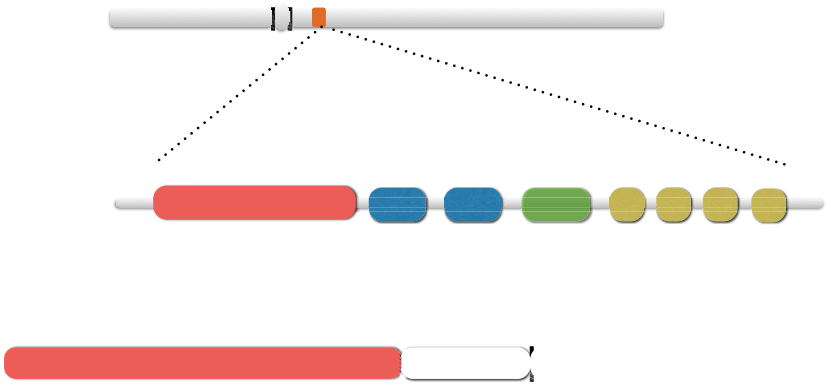
En général, les PTMs peuvent affecter tous les domaines des RNs mais le NTD est le domaine qui compte le plus de sites de PTMs et la phosphorylation est la plus fréquente. Cela s'explique par le caractère désordonné de ce domaine et le rôle de telles protéines au coeur des réseaux de signalisation cellulaire présenté au chapitre 4.

L'organisation des RNs sur l'ADN et l'établissement de chemins allostériques à travers la connection interdomaine conduit à l'intégration et la transduction des signaux provenant de ces différents effecteurs.

Parmi les différents membres de la superfamille des RNs, nous nous sommes intéressés au récepteur des androgènes pour sa particularité structurale et son rôle physiologique dans le développement et le maintien des organes reproducteurs masculins et pathologique.

2.3 Description moléculaire du récepteur des androgènes (AR)

Dans la superfamille des RNs, le récepteur des androgènes (AR) fait partie de la sous famille des récepteurs stéroidiens (NR3C4, *Nuclear Receptor subfamily 3, group C, member 4*) où l'on trouve également le récepteur des glucocorticoïdes (GR), des minéralocorticoïdes (MR), de la progestérone (PR) et des oestrogènes (ER). Le gène de l'AR est localisé en position q11-12 sur le chromosome X chez l'homme [Migeon 1981]. Ce gène possède 8 exons et code pour une protéine de 919 acides aminés avec une masse moléculaire d'environ 110 kDa. L'AR présente l'architecture en 5 domaines commune à la plupart des RNs ainsi qu'une courte région C-terminale adjacente au LBD (Figure 2.4).



résidus respectivement, un polyQ plus petit de 5-6 résidus, et de courtes répétitions d'alanines (polyA) et de prolines (polyP) de 5 et 8 résidus respectivement (Figure 2.5) [Davies 2008].

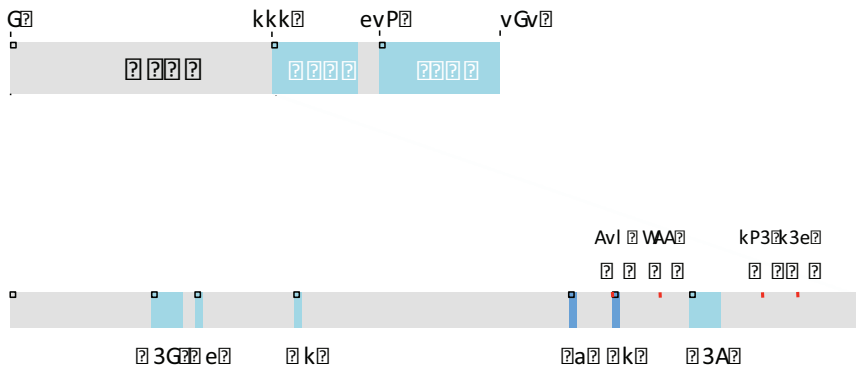


Figure 2.5 – Organisation de AR-NTD présentant les localisations des homopolymères Q21, Q6, Q5, P8, A5, et G23.

Des études biochimiques et biophysiques ont révélé le caractère désordonné de ce domaine et son rôle dans les voies de signalisation de part les nombreuses modifications post-traductionnelles et les interactions avec des co-régulateurs de la transcription [Lavery 2008, McEwan 2012]. Une structuration partielle du NTD en présence d'agents stabilisateurs de structures secondaires comme le TFE (Trifluoroethanol) ou TMAO (TetraMethylAmine-n-Oxyde) a été décrite dans la littérature [Reid 2002] suggérant que ce domaine est susceptible de se replier au cours d'interactions.

Domaine de liaison à l'ADN

Le domaine de liaison à l'ADN de AR est codé par les exons 2 et 3 et reconnaît des éléments de réponse aux androgènes (ARE) palindromique ou inversés répétés séparés de 3 nucléotides (IR3) dont la séquence consensus est 5'- AGAACAnnnTGTTCT- 3'. Ces AREs sont également reconnus par les autres récepteurs stéroïdiens ce qui soulève la question de la spécificité d'expression de gènes cibles vis à vis de chaque récepteur. En effet, le PR et le GR sont souvent présents dans les organes cibles de AR et sont capables de reconnaître les mêmes REs. Une spécificité du DBD de AR est qu'il est l'unique récepteur de cette sous-famille capable de reconnaître des AREs organisés sous forme de répétition directe de la séquence héxamérique 5'- AGAACA- 3' (DR3) [Verrijdt 1999, Claessens 2001]. Cette découverte a permis de classer les AREs en deux groupes, les classiques et les spécifiques permettant en partie d'expliquer la

spécificité d'action des récepteurs stéroïdiens *in vivo*. Ainsi, les gènes androgéno-dépendants contrôlant la masse corporelle et osseuse contiendraient des AREs classiques alors que ceux ayant un rôle dans le développement des organes reproducteurs auraient des AREs spécifiques [Schauwaers 2007].

L'unique structure cristallographique d'un homo-dimère de AR-DBD est celle d'un dimère complexé à un ARE DR3 (PDB :1R4I) (Figure 2.6) [Shaffer 2004].

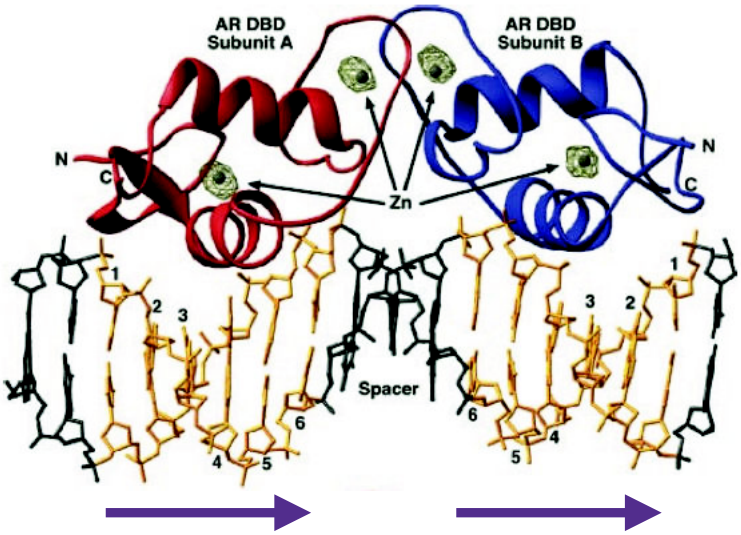


Figure 2.6 – Structure cristallographique d'un dimère de AR-DBD complexé à un élément de réponse de type DR3 (PDB : 1R4I) [Shaffer 2004].

Contrairement à ce qui était attendu par analogie avec la structure de VDR-DBD dont les monomères sont organisés de manière tête-à-queue sur un RE de type DR3 [Shaffer 2002], cette structure révèle une organisation tête-à-tête des monomères. L'organisation du dimère et l'interface de dimérisation impliquant la boîte D du second motif en doigt de zinc de AR-DBD est similaire à celle observée dans le complexe GR-DBD/ADN [Luisi 1991]. Cette organisation résulte de la formation du complexe de manière coopérative, c'est à dire qu'un premier monomère vient se fixer sur un demi-site avec une haute affinité exposant la boîte D, puis le DBD partenaire interagit avec le second demi-site de façon non spécifique [Centenera 2008]. Le dimère est stabilisé sur l'ADN grâce à l'interface de dimérisation dont les résidus impliqués établissent des contacts van der Waals ou des liaisons hydrogènes [Shaffer 2004]. Ces résultats illustrent le rôle de la séquence d'ADN dans l'organisation du dimère de DBD qui peut être

extrapolé dans le cas de la conformation du récepteur entier sur l'ADN.

Enfin, ce domaine contient une partie de la séquence de localisation nucléaire (NLS) bipartite dépendante du ligand.

Région charnière

La région charnière est codée par l'exon 4 et une partie de l'exon 5. Cette région contient la seconde partie de la séquence NLS bipartite dépendante du ligand (^{617}RKCYEAGMTLGARKLKK^{633}) [Saporita 2003].

En plus de son rôle dans la translocation nucléaire du récepteur, cette région joue un rôle dans l'orientation du DBD et du LBD de part sa flexibilité.

La présence du motif basique (^{629}RKLKKLGN^{636}) a un rôle inhibiteur puisque son absence entraîne une augmentation de la transactivation de AR [Haelens 2007].

En outre, la présence d'une séquence PEST riche en proline, glutamate, sérine et thréonine, et située entre les résidus 639-657 peut être impliquée dans la dégradation de AR via le processus d'ubiquitination [Tanner 2004]. En effet, la délétion de cette séquence est associée avec une augmentation de l'activité du récepteur.

Domaine de liaison du ligand

Le LBD de AR est codé par les exons 5 à 8 et se compose de 11 hélices- numérotées de 1 à 12 avec l'hélice H2 manquante. La surface AF-2 formée après liaison du ligand est unique dans le sens où elle présente une affinité plus forte pour les motifs FxxLF du NTD comparé aux motifs LxxLL présents chez les co-activateurs. L'interaction faible avec les motifs LxxL est relié au faible potentiel de transactivation de AF-2 et les co-activateurs sont préférentiellement recrutés au niveau de AF-1 qui est donc la fonction majeure de transactivation. L'interaction N/C permet de stabiliser le ligand et joue un rôle important pour l'expression sélective des gènes sous le contrôle de AR.

La liaison de l'hormone conduit aussi à l'exposition de la surface BF-3 formant une gouttière hydrophobe adjacente à AF-2 et qui régule de manière allostérique l'interaction avec les co-activateurs (Figure 2.7) [Estébanez-Perpiñá 2007].

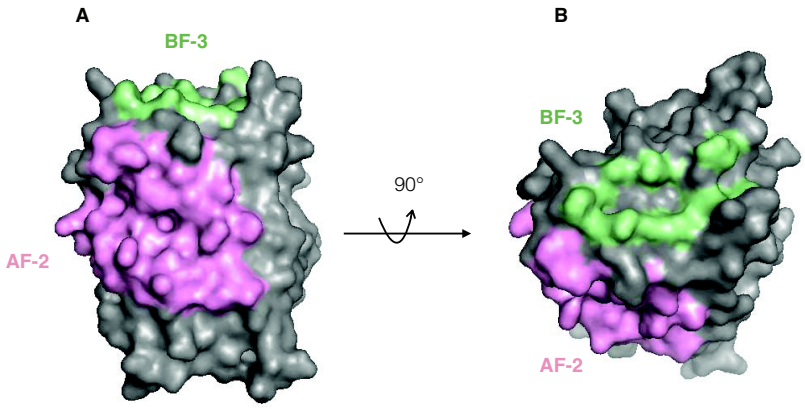


Figure 2.7 – A) Structure de AR-LBD montrant les surfaces d'interaction AF-2 (rose) et BF-3 (vert) (PDB : 2AMA), B) Vue de A) après rotation de 90 ° par rapport à l'axe permettant de visualiser pleinement BF-3. Adapté de [Estébanez-Perpiñá 2007]

Le LBD de AR possède également une séquence d'export nucléaire (NES) localisée entre les résidus 742-817 et inactive en présence du ligand.

De plus, ce domaine est capable de reconnaitre des ligands agonistes et antagonistes menant à l'activation ou la repression de l'activité de AR respectivement. Ces deux types de ligands se lient dans la LBP constituée de 19 résidus hydrophobes. La différence entre le mécanisme d'activation ou de répression est due au repliement de l'hélice H12 dans une orientation différente ce qui empêche la formation d'une surface AF-2 adéquate pour l'interaction avec des co-activateurs (Figure 2.8) [Bourguet 2000a].

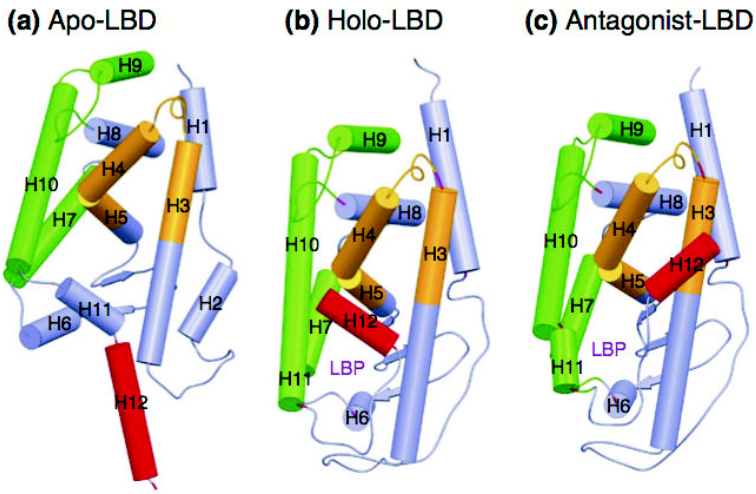


Figure 2.8 – Représentation schématique de trois états conformationnels d'un LBD : (a) forme *apo* en absence de ligand, (b) forme *holo* en présence d'un agoniste et (c) conformation en présence d'un antagoniste. Figure extraite de [Bourguet 2000a].

Agonistes de AR-LBD

La testostérone est une hormone stéroïdienne dont le précurseur dans la voie de biosynthèse est le cholestérol. Dans les cellules cibles, la testostérone est convertie en dihydrotestostérone (DHT) par l'enzyme 5α-réductase (Figure 2.9).

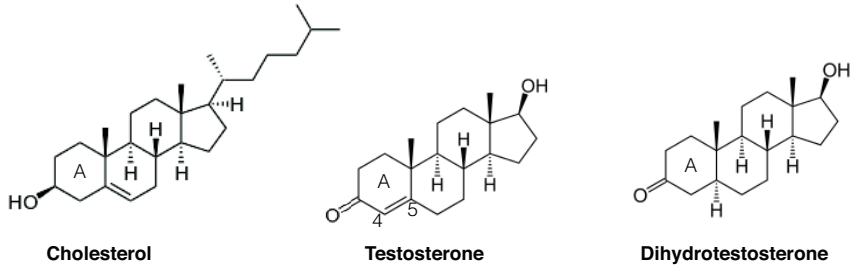


Figure 2.9 – Représentation structurale des agonistes de AR dérivés du cholostérol : testostérone et dihydrotestostérone.

Ces deux molécules diffèrent par la liaison $C_4 - C_5$ sur le cycle A. L'absence de la double liaison dans la DHT augmente son affinité de liaison pour le récepteur et diminue la constante de dissociation (k_{off}) comparé à la testostérone [Tan 2014].

Antagonistes de AR-LBD

Les antagonistes de AR-LBD se nomment anti-androgènes car ils entrent en compétition avec le ligand naturel pour inhiber sa liaison puis l'activité du récepteur.

Les antagonistes de AR utilisés en clinique, pour traiter des patients atteints de cancer de la prostate, sont des agonistes anti-stéroïdiens dérivés de toluidides incluant le nilutamide, flutamide, hydroxyflutamide et bicalutamide (Figure 2.10).

Cependant ces antagonistes sont également capable d'avoir des activités agonistes dont le mécanisme est expliqué dans la section (2.3.3).

Du fait de cet effet agoniste, des antagonistes de seconde génération ont été développés tel que l'enzalutamide connu sous le nom commercial de MDV3100 [Tran 2009].

Toutefois des effets agonistes de cette molécules ont déjà été reportés dans la littérature ce qui démontre la nécessité de développer de nouvelles stratégies thérapeutiques.

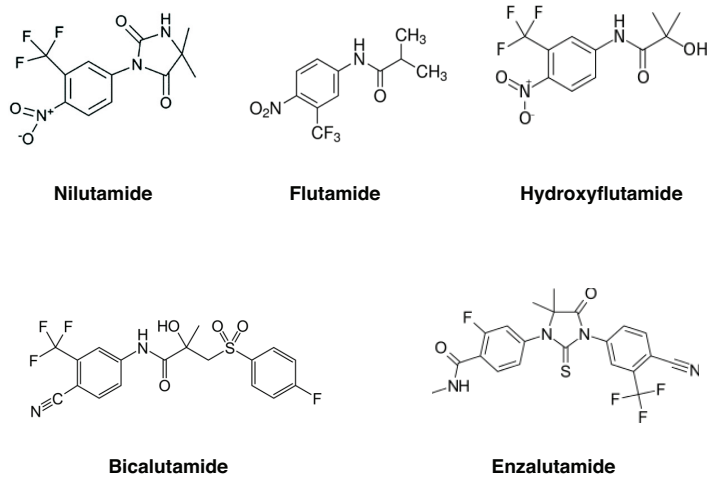


Figure 2.10 – Représentation structurale des antagonistes non stéroïdiens de AR : nilutamide, flutamide, hydroxyflutamide, bicalutamide et enzalutamide un anti-androgène de seconde génération.

2.3.1 Mode d'action de AR

La signalisation des androgènes est médiée par l'AR. Ce récepteur possède deux modes d'action, l'un génomique et l'autre non-génomique (Figure 2.11). Chez l'homme la testostérone est sécrétée par les testicules et circule dans le sang pour aller dans les cellules cibles où elle est convertie en DHT.

Dans le mode génomique et en absence de l'hormone, le récepteur est localisé dans le cytoplasme complexé à des protéines chaperonnes comme les protéines de choc thermique HSP90 et HSP70 qui le maintiennent dans une conformation compétente pour la liaison du ligand.

La fixation de la DHT sur AR entraîne la dissociation des chaperonnes menant à l'exposition de la NLS permettant sa translocation nucléaire. Le changement de conformation induit par le ligand permet l'interaction avec des co-régulateurs et l'exposition de surface pouvant être phosphorylé par des kinases. Dans le noyau, le récepteur interagit avec ses partenaires et reconnaît ses AREs, sur lesquels il s'homodimérise, présents au niveau du promoteur de gènes cibles pour en moduler l'expression (Figure 2.11 **1 et 2**). L'AR peut également interagir avec des facteurs de transcription eux-même complexés à l'ADN pour promouvoir la transcription du gène cible (Figure 2.11 **3**). Les co-régulateurs sont impliqués à la fois dans l'activation ou la répression des gènes probablement dépendant du contexte cellulaire.

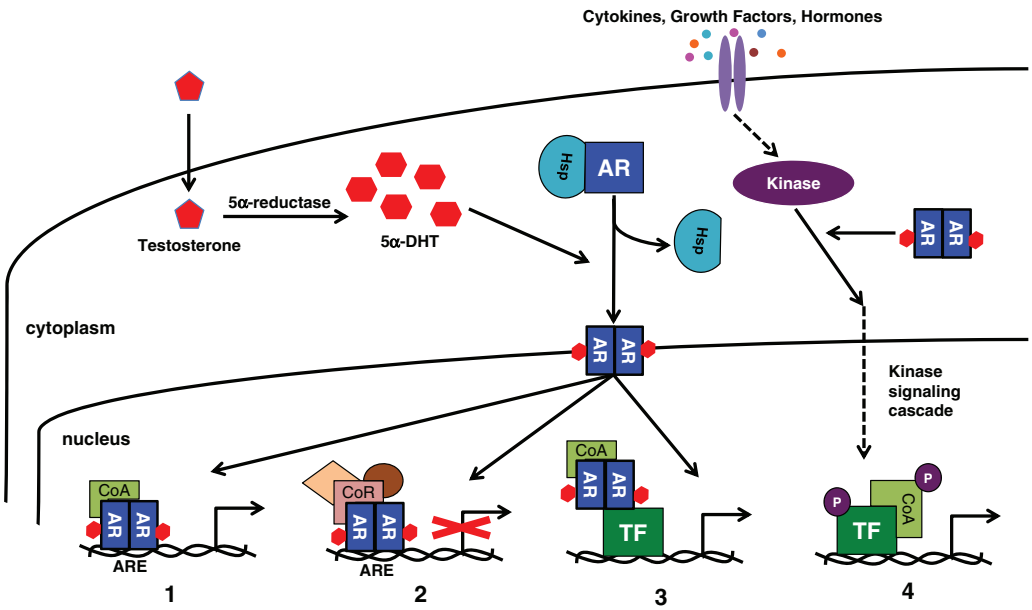


Figure 2.11 – Modes d'action génomique et non-génomique de AR. Figure extraite de [Shafi 2013].

Les co-activateurs

Les co-activateurs sont des protéines qui vont remodeler la chromatine et sont revus en détails dans [Heinlein 2002]. Certaines protéines possèdent des activités histone acétyltransférase (HAT) et vont acétyler les histones telles que les protéines de la familles p160 comprenant les protéines SRC-1, 2 et 3 et la famille CBP/p300. Alors que d'autres, possèdent une activité histone méthyltransférase (HMT) et vont méthyler les histones comme par exemple CARM-1 (Coactivator-associated Arginine Methyl Transferase). Enfin, d'autres protéines possèdent une activité de remodelage de la chromatine ATP-dépendante comme SWI/SNF. Ce type de modification post-traductionnelle des histones permet la décompaction de la chromatine et l'accès de la machinerie transcriptionnelle.

Les co-répresseurs

Les co-répresseurs sont divisés en plusieurs catégories selon leur mode d'action. Certains peuvent inhiber la translocation nucléaire ou la liaison de AR sur ses AREs, et d'autres peuvent compacter la chromatine pour la rendre inaccessible à la machinerie transcriptionnelle. Les plus connus sont NCoR (Nuclear CoRepressor) et SMRT (Silencing Mediator for Retinoic acid anc Thyroid hormon receptor) capables de recruter des protéines à activité histone dé-acétylase (HDAC) permettant la décompaction de la chromatine [Heinlein 2002, Bennett 2010]. De plus,

ces co-répresseurs empêchent l'interaction N/C et peuvent entrer en compétition avec les co-activateurs de la famille p160 [Liao 2003]. La Figure 2.12 présente la liste des co-régulateurs de AR actuellement recensés ainsi que leur domaine préférentiel d'interaction.

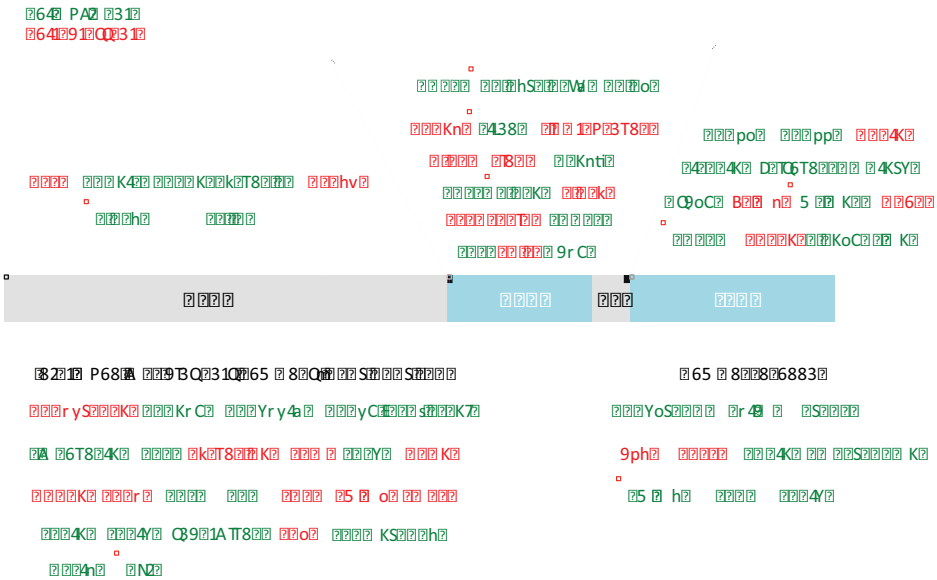


Figure 2.12 – Représentation schématique de AR et de ses différents partenaires d'interaction identifiés. Les co-activateurs sont représentés en vert, les co-répresseurs en rouge, les co-régulateurs exerçant à la fois une activité promotrice ou répressive sont encadrés en rouge. Chaque protéine est localisée au niveau du domaine d'interaction de AR, celles qui interagissent avec plusieurs domaine de AR ou dont le domaine est inconnu sont indiquées en bas. D'après `http://www.androgendb.mcgill.ca`.

Le mode d'action non génomique est quant à lui moins bien étudié (Figure 2.11 **4**), et les androgènes seraient capables d'activer rapidement les voies de signalisation MAPK, PI3K/AKT et PLCγ/PKT/PKA [Kousteni 2001, Heinlein 2004]. Ce mécanisme mène à l'activation de protéine cytoplasmique, de canaux ioniques et de facteurs de transcription. Dans ce mécanisme d'activation, l'AR activé par l'hormone serait présent au niveau de la membrane plasmique.

2.3.2 Régulation post-traductionnelle de AR

L'AR est connecté à des voies de transduction du signal qui régulent son activité par des modifications post-traductionnelles telles que la phosphorylation, la sumoylation, l'ubiquitination et l'acétylation [Gioeli 2012].

La Figure 2.13 présente les nombreuses PTMs reportées dans la littérature.

L'acétylation et la méthylation des résidus lysines de la région charnière serait impliquée dans la translocation nucléaire et le recrutement de co-régulateurs de AR.

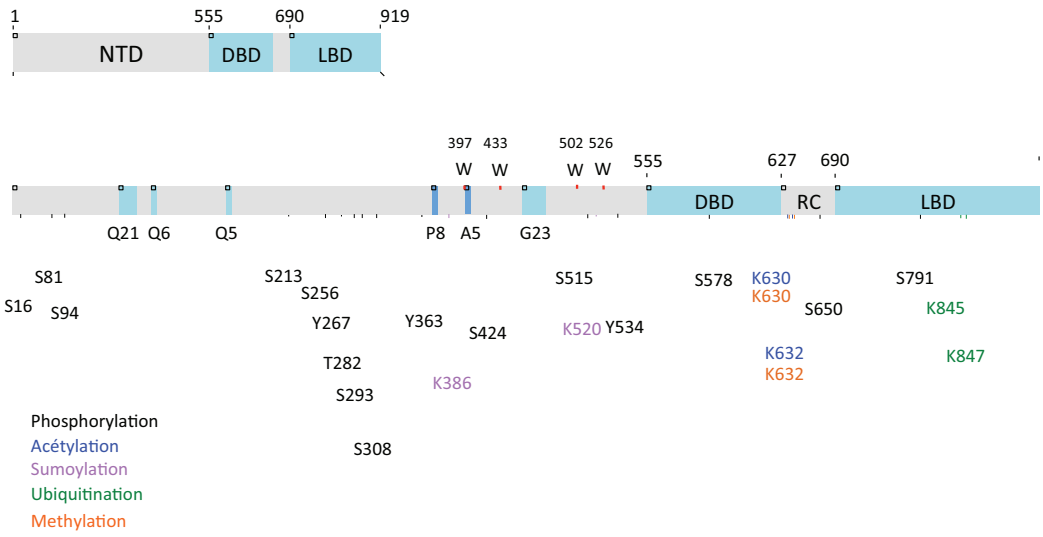


Figure 2.13 – Modifications post-traductionnelles de la protéine humaine AR. La phosphorylation est indiquée en noir, l'acétylation en bleu, la sumoylation en violet, l'ubiquitination en vert et la methylation en orange. Adaptée de [Gioeli 2012]

La sumoylation est une PTM similaire à l'ubiquitination utilisant notamment des voies de signalisation similaires. Cependant, alors que l'ubiquitination est reliée à la stabilité de la protéine, la sumoylation régule plutôt la localisation subcellulaire et l'interaction avec l'ADN, l'activité transcriptionnelle des protéines cibles. La sumoylation des résidus K386 et K520 de AR au niveau des motifs ^{385}IKLE^{388} et ^{519}VKSE^{522} réprime son activité transcriptionnelle [Poukka 2000].

La poly-ubiquitination de AR est associée avec sa dégradation alors que la mono-ubiquitination est associée à une augmentation de la transcription.

Selon le site, la phosphorylation de AR peut se faire en présence ou en absence de ligand et mener à l'augmentation de son activité. Néanmoins, certains sites de phophorylation sont associés avec une diminution de l'activité du récepteur. Les sites de phosphorylation sont largement majoritaires dans le NTD. Les phophorylations sont réalisées par les voies de signalisation PI3K/AKT, Ras/MAPK et ERK.

Ainsi, la capacité de AR à répondre soit à l'hormone soit à l'activation des voies de signalisation met en avant la propriété d'intégration de signaux des RNs.

2.3.3 AR et pathologies

Les androgènes ont des actions multiples comme la différentiation des caractères sexuels mâle, le maintien de la masse osseuse et musculaire. De nombreuses pathologies sont associées à des défauts de fonctionnement de l'AR dû à des mutations faux-sens ou non-sens. Autour de 1000 mutations ont été décrites pour le gène humain de l'AR et sont disponibles sur la base de données Androgen Receptor Gene Mutations Database World Wide Web Server dont le site web est `http://androgendb.mcgill.ca/` [Gottlieb 2012].

Les mutations aboutissant à une perte de fonction sont associées avec le syndrome d'insensibilité aux androgènes qui est un trouble de la différenciation sexuelle aboutissant à un phénotype féminin pour un individu de génotype masculin, des troubles de la spermatogenèse et des maladies dues au vieillissement comme l'ostéoporose.

De plus, une mutation par expansion du triplet de nucléotide CAG codant pour la glutamine supérieure à 38 répétitions dans le NTD entraine l'apparition d'une maladie neurodégénérative nommée atrophie musculaire spino-bulbaire ou syndrome de Kennedy [La Spada 1991, Poletti 2004].

Rôle spécial du récepteur des androgènes dans le cancer de la prostate

Les cellules cancéreuses prostatiques, tout comme les cellules saines de la prostate, sont sous la dépendance des androgènes pour leur survie. C'est historiquement en 1941 par castration chirurgicale que l'hormono-sensibilité du cancer de la prostate (CPa) a été mise en évidence [Huggins 1972]. Charles Huggins a reçu le prix Nobel de médecine et de physiologie en 1966 pour le traitement par thérapie hormonale du CPa. En effet, l'histoire naturelle du cancer de la prostate est marquée par une phase d'hormono-sensibilité, dans laquelle la suppression androgénique s'inscrit comme le traitement efficace. Cependant, malgré une stratégie thérapeutique standardisée, l'entrée en phase d'échappement à la castration androgénique, après une période médiane de 12 à 18 mois, est irrémédiable. Cette phase d'échappement hormonal marque le début d'inefficacité des méthodes de suppression androgénique aboutissant au cancer de la prostate résistant à la castration. Plusieurs facteurs contribuent à la résistance à la castration et l'on peut citer l'amplification du gène de AR, l'émergence de mutations dans le LBD ou encore la perte du domaine LBD. Les mutations observées dans les cas de CPa apportent un gain de fonction au récepteur. Ces mutations touchent tous les domaines du récepteur mais le LBD et le NTD sont les plus ciblés [Bergerat 2009].

Des mutations de substitution d'acides aminés dans le LBD permettent au récepteur de reconnaître les antagonistes utilisés dans les traitements hormonaux comme agonistes pour

activer la transcription de gènes cibles. Par exemple la mutation T877A et F876L permettent une activité agoniste des antagonistes flutamide et enzalutamide [Zhao 1999, Joseph 2013].

Les mutations dans le DBD comme la mutation T575A mènent à une modification de la reconnaissance des AREs pouvant ainsi modifier le spectre des gènes cibles de AR [Monge 2006].

Environ un tiers des mutations de AR ont lieu dans le NTD. De plus, la zone polymorphique codant pour le polyQ est associée à un risque élevé de CPa lorsque le nombre de répétitions est inférieur à 21 [Davies 2008].

La région charnière est la cible de nombreuses mutations non-sens menant à des codons stop prématurés dont les conséquences sont la production d'une protéine dépourvue de LBD (AR- LBD). Ces récepteurs tronqués présentent une activité transcriptionnelle constitutive androgéno-indépendante [Céraline 2004].

Actuellement les mécanismes de la progression tumorale du CPa vers des formes résistantes puis métastatiques restent flous mais il est évident que l'AR est un facteur clé de la croissance et de la survie tumorale. Par ailleurs, le NTD joue un rôle crucial dans l'activité constitutive des récepteurs AR- LBD.

2.3.4 Nouvelles stratégies thérapeutiques ciblant l'AR

Les thérapies actuellement utilisées en clinique ciblent le LBD de l'AR. Arrivé au stade d'échappement au traitement, les armes thérapeutiques à disposition des médecins sont donc considérablement réduites.

Il apparaît que le domaine NTD soit l'instigateur de l'activité persistante de ces récepteurs tronqués. C'est donc tout naturellement que les nouvelles stratégies thérapeutiques tentent de cibler l'activité de ce domaine. La nature désordonnée du NTD est reliée aux interactions avec des partenaires et à son rôle fonctionnel dans la transcription. La présence de désordre rend impossible les études structurales avec les outils conventionnels comme la cristallographie. Ainsi, le manque d'information structurale gène la conception rationnelle de drogues ciblant le NTD.

Néanmoins, des petites molécules interagissant avec ce domaine et inhibant l'activité transcriptionnelle de AR ont été identifiées. La plus prometteuse étant la molécule EPI-001 identifiée lors d'un criblage de composés marins extraits d'éponges [Sadar 2011]. L'interaction de cette molécule avec AR-NTD inhibe le recrutement de CBP et RAP74, une sous-unité du facteur général de la transcription TFIIF. Ce mécanisme suggère que EPI-001 bloque l'interaction du NTD avec des co-activateurs. En outre, il a été montré que cette molécule réduit la prolifération androgéno-dépendante et androgéno-indépendante des cellules cancéreuses qui expriment

l'AR. En revanche, aucun effet n'est observé pour les cellules ne nécessitant pas une activation de AR pour leur croissance [Andersen 2010, McEwan 2012].

Bibliographie

[Andersen 2010] Andersen R.J., Mawji N.R., Wang J., Wang G., Haile S., Myung J.K., Watt K., Tam T., Yang Y.C., Bañuelos C.A., Williams D.E., McEwan I.J., Wang Y. et Sadar M.D. *Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor.* Cancer Cell, vol. 17, no. 6, 2010. – Cité page 26.

[Bennett 2010] Bennett N.C., Gardiner R.A., Hooper J.D., Johnson D.W. et Gobe G.C. *Molecular cell biology of androgen receptor signalling.* The international journal of biochemistry & cell biology, vol. 42, no. 6, 2010. – Cité page 21.

[Bergerat 2009] Bergerat J.P. et C raline J. *Pleiotropic functional properties of androgen receptor mutants in prostate cancer.* Human Mutation, vol. 30, no. 2, 2009. – Cité pages 24 et 131.

[Bour 2005] Bour G., Gaillard E., Bruck N. et Lalevee S. *Cyclin H binding to the RAR activation function (AF)-2 domain directs phosphorylation of the AF-1 domain by cyclin-dependent kinase 7.* In Proceedings of the National Academy of Sciences, 2005. – Cité page 13.

[Bourguet 1995] Bourguet W., Ruff M., Chambon P., Gronemeyer H. et Moras D. *Crystal structure of the ligand-binding domain of the human nuclear receptor RXR- .* Nature, vol. 375, no. 6530, 1995. – Cité pages 9 et 10.

[Bourguet 2000a] Bourguet W., Germain P. et Gronemeyer H. *Nuclear receptor ligand-binding domains : three-dimensional structures, molecular interactions and pharmacological implications.* Trends in pharmacological sciences, vol. 21, no. 10, 2000. – Cité page 18.

[Bourguet 2000b] Bourguet W., Vivat V., Wurtz J.M., Chambon P., Gronemeyer H. et Moras D. *Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains.* Molecular Cell, vol. 5, no. 2, 2000. – Cité page 10.

[Buzón 2012] Buzón V., Carbó L.R., Estruch S.B., Fletterick R.J. et Estébanez-Perpiñá E. *A conserved surface on the ligand binding domain of nuclear receptors for allosteric control.* Molecular and Cellular Endocrinology, vol. 348, no. 2, 2012. – Cité page 11.

[Centenera 2008] Centenera M.M., Harris J.M., Tilley W.D. et Butler L.M. *Minireview : The Contribution of Di erent Androgen Receptor Domains to Receptor Dimerization and Signaling.* Molecular Endocrinology, vol. 22, no. 11, 2008. – Cité page 16.

[Céraline 2004] Céraline J., Cruchant M.D., Erdmann E., Erbs P., Kurtz J.E., Duclos B., Jacqmin D., Chopin D. et Bergerat J.P. *Constitutive activation of the androgen receptor*

by a point mutation in the hinge region : a new mechanism for androgen-independent growth in prostate cancer. International Journal of Cancer, vol. 108, no. 1, 2004. – Cité page 25.

[Chandra 2008] Chandra V., Huang P., Hamuro Y., Raghuram S., Wang Y., Burris T.P. et Rastinejad F. *Structure of the intact PPAR- RXR- nuclear receptor complex on DNA*. Nature, 2008. – Cité page 12.

[Chandra 2013] Chandra V., Huang P., Potluri N., Wu D., Kim Y. et Rastinejad F. *Multi-domain integration in the structure of the HNF-4 nuclear receptor complex*. Nature, vol. 495, no. 7441, 2013. – Cité page 13.

[Chebaro 2013] Chebaro Y., Amal I., Rochel N., Rochette-Egly C., Stote R.H. et Dejaegere A. *Phosphorylation of the Retinoic Acid Receptor Alpha Induces a Mechanical Allosteric Regulation and Changes in Internal Dynamics*. PLoS Computational Biology, vol. 9, no. 4, 2013. – Cité page 13.

[Claessens 2001] Claessens F., Verrijdt G., Schoenmakers E., Haelens A., Peeters B., Verhoeven G. et Rombauts W. *Selective DNA binding by the androgen receptor as a mechanism for hormone-speci c gene regulation.* The Journal of steroid biochemistry and molecular biology, vol. 76, no. 1-5, 2001. – Cité page 15.

[Claessens 2008] Claessens F., Denayer S., Van Tilborgh N., Kerkhofs S., Helsen C. et Haelens A. *Diverse roles of androgen receptor (AR) domains in AR-mediated signaling.* Nuclear Receptor Signaling, vol. 6, page e008, 2008. – Cité page 7.

[Davies 2008] Davies P., Watt K., Kelly S.M., Clark C., Price N.C. et McEwan I.J. *Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain*. Journal of Molecular Endocrinology, vol. 41, no. 5, 2008. – Cité pages 15, 25 et 137.

[del Sol 2009] del Sol A., Tsai C.J., Ma B. et Nussinov R. *The Origin of Allosteric Functional Modulation : Multiple Pre-existing Pathways*. Structure, vol. 17, no. 8, 2009. – Cité page 11.

[Egea 2000] Egea P.F., Mitschler A., Rochel N., Ruff M., Chambon P. et Moras D. *Crystal structure of the human RXR alpha ligand-binding domain bound to its natural ligand : 9-cis retinoic acid*. The EMBO journal, vol. 19, no. 11, pages 2592–2601, 2000. – Cité page 9.

[EP 2002] EP G. *Molecular biology of the androgen receptor*. Journal of Clinical Oncology, vol. 20, no. 13, pages 3001–3015, 2002. – Cité page 9.

[Escriva 1997] Escriva H., Safi R., Hänni C., Langlois M.C., Saumitou-Laprade P., Stehelin D., Capron A., Pierce R. et Laudet V. *Ligand binding was acquired during evolution of nuclear receptors.* Proceedings of the National Academy of Sciences, vol. 94, no. 13, 1997. – Cité page 6.

[Escriva 2000] Escriva H., Delaunay F. et Laudet V. *Ligand binding and nuclear receptor evolution*. BioEssays, vol. 22, no. 8, 2000. – Cité page 6.

[Estébanez-Perpiñá 2007] Estébanez-Perpiñá E., Arnold L.A., Arnold A.A., Nguyen P., Rodrigues E.D., Mar E., Bateman R., Pallai P., Shokat K.M., Baxter J.D., Guy R.K., Webb P. et Fletterick R.J. *A surface on the androgen receptor that allosterically regulates coactivator binding*. Proceedings of the National Academy of Sciences, vol. 104, no. 41, 2007. – Cité pages 11, 17 et 18.

[Gaillard 2006] Gaillard E., Bruck N., Brelivet Y. et Bour G. *Phosphorylation by PKA potentiates retinoic acid receptor activity by means of increasing interaction with and phosphorylation by cyclin H/cdk7*. In Proceedings of the National Academy of Sciences, 2006. – Cité page 13.

[Gampe 2000] Gampe R.T., Montana V.G., Lambert M.H. et Miller A.B. *Asymmetry in the PPAR /RXR crystal structure reveals the molecular basis of heterodimerization among nuclear receptors*. Molecular Cell, vol. 5, no. 3, pages 545–555, 2000. – Cité page 10.

[Germain 2006] Germain P., Staels B., Dacquet C., Spedding M. et Laudet V. *Overview of Nomenclature of Nuclear Receptors*. Pharmacological Reviews, vol. 58, no. 4, 2006. – Cité page 7.

[Germain 2013] Germain P. et Bourguet W. Dimerization of Nuclear Receptors, volume 117. Elsevier Inc., 1 édition, 2013. – Cité page 10.

[Gioeli 2012] Gioeli D. et Paschal B.M. *Post-translational modi cation of the androgen receptor*. Molecular and Cellular Endocrinology, vol. 352, no. 1-2, 2012. – Cité pages 22, 23 et 170.

[Gottlieb 2012] Gottlieb B., Beitel L.K., Nadarajah A., Paliouras M. et Trifiro M. *The androgen receptor gene mutations database : 2012 update*. Human Mutation, vol. 33, no. 5, 2012. – Cité pages 24 et 138.

[Gronemeyer 1995] Gronemeyer H. et Moras D. *Nuclear receptors. How to nger DNA*. Nature, vol. 375, no. 6528, 1995. – Cité page 8.

[Gronemeyer 2004] Gronemeyer H., Gustafsson J.Å. et Laudet V. *Principles for modulation of the nuclear receptor superfamily*. Nature Reviews Drug Discovery, vol. 3, no. 11, 2004. – Cité pages 6 et 9.

[Haelens 2007] Haelens A., Tanner T., Denayer S., Callewaert L. et Claessens F. *The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor.* Cancer research, vol. 67, no. 9, 2007. – Cité pages 9 et 17.

[He 2004] He B., Bai S., Hnat A.T., Kalman R.I., Minges J.T., Patterson C. et Wilson E.M. *An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP).* Journal of Biological Chemistry, vol. 279, no. 29, 2004. – Cité page 14.

[Heinlein 2002] Heinlein C.A. et Chang C. *Androgen receptor (AR) coregulators : an overview.* Endocrine Reviews, vol. 23, no. 2, 2002. – Cité page 21.

[Heinlein 2004] Heinlein C.A. et Chang C. *Androgen Receptor in Prostate Cancer.* Endocrine Reviews, vol. 25, no. 2, 2004. – Cité page 22.

[Huggins 1972] Huggins C. et Hodges C.V. *Studies on prostatic cancer : I. The e ect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate.* CA : a cancer journal for clinicians, vol. 22, no. 4, 1972. – Cité page 24.

[Jenster 1995] Jenster G., van der Korput H.A.G.M., Trapman J. et Brinkmann A.O. *Identi - cation of Two Transcription Activation Units in the N-terminal Domain of the Human Androgen Receptor.* Journal of Biological Chemistry, vol. 270, no. 13, 1995. – Cité page 14.

[Joseph 2013] Joseph J.D., Lu N., Qian J., Sensintaffar J., Shao G., Brigham D., Moon M., Maneval E.C., Chen I., Darimont B. et Hager J.H. *A Clinically Relevant Androgen Receptor Mutation Confers Resistance to Second-Generation Antiandrogens Enzalutamide and ARN-509.* Cancer Discovery, vol. 3, no. 9, 2013. – Cité page 25.

[Kousteni 2001] Kousteni S., Bellido T., Plotkin L.I., O'Brien C.A., Bodenner D.L., Han L., Han K., DiGregorio G.B., Katzenellenbogen J.A., Katzenellenbogen B.S., Roberson P.K., Weinstein R.S., Jilka R.L. et Manolagas S.C. *Nongenotropic, sex-nonspeci c signaling through the estrogen or androgen receptors : dissociation from transcriptional activity.* Cell, vol. 104, no. 5, 2001. – Cité page 22.

[Kumar 2003] Kumar R. et Thompson E.B. *Transactivation Functions of the N-Terminal Domains of Nuclear Hormone Receptors : Protein Folding and Coactivator Interactions.* Molecular Endocrinology, vol. 17, no. 1, 2003. – Cité page 6.

[La Spada 1991] La Spada A.R., Wilson E.M., Lubahn D.B. et Harding A.E. *Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.* Nature, vol. 352, no. 6330, pages 77–79, 1991. – Cité page 24.

[Laudet 1997] Laudet V. *Evolution of the nuclear receptor superfamily : early diversi cation from an ancestral orphan receptor*. Journal of Molecular Endocrinology, vol. 19, no. 3, 1997. – Cité page 6.

[Lavery 2008] Lavery D.N. et McEwan I.J. *Structural Characterization of the Native NH 2-Terminal Transactivation Domain of the Human Androgen Receptor : A Collapsed Disordered Conformation Underlies Structural Plasticity and Protein-Induced Folding* . Biochemistry, vol. 47, no. 11, 2008. – Cité page 15.

[Liao 2003] Liao G., Chen L.Y., Zhang A., Godavarthy A., Xia F., Ghosh J.C., Li H. et Chen J.D. *Regulation of Androgen Receptor Activity by the Nuclear Receptor Corepressor SMRT*. Journal of Biological Chemistry, vol. 278, no. 7, 2003. – Cité page 22.

[Lonergan PE 2011] Lonergan PE T.D. *Androgen receptor signaling in prostate cancer development and progression*. Journal of Carcinogenesis, vol. 10, 2011. – Cité page 14.

[Luisi 1991] Luisi B.F., Xu W.X., Otwinowski Z., Freedman L.P., Yamamoto K.R. et Sigler P.B. *Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with Dna*. Nature, vol. 352, no. 6335, pages 497–505, 1991. – Cité pages 7 et 16.

[Mangelsdorf 1995] Mangelsdorf D.J., Thummel C., Beato M., Herrlich P., Schütz G., Umesono K., Blumberg B., Kastner P., Mark M., Chambon P. et Evans R.M. *The nuclear receptor superfamily : the second decade.* Cell, vol. 83, no. 6, 1995. – Cité page 8.

[McEwan 2007] McEwan I.J., Lavery D., Fischer K. et Watt K. *Natural disordered sequences in the amino terminal domain of nuclear receptors : lessons from the androgen and glucocorticoid receptors.* Nuclear Receptor Signaling, vol. 5, page e001, 2007. – Cité page 6.

[McEwan 2012] McEwan I.J. *Intrinsic disorder in the androgen receptor : identi cation, characterisation and drugability*. Molecular BioSystems., vol. 8, no. 1, pages 82–90, 2012. – Cité pages 15, 26, 131 et 170.

[Meijsing 2009] Meijsing S.H., Pufall M.A., So A.Y., Bates D.L., Chen L. et Yamamoto K.R. *DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity*. Science, vol. 324, no. 5925, 2009. – Cité pages 12 et 170.

[Migeon 1981] Migeon B.R., Brown T.R., Axelman J. et Migeon C.J. *Studies of the locus for androgen receptor : localization on the human X chromosome and evidence for homology with the Tfm locus in the mouse.* Proceedings of the National Academy of Sciences, vol. 78, no. 10, 1981. – Cité page 13.

[Miller 1985] Miller J., McLachlan A.D. et Klug A. *Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.* The EMBO journal, 1985. – Cité page 7.

[Monge 2006] Monge A., Jagla M., Lapouge G., Sasorith S., Cruchant M., Wurtz J.M., Jacqmin D., Bergerat J.P. et Céraline J. *Unfaithfulness and promiscuity of a mutant androgen receptor in a hormone-refractory prostate cancer.* Cellular and Molecular Life Sciences, vol. 63, no. 4, 2006. – Cité pages 25 et 38.

[Moras 1998] Moras D. et Gronemeyer H. *The nuclear receptor ligand-binding domain : structure and function.* Current opinion in cell biology, vol. 10, no. 3, pages 384–391, 1998. – Cité page 9.

[Moutier 2012] Moutier E., Ye T., Choukrallah M.A., Urban S., Osz J., Chatagnon A., Delacroix L., Langer D., Rochel N., Moras D., Benoit G. et Davidson I. *Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology.* Journal of Biological Chemistry, vol. 287, no. 31, 2012. – Cité page 12.

[Orlov 2012] Orlov I., Rochel N., Moras D. et Klaholz B.P. *Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA.* The EMBO journal, vol. 31, no. 2, 2012. – Cité page 12.

[Poletti 2004] Poletti A. *The polyglutamine tract of androgen receptor : from functions to dysfunctions in motor neurons.* Frontiers in Neuroendocrinology, vol. 25, no. 1, 2004. – Cité page 24.

[Poukka 2000] Poukka H., Karvonen U., Jänne O.A. et Palvimo J.J. *Covalent modi cation of the androgen receptor by small ubiquitin-like modi er 1 (SUMO-1).* Proceedings of the National Academy of Sciences, vol. 97, no. 26, 2000. – Cité page 23.

[Rastinejad 1995] Rastinejad F., Perlmann T., Evans R.M. et Sigler P.B. *Structural determinants of nuclear receptor assembly on DNA direct repeats.* Nature, vol. 375, no. 6528, pages 203–211, 1995. – Cité page 8.

[Reid 2002] Reid J., Kelly S.M., Watt K., Price N.C. et McEwan I.J. *Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. In uence of structure-stabilizing solutes and protein-protein interactions.* Journal of Biological Chemistry, vol. 277, no. 22, 2002. – Cité page 15.

[Renaud 1995] Renaud J.P., Rochel N., Ruff M., Vivat V., Chambon P., Gronemeyer H. et Moras D. *Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid.* Nature, vol. 378, no. 6558, 1995. – Cité page 9.

[Ribeiro 1995] Ribeiro R.C., Kushner P.J. et Baxter J.D. *The nuclear hormone receptor gene superfamily.* Annual review of medicine, vol. 46, no. 1, pages 443–453, 1995. – Cité page 5.

[Robinson-Rechavi 2001] Robinson-Rechavi M., Carpentier A.S., Duffraisse M. et Laudet V. *How many nuclear hormone receptors are there in the human genome ?* Trends in genetics, vol. 17, no. 10, 2001. – Cité page 5.

[Sadar 2011] Sadar M.D. *Small molecule inhibitors targeting the "achilles heel" of androgen receptor activity.* Cancer research, vol. 71, no. 4, 2011. – Cité page 25.

[Samarut 2011] Samarut E., Amal I., Markov G.V., Stote R., Dejaegere A., Laudet V. et Rochette-Egly C. *Evolution of Nuclear Retinoic Acid Receptor Alpha (RAR) Phosphorylation Sites. Serine Gain Provides Fine-Tuned Regulation.* Molecular biology and evolution, vol. 28, no. 7, 2011. – Cité page 13.

[Saporita 2003] Saporita A.J., Zhang Q., Navai N., Dincer Z., Hahn J., Cai X. et Wang Z. *Identi cation and Characterization of a Ligand-regulated Nuclear Export Signal in Androgen Receptor.* Journal of Biological Chemistry, vol. 278, no. 43, 2003. – Cité page 17.

[Schauwaers 2007] Schauwaers K., De Gendt K., Saunders P.T.K., Atanassova N., Haelens A., Callewaert L., Moehren U., Swinnen J.V., Verhoeven G., Verrijdt G. et Claessens F. *Loss of androgen receptor binding to selective androgen response elements causes a reproductive phenotype in a knockin mouse model.* Proceedings of the National Academy of Sciences, vol. 104, no. 12, 2007. – Cité page 16.

[Schwabe 1993] Schwabe J.W., Chapman L., Finch J.T. et Rhodes D. *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA : how receptors discriminate between their response elements.* Cell, vol. 75, no. 3, 1993. – Cité page 7.

[Shaffer 2002] Shaffer P.L. et Gewirth D.T. *Structural basis of VDR DNA interactions on direct repeat response elements.* The EMBO journal, vol. 21, no. 9, 2002. – Cité page 16.

[Shaffer 2004] Shaffer P.L., Jivan A., Dollins D.E., Claessens F. et Gewirth D.T. *Structural basis of androgen receptor binding to selective androgen response elements.* Proceedings of the National Academy of Sciences, vol. 101, no. 14, 2004. – Cité pages 16 et 200.

[Shafi 2013] Shafi A.A., Yen A.E. et Weigel N.L. *Androgen receptors in hormone-dependent and castration-resistant prostate cancer.* Pharmacology and Therapeutics, vol. 140, no. 3, 2013. – Cité page 21.

[Tan 2014] Tan M.E., Li J., Xu H.E., Melcher K. et Yong E.l. *Androgen receptor : structure, role in prostate cancer and drug discovery*. Nature Publishing Group, 2014. – Cité page 19.

[Tanner 2004] Tanner T., Claessens F. et Haelens A. *The Hinge Region of the Androgen Receptor Plays a Role in Proteasome-Mediated Transcriptional Activation*. Annals of the New York Academy of Sciences, vol. 1030, no. 1, 2004. – Cité page 17.

[Tran 2009] Tran C., Ouk S., Clegg N.J., Chen Y., Watson P.A., Arora V., Wongvipat J., Smith-Jones P.M., Yoo D., Kwon A., Wasielewska T., Welsbie D., Chen C.D., Higano C.S., Beer T.M., Hung D.T., Scher H.I., Jung M.E. et Sawyers C.L. *Development of a second-generation antiandrogen for treatment of advanced prostate cancer.* Science, vol. 324, no. 5928, 2009. – Cité page 19.

[Verrijdt 1999] Verrijdt G., Schoenmakers E., Alen P., Haelens A., Peeters B., Rombauts W. et Claessens F. *Androgen speci city of a response unit upstream of the human secretory component gene is mediated by di erential receptor binding to an essential androgen response element.* Molecular Endocrinology, vol. 13, no. 9, 1999. – Cité page 15.

[Watson 2013] Watson L.C., Kuchenbecker K.M., Schiller B.J., Gross J.D., Pufall M.A. et Yamamoto K.R. *The glucocorticoid receptor dimer interface allosterically transmits sequence-speci c DNA signals*. Nature Structural & Molecular Biology, vol. 20, no. 7, 2013. – Cité pages 12 et 170.

[Wright 2015] Wright P.E. et Dyson H.J. *Intrinsically disordered proteins in cellular signalling and regulation*. Nature Publishing Group, vol. 16, no. 1, 2015. – Cité pages 7 et 63.

[Wurtz 1996] Wurtz J.M., Bourguet W., Renaud J.P., Vivat V., Chambon P., Moras D. et Gronemeyer H. *A canonical structure for the ligand-binding domain of nuclear receptors.* Nature structural biology, vol. 3, no. 1, 1996. – Cité page 9.

[Zhao 1999] Zhao X.Y., Boyle B., Krishnan A.V., Navone N.M., Peehl D.M. et Feldman D. *Two mutations identi ed in the androgen receptor of the new human prostate cancer cell line MDA PCa 2a.* The Journal of Urology, vol. 162, no. 6, 1999. – Cité page 25.

[Zhou 1994] Zhou Z.X., Sar M., Simental J.A., Lane M.V. et Wilson E.M. *A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor. Requirement for the DNA-binding domain and modulation by NH2-terminal and carboxyl-terminal sequences.* Journal of Biological Chemistry, vol. 269, no. 18, 1994. – Cité page 9.

Chapitre 3

Le motif protéique « doigt de zinc »

3.1 Introduction 35
3.2 Publication 38
Bibliographie 47

3.1 Introduction

L'atome de zinc est impliqué dans des processus biologiques variés où il joue un rôle catalytique, structural et régulateur. Cet atome est essentiel pour stabiliser le repliement des protéines selon un motif protéique appelé doigt de zinc (ZnF). Ce terme tire son origine de la découverte d'une organisation particulière des résidus coordonnant le zinc dans la séquence primaire du facteur de trancription TFIIIA. Cette protéine, présente dans les oocytes de xénope, interagit avec la région codante de l'ARN 5S pour réguler sa transcription. L'analyse de la séquence primaire a montré la présence de sous-domaines composés de 30 acides aminés au sein de la séquence. Deux paires de résidus cystéines et histidines extrêmement conservés qui coordonnent des atomes de zinc sont nécessaires au repliement. La représentation en deux dimensions de ces résidus autour du zinc s'apparente à un doigt d'où la dénomination doigt de zinc . Ce motif protéique comprenant deux cystéines et deux histidines a donné naissance à la famille C2H2 largement représentée aujourd'hui parmi les protéines comprenant des doigts de zinc.

Actuellement, le terme ZnF désigne une vaste famille de motifs et de repliements retrouvés au sein des protéines.

L'analyse bio-informatique des séquences des génomes ainsi que des différents types de repliement stabilisés par l'atome de zinc révèle que celui-ci a joué un rôle important dans l'évolution des protéomes.

Pourquoi l atome de zinc a-t-il été sélectionné par la nature pour être présent dans de nombreux facteurs de transcription interagissant avec l ADN pour réguler l expression de gènes cibles ?

Les motifs à doigt de zinc stabilisent le repliement des petits domaines de liaison à l'ADN.

Plus que la stabilisation d'un domaine structural, l'atome de zinc joue un rôle dans la reconnaissance des interactions spécifiques et non spécifiques avec l'ADN.

La structure cristallographique du facteur de transcription Zif68 lié à l'ADN a permis de mettre en évidence les bases moléculaires de l'interaction avec l'ADN. L'hélice- , des trois doigts de zinc composant Zif68, interagit avec le grand sillon de l'ADN et la liaison successive des doigts de zinc permet à la protéine de s'enrouler autour de l'ADN. Cette interaction est dite canonique. Ainsi, de nombreuses études portant sur différentes protéines à doigt de zinc ont mis en avant le mécanisme d'interaction similaire à Zif68.

Cependant, nombreuses sont les protéines à doigt de zinc qui présentent un mécanisme d'interaction différent qui est alors qualifié de non canonique. En effet, certaines structures cristallographiques des complexes doigt de zinc C2H2/ADN révèlent la présence de contacts avec les phosphates de l'ADN médiés par certains résidus de la protéine.

Uniquement trois types de résidus établissant ces contacts sont conservés parmi ces structures [Wolfe 2000]. Le contact phosphate le plus représenté est médié par le cycle imidazole de l'histidine, particulièrement par le N , qui est impliqué dans une liaison hydrogène avec un oxygène du squelette sucre/phosphate de l'ADN. La coordination du zinc étant établie par le N de l'histidine, le contact entre le N et le phosphate permet d'amener la protéine à proximité immédiate du site de liaison de la protéine au niveau de l'ADN. L'importance de ce contact est démontrée par l'observation d'une altération de la reconnaissance de l'ADN pour des doigts de zinc mutants où ce contact histidine/phosphate est absent [Wolfe 2000].

De plus, le résidu chargé (arginine ou lysine) présent avant la deuxième cystéine du doigt de zinc est également impliqué dans des contacts avec les phosphates de l'ADN.

Enfin, les résidus chargés présents dans les régions permettant de relier et d'espacer deux doigts de zinc voisins sur l'ADN (nommée linker) ont également un rôle dans le mécanisme d'interaction avec l'ADN.

Parmi les protéines comportant un motif à doigt de zinc, les domaines de liaison à l'ADN (DBD) des récepteurs nucléaires (RNs) représentent une famille particulière. Ils sont constitués de deux sites consécutifs de coordination du zinc comportant chacun quatre cystéines conservées parmi la superfamille des RNs, et interagissent avec des séquences d'ADN nom-

mées élément de réponse. De plus, un résidu histidine présent au niveau du premier site de coordination du zinc est également conservé et se trouve au contact du squelette sucre/phosphate de l'ADN. Cette histidine pourrait jouer un rôle de stabilisation du DBD au niveau de l'ADN similaire à celui joué par l'histidine de doigt de zinc de type C2H2.

Comme il a été mentionné dans le chapitre 2, la plupart des RNs sont présents sous forme de dimère au niveau de l'ADN pour réguler l'expression de leur gènes cibles. On peut alors imaginer que la combinaison de quatre sites de coordination du zinc au niveau de l'ADN puisse permettre d'augmenter la spécificité de reconnaissance de la même manière que plusieurs ZnF sont contenus dans les protéines de la famille C2H2.

Les preuves que le site de coordination du zinc joue un rôle dans la reconnaissance de l'ADN sont fournies par les expériences d'interaction DBD/ADN suivies par RMN. En effet, la cartographie des déplacements chimiques suite à l'interaction de la protéine avec l'ADN met en évidence une variation de déplacement chimique (ppm) importante pour les résidus de coordination du zinc (chapitre 5).

Toutefois, ce motif protéique ne se limite pas aux facteurs de transcription pour l'interaction avec l'ADN mais est également impliqué dans des interactions protéine/protéines. Ces interactions interviennent dans des processus biologiques variés tels que la réplication et réparation de l'ADN, la traduction, la prolifération cellulaire et l'apoptose, le métabolisme et la signalisation cellulaire. Des mutations au niveau de ces motifs protéiques entraînant des interactions protéine/protéine ou protéine/ADN aberrantes sont alors associées avec de nombreuses pathologies.

Par exemple, une mutation dans les ZnF de la protéine WT1 est associée au syndrome de Denys-Drash dont l'une des manifestations cliniques est la formation de tumeurs de Wilms [Little 1992]. Un autre exemple est apporté par la protéine CBP qui contient quatre domaines structurés en doigt de zinc. Cette protéine est impliquée dans de gros complexes protéiques et joue un rôle de plateforme dans la cellule. En effet, cette protéine recrute un certain nombre de protéines impliquées dans diverses voies de signalisation (p53 qui est impliquée dans l'apoptose, Mdm2 impliquée dans le sytème de dégradation par le protéasome, c-Fos impliquée dans la prolifération cellulaire) au niveau du promoteur de certains gènes pour en moduler l'expression [Goodman 2000, Matthews 2002]. Ainsi, on comprend bien qu'une anomalie dans le recrutement de protéines au niveau de certains promoteurs peut mener au dévelopement de pathologies. Une illustration de ceci est l'interaction de CBP avec p53 qui se fait en partie via un doigt de zinc TAZ de CBP, ce qui permet l'expression de gènes sous le contrôle de p53 en réponse à un stress cellulaire[Ferreon 2009]. p53 étant un gène suppresseur de tumeur et CBP une protéine plateforme permettant l'interaction avec de nombreux partenaires pour

réguler la transcription de gènes cibles, il en résulte qu'une anomalie dans l'interactions entre ces deux partenaires mène à une division cellulaire anarchique et à la formation de tumeurs [Grossman 2001]. Enfin un dernier exemple concerne la mutation d'un seul résidu (une thréonine substituée par une alanine en position 575, T575A) dans le DBD du récepteur des androgènes. Cette mutation, retrouvée dans des métastases osseuses de cancer de la prostate de certains patients, altère la reconnaissance de l'ADN par ce récepteur [Monge 2006]. L'impact de cette mutation sur les mécanismes de reconnaissance de l'ADN de ce récepteur comparé au récepteur sauvage est étudiée dans le chapitre 5.

3.2 Publication

Dans le cadre de l'étude du DBD du récepteur des androgènes, nous avons accepté d'écrire un chapitre de livre sur les protéines à doigt de zinc dans la série Encyclopedia of Life Sciences (ELS) commandé par l'éditeur Wiley.

Meyer Sandra and Kieffer Bruno (March 2015) ***Protein Motifs : Zinc Fingers***. *In : eLS. John Wiley & Sons, Ltd : Chichester.*

Bibliographie

[Ferreon 2009] Ferreon J.C., Lee C.W., Arai M., Martinez-Yamout M.A., Dyson H.J. et Wright P.E. *Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2.* Proceedings of the National Academy of Sciences, vol. 106, no. 16, 2009. – Cité page 37.

[Goodman 2000] Goodman R.H. et Smolik S. *CBP/p300 in cell growth, transformation, and development.* Genes & development, vol. 14, no. 13, 2000. – Cité page 37.

[Grossman 2001] Grossman S.R. *p300/CBP/p53 interaction and regulation of the p53 response.* European journal of biochemistry, vol. 268, no. 10, 2001. – Cité page 38.

[Little 1992] Little M.H., Prosser J., Condie A., Smith P.J., Van Heyningen V. et Hastie N.D. *Zinc nger point mutations within the WT1 gene in Wilms tumor patients*. Proceedings of the National Academy of Sciences, vol. 89, no. 11, 1992. – Cité page 37.

[Matthews 2002] Matthews J.M. et Sunde M. *Zinc ngers folds for many occasions*. IUBMB Life, 2002. – Cité page 37.

[Monge 2006] Monge A., Jagla M., Lapouge G., Sasorith S., Cruchant M., Wurtz J.M., Jacqmin D., Bergerat J.P. et Céraline J. *Unfaithfulness and promiscuity of a mutant androgen receptor in a hormone-refractory prostate cancer*. Cellular and Molecular Life Sciences, vol. 63, no. 4, 2006. – Cité pages 25 et 38.

[Wolfe 2000] Wolfe S.A., Nekludova L. et Pabo C.O. *DNA recognition by Cys2His2 zinc nger proteins.* Annual Review of Biophysics and Biomolecular Structure, vol. 29, no. 1, pages 183–212, 2000. – Cité page 36.

Chapitre 4

Description des protéines intrinsèquement désordonnées (IDPs) : concept et méthodes

4.1 Découverte et définition des IDPs **50**

4.2 Propriétés des IDPs **52**

4.2.1 Caractéristiques physico-chimiques des IDPs 52

4.2.2 Flexibilité et plasticité structurale des IDPs 53

4.2.3 Transition désordre-ordre 57

4.3 Fonction des IDPs **60**

4.4 Méthodes de caractérisation des IDPs **64**

4.4.1 Outils bio-informatiques de prédiction du désordre 64

4.4.2 SDS-PAGE 66

4.4.3 Dichroïsme circulaire (CD) 66

4.4.4 Spectroscopie de uorescence 68

4.4.5 Diffusion des rayons X aux petits angles (SAXS) 69

4.4.6 Résonance Magnétique Nucléaire (RMN) 75

4.5 Conclusion **86**

Bibliographie **88**

Les protéines intrinsèquement désordonnées (IDPs) sont des protéines dont la fonction requiert un certain degré de désordre conformationnel. Ces protéines ne peuvent pas être représentées par un jeu de cordonnées unique, ce qui rend difficile leur étude par des approches classiques telle que la cristallographie. Le désordre conformationnel confère à ces protéines des propriétés particulières, qui sont indispensables à leur fonction. Par exemple, l'état désordonné multiplie les possibilités d'interaction avec un partenaire moléculaire en exposant l'ensemble des résidus au solvant, alors qu'une protéine globulaire ne peut utiliser à priori que ses résidus

de surface pour interagir avec un partenaire. De ce fait, les IDPs jouent un rôle majeur dans de nombreux processus biologiques. En particulier, elles sont au coeur des réseaux de signalisation et de régulation de processus essentiels tels que la réplication ou transcription des gènes. Il en résulte que ces protéines sont impliquées dans de nombreuses pathologies et constituent de fait des cibles potentielles pour le développement de nouvelles stratégies thérapeutiques.

Ce chapitre présente l'état des connaissances sur les IDPs depuis leur découverte au cours du XXème siècle ainsi que les méthodes permettant de les caractériser.

4.1 Découverte et dé nition des IDPs

L'idée selon laquelle une structure tridimensionnelle rigide détermine la fonction d'une protéine à été introduite en 1894 par Emil Fisher [Fischer 1894]. Le concept communément appelé clé/serrure constituait alors le modèle de référence pour décrire la relation structure/fonction d'une protéine. Les décennies suivantes ont permis aux scientifiques de découvrir que les protéines étaient constituées d'un enchaînement d'acides aminés appelé séquence primaire. Environ un siècle plus tard, des études ont montrés que cette séquence primaire contient l'intégralité de l'information nécessaire au repliement d'une protéine dans sa structure tridimensionnelle et que celle-ci est directement liée à sa fonction biologique [Anfinsen 1972].

De nombreuses études portant sur le repliement des protéines tentent alors d'établir ces liens entre la séquence primaire, la structure tridimensionnelle et la fonction des protéines. La première analyse à ce sujet fut publiée par Hsien Wu en 1931. Son travail montre que la dénaturation des protéines conduit à une perte de leur fonction. Les premières études de repliement des protéines étaient réalisées à l'équilibre thermodynamique mettant ainsi en évidence deux états macroscopiques, l'état natif et l'état dénaturé. Du fait des nombreuses controverses liées à cette théorie, de nombreuses études cinétiques de réactions de dénaturation/renaturation ont été menées à partir des années 70 afin de mettre en évidence des intermédiaires de repliement [Kuwajima 1989]. Le premier état intermédiaire qui a pu être identifié est le *molten globule* (MG) ou globule fondu. Cet état est relativement compact avec des structures secondaires identiques à la structure native d'une protéine globulaire. Cependant, une partie des interactions intramoléculaires sont perdues et la structure tertiaire est, de ce fait, plus flexible. L'existence d'un second état intermédiaire ne possédant pas de structures secondaires stables mais préservant plusieurs contacts intramoléculaires a été décrit par la suite. Cet état de repliement se situe entre la forme complètement dénaturé ou random coil (RC), où la chaîne polypeptidique est dépliée et peut échantillonner toutes les conformations possibles, et la forme MG. La dénomination de *pre-molten globule* (PMG) a été adoptée pour désigner cet état intermédiaire

de repliement [Chaffotte 1997].

C'est à partir des années 80 que des résultats expérimentaux démontrent l'importance fonctionnelle de la flexibilité moléculaire dans le cas des protéines [Schweers 1994]. Ces résultats remettent en question le paradigme de relation structure-fonction qui représentait alors un des dogmes centraux de la biologie moléculaire, toutes exceptions à cette règle étant interprétées comme des artéfacts. Il faudra attendre quelques années et l'accumulation de travaux à ce sujet pour que la communauté scientifique accepte l'idée que des protéines ne présentant pas de structure tridimensionnelle définie puissent être fonctionnelles. Ces protéines sont alors regroupées dans une même classe appelée protéines intrinsèquement désordonnées (IDPs) [Wright 1999, Dunker 2001].

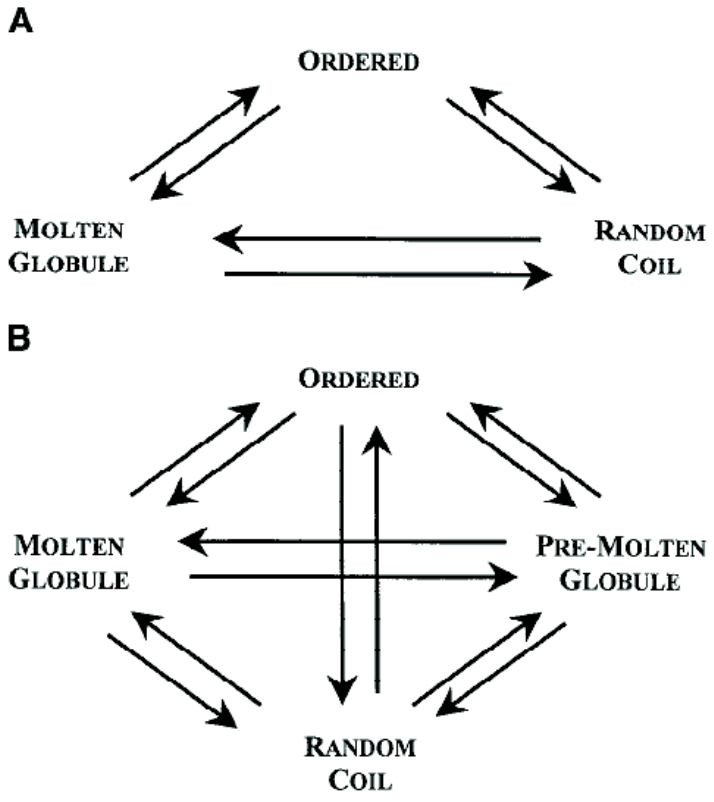


Figure 4.1 – Extension du modèle de la trinité des protéines (A) au modèle du quatuor des protéines (B). Ce modèle propose que l'état natif et fonctionnel d'une protéine peut résulter de quatre conformations possibles (Random coil, pré-molten golbule, molten globule et ordonné) ainsi que des transitions entre ces états. Figure extraite de [Uversky 2002a].

La définition de cette nouvelle classe de protéines, dont la chaîne polypeptidique peut adopter de nombreuses conformations de façon dynamique et interchangeable, nécessite de revoir

le paradigme structure/fonction. Avec les études de dénaturation/renaturation des protéines et la découverte d'un premier état intermédiaire de repliement, le MG, Dunker et coll. proposent le concept de la trinité des protéines [Dunker 2001]. Ce concept suggère que les protéines existent sous trois états thermodynamiques distincts (ordonné, PMG et RC), chacun pouvant correspondre à l'état natif et fonctionnel d'une protéine. L'identification de l'état intermédiaire appelé PMG amènera ensuite les scientifiques à étendre ce concept au modèle du quatuor des protéines. Ce modèle intègre alors l'état PMG et la fonction des protéines résulte des quatre conformations possibles ainsi que des transitions entre ces états (Figure 4.1) [Uversky 2002a].

Cependant, il est possible que ce modèle soit encore trop simpliste pour décrire l'ensemble des états accessibles aux IDPs et à l'origine de leurs diverses propriétés et fonctions. Il existe évidemment un continuum de possibilités entre le passage de l'état RC à l'état replié ordonné d'une protéine.

4.2 Propriétés des IDPs

4.2.1 Caractéristiques physico-chimiques des IDPs

La séquence primaire des IDPs code pour le désordre intrinsèque de la même manière qu'elle contient l'information de structure pour les protéines globulaires. La composition en acides aminés des IDPs favorise les interactions avec les molécules d'eau au détriment de la formation d'un coeur hydrophobe. C'est pourquoi les résidus hydrophobes (I, V, L) et aromatiques (W, Y, F) sont proportionnellement moins représentés que les résidus polaires et chargés (E, K, R, Q). L'analyse de la répartition des paramètres physico-chimiques tels que la charge nette, l'hydrophobicité ou le volume, le long de la séquence donne un première indication sur le caractère ordonné ou désordonné de la protéine [Uversky 2002b, Dunker 2008]. Ces observations ont permis de classer les acides aminés en deux groupes, l'un promoteur de désordre (A, G, R, Q, S, P, E et K), et l'autre promoteur d'ordre (W, C, F, I, Y, V, L, et N) [Dyson 2005]. La complexité de séquence est également un moyen de faire la distinction entre une protéine repliée et une IDP. La complexité de séquence se décrit par la diversité des types de résidus composant une séquence protéique. Les IDPs contiennent fréquemment des séquence de moindre complexité, comme des segments de type poly-proline ou poly-glycine. Les propriétés intrinsèques des séquences protéiques associées aux IDPs ont permis le développement de nombreuses méthodes de prédiction du désordre à partir de la séquence primaire. Une première analyse simple consiste à reporter la charge nette moyenne en fonction de l'hydrophobicité moyenne dans un graphe à deux dimensions. Sur ce type de représentation, proposée par Uversky, les IDPs présentent une hydrophobicité moindre combinée à une charge nette élevée par rapport aux protéines repliées.

(Figure 4.2) [Dunker 2008].

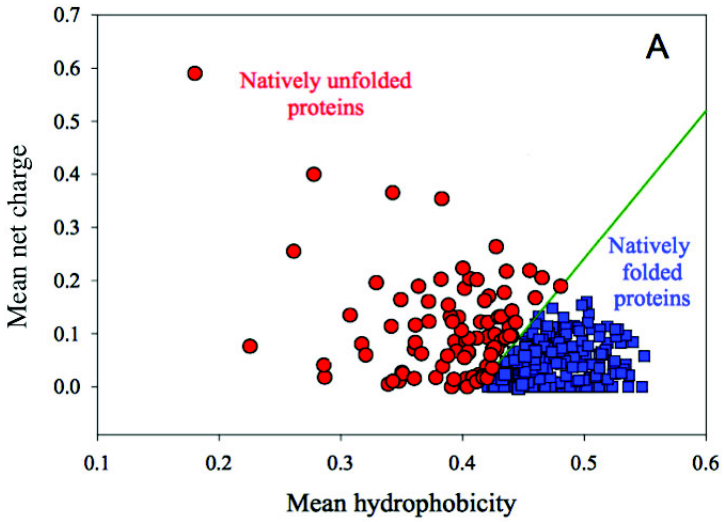


Figure 4.2 – Particularité de la séquence primaire des protéines intrinsèquement désordonnées. Figure extraite de [Dunker 2008].

4.2.2 Flexibilité et plasticité structurale des IDPs

De nombreuses études, théoriques et expérimentales, ont mis en lumière la nature dynamique et flexible de la chaîne peptidique, ainsi que l'importance de ce caractère sur la fonction. Les mesures spectroscopiques (Infra-rouge et résonance magnétique nucléaire) ainsi que les simulations de dynamique moléculaire montrent que les mouvements qui animent la chaîne polypeptidique se produisent sur une très large gamme de temps qui s'étend de la nanosconde à quelques heures (Figure 4.3) [Teilum 2009].

Ces mouvements englobent les vibrations des liaisons chimiques, la rotation des chaînes latérales des acides aminés, ou encore l'adaptation d'une protéine à son ligand.

Si la flexibilité d'une protéine résulte dans la plupart des cas de la dynamique interne et globale de celle-ci, la réciproque n'est pas forcément vraie. Ces deux paramètres sont bel et bien distincts, et, la dynamique reflète les mouvements moléculaires intrinsèques d'une protéine comme par exemple l'échantillonnage de différentes conformations, alors que la flexibilité correspond à la capacité d'une protéine d'adapter sa structure en réponse à la liaison d'un ligand ou des changements d'environnement.

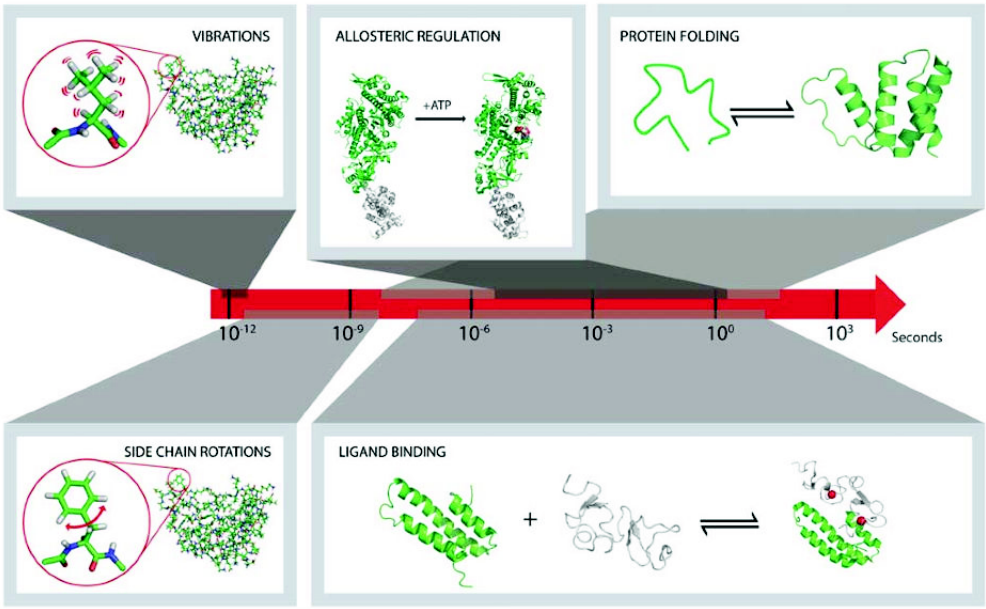


Figure 4.3 – Phénomènes de dynamique et flexibilité à différentes échelles de temps au sein des protéines. Figure extraite de [Teilum 2009]

Pour bien comprendre la différence entre flexibilité et dynamique il est nécessaire d'introduire des notions de thermodynamique.

Les multiples conformations échantillonnées par une protéine lors du processus de repliement peuvent être représentées sous la forme d'un paysage énergétique [Bryngelson 1995]. Ce paysage énergétique adopte une forme d'entonnoir où les valeurs à haute énergie correspondent aux multiples états d'une chaîne polypeptidique dépliée et les valeurs de basse énergie correspondent à l'état natif d'une protéine repliée en une structure 3D unique (Figure 4.4) [Dinner 2000].

Il existe plusieurs chemins de repliement pour une protéine passant par une diminution de l'énergie libre de la protéine et une augmentation du nombre des interactions intramoléculaires. Tous ces chemins convergent vers un point appelé état de transition pour ensuite évoluer spontanément vers le minimum énergétique correspondant à l'état natif. Cela représente une barrière d'énergie et l'énergie fournie au système pour atteindre l'état de transition est appelée énergie d'activation.

De plus, la diminution d'énergie libre du système est également accompagnée d'une diminution d'entropie (Figure 4.4).

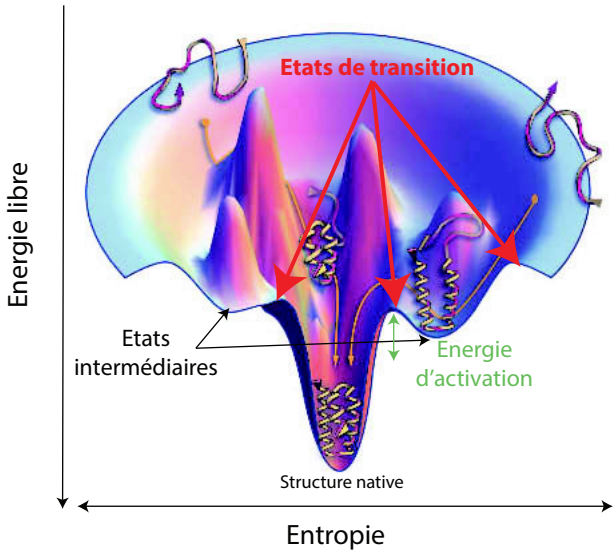


Figure 4.4 – Schéma du processus de repliement des protéines. La théorie du paysage énergétique ou entonnoir de repliement présente les multiples conformations d'une chaîne polypeptidique dépliée, ayant une grande entropie conformationnelle, en haut avec des énergies libres quasi identiques et élevées. Au cours du repliement, la diminution de l'entropie au cours du passage dans l'entonnoir réduit le nombre de conformations de concert avec la diminution de l'énergie libre de la chaîne polypeptidique, permettant d'augmenter le nombre de protéines dans une conformation native. Enfin, l'état natif correspondant à la protéine repliée en une structure bien définie est situé en bas avec un minimum d'énergie unique. Figure extraite de [Dill 2012]

L'énergie libre de Gibbs est reliée à des facteurs structuraux et dynamiques permettant d'évaluer la partie d'énergie potentielle (enthalpie, H) du système non dissipée sous forme de chaleur (entropie, S) :

$$\mathrm{G}^0 = \mathrm{H}^0 - \mathrm{T}\ \mathrm{S}^0 \tag{4.1}$$

H^0 correspond à l'enthalpie standard et peut être rapporté aux aspects structuraux et S^0 est l'entropie qui se rapporte plutôt aux aspects dynamiques, ainsi qu'aux interactions avec l'eau . L'enthalpie donne des informations sur toutes les énergies d'interactions qui stabilisent le système moléculaire et la variation d'enthalpie, H^0, correspond aux interactions perdues ou nouvellement formées.

La mesure de la variation d'entropie, S^0, permet d'obtenir une information sur le désordre du système, qui comprend la protéine dans son solvant (l'eau). Si $S^0 < 0$ le système est plus

ordonné alors que le désordre augmente quand $S^0 > 0$, une situation qui spontanément réalisée.

Les deux contributions enthalpiques et entropiques s'opposent lors du repliement d'une protéine. En effet, l'enthalpie diminue avec le repliement du fait des interactions nouvellement formées au sein du polypeptide alors que le degré d'ordre augmente, ce qui fait diminuer l'entropie conformationelle. Au cours du repliement, l'enfouissement de chaînes latérales non-polaires peut conduire à une augmentation de l'entropie du solvant, qui compense largement le coût entropique lié à la baisse du nombre d'états conformationnels de la protéine. En effet, L'hydratation des groupements apolaires d'une chaîne polypetidique dépliée ou de zones apolaire d'une protéine repliée est réalisée par des molécule d'eau ordonnées. Lorsque l'eau est exclue par le rapprochement des groupements apolaires lors du repliement d'une protéine, le désordre et donc l'entropie de ces molécules d'eau augmente. C'est ce qu'on appelle l'effet hydrophobe.

.

Si l'on revient maintenant aux IDPs, on comprend bien que la notion de flexibilité correspond aux différents états que peuvent adopter ces protéines, caractérisés par une énergie libre quasi identique. Ainsi la différence d'énergie entre les différents états est relativement petite et l'énergie d'activation nécessaire pour passer d'un état à l'autre est faible. Dans ce cas, la chaîne polypeptidique échantillonne rapidement un grand nombre d'états.

Quels sont les avantages d avoir une chaîne exible et dynamique pour une protéine ?

Cette dynamique a pour effets, d'une part, d'augmenter la surface d'interaction comparée à une protéine globulaire, et, d'autre part, de permettre l'association avec différents partenaires du fait de cette plasticité structurale. Alors que les protéines globulaires n'utilisent que 2 à 15% de leur surface, les IDPs en utilisent 70% pour interagir avec leur partenaires [Gunasekaran 2004, Mészáros 2007]. De plus, la nature chimique de ces surfaces d'interaction est une particularité des IDPs puisque les quelques résidus hydrophobes qui composent leur séquence primaire peuvent être utilisés pour l'interaction et non pour le repliement autour d'un coeur hydrophobe comme pour les protéines globulaires. Ces propriétés des IDPs font que la spécificité est indépendante de l'affinité de liaison. En effet, les grandes surfaces d'interaction permettent d'obtenir une haute spécificité. Le coût entropique de l'interaction est généralement important et l'affinité de liaison est faible. Cela permet alors des interactions hautement spécifiques particulièrement importantes dans les voies de signalisation et de régulation. Aussi, grâce au pourcentage élevé de surface d'interaction, les IDPs sont capable d'interagir avec différents partenaires pour adopter alors une structure différentes selon le partenaire, c'est la plasticité.

C'est le cas par exemple de la protéine p53, qui grâce à sa partie C-terminale intrinsèquement désordonnée, peut se lier à plusieurs partenaires différents (Figure 4.5) [Oldfield 2008].

4.2.3 Transition désordre-ordre

La découverte des IDPs va de pair avec la découverte d'un mécanisme d'interaction particulier de ces protéines. Le terme de repliement induit a ensuite été adopté pour caractériser une transition désordre-ordre lors de l'interaction entre une IDP et un partenaire moléculaire. Cette transition peut concerner soit la protéine entière, soit un ou plusieurs domaines qui la composent, soit de petites régions correspondant à des motifs de reconnaissance moléculaire MoRF (Molecular Recognition Feature) ou SLiM (Small Linear Motif) [Vacic 2007, Cheng 2007]. Il s'agit de courts motifs d'environ 3 à 11 résidus et dont le classement est basé sur leur capacité à se replier en hélice , feuillet ou toute autre conformation irrégulière. De plus, un même motif peut servir d'interaction avec différents partenaire et peut adopter plusieurs structures différentes.

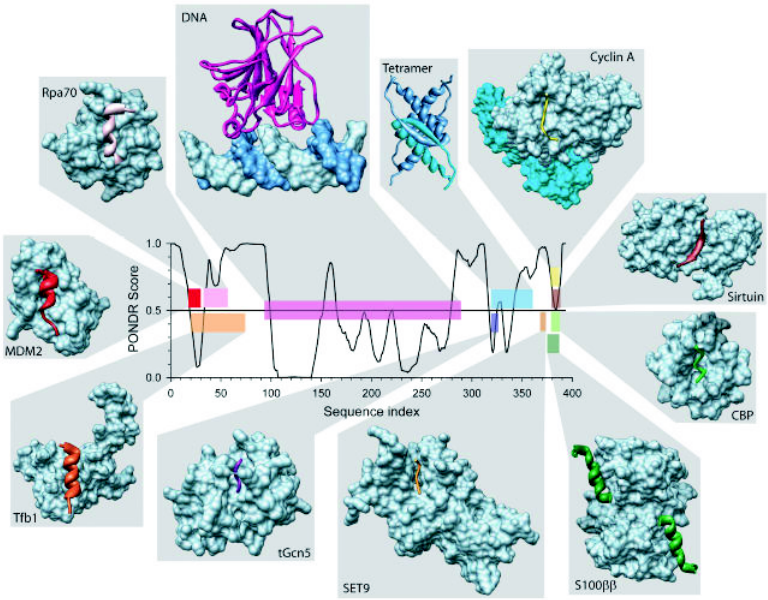


Figure 4.5 – Illustration des multiples interactions de p53. Le graphique correspond à la prédiction de désordre à partir de l'algorithme PONDR. Les régions ordonnées ont des scores inférieurs à 0,5 tandis que les scores supérieurs correspondent à des régions désordonnées. Les différents domaines, ordonnés et non ordonnés, de p53 sont impliqués dans des interactions protéine-protéine, et un même domaine tel que le domaine N- ou C-terminal est capable d'adopter différentes structures selon l'interactant. Figure extraite de [Uversky 2009]

C'est ce que l'on appelle le *one-to-many* et *many-to-one binding* et l'exemple le plus connu est le facteur de transcription p53 cité précédemment (Figure 4.5). La partie N-terminale de cette protéine se replie en hélice après interaction avec Mdm2 ou Tfb1 et la partie C-terminale se replie en hélice après interaction avec S100 , en feuillet après interaction avec Sirtuine et adopte une conformation étendue en présence de CBP et de la Cycline A [Oldfield 2008, Uversky 2009].

Trois mécanismes ont été proposés pour décrire le phénomène de repliement induit :

- *la sélection conformationnelle*. Une conformation particulière de l'IDP est sélectionnée parmi l'ensemble des conformations accessibles à la chaîne. Cette conformation permet la présentation d'un site de liaison au partenaire pour faciliter l'établissement de contacts favorables à l'interaction [Csermely 2010].
- *le repliement après interaction*. L'IDP se replie après interaction avec son partenaire.
- *le y-casting*. Ce mécanisme postule qu'un site de liaison attaché à une chaîne flexible s'apparente à un hameçon sur une canne à pêche permettant l'échantillonnage de grands volumes favorisant la prise du partenaire [Shoemaker 2000, Levy 2004, Trizac 2010]. Les aspects importants de ce mécanisme sont : d'une part, la flexibilité des IDPs permettant la formation d'un état intermédiaire non stable stabilisé par des interactions non spécifiques, suivit par le repliement et l'interaction spécifique avec le partenaire, et d'autre part, le rayon hydrodynamique des IDPs (plus grand qu'une protéine globulaire) étant plus efficace pour les interactions.

En général, la transition désordre-ordre d'une IDP ne résulte pas d'un seul mécanisme mais d'une combinaison de ces trois mécanismes.

La RMN a permis de caractériser en détail le mécanisme de repliement induit correspondant à l'interaction entre le domaine phosporylé KID (pKID, domaine d'activation de la transcription inductible par les kinases) du facteur de transcription CREB [Korzhnev 2008]. Le domaine KID est intrinsèquement désordonné en solution [Radhakrishnan 1997], mais se replie en deux hélices après avoir interagit avec son partenaire, le domaine KIX du co-activateur de la transcription CBP/p300, qui lui, est bien replié [Radhakrishnan 1998]. Lors de leur interaction les domaines forment un complexe de faible affinité stabilisé par des interactions hydrophobes non spécifiques dans lequel pKID est en échange rapide avec la forme libre. Dans ce complexe, pKID reste déstructuré et seule sa partie C-terminale est impliquée dans le complexe. Le complexe est en équilibre avec une forme intermédiaire de repliement où pKID interagit sur toute sa longueur pour finalement se replier en deux hélice orthogonales (Figure 4.6) [Sugase 2007,

Wright 2009].

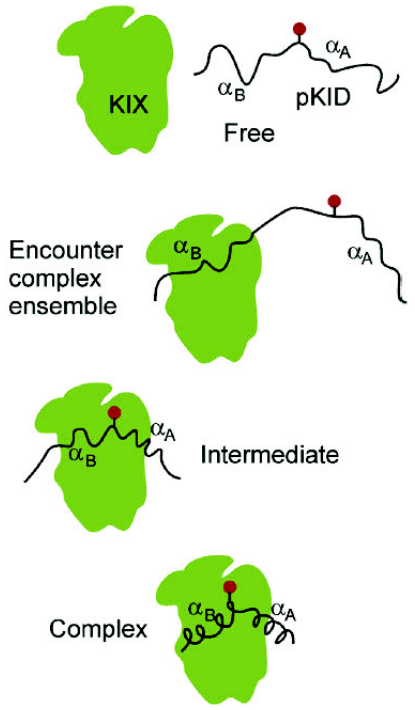


Figure 4.6 – Schéma du repliement induit du domaine phosphorylé KID (pKID) du facteur de transcription CREB lors de l'interaction avec le domaine KIX du coactivateur de la transcription CBP/p300. pKID est désordonné en solution et interagit avec KIX via son extrémité C-terminale pour former un complexe intermédiaire dynamique. S'en suit la liaison sur toute sa longueur et son repliement en deux hélices- . Figure extraite de [Wright 2009].

Cependant, ce modèle de repliement induit n'est pas applicable à toutes les IDPs puisque certaines d'entre elles restent désordonnées lors de leur interaction avec leur partenaire. Ce type de complexe, où le désordre est proéminent, a été qualifié de fuzzy complex [Tompa 2008, Fuxreiter 2012a, Fuxreiter 2012b]. La formation du complexe nécessite des régions mobiles capables d'adopter différentes conformations. Quatre manifestations de ce type de complexes ont été proposées et sont présentés en Figure 4.7 selon le niveau de désordre (local ou global).

Le complexe est qualifié de complexe polymorphique si l'IDP ou le segment échantillonne plusieurs conformations définies dans son état lié. Au contraire, si la transition désordre-ordre ne concerne qu'une partie de l'IDP et que les autres parties restent désordonnées les complexes sont qualifiés de modèle clamp, flanquants ou random. Une interaction protéine-protéine menant à ces fuzzy complex peut être décrite par plusieurs de ces modèles à la fois.

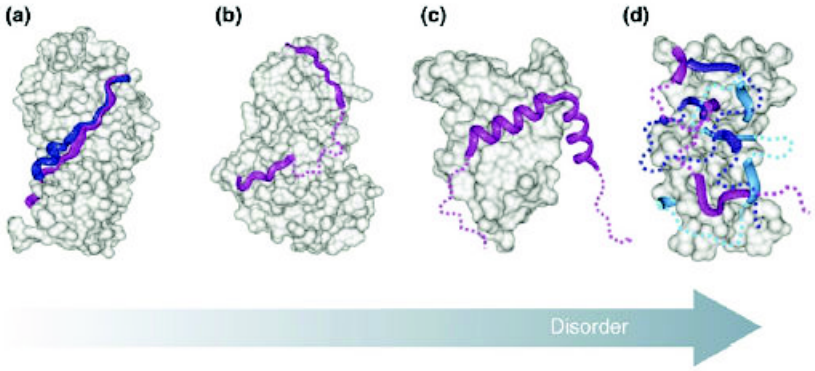


Figure 4.7 – Différents types de complexes impliqués dans les interactions protéine-protéine mettant en jeu des IDPs. L'IDP est représentée en couleur alors que la surface du partenaire est représenté en gris. (a) exemple d'un complexe polymorphique où plusieurs conformations sont présentes dans l'état lié avec le même partenaire. (b) modèle clamp : un segment désordonné sert de linker entre deux motifs d'interaction ordonnés et maintient leur liberté conformationnelle dans l'état lié de l'IDP. (c) modèle flanquant : une région adjacente à un motif d'interaction contribue à la stabilité du complexe en établissant des contacts transitoires, et (d) modèle random : la protéine reste désordonnée sur toute sa longueur pendant l'interaction. Le désordre devient de plus en plus proéminent en allant de (a) vers (d). Figure extraite de [Tompa 2008].

L'exemple de la protéine désordonnée Sic1 interagissant avec son partenaire Cdc4 permet de démontrer que le désordre structural peut être une propriété essentielle pour la fonction d'une IDP [Mittag 2010]. Sic1 contient neufs sites de reconnaissance pour Cdc4 alors que cette dernière ne contient qu'un seul site d'interaction pour Sic1. La phosphorylation multiple des sites de Sic1 entraîne un repliement local permettant d'établir des interactions transitoires avec Cdc4. Le complexe reste dynamique et les sites de phosphorylation sont interchangeables, ce qui permet une régulation très fine du cycle cellulaire.

4.3 Fonction des IDPs

Les IDPs sont retrouvées dans les trois règnes du vivant mais dans des proportions différentes. Ainsi, 2% du génome des Archae et 4% du génome des eubactéries contiennent des IDPs alors que les génomes eucaryotes en contiennent 30% [Oldfield 2005]. On constate donc que la fraction des IDPs augmente avec la complexité des organismes. De plus, il est décrit dans la littérature que plus de la moitié des protéines eucaryotes contiennent des régions désor-

données et que 70% des protéines impliquées dans des voies de signalisation possèdent des régions désordonnées [Iakoucheva 2002].

On peut donc penser que les IDPs jouent un rôle majeur dans les processus de régulation et les voies de signalisations où leur capacité à interagir avec de multiples partenaires de manière spécifique et de faible affinité est nécessaire. Une classification en six grandes classes a été proposé par P. Tompa en 2005. Cinq d'entre elles mettent en jeu des mécanismes de reconnaissance moléculaire soit transitoires soit permanents (Figure 4.8).

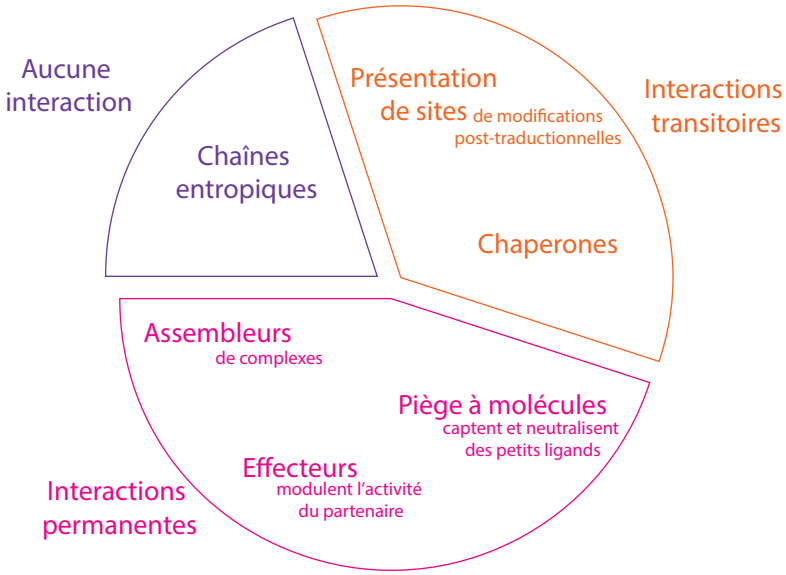


Figure 4.8 – Classification des IDPs selon leur fonctions. Adapté de [van der Lee 2014].

La première classe est celle des chaînes polypeptidiques entropiques dont le désordre est directement relié à leur fonctions. Les caractéristiques dynamiques de ces séquences sont utilisées pour conférer des propriétés mécaniques particulières à des segments de protéines localisés entre deux domaines structurés (segments interstitiels ou linkers). Ces séquences permettent d'ajuster la distance, l'orientation et la mobilité relative des deux domaines, indépendamment de l'interaction avec un partenaire. Un exemple est la région désordonnée riche en proline, glutamine, valine et lysine (PEVK) de la titine qui fonctionne comme un ressort et confère aux muscles leur élasticité [Trombitás 1998]. On peut également citer les protéines associées aux microtubules de type II (MAPs) qui incluent MAP2 et tau. Ces protéines sont constituées de deux domaines : un domaine de projection situé en N-terminal et un domaine associé aux microtubules situé en C-terminal. Le domaine de liaison aux microtubules est globulaire et sa séquence est conservée parmi les MAPs alors que le domaine de projection est désordonné et peu conservé. Sa longueur peut aller d'environ 300 résidus pour la protéine tau

jusqu'à 1400 résidus pour la protéine MAP2. Par exclusion entropique, le domaine de projection permet un espacement suffisant entre les microtubules pour que la structure du cytosquelette soit maintenue [Mukhopadhyay 2001].

Pour les cinq autres classes, la fonction implique des interactions transitoires ou permanentes avec un autre partenaire. De nombreuses IDPs impliquées dans des interactions transitoires présentent des sites de modifications post-traductionnelles (PTMs) comme la phosphorylation, l'acétylation, l'ubiquitination, la méthylation et la sumoylation Ces PTMs modifient l'espace conformationnel accessible à une protéine donnée, ce qui conduit à une modification de leur propriétés de reconnaissance. C'est le cas de la phosphorylation qui est une PTM fréquemment rencontrée dans les processus de signalisation. La signalisation peut être régulée par addition d'un seul groupement phosphate comme par exemple l'activation du facteur de transcription CREB via la phosphorylation d'une sérine appartenant au domaine KID décrit plus haut. La mise en évidence expérimentale d'une transition désordre-ordre à la suite d'une phosphorylation a été possible dans le cas d'un facteur de traduction eucaryote, 4E-BP2, qui interagit avec eIFE4 pour réprimer la traduction. La phosphoryltaion de 4E-BP2 sur plusieurs résidus induit son repliement ce qui abolit l'interaction avec eIFE4 et permet l'initiation de la traduction [Bah 2014].

Les IDPs chaperonnes constituent un second groupe de protéines impliquées dans des interactions transitoires. Ces chaperonnes interagissent avec de nombreuses protéines afin de les replier correctement sans pour autant prendre part à leur structure finale. La fonction chaperonne est directement due au désordre moléculaire puisque la partie dépliée de la chaperonne permet la reconnaissance, la solubilisation et le relâchement de la protéine mal repliée. Le mécanisme de transfert entropique entre la chaperonne et sa protéine cible à été suggéré [Tompa 2004]. Ce transfert permet une exploration plus large du paysage énergétique conformationnel pour la protéine cible allant jusqu'à atteindre son minimum. Un exemple sont les caséines, plus particulièrement l' $_s$caséine qui prévient la précipitation des protéines du petit lait avec l'élévation de la température [Bhattacharyya 1999]. De plus, ces auteurs ont également montré le rôle de chaperone de cette caséine sur l'alcool deshydrogénase ou la citrate synthase pour prévenir l'agrégation thermique.

Les IDPs, dont la fonction requiert une interaction permanente avec un partenaire, appartiennent à l'une des trois classes suivantes : effecteurs , assembleurs ou piège à molécules . Les *e ecteurs* se lient à leur partenaire d'interaction pour en moduler l'activité. Leur action peut être inhibitrice ou activatrice comme l'exemple de la protéine p21^{Cip1} et son homologue p27^{Kip2} [Olashaw 2004, Mitrea 2012]. Selon le cas, ils inhibent ou activent les kinases cyclines dépendantes (Cdks).

Les IDPs appartenant à la classe des *assembleurs* permettent d'assembler des complexes multiprotéiques. On les retrouve dans des processus tels que l'assemblage du ribosome et du cytosquelette ou dans l'assemblage de complexes de régulation de la transcription, comme par exemple le coactivateur de la transcription CBP qui interagit avec des partenaires variés [Wright 2015]. Le domaine SCA7 de l'ATXN7 dont la partie C-terminale est désordonnée fait également parti de cette classe. En effet, cette protéine appartient au module de deubiquitination (mDUB) de SAGA, un complexe multiprotéique ayant une activité de deubiquitination et d'acétylation des histones [Bonnet 2010, Koehler 2014].

Enfin, la dernière classe recense les IDPs *pièges à molécules* qui stockent et/ou neutralisent des petits ligands. Les caséines permettant d'inhiber la formation de phosphate de calcium dans le lait donne un bon exemple de ce type de fonction [Tompa 2005a]. De même, les glycoprotéines salivaires riches en prolines neutralisent les tanins [Bennick 2002].

Cette classification est basée sur le mode d'action des IDPs. Une remarque importante est que ces différents modes de fonctionnements ne sont pas exclusifs puisqu'une même protéine ou encore une même région intrinsèquement désordonnée d'une protéine peut exercer des fonctions différentes comme l'illustrent les différents exemples cités jusqu'ici. Ce phénomène, qui permet d'augmenter la complexité des réseaux intracellulaires sans augmenter le nombre de protéines, a été qualifié de moonlighting [Tompa 2005b]. De ce fait, les IDPs se retrouvent au coeur des réseaux d'interactions intracellulaires et sont qualifiées de concentrateurs ou protéines hub .

Du fait de leur importance dans les processus de signalisation et de régulation au sein de la cellule, les IDPs sont fréquemment associées à de nombreuses pathologies incluant le cancer, les maladies cardiovasculaires et les amyloïdoses. Les études à ce propos ont menées au concept D^2 pour désordre dans le désordre [Uversky 2008]. Pour citer quelques exemples, la protéine p53 est impliquée dans le processus tumoral, l'hirudine dans la maladie cardiovasculaire. p53 est un facteur de transcription suppresseur de tumeur dont la perte de fonction est associée au dévelopement tumoral [Arkin 2005]. La perte de fonction de p53 est principalement due à son interaction avec MDM2, une ubiquitine ligase menant à sa dégradation. Cette interaction s'effectue via la partie N-terminale désordonnée de p53. Les amyloïdoses regroupent les maladies neurodégénératives comme les maladie d'Alzheimer et de Parkinson, les maladies non-dégénératives comme le diabète de type II et aussi les maladies à prions comme la maladie de Creutzfeldt-Jakob. Ces pathologies sont définies par des dépôt insolubles d'agrégats protéiques dans les cellules. Les protéines mises en causes sont l' -synucléine pour la maladie de Parkinson, la protéine Tau hyperphosphorylée dans la maladie d'Alzheimer, et l'amyline dans le diabète de type II. Par exemple, la modification du code de phosphorylation de Tau entraîne une dérégulation des interactions avec ses partenaires (protéines ou ADN) ainsi que sa perte

de fonction observée dans la maladie d'Alzeihmer [Lippens 2012, Qi 2015, Sottejeau 2015].

Les IDPs constituent ainsi une nouvelle classe de cibles thérapeutiques ayant pour but la modulation des interactions protéines-protéines. A l'heure actuelle, il existe deux approches pour la recherche thérapeutique ciblant les IDPs. La première consiste à bloquer l'interaction de l'IDP avec son partenaire structuré. La molécule médicament mime le comportement de l'IDP et entre en compétition avec elle pour l'interaction avec le partenaire [Cheng 2006]. Un exemple bien connu est donné par les inhibiteurs de l'interaction entre p53 et l'ubiquitine ligase MDM2, où le repliement en hélice de la partie N-terminale désordonnée de p53 est impliquée dans l'interaction. Les inhibiteurs utilisés présentent la même structure que les résidus de p53 qui sont impliqués dans l'interaction.

La seconde méthode utilise des petites molécules capables d'interagir directement avec l'IDP. Ces petites molécules forcent l'IDP à adopter une structure 3D différente de celle observée dans le complexe avec son partenaire de façon à empêcher l'interaction avec celui-ci [Hammoudeh 2009, Metallo 2010]. Le facteur de transcription c-Myc, dont la surexpression est récurrente dans les cancers, illustre cette seconde stratégie. L'interaction de c-Myc avec son partenaire via une région désordonnée peut être bloquée par des petites molécules ce qui fait preuve de concept que les IDPs puisse être ciblée dans les approches de Drug-Design .

4.4 Méthodes de caractérisation des IDPs

4.4.1 Outils bio-informatiques de prédiction du désordre

Les motifs d'interaction protéique ainsi que les caractéristiques intrinsèques des séquences primaires des IDPs ont mené au développement d'outils bio-informatiques capable, d'une part, de prédire le désordre structural, et d'autre part, d'identifier ces motifs de reconnaissance moléculaire au sein des régions désordonnées des protéines.

La bio-informatique à joué un rôle crucial dans l'études des IDPs et reste à ce jour le moyen d'estimer le plus rapidement le degré de désordre d'une protéine à partir de sa séquence primaire.

Au cours des dernières années, de nombreux prédicteurs de désordre ont vu le jour dont quelques exemples sont présentés dans le tableau 4.1.
Ces prédicteurs sont classés en trois groupes selon les critères qu'ils utilisent pour faire la distinction entre ordre et désordre [Hazy 2009].

• L'approche la plus simple se base sur la propensité de chaque résidu à être impliqué dans des régions ordonnées ou désordonnées à partir de ses caractéristiques physico-chimiques. Ce

type d'algorithme reprend la combinaison de la charge nette et de l'hydrophobicité suggéré par Uversky et codé dans le logiciel FoldIndex par exemple [Prilusky 2005].

• La seconde approche se base sur un réseau de neurones artificiels qui combinent la charge nette, l'hydrophobicité et la complexité de séquence pour évaluer le degré de désordre d'une protéine. PONDR (Predictor Of Natural Disordered Region) est le premier prédicteur de ce type, et le premier prédicteur de régions désordonnées au sein des protéines, à avoir été développé [Li 1999].

• Enfin la dernière approche estime la capacité d'une chaîne polypeptidique à former des contacts stabilisants comme le programme IUPRED (Prediction of Intrinsically Unstructured Proteins) [Dosztányi 2005].

Dans le cas d'une protéine globulaire, de nombreux contacts inter-résidus sont formés, ce qui fournit l'énergie de repliement. Cet algorithme donne une estimation de l'énergie d'interaction entre les résidus composant la séquence primaire d'une protéine. Les paramètres pris en compte sont la nature du résidu ainsi que son environnement au niveau de la séquence primaire.

Prédicteurs	Source web	Références
PROPENSITÉ		
FoldIndex	http://bip.weizmann.ac.il/fldbin/findex	[Prilusky 2005]
GlobPlot	http://globplot.embl.de	[Linding 2003a]
RÉSEAU NEURAL ARTIFICIEL		
PONDR	http://www.PONDR.com	[Li 1999]
DISOPRED2	http://bioinf.cs.ucl.ac.uk/disopred	[Ward 2004]
DisEMBL	http://dis.embl.de	[Linding 2003b]
CONTACTS INTER-RÉSIDUS		
IUPRED	http://iupred.enzim.hu	[Dosztányi 2005]
MÉTASERVEURS		
MeDOR	http://www.vazymolo.org/MeDOR/	[Lieutaud 2008]
metaPRDOS	http://prdos.hgc.jp/meta/	[Ishida 2008]

Tableau 4.1 – Exemples d'algorithmes et métaserveurs permettant de prédire le désordre à partir de la structure primaire d'une protéine et leur lien url correspondant. D'après [Hazy 2009]

Aucune des méthodes développées ne permet d'obtenir une fiabilité absolue de prédiction. Aussi, il est nécessaire de croiser les résultats obtenus à partir de différents prédicteurs de désordre pour augmenter la fiabilité. C'est pourquoi des métaserveurs combinant les différentes

prédictions ont été développés. Parmi ces métaserveurs, MeDor combine plusieurs algorithmes de prédiction et, à partir de la séquence protéique, fournit une représentation graphique permettant de visualiser les régions désordonnées [Lieutaud 2008].

Cette prédiction du désordre a permis d'estimer l'importance des IDPs au sein des génomes et d'identifier leur potentielles fonctions associées en caractérisant les sites d'interactions ou sites de modifications post-traductionnelles. Actuellement, ELM (Eukaryotic Linear Motif) et MnM (Minimotif Miner) sont les deux bases de données qui recensent tous les motifs d'interactions décrit dans la littérature [Dinkel 2014, Balla 2006]. D'autres méthodes de détection *de novo* de ces motifs ont été mises au point telle que ANCHOR qui utilise les propriétés physico-chimiques de la séquence primaire des IDPs [Dosztányi 2009] pour les identifier directement au niveau des régions désordonnées.

Enfin, ces outils bio-informatiques présentent leur limites quant à l'étude du désordre et d'autres outils plus résolutifs pour étudier l'état déplié sont nécessaires.

4.4.2 SDS-PAGE

La migration sur gel SDS-PAGE fourni une méthode rapide pour mettre en évidence le caractère désordonné d'une protéine. Comme décrit plus haut, la séquence primaire des IDPs est riche en acides aminés chargés, ce qui rend moins efficace la fixation du détergent SDS. Les IDPs possèdent, d'autre part, un rayon hydrodynamique plus grand qu'une protéine compacte de même poids moléculaire théorique.

Ces deux effets conduisent à des altérations de la mobilité électrophorétique.

Ainsi, il a été montré que les IDPs migrent à des masses moléculaires apparentes jusqu'à 1,8 fois plus grandes que leur masse théorique calculée à partir de la séquence protéique [Iakoucheva 2001].

4.4.3 Dichroïsme circulaire (CD)

Le dichroïsme circulaire (CD) correspond à la différence d'absorbance entre une lumière polarisée circulaire droite et gauche. Cette différence d'absorption est convertie en une unité indépendante de la concentration : l'ellipticité molaire, et le spectre CD est obtenu en mesurant l'ellipticité en fonction de la longueur d'onde.

Cette différence d'absorption résulte de la présence d'une organisation asymétrique des chromophores au sein de la macromolécule. La liaison peptidique ainsi que les chaînes latérales des résidus aromatiques sont les deux types de chromophores optiquement actifs dans une protéines. Un spectre CD enregistré dans l'UV lointain (180-250 nm) dépend de la chiralité

de la liaison peptidique induite par les éléments de structures secondaires de type hélices-α, hélice 3_{10} ou polyprolyne II, feuillets-β, boucles-β ou structure irrégulière communément appelée random coil. Des spectres de référence ont été établis pour chaque type de structure secondaire (Figure 4.9).

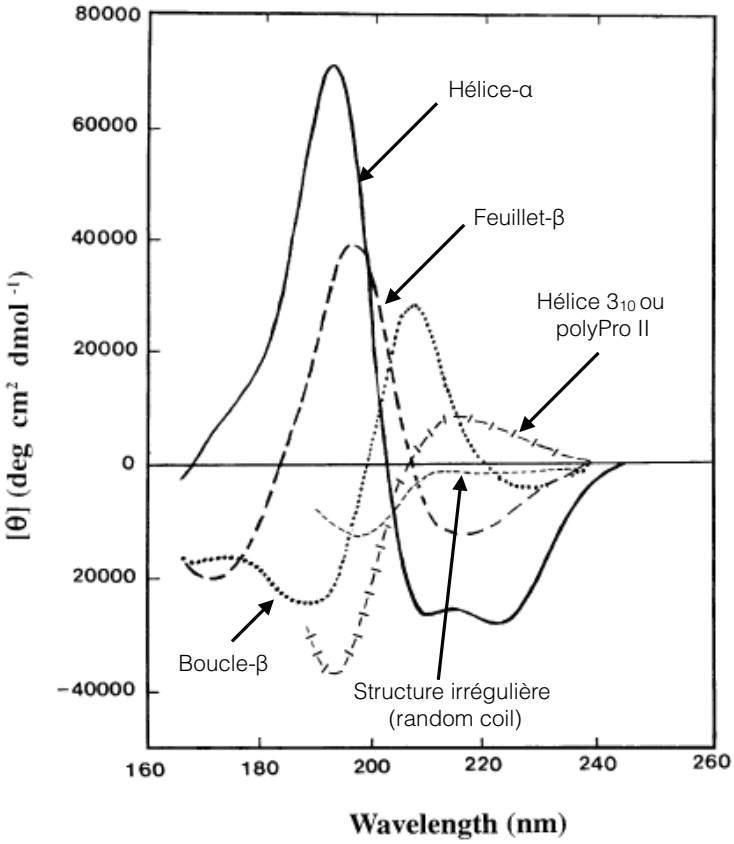


Figure 4.9 – Spectre de dichroïsme circulaire dans l'UV lointain pour chaque type de structure secondaire : hélice-α, hélice 3_{10} ou polyprolyne II, feuillet-β, boucles-β et random coil. Adapté de [Kelly 2005].

Ainsi, une mesure de dichroïsme circulaire dans l'UV lointain permet de détecter la présence d'éventuelles structures secondaires dans une protéine à partir du spectre de la protéine d'intérêt et d'une base de spectres de référence constituée de spectres CD de protéines dont la structure cristallographique est connue. Plusieurs méthodes permettant d'estimer la proportion de structures secondaires dans une protéine sont disponibles, dont la plus connue est le programme CONTIN [Provencher 1981].

En ce qui concerne les mesures de dichroïsme circulaire dans la région de l'UV proche (250-350 nm), les résultats découlent de la symétrie de l'environnement des acides aminés

aromatiques ainsi que des ponts disulfures [SYu 1994]. Ceci permet de caractériser la structure tertiaire d'une protéine. En général, le signal dans l'UV-proche d'une IDP est quasi nul néanmoins il arrive que des IDPs possèdent un spectre relativement complexe indiquant que les résidus aromatiques sont dans un environnement asymétrique du fait de la présence de structures résiduelles au sein de ces protéines [Kelly 2000].

Cette technique a pour avantage de ne consommer qu'une quantité réduite de matériel puisque les concentrations requises sont de l'ordre de la dizaine de micromolaire. Cependant, les informations structurales obtenues grâce à cette méthode sont uniquement d'ordre global.

4.4.4 Spectroscopie de uorescence

La spectroscopie de fluorescence donne des informations sur l'environnement local d'un fluorophore et permet de détecter d'éventuelles structures tertiaires au sein des protéines. Les résidus aromatiques sont à l'origine de la fluorescence intrinsèque des protéines et l'émission de fluorescence est due principalement à l'excitation des tryptophanes (Trp). Pour obtenir une excitation sélective du Trp nous choisirons une excitation à 295 nm. Le Trp constitue une bonne sonde de structures tertiaires locales et est également utilisé pour l'étude de la transition désordre/ordre ou ordre/désordre d'une protéine. Le maximum d'émission du Trp dans un solvant fortement polaire comme l'eau se situe à 355 nm, alors qu'un décalage vers 325 nm est observé dans une solution apolaire comme le dioxane [Eftink 1994]. De même, le rendement quantique augmente dans un environnement hydrophobe comme dans un coeur hydrophobe de protéine repliée, alors qu'une exposition au solvant peut entraîner une diminution importante du rendement quantique comme pour une IDP. Ce rendement quantique détermine l'efficacité de fluorescence et correspond au rapport entre le nombre de photons émis et absorbés.

Les résultats obtenus par cette méthode sont à prendre avec précaution puisqu'un décalage du maximum de fluorescence vers de basses longueur d'onde ou une augmentation de l'intensité ne sont pas forcément synonymes de repliement natif de la protéine mais peuvent correspondre à des agrégats de celle-ci en solution.

Les protéines ne possédant pas de Trp dans leur séquence primaire peuvent être couplées à des fluorophores soit par mutagenèse soit par greffage avec un fluorophore comme l'ANS, (1-anilino-naphtalène-8-sulfonate) [Semisotnov 1991, Hawe 2008]. Dans ce cas, on s'assurera que la présence du fluorophore n'altère pas les propriétés de repliement ou de désordre dans la protéine étudiée.

Cette méthode est très utilisée du fait de sa sensibilité et de sa simplicité. Cependant,

comme pour le CD elle ne permet d'accéder qu'à une information plutôt qualitative et globale.

4.4.5 Di usion des rayons X aux petits angles (SAXS)

La diffusion des rayons X aux petits angles, communément appelée SAXS, permet d'obtenir des informations sur la forme globale, la taille et masse moléculaire d'une protéine ou d'un complexe en solution. Par ailleurs, le SAXS permet également d'étudier la flexibilité d'une protéine et ses différentes conformations.

Cette technique n'est pas limitée par la masse moléculaire et peut être utilisée dans des conditions proches des conditions physiologiques. Cependant, c'est une méthode de basse résolution (10-15 Å) et il est souvent nécessaire de combiner les résultats obtenus avec ceux résultant d'autres méthodes structurales comme la cristallographie ou la RMN.

Principe de la mesure de di usion

Les rayons X sont des ondes électromagnétiques dont la longueur d'onde varie de 0,01 à 10 nm. L'absorption et la diffusion sont les deux types d'interactions des rayons X avec la matière. Lorsque les rayons X viennent frapper un échantillon, une partie passe au travers de celui-ci sans interagir avec la matière, une autre partie est absorbée pour être transformée en une autre forme d'énergie et, enfin, une dernière partie est diffusée dans une autre direction de propagation. En ce qui concerne la diffusion, l'onde incidente et l'onde de diffusion sont décrites par leur vecteurs d'onde respectifs k_0 et k_1. La diffusion peut être accompagnée d'une perte d'énergie, dans le cas d'une diffusion incohérente, ou non, dans le cas d'une diffusion cohérente ou élastique. Le SAXS ne considère que la diffusion élastique et les vecteurs d'ondes incidents et diffusés ont tous deux une magnitude de 2 / avec correspondant à la longueur d'onde des rayons X. La différence entre ces deux vecteurs d'onde correspond au vecteur de diffusion $k_1 - k_0 = q$ (Figure 4.10). L'intensité des rayons diffusés (I) est mesurée à différents angles de diffusion (2) par un détecteur situé à une certaine distance de l'échantillon. Puis les données expérimentales sont représentées sous forme d'une courbe d'intensité (I) en fonction du vecteur de diffusion q (en nm^{-1} ou $^{-1}$) défini par la relation :

$$q = 2\,|k|\sin \quad = \frac{4}{} \sin \tag{4.2}$$

où est l'angle incident, la longueur d'onde du faisceau de rayon X.

On peut dire que le SAXS est une méthode de contraste puisque cette technique permet de mesurer une différence entre les densités électroniques de la protéine d'intérêt et du solvant, λ(r) et λ(s) respectivement, correspondant à la différence de densité électronique moyenne λ(r).

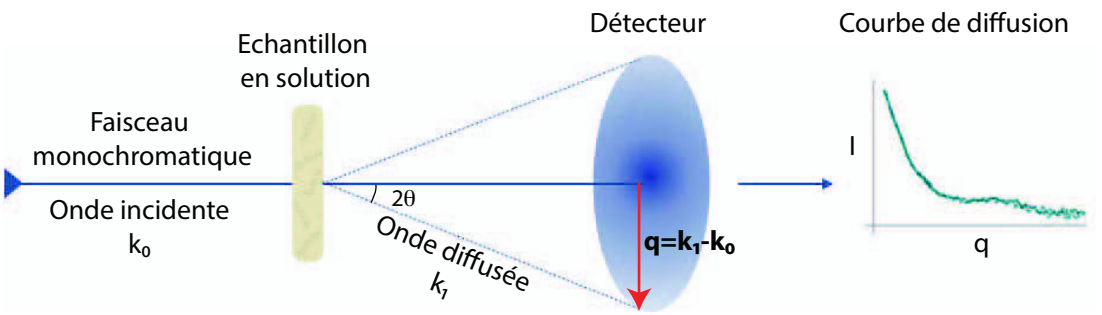


Figure 4.10 – Mesure de diffusion des rayons X aux petits angles (SAXS).

Les particules en solutions sont orientées de façon aléatoire. Si l'échantillon est monodisperse et dilué et que les particules n'interagissent pas entres elles, alors l'intensité totale résultera de la diffusion moyenne d'une seule particule dans toutes les orientations, c'est ce que l'on appelle le facteur forme. Au contraire, si l'échantillon est concentré et que les particules interagissent entre elles, la distance entre chaque particules est à prendre en compte. Cette interférence inter-particules est nommée facteur structure. Ainsi la diminution de l'intensité à petit q correspond aux forces de répulsions alors qu'une augmentation de l'intensité correspond à des forces d'attraction indiquant un phénomène d'agrégation. La courbe de diffusion $I(q)$ résume le type d'information obtenu à partir d'une expérience de SAXS. L'intensité de la lumière diffusée à un angle donné q est donné par :

$$I(q) = N(\Delta\rho V)^2 P(q) S(q) \tag{4.3}$$

où N est le nombre de molécules, V le volume spécifique de la protéine, $\Delta\rho$ le contraste, $P(q)$ le facteur de forme et $S(q)$ le facteur de structure.

Cette courbe peut être interprétée à différents niveaux qui dépendent des angles considérés (Figure 4.11).

Interprétation des données SAXS

Les premières informations directement obtenues à partir des données SAXS et ne nécessitant pas de modélisation sont : le rayon de giration, la fonction de distribution des distances P(r) et le profil de diffusion en représentation Kratky d'un protéine.

Le rayon de giration, R_g, renseigne sur la taille moyenne d'une molécule. Il est obtenu en analysant la portion de la courbe de diffusion $I(q)$ à faible valeurs de q.

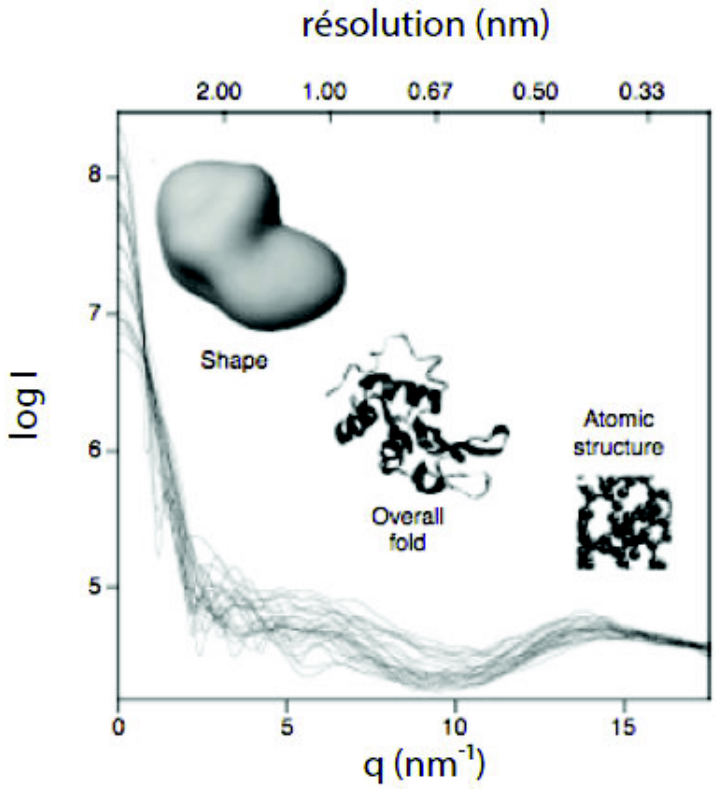


Figure 4.11 – Courbes de diffusion aux rayons X de macromolécules de tailles et de formes différentes. Les courbes se distinguent à petits angles à une résolution de 2-3 nm. L'intensité I_0 permet de déduire la masse moléculaire et la pente de la courbe, à petit angle, dépend de la forme globale d'une macromolécule. En revanche, à grand angle, les courbes deviennent indiscernables, dû au fait que la méthode SAXS ne fournit pas une résolution atomique. Figure extraite de [Svergun 2002]

Pour un échantillon monodisperse, **la représentation de Guinier** ($\ln(I(q))\ vs\ q^2$) est une fonction linéaire dont l'intersection de la droite avec l'axe des ordonnées correspond à I_0 et la pente au R_g de la particule .

Ainsi, la représentation de Guinier renseigne directement sur la qualité des données et sur l'état d'agrégation de la protéine. La loi de Guinier est valable uniquement pour des petits angles q, c'est à dire jusqu'à $qR_g<$ 1,3 pour les protéines globulaires et seulement jusqu'à $qR_g<$ 0,8-1 pour les IDPs du fait de nombreuses conformations et tailles qu'elles peuvent adopter [Receveur-Brechot 2012].

L'intensité de diffusion transmise I_0 est reliée à la masse moléculaire de l'objet en solution :

$$I_0 = \frac{cM_w}{\mathcal{N}}[(\lambda_p - \lambda_s)v_p]^2 \tag{4.4}$$

où c est la concentration de l'échantillon en solution, $\mathcal{N}$ le nombre d'Avogadro, λ_p et λ_s la densité électronique de la protéine et du solvant, respectivement, et v_p le volume de la protéine. La détermination de la masse moléculaire dépend de la précision de la mesure de I_0 qui est sensible et dépend directement, entre autres, de la présence d'agrégats et des interactions intermoléculaires et de la précision de la mesure de concentration de la protéine.

Par ailleurs, les R_g expérimentaux peuvent être comparés à des R_g théoriques calculés pour des protéines globulaires ou dénaturées afin d'évaluer le degré de structuration de nos différentes protéines.

Plusieurs études basées sur la théorie de Flory, provenant de la physique des polymères, tentent d'estimer le R_g attendu pour une protéine globulaire ou désordonnée en fonction du nombre de résidus. Ces études reposent l'observation que le R_g évolue en fonction du nombre de résidus N en suivant une loi de puissance :

$$R_g = R_0 N^{\nu} \tag{4.5}$$

où R_0 est une constante, correspond à l'inverse de la dimension fractale de l'objet étudié, et N représente le nombre de monomère constituant le polymère. Théoriquement, pour un objet compact et sphérique =1/3 alors que pour un random coil =0,5 et =0,588 pour un polymère semblable à une protéine dénaturée. Ensuite, Plaxco et coll. ont recensé les dimensions d'un large nombre de protéines globulaires et dénaturées de la littérature et déterminé des valeurs de = 0,38 $\pm$ 0,05 et =0,598 $\pm$ 0,028 pour une protéine globulaire ou complètement dénaturée, respectivement, et un R_0=1,93 dans les deux cas [Kohn 2004]. Leur valeur obtenue pour une protéine dénaturée est alors proche de la valeur obtenue par Flory. Une autre étude, sur un nombre restreint de protéine globulaires et dénaturées, obtient des valeurs similaires avec le couple R_0=($\sqrt{3\ 5}$)4,75/ =0,29 et R_0=2,23/ =0,58 en condition native ou dénaturante [Wilkins 1999]. Enfin, à partir de données d'IDP en solution aqueuse étudiée par RMN et SAXS, Bernardo et Blackledge déterminent d'autres valeurs qui sont R_0=2,54 et =0,522 [Bernadó 2009]. Ces prédictions peuvent être utilisées comme limites supérieures et inférieures du R_g et un écart significatif du R_g expérimental par rapport à la limite supérieure théorique indique la présence de structure secondaire résiduelle. En effet, la présence de structure résiduelle a tendance à compacter une protéine alors que la présence d'une hélice de type polyproline II souvent présente dans les IDPs à tendance à augmenter la dimension.

La **représentation de Kratky** ($q^2 I(q)\ vs\ q$) permet d'obtenir rapidement des informations sur la nature compacte ou désordonnées d'une protéine. Cette représentation découle directement de la courbe expérimentale [Kratky 1949]. Pour une protéine globulaire, la représentation de Kratky correspondante sera en forme de cloche avec un maximum bien défini. Les chaînes polypeptidiques dépliées présentent, quant à elles, une courbe qui augmente avec les valeurs de q jusqu'à atteindre un plateau à grands q, suivi parfois par une remontée indiquant la présence de rigidités locales. Cependant avec cette représentation il est difficile voir impossible de faire la distinction entre une protéine globulaire et une protéine partiellement désordonnées possédant des domaines ou régions structurées puisque la représentation de Kratky est une courbe en forme de cloche dans les deux cas. Il est donc fortement recommandé d'utiliser la représentation de Kratky sans dimension ou Kratky normalisé ($(qR_g)^2 I(q)\ \ I_0\ vs\ qR_g$) [Durand 2010]. La multiplication par le R_g de la protéine et la division par le I_0 rend les données indépendantes de la taille et du poids moléculaire de la protéine respectivement.

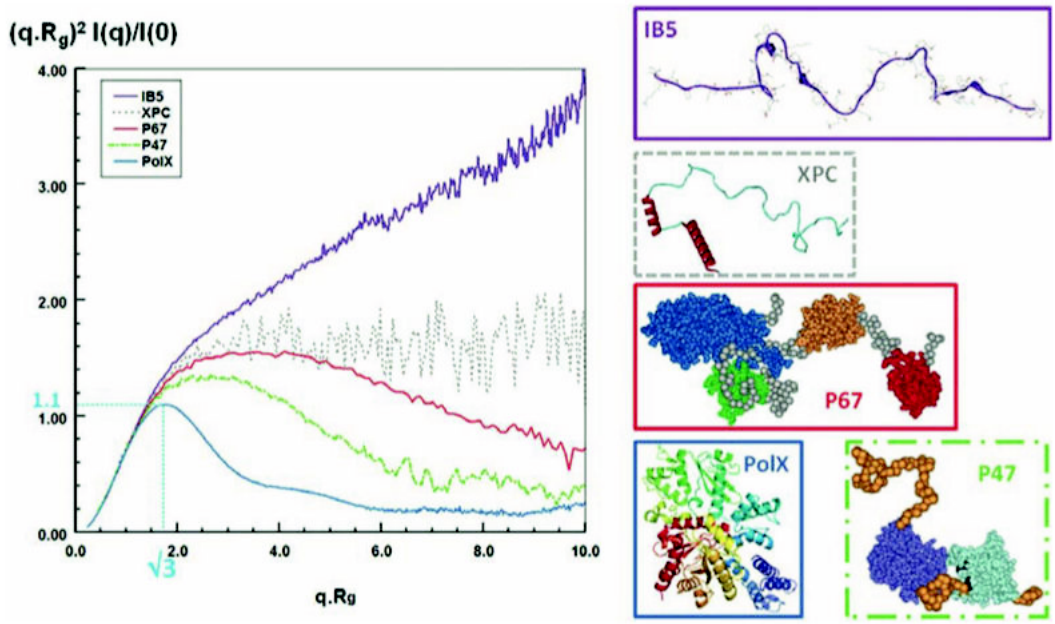


Figure 4.12 – Représentation de Kratky normalisée. Figure extraite de [Receveur-Brechot 2012]

Il est alors possible de comparer les protéines entres elles, qu'elles soient globulaires ou déstructurées, et ce quelle que soit leur taille. La courbe de diffusion d'une protéine globulaire dans une représentation de Kratky normalisée présente un maximum de 1,1 pour $qR_g = \sqrt{3}$ indépendamment de la taille de la protéine. Pour une protéine complètement désordonnée, la courbe de diffusion augmente avec les valeurs croissantes de qR_g pour atteindre un plateau entre 1,5 et 2 [Receveur-Brechot 2012]. Une augmentation de cette courbe pour des grandes valeurs de

qR_g indique la présence de rigidités locales au sein de l'IDP. Enfin, les protéines partiellement désordonnées présentent des profils de diffusion intermédiaires entre les deux extrêmes cités précédemment.

La **fonction de distribution des distances**, P(r), est un histogramme de toutes les distances interatomiques, r, au sein d'une protéine. Le programme GNOM permet de calculer la fonction de distribution des distances [Svergun 1992]. La valeur maximale de r pour la quelle P(r) est non nulle, D_{max}, correspond au diamètre maximal de la protéine. Cet histogramme fournit une information sur la forme et le degré de compaction de la protéine en solution. La fonction P(r) d'une IDP est asymétrique et possède souvent des valeurs non nulles pour des grandes valeurs de q due aux nombreuses conformations alors qu'elle est symétrique pour une protéine globulaire. De plus, pour une protéine constituée de deux domaines globulaires séparés par une région flexible, cette fonction est caractérisée par un premier pic à petites distances, correspondants aux distances interatomiques au sein du domaine, puis s'étend en présentant un épaulement plus ou moins prononcé, correspondant aux distances interdomaines. La Figure 4.13 présente différentes fonctions de distribution de distance pour des objets de même taille mais de forme différente.

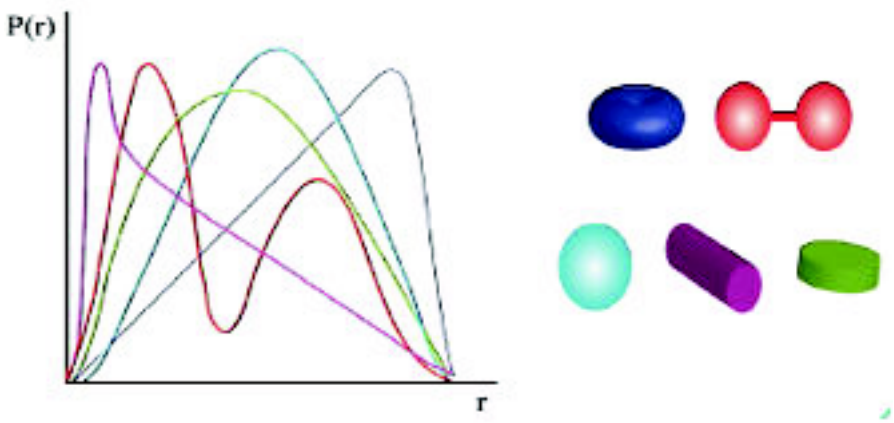


Figure 4.13 – Fonction de distribution de distance pour des objets de taille identique mais de formes différentes.

D'autre part, le R_g et le I_0 peuvent aussi être déduits de la fonction P(r). Cette méthode alternative n'est pas reliée à un modèle et utilise l'ensemble de la courbe de diffusion de la protéine. La comparaison des valeurs de R_g et I_0 obtenues à partir de la loi de Guinier et de la fonction P(r) fournit une méthode de validation de la qualité des données.

4.4.6 Résonance Magnétique Nucléaire (RMN)

En l'absence de repliement tridimensionnel, et donc de possibilité de cristalliser la protéine, la spectroscopie de résonance magnétique nucléaire (RMN) est la seule méthode permettant d'obtenir des informations sur les régions désordonnées des protéines, avec une résolution atomique [Dyson 2005].

Cette méthode utilise les propriétés magnétiques des spins nucléaires pour obtenir des informations structurales et dynamiques.

Les principaux noyaux utilisés en RMN biomoléculaire sont caractérisés par un spin $\frac{1}{2}$ et sont présentés dans le tableau 4.2.

Noyau	$(10^7$rad.$s^{-1}.T^{-1})$	Abondance naturelle
^{1}H	26,75	100%
^{15}N	-2,71	0,4%
^{13}C	6,73	1,1%
^{31}P	10,84	100%
^{19}F	25,18	100%

Tableau 4.2 – Rapport gyromagnétique ( ) et abondance naturelle des isotopes les plus couramment étudiés en RMN biomoléculaire.

L'azote 15 et le carbone 13 ayant une abondance naturelle de 0,4% et 1,1%, respectivement, il est nécessaire d'enrichir les protéines en isotopes pour augmenter la sensibilité de détection [Greenbaum 2001]. Le développement des techniques d'enrichissement ou marquage isotopique des protéines a permis l'amélioration et le développement d'expériences de RMN multidimensionelles permettant l'attribution des fréquences de résonance d'une protéine.

La RMN est une méthode non destructive qui permet, à partir d'un seul échantillon, de réaliser différentes mesures pour étudier différentes propriétés du système moléculaire (structure, dynamique, interactions, diffusion, ...). Ces propriétés s'expriment à travers les différentes observables RMN.

Les observables RMN

Une des spécificité de la RMN réside dans la multiplicité des observables. On peut définir trois échelles de temps en RMN :

- l'échelle de temps de Larmor , de l'ordre de la nanoseconde
- l'échelle de temps spectrale, de l'ordre de la milliseconde

- l'échelle de temps de relaxation de l'ordre de la seconde

L'interconversion des différents conformères d'une IDP ou son interaction avec un partenaire (ligand ou protéine) est un système dynamique pour lequel il y a sans cesse un échange entre plusieurs états (qui peuvent être une conformation par rapport à une autre ou la forme libre et liée lors d'une interaction) sur une échelle de temps inférieure à la milliseconde.

Fréquence de résonances

Les fréquences de résonance (ou déplacement chimique) pour les différents noyaux composants une protéine ou un acide nucléique permettent dès le premier abord d'évaluer le caractère replié de la protéine étudiée.

Pour une protéine désordonnée, les fréquences de résonance des différents groupes de protons présents dans les acides aminés sont essentiellement dépendantes de la nature chimique de l'acide aminé. Par exemple, les protons amides H_N résonnent entre 7,5 et 8,5 ppm alors que les protons correspondant à des groupes méthyles résonnent entre 0,8 et 1,5 ppm. Dans le cas d'une protéine repliée les fréquences correspondant aux H_N se dispersent sur une large gamme de valeurs comprises entre 6 et 10 ppm (Figure 4.14). Des tables de déplacement chimique correspondant aux fréquences caractéristiques des acides aminés sous forme désordonnée ont été établies. Ces tables sont basées, soit sur une analyse statistiques réalisée à partir d'une base de donnée de protéines [Gro 1988, Wishart 1991, Zhang 2003], soit à partir de mesures réalisés sur des petits peptides désordonnés [Thanabal 1994, Braun 1994, Merutka 1995, Wishart 1995b, Wishart 1995a]. Les différences entre ces tables sont les conditions expérimentales telles que le choix du solvant, de la température, du référencement des spectres ainsi que de la prise en compte ou non de l'effet des proches voisins dans la séquence primaire.

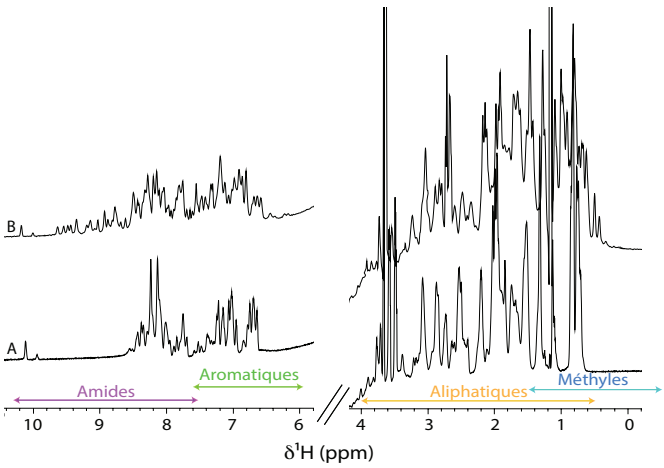


Figure 4.14 – Spectre 1D proton d'une protéine dépliée (A) et d'une protéine contenant la même partie dépliée en fusion avec un domaine globulaire (B).

Pour pouvoir attribuer les fréquences observées sur le spectre d'une protéine à chaque noyau qui la compose, il est nécessaire d'avoir recours à la RMN multidimensionnelle. L'expérience à deux dimension $^{1}H-^{15}N$ HSQC corrèle la fréquence d'un proton amide (H_N) avec celle de l'atome d'azote (^{15}N) qui le porte. Cela se traduit, dans le spectre, par la présence de tâches de corrélation correspondant à chaque résidu faisant parti de la séquence primaire exception faite des prolines et du premier résidu. Ce spectre constitue l'empreinte digitale de la protéine où chaque tâche de corrélation observée dépend de l'environnement chimique du groupement amide et reflète ainsi la structure 3D de la protéine.

Les déplacements chimiques des noyaux du squelette peptidique peuvent être interprétés en terme de structures secondaires. La déviation des valeurs de déplacements chimiques du squelette peptidique par rapport à ceux attendus pour une protéine désordonnée permet d'estimer la présence de structures secondaires au sein d'une protéine [Wishart 1992]. En particulier, le déplacement chimique des C est très sensible aux deux angles de torsion γ et . la valeur du déplacement chimique augmente dans une hélice- alors qu'ils diminuent dans les feuillets- . Les déplacement chimiques de différents noyaux, possédant des sensibilités différentes, peuvent être pris en compte pour l'analyse. Plusieurs programmes ont été développés pour utiliser l'information structurale contenue dans les spectres RMN et détecter les structures secondaires au sein des IDPs. Le programme SSP (Secondary Structure Propensities) permet d'estimer la fraction de structure secondaire contenue dans une IDP en combinant les valeurs de déplacements chimiques de différents noyaux du squelette peptidique selon leur sensibilité à la structure locale [Marsh 2006]. Le résultat obtenu est un score représentant le pourcentage de structure secondaire, des valeurs positives indiquant la présence d'une hélice- et des valeurs négatives indiquant la présence de feuillet- ou structure étendue.

Déplacements chimiques et phénomènes dynamiques

Une IDP étant un ensemble de conformères interconvertibles, les noyaux qui la composent observent alors des environnements chimiques différents. Ceci est également valable lors du repliement après interaction d'une IDP.

La mesure des déplacements chimique pour chaque noyau sera dépendante des gammes de temps du mouvement par rapport à la différence de fréquence entre deux conformations.

Le régime d'échange sera rapide, intermédiaire ou lent selon la valeur de la constante d'échange entre les différentes conformations, définie par k_{ex}, par rapport à la différence de déplacement chimique , entre la forme libre et liée (Figure 4.15).

- Pour $k_{ex} >>$: le régime d'échange est rapide car la vitesse d'échange est beaucoup

plus rapide que la différence de déplacement chimique en Hz, , entre la forme libre et liée. Le signal observé est une seule résonance correspondant à la moyenne des valeurs de déplacement chimique de l'état libre et lié moyenné par la population de ces états.

- Pour $k_{ex} \simeq$: la raie de résonance s'élargie et peut disparaître. Le régime d'échange est intermédiaire.
- Pour $k_{ex} <<$: le régime d'échange est lent et les espèces en échange sont discriminées par des valeurs de déplacement chimiques distinctes.

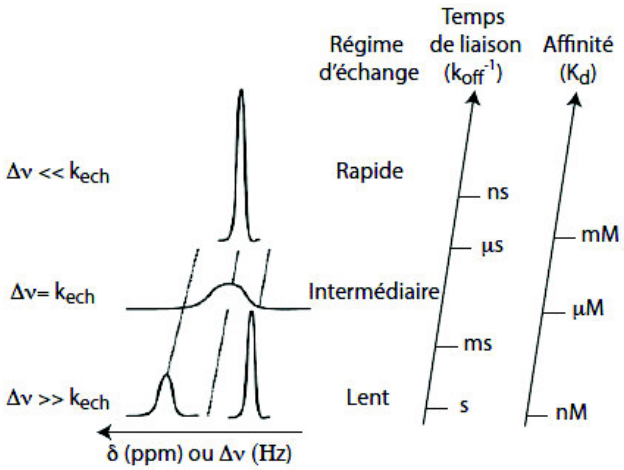


Figure 4.15 – Impact de la vitesse d'échange, définie par k_{ex}, entre les espèces libres et liées d'un complexe au niveau des spectres RMN. Lorsque l'interaction est faible ($K_d > 100 \ M$), l'échange rapide mène à l'observation d'une seule résonance. Pour les interactions fortes ($K_d < 1 \ M$), le régime d'échange lent entraîne l'oberservation de deux raies de résonances distinctes. Enfin, les échanges intermédiaires, dont la vitesse d'échange coïncide avec la vitesse de l'observable RMN, sont caractérisés par un élargissement des raies de résonance. Figure extraite de [Kieffer 2011]

Les différents conformères d'une IDP sont en échange rapide et part conséquent les spectres enregistrés présentent une seule corrélation par résidus composant la protéine. Ainsi, le déplacement chimique mesuré correspond à la moyenne des différentes conformations adoptées par l'IDP.

Utilisation des déplacements chimiques pour l'étude des interactions

Les interactions protéine/ligand mettant en jeu des IDPs sont des interactions spécifiques et de faible affinité. L'affinité se situe au dessus de quelques dizaines de micromolaire ce qui

correspond en général à un régime rapide entre espèces libres et liées.

L'interaction entre une protéine [P] et son ligand [L] est gouvernée par des vitesses :

$$[\mathrm{P}] + [\mathrm{L}] \underset{\mathrm{k_{off}}}{\overset{\mathrm{k_{on}}}{\rightleftharpoons}} [\mathrm{PL}] \tag{4.6}$$

où k_{on} et k_{off} sont respectivement les constantes de vitesses d'association et de dissociation permettant de passer d'une forme à l'autre, [P][L] et [PL] les concentrations de protéine et ligand libre et de complexe.

Le k_{on} et le k_{off} sont rapides du fait des mécanismes d'interaction de type *y-casting* et des interactions de faible affinité respectivement. Il est alors possible de réaliser un titrage permettant de cartographier la zone d'interaction.

Au sein d'une molécule, la valeur de la fréquence de résonance d'un noyau donné dépend de son environnement chimique, de la structure locale et de la nature des noyaux à proximité mais également des effets à longue portée dû au repliement tri-dimensionnel. La perturbation des déplacements chimiques (ou CSP, *chemical shifts perturbation*) est une méthode expérimentale relativement simple permettant de cartographier un site d'interaction sur une protéine qu'elle soit dépliée ou non [Zuiderweg 2002, Clarkson 2003, Williamson 2013].

En pratique, la protéine généralement marquée à l'azote 15 et/ou au carbone 13 est titrée par le ligand non marqué et inversement. Des spectres 2D HSQC sont enregistrés à chaque ajout de ligand. Des échantillons de protéines marqués ^{15}N et/ou ^{13}C sont nécessaire pour mettre en oeuvre l'expérience. L'analyse consiste en la mesure précise des déplacement chimiques à chaque point du titrage des noyaux ^{1}H, ^{15}N et ^{13}C (Figure 4.16).

Les perturbations des déplacements chimiques sont calculées avec l'équation :

$$_{\mathrm{max}} = \sqrt{(\quad {}^{1}\mathrm{H})^2 + \mathrm{k}(\quad {}^{15}\mathrm{N},{}^{13}\mathrm{C})^2} \tag{4.7}$$

où ^{1}H correspond à la différence de déplacement chimique en proton, ^{15}N,^{13}C à la différence de déplacement chimique en azote ou carbone, selon l'expérience RMN utilisée, entre la forme libre et liée d'une protéine et k est un facteur de proportionnalité permettant de donner le même poids à la contribution proton et azote ou carbone [Stockman 2002].

Constantes de couplage

Le couplage scalaire noté J correspond à une interaction entre les spins au travers des liaisons chimiques. Le couplage se traduit par l'apparition de raies multiples sur le spectre proton à une dimension d'une molécule simple.

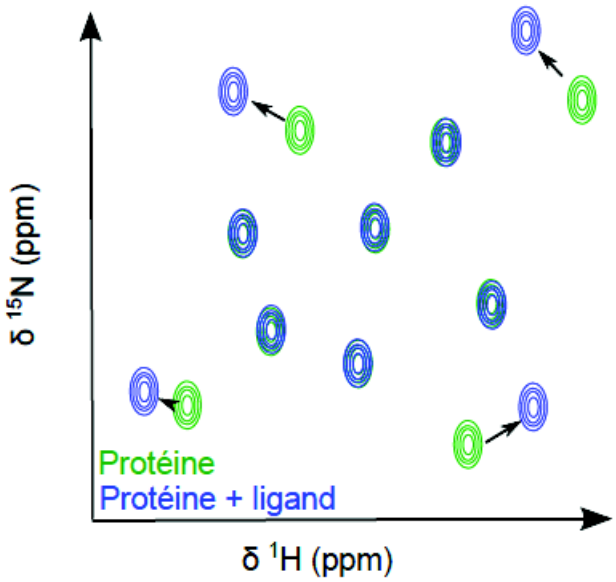


Figure 4.16 – Interaction d'une protéine avec son ligand. La protéine uniformément marquée ^{15}N est titrée par le ligand non marqué. La superposition des spectres HSQC au cours du titrage (protéine libre en vert, protéine + ligand en bleu) permet de suivre les variations des déplacements chimiques au cours de la formation du complexe protéine-ligand, et ainsi de cartographier la surface d'interaction.

Par exemple le couplage scalaire de deux noyaux A et X de spins $\frac{1}{2}$ se traduit par un doublet dont la séparation entre les pics correspond à la constante de couplage J_{AX}. La constante de couplage se note $^{n}J_{ij}$ où n représente le nombre de liaisons chimiques séparant les noyaux i et j. Le couplage scalaire peut être détecté jusqu'à une distance de 3 à 4 liaisons. Les constantes ^{3}J dépendent des angles de torsion selon l'équation de Karplus [Karplus 1959]. La valeur des angles dihèdres γ et dépend de la conformation du squelette peptidique et de fait il est possible d'en déduire la présence de structures secondaires. Plusieurs stratégies ont été proposées pour mesurer les constantes J, et en pratique la mesure de $^{3}J_{H_N,H_\alpha}$ permet de décrire l'échantillonnage des conformations d'une chaîne polypeptidique en particulier pour les IDPs [Vuister 1993, Otten 2009]. Les résultats expérimentaux sont alors comparés aux résultats obtenus pour des peptides désordonnés [Smith 1996]. Une différence positive indique la présence de structure secondaire de type feuillets- alors qu'une différence négative indique la présence de structure secondaire de type hélice- ou hélice polyproline de type II. Cependant, l'écart par rapport aux valeurs des J mesurés pour les peptides désordonnés est en général très faible et il est donc pratiquement difficile de distinguer la présence de structures secondaires

résiduelles uniquement à partir des valeurs de J pour une IDP. La combinaison des résultats observés pour les mesures des constantes de couplages scalaires et des fréquences de résonances de chaque résidu permet alors de renforcer l'interprétation des résultats.

Le couplage dipolaire est une interaction entre différents noyaux à travers l'espace. Cette interaction est en principe moyennée à 0 du fait de la réorientation rapide (échelle de temps de la nanoseconde) des molécules en solution. Par contre, la modulation de cette interaction sous l'effet des mouvements moléculaires est à l'origine des phénomènes de relaxation. En particulier, cette interaction est à l'origine de l'effet Overhauser Nucléaire (NOE) [Overhauser 1953]. La quantité d'aimantation transférée dépend de la distance entre deux noyaux, A et X, en $1/r^6$ ainsi que de la dynamique du système. Les distances inter-atomiques, jusqu'à 6 Å, sont obtenues à partir des expériences NOESY (Nuclear Overhauser Effect SpectroscopY). En 2002, Kurt Wüthrich à obtenu le prix nobel de chimie pour ses travaux démontrant l'intérêt de l'utilisation des expériences NOESY pour déterminer la structure 3D des protéines en solution [Wüthrich 1982].

Il est possible de perturber l'orientation des molécules en utilisant des milieux anisotropes (gels, cristal liquide, phages) [Tjandra 1997]. Dans ce cas, le couplage dipolaire entre deux noyaux n'est plus moyenné à 0 et sa valeur résiduelle (RDC, residual dipolar coupling) (10^{-4} fois la valeur de l'interaction) renseigne sur l'orientation du vecteur liant les deux noyaux en interaction par rapport au champ magnétique [Tjandra 1997]. Les RDCs sont de plus longue portée que l'effet NOE et peuvent être utilisés pour déterminer la conformation moléculaire via la mesure des distances intra-moléculaires. Pour les protéines repliées les RDCs servent à l'affinement de la structure calculée par RMN [Bax 2003] alors que dans le cas des IDPs ces mesures informent sur la présence de structures secondaires résiduelles et de contacts à longue distance (Figure 4.17) [Bernadó 2005a, Mukrasch 2007, Jensen 2009].

Relaxation

Lorsqu'un échantillon est placé dans un spectromètre RMN, l'aimantation macroscopique s'aligne le long du champ magnétique de l'appareil (axe z) et le système est à l'équilibre. L'application d'une impulsion radiofréquence de quelques s perturbe l'équilibre et bascule l'aimantation dans le plan transverse (axes x, y). La relaxation est caractérisée par le retour à l'état initial de l'aimantation. Ce phénomène est à la base de toute mesure RMN où l'on observe le mouvement de précession libre de l'aimantation ou FID pour Free Induction Decay. L'efficacité du processus de relaxation, c'est à dire la vitesse à laquelle l'aimantation retourne à sa position d'équilibre, dépend de la dynamique des molécules en solutions.

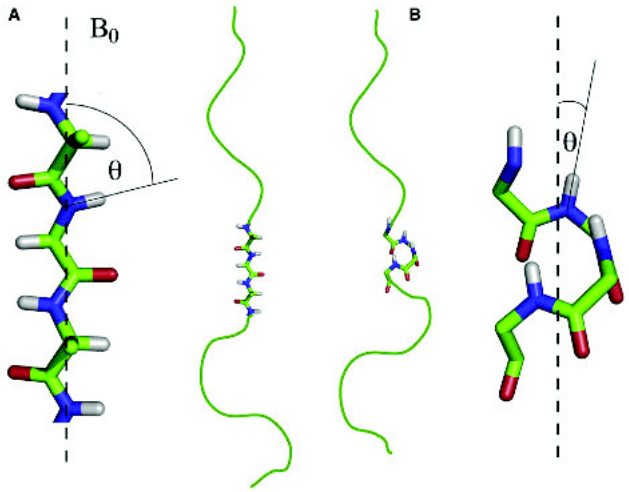


Figure 4.17 – Les RDCs diffèrent selon le niveau de structuration au sein d'une protéine. En effet, le vecteur internucléaire NH est orienté perpendiculairement pour une IDP (A) alors qu'il est quasiment aligné lors d'une structure en hélice (B) par rapport au tenseur d'alignement de la chaîne polypeptidique (aligné sur le champ magnétique B_0) dans un milieux orientant lors de la mesure de RDCs. Figure extraite de [Jensen 2009]

Les interactions dipolaires entre les spins proches dans l'espace et l'anisotropie du déplacement chimique dépendant du champ magnétique appliqué interviennent dans les processus de relaxation. Cela mène à deux vitesses de relaxation distinctes :

- la vitesse relaxation transverse ou R2 ($R2 = 1\ T2$), qui décrit la perte de la composante transverse de la relaxation et donc du signal RMN. Le temps de relaxation T2 est caractéristique de l'interaction entre les noyaux voisins au sein d'une molécule, est aussi appelé temps de relaxation spin-spin.
- la vitesse de relaxation longitudinale ou R1 ($R1 = 1\ T1$), qui décrit le retour à l'équilibre d'une population de spin après une perturbation, est caractérisée par le temps de relaxation T1. Ce temps, caractéristique de l'interaction des molécules avec l'environnement, est aussi appelé temps de relaxation longitudinale spin-réseau.

La relaxation est un paramètre important en RMN puisqu'elle permet de déterminer le délai entre les impulsions lors d'une expérience et détermine également la largeur de raie du signal observé. En effet, la largeur de raie est proportionnelle à la vitesse de relaxation R2. Ainsi, les molécules de haut poids moléculaire sont caractérisées par un élargissement des raies de résonance dû à leur temps de corrélation plus lent qu'une petite molécule qui présenterait des raies fines. Néanmoins, pour un même poids moléculaire, les largeurs de raies sont plus fines dans le cas des protéines désordonnées comparé aux protéines repliées du fait de la flexibilité de

la chaîne latérale qui est plus rapide que la rotation globale. Les expériences de relaxation ^{15}N permettent de séparer la dynamique du squelette peptidique d'une protéine de l'ordre de la ps de la dynamique globale de la protéine dans sa globalité de l'orde de la ns [Mittermaier 2009].

Mesure de la di usion transitionnelle

Des expériences de diffusion sont utiles pour observer le comportement hydrodynamique d'une protéine. La diffusion translationnelle est reliée au mouvement *Brownien* et peut être évaluée par la mesure du coefficient de diffusion translationnel D_T ou du rayon hydrodynamique R_h , inversement proportionnels, d'après la relation de Stockes-Einstein :

$$\mathrm{R_h} = \frac{\mathrm{kT}}{6 \quad \mathrm{D_T}} \tag{4.8}$$

où k est la constante de Boltzmann, T la température absolue et la viscosité. Cependant, cette équation est valable uniquement pour un modèle sphérique. Pour une IDP, on peut donc s'attendre à un D_T relativement grand comparé à celui d'une protéine globulaire de même poids moléculaire du fait d'un rayon hydrodynamique apparent plus important. D'autre part, si l'IDP présente des structures résiduelles ou un degré de compaction la valeur de D_T sera intermédiaire à celle d'une IDP pure et celle d'une protéine globulaire.

Il est possible de mesurer le coefficient de diffusion translationnel (D_T) d'une molécule en solution par RMN en appliquant un gradient de champ pulsé (PFG, Pulse Field Gradient) le long de l'axe de l'échantillon (axe z). La séquence d'impulsions de base a été décrite par Stejskal et Tanner en 1965 et consiste en une série d'échos de spin enregistrés pour différents gradients (PFG-SE) [Stejskal 1965]. Un premier gradient est appliqué pendant une durée après une première impulsion de 90 °. Cela introduit un déphasage de l'aimantation qui va dépendre de la position z d'un noyau donné dans l'échantillon. Ceci est appelé le codage spatial des spins. Ensuite, un second gradient, identique au précédent, nommé gradient de décodage est appliqué après une impulsion de 180 ° qui inverse l'aimantation. Si les molécules auxquelles appartiennent les spins sont stationnaires pendant le délai de diffusion séparant les deux gradients, alors le codage et le décodage se compensent et l'aimantation est entièrement refocalisée. L'intensité du signal mesuré est alors maximale. Toutefois, les molécules en solutions diffusent pendant le délai et la positions des spins le long de l'axe du gradient de champ appliqué (axe z) est modifiée, engendrant un décalage des phases. Ce décalage dégrade la refocalisation de l'aimantation totale et le signal mesuré décroît selon la vitesse de diffusion des molécules en suivant une loi exponentielle :

$$\mathrm{I} = \mathrm{I_0 e^{-D_T \ \ ^2 g^2 \ \ ^2 (\ \ - \frac{}{3})}} \tag{4.9}$$

où D_T est le coefficient de diffusion, le rapport gyromagnétique, g et sont respectivement l'intensité et la durée du gradient de champ pulsé et correspond au délai entre les deux gradients.

Environ 30 ans plus tard Morris et Johnson [Morris 1992] proposent d'utiliser la diffusion comme un axe spectroscopique dans l'expérience qu'ils nomment DOSY (Diffusion Ordered SpectroscopY). Cette élégante méthode permet de séparer les molécules de différentes tailles au sein d'un mélange et est souvent comparée à la technique de chromatographie . L'acquisition d'une DOSY consiste à accumuler une série de spectre 1D PFG-SE en gardant le délai constant. Il en résulte un spectre à deux dimensions, l'une correspondant au D_T (axe vertical) et l'autre correspondant aux déplacements chimique des molécules étudiées (axe horizontal). Dans cette représentation toutes les résonances correspondant à la même espèce sont, en principe, alignées.

Depuis, de nombreuses améliorations ont été proposées notamment l'expérience p-DOSY (permutated DOSY), développée au laboratoire cette année, qui permet de mesurer le coefficient de diffusion transitionnel pour des systèmes hors équilibre sans qu'il n'y ait aucun biais par l'application d'un gradient aléatoire le long de l'axe de l'échantillon [Oikonomou 2015].

La RMN DOSY est largement utilisée dans divers domaines tels que l'industrie agro-alimentaire ou pharmaceutique pour l'analyse de mélanges complexes [Gostan 2004, Balayssac 2009] mais également pour la détermination de la masse moléculaire [Augé 2009]. Il est également possible d'étudier les interactions moléculaires [Cohen 2012] et de déterminer l'indice de polydispersité des polymères [Viéville 2011].

Stratégie d'attribution des IDPs

L'attribution des résonances est un pré-requis pour l'étude de la dynamique et des structures résiduelles des IDPs par RMN. La faible dispersion des fréquences de résonances protons rend l'attribution plus compliquée que pour une protéine bien repliée. Cependant, les noyaux ^{15}N, qui présentent des raies de résonances très fines de part leur dynamique rapide (T2 faible), et les noyaux CO présentent la meilleure dispersion spectrale pour les IDPs. De nombreuses expériences RMN ont été développées et proposées pour l'étude des IDPs.

Les IDPs étant très exposées au solvant, l'échange des $^{1}H_N$ avec les protons du solvant peut entraîner une diminution de l'intensité du signal. Des expériences facilitant l'attribution des fréquences de résonances des IDPs et basées sur la détection des H_α (alors plus avantageuse que la détection H_N) ont été développées [Mäntylahti 2011, Lopez 2013]. Des avantages similaires sont obtenus en réalisant une détection des noyaux ^{13}C à condition de disposer de sonde adaptée. Du fait d'une relaxation plus lente de ces noyaux par rapport au proton amide, les

raies sont plus fines et les informations sur le squelette peptidique et chaînes latérales de chaque résidus y compris les prolines peuvent être obtenus à partir des CO, C_{α}, $C_{aliphatique}$ et $C_{aromatique}$ [Bermel 2006, Csizmok 2008]. De plus, les noyaux ^{13}C sont plus sensibles à la présence de structures secondaires et présentent donc une meilleure dispersion spectrale. Il est possible de substituer la carte de corrélation CO-N à $H_N - N$ grâce aux expérience de type CON dans lesquelles la fréquence de résonance de noyaux ^{13}C est corrélée avec la fréquence de résonance de l'amide de chaque résidu et de faciliter l'attribution. C'est par exemple le cas de l'attribution de l' -synucléine réalisée grâce à l'acquisition d'uniquement deux expériences basées sur la détection du carbonyle, CO, comparé à sept expériences en détection $^{1}H_N$ [Bermel 2006].

Aussi, il est nécessaire d'augmenter la résolution spectrale pour pallier à la faible dispersion des déplacements chimiques. Dans ce sens, l'échantillonnage non uniforme (NUS, non uniform sampling) a permis l'augmentation du nombre de dimension (jusqu'à 7) pour gagner en résolution sans augmenter dramatiquement le temps d'acquisition d'une expérience. Cela a mené au développement de protocole d'attribution automatisé permettant l'attribution d'IDPs de haut poids moléculaire et présentant des répétitions d'acides aminés comme le très bel exemple de la protéine Tau faisant 441 résidus [Hiller 2007, Narayanan 2010].

Modélisation : génération d'ensembles et conformères

La description atomique d'une protéine ou d'un complexe protéique est l'objectif principal de la biologie structurale. En effet, ces informations donnent accès aux mécanismes moléculaires impliqués dans le fonctionnement de la protéine ou du complexe impliquant la protéine et peuvent considérablement aider à la mise au point de nouvelles molécules thérapeutiques pour traiter certaines pathologies.

Du fait de leur flexibilité les IDPs présentent de nombreux degrés de liberté à l'origine des nombreuses conformations interchangeables en solution. Des outils bio-informatique permettant la description par ensemble ont été développé, pour obtenir des informations concernant la relation structure/fonction particulière à ces protéines, et intégrant les grandes quantités des données résultant d'expérience de SAXS ou RMN.

Ainsi, les outils supposent que les données expérimentales peuvent être représentées par un nombre moyen de conformères ou ensembles. La procédure se fait en deux étapes qui sont la pré-génération d'ensembles conformationnels puis la sélection d'un sous-groupe d'ensemble qui décrit au mieux les données expérimentales. Un tel processus peut-être totalement randomisé ou réalisé sous contraintes géométriques comme par exemple les contraintes d'angles dihèdres γ et ou la présence de structures secondaires.

L'algorithme *Flexible-meccano* (FM) fournit des ensembles conformationnels représentant

toutes les possibles conformations que peut adopter une IDP [Ozenne 2012] ce qui constitue la première étape. La stratégie de cet algorithme est randomisée et échantillonne l'espace conformationnel en créant un modèle statistique de la protéine dépliée en se basant sur le comportement conformationnel de chaque acide aminés à partir des distributions des angles dihèdres trouvés dans les boucles des structures de protéines.

L'étape suivante consiste à choisir l'ensemble conformationnel pertinent et est nommée back-calculation . Comme son nom l'indique, il s'agit de calculer des paramètres physico-chimiques pour toutes les conformations obtenues avec FM et de comparer les moyennes de ces paramètres aux données expérimentales et ainsi de sélectionner les sous-ensembles de conformères pour lesquels les moyennes s'en rapprochent le plus. L'algorihme ASTEROIDS (A Selection Tool for Ensemble Representation Of Intrinsically Disordered States), parmi d'autres algorithmes disponibles, est un exemple qui permet d'utiliser différents types de données pour réaliser cette sélection [Nodet 2009, Salmon 2010]. La combinaison de données résultants de plusieurs méthodes différentes telles que le SAXS (R_g, D_{max}) et les nombreux paramètres RMN (déplacement chimique, D_T, RCD) cités plus hauts permet d'appréhender au mieux la description du comportement d'une IDP en solution [Bernadó 2005b]. Cette démarche est illustrée dans l'étude de la protéine tau et de l' -synucléine en solution où l'utilisation des données RMN telles que les mesures de RCD, déplacements chimiques et les données SAXS combinées avec une approche FM/ASTEROIDS a permit la caractérisation de la structure locale et à longue distance de ces deux IDPs [Schwalbe 2014].

La plus grande difficulté est de choisir l'ensemble conformationnel qui décrit le mieux le système flexible que l'on étudie. Enfin, la qualité de la modélisation dépend de la qualité des données expérimentales.

4.5 Conclusion

La découverte des IDPs a boulversé le domaine de la biologie structurale. Pendant longtemps le concept de la relation structure fonction a été la référence de la communauté scientifique.

Il est maintenant admis que pour certaines protéines l'état totalement ou partiellement désordonné soit l'état natif et que de telles protéines puissent exercer plusieurs fonctions biologiques. Le désordre inhérent à ces protéines leur confère une certaine flexibilité qui est alors un atout fonctionnel. De nombreuses IDPs présentent des structures secondaires résiduelles et nécessitent l'interaction avec un partenaire pour se replier. Ce processus est nommé repliement induit et confère de nombreux avantages tel qu'une interaction hautement spécifique sans

nécessiter une forte affinité pour le partenaire. Une IDP présente un ensemble de conformations interconvertibles permettant de fournir de grandes surfaces d'interaction pour différents partenaires. On parle alors de plasticité structurale. De part leur propriétés, ces protéines ont un rôle central dans les processus biologiques tels que les mécanismes de régulation et de signalisation mais également dans les pathologies. Ces protéines sont donc d'importantes cibles pour la recherche et le développement de nouvelles thérapeutiques.

A l'heure actuelle il n'existe pas de méthode unique permettant d'étudier ces protéines, mais il est nécessaire de combiner plusieurs technologies. Les études bio-informatiques sont essentielles pour repérer des séquences conservées dans l'évolution et des sites potentiels d'interaction. Le dichroïsme circulaire et la spectroscopie de fluorescence permettent de sonder la présence de structure secondaires et tertiaires au sein d'une IDP. Le SAXS et les nombreux paramètres qui peuvent être mesurés par RMN permettent les études structurales des ces protéines flexibles avec une résolution atomique.

Bibliographie

[Anfinsen 1972] Anfinsen C.B. *Studies on the principles that govern the folding of protein chains*. Science, 1972. – Cité page 50.

[Arkin 2005] Arkin M. *Protein protein interactions and cancer : small molecules going in for the kill*. Current Opinion in Chemical Biology, vol. 9, no. 3, 2005. – Cité page 63.

[Augé 2009] Augé S., Schmit P.O., Crutchfield C.A., Islam M.T., Harris D.J., Durand E., Clemancey M., Quoineaud A.A., Lancelin J.M., Prigent Y., Taulelle F. et Delsuc M.A. *NMR Measure of Translational Di usion and Fractal Dimension. Application to Molecular Mass Measurement*. The Journal of Physical Chemistry B, vol. 113, no. 7, 2009. – Cité pages 84 et 155.

[Bah 2014] Bah A., Vernon R.M., Siddiqui Z., Krzeminski M., Muhandiram R., Zhao C., Sonenberg N., Kay L.E. et Forman-Kay J.D. *Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch*. Nature, 2014. – Cité page 62.

[Balayssac 2009] Balayssac S., Trefi S., Gilard V., Malet-Martino M., Martino R. et Delsuc M.A. *2D and 3D DOSY 1H NMR, a useful tool for analysis of complex mixtures : Application to herbal drugs or dietary supplements for erectile dysfunction*. Journal of Pharmaceutical and Biomedical Analysis, vol. 50, no. 4, 2009. – Cité page 84.

[Balla 2006] Balla S., Thapar V., Verma S., Luong T., Faghri T., Huang C.H., Rajasekaran S., del Campo J.J., Shinn J.H., Mohler W.A., Maciejewski M.W., Gryk M.R., Piccirillo B., Schiller S.R. et Schiller M.R. *Minimotif Miner : a tool for investigating protein function*. Nature Methods, vol. 3, no. 3, 2006. – Cité page 66.

[Bax 2003] Bax A. *Weak alignment o ers new NMR opportunities to study protein structure and dynamics*. Protein science : a publication of the Protein Society, vol. 12, no. 1, 2003. – Cité page 81.

[Bennick 2002] Bennick A. *Interaction of plant polyphenols with salivary proteins*. Critical Reviews in Oral Biology & Medicine, vol. 13, no. 2, 2002. – Cité page 63.

[Bermel 2006] Bermel W., Bertini I., Felli I.C., Lee Y.M., Luchinat C. et Pierattelli R. *Protonless NMR Experiments for Sequence-Speci c Assignment of Backbone Nuclei in Unfolded Proteins*. Journal of the American Chemical Society, vol. 128, no. 12, 2006. – Cité page 85.

[Bernadó 2005a] Bernadó P., Bertoncini C.W., Griesinger C., Zweckstetter M. et Blackledge M. *De ning Long-Range Order and Local Disorder in Native -Synuclein Using Residual*

Dipolar Couplings. Journal of the American Chemical Society, vol. 127, no. 51, 2005. – Cité page 81.

[Bernadó 2005b] Bernadó P., Blanchard L., Timmins P., Marion D., Ruigrok R.W.H. et Blackledge M. *A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering*. Proceedings of the National Academy of Sciences, vol. 102, no. 47, 2005. – Cité page 86.

[Bernadó 2009] Bernadó P. et Blackledge M. *A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering*. Biophysical journal, vol. 97, no. 10, 2009. – Cité pages 72 et 168.

[Bhattacharyya 1999] Bhattacharyya J. et Das K.P. *Molecular chaperone-like properties of an unfolded protein, alpha(s)-casein.* Journal of Biological Chemistry, vol. 274, no. 22, 1999. – Cité page 62.

[Bonnet 2010] Bonnet J., Wang Y.H., Spedale G., Atkinson R.A., Romier C., Hamiche A., Pijnappel W.W.M.P., Timmers H.T.M., Tora L.a.s.o., Devys D. et Kieffer B. *The structural plasticity of SCA7 domains de nes their di erential nucleosome-binding properties*. EMBO reports, vol. 11, no. 8, 2010. – Cité page 63.

[Braun 1994] Braun D., Wider G. et Wuethrich K. *Sequence-corrected 15N" random coil" chemical shifts*. Journal of the American Chemical Society, vol. 116, no. 19, pages 8466–8469, 1994. – Cité page 76.

[Bryngelson 1995] Bryngelson J.D., Onuchic J.N. et Socci N.D. *Funnels, pathways, and the energy landscape of protein folding : a synthesis*. Proteins : Structure, 1995. – Cité page 54.

[Chaffotte 1997] Chaffotte A.F., Guijarro J.I., Guillou Y., Delepierre M. et Goldberg M.E. *The "pre-molten globule," a new intermediate in protein folding.* Journal of protein chemistry, vol. 16, no. 5, 1997. – Cité page 51.

[Cheng 2006] Cheng Y., LeGall T., Oldfield C.J., Mueller J.P., Van Y.Y.J., Romero P., Cortese M.S., Uversky V.N. et Dunker A.K. *Rational drug design via intrinsically disordered protein*. Trends in Biotechnology, vol. 24, no. 10, 2006. – Cité page 64.

[Cheng 2007] Cheng Y., Oldfield C.J., Meng J., Romero P., Uversky V.N. et Dunker A.K. *Mining -Helix-Forming Molecular Recognition Features with Cross Species Sequence Alignments* †. Biochemistry, vol. 46, no. 47, 2007. – Cité page 57.

[Clarkson 2003] Clarkson J. et Campbell I.D. *Studies of protein-ligand interactions by NMR.* Biochemical Society Transactions, vol. 31, no. Pt 5, 2003. – Cité page 79.

[Cohen 2012] Cohen Y., Avram L., Evan Salem T., Slovak S., Shemesh N. et Frish L. *Di usion NMR in Supramolecular Chemistry and Complexed Systems*. In Analytical Methods in Supramolecular Chemistry, pages 197–285. Wiley-VCH Verlag GmbH & Co. KGaA, 2012. – Cité page 84.

[Csermely 2010] Csermely P., Palotai R. et Nussinov R. *Induced t, conformational selection and independent dynamic segments : an extended view of binding events*. Trends in Biochemical Sciences, vol. 35, no. 10, 2010. – Cité page 58.

[Csizmok 2008] Csizmok V., Felli I.C., Tompa P., Banci L. et Bertini I. *Structural and Dynamic Characterization of Intrinsically Disordered Human Securin by NMR Spectroscopy*. Journal of the American Chemical Society, vol. 130, no. 50, 2008. – Cité page 85.

[Dill 2012] Dill K.A. et MacCallum J.L. *The protein-folding problem, 50 years on.* Science, vol. 338, no. 6110, 2012. – Cité page 55.

[Dinkel 2014] Dinkel H., Van Roey K., Michael S., Davey N.E., Weatheritt R.J., Born D., Speck T., Krüger D., Grebnev G., Kuban M., Strumillo M., Uyar B., Budd A., Altenberg B., Seiler M., Chemes L.B., Glavina J., Sánchez I.E., Diella F. et Gibson T.J. *The eukaryotic linear motif resource ELM : 10 years and counting.* Nucleic Acids Research, vol. 42, no. Database issue, 2014. – Cité pages 66 et 137.

[Dinner 2000] Dinner A.R., Sali A., Smith L.J., Dobson C.M. et Karplus M. *Understanding protein folding via free-energy surfaces from theory and experiment.* Trends in Biochemical Sciences, vol. 25, no. 7, 2000. – Cité page 54.

[Dosztányi 2005] Dosztányi Z., Csizmok V., Tompa P. et Simon I. *IUPred : web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, vol. 21, no. 16, 2005. – Cité pages 65 et 147.

[Dosztányi 2009] Dosztányi Z., Mészáros B. et Simon I. *ANCHOR : web server for predicting protein binding regions in disordered proteins.* Bioinformatics, vol. 25, no. 20, 2009. – Cité page 66.

[Dunker 2001] Dunker A.K., Lawson J.D., Brown C.J. et Williams R.M. *Intrinsically disordered protein*. Journal of Molecular Graphics and Modelling, vol. 19, no. 1, pages 26–59, 2001. – Cité pages 51 et 52.

[Dunker 2008] Dunker A.K., Oldfield C.J., Meng J., Romero P., Yang J.Y., Chen J., Vacic V., Obradovic Z. et Uversky V.N. *The unfoldomics decade : an update on intrinsically disordered proteins*. BMC Genomics, vol. 9, no. Suppl 2, pages S1–26, 2008. – Cité pages 52 et 53.

[Durand 2010] Durand D., Vivès C., Cannella D., Pérez J., Pebay-Peyroula E., Vachette P. et Fieschi F. *NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi- exible linkers*. Journal of Structural Biology, vol. 169, no. 1, 2010. – Cité pages 73 et 166.

[Dyson 2005] Dyson H.J. et Wright P.E. *Intrinsically unstructured proteins and their functions*. Nature Reviews Molecular Cell Biology, vol. 6, no. 3, 2005. – Cité pages 52 et 75.

[Eftink 1994] Eftink M.R. *The use of uorescence methods to monitor unfolding transitions in proteins.* Biophysical journal, vol. 66, no. 2 Pt 1, 1994. – Cité page 68.

[Fischer 1894] Fischer E. *Ein uss der Con guration auf die Wirkung der Enzyme*. Berichte der deutschen chemischen Gesellschaft, vol. 27, no. 3, 1894. – Cité page 50.

[Fuxreiter 2012a] Fuxreiter M. *Fuzziness : linking regulation to protein dynamics*. Molecular BioSystems., vol. 8, no. 1, pages 168–177, 2012. – Cité page 59.

[Fuxreiter 2012b] Fuxreiter M. et Tompa P. *Fuzzy complexes : a more stochastic view of protein function.* Advances in experimental medicine and biology, vol. 725, no. Chapter 1, pages 1–14, 2012. – Cité page 59.

[Gostan 2004] Gostan T., Tramesel D., Brun E., Prigent Y. et Delsuc M.A. *L experience Dosy, une puissante méthode RMN pour l analyse de mélanges complexes et la détection de traces*. Spectra analyse, 2004. – Cité page 84.

[Greenbaum 2001] Greenbaum N. et Ghose R. *Nuclear Magnetic Resonance (NMR) Spectroscopy : Structure Determination of Proteins and Nucleic Acids*. John Wiley & Sons, Ltd, 2001. – Cité page 75.

[Gro 1988] Gro K.H. et Kalbitzer H.R. *Distribution of chemical shifts in 1 H nuclear magnetic resonance spectra of proteins*. Journal of Magnetic Resonance (1969), 1988. – Cité page 76.

[Gunasekaran 2004] Gunasekaran K., Tsai C.J. et Nussinov R. *Analysis of Ordered and Disordered Protein Complexes Reveals Structural Features Discriminating Between Stable and Unstable Monomers*. Journal of Molecular Biology, vol. 341, no. 5, 2004. – Cité page 56.

[Hammoudeh 2009] Hammoudeh D.I., Follis A.V., Prochownik E.V. et Metallo S.J. *Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein c-Myc*. Journal of the American Chemical Society, vol. 131, no. 21, 2009. – Cité page 64.

[Hawe 2008] Hawe A., Sutter M. et Jiskoot W. *Extrinsic Fluorescent Dyes as Tools for Protein Characterization*. Pharmaceutical Research, vol. 25, no. 7, 2008. – Cité page 68.

[Hazy 2009] Hazy E. et Tompa P. *Limitations of Induced Folding in Molecular Recognition by Intrinsically Disordered Proteins*. ChemPhysChem, vol. 10, no. 9-10, 2009. – Cité pages 64 et 65.

[Hiller 2007] Hiller S., Wasmer C., Wider G. et Wüthrich K. *Sequence-speci c resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy.* Journal of the American Chemical Society, vol. 129, no. 35, 2007. – Cité page 85.

[Iakoucheva 2001] Iakoucheva L.M., Kimzey A.L., Masselon C.D., Smith R.D., Dunker A.K. et Ackerman E.J. *Aberrant mobility phenomena of the DNA repair protein XPA.* Protein science : a publication of the Protein Society, vol. 10, no. 7, 2001. – Cité page 66.

[Iakoucheva 2002] Iakoucheva L.M., Brown C.J., Lawson J.D., Obradovic Z. et Dunker A.K. *Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins*. Journal of Molecular Biology, vol. 323, no. 3, 2002. – Cité page 61.

[Ishida 2008] Ishida T. et Kinoshita K. *Prediction of disordered regions in proteins based on the meta approach*. Bioinformatics, vol. 24, no. 11, 2008. – Cité pages 65 et 132.

[Jensen 2009] Jensen M.R., Markwick P.R.L., Meier S., Griesinger C., Zweckstetter M., Grzesiek S., Bernadó P. et Blackledge M. *Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings*. Structure, vol. 17, no. 9, 2009. – Cité pages 81 et 82.

[Karplus 1959] Karplus M. *Contact Electron-Spin Coupling of Nuclear Magnetic Moments*. The Journal of Chemical Physics, vol. 30, no. 1, pages 11–6, 1959. – Cité page 80.

[Kelly 2000] Kelly S.M. et Price N.C. *The use of circular dichroism in the investigation of protein structure and function*. Current protein & peptide science, vol. 1, no. 4, pages 349–384, 2000. – Cité page 68.

[Kelly 2005] Kelly S.M., Jess T.J. et Price N.C. *How to study proteins by circular dichroism*. Biochimica et Biophysica Acta, vol. 1751, no. 2, 2005. – Cité page 67.

[Kieffer 2011] Kieffer B., Homans S. et Jahnke W. *Nuclear magnetic resonance of ligand binding to proteins*. Biophysical Approches Determining Ligand Binding to Biomolecular Targets : Detection, Measurement and Modelling., 2011. – Cité page 78.

[Koehler 2014] Koehler C., Bonnet J., Stierle M., Romier C., Devys D. et Kieffer B. *DNA Binding by Sgf11 Protein A ects Histone H2B Deubiquitination by Spt-Ada-Gcn5-Acetyltransferase (SAGA)*. Journal of Biological Chemistry, vol. 289, no. 13, 2014. – Cité page 63.

[Kohn 2004] Kohn J.E., Millett I.S., Jacob J., Zagrovic B., Dillon T.M., Cingel N., Dothager R.S., Seifert S., Thiyagarajan P., Sosnick T.R., Hasan M.Z., Pande V.S., Ruczinski I., Doniach S. et Plaxco K.W. *Random-coil behavior and the dimensions of chemically unfolded proteins*. Proceedings of the National Academy of Sciences, vol. 101, no. 34, pages 12491–12496, 2004. – Cité pages 72, 168 et 169.

[Korzhnev 2008] Korzhnev D.M. et Kay L.E. *Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy : An Application to Protein Folding*. Accounts of Chemical Research, vol. 41, no. 3, 2008. – Cité page 58.

[Kratky 1949] Kratky O. et Porod G. *Di use small-angle scattering of X-rays in colloid systems.* Journal of colloid science, vol. 4, no. 1, 1949. – Cité page 73.

[Kuwajima 1989] Kuwajima K. *The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure.* Proteins : Structure, Function, and Bioinformatics, vol. 6, no. 2, pages 87–103, 1989. – Cité page 50.

[Levy 2004] Levy Y., Wolynes P.G. et Onuchic J.N. *Protein topology determines binding mechanism.* Proceedings of the National Academy of Sciences, vol. 101, no. 2, 2004. – Cité page 58.

[Li 1999] Li X., Romero P., Rani M., Dunker A.K. et Obradovic Z. *Predicting protein disorder for N-, C-and internal regions*. Genome Informatics, 1999. – Cité page 65.

[Lieutaud 2008] Lieutaud P., Canard B. et Longhi S. *MeDor : a metaserver for predicting protein disorder*. BMC Genomics, vol. 9, no. Suppl 2, pages S25–5, 2008. – Cité pages 65, 66 et 133.

[Linding 2003a] Linding R. *GlobPlot : exploring protein sequences for globularity and disorder*. Nucleic Acids Research, vol. 31, no. 13, 2003. – Cité pages 65 et 147.

[Linding 2003b] Linding R., Jensen L.J., Diella F., Bork P., Gibson T.J. et Russell R.B. *Protein Disorder Prediction*. Structure, vol. 11, no. 11, 2003. – Cité pages 65 et 147.

[Lippens 2012] Lippens G., Amniai L., Wieruszeski J.M., Sillen A., Leroy A. et Landrieu I. *Towards understanding the phosphorylation code of tau*. Biochemical Society Transactions, vol. 40, no. 4, 2012. – Cité page 64.

[Lopez 2013] Lopez J., Ahuja P., Gerard M., Wieruszeski J.M. et Lippens G. *A new strategy for sequential assignment of intrinsically unstructured proteins based on 15N single isotope labelling*. Journal of Magnetic Resonance, vol. 236, 2013. – Cité page 84.

[Mäntylahti 2011] Mäntylahti S., Hellman M. et Permi P. *Extension of the HA-detection based approach : (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain*

assignment of intrinsically disordered proteins. Journal of biomolecular NMR, vol. 49, no. 2, 2011. – Cité page 84.

[Marsh 2006] Marsh J.A., Singh V.K., Jia Z. et Forman-Kay J.D. *Sensitivity of secondary structure propensities to sequence di erences between alpha- and gamma-synuclein : implications for brillation.* Protein science : a publication of the Protein Society, vol. 15, no. 12, 2006. – Cité pages 77 et 203.

[Merutka 1995] Merutka G., Dyson H.J. et Wright P.E. *Random coil 1H chemical shifts obtained as a function of temperature and tri uoroethanol concentration for the peptide series GGXGG.* Journal of biomolecular NMR, vol. 5, no. 1, 1995. – Cité page 76.

[Mészáros 2007] Mészáros B., Tompa P., Simon I. et Dosztányi Z. *Molecular Principles of the Interactions of Disordered Proteins*. Journal of Molecular Biology, vol. 372, no. 2, 2007. – Cité page 56.

[Metallo 2010] Metallo S.J. *Intrinsically disordered proteins are potential drug targets*. Current Opinion in Chemical Biology, vol. 14, no. 4, 2010. – Cité page 64.

[Mitrea 2012] Mitrea D.M., Yoon M.K., Ou L. et Kriwacki R.W. *Disorder-function relationships for the cell cycle regulatory proteins p21 and p27*. Biological Chemistry, vol. 393, no. 4, 2012. – Cité page 62.

[Mittag 2010] Mittag T., Marsh J., Grishaev A., Orlicky S., Lin H., Sicheri F., Tyers M. et Forman-Kay J.D. *Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase*. Structure, vol. 18, no. 4, 2010. – Cité page 60.

[Mittermaier 2009] Mittermaier A.K. et Kay L.E. *Observing biological dynamics at atomic resolution using NMR*. Trends in Biochemical Sciences, vol. 34, no. 12, 2009. – Cité page 83.

[Morris 1992] Morris K.F. et Johnson Jr C.S. *Di usion-ordered two-dimensional nuclear magnetic resonance spectroscopy*. Journal of the American Chemical Society, vol. 114, no. 8, pages 3139–3141, 1992. – Cité page 84.

[Mukhopadhyay 2001] Mukhopadhyay R. et Hoh J.H. *AFM force measurements on microtubule-associated proteins : the projection domain exerts a long-range repulsive force.* FEBS Letters, vol. 505, no. 3, 2001. – Cité page 62.

[Mukrasch 2007] Mukrasch M.D., Markwick P., Biernat J., von Bergen M., Bernadó P., Griesinger C., Mandelkow E., Zweckstetter M. et Blackledge M. *Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identi ed from Residual Dipolar Cou-*

plings and Molecular Simulation. Journal of the American Chemical Society, vol. 129, no. 16, 2007. – Cité page 81.

[Narayanan 2010] Narayanan R.L., Dürr U.H.N., Bibow S., Biernat J., Mandelkow E. et Zweckstetter M. *Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues*. Journal of the American Chemical Society, vol. 132, no. 34, 2010. – Cité page 85.

[Nodet 2009] Nodet G., Salmon L., Ozenne V., Meier S., Jensen M.R. et Blackledge M. *Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings*. Journal of the American Chemical Society, vol. 131, no. 49, 2009. – Cité page 86.

[Oikonomou 2015] Oikonomou M., Asencio-Hernández J., Velders A.H. et Delsuc M.A. *Accurate DOSY measure for out-of-equilibrium systems using permutated DOSY (p-DOSY)*. Journal of Magnetic Resonance, vol. 258, no. C, 2015. – Cité pages 84 et 155.

[Olashaw 2004] Olashaw N., Bagui T.K. et Pledger W.J. *Cell cycle control : a complex issue*. Cell Cycle, vol. 3, no. 3, pages 261–262, 2004. – Cité page 62.

[Oldfield 2005] Oldfield C.J., Cheng Y., Cortese M.S., Brown C.J., Uversky V.N. et Dunker A.K. *Comparing and Combining Predictors of Mostly Disordered Proteins* . Biochemistry, vol. 44, no. 6, 2005. – Cité page 60.

[Oldfield 2008] Oldfield C.J., Meng J., Yang J.Y., Yang M.Q., Uversky V.N. et Dunker A.K. *Flexible nets : disorder and induced t in the associations of p53 and 14-3-3 with their partners*. BMC Genomics, vol. 9, no. Suppl 1, pages S1–20, 2008. – Cité pages 57 et 58.

[Otten 2009] Otten R., Wood K. et Mulder F.A.A. *Comprehensive determination of 3JHNH for unfolded proteins using 13C-resolved spin-echo di erence spectroscopy*. Journal of biomolecular NMR, vol. 45, no. 4, 2009. – Cité page 80.

[Overhauser 1953] Overhauser A.W. *Polarization of nuclei in metals*. Physical Review, vol. 92, no. 2, pages 411–415, 1953. – Cité page 81.

[Ozenne 2012] Ozenne V., Bauer F., Salmon L., Huang J.r., Jensen M.R., Segard S., Bernado P., Charavay C. et Blackledge M. *Flexible-meccano : a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables*. Bioinformatics, vol. 28, no. 11, 2012. – Cité page 86.

[Prilusky 2005] Prilusky J., Felder C.E., Zeev-Ben-Mordehai T., Rydberg E.H., Man O., Beckmann J.S., Silman I. et Sussman J.L. *FoldIndex : a simple tool to predict whether a*

given protein sequence is intrinsically unfolded. Bioinformatics, vol. 21, no. 16, 2005. – Cité pages 65 et 147.

[Provencher 1981] Provencher S.W. et Glöckner J. *Estimation of globular protein secondary structure from circular dichroism.* Biochemistry, vol. 20, no. 1, 1981. – Cité page 67.

[Qi 2015] Qi H., Cantrelle F.X., Benhelli-Mokrani H., Smet-Nocca C., Buée L., Lippens G., Bonnefoy E., Galas M.C. et Landrieu I. *Nuclear Magnetic Resonance Spectroscopy Characterization of Interaction of Tau with DNA and Its Regulation by Phosphorylation.* Biochemistry, vol. 54, no. 7, 2015. – Cité page 64.

[Radhakrishnan 1997] Radhakrishnan I., Pérez-Alvarado G.C., Dyson H.J., Wright P.E., David P. et R M.M. *Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB : a Model for Activator :coactivator Interactions.* Cell, vol. 91, 1997. – Cité page 58.

[Radhakrishnan 1998] Radhakrishnan I., Pérez-Alvarado G.C., Dyson H.J. et Wright P.E. *Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB.* FEBS Letters, vol. 430, no. 3, 1998. – Cité page 58.

[Receveur-Brechot 2012] Receveur-Brechot V. et Durand D. *How random are intrinsically disordered proteins ? A small angle scattering perspective.* Current protein & peptide science, vol. 13, no. 1, 2012. – Cité pages 71, 73 et 166.

[Salmon 2010] Salmon L., Nodet G., Ozenne V., Yin G., Jensen M.R., Zweckstetter M. et Blackledge M. *NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins.* Journal of the American Chemical Society, vol. 132, no. 24, 2010. – Cité page 86.

[Schwalbe 2014] Schwalbe M., Ozenne V., Bibow S., Jaremko M., Jaremko L., Gajda M., Jensen M.R., Biernat J., Becker S., Mandelkow E., Zweckstetter M. et Blackledge M. *Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and &alpha ;-Synuclein in Solution from NMR and Small Angle Scattering.* Structure, vol. 22, no. 2, 2014. – Cité page 86.

[Schweers 1994] Schweers O., Schönbrunn-Hanebeck E., Marx A. et Mandelkow E. *Structural studies of tau protein and Alzheimer paired helical laments show no evidence for beta-structure.* Journal of Biological Chemistry, vol. 269, no. 39, 1994. – Cité page 51.

[Semisotnov 1991] Semisotnov G.V., Rodionova N.A., Razgulyaev O.I., Uversky V.N., Gripas A.F. et Gilmanshin R.I. *Study of the "molten globule" intermediate state in protein*

folding by a hydrophobic uorescent probe. Biopolymers, vol. 31, no. 1, 1991. – Cité page 68.

[Shoemaker 2000] Shoemaker B.A., Portman J.J. et Wolynes P.G. *Speeding molecular recognition by using the folding funnel : the y-casting mechanism.* Proceedings of the National Academy of Sciences, vol. 97, no. 16, 2000. – Cité page 58.

[Smith 1996] Smith L.J., Bolin K.A., Schwalbe H., MacArthur M.W., Thornton J.M. et Dobson C.M. *Analysis of main chain torsion angles in proteins : prediction of NMR coupling constants for native and random coil conformations.* Journal of Molecular Biology, vol. 255, no. 3, 1996. – Cité page 80.

[Sottejeau 2015] Sottejeau Y., Bretteville A., Cantrelle F.X., Malmanche N., Demiaute F., Mendes T., Delay C., Dos Alves H.A., Flaig A., Davies P., Dourlen P., Dermaut B., Laporte J., Amouyel P., Lippens G., Chapuis J., Landrieu I. et Lambert J.C. *Tau phosphorylation regulates the interaction between BIN1 s SH3 domain and Tau s proline-rich domain.* Acta Neuropathologica Communications, 2015. – Cité page 64.

[Stejskal 1965] Stejskal E.O. et Tanner J.E. *Spin Di usion Measurements : Spin Echoes in the Presence of a Time-Dependent Field Gradient.* The Journal of Chemical Physics, vol. 42, no. 1, pages 288–6, 1965. – Cité page 83.

[Stockman 2002] Stockman B.J. et Dalvit C. NMR screening techniques in drug discovery and drug design. Progress in Nuclear Magnetic Resonance Spectroscopy, 2002. – Cité page 79.

[Sugase 2007] Sugase K., Dyson H.J. et Wright P.E. *Mechanism of coupled folding and binding of an intrinsically disordered protein.* Nature, vol. 447, no. 7147, 2007. – Cité page 59.

[Svergun 1992] Svergun D.I. *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* Journal of applied crystallography, vol. 25, no. 4, 1992. – Cité pages 74, 165 et 174.

[Svergun 2002] Svergun D.I. et Koch M. *Advances in structure analysis using small-angle scattering in solution.* Current Opinion in Structural Biology, vol. 12, no. 5, pages 654–660, 2002. – Cité page 71.

[SYu 1994] SYu V. et Vassilenko K.S. *Determination of protein tertiary structure class from circular dichroism spectra.* Analytical biochemistry, vol. 222, no. 1, 1994. – Cité page 68.

[Teilum 2009] Teilum K., Olsen J.G. et Kragelund B.B. *Functional aspects of protein exibility.* Cellular and Molecular Life Sciences, vol. 66, no. 14, 2009. – Cité pages 53 et 54.

[Thanabal 1994] Thanabal V., Omecinsky D.O., Reily M.D. et Cody W.L. *The 13C chemical shifts of amino acids in aqueous solution containing organic solvents : application to the secondary structure characterization of peptides in aqueous tri uoroethanol solution.* Journal of biomolecular NMR, vol. 4, no. 1, 1994. – Cité page 76.

[Tjandra 1997] Tjandra N. et Bax A. *Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium.* Science, vol. 278, no. 5340, pages 1111–1114, 1997. – Cité page 81.

[Tompa 2004] Tompa P. *The role of structural disorder in the function of RNA and protein chaperones.* The FASEB Journal, vol. 18, no. 11, 2004. – Cité page 62.

[Tompa 2005a] Tompa P. *The interplay between structure and function in intrinsically unstructured proteins.* FEBS Letters, vol. 579, no. 15, 2005. – Cité page 63.

[Tompa 2005b] Tompa P., Szász C. et Buday L. *Structural disorder throws new light on moonlighting.* Trends in Biochemical Sciences, vol. 30, no. 9, 2005. – Cité page 63.

[Tompa 2008] Tompa P. et Fuxreiter M. *Fuzzy complexes : polymorphism and structural disorder in protein protein interactions.* Trends in Biochemical Sciences, vol. 33, no. 1, 2008. – Cité pages 59 et 60.

[Trizac 2010] Trizac E., Levy Y. et Wolynes P.G. *Capillarity theory for the y-casting mechanism.* Proceedings of the National Academy of Sciences, vol. 107, no. 7, 2010. – Cité page 58.

[Trombitás 1998] Trombitás K., Greaser M., Labeit S., Jin J.P., Kellermayer M., Helmes M. et Granzier H. *Titin extensibility in situ : entropic elasticity of permanently folded and permanently unfolded molecular segments.* The Journal of cell biology, vol. 140, no. 4, 1998. – Cité page 61.

[Uversky 2002a] Uversky V.N. *Natively unfolded proteins : A point where biology waits for physics.* Protein Science, vol. 11, no. 4, 2002. – Cité pages 51 et 52.

[Uversky 2002b] Uversky V.N. *What does it mean to be natively unfolded ?* European journal of biochemistry, vol. 269, no. 1, 2002. – Cité page 52.

[Uversky 2008] Uversky V.N., Oldfield C.J. et Dunker A.K. *Intrinsically Disordered Proteins in Human Diseases : Introducing the D2 Concept.* Annual Review of Biophysics, vol. 37, no. 1, 2008. – Cité page 63.

[Uversky 2009] Uversky V.N., Oldfield C.J., Midic U., Xie H., Xue B., Vucetic S., Iakoucheva L.M., Obradovic Z. et Dunker A.K. *Unfoldomics of human diseases : linking protein intrinsic disorder with diseases.* BMC Genomics, vol. 10, no. Suppl 1, pages S7–17, 2009. – Cité pages 57 et 58.

[Vacic 2007] Vacic V., Oldfield C.J., Mohan A., Radivojac P., Cortese M.S., Uversky V.N. et Dunker A.K. *Characterization of Molecular Recognition Features, MoRFs, and Their Binding Partners*. Journal of Proteome Research, vol. 6, no. 6, 2007. – Cité page 57.

[van der Lee 2014] van der Lee R., Buljan M., Lang B., Weatheritt R.J., Daughdrill G.W., Dunker A.K., Fuxreiter M., Gough J., Gsponer J., Jones D.T., Kim P.M., Kriwacki R.W., Oldfield C.J., Pappu R.V., Tompa P., Uversky V.N., Wright P.E. et Babu M.M. *Classi cation of Intrinsically Disordered Regions and Proteins*. Chemical Reviews, vol. 114, no. 13, 2014. – Cité page 61.

[Viéville 2011] Viéville J., Tanty M. et Delsuc M.A. *Polydispersity index of polymers revealed by DOSY NMR.* Journal of magnetic resonance (San Diego, Calif. : 1997), vol. 212, no. 1, 2011. – Cité page 84.

[Vuister 1993] Vuister G.W. et Bax A.D. *Quantitative J correlation : a new approach for measuring homonuclear three-bond J (HNH. alpha.) coupling constants in 15N-enriched proteins*. Journal of the American Chemical Society, 1993. – Cité page 80.

[Ward 2004] Ward J.J., McGuffin L.J., Bryson K., Buxton B.F. et Jones D.T. *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, vol. 20, no. 13, 2004. – Cité pages 65 et 147.

[Wilkins 1999] Wilkins D.K., Grimshaw S.B., Receveur V., Dobson C.M., Jones J.A. et Smith L.J. *Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques* . Biochemistry, vol. 38, no. 50, 1999. – Cité pages 72 et 168.

[Williamson 2013] Williamson M.P. *Using chemical shift perturbation to characterise ligand binding*. Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 73, no. C, 2013. – Cité page 79.

[Wishart 1991] Wishart D.S., Sykes B.D. et Richards F.M. *Relationship between nuclear magnetic resonance chemical shift and protein secondary structure*. Journal of Molecular Biology, 1991. – Cité page 76.

[Wishart 1992] Wishart D.S., Sykes B.D. et Richards F.M. *The chemical shift index : a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy.* Biochemistry, vol. 31, no. 6, 1992. – Cité page 77.

[Wishart 1995a] Wishart D.S., Bigam C.G., Holm A., Hodges R.S. et Sykes B.D. *(1)H, (13)C and (15)N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor e ects.* Journal of biomolecular NMR, vol. 5, no. 3, 1995. – Cité page 76.

[Wishart 1995b] Wishart D.S., Bigam C.G., Yao J., Abildgaard F., Dyson H.J., Oldfield E., Markley J.L. et Sykes B.D. *1H, 13C and 15N chemical shift referencing in biomolecular NMR.* Journal of biomolecular NMR, vol. 6, no. 2, 1995. – Cité page 76.

[Wright 1999] Wright P.E. et Dyson H.J. *Intrinsically unstructured proteins : re-assessing the protein structure-function paradigm.* Journal of Molecular Biology, vol. 293, no. 2, 1999. – Cité page 51.

[Wright 2009] Wright P.E. et Dyson H.J. *Linking folding and binding.* Current Opinion in Structural Biology, vol. 19, no. 1, 2009. – Cité page 59.

[Wright 2015] Wright P.E. et Dyson H.J. *Intrinsically disordered proteins in cellular signalling and regulation.* Nature Publishing Group, vol. 16, no. 1, 2015. – Cité pages 7 et 63.

[Wüthrich 1982] Wüthrich K., Wider G., Wagner G. et Braun W. *Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance.* Journal of Molecular Biology, vol. 155, no. 3, 1982. – Cité page 81.

[Zhang 2003] Zhang H., Neal S. et Wishart D.S. *RefDB : a database of uniformly referenced protein chemical shifts.* Journal of biomolecular NMR, 2003. – Cité pages 76 et 203.

[Zuiderweg 2002] Zuiderweg E.R.P. *Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy* . Biochemistry, vol. 41, no. 1, 2002. – Cité page 79.

Chapitre 5

Mécanismes moléculaires de reconnaissance de l'ADN par AR-DBD

5.1 Introduction 101
5.2 Publication 101

5.1 Introduction

L'article en préparation intitulé Molecular basis of altered DNA recognition by the androgen receptor DNA-binding domain propose un mécanisme permettant d'expliquer la perte de spécificité de reconnaissance de l'ADN observée pour la mutation T575A au sein du DBD. Ce travail a été initié au laboratoire par Mme Ying-Hui Wang. Les premiers résultats avaient mis en évidence le rôle joué par le résidu His570 ce processus. Cependant, les travaux réalisés à partir de domaines DBD marqués à l'azote 15 ne permettaient pas l'étude de l'interaction entre les domaines DBD et des éléments de réponse de type IR3. Mon rôle a été de compléter l'attribution des carbones des deux protéines (sauvage et mutante) afin d'utiliser les corrélations des groupes méthyles comme sondes dans les complexes ternaires protéines :ADN à l'aide d'expériences de type ^{1}H-^{13}C Methyls-SOFAST. Ce type d'expériences est en effet particulièrement bien adapté pour l'étude de complexes de grande taille. Nous avons également conduit des mesures de microcalorimétrie de type ITC. L'ensemble des résultats obtenus montre que la mutation T575A conduit à une légère baisse d'affinité pour les éléments de réponse IR3, ainsi qu'à une déstabilisation de l'interface de dimérisation.

5.2 Publication

Molecular basis of altered DNA recognition by the androgen receptor DNA-binding domain

Yinghui Wang, Isabelle Lebars, Sandra Meyer, Emmanuel Oppong, Justine Wolf, Didier Busso, R. Andrew Atkinson, Jocelyn Céraline and Bruno Kieffer

[1]Département de Biologie et de Génomique Structurales, Institut de Génétique et de Biologie Moléculaire et Cellulaire (Université de Strasbourg / CNRS UMR 7104, INSERM U964), 1 rue Laurent Fries, 67404 Illkirch, France.

[2]Equipe Signalisation et Cancer de la prostate (EA4438), Université de Strasbourg, Faculté de Médecine, 1 place de l'Hôpital, BP426 - F67091 Strasbourg cedex

Address correspondence to: Bruno Kieffer, IGBMC, 1 rue Laurent Fries, 67404 Illkirch, France. Fax: +33368854718; E-mail: kieffer@igbmc.fr

Abbreviations. AR, Androgen Receptor; ARE, Androgen Response Elements; GR, Glucocorticoïd Receptor; GRE, Glucocorticoïd Response Element; DBD, DNA-Binding Domain; LBD, Ligand-Binding Domain; DHT, Dihydrotestosterone; NMR, Nuclear Magnetic Resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; HSQC: heteronuclear single quantum coherence; TOCSY, Total Correlation Spectroscopy; GARP, Globally optimized alternating phase rectangular pulse; r.m.s.d, root-mean-square deviation.

Introduction

The androgen receptor (AR) is a ligand-activated transcription factor and a member of the super-family of nuclear receptors [1,2]. It mediates the biological actions of the male sexual hormone dihydrotestosterone (DHT), including the normal development of the prostate gland [3]. Since AR is also implicated in the progression of prostate cancer, conventional therapy includes androgen ablation, by castration alone or combined with anti-androgens. After an initial clinical response, tumour growth will eventually recur. A number of molecular events are thought to be involved in escape from androgen ablation therapy, including mutations that alter the function of expressed AR [4-6].

AR shares a common structural and functional architecture with other members of the super-family of nuclear receptors. The DNA-binding domain (DBD) and ligand-binding domain (LBD) are preceded by an unusually long A/B domain. AR binds to specific response elements as a homodimer, under the control of dihydrotestosterone binding to the C-terminal LBD. Certain mutations can be rationalised readily in terms of effects on this structural/functional organisation [7]. Thus, in the LBD, the L701H mutation allows activation by adrenocorticoids while mutations of W741 and T877 allow recognition of the anti-androgens bicalutamide and flutamide, respectively, as agonists. Indeed, the T877A mutation is most frequently found in flutamide-treated patients and is thought to arise in response to the selective pressure of the treatment. Nonsense mutations or aberrant pre-mRNA splicing, on the other hand, leads to the expression of a truncated AR that lacks a LBD and is thus constitutively active [8-14]. Yet other mutations affect interactions with co-activator or co-repressor proteins.

The DNA-binding domain interacts with response elements in the promoters of target genes, when AR homo-dimerises in a DNA-dependent manner. However, while AR and other members of the same subgroup of the nuclear receptor super-family recognise the same inverted repeats in such promoters, specificity may arise from binding to direct repeats. AR DBD is expected to bind to inverted repeats in a "head-to-head" manner, as observed for the glucocorticoid receptor. In the only crystal structure of the AR DBD, the homo-dimer binds to direct repeats in the same manner, rather than the "head-to-tail" arrangement observed for the vitamin D receptor binding to direct repeats (ref). Further specificity is thought to derive from the recruitment of co-activator or co-repressor proteins in the presence of agonist or antagonist ligands, respectively. The DBD itself is formed by two C4 zinc fingers in which the first set of four cysteine residues bind one zinc ion while the second set bind another zinc ion (Figure 1A).

Given the primary function of the DNA-binding domain, it is not surprising that mutations may arise in this domain in the context of escape from androgen ablation therapy. In practice, however, the number of reported mutations in the DBD in prostate cancer patients is relatively limited [15]. A splicing variant that contains an additional 23 amino acids between the two zinc fingers has been found to remain cytosolic upon stimulation with DHT [16]. Among the point mutations that have been identified (T575A, K580R, A586V, A587S and C619Y), the positively-charged side-chain K580 that may interact directly with DNA is replaced by another positively-charged side-chain, while A586 and A587 lie on the back of the α-helix expected to lie in the major groove, such that the mutations may disrupt the interaction. Both T575 and C619, which is not a zinc-coordinating residue, are located distal to the DNA-binding interface of the DBD and the C619Y mutant has been shown to be unable to bind DNA and to be transcriptionally inactive [17]. The site of the T575A mutation precedes the third cysteine residue of the first zinc-binding site and is not expected to influence DNA-binding properties directly. It does, however, leading to a loss of specificity for AR-specific response elements. The mutation is also found coupled to the T877A mutation in the LBD, yielding an AR that it both "promiscuous" in its ligand binding and "unfaithful" in its DNA-binding properties [18].

We set out to characterize the effects of the T575A mutation on the biophysical properties of AR DBD, with the aim of offering an atomic level explanation of effects observed in tissues from prostate cancer patients. For that purpose, we solved the solution structure of both proteins (wild-type and mutant T575A) and we investigated their binding properties to both specific and non-specific half-site response elements using NMR. Our results highlight a subtle DNA binding mechanism where the remote residue at position 575 modulates the DNA binding specificity by destabilizing the AR DBD homodimer. Sequence analysis of nuclear receptor DBD suggest that this mechanism, which involves the strictly conserved His570 is shared by all steroid nuclear receptors.

Results

Wild-type and mutant structures are highly similar.

Full ^{1}H, ^{13}C and ^{15}N resonance assignments were obtained for both wild-type and mutated AR DBD and structural restraints were derived from NOESY spectra, ^{1}H-^{2}H exchange experiments, TALOS software for prediction of backbone Φ and Ψ angles, and measurement of residual dipolar couplings in gels. Analysis of the restraints allowed identification of elements of secondary structure which corresponded to those expected for nuclear receptor DBDs. Chemical shift differences between wild-type and mutant domains are concentrated around the site of mutation but are also observed for residues at the end of the second α-helix (Figure 1). The three-dimensional structures of AR DBD-WT and ARDBD-T575A are well-defined over much of the primary sequence (Figure 2), and this is reflected in low values of the average pairwise r.m.s.d. values calculated using a 5-residue window (Figure 2C, 2D). In both structures, the N-terminal 7 residues were poorly defined due to a lack of restraints, as expected from the dynamic properties of these residues (see below). In contrast, the C-terminus of the domain is well defined, and again this correlates with the observed dynamic properties. The first zinc-binding site appears more tightly defined by experimental restraints than the second site, in both wild-type and mutant domains. The tertiary structures of AR DBD-WT and AR DBD-T575A proved to be highly similar with an average pairwise backbone r.m.s.d. value between the two sets of structures of 1.11 +/- 0.12 Å, calculated using the backbone atoms of residues Cys559-Glu588 and Thr602-Thr625. Differences in orientation of the α-helices are minor and the backbone conformations around the site of mutation are very close. Both structures adopt a fold similar to that of other DNA-binding domains on nuclear receptors and that of rat AR DBD in a complex with a direct repeat of an AR-specific response element.

Figure 1

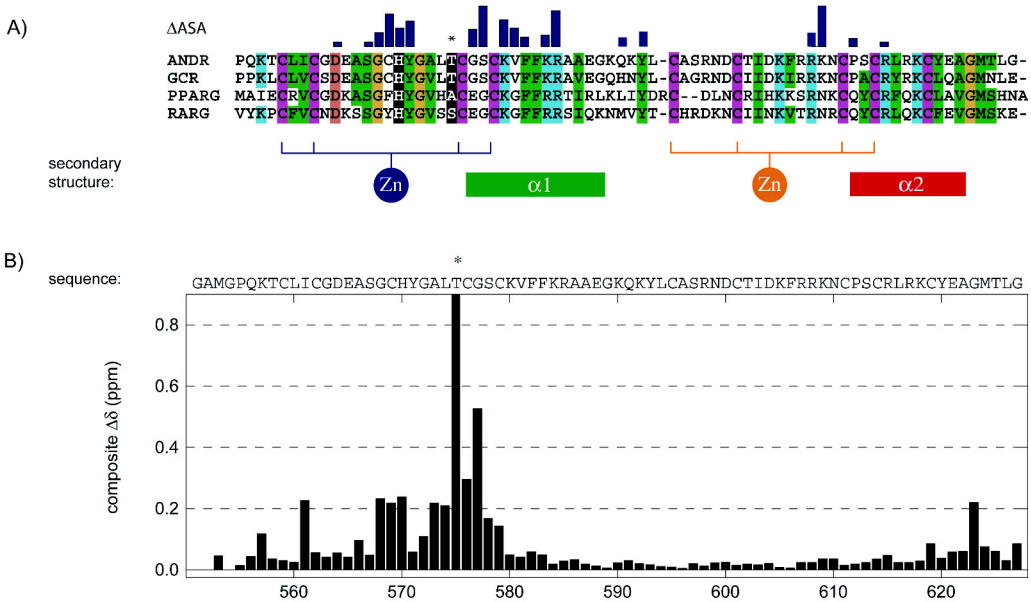


Figure 1. A) Nuclear receptor DNA-binding domain sequences, grouped according to sequence conservation with respect to the sequence of AR DBD. An asterisk marks the site of mutation. Above the sequences is the change in accessible surface area per residue upon binding to DNA: values range from 0.0 to 0.35. Accessible Surface Areas (ASA) values were calculated for each residue as a fraction of the total surface area of each amino acid type, for one monomer in the crystal structure (1R4I.pdb) in the presence and absence of the double-stranded DNA. Below the sequences are indicated the α-helices, inferred for AR DBD from chemical shifts and slowly exchanging $^{1}H_{N}$ resonances and the set of cysteine residues that coordinate the two zinc ions. The colours of the zinc-binding sites and elements of secondary structure are retained in subsequent figures. B) Composite changes in chemical shift upon mutation of T575 to alanine, calculated for $^{1}H_{N}$ and ^{15}N resonances (upper panel) and for all assigned ^{1}H, ^{13}C and ^{15}N resonances (lower panel). The central portion shows the primary sequence of the exact construct of AR DBD studied here. The α-helices and the sets of cysteine residues that coordinate the two zinc ions are indicated. An asterisk marks the site of mutation.

Figure 2

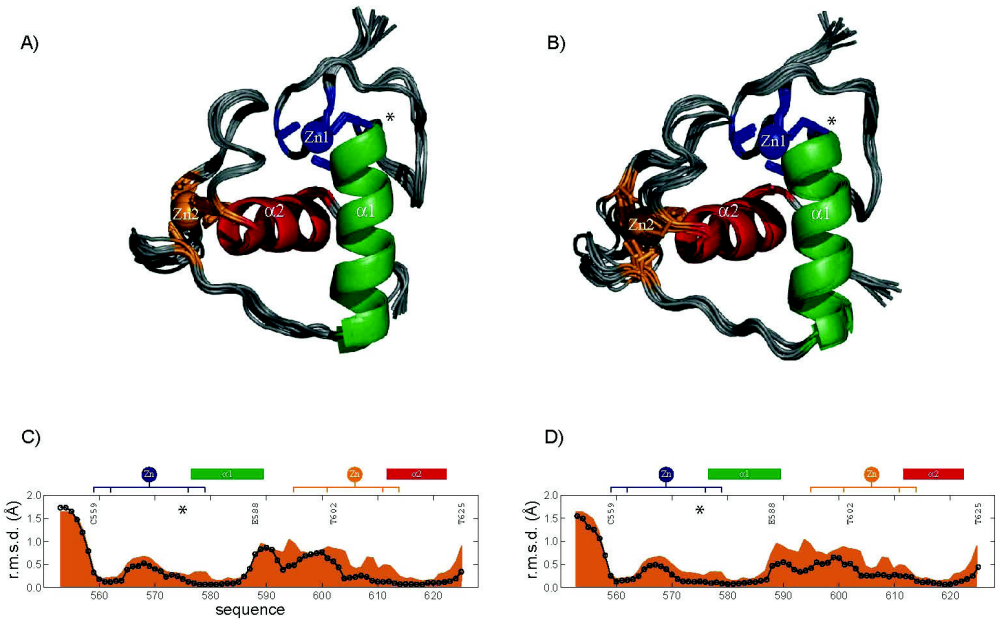


Figure 2. Bundles of 10 representative structures calculated for A) AR DBD-WT and B) AR DBD-T575A. The zinc-binding sites and elements of secondary structure are coloured as in Figure 1 and labelled. For clarity, residues preceding Gln6 are not shown. An asterisk marks the approximate site of mutation (see Figure 3A). A measure of local backbone definition is given by average pairwise r.m.s.d. values calculated over five-residue segments of the primary sequence for 20 calculated structures of C) AR DBD-WT and D) AR DBD-T575A. Residues mentioned in the text are labelled. An asterisk marks the site of mutation.

T575A mutation do not alter ARE half site interaction specificity.

The DNA binding properties of the two DBD domains was then probed in-vitro by titrating ^{15}N labelled domains with two oligonucleotides encompassing either the classical specific half-site of the androgen response element (ARE) d(5'-GCAGAACAGC-3') or the non-specific half-site of the glucocorticoid receptor response element (GRE) d(5'-GCTGTACAGC-3'). The analysis of the four sets of ^{1}H-^{15}N HSQC (two protein domains times two oligonucleotides) revealed that the binding affinity value is determined rather by the oligonucleotide sequence than the DBD mutation. Indeed, progressive addition of ARE oligonucleotides to either the wild-type or T575A mutant DBD do-

mains led to intermediate to slow exchange between bound and free states of the proteins, indicative of a *Kd* value lower than 1 µM (Figures 3A and 3B). In contrast, addition of non-specific GRE oligonucleotide led to a fast exchange between bound and free states, enabling the determination of *Kd* values of 31 ± 5 µM and 42 ± 9 µM for the wild-type and the T575A mutant DBD, respectively (Figures 3C to 3F).

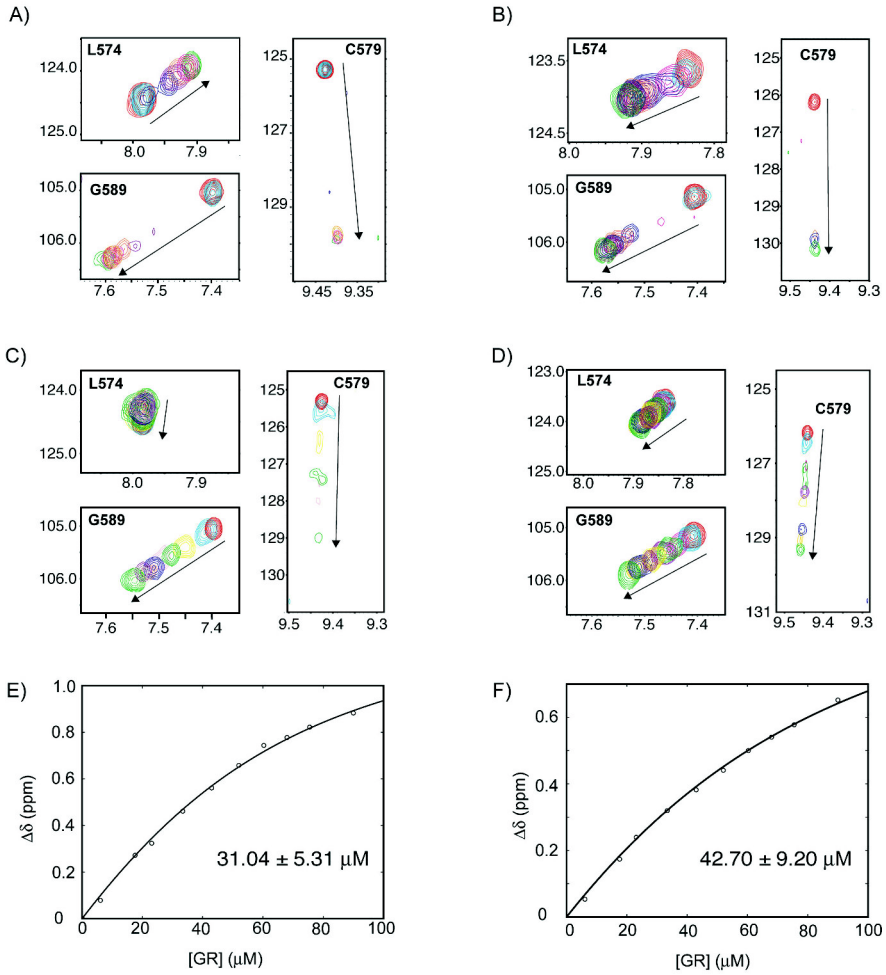


Figure 3

Figure 3. Overlays of regions of ^{1}H-^{15}N HSQC spectra of ^{15}N labeled (A) AR DBD wt, free in solution and upon successive additions of ARE, (B) AR DBD T575A, free in solution and upon successive additions of ARE, (C) AR DBD wt, free in solution and upon successive additions of GR and (D) AR DBD T575A, free in solution and upon successive additions of GR. Arrows indicate the cross-peaks which shifted upon addition of DNA. Chemical shift changes of the backbone $^{1}H_N$ and ^{15}N ($\Delta\delta_{ave} = [(\Delta\delta^{1}H)^2 + (\gamma_N/\gamma_H)^2*(\Delta\delta^{15}N)^2]^{1/2}$) upon titrations of AR DBD wt (E) and AR DBD T575A (F) with GR are plotted as a function of GR concentration. The continuous line is the best fit of the data to the equation described in Materials and Methods.

A detailed analysis of individual protein residue behavior revealed however some striking differences between mutant and wild type proteins upon titration as illustrated by Leu574 whose H_N chemical shift undergoes an opposite shift upon titration by the ARE oligonucleotide (Figures 3A and 3B). The weighted average of ^{1}H and ^{15}N maximal chemical shift changes observed for the four titration experiments is shown as a function of the sequence in figure 4. The overall pattern of frequency shifts correlates the changes expected from the proximity of the DNA molecule, with the larger changes observed for residues located close to the first zinc-binding site and within helix $\alpha 1$. The largest changes were observed for His570, Cys579 and Val581 ($\Delta\partial_{max} > 280$ Hz). The comparison between the profiles obtained for the wild-type and the T575A mutant proteins revealed a striking difference for His570 (indicated by a star in figure 4). The significant frequency shift observed for this residue in the wild-type protein upon addition of the ARE oligonucleotide is fully abolished in the T575A mutant. Moreover, this residue exhibited one of the largest differences between specific ARE and non-specific GR oligonucleotides, indicating a possible role in the specificity of the protein-DNA interaction. The behavior of His570 contrasted with those of Val581, where the frequency shifts were determined by the DNA sequence and insensitive to the T575A mutation.

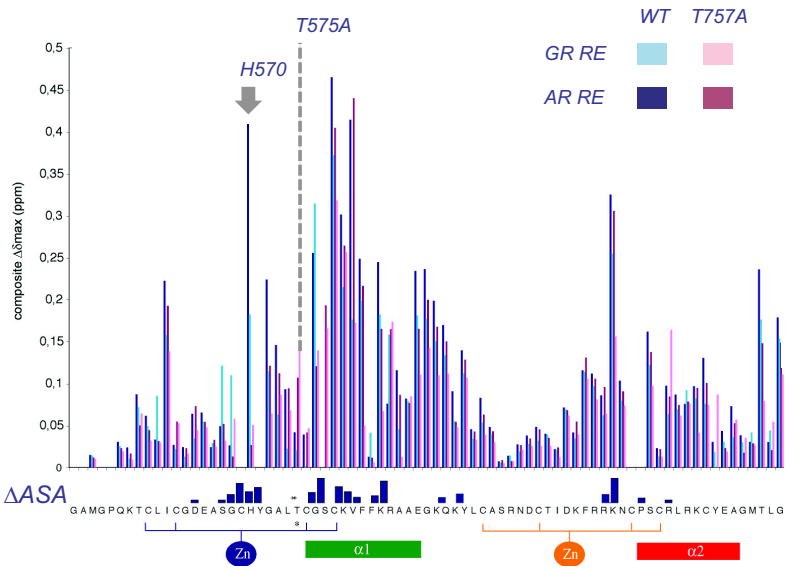


Figure 4. The chemical shift differences ($\Delta\delta_{max}$) between the free state and bound states calculated in titrations of AR DBDwt with ARE (dark blue), AR DBDwt with GR (light blue), AR DBD T575A with ARE (purple) and AR DBD T575A with GR (pink) are plotted as a function of the primary sequence. The α-helices and the sets of cysteine residues that coordinate the two zinc ions are indicated. The positions of T575 and H570 residues are indicated by a dashed line and an arrow, respectively. Residues of the DBD in direct contact with DNA are indicated by the change in accessible surface area per residue upon binding to DNA (ΔASA).

Local environment of H570 is altered upon T575A mutation

A closer view of the local environment of His570 was provided by the measurement of ^{15}N chemical shifts of the histidine imidazole ring. This measurement is performed using long-range ^{1}H-^{15}N HSQC spectra, that correlates the frequencies of ^{15}Nε2 and ^{15}Nδ1 nuclei with those of the Hε and Hδ non-exchangeable protons. The pattern of correlation peaks measured for both the wild type and the T575A mutant at pH 6.5 was clearly different, evidencing an alteration of the local environment of the Nδ1 atom upon T575A mutation (Figures 5A, 5B).

Figure 5

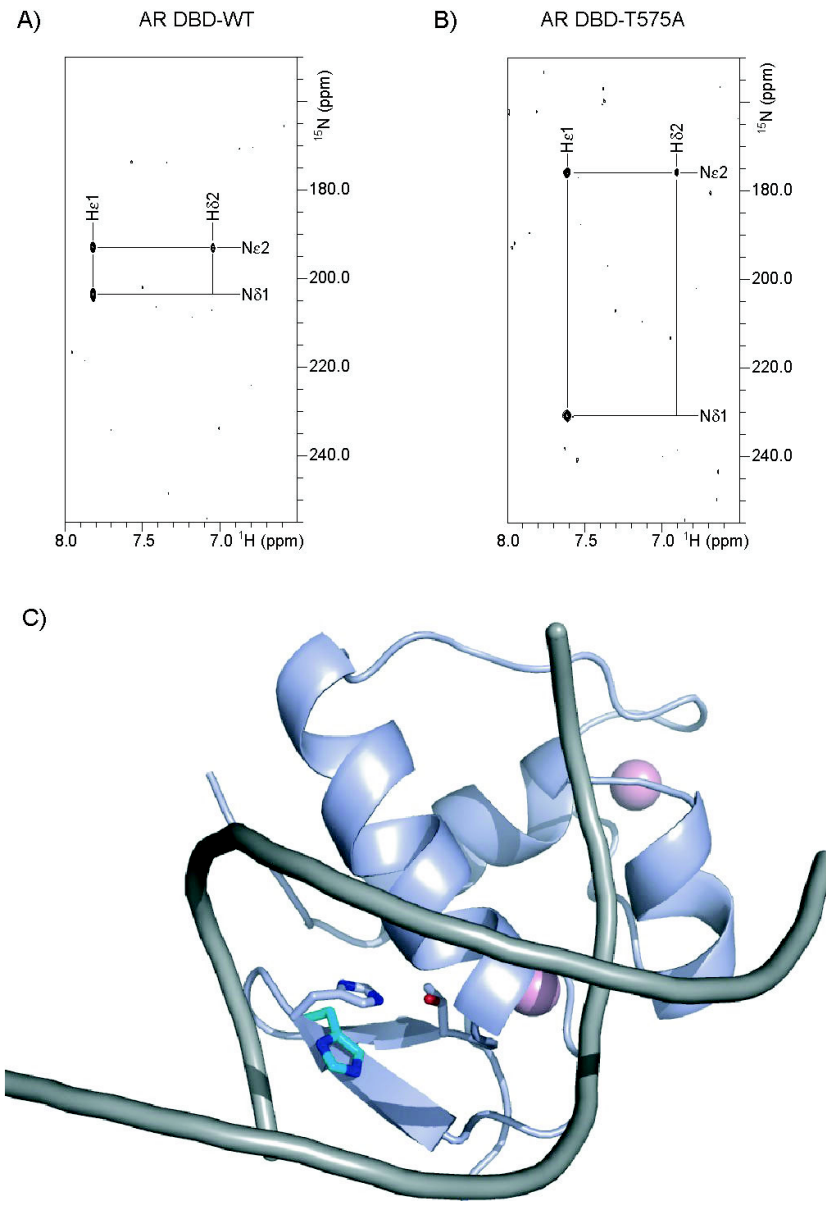


Figure 5. Long-range ^{1}H-^{15}N HSQC spectra correlating Hδ1 and Hε2 resonances with Nδ1 and Nε2 resonances of His570 for A) AR DBD-WT and B) AR DBD-T575A. C) Overlay of the His570 side chain from the solution structure of free wild-type AR DBD (in cyan) and the crystal structure of the DNA bound AR DBD (in grey, PDBID: 1R4I).

This change could either be due to a shift of the histidine pKa, resulting to a different protonation state at pH 6.5, or to a change of the hydrogen bounding pattern involving either Nδ1 or Nε2 atom as an acceptor. pH titration experiments were performed and did not show a significant change of the H570 pKa values (data not shown). ^{15}N relaxation measurements were performed on both AR DBD-WT and AR DBD-T575A domains at 298K to seek for possible differences introduced by the T575A mutation on the backbone dynamics of the two proteins (Supplementary figure 1). These data revealed increased µs-ms time-scale dynamics for His570 and Gly572 residues upon T575A mutation. These data suggest that the T575A mutation affects the dynamical behaviour of His570, probably by disrupting transient hydrogen bound between the Nε2 of His570 and the Hγ1 of Thr575. Examination of the three-dimensional structures of AR DBD-WT and AR DBD-T575A reveals that the side-chains of His570 and Thr575 (or Ala575) are in close proximity (Figure 5C).

The T575A mutation leads to a loss of affinity for IR3 response elements

The apparent affinity of AR DBD for two natural IR3 response elements, AREc12 and SLP, was measured using Isothermal Titration Calorimetry experiments (Figure 6). AREc12 is recognized by AR but also by the Glucocorticoid Receptor (GR) and the Progesterone Receptor (PR) while SLP is specifically recognized by AR and not by the other steroid receptors [19]. The ITC experiments indicated that both wild-type and T575A DBDs bind to both response elements with micro-molar affinity as an homodimer (with DNA:DBD stoichiometry parameters varying between 0.3 and 0.6, see Table 1). Both proteins bind AREc12 response elements with a slightly better affinity (a consistent affinity ratio of 2.3 is observed between the Kd values for AREc12 and SLP). The T575A mutation leads to a decrease of affinity for both response elements by a factor of about 6 (5.7 and 6.5 for the SLP and AREc12 response elements, respectively). Since no significant difference was observed between the wild-type and T575A DBDs in their binding affinity for DNA half-sites, the differences observed for the binding to complete response elements suggest that the effect of the T575A mutation on DNA recognition is related to the homo-dimerization of AR DBD on response elements. This aspect was further investigated by performing titration experiments where both DNA and protein were monitored using imino-proton chemical shifts and methyl resonances, respectively.

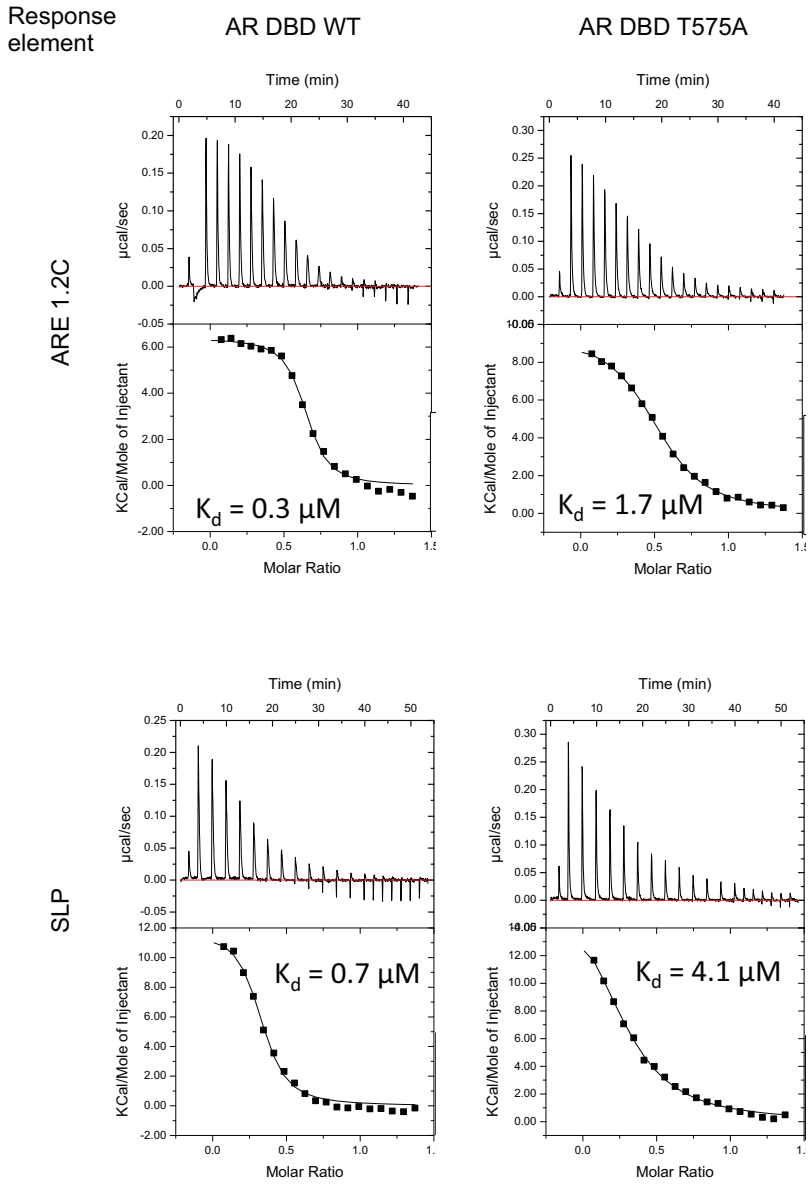


Figure 6: Isothermal Titration Calorimetry experiments performed at 15 °C for the (upper panel) and the SLP (lower panel) response elements. Left and right columns correspond to wild-type and T575 mutant AR DBD, respectively.

Table 1: Thermodynamic parameters for the binding of wild-type and T575 mutant AR DBD to SLP and AREc12 13 base pairs oligonucleotides. Experiments were performed at 15 °C in 50 mM sodium phosphate (pH 6.5) and 50 mM NaCl.

Interactions	K_d (μM)	n	ΔH kcal.mol^{-1}	-TΔS kcal.mol^{-1}
SLP - wt AR DBD	0.7	0.3	11.9 ± 0.4	-19.9 ± 0.4
SLP - T575A AR DBD	4.1	0.3	18.0 ± 1.0	-25.1 ± 1.0
AREc12 - wt AR DBD	0.3	0.6	6.4 ± 0.1	-15.0 ± 0.1
AREc12 - T575A AR DBD	1.7	0.5	9.4 ± 0.1	-17.0 ± 0.1

The T575A mutation affects the cooperative binding to IR3 response elements

The comparison of imino-proton resonances of the SLP and AREc12 DNA oligonucleotides upon titration with either wild type or T575A AR DBD proteins provide a first insight on the effect of the mutation on DNA recognition process (Figure 7). In particular, the imino-protons of G4 and G13, whose resonances are well resolved and located in the two response element half-sites, provide appropriate probes to report on local changes upon protein binding. For the SLP response element, successive additions of AR DBD lead to a progressive broadening of all imino-proton resonances. The broadening of the G4 imino-proton resonance is observed earlier than those of G13 in the titration with both wild-type and mutant proteins suggesting a stepwise mechanism for the recognition of SLP by AR DBD (Figure 7A and 7C and Supplementary figures 2 and 3). In contrast, additions of wild-type AR DBD to the AREc12 oligonucleotide lead to both a shift and a broadening of G4 and G13 imino-proton resonances indicative of a cooperative binding in both half-sites. This behaviour is clearly affected by the T575A mutation, where sequential broadening of G4 and G13 imino-proton resonances is observed. The dimerization upon DNA binding was also followed from the protein side, by monitoring chemical shift perturbations of methyl resonances during the titration experiment (Figure 7E and Supplementary figure 4). In particular, we found that the methyls groups of Ala587 and Ala596 were appropriate probes to monitor the protein-DNA and protein-protein interfaces, respectively, during the interaction. The comparison of the methyl resonances evolution of Ala587 of the wild-type DBD upon binding to SLP and AREc12 indicates that this residue, located in the DNA recognition helix, is sensitive to the A to T transversion within the second re-

sponse element half-site (Supplementary figure 4). This breaking of symmetry, that provides a clear evidence of the dimer formation of both wild-type and mutant DBD upon binding to AREc12 oligomer, is not observed for the binding to SLP due to identical chemical shifts environments of Ala587 in the two half-sites. Dimerization of DBD is also evidenced by the methyl resonances evolution of Ala596 in the wild-type protein. The corresponding correlation peak fades out very quickly in presence of SLP oligonucleotide while a second peak, corresponding to the protein-DNA complex is observed for the AREc12 oligonucleotide. This indicates that significant dynamics of the protein-protein interface is present within the AR DBD-SLP ternary complex, contrasting with a more rigid interface for the AREc12 complex. For the T575A mutant, the methyl correlation peak corresponding to Ala596 in the symmetric ternary complex is observed, albeit the peak corresponding to the free form of the DBD still remains observable at high DNA concentrations.

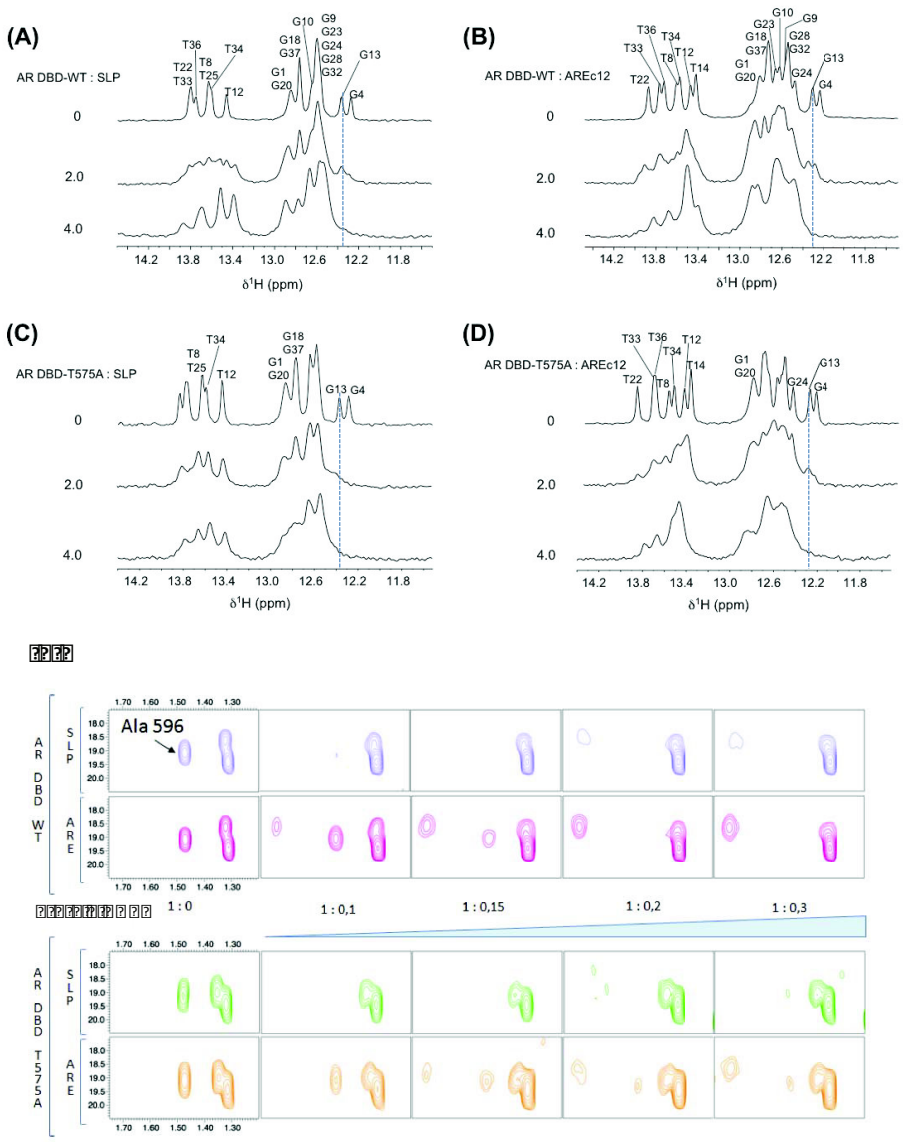


Figure 7: AR DBD - DNA titration experiments with wild-type DBD and SLP (A) or AREc12 (B) response elements and T575A DBD and SLP (C) or AREc12 (D) response elements showing the imino-proton region of the oligonucleotides upon addition of increasing amounts of protein. Protein to DNA stoichiometric ratios are indicated on the left side. (E) Evolution of Ala596 methyl correlation shown for increasing amounts of DNA (indicated in the middle). Titration experiments between the wild-type DBD and SLP or AREc12 oligonucleotides are shown in violet and pink, re-

spectively. Titration between the T575A DBD and SLP or AREc12 oligonucleotides are shown in green and orange, respectively.

Discussion

In attempting to rationalise the clinical effects of a point mutation such as T575A from biophysical measurements, it is important to consider the context of the mutation. DNA-binding domains of nuclear receptors are very highly conserved across species and well-conserved across the set of receptors of a particular species (Figure 1A). This results in a conserved fold around two zinc ions and a common mode of interaction with DNA. The set of residues that are strictly conserved across the super-family of human nuclear receptors contains, not surprisingly, the eight coordinating cysteine residues The remainder of the strictly and very highly conserved residues are gathered rather centrally and include side-chains fulfilling a structural role, such as the pair of phenylalanine residues on the interior of the first α-helix, and positively charged side-chains on the same α- helix that interact in the major groove of the DNA. Analysis of sequence alignments also reveals blocks of residues that are conserved only among the steroid receptors. These blocks are situated more peripherally than the residues conserved across all DNA-binding domains: some are seen to cluster around the zinc binding sites, while another set are grouped in the loop connecting α1 to the second zinc binding site. With the possible exception of Tyr593, none of the residues in this latter group are expected to contribute directly to the interaction with DNA. A further set of residues are seen to be specific to AR, with amino acid types that differ from the other steroid receptors and that are not observed frequently at corresponding positions in other classes of receptor: Thr558, Ala573, Ala587, Lys590, Gln591, Lys592, Ser597, Thr602, Phe606, Ser613 and Glu621. None of these sites contributes directly to interactions with DNA and most are situated on the opposite face of the domain to α1.

It is rather difficult to rationalise AR specificity for DNA response elements in terms of a precise set of interactions. Since the blocks of residues specific to steroid receptors are associated with similarity of ligand, these may reflect evolutionary divergence of nuclear receptor classes rather than factors contributing to DNA recognition. The clustering of these blocks around the zinc binding sites, however may suggest a role in modulating the stability of the sites. The set of residues specific to AR are largely polar or charged (with the notable exception of Phe606) and may correspond to

sites that define interactions with the ligand binding domain of AR or with other proteins. Some of these residues lie within the steroid receptor-specific blocks, suggesting AR-specific modulation of the role of these blocks. AR specificity therefore seems to be rather subtly defined: it can, however, be altered in cancerous cells by a single mutation in the DNA-binding domain.

The results of this study suggest several possible effects at the atomic level that may contribute to reduced specificity of AR for response elements. Firstly, the increased ps-ns time-scale motions, if quenched upon binding to DNA sequences, incur a small entropic penalty for the mutant domain which should raise the dissociation constant. Secondly, the alteration of histidine side-chain properties (such as charge or polarity) may have direct or indirect effects on binding to DNA. The importance of the protonation state of this residue was noted in the study of the effects of pH on the DNA-binding properties of the glucocorticoid receptor [21]. While His570 is not expected to make contacts with DNA that define specificity, it may contribute to the stabilization of the symmetric dimer at specific sites as suggested by our titration experiments. This hypothesis is further reinforced by the phenotype of mutants where His570 was replaced by an alanine in full-length AR. The response of mutated androgen receptors to dihydrosterone hormone was probed using luciferase assays performed in prostate cell line LNCaP cells, using either a specific (PSA61) or a non-specific (MMTV) AR promoter reporter constructs (Supplementary figure 5). As previously observed, the T575A mutation leads to an enhanced response for the MMTV non-specific reporter, whereas the specific response remains unaffected. The activity of H570A AR on the same construct was 1.5-fold higher to that one of the wild type AR, but was reduced to background levels on the PSA61 specific reporter construct, reflecting an inversion of specificity with respect to the wild-type protein.

In a recent elegant study of the binding of the glucocorticoid receptor to DNA sequences, Miejsing et al. suggested "that DNA functions as a sequence-specific allosteric ligand of GR that tailors the activity of the receptor towards target genes". A similar crystallographic study of AR would provide valuable insight into AR function and allow direct comparison of AR and GR response to DNA-binding. In this context, the analysis of the T575A mutant might shed further light on the effects of this mutation, although if the effects are as subtle as they appear to be from our data, structural data may not provide a full explanation. NMR spectroscopy may again contribute by allowing changes in dynamics and stability of the DNA binding domain to be monitored as a function of DNA sequence.

Materials and methods

Samples preparation

AR DBD-WT ($AR_{555-627}$) and AR DBD-T575A were cloned into $pETM_{30}$ containing the GST tag. 50⌈L of BL21 (DE3) *star* competent cells were transformed with plasmid DNA by heat-shock process. The proteins were expressed in *E.coli* BL21 (DE3) star cells. ^{15}N- and ^{15}N,^{13}C-labelled proteins were produced in M9 minimal medium supplemented with 1.0 $g.L^{-1}$ of $^{15}NH_4Cl$ and 2.0 $g.L^{-1}$ of ^{13}C glucose. Unlabelled proteins were expressed in LB medium. Single colonies of transformed BL21 (DE3) star cells were grown overnight in 200 mL LB medium at 37°C, with 50 $\mu L.mL^{-1}$ kanamycin. When the cells had grown to an OD_{600} of 0.6, expression was induced by 1mM IPTG at 18°C and incubated overnight. The expressed cultures were harvested by centrifugation at 4000 rpm for 20 min at 4°C. Pellets were stored at -20°C.

Pellets were lysed in 40 mL binding buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM DTT), sonicated and ultracentrifuged at 40000 rpm for 1 hour at 4°C. The fusion protein was incubated with resin for 2 hours at 4°C and washed with washing buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 1 mM DTT). GST was cleaved from the fusion protein by addition of TEV protease and incubation at room temperature for 2 hours. The protein was eluted with the washing buffer and purified on a Superdex G75 (HiLoad 16/60) gel fitration column, equilibrated with buffer (50 mM phosphate at pH 6.5, 150 mM NaCl, 1 mM DTT). Fractions were concentrated to a volume of 1 ml using 5K amicon (Vivaspin) and transferred to 50mM phosphate pH 6.5 and concentrated further to 500 µL. Final concentrations of AR DBD-WT and AR DBD- T575A were about 600 µM and 1 mM for labelled and unlabelled samples, respectively.

DNA used for titration experiments were 10 bp oligonucleotides: ARE, d(5'-GCAGAACAGC-3')/(5'-GCTGTTCTGC-3'), which the six central base-pairs were reported to be a specific response element for androgen receptor and GR, d(5'-GCTGTACAGC-3'), which the six central base-pairs were described to be a non-specific response element. Titrations with full-length response elements were performed with 13 bp oligonucleotides: SLP, d(5'-GCAGAACTGGCTGACCAGC-3')/ d(5'-GCTGGTCAGCCAGTTCTGC-3') and AREc12, d(5'-GCAGAACTGGCTGTCCAGC-3')/ d(5'-GCTGGACAGCCAGTTCTGC-3'). DNA samples were purchased from Eurogentec (Belgium). The duplexes were annealed in the NMR buffer by heating at 90°C (2 min), and snap-cooled at 4°C for 5 min. The concentrations of DNA samples were determined on a Nanodrop Spectrometer using molar extinction coefficients calculated with the Applied Biosystems calculator (www.ambion.com/techlib/misc/oligocalculator.html).

NMR spectroscopy

Most NMR spectra were recorded at 600 MHz (^{1}H frequency) on a Bruker DRX600 spectrometer equipped with a triple-resonance cryoprobe with z-gradients. All spectra were recorded at 25°C, unless specifically stated otherwise. One-dimensional ^{1}H spectra were recorded with appropriate spectral widths for the sample, using a WATERGATE sequence [22] for water suppression. Additional homonuclear TOCSY and NOESY spectra were recorded at 800 MHz.

^{1}H-^{15}N HSQC were recorded with spectral widths set to avoid folding of ^{15}N resonances. Typically, 256 t_1 increments were acquired and quadrature was achieved using the method of States et al. [23]. Spectra were processed applying a 90°-shifted sine bell and zero-filling at least to the next power of two in each dimension prior to Fourier transformation. Spectra were processed using NMRPipe [24] and analysed in CARA [25] and the NEASY module of CARA. ^{15}N-edited TOCSY and NOESY were recorded with 50 increments in the indirect ^{15}N dimension and 320 increments in the indirect ^{1}H dimension. Spectra were processed as above with linear prediction in the indirect ^{15}N dimension.

A set of three-dimensional triple-resonance experiments (HN(CO)CA, HNCA, HN(CO)CACB, HNCACB, HNCO) were recorded to obtain backbone resonance assignments. Spectra were processed as above with linear prediction in the indirect ^{13}C dimension. Predictions of backbone Φ and Ψ angles were obtained from resonance assignments using the programme TALOS [26]. Residual dipolar couplings for backbone ^{1}H-^{15}N pairs were measured in polyacrylamide gels using IPAP pulse sequences [27].

Three-dimensional structure determination was performed using the semi-automatic ATNOS/CANDID procedure [28,29], with XPLOR-NIH [30,31] as molecular dynamics programme. Hydrogen bonds that could be unambiguously identified from hydrogen/deuterium exchange data and NOE patterns were introduced, together with constraints on dihedral angles predicted with a high degree of confidence by TALOS. The resulting structures were refined in XPLOR-NIH using a standard protocol (refine.inp), modified to start from a higher initial temperature (2000K), with a larger number of cooling steps (10000 steps) and a longer final energy minimisation step (1200 steps). The structures were then refined using parallhdg5.3 parameters and stereospecific assignments made where possible, before final refinement in explicit solvent [32,33]. Local and global pairwise r.m.s.d. values were calculated in XPLOR-NIH using in-house scripts. Figures were produced using Pymol [34].

^{15}N R_1 and R_2 relaxation rates were measured using pulse sequences based on those of Farrow et al. [35], with 8 scans per increment and 200 increments in the indirect dimension, with a delay of 2.2 s between scans. For ^{15}N R_1 relaxation, intensities were extracted from a set of 11 spectra recorded with values of the longitudinal relaxation delay of 0, 150.5, 301.0, 451.4 (x2), 601.9, 752.4, 902.9, 1053, 1204 and 1354 ms, with 180° proton pulses every 2 ms to suppress cross-correlated relaxation [36]. For ^{15}N R_2 relaxation, intensities were extracted from a set of 12 spectra recorded with values of the transverse relaxation delay of 0, 15.8, 31.7, 47.5 (x2), 63.4, 79.2, 95.0, 126.7, 158.4, 190.1 and 221.8 ms, with ^{15}N 180° pulses applied every 1.2 ms at a field strength of 4.2 kHz and ^{1}H 180° pulses applied every 2.4 ms to suppress cross-correlated relaxation [36] . Exponential decay rates were obtained from a non-linear least-squares two-parameter fit using a Levenberg-Marquardt algorithm implemented in Matlab (The Mathworks, Inc.).

^{1}H-^{2}H exchange rates of exchangeable hydrogen atoms were measured by diluting ^{15}N-labelled samples 10-fold into $^{2}H_2O$ and recording ^{1}H-^{15}N HSQC spectra over the following 14-18 hours. Intensities were fitted using a single exponential as for relaxation data. Long-range HSQC experiments were performed to correlate the side-chain ^{15}N resonances of His570 with the non-exchangeable ^{1}H resonances of the side-chain.

Titration experiments

NMR experiments were recorded at 700 MHz on a Avance III Bruker spectrometer equipped with a TCI z-gradient cryoprobe. NMR data were processed using TopSpin (Bruker) and analyzed with Sparky software packages (Goddard, T.D. & Kneller, D.G., SPARKY 3, University of California, San Francisco). NMR experiments were performed in 50 mM sodium phosphate buffer (pH 6.5) and 1 mM DTT in 90/10 H_2O/D_2O.

DNA samples volumes were 280 µl in Shigemi NMR tubes. ^{1}H assignments were obtained using standard homonuclear experiments. NMR data were acquired at 10°C, 20°C and 25°C. Solvent suppression was achieved using the "Jump and Return" sequence combined to WATERGATE (Plateau *et al.*, 1982; Sklenar *et al.*, 1993; Piotto *et al.*, 1993). Two-dimensional NOESY spectra were acquired with mixing time of 400 ms and 50 ms. Base-pairing was established via sequential nuclear Overhauser effects (nOes) observed in 2D NOESY spectra at different mixing times.

All titrations experiments were performed at 25°C. Complexes were formed by addition of DNA to protein monitoring the amide region of ^{1}H-^{15}N HSQC spectra. ^{15}N decoupling during acquisition time was achieved using GARP composite pulse sequence. Complexes were also formed by addi-

tion of protein to DNA monitoring the imino protons region of 1D spectra. Samples volumes were 200 µl in 3 mm NMR tubes. The concentrations of proteins samples ranged from 10 to 100 µM.

Dissociation constants (K_d) values were calculated by fitting the experimental data (chemical shift changes of 10 residues upon DNA binding) to the equation:

$$K_d = ([R_0] - [RL])*([L_0]-[RL])/[RL]$$

$$[RL] = [R_0] * \Delta\delta_{obs} / \Delta\delta_{max}$$

where $[R_0]$, $[L_0]$, and [RL] are the concentrations of the protein (total), DNA (total) and the protein DNA-complex, respectively; $\Delta\delta_{obs}$ is the difference between the observed chemical shift and the chemical shift of the free state; and $\Delta\delta_{max}$ is the chemical shift difference between the free state and bound states. K_d and Δ_{max} were treated as fitting parameters during the curve fitting.

Isothermal Titration Calorimetry

Isothermal titration calorimetry experiments were performed with a MicroCal instrument ITC200 at 15 °C. Protein and DNA samples were dialyzed in a buffer containing 50 mM phosphate at pH 6.5 and 50 mM NaCl. 200 µL of a solution containing 60 µM of AR DBD was titrated (in the cell) by a 160 µM double stranded DNA solution. Control experiments were recorded to correct for the heats of DNA dilution. MicroCal Origin 7.0 was used to correct for the heats of mixing and of dilution, and to analyze the data. The thermodynamic parameters as enthalpy change ΔH, entropy change ΔS, affinity K_D and stoechiometry *n*, were obtained using the model of one binding site.

Cell lines and cell cultures

The human androgen-sensitive prostate cancer cell line LNCaP clone FCG (European Collection of Cell Cultures, Salisbury, UK) (ECACC 89110211, passage number 08-11) was cultured in phenol red-free RPMI 1640 medium supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, 10 mM HEPES, 1 mM sodium pyruvate (Sigma-Aldrich), 100 U/ml penicillin and 100 µg/ml streptomycin (Invitrogen). Cells were cultured in a humidified atmosphere enriched in 5% CO_2 at 37°C.

Plasmids

The pEGFP-C3 plasmid was from BD Clontech (Ozyme, France). The pEGFP-AR-WT construct was as described previously [9]. The pEGFP-AR-T575A expression plasmid was derived from

pEGFP-AR-WT by replacing the wild type BstEII - BamHI fragment with the corresponding fragment from pAR-T575A [18]. The pEGFP-AR-H570A expression plasmid was constructed by site-directed mutagenesis performed on pEGFP-AR-WT. The MMTV-Luc reporter plasmid containing steroid hormone responsive elements was a gift from Prof. P. Chambon (IGBMC, Illkirch, France). The PSA-61-Luc reporter plasmid containing a 6-kb PSA promoter fragment with the three AREs (AREI, AREII, and AREIII) was from Dr. Trapman (Erasmus University, Rotterdam). The control luciferase reporter vector pGL4.70[hRLuc] was purchased from Promega (Promega, Charbonnières, France).

Transfection and Luciferase reporter assay

Transient transfection was performed in LNCaP cells by using the JetPEI reagent (PolyPlus Transfection, Polyplus, France) according to instructions provided by the manufacturer. LNCaP cells were seeded in 96-well plates (10^4 cells/well) and transfected for 48 hours with 75 ng of the indicated pEGFP plasmid plus 150 ng of the reporter plasmid and 10 ng pGL4.70[hRLuc]. Twenty-four hours later, cells were switched to phenol red- and serum-free medium containing 100 nM DHT (Sigma-Aldrich, Saint-Quentin Fallavier, France) or vehicle. Luciferase activity was measured 48 hours later with the Dual-Glo luciferase reporter assay kit (Promega, Charbonnières, France). Transcriptional activity of the wild type AR was considered as control. Statistical analysis was performed with Graphpad Quickcalcs (Graphpad.com) using a paired t test from at least two independent experiments. A P value of less than 0.05 was considered significant

References

1. Evans, R. M. (1988) *Science* **240**(4854), 889-95

2. Robinson-Rechavi, M., Escriva Garcia, H., and Laudet, V. (2003) *J Cell Sci* **116**(Pt 4), 585-6

3. Lee, H. and Chang, C. (2003) *Cell Mol Life Sci* **60**(8), 1613-22

4. Culig, Z. (2003) *Urology* **62**(5 Suppl 1), 21-6

5. Taplin, M. and Balk, S. P. (2004) *J Cell Biochem* **91**(3), 483-90

6. Bergerat, J. and Céraline, J. (2009) *Hum Mutat* **30**(2), 145-57

7. Brooke, G. N. and Bevan, C. L. (2009) *Curr Genomics* **10**(1), 18-25

8. Céraline, J., Cruchant, M. D., Erdmann, E., Erbs, P., Kurtz, J., Duclos, B., Jacqmin, D., Chopin, D., and Bergerat, J. (2004) *Int J Cancer* **108**(1), 152-7

9; Lapouge, G., Erdmann, E., Marcias, G., Jagla, M., Monge, A., Kessler, P., Serra, S., Lang, H., Jacqmin, D., Bergerat, J., and Céraline, J. (2007) *Int J Cancer* **121**(6), 1238-44

10. Lapouge, G., Marcias, G., Erdmann, E., Kessler, P., Cruchant, M., Serra, S., Bergerat, J., and Céraline, J. (2008) *Adv Exp Med Biol* **617**, 529-34

11. Dehm, S. M., Schmidt, L. J., Heemers, H. V., Vessella, R. L., and Tindall, D. J. (2008) *Cancer Res* **68**(13), 5469-77

12. Hu, R., Dunn, T. A., Wei, S., Isharwal, S., Veltri, R. W., Humphreys, E., Han, M., Partin, A. W., Vessella, R. L., Isaacs, W. B., Bova, G. S., and Luo, J. (2009) *Cancer Res* **69**(1), 16-22

13. Guo, Z., Yang, X., Sun, F., Jiang, R., Linn, D. E., Chen, H., Chen, H., Kong, X., Melamed, J., Tepper, C. G., Kung, H., Brodie, A. M. H., Edwards, J., and Qiu, Y. (2009) *Cancer Res* **69**(6), 2305-13

14. Marcias, G., Erdmann, E., Lapouge, G., Siebert, C., Barthélémy, P., Duclos, B., Bergerat, J., Céraline, J., and Kurtz, J. (2010) *Hum Mutat* **31**(1), 74-80

15. Gottlieb, B., Beitel, L. K., Wu, J. H., and Trifiro, M. (2004) *Hum Mutat* **23**(6), 527-33

16; Jagla, M., Fève, M., Kessler, P., Lapouge, G., Erdmann, E., Serra, S., Bergerat, J., and Céraline, J. (2007) *Endocrinology* **148**(9), 4334-43

17. Nazareth, L. V., Stenoien, D. L., Bingman, 3rd, W. E., James, A. J., Wu, C., Zhang, Y., Edwards, D. P., Mancini, M., Marcelli, M., Lamb, D. J., and Weigel, N. L. (1999) *Mol Endocrinol* **13**(12), 2065-75

18. Monge, A., Jagla, M., Lapouge, G., Sasorith, S., Cruchant, M., Wurtz, J., Jacqmin, D., Bergerat, J., and Céraline, J. (2006) *Cell Mol Life Sci* **63**(4), 487-97

19. Denayer, S., Helsen, C., Thorrez, L., Haelens, A. and Claessens, F. (2010) Mol Endocrinology 24:898-913.

21. Lundbäck, T., van Den Berg, S., and Härd, T. (2000) *Biochemistry* **39**(30), 8909-16

22. Piotto, M., Saudek, V., and Sklenár, V. (1992) *J Biomol NMR* **2**(6), 661-5

23. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J Magn Reson* **48**(2), 286–292

24. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J Biomol NMR* **6**(3), 277–293

25. Keller, R. (2004). *The Computer Aided Resonance Assignment Tutorial.* CANTINA Verlag.

26. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J Biomol NMR* **13**(3), 289–302

27. Hall, J. B., Dayie, K. T., and Fushman, D. (2003) *J Biomol NMR* **26**(2), 181–186

28. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Mol Biol* **319**(1), 209–227

29. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Biomol NMR* **24**(3), 171–189

30 Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J Magn Reson* **160**(1), 65-73

31. Schwieters, C. D., Kuszewski, J. J., and Clore, G. M. (2006) *Prog NMR Spectr* **48**(1), 47–62

32. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J., and Nilges, M. (2003) *Proteins* **50**(3), 496-506

33. Nabuurs, S. B., Nederveen, A. J., Vranken, W., Doreleijers, J. F., Bonvin, A. M. J. J., Vuister, G. W., Vriend, G., and Spronk, C. A. E. M. (2004) *Proteins* **55**(3), 483-6

34. DeLano, W. (2002). The PyMol Molecular Graphics System. . Retrieved from http://www.pymol.org.

35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33**(19), 5984-6003

36. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) *J Magn Reson* **97**(2), 359–375

Molecular basis of altered DNA recognition by the androgen receptor DNA-binding domain

Supplementary Material

Supplementary Figure 1

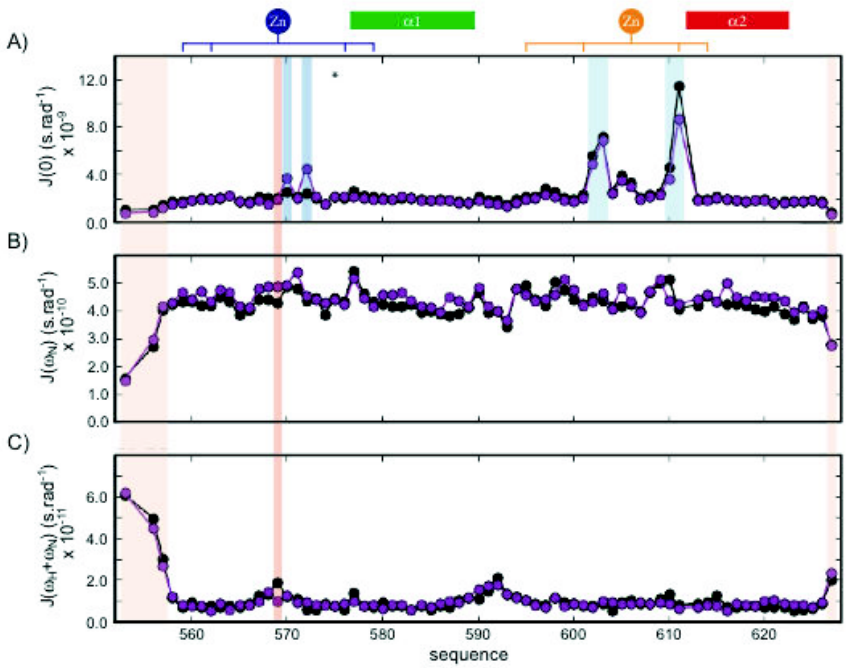


Figure S1. Spectral density values, J(ω), at frequencies A) 0, B) ω_N and C) $\omega_H+\omega_N$, plotted against sequence. J(ω) values were determined from relaxation measurements for backbone ^{15}N nuclei in AR DBD-WT (black) and AR DBD-T575A (magenta). Light pink shading indicates residues affected by ps-ns time-scale motions in both domains; dark pink shading indicates residues affected only in one domain. Light blue shading indicates residues affected by µs-ms time-scale conformational exchange in both domains; dark blue shading indicates residues affected only in AR DBD-T575A. An asterisk marks the site of mutation.

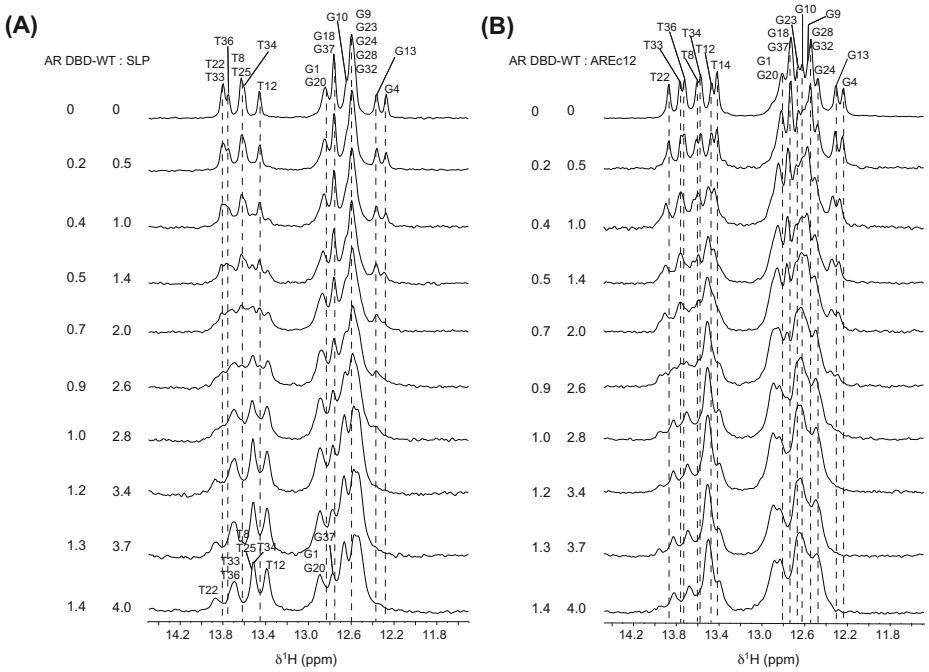


Figure S2. Imino-proton 1D spectra of the SLP (A) and AREc12 (B) oligonucleotides upon addition of increasing amounts of wild-type AR DBD.

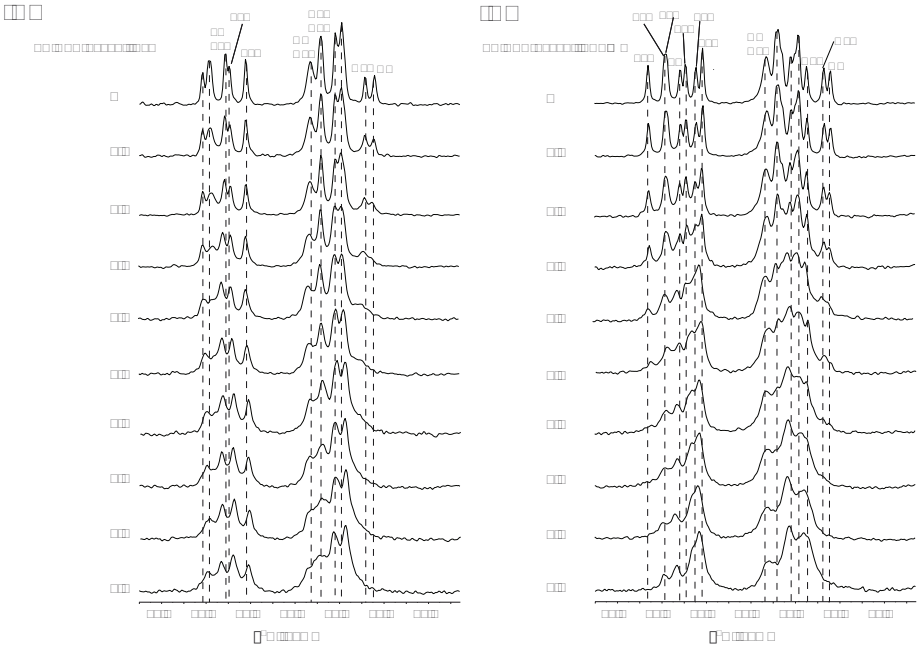


Figure S3. Imino-proton 1D spectra of the SLP (A) and AREc12 (B) oligonucleotides upon addition of increasing amounts of T575A AR DBD.

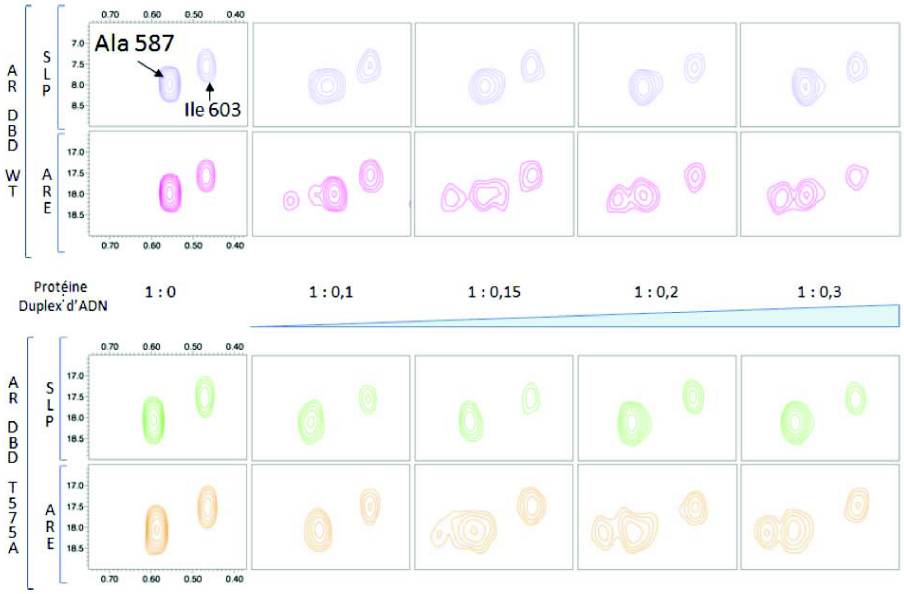


Figure S4. Evolution of Ala587 and Ile603 methyl correlations for increasing amounts of DNA (indicated in the middle). Titration experiments between the wild-type DBD and SLP or AREc12 oligonucleotides are shown in violet and pink, respectively. Titration between the T575A DBD and SLP or AREc12 oligonucleotides are shown in green and orange, respectively.

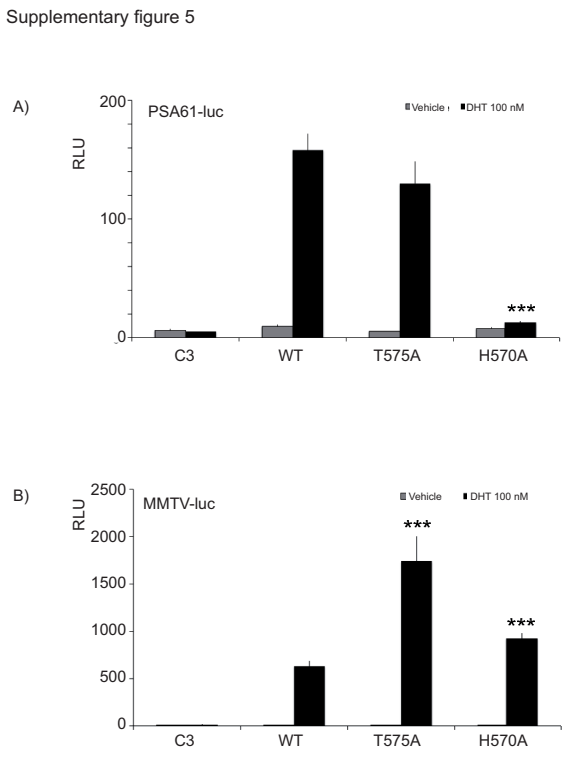


Figure S5. Luciferase reporter assay for AR-WT, AR-T575A and AR-H570A. The H570A mutation orientates androgen receptor transactivation towards the non-specific responsive promoter. LNCaP cells were cotransfected with AR DBD-WT or the indicated mutant expression plasmid together with PSA61-Luc (A) or MMTV-luc (B) firefly luciferase construct and pGL4.70[hRLuc] (Renilla) for transfection control. Cells were then treated with 100nM DHT or vehicle in serum-free medium for 48 hours. Luminescence (relative light units) was normalized to the Renilla luciferase expression from the pGL4.70[hRLuc]. P values of less than 0.05 were considered significant (***).

Chapitre 6

Analyse bio-informatique de la séquence protéique humaine de AR-NTD

6.1 Prédiction ordre/désordre dans le NTD du récepteur des androgènes . 132

6.1.1 metaPrDOS 132

6.1.2 Analyse des clusters hydrophobes (HCA) 133

6.2 Conservation de séquence 135

6.3 Analyse des motifs d'interaction dans AR-NTD 137

6.4 Production et purification 139

6.5 Conclusion 143

Bibliographie 144

Le domaine N-terminal des RNs est capable d'activer l'expression de gènes indépendamment de la présence du ligand. Cette propriété explique l'émergence de variants du récepteur des androgènes dépourvus de LBD au cours de la progression du cancer de la prostate qui rendent inopérantes les thérapies basées sur la privation hormonale. [Bergerat 2009]. McEwan et coll. font partis des premiers à avoir travaillé sur les régions désordonnées des RNs, particulièrement le domaine NTD, comme les récepteur des androgènes, le récepteur des minéralocorticoïdes et le récepteur des glucocorticoïdes [Lavery 2005, Fischer 2010, McEwan 2012]. Le NTD de AR est considéré comme intrinsèquement désordonné du fait de la faible complexité de séquence, le ratio élevé de résidus hydrophyles/hydrophobes et la présence d'homopolymères comme le polyglutamine, polyglycine et polyproline. La raison fonctionnelle de ces particularités reste encore inconnue. Ces propriétés font qu'il est difficile d'exprimer cette région entière en quantité suffisante pour des analyses structurales et fonctionnelles et il est nécessaire de développer de nouvelles approches pour établir des relations séquence/fonction. Pour cela, nous avons analysé la séquence primaire de AR-NTD pour identifier des régions particulières.

6.1 Prédiction ordre/désordre dans le NTD du récepteur des androgènes

6.1.1 metaPrDOS

L'analyse bioinformatique de AR-NTD est basée sur des prédictions de désordre et la conservation au sein des différents espèces.

Une prédiction de désordre a été réalisée de AR-NTD à l'aide du métaserveur metaPrDOS (Figure 6.1) [Ishida 2008]. Cette analyse affecte un score variant de 0 à 1 pour les régions prédites respectivement complètement ordonnées à complètement désordonnées.

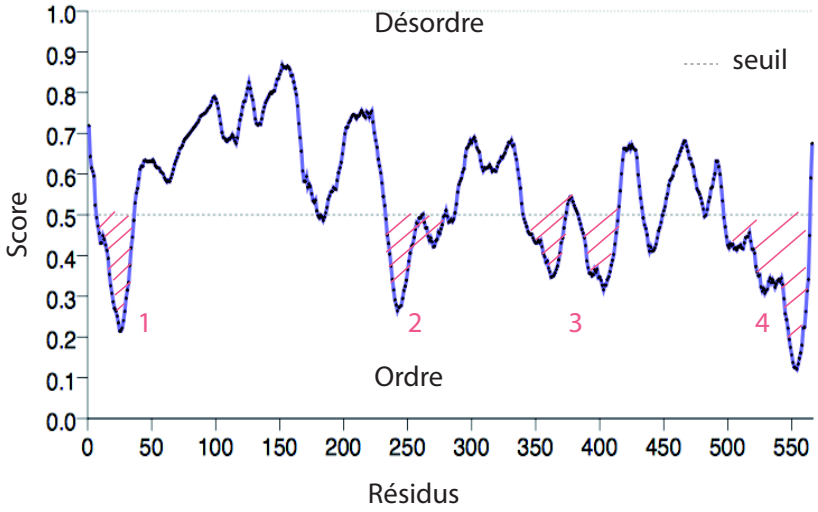


Figure 6.1 – Prédiction de désordre de la région NTD du récepteur des androgènes obtenue avec le métaserveur metaPrDOS. Les régions sont prédites désordonnées pour des scores se rapprochant de 1 et ordonnées pour des scores se rapprochant de 0 avec un seuil à 0,5. La numérotation des résidus correspond à la séquence protéique humaine de AR.

A partir de la Figure 6.1, nous pouvons définir deux grandes régions : un première région très désordonnée de façon homogène dont les bornes sont les résidus [50-200], puis le reste de la séquence qui est caractérisée par une alternance plus fréquente entre des régions plus ordonnées et désordonnées. Nous constatons également que quatre régions (que nous définirons région 1, 2, 3 et 4) présentent un degré d'ordre plus élevé.

6.1.2 Analyse des clusters hydrophobes (HCA)

L'analyse issue des métaserveurs de désordre peut être affinée en examinant la distribution des résidus hydrophobes le long de la séquence. Le métaserveur meDor [Lieutaud 2008] fournit, de haut en bas, une prédiction de structure secondaire, la structure protéique numérotée sur une ligne, un tracé HCA (Hydrophobic Cluster Analysis) ainsi que les résultats de plusieurs prédicteurs de régions désordonnées qui ont été interrogés.

Le tracé HCA permet d'apprécier de façon visuelle la présence de motifs réguliers le long de la séquence primaire. La séquence est organisée de façon à faire apparaître une organisation périodique de motifs hydrophobes caractéristiques des hélices amphiphiles (périodicité 3-4) ou de feuillet. Cette séquence est organisée en segments diagonaux alternativement de 7 ou 8 résidus pour lesquels 3 ou 4 résidus sont répétés sur le segments suivant (cf. schéma d'aide à la lecture, Figure 6.2).

Les résultats issus de cette analyse pour le domaine NTD sont présentés en Figure 6.2.

Nous retrouvons les deux régions identifiées avec metaPrDOS, à savoir une région prédite désordonnée de façon homogène entre les résidus 50-200 ainsi que le reste de la séquence où la prédiction de régions désordonnées et ordonnées est alternée. Les régions 1 à 4 identifiées précédemment sont replacées sur la Figure. La présence des homopolymères de polyQ (Q21 et Q6, Q5), polyP (P8), polyA (A5) et polyG (G23) est également indiquée. Les régions non identifiées précédemment sont nommées A, B et C.

La région 1 est riche en prolines et en résidus hydrophobes. Dans cette région, l'organisation périodique de résidus hydrophobes dans la région [20-35] peut correspondre à une structure en hélice- De plus, cette région est flanquée par deux zones enrichies en prolines, qui sont des résidus casseurs d'hélice, puis suivie par le polyQ.

Les régions A [160-200], B [280-330] et C [380-450] se ressemblent beaucoup de par leur richesse en résidus proline et glycine caractéristique des régions désordonnées.

De plus, nous constatons que les prédicteurs indiquent les régions désordonnées de façon cohérente.

Les résidus appartenant à la région 2 sont prédits ordonnés et présentent une période de deux caractéristiques d'un feuillet- . La régions 3 contient des résidus hydrophobes (leucine, tyrosine, phénylalanine) qui pourraient s'organiser en feuillet- ou hélice- d'après le prédicteur de structure secondaire. Enfin, nous observons que la régions 4 est riche en résidus hydrophobes et aromatiques (methionine, valine, tyrosine, tryptophane, leucine et isoleucine) habituellement absents des régions désordonnées. Bien que cette région soit prédite désordonnée par l'un des prédicteurs, il est possible que ces résidus soient impliqués dans une structure secondaire.

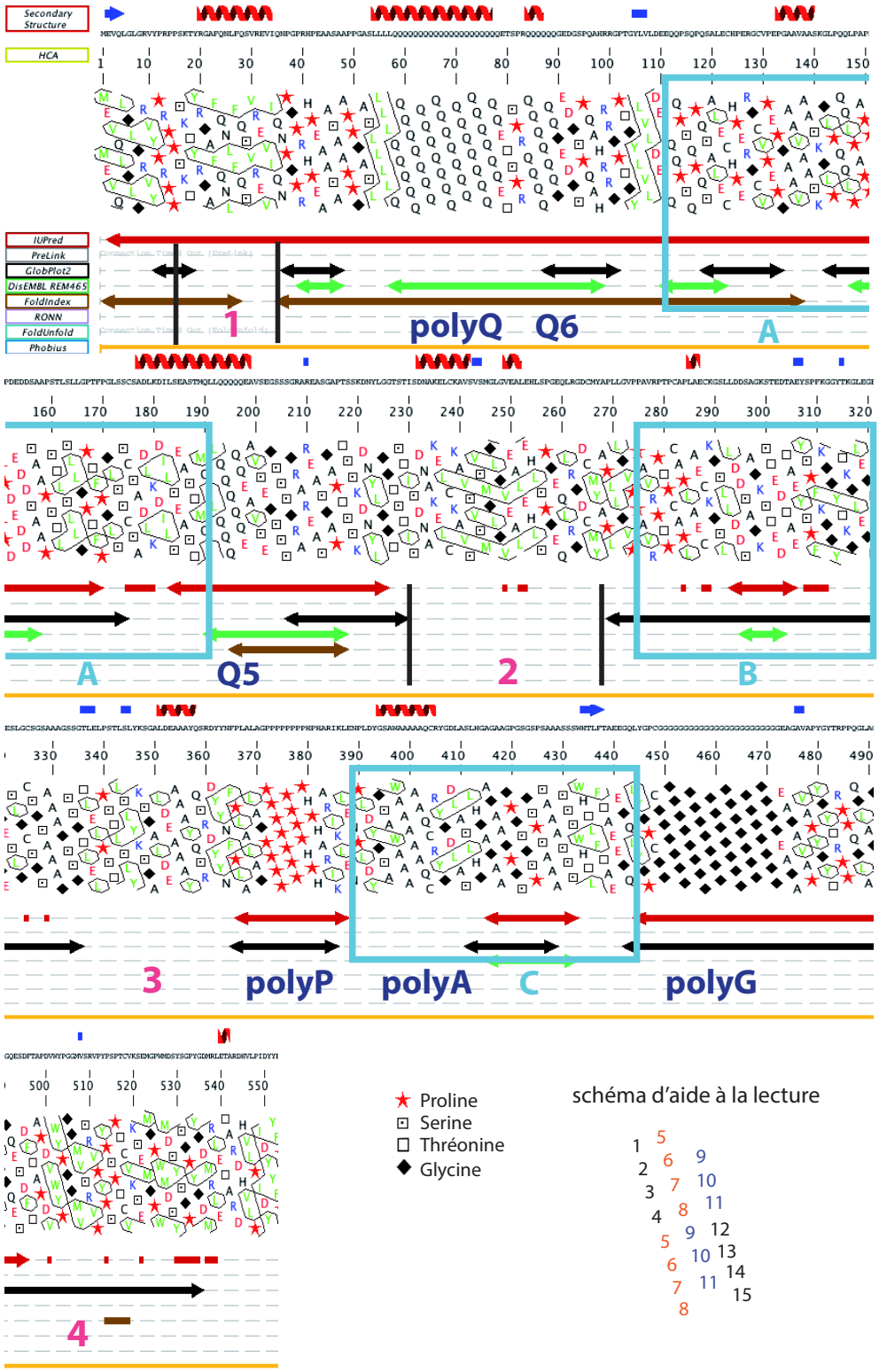


Figure 6.2 – Analyse bioinformatique réalisée avec MeDor. La figure montre la prédiction de structure secondaire le long de la séquence protéique de AR-NTD, un tracé HCA permettant l'identification de zones contenant des résidus hydrophobes habituellement impliqués dans des structures secondaires ainsi que la prédiction des régions désordonnées par les différents prédicteurs interrogés.

6.2 Conservation de séquence

Une première analyse du profil de conservation de séquence met en évidence deux régions identifiées précédemment dans AR-NTD : la région 2 conservée à hauteur de 100% et la région 4, située à l'extrémité C-terminale du NTD, conservée à hauteur de 75% (Figure 6.3).

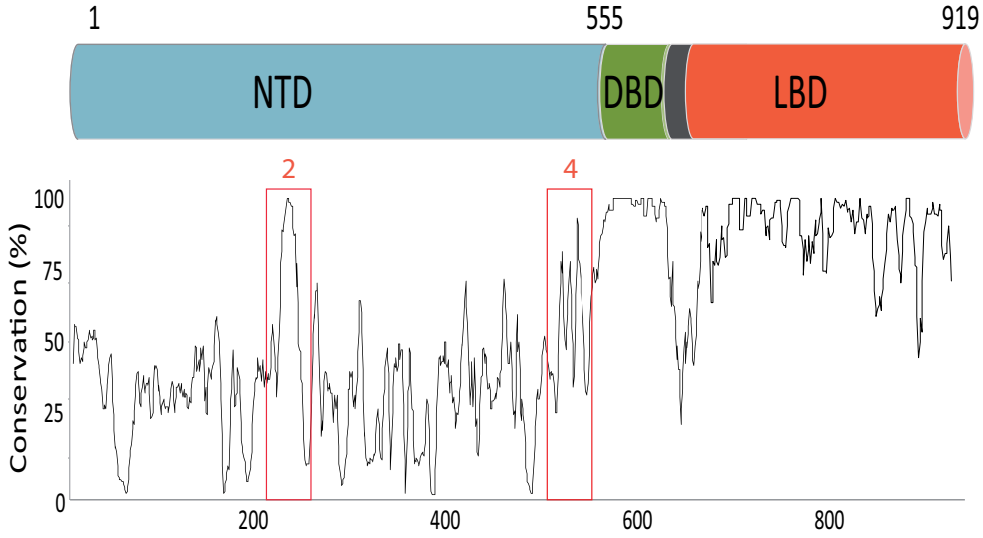


Figure 6.3 – Représentation quantitative de la conservation de séquence primaire des domaines de AR parmi 12 vertébrés. Le DBD et LBD ainsi que 2 courtes régions du NTD sont extrêmement conservés et montrent un taux de conservation proche de 100%. Figure extraite de [Asencio-Hernández 2014].

L'alignement des régions 2 et 4 parmi 12 vertébrés est présenté en Figure 6.4.

Nous observons un bloc d'une vingtaine de résidus conservés appartenant à la région 2. De plus, nous constatons que les charges négatives sont également conservées, de même que la conservation de séquence, comme par exemple les positions 252 et 255 acceptant une variation du résidu glutamate en aspartate (Figure 6.4). Le résidu C240 qui est strictement conservé au sein de cette région est impliqué dans un processus d'oligomérisation formant des fibres amyloïdes [Asencio-Hernández 2014].

La région 4 se compose d'une cinquantaine de résidus qui se distinguent par une succession de 4 blocs conservés, mais plus court, 5-7 résidus. Les charges négatives et positives sont conservées aussi bien que le type de résidu. De plus, nous observons une alternance de ces charges et la présence de trois tryptophanes dans cette région.

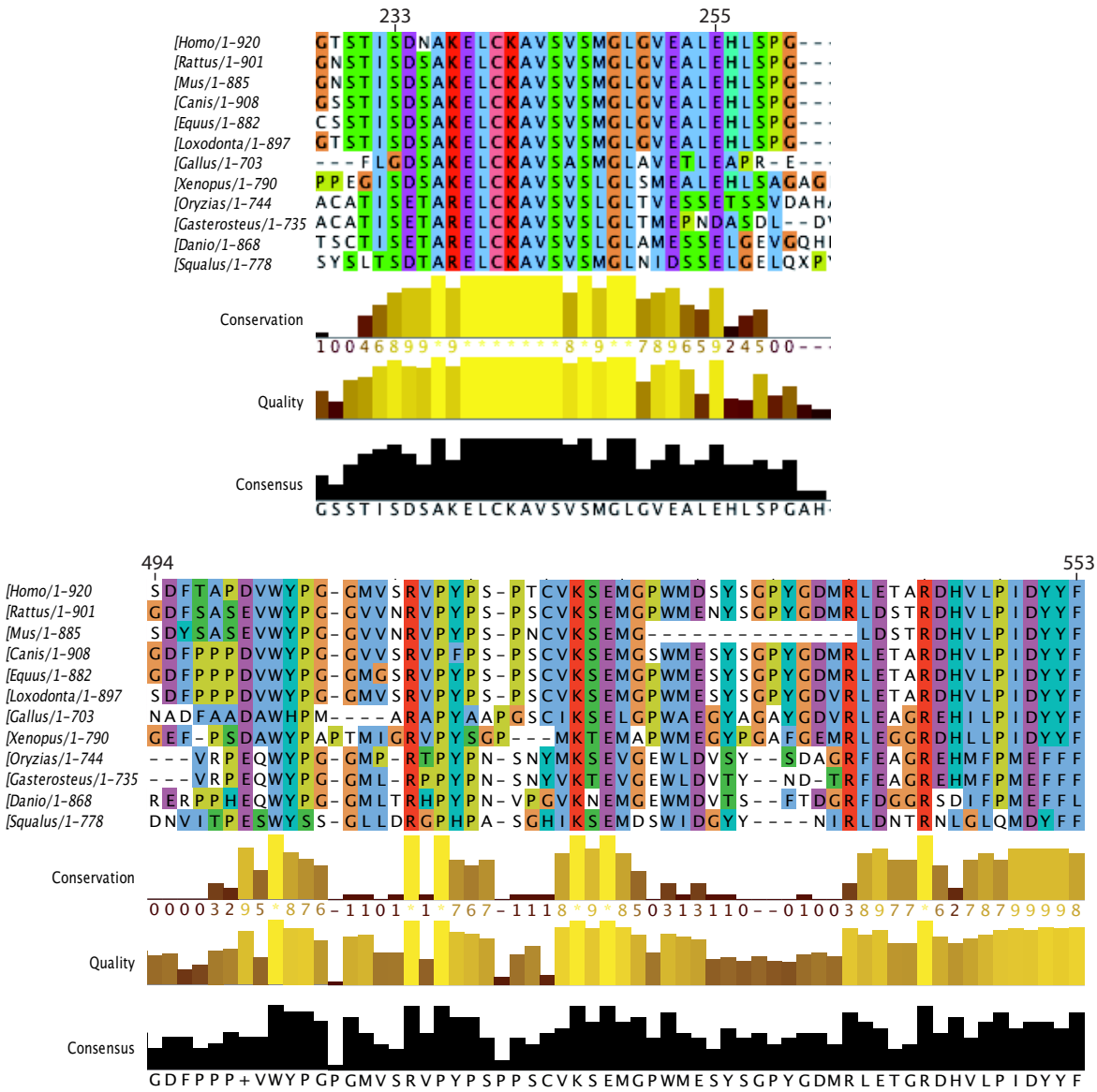


Figure 6.4 – Alignement des séquences conservées de AR-NTD parmi 12 vertébrés. La numérotation des résidus correspond à la séquence protéique humaine de AR. D'après [Asencio-Hernández 2014].

Par ailleurs, la présence de régions de faible conservation au sein du NTD doit être analysée avec précaution. En effet, une région-non conservée peut correspondre à des variations dans le positionnement d'éléments de séquence (comme le polyQ) ou des variations de leur longueur (Figure 6.5).

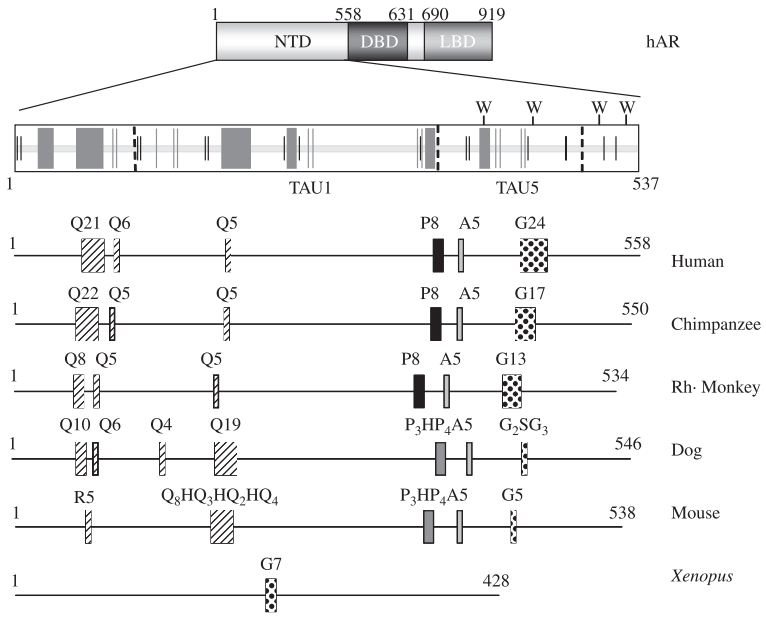


Figure 6.5 – Disposition des homopolymères dans AR-NTD parmi l'Homme, le Chimpanzé, le Singe, le Chien, la Souris et le Xénope. D'après [Davies 2008].

La Figure 6.5 correspond à une représentation schématique de AR-NTD montrant les positions des différents homopolymères présents dans ce domaine. Ainsi, les positions des polyQ sont conservées chez l'homme, le chimpanzé et le singe alors qu'elles sont inversées chez le chien et la souris. La présence de ce domaine est donc importante mais sa position l'est moins. Par contre, nous constatons que cet homopolymère est toujours positionné dans la première moitié de AR-NTD.

6.3 Analyse des motifs d interaction dans AR-NTD

La présence de sites de reconnaissance et d'interactions moléculaires est également prise en compte dans l'analyse bioinformatique. Ceci a été réalisé à l'aide de l'outil de prédiction Eukaryotic Linear Motif (ELM) [Dinkel 2014] permettant de repérer de petits motifs tels que les SLiMs ou MoRFs. Ces motifs peuvent être impliqués dans le repliement après interaction de la région ou dans des modifications post-traductionnelles (PTMs). Par exemple, des motifs du type LxxLL, WxxLF ou FxxLF sont présents dans le domaine AR-NTD. Comme mentionné dans le chapitre 2, ces motifs sont présents au niveau de co-régulateurs de AR. De plus, les motifs FxxLF et WxxLF sont impliqués dans l'interaction NTD/LBD de AR [He 2000].

A cela s'ajoute la présence ou non de mutations dans l'AR-NTD et leurs liens avec des

pathologies décrites dans la littérature. L'ensemble des mutations recensées pour l'AR se trouve sur la base de données androgendb [Gottlieb 2004, Gottlieb 2012] à l'adresse url http://androgendb.mcgill.ca/.

Les différents motifs d'interactions, les PTMs ainsi que les mutations identifiées dans des échantillons de cancer de la prostate sont indiqués dans le tableau 6.1.

La combinaison de l'ensemble des ces résultats nous a mené à concentrer nos efforts sur la région C-terminale de AR-NTD qui est prédite ordonnée, conservée parmi les vertébrés et contenant plusieurs sites d'interactions protéines/protéines ainsi que des sites de PTMs.

La stratégie utilisée est de produire différents sous-domaines (au nombre de 5) contenant chacun un résidu tryptophane supplémentaire servant d'identifiant, et qui seront produits en fusion ou non avec le DBD.

La fusion du DBD offre la possibilité d'établir un lien entre le domaine NTD et la liaison à l'ADN ainsi qu'un lien sur une éventuelle communication interdomaine NTD/DBD. De plus, la présence du DBD nous sert de sonde RMN puisque ce domaine est bien étudié au laboratoire.

Régions	Motifs d'interaction/Résidus hydrophobes, PTMs/Mutations		
Région 1 [1-150]	FQNLF (FxxLF)	(Y, F, L, V, M, I) S16/S81/S94 (Phosphorylation)	E43G/V / Q58L S121S
Région A [160-200]	LKDIL/TMQLL (LxxLL)	(L, M, F, V))	L194/F/R
Région 2 [235-280]		(L, V, M, I, L) alternance hydrophobe/hydrophile S256/Y267 (Phosphorylation)	T229C/A /A253V Q255K
Région B [280-330]	PLLGVPPA GESLG	(Y, F, M) T282/S293/S308 (Phosphorylation)	
Région 3 [330-380]		(Y, L, F) polyP Y363 (Phosphorylation)	P342L/S343* A358V/T/R362H
Région C [380-450]	WHTLF (WxxLF) GPGSGSPS	(Y, F, L, I, W) polyA S424/S426 (Phosphorylation) K386 (Sumoylation)	G416S/D/ P392L W435L/C/T440P
Région 4 [480-554]	SDFTAP PYSPT/LICGD	(V, W, Y, F, M) S515/K520/Y534 (Phosphorylation) K520 (Sumoylation)	R485C/T/ T498I A499T/V /V509L/G P515S/G525S/D D529G/P534S

Tableau 6.1 – Résumé des motifs d'interaction, résidus hydrophobes, modifications post-traductionnelles (PTMs) et mutations identifiées dans des échantillons de cancer de la prostate pour les différentes régions de AR-NTD identifiées à l'aide du metaserveur MeDor.

La Figure 6.6 présente les 7 polypeptides étudiés ainsi que la nomenclature correspondante utilisée tout au long de ce manuscrit, à savoir W1, W1-DBD, W2, W2-DBD, W3-DBD et W4-DBD.

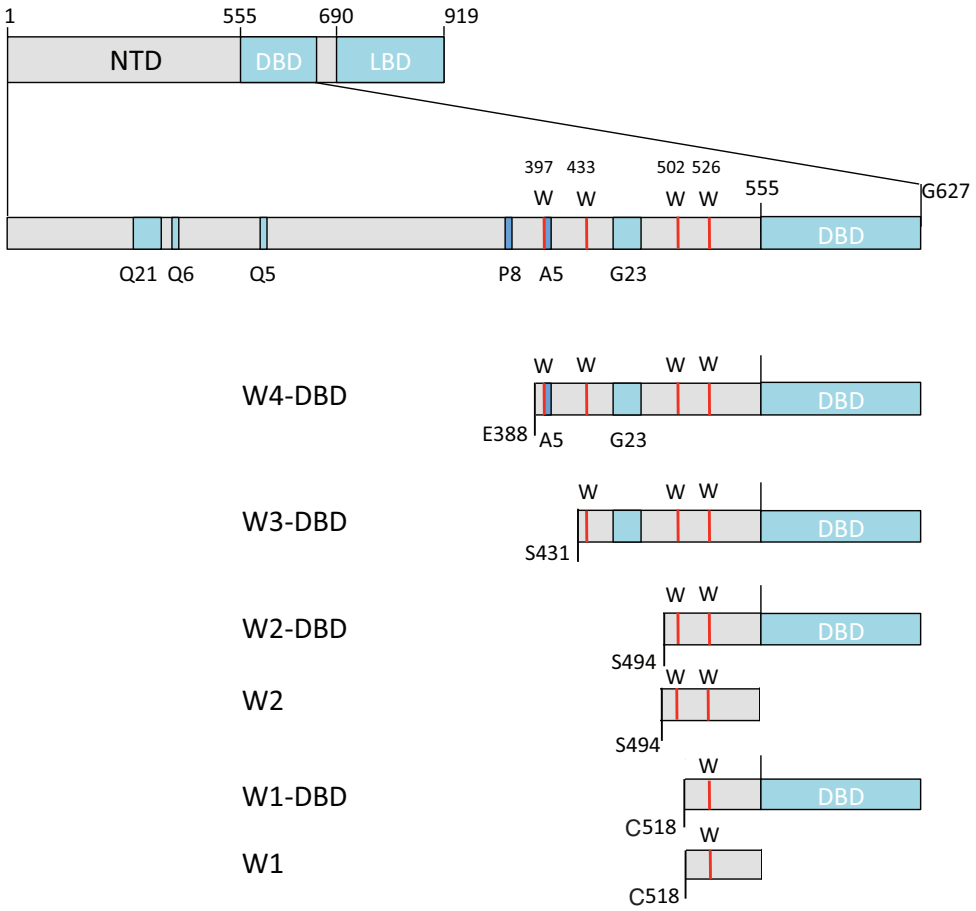


Figure 6.6 – Carte des 5 régions présentant des caractéristiques de séquence particulières au sein du NTD identifiées à l'aide des outils bio-informatiques metaPrDOS, MeDor et ELM.

6.4 Production et puri cation

L'étape suivante vise à définir la possibilité de travailler en solution avec ces sous-domaines par des études biochimiques.

Les différents sous-domaines ont été clonés en fusion avec une étiquette histidine suivie d'une protéine étiquette GST ou MBP et d'un site de clivage à la TEV situé en position N-terminale. Cette approche permet de purifier les protéines recombinantes produites en deux étapes avec une grande efficacité, d'une part, et d'utiliser l'étiquette de fusion pour limiter

les phénomènes d'agrégation. Nous avons enfin systématiquement cherché à caractériser une différence entre des fragments du NTD produits de façon isolée et ces mêmes fragments produits en fusion avec le domaine de liaison à l'ADN (DBD) du récepteur.

Afin de cribler différentes conditions d'expression pouvant conduire à l'obtention de fragments solubles du NTD, l'expression systématique de ces fragments ainsi que les premières étapes de purification (purification des protéines recombinantes) ont été réalisées en parallèle dans des plaques 24 puits avec un volume de culture de 2ml.

La Figure 6.7 montre que les taux d'expression sont meilleurs pour les fusions MBP et que tous les fragments conduisent à l'obtention d'une protéine soluble.

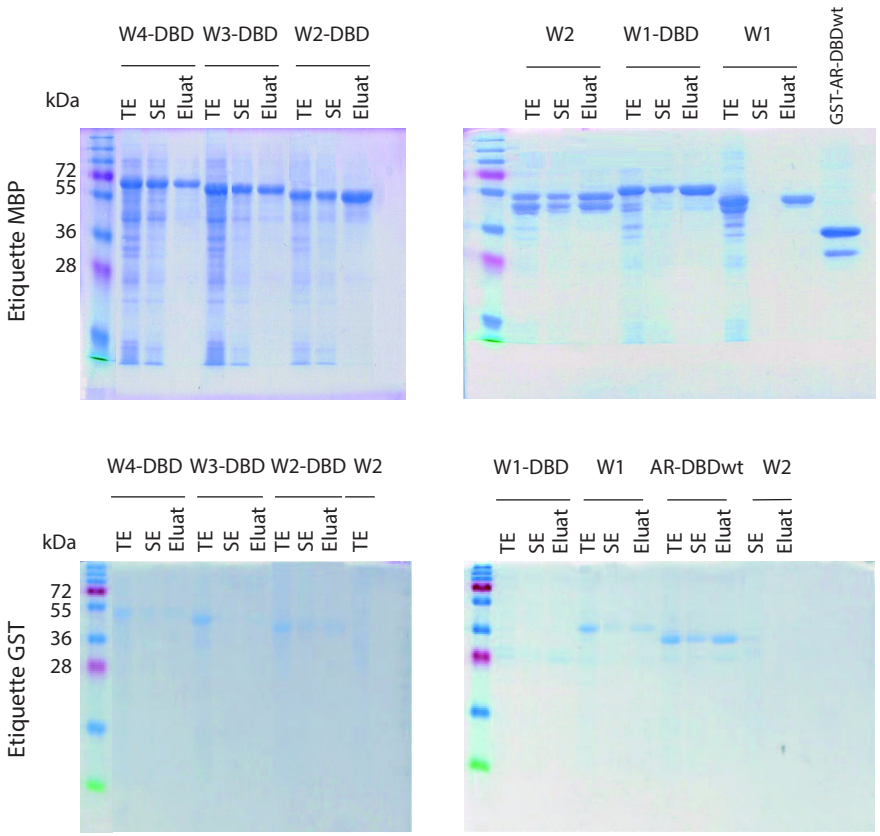


Figure 6.7 – Criblage des conditions d'expression des différents polypeptides.

Ainsi, à partir des polypeptides en fusion avec l'étiquette MBP, le protocole d'expression a été optimisé pour la production de fragments dans un volume de culture d'1l en vue des analyses biophysiques. Le protocole est détaillé en annexe de ce chapitre. L'étape d'ultracentrifugation sur la nuit, pour les échantillons W4-DBD, W3-DBD, W2-DBD et W1-DBD, est cruciale pour éliminer les agrégats solubles qui pourraient recruter la protéine soluble et entraîner son

agrégation (Figure 6.9). Ce faisant, la concentration de protéine soluble diminue d'une part, et d'autre part la polydispersité des échantillons serait un biais pour les études biophysiques.

Une étape supplémentaire dans le protocole de purification est ajoutée pour les fragments W3-DBD et W4-DBD. En effet, les volumes d'élution de ces fragments sont proches du volume d'élution de l'étiquette MBP rendant alors difficile la séparation de ces protéines sur la colonne de gel filtration utilisée. Ainsi, une purification sur colonne héparine présentant une affinité pour les domaines de liaison à l'ADN permet de séparer nos fragments d'intérêts retenus sur la colonne alors que la MBP passe au travers de la résine.

Les profils d'élution issus de l'étape de gel filtration pour les fragments W1, W1-DBD, W2, W2-DBD et DBD-wt ainsi que le profil d'élution correspondant au fragment W3-DBD ou W4-DBD obtenu après purification sur colonne héparine sont présenté en Figure 6.8.

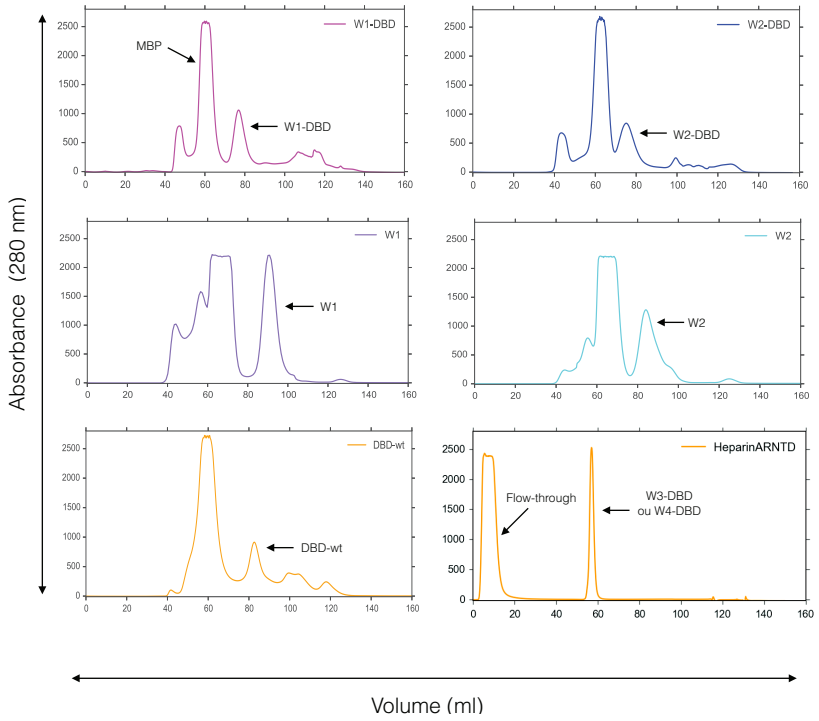


Figure 6.8 – Profils d'élution issus de l'étape de gel filtration pour les fragments W1, W1-DBD, W2, W2-DBD et DBD-wt et profil d'élution correspondant au fragment W3-DBD ou W4-DBD obtenu après purification sur colonne héparine.

Le tableau 6.2 résume les quantités moyennes obtenues pour chaque polypeptide à partir d'1L de culture.

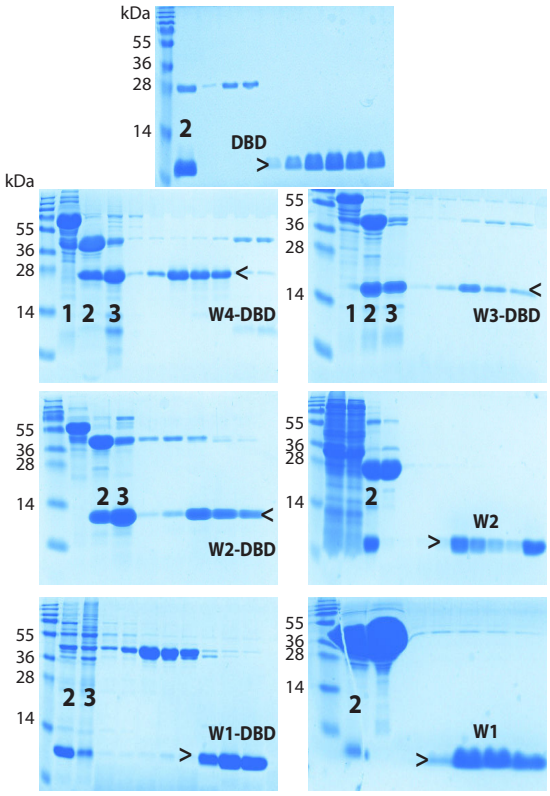


Figure 6.9 – Etapes de purification des différents polypeptides en fusion avec l'étiquette MBP. La piste 1 correspond à la protéine recombinante ; 2 correspond au clivage par l'enzyme TEV et 3 à l'étape d'ultracentrifugation. Les noms des différents polypeptides sont indiqués sur les gels SDS-PAGE au niveau des pistes correspondantes à l'étape de gel filtration.

Polypeptide	Taux d'expression (mg/L de culture)
DBD-wt	7,0
W4-DBD	1,2
W3-DBD	$\simeq$1
W2-DBD	1,8
W2	4,6
W1-DBD	4,6
W1	7,36

Tableau 6.2 – Tableau des quantités de protéines (en mg/L de culture) obtenues après purification de chaque échantillon.

6.5 Conclusion

Nous avons sélectionné cinq sous-régions situées au niveau de l'extrémités C-terminale du domaine NTD de AR. Ces régions sont la cibles de mutations associées au cancer de la prostate et présentent des sites de modifications post-traductionnelles. De plus, elles ne présentent pas toutes les caractéristiques des protéines intrinsèquement désordonnées puisqu'elles sont notamment enrichies en résidus hydrophobes.

Notre capacité à produire ces différentes régions en quantité suffisante nous a permis des les caractériser par des approches biophysiques dont les résultats sont présentés dans les chapitres suivants.

Bibliographie

[Asencio-Hernández 2014] Asencio-Hernández J., Ruhlmann C., McEwen A., Eberling P., Nominé Y., Céraline J., Starck J.P. et Delsuc M.A. *Reversible Amyloid Fiber Formation in the N Terminus of Androgen Receptor*. ChemBioChem, vol. 15, no. 16, 2014. – Cité pages 135 et 136.

[Bergerat 2009] Bergerat J.P. et C raline J. *Pleiotropic functional properties of androgen receptor mutants in prostate cancer*. Human Mutation, vol. 30, no. 2, 2009. – Cité pages 24 et 131.

[Davies 2008] Davies P., Watt K., Kelly S.M., Clark C., Price N.C. et McEwan I.J. *Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain*. Journal of Molecular Endocrinology, vol. 41, no. 5, 2008. – Cité pages 15, 25 et 137.

[Dinkel 2014] Dinkel H., Van Roey K., Michael S., Davey N.E., Weatheritt R.J., Born D., Speck T., Krüger D., Grebnev G., Kuban M., Strumillo M., Uyar B., Budd A., Altenberg B., Seiler M., Chemes L.B., Glavina J., Sánchez I.E., Diella F. et Gibson T.J. *The eukaryotic linear motif resource ELM : 10 years and counting.* Nucleic Acids Research, vol. 42, no. Database issue, 2014. – Cité pages 66 et 137.

[Fischer 2010] Fischer K., Kelly S.M., Watt K., Price N.C. et McEwan I.J. *Conformation of the Mineralocorticoid Receptor N-terminal Domain : Evidence for Induced and Stable Structure*. Molecular Endocrinology, vol. 24, no. 10, 2010. – Cité page 131.

[Gottlieb 2004] Gottlieb B., Beitel L.K., Wu J.H. et Trifiro M. *The androgen receptor gene mutations database (ARDB) : 2004 update*. Human Mutation, vol. 23, no. 6, pages 527–533, 2004. – Cité page 138.

[Gottlieb 2012] Gottlieb B., Beitel L.K., Nadarajah A., Paliouras M. et Trifiro M. *The androgen receptor gene mutations database : 2012 update*. Human Mutation, vol. 33, no. 5, 2012. – Cité pages 24 et 138.

[He 2000] He B., Kemppainen J.A. et Wilson E.M. *FXXLF and WXXLF Sequences Mediate the NH2-terminal Interaction with the Ligand Binding Domain of the Androgen Receptor*. Journal of Biological Chemistry, vol. 275, no. 30, 2000. – Cité page 137.

[Ishida 2008] Ishida T. et Kinoshita K. *Prediction of disordered regions in proteins based on the meta approach*. Bioinformatics, vol. 24, no. 11, 2008. – Cité pages 65 et 132.

[Lavery 2005] Lavery D.N. et McEwan I.J. *Structure and function of steroid receptor AF1 transactivation domains : induction of active conformations*. Biochemical Journal, vol. 391, no. 3, 2005. – Cité page 131.

[Lieutaud 2008] Lieutaud P., Canard B. et Longhi S. *MeDor : a metaserver for predicting protein disorder*. BMC Genomics, vol. 9, no. Suppl 2, pages S25–5, 2008. – Cité pages 65, 66 et 133.

[McEwan 2012] McEwan I.J. *Intrinsic disorder in the androgen receptor : identi cation, characterisation and drugability*. Molecular BioSystems., vol. 8, no. 1, pages 82–90, 2012. – Cité pages 15, 26, 131 et 170.

Annexe A

A.1 Analyse bioinformatique

La prédiction de structure a été réalisée avec les métaserveurs MeDor et metaPrDOS. Ces derniers interrogent plusieurs prédicteurs et combinent les différentes informations pour fournir une prédiction de structure au sein d'une séquence protéique.

Prédicteurs interrogés	MeDOr	metaPrDOS	référence
IUPred	X	X	[Dosztányi 2005]
PrDOS		X	[Ishida 2007]
DisEMBL ^{TM}	X	X	[Linding 2003b]
DISPROT (VSL2)	X	X	[Peng 2006]
DISOPRED2		X	[Ward 2004]
POODLE-S		X	[Shimizu 2007]
DISOclust		X	
DISpro		X	[Cheng 2005]
Prelink	X		[Coeytaux 2005]
RONN	X		[Yang 2005]
FlodUnfold	X		[Garbuzynskiy 2008]
FoldIndex	X		[Prilusky 2005]
GlobPlot2	X		[Linding 2003a]
DISPROT VL3 et VL3H	X		[Obradovic 2003]
Phobius	X		[Käll 2004]

A.2 Production et puri cation des sous-domaines de AR-NTD

Plasmides

Les ADNc des sous-domaines de AR-NTD avec ou sans fusion au DBD (W4-DBD [388-627] ; W3-DBD [431-627] ; W2-DBD [494-627] ; W2 [494-554] ; W1-DBD [518-627] et W1

[518-554]) et de AR-DBD [555-627] du gène humain du récepteur des androgènes ont été clonés entre les sites de restriction BamHI/XhoI des vecteurs d'expression pETM41 et pETM30 respectivement.

Ces vecteurs permettent l'expression d'une protéine recombinante composée de la protéine d'intérêt en fusion avec une double étiquette de purification située en N-terminal de la protéine d'intérêt, une étiquette poly-His suivie de la MBP (Maltose Binding Protein) ou de la GST (Glutathione-S-Transférase) pour le vecteur pETM41 et pETM30 respectivement. La présence d'un site de clivage à la protéase TEV permet d'éliminer les étiquettes de purification et d'obtenir uniquement la protéine d'intérêt.

Expression et puri cation

Les souches bactériennes *E. coli* BL21 (DE3) et BL21 (DE3) Star ont été transformées avec les plasmides recombinant AR-NTDs et DBD respectivement. Les cellules sont cultivées en milieu riche LB (en absence de marquage), ou en milieu minimum M9 enrichi en isotope ^{15}N (lors du marquage isotopique des protéines), à 37 °C jusqu'à obtenir une DO_{600nm} de 0,6.

L'expression de la protéine est induite en ajoutant 1 mM d'IPTG dans le milieu de culture, et se poursuit sous agitation à 16 °C (polypeptides AR-NTDs) ou 18 °C (AR-DBD) sur la nuit. Le lendemain, après une centrifugation de 30 minutes / 3500 rpm / 4 °C les culots bactériens sont suspendus dans le tampon de lyse (50 mM Tris pH 7.5, 150mM NaCl, 10% Glycérol, 1 mM dithithreitol (DTT), 2.5 U mL DNAse I, 2.5 g/mL RNAse A et un comprimé d'inhibiteur de protéase EDTA-free (protease inhibitor cocktail, Roche)).

Les cellules sont lysées à l'aide de 6 cycles de sonication d'une minute sur un Branson Digital Sonifier. Le lysat cellulaire est ultracentrifugé 1 h / 36000 rpm / 4 °C, puis la fraction soluble est filtrée sur une membrane de 0,20 m Minisart High Flow, (Sartorius Stedium Biotech), et la protéine recombinante est purifiée par chromatographie d'affinité à partir de la fraction soluble. La colonne d'affinité contenant 30 mL de résine Amylose (New England Biolabs) ou 10 ml de résine Glutathione reduced est préalablement équilibrée à l'aide du tampon d'équilibrage (50 mM Tris pH 7.5, 150 mM NaCl, 10% Glycérol, 1 mM dithiothreitol (DTT), 2.5 U mL DNAse I, 2.5 /mL RNAse A et un comprimé d'inhibiteur de protéase EDTA-free (protease inhibitor cocktail, Roche)). La protéine recombinante liée à la résine est ensuite lavée avec le tampon de lavage (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT) puis éluée avec ce même tampon supplémenté de 20 mM de maltose ou 10 mM de Glutathione. L'étiquette His_6-GST/ His_6-MBP est ensuite clivée par la protéase TEV pendant une nuit à 4 °C.

Une étape de purification intermédiaire consistant en une chromatographie d'affinité au nickel a été réalisée pour les fragments en fusion au DBD. Un colonne contenant 4ml de résine

Ni-NTA a été préalablement équilibrée avec le tampon de lavage cité ci-dessus. Puis la protéine dépourvue de son étiquette MBP est passée à travers la colonne et récupérée dans le flow-through . L'étiquette MBP et la protéase TEV sont retenues sur la résine. Cette étape est répétée trois fois pour éliminer au maximum l'étiquette MBP.

Puis les échantillons sont concentrés à 2 mL et purifiés par chromatographie d'exclusion de taille sur une colonne de gel filtration Superdex G75 (HiLoad 16/60) (Amersham Biosciences). La colonne est pré-équilibrée avec le tampon de gel filtration (50 mM Na3PO4 pH 6.5, 150 mM NaCl, 1 mM DTT). Les fractions correspondant à la protéine d'intérêt sont déposées sur un gel 15% SDS-PAGE puis rassemblées.

Une étape de purification supplémentaire sur colonne héparine est réalisée pour les fragment W2-DBD et W3-DBD puisque ceux-ci sont élués au même volume que la MBP au moment de l'étape de gel filtration. La résine héparine (HiTrap Heparin HP 5ml) est bien adaptée à nos fragments puisqu'elle permet de retenir les protéines interagissant avec l'ADN. La protéine retenue sur la résine est éluée à l'aide d'un gradient de NaCl allant de 50 mM à 1 M. Après purification, la protéine est dialysée dans le tampon de gel filtration (50 mM Na Phosphate, pH 6.5, 150 mM NaCl et 1 mM DTT)

Enfin l'échantillon est concentré à 0,5-1 mL dans un concentrateur Vivaspin ayant un cut-off dont la taille correspond à la moitié du poids moléculaire de la protéine en question. Pour les polypeptides AR-NTDs qui contiennent des W dans leur séquence primaire, les concentrations finales des échantillons ont été déterminées par mesure de l'absorbance à 280 nm à l'aide du spectromètre NanoDrop 2000 (Thermo Fisher Scientific).

Pour AR-DBD ne contenant aucun tryptophane, la concentration est déterminée par RMN par quantification externe en appliquant la méthode PULCON (PUlse Length CONcentration determination) [Wider 2006]. Cette méthode permet de déterminer une concentration grâce à la comparaison des spectres RMN, enregistrés sur un échantillon de tryptophane dont la concentration est connue avec précision et sur l'échantillon qui doit être quantifié. Les concentrations sont ensuite calculées en faisant le rapport des intégrales de pics normalisées par le nombre de protons contribuant à l'intégrale. Le tube de référence utilisé pour la quantification est un tube de 3 mm de diamètre contenant une solution de 1,18 mM de tryptophane dans du tampon phosphate (50 mM Na Phosphate, pH 6.5, 150 mM NaCl, 1 mM DTT et 10 % D2O).

Bibliographie

[Cheng 2005] Cheng J., Sweredoski M.J. et Baldi P. *Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data*. Data Mining and Knowledge Discovery, vol. 11, no. 3, 2005. – Cité page 147.

[Coeytaux 2005] Coeytaux K. et Poupon A. *Prediction of unfolded segments in a protein sequence based on amino acid composition*. Bioinformatics, vol. 21, no. 9, 2005. – Cité page 147.

[Dosztányi 2005] Dosztányi Z., Csizmok V., Tompa P. et Simon I. *IUPred : web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, vol. 21, no. 16, 2005. – Cité pages 65 et 147.

[Garbuzynskiy 2008] Garbuzynskiy S.O., Lobanov M.Y. et Galzitskaya O.V. *To be folded or to be unfolded ?* Protein Science, vol. 13, no. 11, 2008. – Cité page 147.

[Ishida 2007] Ishida T. et Kinoshita K. *PrDOS : prediction of disordered protein regions from amino acid sequence*. Nucleic Acids Research, vol. 35, no. Web Server, 2007. – Cité page 147.

[Käll 2004] Käll L., Krogh A. et Sonnhammer E.L.L. *A Combined Transmembrane Topology and Signal Peptide Prediction Method*. Journal of Molecular Biology, vol. 338, no. 5, 2004. – Cité page 147.

[Linding 2003a] Linding R. *GlobPlot : exploring protein sequences for globularity and disorder*. Nucleic Acids Research, vol. 31, no. 13, 2003. – Cité pages 65 et 147.

[Linding 2003b] Linding R., Jensen L.J., Diella F., Bork P., Gibson T.J. et Russell R.B. *Protein Disorder Prediction*. Structure, vol. 11, no. 11, 2003. – Cité pages 65 et 147.

[Obradovic 2003] Obradovic Z., Peng K., Vucetic S., Radivojac P., Brown C.J. et Dunker A.K. *Predicting intrinsic disorder from amino acid sequence.* Proteins : Structure, Function, and Bioinformatics, vol. 53 Suppl 6, no. 6, pages 566–572, 2003. – Cité page 147.

[Peng 2006] Peng K., Radivojac P., Vucetic S., Dunker A.K. et Obradovic Z. *1471-2105-7-208.fm*. BMC Bioinformatics, vol. 7, no. 1, pages 208–17, 2006. – Cité page 147.

[Prilusky 2005] Prilusky J., Felder C.E., Zeev-Ben-Mordehai T., Rydberg E.H., Man O., Beckmann J.S., Silman I. et Sussman J.L. *FoldIndex : a simple tool to predict whether a given protein sequence is intrinsically unfolded.* Bioinformatics, vol. 21, no. 16, 2005. – Cité pages 65 et 147.

[Shimizu 2007] Shimizu K., Hirose S. et Noguchi T. *POODLE-S : web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-speci c scoring matrix*. Bioinformatics, vol. 23, no. 17, 2007. – Cité page 147.

[Ward 2004] Ward J.J., McGuffin L.J., Bryson K., Buxton B.F. et Jones D.T. *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, vol. 20, no. 13, 2004. – Cité pages 65 et 147.

[Wider 2006] Wider G. et Dreier L. *Measuring Protein Concentrations by NMR Spectroscopy*. Journal of the American Chemical Society, vol. 128, no. 8, 2006. – Cité page 149.

[Yang 2005] Yang Z.R., Thomson R., McNeil P. et Esnouf R.M. *RONN : the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins*. Bioinformatics, vol. 21, no. 16, 2005. – Cité page 147.

Chapitre 7

Caractérisation biophysique de la région C-terminale de AR-NTD

7.1 Résonance magnétique nucléaire (RMN) **153**
7.1.1 Analyse des spectres protons 153
7.1.2 Mesure de la diffusion translationnelle 155
7.1.3 Marquage isotopique des fragments 157
7.2 Spectroscopie de fluorescence **159**
7.3 Dichroïsme circulaire (CD) **161**
7.4 Diffusion des rayons X aux petits angles (SAXS) **164**
7.4.1 Structures résiduelles 165
7.4.2 Dimensions et forme des protéines 166
7.4.3 Estimation du désordre par la physique des polymères 167
7.5 Conclusion **170**
Bibliographie **171**

Le comportement des sous-domaines AR-NTD a été caractérisé par des approches biophysiques pour obtenir des informations sur le comportement global de ses protéines et sur les propriétés locales telles que la présence de structures secondaires résiduelles. Ces caractérisations incluent systématiquement les expériences de dichroïsme circulaire, spectroscopie de fluorescence, SAXS et RMN.

7.1 Résonance magnétique nucléaire (RMN)

7.1.1 Analyse des spectres protons

Dans un premier temps, nous avons réalisé une analyse RMN des 7 fragments afin de caractériser leur état de repliement et leur propension à s'agréger. Les spectres 1D ^{1}H des 7 fragments produits en milieu LB sont présentés sur la Figure 7.1. Certains fragments (W2-DBD

et W3-DBD) n'ont pas pu être obtenus en grande quantité ce qui explique la différence de qualité des spectres. Nous pouvons cependant constater que l'extension du NTD coïncide avec l'augmentation de l'intensité des pics résonant autour de 8-8,5 ppm, caractéristiques de protons situés dans des régions désordonnées. Ceci indique que le domaine W4 tel qu'il est défini dans notre étude n'est pas capable de se replier de façon autonome. Les fragments étudiés ayant été conçus de façon à incorporer successivement les 4 résidus tryptophanes conservés de la région TAU5, la comparaison des spectres 1D ^{1}H dans la région autour de 10 ppm a permis d'attribuer les fréquences des protons amides du cycle indole de ces résidus (Figure 7.2)

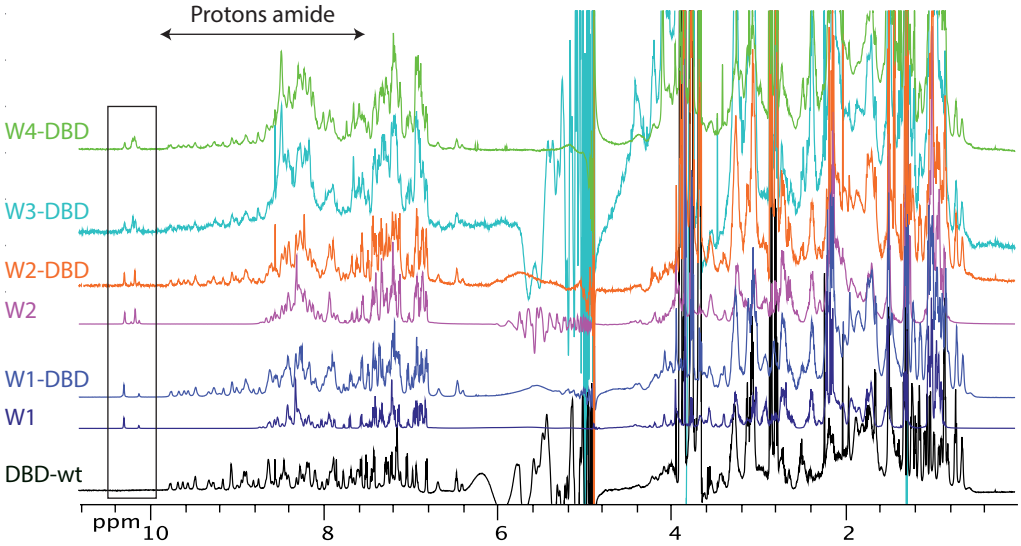


Figure 7.1 – Spectres 1D ^{1}H pour chaque polypeptides enregistrés à 700 MHz et 298 K dont la région encadrée correspond aux fréquences de résonances des H_{ϵ} des résidus W encadrée. Un agrandissement de cette région est présentée dans la Figure 7.2.

Pour le fragment W1, nous constatons la présence systématique d'un pic minoritaire qui représente 25% de la population totale. Cette espèce est attribuée à la présence d'une proline (P525) au voisinage immédiat. Nous remarquons également que la présence du DBD ne modifie pas les déplacements chimiques ce qui suggère l'absence d'interaction entre le NTD et le DBD.

De plus, chaque ajout de tryptophane conduit à l'apparition de pic supplémentaire à 10,18 ppm, une fréquence bien distincte de celle observée pour W526. Ceci suggère que le cycle indole de ce dernier résidu se trouve dans un environnement bien distinct des 3 autres.

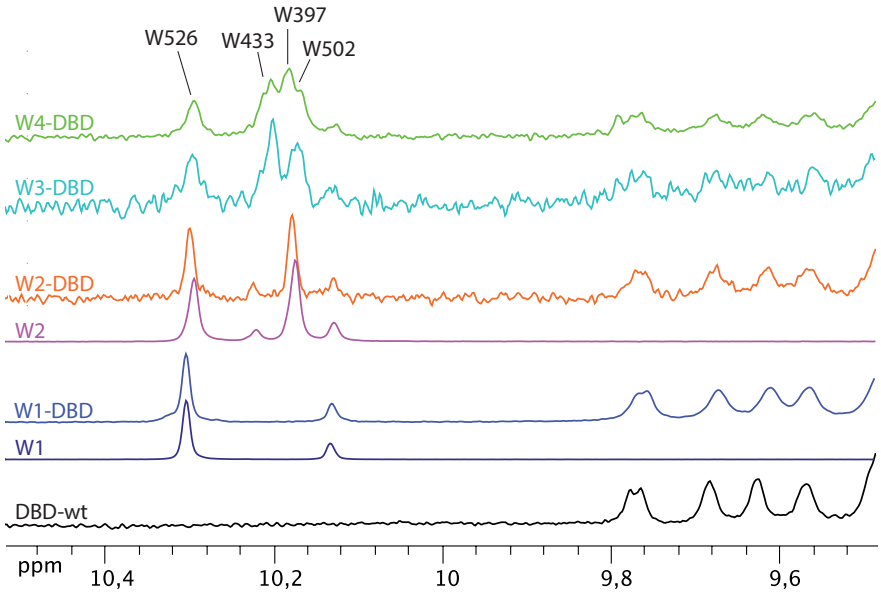


Figure 7.2 – Agrandissement de la région correspondant aux fréquences de résonance des H_ϵ des résidus W dans les spectres 1D ^{1}H.

7.1.2 Mesure de la di usion translationnelle

Le coefficient de diffusion translationnel (D_T) a été mesuré pour les fragments qui ont pu être produits en quantité suffisante (DBD-wt, W1, W2, W1-DBD et W2-DBD). Ce coefficient a été mesuré à l'aide de l'expérience p-DOSY développée au laboratoire [Oikonomou 2015] qui permet une mesure exacte du coefficient de diffusion dans le cas de systèmes hors équilibre. Dans cette expérience, l'intensité des gradients varie de façon aléatoire, ce qui permet d'éliminer les artéfacts comme un décalage de température par exemple.

Le coefficient de diffusion est relié à la masse d'un objet par la loi de puissance D $\propto$ M$^{1/\alpha}$, où correspond à la dimension fractale de l'objet [Augé 2009]. La dimension fractale est comprise entre 5/3, pour des objets complètement désordonnés, et 3 pour des objets compacts. Cependant, le coefficient de diffusion est aussi relié au rayon hydrodynamique (équation (4.8)) de l'objet ce qui rend la relation précédente valable uniquement pour des objets de même forme géométrique. Notre série de fragments n'étant pas homogène, puisque constituée de fragments soit dépliés, soit repliés ou composites (c'est à dire constitués d'une partie désordonnée et d'une partie repliée) nous ne pouvons pas relier directement le coefficient de diffusion à la masse moléculaire pour les comparer entre eux.

Néanmoins, cette expérience nous a permis tout de même d'obtenir des informations sur le comportement hydrodynamique de chaque polypeptide.

La superposition des spectres de diffusion des différents fragments est présentée en Figure 7.3.

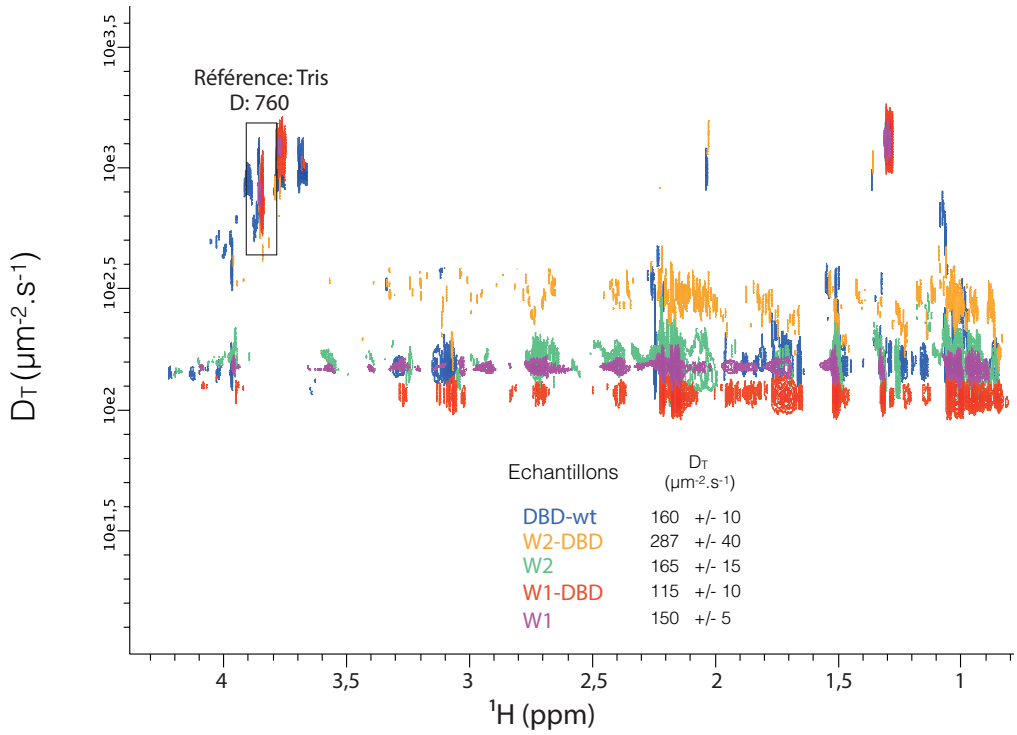


Figure 7.3 – Superposition des spectres de diffusion des fragments DBD-wt (bleu), W1 (violet), W1-DBD (rouge), W2 (vert) et W2-DBD (orange) enregistrés à 700 MHz et 25 °C.

Nous observons que le fragment W1 constitué de 41 résidus possède un coefficient de diffusion de 150 ± 5 $m^2.s^{-1}$ légèrement inférieur à celui du au DBD-wt de 77 résidus (D= 160 ± 10 $m^2.s^{-1}$). Ceci indique que le fragment W1 est relativement déplié en solution, ce qui explique une diffusion plus lente ou égale à celle du DBD-wt, replié. Le fragment W1-DBD présente une diffusion plus lente (115 ± 10 $m^2.s^{-1}$) que W1 et DBD-wt ce qui indique que l'extension NTD reste majoritairement dépliée dans le contexte de la fusion.

Le fragment W2, composé de 65 résidus, présente un coefficient de diffusion (165 ± 15 $m^2.s^{-1}$) similaire à celui du DBD-wt qui laisse penser que ce fragment n'est pas totalement déplié en solution. Ceci est confirmé d'une part, par le fait que W2 diffuse plus rapidement que W1 et d'autre part, par la diffusion plus rapide de W2-DBD comparé au DBD-wt.

Toutefois, il faut rester prudent à propos des interprétations réalisées au sujet des fragments W2 et W2-DBD, puisque nous constatons que les spectres ne sont pas homogènes indiquant une instabilité de ces fragments dans le temps menant éventuellement à la protéolyse.

7.1.3 Marquage isotopique des fragments

Dans un second temps, le marquage isotopique en milieu minimum des fragments en fusion avec le DBD (W4-DBD, W3-DBD, W2-DBD et W1-DBD) a été réalisé dans l'optique d'attribuer les fréquences de résonances des résidus de la partie NTD.

L'impact de la présence de fragments du NTD de plus en plus longs sur le domaine DBD a été étudié en comparant les spectres de corrélation $^{1}H-^{15}N$ enregistrés sur les quatres protéines (Figure 7.4).

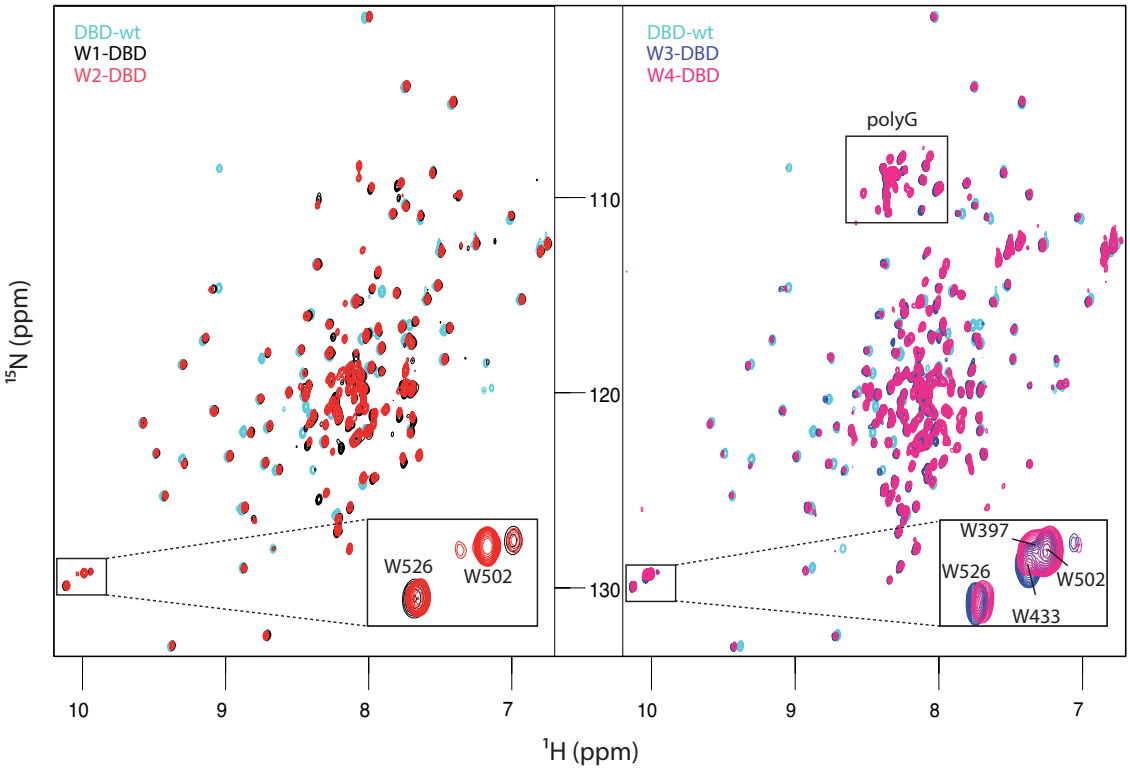


Figure 7.4 – Spectres 2D $^{1}H-^{15}N$-HSQC, pour chaque polypeptide, enregistrés à 700 MHz et 25 °C pour les échantillons DBD-wt, W1-DBD et W2-DBD et à 20 °C pour W3-DBD et W4-DBD.

Le spectre $^{1}H-^{15}N$ HSQC permet, en premier lieu, d'identifier la présence de certains types de résidus dans la protéine.

Les spectres $^{1}H-^{15}N$ HSQC de ces fragments présentent à la fois des caractéristiques de protéine repliée du fait de la dispersion importante de certaines fréquences qui correspondent à des résidus appartenant au DBD, et des caractéristiques de protéine désordonnée avec une faible dispersion des fréquences dans la dimension proton mais une bonne dispersion dans la dimension azote correspondant à des résidus appartenant au NTD. Nous observons également

la présence de l'homo-polymère de glycines (polyG) dans les fragments W3-DBD et W4-DBD (Figure 7.4).

Pour les fragments NTD de grande taille, nous observons l'apparition de corrélation de largeur importante qui pourraient correspondre à la présence de multiples conformations en échange intermédiaire-lent.

Un décalage des fréquences vers la gauche est observé pour le spectre correspondant à W3-DBD et W4-DBD par rapport au spectre du DBD-wt dû aux conditions expérimentales. En effet, les spectres ont été enregistrés à 20 °C (5 degrés de moins) pour limiter les phénomènes d'agrégation et de dégradation des échantillons.

Aussi, ces spectres ont permis de confirmer l'attribution des protons H_ϵ des cycles indoles des 4 tryptophanes.

Puis, nous avons examiné l'effet de l'ajout de fragments du domaine NTD de plus en plus longs sur les fréquences du DBD. La disponibilité des attributions du domaine DBD obtenues pour le domaine isolé et au sein des fragments W1-DBD et W2-DB permet d'évaluer la possibilité d'une interaction intra-moléculaire entre les acides aminés appartenant au NTD avec ceux du DBD. Les différences de déplacement chimique sont représentées sous la forme d'un déplacement chimique composite calculé selon l'équation (4.7) en fonction de la séquence (Figure 7.5).

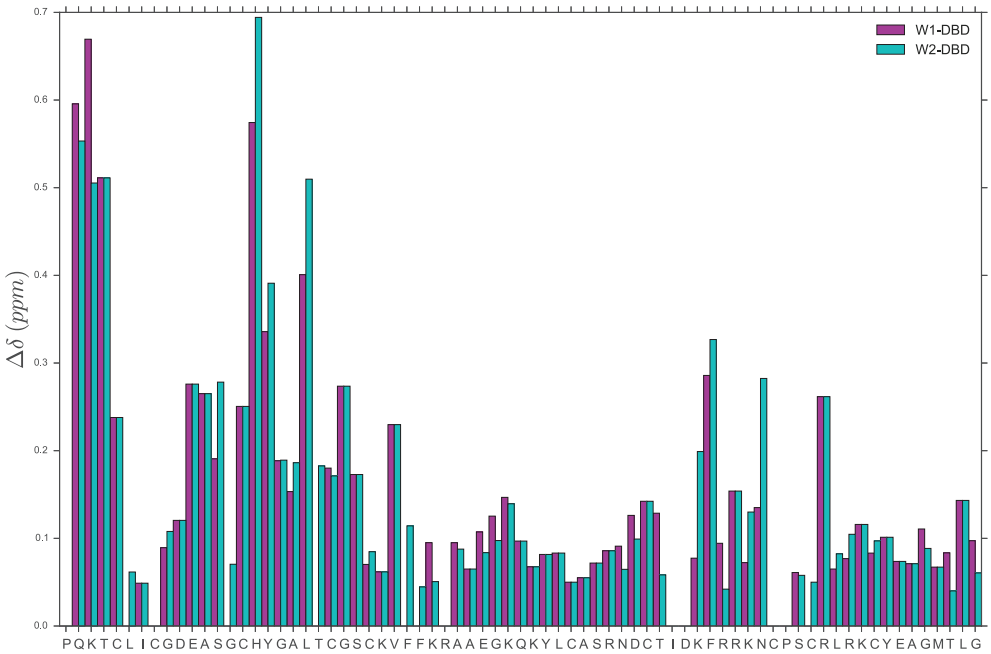


Figure 7.5 – Perturbation des déplacements chimiques des résidus appartenant à AR-DBD dans les fragments W1-DBD (magenta) et W2-DBD (cyan) comparé au DBD isolé.

Nous observons une différence systématique de 0,1 ppm pour l'ensemble des résidus, sans doute liée à une différence de calibration de température. Quelques résidus montrent des différences plus prononcées et ceux présentant un déplacement chimique supérieur à 0,2 ppm sont représentés sur la structure tridimensionnelle de AR-DBD (Figure 7.6).

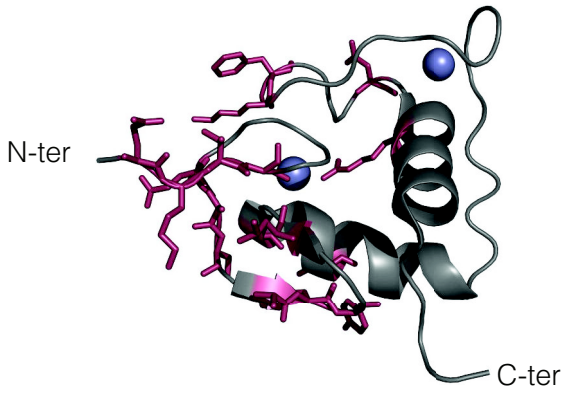


Figure 7.6 – Cartographie des déplacements chimiques supérieurs à 0,2 ppm sur la structure cristallographique de AR-DBD (PDB : 1R4I).

Ces résidus sont regroupés dans la région correspondant à l'interface entre la région NTD et le DBD. L'observation d'un effet sur le résidu K605 situé plus loin sur la surface du DBD suggère la possibilité d'interactions transitoires entre des résidus chargés négativement sur le NTD et le groupe d'acides aminés positifs constitué des résidus RKK à la surface du DBD. De plus, le résidu H570 établissant une liaison hydrogène avec un phosphate de l'ADN présente un déplacement chimique important.

7.2 Spectroscopie de fluorescence

Nous avons utilisé la fluorescence intrinsèque du tryptophane pour sonder l'environnement des 4 résidus tryptophane de AR-NTD. En effet, le maximum d'émission de fluorescence (λ_{max}) est situé à 360 nm lorsque le W est totalement exposé au solvant alors qu'il est situé à 325 nm dans une solution apolaire.

Les spectres ont été enregistrés pour des concentrations identiques (2 μM) de protéines à 25 °C.

La Figure 7.7 montre les spectres d'émissions des différents polypeptides après excitation à 295 nm.

Le pic à 295 nm correspond à la diffusion de la lumière d'excitation par les différents

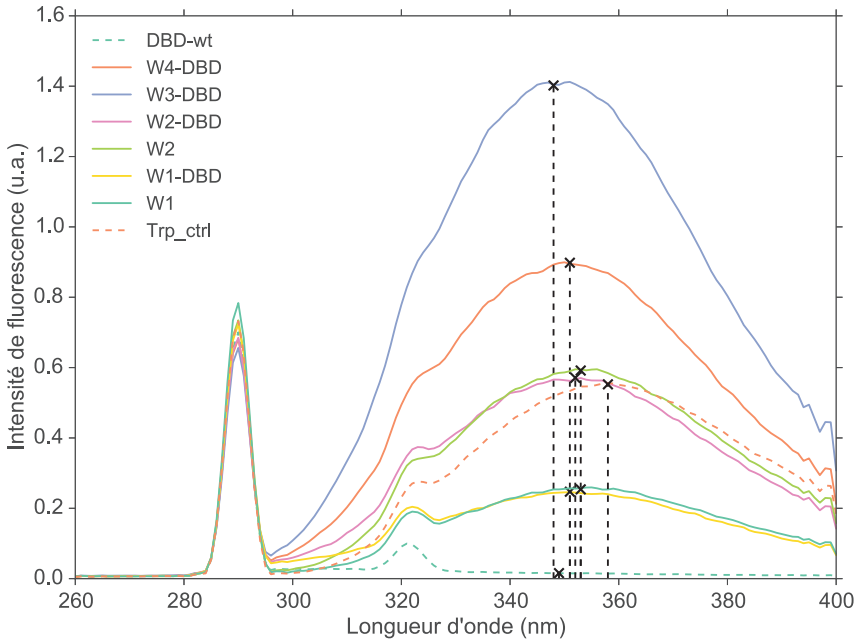


Figure 7.7 – Spectres d'émission de fluorescence obtenus pour chaque échantillon après excitation à 295 nm .

échantillons. Ce pic présente des caractéristiques (largeur et intensité) similaires pour chaque échantillon ce qui indique l'absence d'agrégats de grande taille en solution.

Le λ_{max} de chaque polypeptide a été déterminé avec un algorithme d'ajustement (Savitsky-Golay) [Savitzky 1964].

Echantillon	λ_{max} (nm)	Intensité max de fluorescence (u.a)
$\text{Trp}_{-\text{ctrl}}$	358,0	0,54
W4-DBD	351,0	0,90
W3-DBD	348,7	1,4
W2-DBD	352,0	0,54
W1-DBD	351,3	0,24
W2	353,4	0,59
W1	354,0	0,26

Tableau 7.1 – Maximum d'émission et d'intensité de fluorescence pour chaque échantillon après une excitation à 295 nm.

Le rendement quantique de W1 diminue d'un facteur 2,5-3 par rapport au tryptophane

libre. Puis, nous n'observons aucune différence entre les spectres de fluorescence de W1 et W1-DBD, ce qui indique que la fluorescence du W526 est liée à des propriétés intrinsèques de séquence et non à la présence du DBD.

En ce qui concerne W2, le rendement quantique des 2 tryptophanes de ce polypeptide (W526 et W502) correspond à celui du tryptophane libre. Comme précédemment, aucune différence n'est observée entre les spectres de fluorescence de W2 et W2-DBD signifiant que le DBD n'a pas d'influence sur le NTD.

Les résultats pour W3-DBD montrent une augmentation importante du rendement quantique alors que l'ajout du W397 dans le polypeptide W4-DBD conduit quant à lui à une diminution du rendement quantique. De plus, nous observons un décalage du $_{max}$ de W3 vers le bleu de la longueur d'onde d'émission comparé au $_{max}$ de W4 qui se situe dans la même zone que W1, W1-DBD, W2 et W2-DBD.

Les caractéristiques de W3-DBD et W4-DBD peuvent correspondre à des interactions intermoléculaires, comme par exemple des échanges entre des formes multimériques et monomériques.

7.3 Dichroïsme circulaire (CD)

La caractérisation des structures secondaires partielles au sein du NTD a été réalisée par dichroïsme circulaire (CD). Les Figures 7.8 A et 7.8 B montrent les spectres obtenus pour des concentrations identiques (35 M) de protéines à 25 °C. La forme du spectre de CD dans l'UV lointain (185-260nm) indique la proportion des molécules repliées en hélice- , feuillets- ou complètement désordonnées.

Le spectre CD du DBD-wt présente deux pics d'ellipticité molaire négative à 208 et 222 nm caractéristique d'une structure secondaire de type hélice- principalement. Cependant, d'après la structure 3D de AR-DBD, ce domaine se compose de deux hélices- , deux petits feuillets- et d'un certain nombre de boucles. Ainsi, les spectres CD peuvent être difficiles à interpréter du fait des structures secondaires qui coexistent au sein d'une protéine.

Les spectres CD de W1 et W2 sont assez similaires avec une forte ellipticité molaire négative vers 220 nm caractéristique de protéines désordonnées de type random coil (RC). Néanmoins, le fragment W1 se différencie par la présence d'un pic d'ellipticié à 220 nm indiquant une petite proportion d'hélice- (Figure 7.8 A).

L'allure des spectres de CD pour les polypeptides comprenant une région NTD fusionnée au DBD est relativement similaire et nous pouvons remarquer que l'allongement du NTD conduit à un doublement du contenu en hélice- par rapport au DBD isolé (Figure 7.8 B).

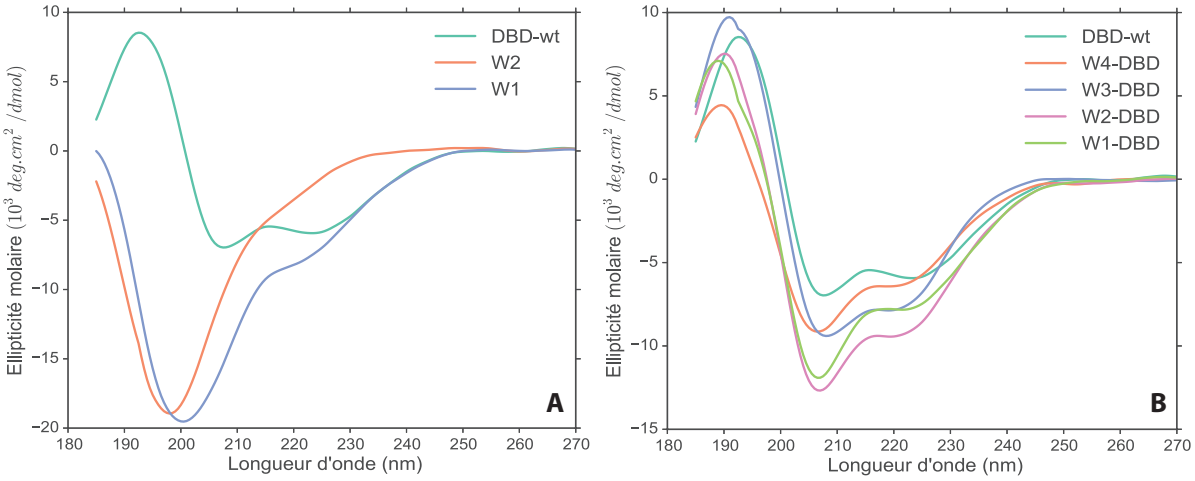


Figure 7.8 – Spectres de dichroïsme circulaire des polypeptides W1, W2 et DBD-wt (A) et W1-DBD, W2-DBD, W3-DBD et W4-DBD (B) enregistrés pour 35 M de protéine à 25 °C.

De plus, la comparaison des spectres obtenus pour les fragments contenant uniquement les parties désordonnées (W1 et W2) avec ceux obtenus pour ces régions du NTD en fusion avec le DBD (W1-DBD et W2-DBD) nous ont permis d'explorer l'impact de la présence du DBD sur les conformations résiduelles du NTD. Nous avons enregistré les spectres dichroïques de mélanges équimolaires du domaine DBD et d'une des régions W1 ou W2 (notés $W1+DBD_{exp}$ et $W2+DBD_{exp}$ sur la Figure 7.9). Nous observons que l'allure de ces spectres est similaire à la somme des spectres enregistrés pour ces domaines individuels (courbes W1+DBD et W2+DBD sur la Figure 7.9). Le spectres de CD obtenus pour W1-DBD et W2-DBD sont différents de ceux obtenus pour le mélange équimolaire et de ceux obtenus après addition des spectres des domaines isolés. Ces résultats indiquent que le domaine DBD n'interagit pas avec le fragments W1 et W2 lorsqu'ils existent sous forme de domaines isolés. Par contre, il semble y avoir une interaction entre le domaine DBD et les régions W1 et W2, dans le contexte de la fusion, qui favorise un repliement partiel de ces dernières.

Les données de CD ont ensuite été analysées pour extraire les proportions respectives des différents éléments de structure secondaire à l'aide de la méthode CONTIN/LL et de la base de données de spectres de référence SMP56. La Figure 7.10 représente le nombre de résidus de chaque polypeptide impliqués dans un type de structure secondaire obtenu après déconvolution de chaque spectre CD.

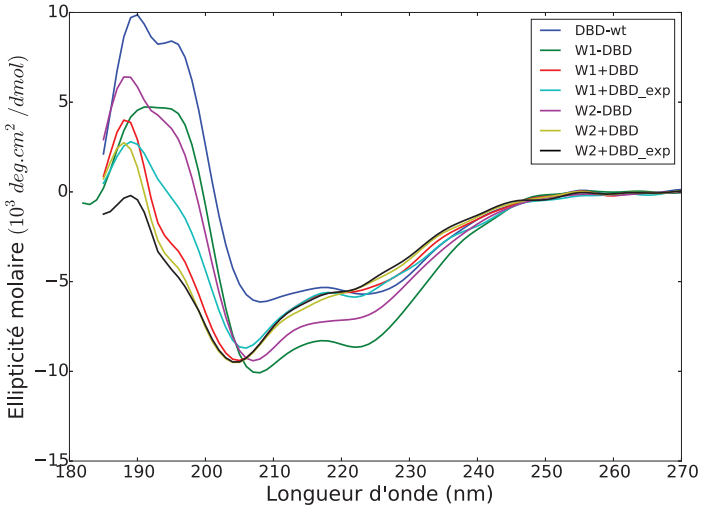


Figure 7.9 – Spectres CD correspondant soit à la somme des spectres des domaines isolés DBD-wt, W1 et W2 soit à leur mélange équimolaire comparé aux spectres CD de ces domaines isolés.

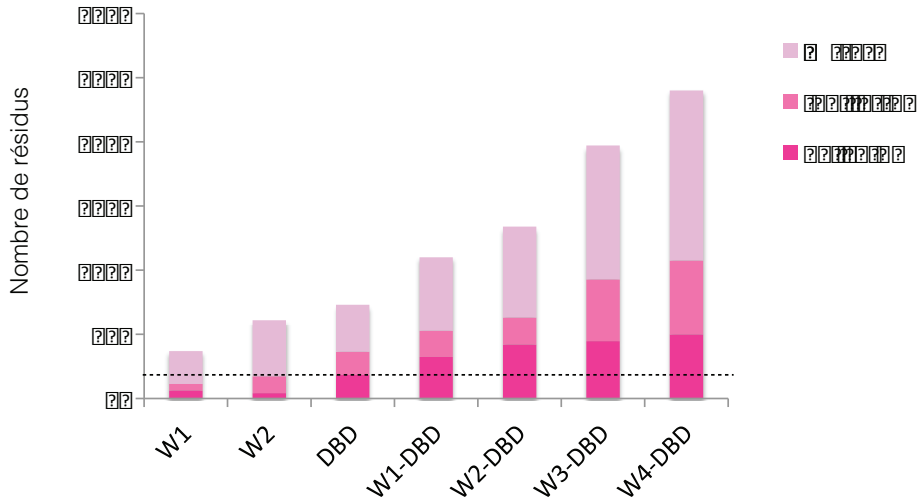


Figure 7.10 – Nombre de résidus impliqués dans les structures secondaires α, β ou autre (random coil et turn) pour chaque polypeptide.

La proportion de résidus du DBD impliqués dans chaque type de structure secondaire correspond à celle attendue par rapport à sa structure 3D. Pour W1 et W2, la différence d'ellipticité molaire à 220 nm indiquant une structuration en hélice-α pour W1 est confirmée,

suggérant que les résidus ajoutés dans W2 sont majoritairement dépliés.

Aussi, pour les fragments W1-DBD et W2-DBD nous observons que davantage de résidus des régions NTD sont impliqués dans des hélices-α dans le contexte de la fusion au DBD comparé aux fragments W1 et W2.

La comparaison des résultats obtenus pour les polypeptides W2-DBD, W3-DBD et W4-DBD montrent que le nombre de résidus impliqués dans une structure secondaire de type α reste quasi constant. Cela indique que les résidus du NTD ajoutés dans les polypeptides W3-DBD et W4-DBD sont principalement impliqués dans des structures de type β, random coil ou turn et très peu d'entres eux sont impliqués dans une structure secondaire de type α.

7.4 Diffusion des rayons X aux petits angles (SAXS)

La courbe de diffusion renseigne sur l'intensité I_0, reliée à la masse moléculaire de la particule en solution, ainsi que sur le rayon de giration dont la valeur provient de la pente aux petits angles. I_0 étant dépendant de la concentration d'échantillon (équation 4.4), nous ne pouvons pas utiliser cette valeur dans notre cas puisque chaque expérience à été réalisée pour une concentration différente de protéine. La Figure 7.11 présente les courbes de diffusion représentant $lnI(q)$ en fonction de q pour le domaine DBD-wt replié et le fragment W1 qui apparaît plutôt désordonné au vu des résultats jusqu'à présent.

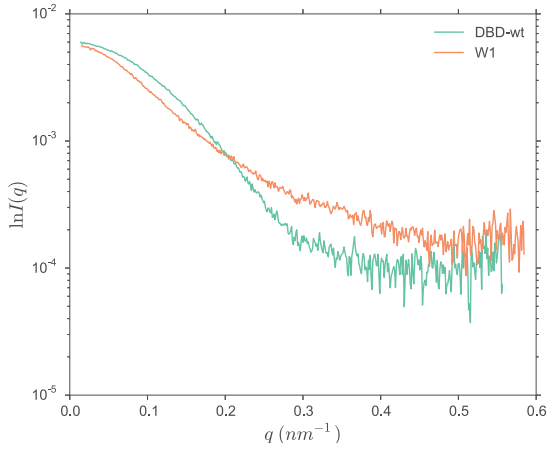


Figure 7.11 – Courbe de diffusion correspondant aux polypeptides W1 et DBD-wt.

La similarité de ces courbes nous indique que cette représentation de la courbe de diffusion n'est pas adaptée pour observer des différences subtiles entre l'état replié, désordonné et la présence de structures secondaires résiduelles.

Ainsi, à partir des données enregistrées au synchrotron SOLEIL et sans avoir recours à des étapes de modélisation moléculaire, nous avons déterminé le rayon de giration (R_g), la représentation de Kratky et la fonction de distribution de distance P(r) à l'aide du logiciel PRIMUS et GNOM faisant partie de la suite ATSAS [Svergun 1992, Konarev 2003, Petoukhov 2012].

7.4.1 Structures résiduelles

La représentation de Kratky normalisée de la courbe de diffusion renseigne directement sur la nature compacte ou non structurée d'un échantillon (chapitre 4). Ce type de représentation est donnée pour chaque échantillon dans la Figure 7.12.

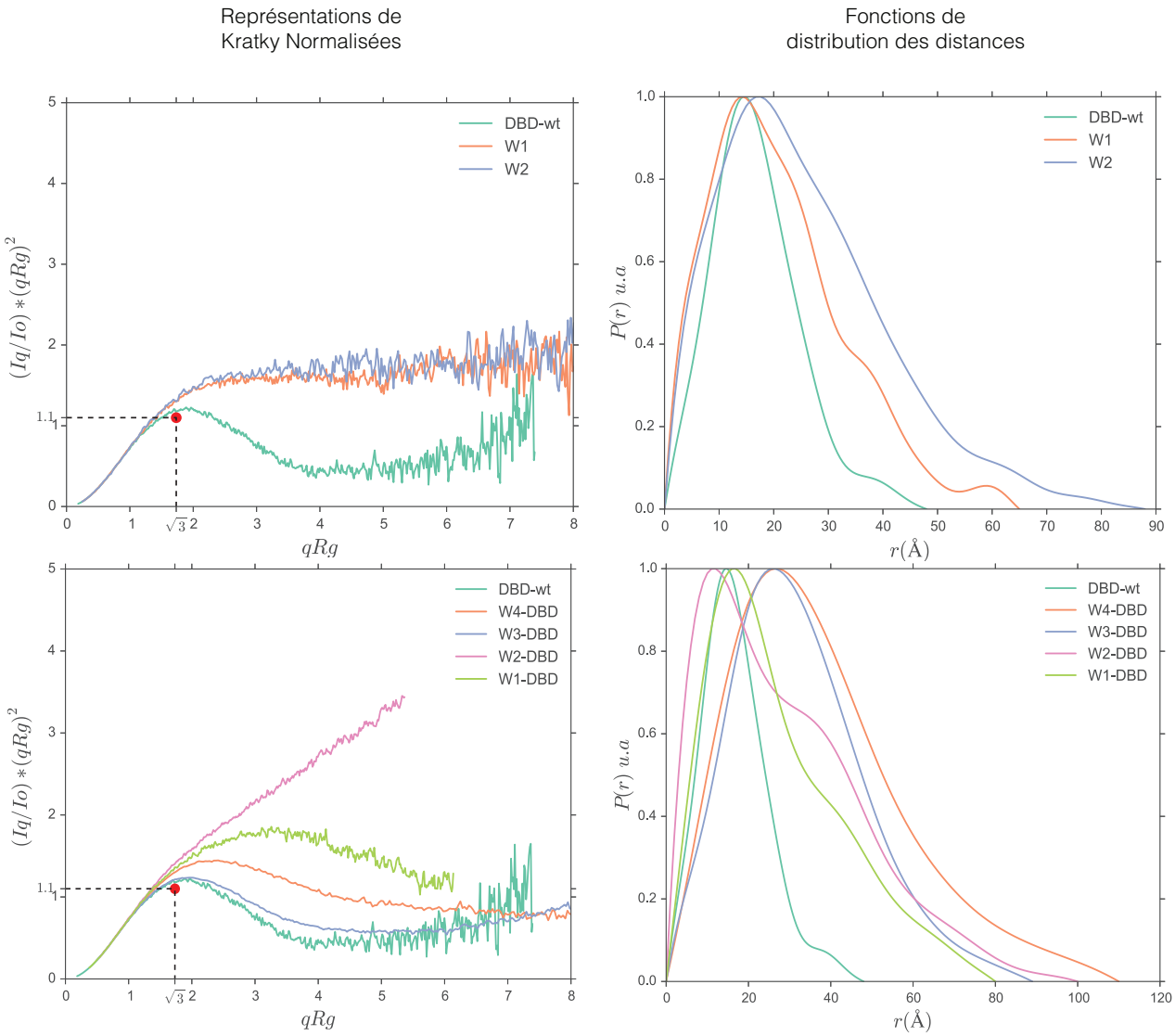


Figure 7.12 – Courbe de diffusion de chaque polypeptide en représentation de Kratky normalisée et fonction de distribution des distances (P(r)) correspondantes.

Cette représentation pour le DBD possède une allure caractéristique de protéine globulaire en forme de cloche, symétrique et avec un maximum bien défini dont les coordonnées sont ($\sqrt{3}$; 1,1) [Durand 2010].

A l'inverse, les polypeptides W1 et W2 dont les courbes présentent un profil caractéristique d'IDP avec une augmentation avec la valeur $I_q \;\; I_0(qR_g)^2$ jusqu'à atteindre un plateau dont la valeur maximale et située entre 1,5-2.

Les profils de Kratky normalisés pour les polypeptides W1-DBD, W3-DBD et W4-DBD présentent un début de cloche, correspondant au domaine DBD, puis semblent atteindre un plateau à grande valeur de q dont la valeur est intermédiaire aux protéines globulaires (DBD) et IDPs (W1 et W2). Cela indique la présence de structures résiduelles au sein des régions NTD correspondantes dont le degré est fonction de la position du maximum et du plateau entre les deux extrêmes (protéine globulaire et IDP).

Enfin, le profil correspondant au polypeptide W2-DBD est caractéristique d'une protéine étendue présentant des rigidités locales [Receveur-Brechot 2012].

7.4.2 Dimensions et forme des protéines

Les dimensions, R_g, de chaque polypeptide ont été calculée dans la zone de $qR_g <$ 0,8-1,3, la courbe de diffusion étant linéaire en représentation de Guinier dans cette zone. Le tableau 7.2 résume les différentes valeurs de I_0, R_g ainsi que les limites de l'approximation de Guinier et Dmax obtenues.

Polypeptide	MM (kDa)	I_0 (10^{-3})	R_g (Å)	limites qR_g		D_{max} (Å)
DBD-wt	8,4	6,0 ±0,006	13,0 ± 0,02	0,18	1,30	48
W4-DBD	24,8	41,0 ± 0,080	29,6± 0,10	0,70	1,00	110
W3-DBD	20,8	48,3 ± 0,050	25,8± 0,07	0,49	0,90	89
W2-DBD	15,3	16,2 ± 0,028	22,8± 0,08	0,50	1,00	100
W2	7,2	3,5 ± 0,011	21,3 ±0,19	0,24	0,98	88
W1-DBD	12,7	8,2 ± 0,023	22,5 ± 0,21	0,38	0,84	80
W1	4,6	5,7 ± 0,008	16,2± 0,05	0,23	1,20	65

Tableau 7.2 – Valeurs de I_0, R_g, qR_g et D_{max} obtenues d'après l'approximation de Guinier et la fonction de distribution des distances.

Dans un premier temps, la fonction P(r) nous fournit l'histogramme des distances intermoléculaires au sein de chaque échantillon. Le maximum correspond à la distance la plus représentée dans l'échantillon et la valeur maximale de r correspond à la distance maximale

(D_{max}) dans l'échantillon. Les fonctions P(r) pour chaque échantillon sont tracées dans la Figure 7.12 et les valeurs de (D_{max}) correspondantes sont reportées dans le tableau 7.2 .

Le profil de la fonction P(r) correspondant au DBD-wt est symétrique avec une petite extension. La distance la plus représentée au sein de ce domaine est de 15 Å et la distance maximale de 48 Å.

Pour W1 la distance la plus représentée est identique à celle du DBD-wt et la distance maximale est 65 Å pour un profil asymétrique représentatif d'une protéine étendue. Par ailleurs, un léger épaulement est observé à 40 Å témoignant d'une éventuelle structuration. Le profil de W2 est similaire à celui de W1, mais la distance maximale et la distance la plus représentée sont de 88 Å et 18,6 Å respectivement. Ceci indique que les 24 résidus ajoutés à W2 sont majoritairement désordonnés.

Le profil P(r) obtenu pour W1-DBD contient un sommet correspondant au domaine DBD-wt et une valeur D_{max} de 80 Å correspondant à la région NTD. De plus, nous observons un léger épaulement à 45 Å qui pourrait correspondre à la distance séparant le DBD de la zone partiellement structurée de la région W1.

Concernant les fragments W3-DBD et W4-DBD, la fonction P(r) possèdent un sommet décalé à 25 Å comparé au DBD-wt suggérant l'existence d'un domaine plus large. Les valeurs D_{max} (89 et 110 Å pour W3-DBD et W4-DBD respectivement) indiquent que les résidus appartenant à la région W4 sont plutôt étendus.

Le profil obtenu pour W2-DBD est surprenant car il présente un sommet correspondant à une distance majoritaire inférieure à celle du DBD-wt. Cela peut se traduire par une déstructuration de ce dernier. Ce profil est triangulaire avec une valeur de D_{max} de 100 Å indiquant l'existence d'une conformation étendue en solution. Ce type de profil est caractéristique d'un cylindre pour lequel la valeur de D_{max} correspond à sa longueur et la distance majoritairement représentée à son diamètre. Cependant, on ne peut pas exclure un phénomène de protéolyse (non contrôlée au cours de l'expérience) qui aboutirait à une distribution des distances similaire à celle que nous observons.

7.4.3 Estimation du désordre par la physique des polymères

Nous avons ensuite utilisé une analyse issue de la physique des polymères simples. Cette analyse repose sur l'observation que le rayon de giration R_g évolue en fonction du nombre de résidus N en suivant une loi de puissance :

$$R_g = R_0 N^{\nu} \tag{7.1}$$

L'exposant représente l'inverse de la dimension fractale de l'objet étudié. Le tableau 7.3 présente les valeurs calculées de rayon de giration prédites pour des protéines globulaires ou IDP à partir des trois études citées dans le chapitre 4 comparées aux valeurs expérimentales.

Polypeptide	Résidus	Plaxco et coll.		Wilkins et coll.		Bernadó et Blackledge	Données expérimentales
		R_g (Å) G	R_g (Å) IDP	R_g(Å) G	R_g(Å) IDP	R_g (Å) IDP	R_g (Å)
DBD-wt	77	10,1	25,9	13,0	27,7	24,5	13,0 ± 0,02
W4-DBD	244	15,6	51,7	18,1	54,1	44,8	29,6 ± 0,10
W3-DBD	201	14,5	46,0	17,1	48,3	40,5	25,8 ± 0,07
W2-DBD	138	12,6	36,7	15,4	38,9	33,3	22,8 ± 0,08
W2	65	9,4	23,4	12,3	25,1	22,4	21,3 ± 0,19
W1-DBD	114	11,7	32,8	14,5	34,8	34,1	22,5 ± 0,21
W1	41	7,9	17,8	10,8	19,2	17,6	16,2 ± 0,05

Tableau 7.3 – Valeurs calculées des rayons de giration (en Å) pour des protéines globulaires (G) ou IDPs à partir du nombre de résidu pour chaque polypeptide d'après les méthodes [Kohn 2004, Wilkins 1999] et uniquement IDP d'après [Bernadó 2009] comparées aux valeurs expérimentales.

Ces prédictions peuvent être utilisées comme limites supérieures et inférieures, et un écart significatif de la valeur expérimentale du R_g par rapport à la limite supérieure théorique indique la présence de structure secondaire résiduelle. Les valeurs de R_g calculées à partir des paramètres de ces trois études comparées aux valeurs expérimentales indiquent que les fragments W1 et W2 sont plutôt désordonnés alors que le domaine DBD-wt est replié. Les fragments composés d'une région NTD ainsi que du DBD-wt étant des fragments composites, les valeurs de R_g expérimentales sont situées entre les deux extrémités.

La Figure 7.13 est une représentation de la loi de puissance utilisant uniquement les paramètres déterminés par [Kohn 2004], correspondant à la méthode la plus récente permettant de calculer des valeurs de rayon de giration pour des protéines globulaires et IDP. Cette représentation correspond à une droite dont la pente correspond à la valeur de .

Si le comportement des échantillons est régulier, alors leurs valeurs de R_g s'alignent sur une droite. Les résultats indiquent que le rayon de giration du DBD-wt passe par la droite correspondant aux protéines structurées conformément à ce qui est attendu.

Les rayons de giration des polypeptides W1 et W2 s'alignent quasiment avec la droite correspondant aux protéines désordonnées. Néanmoins le fragment W1 est légèrement plus distant de cette droite que W2 signe d'une structuration résiduelle légèrement supérieure à ce

dernier.

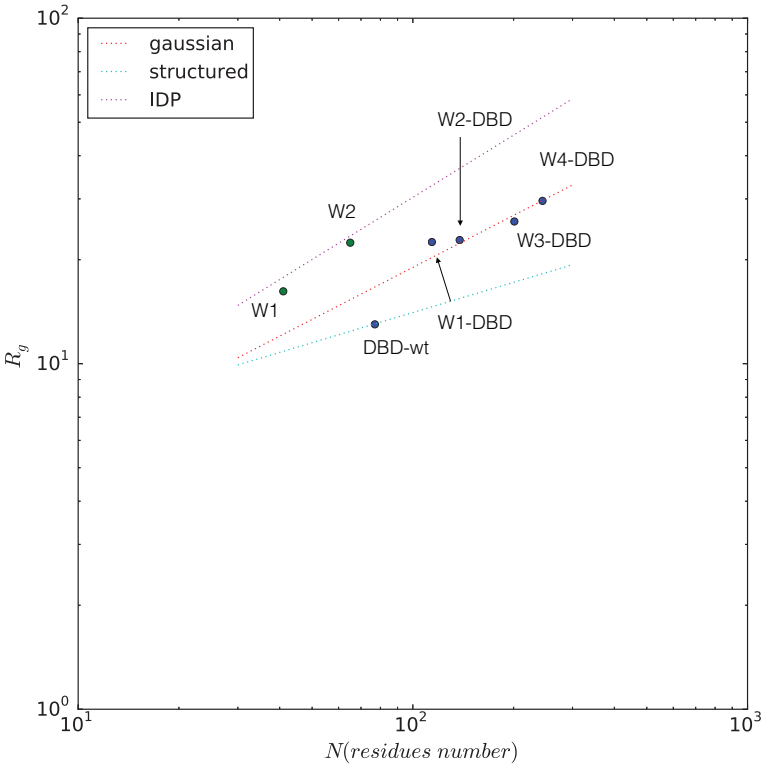


Figure 7.13 – Représentation logarithmique de la loi de puissance suivie par chaque polypeptide d'après [Kohn 2004]. Les droites attendues pour des protéines structurées et désordonnées (IDP) en pointillés bleu et violet respectivement, et la droite indiquant un comportement aléatoire en pointillés rouge. Les valeurs expérimentales de R_g pour chaque polypeptide sont représentées par un point.

Les valeurs expérimentales des polypeptides W1-DBD, W2-DBD, W3-DBD et W4-DBD ne s'alignent pas sur ces deux droites. Ces échantillons étant irréguliers dû à la fusion du DBD qui est globulaire et d'une partie étendue plus ou moins longue, cela rend l'interprétation des données plus complexe. Tout de même, nous pouvons remarquer que le fragment W2-DBD semble plus structuré que W1-DBD. En effet, en passant de W1-DBD à W2-DBD la valeur du rayon de giration n'augmente que de 0,3 Å alors que 24 résidus sont ajoutés. Cette observation est cohérente avec les résultats de CD.

7.5 Conclusion

L'ensemble des données biophysiques obtenus sur les fragments W1, W1-DBD, W2, W2-DBD, W3-DBD et W4-DBD montrent que, en présence du DBD, ces régions qui sont principalement désordonnées présentent des signes de structure locale ou partielles. Le NTD de AR, en particulier la région AF-1 (constituée des sous-unités TAU1 et TAU5) contient la fonction de transactivation majeure du récepteur mais les mécanismes moléculaires sont encore inconnus. Des études ont montrées que cette région présente une transition désordre/ordre en hélice- en présence d'agents stabilisateurs comme le TFE ou TMAO ou lors de l'interaction avec la sous-unité RAP74 du facteur de transcription TFIIH [Kumar 2012, McEwan 2012]. Cette région est également le siège de nombreuses mutations et de modifications post-traductionnelles qui sont la phosphorylation et la sumoylation et qui permettent de moduler l'activité du récepteur [Gioeli 2012].

Actuellement les récepteurs nucléaires dans leur ensemble sont considérés comme des intégrateurs de signaux à l'origine de la présence de différentes communications inter-domaine dans le mécanisme de transduction du signal a déjà été démontrée pour plusieurs d'entre eux [Meijsing 2009, Watson 2013].

Dans le cas de AR, le rôle de l'interaction NTD/LBD est démontré dans la régulation de son activité. La communication entre les domaines NTD et DBD est étudiée depuis peu et pour le moment aucune étude à résolution atomique n'a été réalisée. La comparaison des déplacements chimiques des fréquences de résonances de AR-DBD, pris de façon isolé, et au sein des fragments W1-DBD et W2-DBD révèle que des résidus situés à l'interface protéine-ADN comme H570 présente un déplacement chimique important en présence de la région NTD. Ces résultats suggèrent l'existence d'un couplage allostérique entre le DBD et les résidus adjacents sur le NTD. Ce couplage modifie l'ensemble conformationnel accessible au NTD en favorisant une conformation en hélice- .

Les modifications de la région TAU5 sous l'effet de PTMs ou des différentes interactions avec des co-régulateurs modulent l'activité transcriptionnelle de AR en utilisant les mécanismes de couplage allostériques entre les domaines repliés et désordonnés qui ont été proposés dans le cas du récepteur GR.

Bibliographie

[Augé 2009] Augé S., Schmit P.O., Crutchfield C.A., Islam M.T., Harris D.J., Durand E., Clemancey M., Quoineaud A.A., Lancelin J.M., Prigent Y., Taulelle F. et Delsuc M.A. *NMR Measure of Translational Di usion and Fractal Dimension. Application to Molecular Mass Measurement*. The Journal of Physical Chemistry B, vol. 113, no. 7, 2009. – Cité pages 84 et 155.

[Bernadó 2009] Bernadó P. et Blackledge M. *A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering*. Biophysical journal, vol. 97, no. 10, 2009. – Cité pages 72 et 168.

[Durand 2010] Durand D., Vivès C., Cannella D., Pérez J., Pebay-Peyroula E., Vachette P. et Fieschi F. *NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi- exible linkers*. Journal of Structural Biology, vol. 169, no. 1, 2010. – Cité pages 73 et 166.

[Gioeli 2012] Gioeli D. et Paschal B.M. *Post-translational modi cation of the androgen receptor*. Molecular and Cellular Endocrinology, vol. 352, no. 1-2, 2012. – Cité pages 22, 23 et 170.

[Kohn 2004] Kohn J.E., Millett I.S., Jacob J., Zagrovic B., Dillon T.M., Cingel N., Dothager R.S., Seifert S., Thiyagarajan P., Sosnick T.R., Hasan M.Z., Pande V.S., Ruczinski I., Doniach S. et Plaxco K.W. *Random-coil behavior and the dimensions of chemically unfolded proteins*. Proceedings of the National Academy of Sciences, vol. 101, no. 34, pages 12491–12496, 2004. – Cité pages 72, 168 et 169.

[Konarev 2003] Konarev P.V., Volkov V.V., Sokolova A.V., Koch M.H.J., Svergun D.I. et IUCr. *PRIMUS : a Windows PC-based system for small-angle scattering data analysis*. Journal of applied crystallography, vol. 36, no. 5, 2003. – Cité page 165.

[Kumar 2012] Kumar R. et McEwan I.J. *Allosteric Modulators of Steroid Hormone Receptors : Structural Dynamics and Gene Regulation*. Endocrine Reviews, vol. 33, no. 2, 2012. – Cité page 170.

[McEwan 2012] McEwan I.J. *Intrinsic disorder in the androgen receptor : identi cation, characterisation and drugability*. Molecular BioSystems., vol. 8, no. 1, pages 82–90, 2012. – Cité pages 15, 26, 131 et 170.

[Meijsing 2009] Meijsing S.H., Pufall M.A., So A.Y., Bates D.L., Chen L. et Yamamoto K.R. *DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity*. Science, vol. 324, no. 5925, 2009. – Cité pages 12 et 170.

[Oikonomou 2015] Oikonomou M., Asencio-Hernández J., Velders A.H. et Delsuc M.A. *Accurate DOSY measure for out-of-equilibrium systems using permutated DOSY (p-DOSY).* Journal of Magnetic Resonance, vol. 258, no. C, 2015. – Cité pages 84 et 155.

[Petoukhov 2012] Petoukhov M.V., Franke D., Shkumatov A.V., Tria G., Kikhney A.G., Gajda M., Gorba C., Mertens H.D.T., Konarev P.V. et Svergun D.I. *New developments in the ATSASprogram package for small-angle scattering data analysis.* Journal of applied crystallography, vol. 45, no. 2, 2012. – Cité page 165.

[Receveur-Brechot 2012] Receveur-Brechot V. et Durand D. *How random are intrinsically disordered proteins ? A small angle scattering perspective.* Current protein & peptide science, vol. 13, no. 1, 2012. – Cité pages 71, 73 et 166.

[Savitzky 1964] Savitzky A. et Golay M. *Smoothing and di erentiation of data by simpli ed least squares procedures.* Analytical Chemistry, 1964. – Cité page 160.

[Svergun 1992] Svergun D.I. *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* Journal of applied crystallography, vol. 25, no. 4, 1992. – Cité pages 74, 165 et 174.

[Watson 2013] Watson L.C., Kuchenbecker K.M., Schiller B.J., Gross J.D., Pufall M.A. et Yamamoto K.R. *The glucocorticoid receptor dimer interface allosterically transmits sequence-speci c DNA signals.* Nature Structural & Molecular Biology, vol. 20, no. 7, 2013. – Cité pages 12 et 170.

[Wilkins 1999] Wilkins D.K., Grimshaw S.B., Receveur V., Dobson C.M., Jones J.A. et Smith L.J. *Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques* . Biochemistry, vol. 38, no. 50, 1999. – Cité pages 72 et 168.

Annexe B

B.1 Expériences RMN

Les expériences RMN on été réalisées sur un spectromètre 700 MHz Buker Avance III équipée d'une cryosonde Z-gradient à 25 °C. Les spectres 1D et DOSY ont été traités avec le logiciel NMR-notebook 2.7 (NMRTEC Illkirch, France). Les spectres 2D ont été traité avec le logiciel Topspin 2.1 (Bruker) et analysés avec le logiciel CcpNmr Analysis [Vranken 2005]. Les échantillons consistaient en 150 l de protéine dont la concentration est notée dans le tableau B.1 dans le tampon phosphate (50 mM Na Phosphate pH 6.5, 150 mM NaCl, 1mM DTT et 10% D2O) en tube 3 mm. La séquence d'impulsion utilisée pour les spectres 1D ^{1}H et DOSY étaient `zgespgp` et `led-es` respectivement. Les expériences 2D ^{1}H-^{15}N-HSQC ont été enregistrées avec la séquence d'impulsion `hsqcf3gpph19CN`.

Polypeptide	Concentration (mM)
DBD-wt	0,25
W4-DBD	0,09
W3-DBD	0,03
W2-DBD	0,10
W2	0,80
W1-DBD	0,36
W1	2,00

Tableau B.1 – Concentrations des échantillons utilisées lors des expériences RMN.

B.2 Expériences SAXS

Les expériences de SAXS ont été réalisées au Synchrotron SOLEIL (Saint Aubin, Gif-sur-Yvette, France) sur la ligne SWING SAXS/WAXS équipée d'un système SE-HPLC (Agilent). Les échantillons de plus haute concentration ont été injectés sur une colonne Bio-SEC-3 (Agilent) pré-équilibrée avec le tampon de gel filtration supplémenté de 1 % de glycérol. Les échantillons de faible concentration ont été injectés en batch, c'est à dire directement injecté dans le capillaire relié au faisceau de rayons X. Les volumes injectés et les concentrations utilisées pour chaque expérience sont reportés dans le tableau B.2. La courbe de diffusion du

tampon a été enregistrée et utilisée comme référence pour être soustraite à la courbe de diffusion de chaque échantillon. 250 expositions d'1 ms ont été enregistré pour chaque échantillon. Les données ont été analysés avec le logiciel PRIMUS. La diffusion transmise I_0 et le rayon de giration R_g ont été obtenus par l'approximation de Guinier, $I(q) = I_0\, exp(-q^2 R^2{}_g\ 3)$ pour $qR_g <$ 0,8-1,3. La fonction de distribution des distances P(r) est obtenue par transformée de Fourier indirecte implémentée dans le programme GNOM [Svergun 1992]. Le diamètre maximum de chaque échantillon, D_{max}, a également été déterminé.

Polypeptide	Volume injecté (μl)	Concentration (mg/ml)	
		Batch	HPLC
DBD-wt	50		4
W4-DBD	80	1,8	
W3-DBD	80	2,4	
W2-DBD	80	1	
W2	70		2,5
W1-DBD	70		2,8
W1	50		6,8

Tableau B.2 – Concentrations et volumes injectés sur la colonnes HPLC ou directement dans le capillaire lors des expériences de SAXS pour chaque échantillon.

B.3 Dichroïsme circulaire

Les expériences de dichroïsme circulaire ont été enregistrées avec un spectropolarimètre Jasco J-815 (Easton, MD). Les données dans l'UV lointain ont été obtenues avec des cuvettes en quartz de 0,1 mm (Quartz-Suprasil, Helma UK Ltd). Les spectres ont été enregistrés à 25 °C avec des échantillons de 35 M dans le tampon phosphate (50 mM Na Phosphate/ pH 6,5/ 50 mM NaCl/ 0,3 mM DTT) entre 180-260 nm. Le spectre du tampon, obtenu dans des conditions identiques, a été soustrait des spectres de chaque échantillon. Le contenu en structure secondaire a été estimé avec le logiciel CDpro suite [Sreerama 2000] en utilisant la méthode CONTIN/LL. Les spectres ont été comparés à plusieurs bases de données de spectres de références et la base de données SMP56, donnant le plus petit RMSD, a été choisie.

Polypeptide	Structure secondaire	Contenu en (%)	RMS (%)
DBD-wt	α	25,1	1,1
	β	24,7	1,5
	autres	50,2	1,2
W4-DBD	α	20,8	1,2
	β	24,0	1,1
	autres	55,2	1,4
W3-DBD	α	22,7	1,3
	β	24,4	1,9
	autres	53,0	2,9
W2-DBD	α	31,4	1,6
	β	15,7	1,4
	autres	53,0	2,5
W2	α	7,0	2,1
	β	21,5	6
	autres	71,5	6
W1-DBD	α	29,5	1,8
	β	18,5	2,4
	autres	52,0	2,7
W1	α	17,0	2,8
	β	13,7	4,6
	autres	69,3	6,6

Tableau B.3 – Pourcentage de structures secondaires contenues dans les différents polypeptides.

B.4 Spectroscopie de fluorescence

Les mesures de fluorescence ont été réalisées avec un spectrofluorimètre SPEX Fluorolog 3 (Jobin Yvon Inc., Edison, NJ, USA) équipé d'une lampe à Xenon de 450 Watt, un monochromateur double d'excitation et un monochromateur simple d'émission. Les données ont été acquises avec un photomultiplicateur compteur de photon (linéaire jusqu'à 10^8 comptage/s) à 950 Volts. Les largeurs de fente ont été ajustées à 4 nm pour l'excitation et l'émission. Les échantillons de concentration 2 M ont été placés dans une cuvette en quartz maintenue à 25 °C. Les spectres de fluorescence ont été enregistrés en excitant l'échantillon à 295 nm et le signal d'émission a été enregistré entre 260-400 nm. Les spectres ont été enregistrés séquentiellement dans la même cuvette et systématiquement corrigés des fluctuations d'intensité provenant de la lampe.

Bibliographie

[Sreerama 2000] Sreerama N. et Woody R.W. *Estimation of Protein Secondary Structure from Circular Dichroism Spectra : Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set*. Analytical biochemistry, vol. 287, no. 2, 2000. – Cité page 174.

[Svergun 1992] Svergun D.I. *Determination of the regularization parameter in indirect-transform methods using perceptual criteria*. Journal of applied crystallography, vol. 25, no. 4, 1992. – Cité pages 74, 165 et 174.

[Vranken 2005] Vranken W.F., Boucher W., Stevens T.J., Fogh R.H., Pajon A., Llinas M., Ulrich E.L., Markley J.L., Ionides J. et Laue E.D. *The CCPN data model for NMR spectroscopy : Development of a software pipeline*. Proteins : Structure, Function, and Bioinformatics, vol. 59, no. 4, 2005. – Cité pages 173 et 204.

Chapitre 8

Etude du fragment W1-DBD par RMN

8.1 Production et purification du fragment **178**

8.2 Stratégie d'attribution des fréquences de résonances **179**

8.3 Publication : Attribution et étude de la structure secondaire résiduelle **182**

8.4 Etude de l'interaction du fragment W1-DBD avec un élément de réponse IR3 . . . . . **193**

8.4.1 Mise en place des expériences de titrage 193

8.4.2 Résultats 193

8.5 Conclusion **198**

Bibliographie **201**

Pour aller plus loin dans la caractérisation par RMN du domaine N-terminal du récepteur des androgènes (AR) nous avons décidé d'entamer une campagne d'attribution des fréquences de résonances de chaque sous-domaine sélectionné. Ce chapitre présente les résultats dont nous disposons actuellement, à savoir l'attribution des fréquences de résonances pour le fragment W1-DBD ou $NTD-DBD_{518-627}$ ainsi que la contribution du fragment NTD dans l'interaction du DBD avec l'ADN. Ce fragment est situé à l'extrémité C-terminale du NTD et fait partie de la région TAU5 de AR qui est impliquée dans l'activité ligand-indépendante du récepteur à travers des mécanismes encore inconnus.

L'attribution des fréquences de résonances du fragment W1-DBD a été réalisée dans le cadre d'un projet de stage pour un étudiant de deuxième année de l'Ecole Supérieure de Biotechnologie de Strasbourg (ESBS), Pau Pérez-Escrivà, dont j'ai encadré le travail.

La disponibilité de l'attribution du fragment W1-DBD a permis l'étude de la structure secondaire par analyse des déplacements chimiques des noyaux du squelette peptidique, ainsi que l'étude de l'interaction avec l'ADN par perturbation des déplacements chimiques. Ces résultats font l'objet d'une publication acceptée dans la revue *Journal of Biomolecular Assignments* .

8.1 Production et puri cation du fragment

La culture bactérienne, la surexpression et la purification du fragment W1-DBD doublement marquée $^{15}N/^{13}C$ ont été réalisées d'après le protocole décrit dans l'annexe A du chapitre 6.

La Figure 8.1 présente le gel SDS-PAGE des différentes étapes de purification.

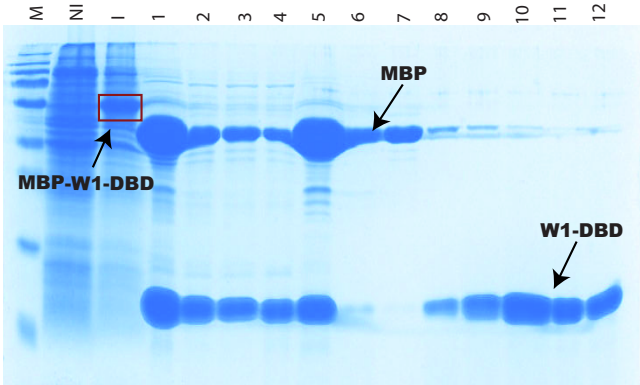


Figure 8.1 – Gel SDS-PAGE indiquant les différentes étapes de purification du fragment W1-DBD. NI : expression non-induite ; I : induction de l'expression ; 1 : chromatographie d'affinité sur résine amylose après coupure à la tev ; 2-5 : chromatographie d'affinité sur résine nickel ; 7-12 : gel filtration.

A savoir, les chromatographies d'affinité successives sur résine amylose et résine nickel suivie d'une chromatographie d'exclusion de taille ou gel filtration sur colonne Superdex75 16/60.

La chromatographie sur résine nickel est utilisée pour éliminer au maximum l'étiquette MBP et la Tev avant l'étape de gel filtration. En effet, ces deux protéines présentent une étiquette poly-His située en N-terminale leur permettant d'être retenues sur une résine nickel. Ainsi, le fragment W1-DBD passe au travers de la résine pour être récupéré dans le flow-through (pistes 2-4). La Figure 8.1 montre que, d'une part, toute la MBP n'est pas retenue sur la résine, et d'autre part, une partie de notre protéine d'intérêt est perdue sur la résine (piste 5).

L'étape de gel filtration permet alors de séparer le restant de MBP et le fragment W1-DBD en fonction du poids moléculaire.

Les fractions les plus pures (10-12) ont été réunies puis un contrôle qualité de l'échantillon a été réalisé par spectrométrie de masse. La masse moyenne attendue pour une protéine enrichie à 100% en isotope ^{15}N et ^{13}C est de 13416,56 Da. L'analyse révèle la présence d'une espèce majoritaire en solution de masse moyenne 13408,41 Da correspondant à un enrichissement d'environ 98,8% en isotopes ^{15}N et ^{13}C (figure 8.2).

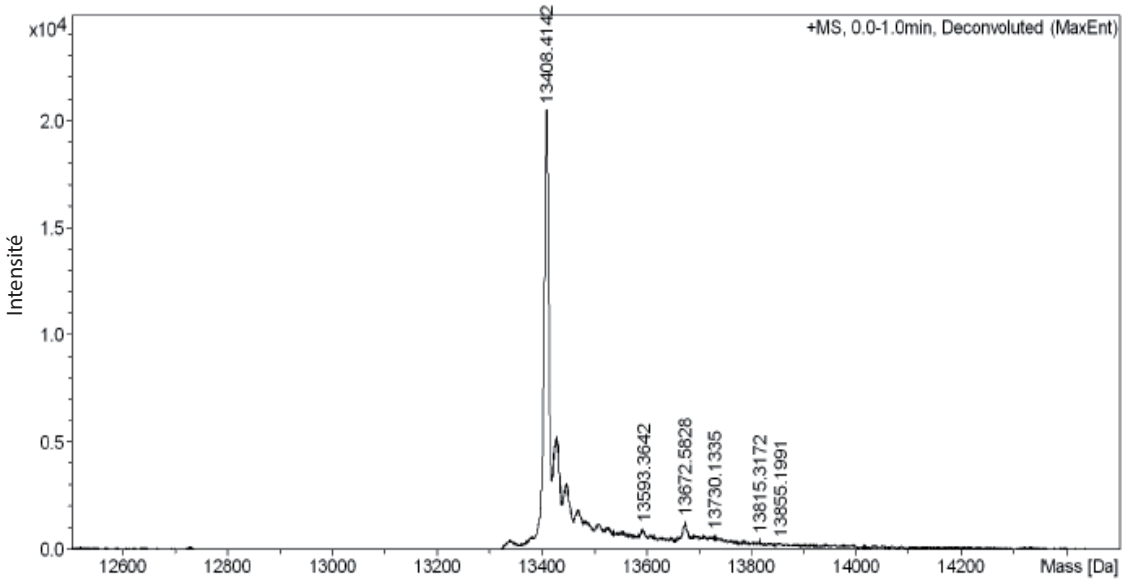


Figure 8.2 – Analyse par spectrométrie de masse du fragment W1-DBD. Une espèce majoritaire est présente en condition dénaturante dont la masse moyenne est 13408,41 Da (masse moyenne attendue pour une protéine enrichie en isotope ^{15}N et ^{13}C : 13416,56 Da).

8.2 Stratégie d attribution des fréquences de résonances

L'attribution est une étape clé dans la détermination de la structure d'une protéine mais également lors des études d'interactions afin de déterminer les résidus mis en jeu. Pour les protéines de grandes tailles, le marquage isotopique ^{15}N et/ou ^{13}C permet de simplifier les spectres et donc le travail d'attribution qui en découle. En effet, les expériences RMN hétéronucléaires multidimensionnelles, à deux ou trois dimensions, corrèlent la résonance d'un proton avec celle d'un ou plusieurs hétéronoyaux comme l'azote 15 ou le carbone 13. La grande diversité des expériences RMN disponibles diffèrent par le type de noyau impliqués dans les corrélations ainsi que le type d'interaction mise en oeuvre dans l'expérience. La stratégie classique d'attribution séquentielle pour une protéine marquée à l'azote 15 ou le carbone 13 utilise des couplages $^1J_{CC}$, $^1J_{NC}$ et $^2J_{NC}$ pour établir corrélations entre les atomes voisins au sein d'un acide aminé, ainsi qu'au travers de la liaison peptidique. La carte d'identité de la protéine, point de départ de l'attribution, est fournie par le spectre ^{1}H-^{15}N-HSQC. Ce spectre est présenté dans la section 8.3 correspondant à la publication des fréquences de résonances du squelette peptidique pour ce fragment.

Les expériences triple résonance permettent d'attribuer les fréquences de résonances des

atomes composant le squelette peptidique de la protéine, à savoir le proton amide (H_N), l'azote (N), les carbones alpha et beta (C , C) ainsi que le carbonyle (CO). Les expériences utilisées sont la HNCA, HNCACB, CBCA(CO)NH, et la HNCO (Figure 8.3).

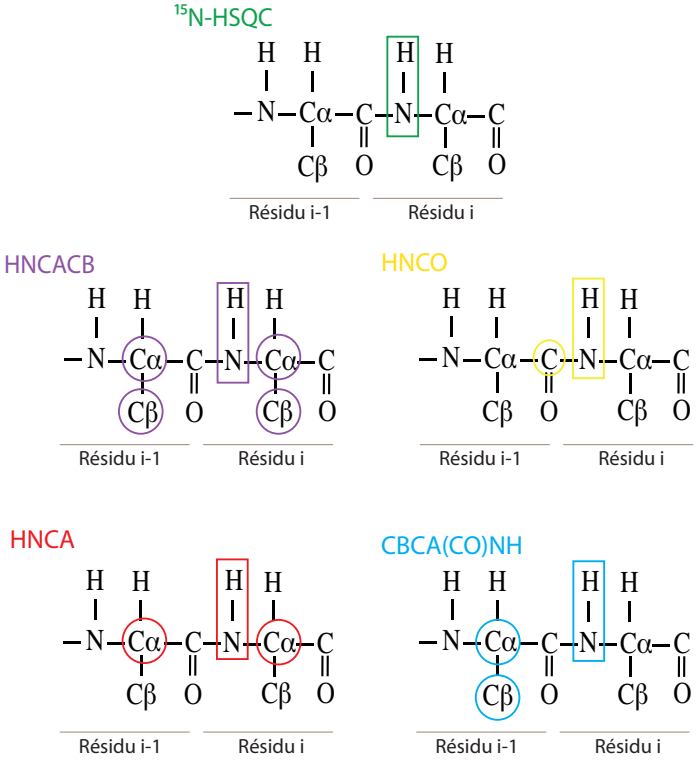


Figure 8.3 – Jeu d'expériences RMN multidimensionnelles permettant l'attribution séquentielle des fréquences de résonances du squelette peptidique. Les noyaux corrélés sont entourés pour chacune des expériences utilisées :^{15}N-HSQC, HNCA, HNCACB CBCA(CO)NH, HNCO.

Ces expériences sont complémentaires les unes des autres et possèdent toutes un plan de base ^{1}H-^{15}N-HSQC permettant de relier le groupement amide d'un résidu (i) à des noyaux ^{1}H et ^{13}C du résidu (i) ou (i-1). La troisième dimension permet de corréler les fréquences des noyaux C , C ou CO à un couple H_N-N selon l'expérience. Les corrélations obtenues par ces différentes expériences sont résumées dans le tableau 8.1

Expérience RMN	Noyaux observés
^{15}N-HSQC	H_N(i) et ^{15}N(i)
HNCO	H_N(i), ^{15}N(i), ^{13}CO(i-1)
HNCA	H_N(i), ^{15}N(i), ^{13}C (i) et ^{13}C (i-1)
HNCACB	H_N(i), ^{15}N(i), ^{13}C (i-1), ^{13}C (i-1),^{13}C (i) et ^{13}C (i)
CBCA(CO)NH	H_N (i), ^{15}N(i), ^{13}C (i-1), ^{13}C (i-1)

Tableau 8.1 – Expériences RMN hétéronucléaires multidimensionnelles utilisées pour l'attribution des fréquences de résonances du fragment W1-DBD et corrélations observées.

Les attributions partielles de AR-DBD [555-627] étaient disponibles à la suite des travaux de thèse de Ying-Hui Wang au laboratoire, et ont facilité le travail d'attribution du fragment plus long. En pratique, à partir d'une tâche de corrélation provenant d'un résidu (i) dans le spectre ^{1}H$-^{15}$N-HSQC et en se plaçant dans le plan correspondant à la fréquence azote dans les expériences 3D, il est possible d'identifier les déplacements chimiques des C (i), C (i-1), C (i-1) et CO (i-1). Les valeurs de ces déplacements chimiques, obtenues expérimentalement, sont ensuite comparées aux valeurs statistiques de la banque de données BMRB (*Biological Magnetic Resonance Data Bank*) pour le type de résidu considéré.

Cependant, la faible dispersion des signaux des protons amides représente une difficulté pour l'attribution des parties désordonnées. Nous nous sommes basés sur l'identification non-ambigue de résidus de type particuliers sur la base des fréquences de résonances de leur C_β. Ces types de résidus sont notamment la sérine et thréonine dont la valeur de déplacement chimique de leur C est supérieure à celle de leur C . De plus, un déplacement chimique pour un C aux alentours de 20 ppm voir en dessous correspond typiquement à l'acide aminé alanine. Enfin, la glycine est également facilement identifiable puisque c'est l'unique résidu ne possédant pas de C et la valeur de déplacement chimique de son C se situe aux environs de 45 ppm alors que pour les autres résidus cette valeur est comprise entre 50 et 60 ppm.

La séquence primaire de la région NTD du fragment W1-DBD contient 3 glycines, 3 sérines, 1 thréonine et 1 alanine. Il faut ensuite repérer des motifs particuliers de type S-G et T-A uniques au niveau de la séquence pour démarrer l'attribution séquentielle en allant d'un résidu (i) au précédent (i-1). Ainsi, de proche en proche nous sommes capables de parcourir l'ensemble de la séquence primaire d'intérêt. L'attribution des prolines nécessite des expériences 3D complémentaires qui n'ont pas été mises en oeuvre ici. De ce fait, l'attribution a été faite par segments correspondants au nombre de résidus proline +1 présents dans la séquence. Dans notre cas, 6 prolines sont présentes sur l'ensemble de la séquence dont 2 qui se suivent et nous

avons donc 6 segments attribués.

Le squelette peptidique a été attribué à raison de 98% pour les proton amides, 96% et 94% pour les C et C respectivement et 92% pour les CO. Hormis les résidus prolines, les résidus D604 et C611 n'ont put être attribués du fait de l'absence de corrélation. Des expériences conduites à 15 °C ont permis de retrouver ces corrélations mais les résonances correspondantes étaient plus larges témoignant un échange conformationnel et donc un mouvement lent de ces résidus. De plus, pour certains résidus, nous observons la présence de tâches de corrélation secondaires de faible intensité et avec un décalage de la fréquence en proton qui suggère l'existence d'une autre conformation minoritaire de la protéine en solution. Enfin, l'attribution des fréquences de résonances du fragment W1-DBD à 25 °C a été déposée sur BioMagResBank (http://www.bmrb.wisc.edu/) sous le numéro d'accession 26645.

8.3 Publication : Attribution et étude de la structure secondaire résiduelle

Le domaine AR-NTD est le siège de nombreuses modifications post-traductionnelles et d'interactions protéine/protéine qui font que ce récepteur est impliqué dans de nombreuses voies de signalisations pour réguler l'expression des gènes cibles. Nous avons attribué les fréquences de résonances du squelette peptidique du fragment W1-DBD, nommé $\mathrm{NTD-DBD}_{518-627}$ dans la publication acceptée dans la revue *Journal of Biomolecular Assignments* .

Les résultats indiquent que les 37 derniers résidus de AR-NTD présentent une propensité d'environ 20% à former une structure secondaire en hélice- . Ces résultats pourront être utilisés dans des futures études d'interactions avec des partenaires de cette région du récepteur mais également avec des séquences d'ADN pour établir le rôle de cette région dans l'interaction de AR avec l'ADN.

Meyer Sandra, Wang Ying-Hui, Pérez-Escrivà Pau and Kieffer Bruno (2015), ***Backbone ^{1}H, ^{15}N, ^{13}C NMR assignment of a part of the N-terminal and DNA binding domain of the androgen receptor***. *Biomolecular NMR Assignments, DOI 10.1007/s12104-015-9661-8.*

8.4 Etude de l interaction du fragment W1-DBD avec un élément de réponse IR3

8.4.1 Mise en place des expériences de titrage

L'interaction du fragment W1-DBD avec l'élément de réponse IR3 de 19 paires de base Slp (Sex-limited protein), issu du promoteur d'une protéine du sérum chez la souris, a été suivie par RMN à 25 °C. L'expérience de titrage protéine-ADN peut être réalisée de deux façons différentes : soit par ajouts successifs de petits volumes d'une solution concentrée d'ADN dans un échantillon de protéine, soit par ajout successif d'une solution concentrée de protéine à un échantillon d'oligonucléotide.

Dans notre étude, nous avons opté pour la première méthode en raison de la difficulté à concentrer la protéine du fait de la propension à s'agréger à haute concentration. La formation du complexe protéine-ADN est suivie en enregistrant un spectre proton adapté pour l'observation des protons-imino de l'ADN, ainsi qu'une expérience $^{1}H-^{15}N$-SOFAST-HMQC à chaque ajout.

Le signal correspondant aux protons-imino offre une signature des appariements Watson-Crick, caractéristiques d'une séquence d'ADN double brin. En effet, l'échange des protons-imino avec les protons du solvant est trop rapide pour permettre leur observation lorsque ces protons ne sont pas engagés dans une liaison hydrogène. La séquence d'impulsion utilisée est de type Jump & Return (JR) [Plateau P. 1982] qui permet d'éviter la saturation de l'aimantation de l'eau. Le délai caractéristique de la séquence JR (= 43,5 s) a été ajusté pour obtenir un maximum d'excitation autour de 13 ppm.

8.4.2 Résultats

Evolution des résonances des protons-imino au cours de l'interaction

L'évolution des spectres JR lors du titrage est représentée sur la Figure 8.4 dont l'agrandissement correspond à la région des protons-imino. Lors des premiers ajouts, la faible quantité d'oligonucléotides, ainsi que la largeur de raies importante (l'ensemble de l'ADN étant sous forme complexée du fait de l'excès de protéine) ne permet pas l'observation d'un signal.

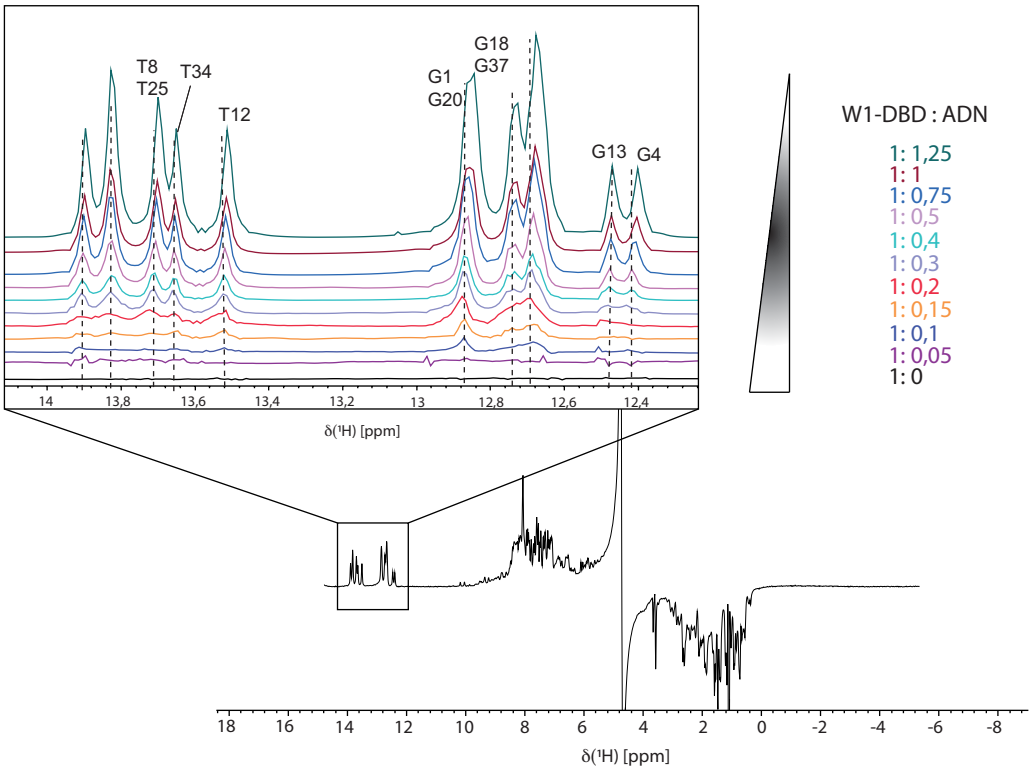


Figure 8.4 – Evolution du signal des protons-imino lors du titrage W1-DBD :ADN. L'encadré correspond à un agrandissement de la région des protons-imino où sont superposés les spectres 1D ^{1}H Jump & Return en présence de concentration croissantes d'ADN. Les spectres ont été enregistrés à 700MHz, 298K et pH=6,5.

Nous observons une augmentation de l'intensité des raies jusqu'à un rapport protéine-ADN de 1 : 0,5 (ce qui correspond à la fixation de deux molécules de DBD sur un oligonucléotides). L'ajout de quantités supplémentaires d'oligonucléotides conduit à un déplacement des raies vers les fréquences correspondant à la forme libre de l'oligonucléotide SLP. L'attribution des fréquences de résonances des protons-imino de cet oligonucléotide a été réalisé par Isabelle Lebars dans le cadre de l'encadrement d'un stage de master 2 (Wissem Medjaoui, 2011). Une partie de ces attributions est représentée sur la Figure 8.4. Les déplacements observés pour les protons-imino G_{13} et G_4 indiquent que les deux demi-sites de l'élément de réponse sont occupés.

Evolution des résonances de la protéine au cours de l'interaction

Les spectres ^{1}H-^{15}N-SOFAST-HMQC permettent de suivre l'interaction du côté de la protéine. Les spectres correspondant aux ratios protéine : ADN de 1 : 0, 1 : 0,5, 1 : 1 et 1 : 1,25

ont été superposés sur la Figure 8.5.

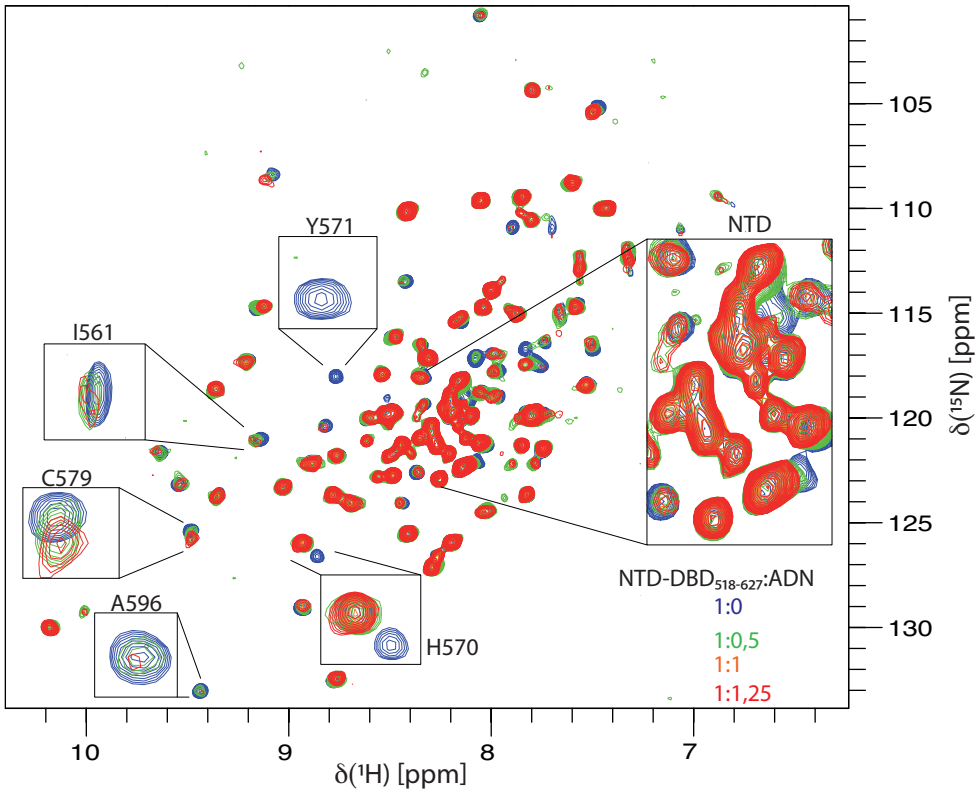


Figure 8.5 – Interaction suivie du côté de la protéine. Superposition des spectres 2D ^{1}H-^{15}N-SOFAST-HMQC du fragment W1-DBD en présence de concentrations croissantes d'ADN. Les différents agrandissements montrent le comportement différent que peuvent avoir les résidus au cours d'une expérience de titrage. Les spectres ont été enregistrés à 700MHz, 298K et pH=6,5.

Les résultats montrent que les résidus ne présentent pas tous le même comportement au sein du complexe. En effet, l'intensité diminue pour certains résidus jusqu'à disparaître du spectre car le pic s'élargit de façon importante. Ce phénomène est observé pour les résidus appartenant au DBD se trouvant à l'interface protéine/ADN, comme Y571 par exemple, ou dans l'interface de dimérisation protéine/protéine entre les deux DBDs comme par exemple A596.

L'élargissement des pics de corrélation ^{1}H$-^{15}$N est lié à une contribution de l'équilibre entre les formes libres et complexées à l'ADN de la protéine. En effet, la constante de dissociation apparente mesurée pour l'oligonucléotide SLP est de l'ordre de 10 M (résultats préliminaires).

L'effet de l'échange entre les formes libres et liées sur les fréquences de résonance dépend du carré de la différence de fréquence entre les deux formes. Pour un échange dit rapide

nous pouvons ainsi observer un décalage en fréquence pour une faible différence de fréquence entre la forme libre est liée, et un élargissement des raies pour une différence importante.

Enfin, l'observation d'un pic double pour I561 est plus surprenante, car ce résidu est situé entre deux cystéines impliquées dans la coordination du premier doigt de zinc, localisé à distance de l'oligonucléotide. Les deux formes pourraient être liées à une légère asymétrie du premier doigt de zinc. En effet, le proton amide de I561 est engagé dans une liaison hydrogène avec le souffre de C559, ce qui le rend très sensible à toute modification de la sphère de coordination du zinc.

Pour une analyse détaillée de cette expérience, les différences de déplacement chimique composites (ppm) ont été calculées pour chaque résidu d'après l'équation présentée dans le chapitre 3, pour les ratios protéine : ADN 1 : 0,5 et 1 : 1. La Figure 8.6 montre la différence de déplacement chimique pour le ratio 1 : 0,5 (ces différences n'évoluent pas de manière significative entre les deux points de titrage).

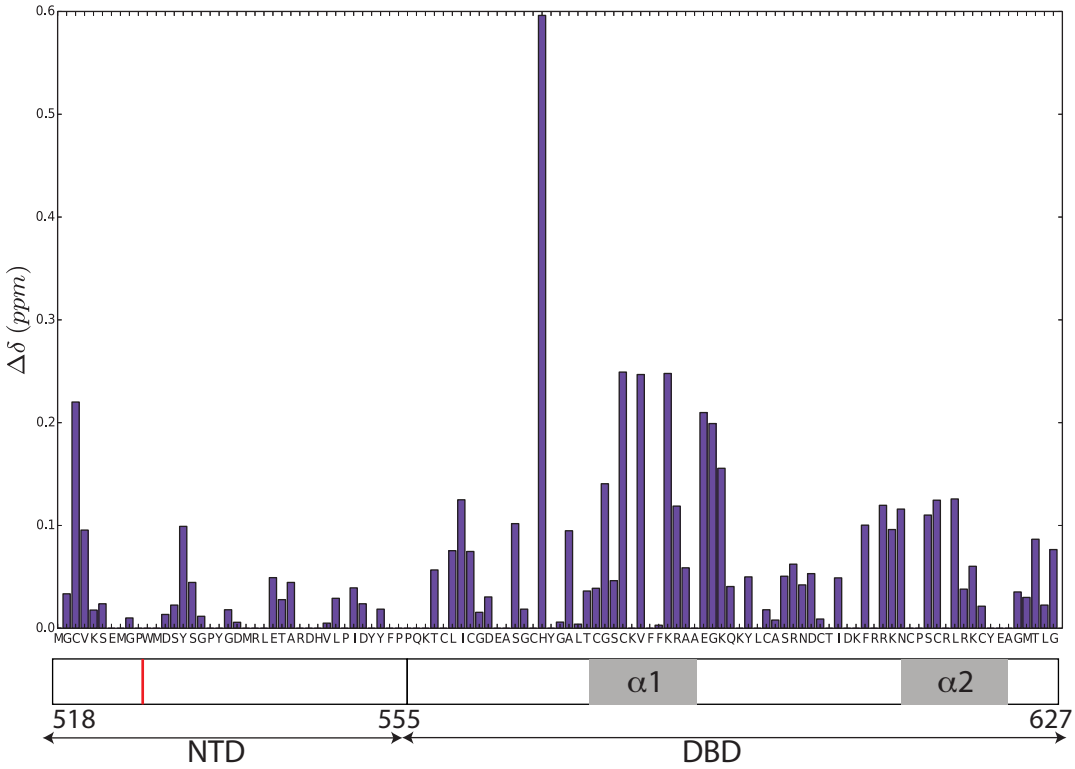


Figure 8.6 – Perturbation des déplacements chimiques pour chaque résidu entre l'état libre et lié du fragment W1-DBD. La structure secondaire du DBD provenant de la structure cristallographique du DBD lié à l'ADN est représenté par les hélices $_1$ et $_2$ (PDBID : 1R4I)

Comme attendu, les déplacements chimiques, correspondant à une modification de l'environnement chimique local d'un noyau, sont plus importants pour les résidus appartenant au DBD et plus particulièrement ceux situés aux interfaces protéine/ADN et protéine/protéine. En particulier, le résidu H570 qui établit une liaison hydrogène avec le squelette sucre/phosphate de l'ADN présente le déplacement chimique le plus important (chapitre 5). Des déplacements de fréquence sont également observés pour les protons amides des résidus du NTD. Ceci pourrait être lié à des contacts transitoire entre des régions du NTD et l'ADN ou le DBD, et concerne en particulier les résidus autour des positions C520 et Y532.

Dans l'expérience SOFAST, l'intensité des pics de corrélation est sensible à la relaxation T2 des noyaux H_N et ^{15}N, mais également à la relaxation T1 de l'ensemble des protons non-échangeables de la protéine, qui constitue un réservoir d'aimantation, favorisant un retour à l'équilibre rapide de l'aimantation des H_N après une excitation. L'analyse comparée des intensités des raies pour les formes libres et liées de la protéine met en évidence une réponse contrastée entre la partie repliée de la protéine, le DBD, et sa partie désordonnée, le NTD (Figure 8.7).

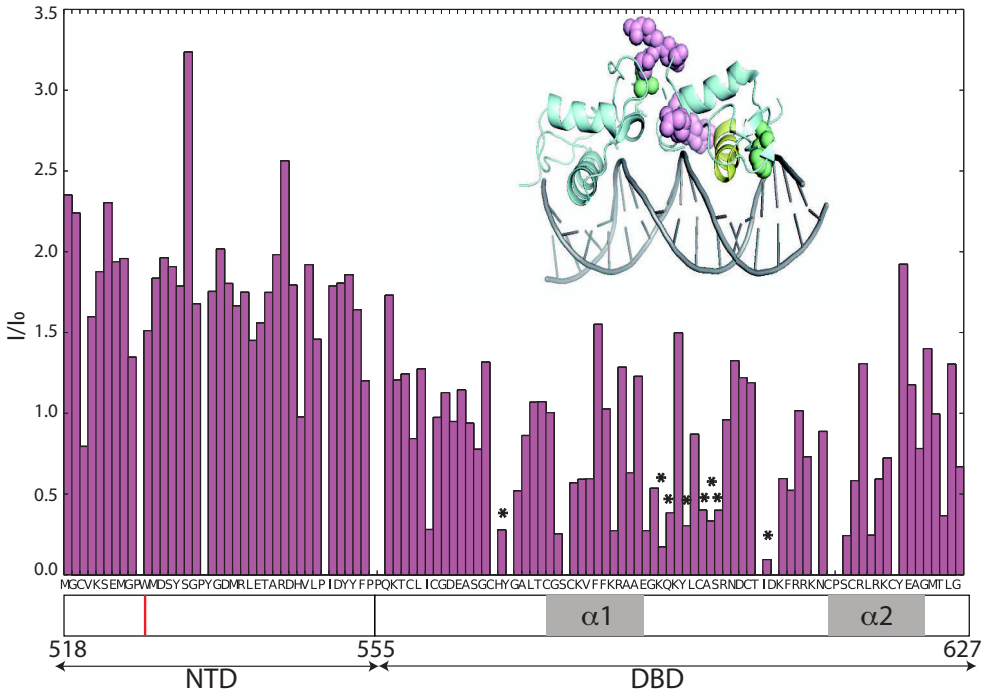


Figure 8.7 – Rapport des intensités de chaque résidu entre l'état libre et lié du fragment W1-DBD. Le structure cristallographique correspondant à un homodimère de DBD de AR sur l'ADN est représentée (PDB : 1R4I). L'hélice $\alpha 1$ du DBD est représentée en jaune. Les étoiles correspondent à des résidus se trouvant dans l'interface de dimérisation (sphères roses), les résidus rapporteurs de l'interaction, H570, au contact de l'ADN,et l'A596 à l'interface de dimérisation (sphères vertes.)

Alors que la formation du complexe conduit à une baisse d'intensité pour la plupart des résidus du DBD, une augmentation significative de l'intensité des raies est observée pour les résidus du NTD. Cette augmentation est liée à une meilleure efficacité de l'expérience SOFAST du à une relaxation T1 plus rapide des H_N. Cet effet peut s'expliquer par l'existence de contacts transitoires entre la partie NTD de AR et l'oligonucléotide ou bien une modification de la dynamique de l'ensemble conformationnel du NTD.

Des expériences complémentaires de type hetSOFAST devront être menées pour établir la source supplémentaire de relaxation T1 conduisant à des expériences SOFAST plus efficaces [Schanda 2005].

8.5 Conclusion

La contribution du domaine NTD dans l'interaction DBD/ADN a déjà été décrite dans la littérature pour l'AR [Palvimo 1993]. Les études à ce sujet montrent que la présence du NTD diminue l'affinité de AR-DBD pour ses éléments de réponse.

Gast et coll. ont montré que la délétion des résidus 510-536, 488-536 ou 447-536 au niveau de l'extrémité C-terminale du NTD mène à une augmentation de l'activité de liaison du DBD à l'ADN par rapport au récepteur sauvage [Gast 1998]. Ceci montre que la région 447-536 de AR-NTD est probablement impliquée dans la modulation de l'interaction de AR-DBD avec l'ADN. Ils proposent que la contribution du NTD dans l'interaction DBD/ADN soit due à la présence de structure secondaire permettant le recrutement de co-régulateurs plutôt qu'à un effet de séquence.

Une autre étude a permis d'identifier 81 résidus du NTD directement adjacent au DBD ayant un effet inhibiteur de la liaison du DBD sur une séquence consensus d'ADN non spécifique de AR [Liu 2003].

Brodie et coll. identifient une régulation allostérique de AR qui passe par un changement de conformation du domaine NTD au contact de l'ADN par dichroïsme circulaire et protéolyse [Brodie 2005]. Les auteurs suggèrent également une communication intra-domaine entre le DBD et NTD pour expliquer la diminution de l'affinité de liaison du récepteur en présence du NTD. Ceci a déjà été suggéré auparavant lors d'études utilisant des récepteurs chimériques AR/GR. En effet, l'utilisation d'une protéine composée de AR-NTD/GR-DBD conduit à l'activation de la transcription du gène de AR qui contient un élément de réponse pour AR dans sa région promotrice, alors que ceci n'est pas observé dans le cas d'une protéine chimérique GR-NTD/AR-DBD [Grad 2001]. Ces études mettent en avant la nécessité de la présence de AR-NTD dans la régulation des gènes cibles de AR.

Des effets similaires ont également été observés pour le récepteur des glucocorticoïdes (GR) [Eriksson 1990, Dahlman-Wright 1992].

Bien que toutes ces études montrent l'implication du NTD dans la liaison d'un récepteur nucléaire à l'ADN, aucune d'entre elle n'a encore montré clairement l'existence d'une interaction NTD/DBD ou NTD/ADN.

L'attribution des fréquences de résonances du fragment W1-DBD fournit les premières informations sur la conformation et la dynamique de la région TAU5, qui est importante pour l'activité transcriptionnelle ligand-indépendante de AR.

Nous montrons que ce fragment du NTD présente une propensité d'environ 20% à former une structure secondaire en hélice, ce qui est cohérent avec la présence de nombreux résidus hydrophobes conservés. Ces résultats corroborent ceux obtenus par dichroïsme circulaire au chapitre 7. Nous mentionnons la présence d'une forme minoritaire en solution qui pourrait provenir de l'isomérisation cis/trans d'une ou plusieurs prolines situées dans le NTD.

La comparaison des déplacements chimiques mesurés pour le DBD et pour le fragment W1-DBD indique que la présence de ce fragment NTD n'altère pas la structure tridimensionnelle du DBD. Cependant, le fragment NTD peut être impliqué dans des contacts transitoires avec la surface du DBD. De plus, la présence de la partie désordonnée n'altère pas la capacité de lier l'ADN sous la forme d'un homo-dimère symétrique du DBD puisque l'intensité du signal de l'A596, qui est notre rapporteur de la dimérisation du DBD de AR, diminue au cours du titrage de la même façon que pour le récepteur sauvage.

Les différences de relaxation T1 dans l'expérience SOFAST sont le reflet des différences structurales et dynamiques au sein d'une même protéine. Nos résultats laissent penser que cette courte région NTD contribue à l'interaction de AR-DBD avec l'ADN, bien qu'elle reste fondamentalement désordonnée. Les variations de l'affinité seraient expliquées par une modulation de l'ensemble conformationnel accessible par le NTD. Cette hypothèse est inspirée des travaux récents de Vincent Hilser, qui propose un mécanisme par lequel une région désordonnée d'une protéine peut être un effecteur allostérique d'une partie repliée [Hilser 2007].

Une analyse plus fine devra permettre de déterminer l'origine de cet effet qui peut être dû soit à une modification de l'ensemble conformationnel décrit par le NTD, et/ou des variations de la vitesse d'échange des protons amides avec ceux de l'eau. Les expériences RMN de type EXSY ou CLEANEX permettent de mesurer la constante d'échange entre les H_N et les protons de l'eau, qui renseigne sur l'accessibilité des H_N par le solvant [Hwang 1998].

Ce type d'expérience renseigne sur l'accessibilité de la protéine au solvant mais également sur la densité de proton et donc sur la compaction globale d'une chaîne polypeptidique. Il est alors possible de faire la distinction entre une partie globulaire et une partie flexible au sein

d'une même protéine. Ainsi, si le fragment NTD se trouve à proximité de l'ADN les H_N seront alors protégés du solvant.

Enfin, des données quantitatives sur les contributions énergétiques de la fixation à l'ADN devront être mesurées pour déterminer l'impact de la région NTD sur l'affinité de liaison à l'ADN. Ces résultats seront dans un second temps comparés à ceux obtenus lors de l'interaction de AR-DBD isolé avec ses éléments de réponse. Des résultats préliminaires obtenus par Emmanuel Oppong indiquent que la présence du NTD entraine une diminution de l'affinité du DBD pour l'ADN.

L'ensemble de ces résultats indiquent que cette région du AR-NTD présente des caractéristiques structurales et biochimiques suggérant son implication dans les interactions protéines-protéines avec des co-régulateurs du récepteur mais également dans l'interaction du récepteur avec l'ADN et probablement dans la régulation de la spécificité de reconnaissance des types d'éléments de réponse. Une combinaison de plusieurs mécanismes est sans doute à l'origine de la spécificité de reconnaissance de l'ADN par AR, parmi lesquels nous pouvons supposer un changement de l'environnement du résidu H570 de AR-DBD induit par la présence de la région NTD. En effet, ce résidu présente un déplacement chimique important en présence de la région NTD. De plus, ce résidu établit une liaison hydrogène avec un phosphate de l'ADN lors de l'interaction [Shaffer 2004].

L'attribution des fréquences de résonances de ce fragment sont utiles pour étudier les interactions protéines-protéines et protéines-ADN qui interviennent dans la régulation fine de la fonction de AR. Ce fragment permet également d'envisager l'étude des mécanismes moléculaires de signalisation par modification post-traductionnelle du NTD. En effet, le fragment W1-DBD contient deux sites de phosphorylation conservés (sur des tyrosines) ainsi qu'un site de sumoylation.

Bibliographie

[Brodie 2005] Brodie J. et McEwan I.J. *Intra-domain communication between the N-terminal and DNA-binding domains of the androgen receptor : modulation of androgen response element DNA binding*. Journal of Molecular Endocrinology, vol. 34, no. 3, 2005. – Cité page 198.

[Dahlman-Wright 1992] Dahlman-Wright K., Wright A. et Gustafsson J.A. *Determinants of high-a nity DNA binding by the glucocorticoid receptor : evaluation of receptor domains outside the DNA-binding domain*. Biochemistry, 1992. – Cité page 199.

[Eriksson 1990] Eriksson P. et Wrange O. *Protein-protein contacts in the glucocorticoid receptor homodimer in uence its DNA binding properties.* Journal of Biological Chemistry, vol. 265, no. 6, 1990. – Cité page 199.

[Gast 1998] Gast A., Schneikert J. et Cato A. *N-terminal sequences of the human androgen receptor in DNA binding and transrepressing functions*. The Journal of steroid biochemistry and Molecular Biology, vol. 65, no. 1-6, pages 117–123, 1998. – Cité page 198.

[Grad 2001] Grad J.M., Lyons L.S., Robins D.M. et Burnstein K.L. *The androgen receptor (AR) amino-terminus imposes androgen-speci c regulation of AR gene expression via an exonic enhancer.* Endocrinology, vol. 142, no. 3, 2001. – Cité page 198.

[Hilser 2007] Hilser V.J. et Thompson E.B. *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins.* Proceedings of the National Academy of Sciences, vol. 104, no. 20, 2007. – Cité page 199.

[Hwang 1998] Hwang T.L., van Zijl P.C. et Mori S. *Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme.* Journal of biomolecular NMR, vol. 11, no. 2, 1998. – Cité page 199.

[Liu 2003] Liu G.Z., Wang H. et Wang Z. *Identi cation of a highly conserved domain in the androgen receptor that suppresses the DNA-binding domain-DNA interactions.* Journal of Biological Chemistry, vol. 278, no. 17, 2003. – Cité page 198.

[Palvimo 1993] Palvimo J.J., Kallio P.J., Ikonen T., Mehto M. et Jänne O.A. *Dominant negative regulation of trans-activation by the rat androgen receptor : roles of the N-terminal domain and heterodimer formation.* Molecular Endocrinology, vol. 7, no. 11, 1993. – Cité page 198.

[Plateau P. 1982] Plateau P. G.M. *Exchangeable proton NMR without Base-line distorsion using new strong-pulse sequences*. Journal of American Chemical Society, vol. 104, pages 7310–7311, 1982. – Cité page 193.

[Schanda 2005] Schanda P., Kup e E. et Brutscher B. *SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds*. Journal of biomolecular NMR, vol. 33, no. 4, 2005. – Cité page 198.

[Shaffer 2004] Shaffer P.L., Jivan A., Dollins D.E., Claessens F. et Gewirth D.T. *Structural basis of androgen receptor binding to selective androgen response elements*. Proceedings of the National Academy of Sciences, vol. 101, no. 14, 2004. – Cité pages 16 et 200.

C.1 Acquisition, interprétation des spectres RMN du fragment W1-DBD

Etude de la structure secondaire

Pour déterminer la présence de structure de secondaire du fragment NTD−DBD$_{518-627}$ nous avons utilisé le programme SSP (Secondary Structure Propensities) [Marsh 2006]. Ce programme de prédiction combine les déplacements chimiques des noyaux du squelette peptidique et renvoie un score représentant la fraction de structure / . Les scores de -1 et +1 correspondent à des structures secondaires complète de type feuillets- et hélices- respectivement.

Le score est calculé pour un résidu i et prend en compte les deux résidus précédents et suivants dans la séquence.

$$\mathrm{SSPi} = \frac{\sum_{\mathrm{j=i-2}}^{\mathrm{i+2}} \sum_{\mathrm{X}} \begin{cases} \frac{|\quad \mathrm{X_{jobs}}|}{\rho_{\mathrm{j}}} \text{ si } (\quad \mathrm{X_{j}obs})(\Delta\delta \mathrm{X_{j}}\alpha) > 0 \\ -\frac{|\quad \mathrm{X_{jobs}}|}{\rho_{\mathrm{j}}} \text{ si } (\quad \mathrm{X_{j}obs})(\Delta\delta \mathrm{X_{j}}\beta) > 0 \end{cases}}{\sum_{\mathrm{j=i-2}}^{\mathrm{i+2}} \sum_{\mathrm{X}} \begin{cases} \frac{|\quad \mathrm{X_{j}}\quad |}{\rho_{\mathrm{j}}} \text{ si } (\quad \mathrm{X_{j}obs})(\Delta\delta \mathrm{X_{j}}\alpha) > 0 \\ -\frac{|\quad \mathrm{X_{j}}\quad |}{\rho_{\mathrm{j}}} \text{ si } (\quad \mathrm{X_{j}obs})(\Delta\delta \mathrm{X_{j}}\beta) > 0 \end{cases}} \tag{C.1}$$

X_{jobs} correspond au déplacement chimique secondaire du noyau X du résidu j. $X_{j\alpha/\beta}$ correspond au déplacement chimique secondaire moyen du noyau X du résidu j s'il était dans une structure secondaire de type ou et ρ_j est la déviation standard de $X_{j\alpha/\beta}$. Les déplacements chimiques moyens pour des résidus présents dans une hélice- , feuillet- ou random coil proviennent de la base de données des déplacements chimiques référencés (RefDB)[Zhang 2003].

Les noyaux H_N et ^{15}N étant très sensibles aux conditions expérimentales tels que le pH et la température, nous avons utilisé les déplacements chimiques des $^{13}C_{\alpha}$, $^{13}C_{\beta}$.

Titration du fragment W1-DBD

Les expériences d'interaction ont été réalisées dans un tube RMN de 3mm. Des quantités croissantes de oligonucléotides double brin SLP (concentration stock = 0,96 mM) sont ajoutées

à 150 l de protéine W1-DBD doublement marquée ^{15}N, ^{13}C à 100 M dans le tampon phosphate (50 mM Na Phosphate/ pH 6.5/ 150 mM NaCl/ 1mM DTT/ 10% D_2O).

L'oligonucléotide double brin SLP a été obtenu d'après le protocole décrit dans le chapitre 5.

L'interaction est suivie par l'enregistrement d'expériences 1D ^{1}H Jump & Return et 2D ^{1}H-^{15}N-SOFAST-HMQC à chaque point de titration.

Les spectres RMN ont été enregistrés à 298 K sur un spectromètre Bruker Avance III 700 MHz équipé d'une cryosonde Z-gradient. Le traitement des spectres a été réalisé avec le logiciel TopSpin 2.1 (Bruker) et analysés avec les logiciels CcpNMR Analysis [Vranken 2005] et NMR-notebook 2.7. Pour augmenter le rapport signal/bruit des spectres 1D ^{1}H Jump & Return , un l'algorithme de débruitage urQRd (uncoiled random QR denoising) a été appliqué par Lionel Chiron [Chiron 2014].

Bibliographie

[Chiron 2014] Chiron L., van Agthoven M.A., Kieffer B., Rolando C. et Delsuc M.A. *E cient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry.* Proceedings of the National Academy of Sciences, vol. 111, no. 4, 2014. – Cité page 204.

[Marsh 2006] Marsh J.A., Singh V.K., Jia Z. et Forman-Kay J.D. *Sensitivity of secondary structure propensities to sequence di erences between alpha- and gamma-synuclein : implications for brillation.* Protein science : a publication of the Protein Society, vol. 15, no. 12, 2006. – Cité pages 77 et 203.

[Vranken 2005] Vranken W.F., Boucher W., Stevens T.J., Fogh R.H., Pajon A., Llinas M., Ulrich E.L., Markley J.L., Ionides J. et Laue E.D. *The CCPN data model for NMR spectroscopy : Development of a software pipeline*. Proteins : Structure, Function, and Bioinformatics, vol. 59, no. 4, 2005. – Cité pages 173 et 204.

[Zhang 2003] Zhang H., Neal S. et Wishart D.S. *RefDB : a database of uniformly referenced protein chemical shifts*. Journal of biomolecular NMR, 2003. – Cité pages 76 et 203.

Chapitre 9

Conclusion générale

L'étude présentée ici a pour but de comprendre les mécanismes de reconnaissance de l'ADN et le rôle du domaine N-terminal (NTD) désordonné dans la fonction des récepteurs nucléaires.

Dans une première partie, nous avons vu qu'une mutation faux-sens dans le domaine de liaison à l'ADN (DBD) du récepteur des androgènes (AR) (T575A), située à distance de l'ADN, altérait son mode d'interaction avec l'ADN. Les études préliminaires menée par Ying-Hui Wang, dans le cadre de ses travaux de thèse, montrent que les structures tridimensionnelles des DBDs wt et T575A sont quasiment identiques et que l'unique différence est l'état de protonation du résidu H570. Ces résultats n'expliquent pas les différences observées au niveau cellulaire et il n'est pas possible d'établir un lien structure/fonction. C

Pour comprendre ce phénomène de dérégulation nous avons entrepris d'étudier la formation du complexe DBD/ADN en solution par RMN. Pour cela, nous avons utilisé deux séquences d'ADN provenant d'ARE ne différant que d'une seule base. L'une de ces séquences est reconnue uniquement par AR alors que l'autre est reconnue également par GR.

La caractérisation des interactions AR-DBD/ADN repose sur des expériences de titrage protéines-oligonucléotides suivies par RMN. Le marquage isotopique de la protéine permet de suivre l'évolution des espèces en présence (DBD libre ou lié, oligonucléotide) au cours de la titration.

Ces expériences indiquent finalement que les monomères de DBD (sauvage et mutant) forment un homodimère de type tête à tête pour les deux types d'éléments de réponse (AR et AR/GR). Les différences de déplacements chimiques se situent essentiellement au sein de l'interface de dimérisation du DBD, ainsi qu'au niveau des acides-aminés en contact direct avec l'ADN.

Ces données structurales ont été complétées par des mesures de titration calorimétrique isotherme (ITC) et de spectrométrie de masse dans le cadre du stage de Master 2 de *Justine WOLF* que j'ai eu l'occasion d'encadrer. Les résultats indiquent également que les DBDs mutant et sauvage présentent une affinité plus grande pour l'ADN non spécifique, préférence qui n'est pas altérée par la mutation T575A. La stoechiométrie des complexes a été vérifiée par spectrométrie de masse en condition native, qui montre l'existence d'un complexe ternaire

dans tous les cas.

L'ensemble de ces résultats suggère que les différences de propriétés de reconnaissance de l'ADN observées pour le mutant T575A sont liées à des altérations de nature dynamique, plutôt que structurale. Ainsi, *in vitro* nous observons que la modification des énergies faibles entraîne un changement dans la dynamique de l'interaction. A partir de ces résultats, nous pouvons formuler l'hypothèse que les modifications des mécanismes de reconnaissance de l'ADN sont liées à une déstabilisation qui résulte d'une modification de l'interaction entre l'histidine H570 et le squelette phosphate de l'ADN. Dans le contexte cellulaire, ces modifications d'énergies faibles sont associées à des pathologies comme le cancer de la prostate pour la mutation T575A, ou le syndrome d'insensibilité partielles aux androgènes nommé syndrome de Reifenstein lors de la mutation A596T située dans l'interface de dimérisation [1].

Ces observations corroborent les résultats issus de plusieurs études récentes qui montrent que les récepteurs nucléaires constituent des effecteurs allostériques complexes, capables d'intégrer des signaux moléculaires de natures différentes (modifications post-traductionnelles, séquence d'ADN, interactions avec des ligands).

Dans une deuxième partie de ma thèse, nous nous sommes intéressés à la caractérisation du domaine NTD de AR. Après une première étude bio-informatique, nous avons identifié 5 régions présentant des caractéristiques de séquence particulières au sein du NTD. La mise au point d'un protocole de purification robuste a démontré notre capacité à produire des fragments de AR-NTD avec une qualité et une reproductibilité permettant leur étude à l'aide de méthodes biophysiques. Ces fragments ont été étudiés pour la première fois par une combinaison de méthodes biophysiques telles que la fluorescence, le dichroïsme circulaire, le SAXS et la RMN. Nos résultats indiquent que les 5 régions identifiées sont majoritairement désordonnées mais présentent un comportement hydrodynamique différent en fonction de la zone considérée. En effet, les résultats contradictoires obtenus pour les fragments W2 et W2-DBD suggèrent la présence de plusieurs conformations instables pour cette région du NTD. De plus, les données RMN nous indiquent que le résidu tryptophane W526 se trouve dans un contexte de séquence particulier puisque sa fréquence de résonance est bien distincte de celle des 3 autres tryptophanes.

La présence du DBD en fusion avec les différentes régions NTD nous a permis de mettre en évidence une stabilisation de structure en hélice- de la région C-terminale de AR-NTD par CD qui a ensuite été confirmée par l'analyse des déplacements chimiques obtenus par RMN. Ces résultats suggèrent une communication interdomaine avec la formation de contacts

1. Kaspar et al., Molecular and Cellular Biology, Vol. 13, No.12, p. 7850-7858

transitoires entre le NTD et le DBD de AR. Le DBD nous sert également de sonde RMN pour étudier le rôle du domaine NTD dans l'interaction du récepteur avec l'ADN. En effet, le rôle de cette région désordonnée dans le mécanisme de reconnaissance de l'ADN par AR-DBD n'a pas encore été décrit Le gain de relaxation T1 observé pour les résidus du NTD lors des expériences d'interaction avec l'ADN du fragment W1 suggère que le NTD se trouve à proximité de l'ADN ou du DBD. Bien que ces résultats nécessitent d'être confirmés par des expériences de mesure du temps de relaxation et des expériences de type hetSOFAST, CLEANEX ou EXSY, nos résultats indiquent que l'extrémité C-termnale de AR-NTD est capable d'adopter une structure en hélice- , phénomène qui peut avoir lieu soit par communication interdomaine NTD/DBD, soit lors de l'interaction avec des partenaires comme l'ADN ou des protéines régulatrices.

Par ailleurs, de nombreuses études mettent en avant l'importance des régions désordonnées dans des processus complexes comme le contrôle de l'expression des gènes cibles. Dans leur revue, Hilser et Thompson décrivent du rôle clé du domaine NTD dans la régulation allostérique des récepteurs nucléaires[2]. Ainsi, le domaine NTD facilite le couplage allostérique entre les différents domaines des récepteurs nucléaires et leurs interactants. Enfin, des travaux récents suggèrent que ce domaine NTD existe comme un ensemble de conformations dont certaines peuvent être stabilisées par des interactions allostériques comme par exemple les modifications post-traductionnelles.

A court terme les perspectives de ce travail sont l'étude de l'interaction de cette région AR-NTD avec des co-régulateurs du récepteur induisant des modifications post-traductionnelles et l'observation de l'interaction à l'ADN. Il sera également intéressant d'étudier la partie N-terminale du NTD, notamment le rôle des différents homopolymères et l'importance de leur topologie. A long terme, l'étude du récepteur entier pourra permettre de disséquer les mécanismes allostériques ayant lieu entre les différents domaines et les effecteurs allostériques que sont le ligand, la séquence d'ADN, les modifications post-traductionnelles et les co-régulateurs. La compréhension du fonctionnement de ce domaine pourra permettre la développement de nouvelles stratégies thérapeutiques pour les patients atteint de cancer de la prostate en situation d'échappement au traitement.

2. Hilser V. J. and Thompson E. B., The Journal of Biological Chemistry, Vol. 286, No.46, p. 39675-39682

UNIVERSITÉ DE STRASBOURG

Sandra MEYER

igbmc
institut de génétique et de biologie moléculaire et cellulaire

Caractérisation des domaines NTD et DBD du récepteur des androgènes par des approches biophysiques.

Résumé

Cette thèse se situe à l'interface entre la biologie et la biophysique. Les méthodologies utilisées recouvrent principalement la résonance magnétique nucléaire (RMN), la diffusion des rayons X aux petits angles (SAXS), le dichroïsme circulaire (CD) et la spectroscopie de fluorescence. Une première partie vise à étudier le domaine de liaison à l'ADN (DBD) du récepteur des androgènes (AR) et les déterminants de l'interaction avec l'ADN. Une mutation faux-sens dans le DBD altère la spécificité de reconnaissance de l'ADN du récepteur bien que la structure tridimensionnelle soit identique au DBD sauvage. Les résultats montrent un changement dans la dynamique du récepteur mutant entrainant une déstabilisation de l'homo-dimère.
La seconde partie de ma thèse consiste à établir un lien séquence/fonction au niveau du domaine N-terminal (NTD) de AR.
D'après la littérature, cette région joue un rôle important pour l'activité du récepteur, et elle est également décrite comme étant intrinsèquement désordonnée. Les résultats révèlent que cette région établit des contacts transitoires avec le DBD. Ceci suggère l'existence d'un couplage allostérique entre le DBD et les résidus adjacents sur le NTD.
Ce couplage modifie l'ensemble conformationnel accessible au NTD en favorisant une conformation en hélice-α.

Mots-clés : AR, NTD, DBD, IDP, RMN

Abstract

My PhD project is at the boundary between biology and biophysic. Methods used include nuclear magnetic resonance (NMR), small ange X-ray scattering (SAXS), circular dichroïsm (CD) and fluorescence spectroscopy. The androgen receptor (AR) DNA binding domain (DBD) and its interaction with DNA was studied in a first part. A mutation in the DBD leads to a modified DNA recognition by the mutant compared to the wild-type. Our results indicate changes in dynamic of the mutant receptor that leads to the homodimer destabilisation.
The second part of my project aim to establish a link between sequence and function of the AR N-terminal domain (NTD).
As described in literature, this region is involved in the activity of the receptor and is also an intrinsically disordered protein (IDP). The results obtained during my thesis indicate that this region is involved in transient contact with the DBD. This suggest an allosteric coupling between the DBD and the neighboring residues on the NTD.
This coupling modifies the conformational ensemble accessible to the NTD by stabilizing a α-helix conformation.

Keywords: AR, NTD, DBD, IDP, NMR