



Synthèse et caractérisations de nouvelles sondes fluorescentes et réactifs chimiques pour l'étude structurale de l'ARN

Patrick Barhoum

► To cite this version:

Patrick Barhoum. Synthèse et caractérisations de nouvelles sondes fluorescentes et réactifs chimiques pour l'étude structurale de l'ARN. Autre. Université de Strasbourg, 2013. Français. NNT : 2013STRAJ032 . tel-00998937

HAL Id: tel-00998937

<https://theses.hal.science/tel-00998937>

Submitted on 3 Jun 2014

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.



Thèse de doctorat de l'Université de Strasbourg
École Doctorale des Sciences de la Vie et de la Santé
Spécialité : Chimie-Biologie

Présentée par: Patrick BARHOUM

Pour l'obtention du titre de
Docteur de l'Université de Strasbourg

**SYNTHÈSE ET CARACTÉRISATION DE NOUVELLES
SONDES FLUORESCENTES ET RÉACTIFS CHIMIQUES
POUR L'ÉTUDE STRUCTURALE DE L'ARN**

Présentée le 30 septembre 2013, à Strasbourg

Membres du jury :

Pr. Alain BURGER	Univ. Nice-Sophia Antipolis	Rapporteur externe
Pr. Mark HELM	Univ. Mainz-Allemagne	Rapporteur externe
Dr. Andrey KLYMCHENKO	Univ. Strasbourg	Examineur interne
Dr. Valérie VIVET-BOUDOU	Univ. Strasbourg	Directrice de thèse

A Papa, Mamam, Zouz, Edy, Fido, Aline et Joe-lynn

Remerciements

Tout d'abord j'aimerais remercier l'ensemble des membres du jury d'avoir accepté de juger mon travail de thèse, et pour l'intérêt qu'ils ont porté à ce travail : Dr. Alain BURGER professeur de l'Université de Nice-Sophia Antipolis, Dr. Mark HELM professeur de l'Université de Mainz en Allemagne, et Dr. Andrey KLYMCHENKO du Laboratoire de Biophotonique et Pharmacologie de l'Université de Strasbourg.

Je remercie également mon co-directeur de thèse Roland MARQUET de m'avoir accepté dans son équipe de recherche et de m'avoir encadré pendant ces trois années de thèse. Mes remerciements vont aussi à ma directrice de thèse Valérie VIVET-BOUDOU de m'avoir encadré et pour les nombreuses corrections du manuscrit de thèse.

Je remercie Eric WESTHOF de m'avoir accepté dans son unité de recherche et Catherine FLORENTZ de m'avoir accepté au sein de l'école doctorale des sciences de la vie et de la santé.

Merci à tous les membres de l'équipe « Rétrovirus et Virus à ARN » ceux d'aujourd'hui et ceux qui sont partis vers de nouvelles aventures, qui ont partagé avec moi cette expérience de thèse. Je vous remercie aussi pour votre bonne humeur: JC, Catherine, Julien, Marie, Anne et les jeunes doctorantes Camille et Élodie.

J'adresse mes remerciements à Serena de m'avoir encadré pour la partie fluorescence et caractérisation des sondes fluorescentes. Merci pour ta relecture enrichissante de la thèse et pour tes nombreux conseils. Merci aussi pour l'horoscope du jour, les fous rires, les pauses café, les repas Libanais, le vin Italien et beaucoup d'autres souvenirs...

Je tiens également à remercier Delphine. R, Red et Santi pour les moments qu'on a passés ensemble et toutes les discussions (en anglais ou en français) autour des repas de midi. Red, merci pour les soirées de fous chez toi et de m'avoir appris à faire du snowboard même si je garde un mauvais souvenir de ce week-end de ski. Santi, tu sais quoi on l'a faite enfin !!! On a finis la thèse !!! Et Delph une amie spéciale, on a passé des moments inoubliables chez toi avec ta petite famille adorable.

Merci à tous les membres de l'UPR 9002 et surtout le quatrième étage : les jeunes et les moins jeunes... Je remercie plus particulièrement « Eric ENNIFAR » pour la partie modélisation et « Philippe WOLF » pour la partie Spectrométrie de masse. Pour passer à la partie sérieuse de ces remerciements, je vous remercie tous de m'avoir appris à boire autant d'alcool...je n'oublierais jamais les pots de vendredi soir « et surtout les soirées bourrage de gueule ».

Je remercie aussi tous les membres des services RMN et spectrométrie de masse de la tour de chimie. Je tiens également à remercier le service d'analyses et de spectroscopie optique et plus particulièrement Martine HEINRICH.

Je tiens également à remercier toutes les sources qui ont aidé à financer ce projet de recherche.

Mes remerciements vont aussi à ma famille et mon soutien au quotidien, qui m'ont permis d'entreprendre des études supérieures dans les meilleures conditions. Je leur en suis très reconnaissant pour tous les sacrifices qu'ils ont dû accomplir pour subvenir à mes besoins. Maman et papa, j'espère que vous êtes fiers de moi. Zouz, Edy, Fido et Aline mes premiers fans, vos encouragements sont pour moi les piliers fondateurs de ce que je suis et de ce que je fais. Vous avez tous cru en moi et ouf ! Maintenant j'y suis ! Je vous remercie pour l'enthousiasme, la confiance, l'amour et toute l'attention avec une générosité sans compter que vous m'avez su apporter. Ma petite Joe-Lynn, je t'aime tellement ma princesse adorable.

Ces remerciements ne peuvent s'achever sans une pensée à la fille que j'aime tellement, à toi Ranine. Nannouna « Habibi » t'es quelqu'un d'unique et j'ai de la chance de t'avoir rencontré. La réponse à ta question : « parce que tu m'aimes autant, au-delà de toutes mes espérances ». T'étais toujours présente pour me soutenir mêmes dans les moments les plus difficiles de cette thèse et plus particulièrement durant les derniers mois de rédaction qui n'ont pas toujours été les plus agréables, alors pour tout cela un grand merci.

À Nancy, Vincent et Lynn. Nancy je te remercie pour tous les moments qu'on a passé ensemble à râler, à faire du shopping, pour les fous rires et les éclats de rire. Je te remercie aussi de m'avoir écouté et soutenu « ma psychothérapeute préférée ». Je me rappelle la première fois qu'on s'est croisé et ça me fait toujours rigoler...

À Marie, Philippe et Joe de m'avoir considéré membre de la famille et pour tous les bons moments partagés à l'Eden.

À Hassib, Daniel et Khaled, les amis de toujours d'avoir été là pour moi et de me rappeler qu'il n'y a pas que la thèse dans la vie et que les amis sont indispensables pour tenir le coup.

Et pour finir, je n'oublie jamais tous mes amis de Strasbourg, Charbel HAJJ, Joëlle, Marianne, Mira, Elie, Dalya, Nour et Omar. Merci pour tous les moments partagés en dehors de la thèse.

LISTE DES ABRÉVIATIONS

A

A	Adénosine
ABCE1	<i>ATP Binding Cassette Protein</i>
AND	Acide DésoxyriboNucléique
ALIX	<i>Apoptosis-Linked gene2-interacting protein X</i>
AMV	<i>Avian Myeloblastosis Virus</i>
APC	<i>Antigen Presenting Cells</i>
APS	Ammonium Persulfate
AP-2/3	Complexes Protéiques Adaptateurs clatherine-associés
ARN	Acide RiboNucléique
ARNg	ARN génomique
ARNm	ARN messenger
ATP	Adénosine TriPhosphate

B

BET	Bromure d'Éthydium
BzCN	Cyanure de Benzoyle

C

C	Cytosine
CA	Capside
CAFA	<i>Capillary-Automated Footprinting Analysis</i>
CCM	Chromatographie Couche Mince
Cdk9	<i>Cyclin-Dependent Protein Kinase 9</i>
cDNA	<i>Complementary Desoxyribonucleic Acid</i>
CD4	<i>Cluster of Differentiation</i>
CHMP	<i>Charged Multivesicular body Proteins</i>
CMCT	1-Cyclohexyl-3-(2-morpholinoethyl) carbodiimide
CRM1	<i>Chromosome Maintenance Region 1</i>
CTD	<i>C-Terminal Domain</i>
CuCAAC	<i>CuI-Catalyzed Azide-Alkyne 1,3 dipolar Cycloaddition</i>

D

DDX6	<i>ATP-dependent RNA helicase/ DEAD box protein</i>
DEPC	DiEthyl PyroCarbonate
DIS	<i>Dimerization Initiation Site</i>
DLS	<i>Dimer Linkage Structure</i>
DMS	DiMéthyle Sulfate
DMSO	DiMéthyle Sulfoxide

E

EDTA		Éthylène Diamine TétraAcétate
ENU		EthylNitrosoUrea
Env		<i>Envelop Protein</i>
EO		Propargyl Ethoxy-amine
ERANGE		<i>Enhanced Read Analysis of Gene Expression</i>
ESCRT-0/I/II/III		<i>Endosomal Sorting Complex Required for Transport</i>
	F	
FragSeq		<i>Fragmentation Sequencing</i>
FRET		<i>Fluorescence Resonance Energy Transfer</i>
	G	
G		Guanosine
Gag		<i>Group antigen specific</i>
GTP/GDP		Guanosine TriPhosphate/ Guanosine DiPhosphate
	H	
HBR		<i>Highly Basic Region</i>
HMG-I(Y)		<i>High Mobility Group Protein Isoform Y</i>
HPLC		<i>High Performance Liquid Chromatography</i>
	L	
LTR		<i>Long Terminal Repeat</i>
	M	
1M7		<i>1-Methyl-7-Nitroisatoic anhydride</i>
MA		Matrice
MuLV		<i>Moloney murine Leukemia Virus</i>
	N	
NC		Nucléocapside
Nef		<i>Negative factor</i>
NELF		<i>Negative Elongation Factor</i>
NEDD4-1/2		<i>Neural Precursor Cell Expressed, Developmentally Down-Regulated 4</i>
NES		<i>Nuclear Export Signal</i>
NF-κB		<i>Nuclear Factor-κ-light-chain-enhancer of activated B cells</i>
NGS		<i>Next Generation Sequencing</i>
NLS		<i>Nuclear Localization Signal</i>
NMIA		<i>N-Methylisatoic Anhydride</i>
	O	
ORF		<i>Open Reading Frame</i>
	P	
PA		Propargylamine

PARS	<i>Parallel Analysis of RNA Structure</i>
PARTE	<i>Parallel Analysis of RNA Structure with Temperature Elevation</i>
PBP	<i>Processing Body Proteins</i>
PBS	<i>Primer Binding Site</i>
PCNA	<i>Proliferating Cell Nuclear Antigen</i>
PCR	<i>Polymerase Chain Reaction</i>
PEG₃-azoture	<i>PolyEthylenGlycol-azoture</i>
pH	<i>Potential Hydrogen</i>
Pka	<i>Constante de dissociation acide</i>
Pol	<i>Polymérase</i>
Poly-A	<i>PolyAdénosine</i>
PNK	<i>PolyNucléotide Kinase</i>
PPi	<i>Pyrophosphate</i>
PPT	<i>PolyPurine Tract</i>
PR	<i>Protéase</i>
Psi (Ψ)	<i>Putative packaging domain</i>
P-TEFb	<i>Positive Transcription Elongation Factor</i>

R

Rev	<i>Viral Regulatory Protein</i>
RMN	<i>Résonnance Magnétique Nucléaire</i>
RNase	<i>Ribonucléase</i>
RNP	<i>RiboNucléoProtéine</i>
RRE	<i>Rev Responsive Element</i>
RT	<i>Rétro-Transcriptase</i>

S

SAFA	<i>Semi-Automated Footprinting Analysis</i>
SD	<i>Splicing Domain</i>
SD1	<i>Major Splice Donor site</i>
SDS	<i>Sodium Dodecyl Sulfate</i>
SHAPE	<i>Selective 2'-Hydroxyl acylation Analyzed by Primer Extension</i>
SIDA	<i>Syndrome d'ImmunoDéficiency Acquis</i>
SL	<i>Stem Loop</i>
SM	<i>Spectrométrie de masse</i>
SOLiD3	<i>Sequencing by Oligonucleotide Ligation and Detection</i>
SP ½	<i>Small Peptide</i>
SU	<i>Surface Subunit</i>

T

TAR	<i>Tat Responsive element</i>
Tat	<i>Transcriptional Trans-Activator</i>
TBE	<i>Tris Borate EDTA</i>
TBP	<i>TATA-Binding Protein</i>
TBTA	<i>Tris[(1-Benzyl-1H-1,2,3-Triazol-4-yl)methyl]Amine</i>
TEAA	<i>TriÉthylAmonium Acétate</i>

TEMED		N,N,N',N'-Tétraméthyléthylènediamine
TFA		Acide TriFluoroacétique
THF		TétraHydroFurane
TM		Transmembranaire
Tris		Tris(hydroxyméthyl)amino méthane
TSG101		<i>Tumor Susceptibility Gene</i>
TYMV		<i>Turnip Yellow Mosaic Virus</i>
	U	
U		Uracil
UPF1/3b		<i>UP-Frameshift 1/3b</i>
	V	
Vif		<i>Viral infectivity factor</i>
Vpr		<i>Viral protein R</i>
VPS4		<i>Vacuolar Protein Sorting-associated protein 4A</i>
Vpu		<i>Unique Viral Protein</i>
	X	
XMRV		<i>Xenotropic Murine leukemia Related-Virus</i>

TABLE DES MATIÈRES

Avant propos	3
CHAPITRE I. INTRODUCTION GÉNÉRALE.....	6
I.1 Le Virus de l'Immunodéficience Humaine de type 1 (VIH-1).....	6
I.1.1 Structure et morphologie des virions matures infectieux.....	6
I.1.2 Le cycle viral du VIH-1	7
I.1.2.1 La reconnaissance et l'entrée dans la cellule hôte.....	8
I.1.2.2 La rétro-transcription de l'ARN génomique du VIH-1	9
I.1.2.3 L'import et l'intégration de l'ADN proviral dans le génome de la cellule hôte	11
I.1.2.4 La transcription de l'ADN proviral en ARNm.....	12
I.1.2.5 L'épissage et l'export des ARN	14
I.1.2.6 La traduction	15
I.1.2.7 L'encapsidation de l'ARNg et l'incorporation des protéines de l'enveloppe.....	15
I.1.2.7.1 Encapsidation de l'ARNg.....	15
I.1.2.7.2 Incorporation des protéines dans la particule virale	19
I.1.2.8 Le bourgeonnement et la libération des particules virales	19
I.1.2.9 La maturation des particules virales.....	21
I.1.2.9.1 Le génome du VIH-1 est encapsidé sous-forme de dimère	23
I.1.3 Organisation génétique et structurale du génome	24
I.2 Les Acides RiboNucléiques « ARN ».....	27
I.2.1 Différentes familles d'ARN et leurs rôles dans les différents processus cellulaires	27
I.2.2 Structure des ARN : capacité de l'ARN à former des structures complexes.....	28
I.2.2.1 Diversité d'appariement entre les bases	29
I.2.2.2 Structures secondaires	31
I.2.2.3 Structures tertiaires.....	32
I.3 L'étude de la structure secondaire des ARN.....	34
I.3.1 Les sondes enzymatiques	36
I.3.2 Les sondes chimiques	39
I.3.3 Les techniques de détection à haut débit.....	49
I.3.3.1 Séquenceur d'ADN et électrophorèse sur capillaire (hSHAPE).....	49
I.3.3.2 SHAPE-Sequencing ou « SHAPE-Seq »	54
I.3.3.3 RNA-Sequencing ou « RNA-Seq ».....	56
I.3.3.4 Parallel Analysis of RNA Structure ou « PARS »	57
I.3.3.5 Fragmentation Sequencing « FragSeq ».....	59
I.3.3.6 PARS contre Fragseq : comparaison entre les deux méthodes	61
I.4 État des lieux	63
CHAPITRE II : SYNTHÈSE ET CARACTÉRISATION DE NOUVELLES	
SONDES FLUORESCENTES.....	64
II.1 Problématique et objectif.....	64
II.1.1 La principale limitation de la technique de hSHAPE	64
II.1.2 L'optimisation de Kevin Weeks	65
II.1.3 Objectif du travail de thèse	67
II.2 Les sondes fluorescentes	68
II.2.1 Une seule entité fluorescente	69
II.2.2 Deux entités fluorescentes	71
II.2.2.1 Amorces à transfert d'énergie.....	71
II.2.2.2 Sondes fluorescentes à transfert d'énergie.....	73
II.2.2.2.1 Bras espaceur trifonctionnel.....	73
II.3 Conception et synthèse de nouveaux BigDyes	78
II.4 MATÉRIELS ET MÉTHODES	81
II.4.1 Préparation du matériel, techniques de purification et de caractérisation	81
II.4.2 Synthèses	83

II.5 RÉSULTATS	94
II.5.1 Synthèse des dérivés de la fluorescéine modifiés	94
II.5.2 Synthèse de la Carboxy-rhodamine et de ses dérivés modifiés	95
II.5.3 Synthèse du bras espaceur : couplage de la 6-FAM-alcyne avec la 5-ROX-azoture	96
II.5.4 Analyse du BigDye par spectroscopie de fluorescence	97
II.5.5 Étude de modélisation du BigDye	98
II.5.6 Caractérisation des interactions entre les sondes du BigDye	99
II.5.6.1 Le couplage excitonique	99
II.5.7 Caractérisation des interactions entre les sondes du BigDye: transfert d'énergie non radiatif de fluorescence (FRET)	103
II.5.7.1 Efficacité de transfert	106
II.5.7.2 Recouvrement spectral	107
II.5.8 Effets de la température vis-à-vis des interactions entre les sondes fluorescentes	108
II.5.9 Stabilité de l'émission fluorescente du BigDye	109
II.6 Discussion	111
II.6.1 Choix et synthèse de la carboxy-fluorescéine	111
II.6.1.1 Dérivé de la fluorescéine : Substitution de Mannich	112
II.6.2 Choix et synthèse de la carboxy-rhodamine et ses dérivés	113
II.6.3 Choix et synthèse du bras espaceur	116
II.6.4 Couplage excitonique intramoléculaire: système fluorescéine-rhodamine	119
CHAPITRE III: SYNTHÈSE DE NOUVEAUX RÉACTIFS CHIMIQUES POUR L'ÉTUDE STRUCTURALE D'UN GÉNOME ENTIER IN VIRIO	121
III.1 Objectif	121
III.2 État des lieux	122
III.3 Les anhydrides isatoïques	123
III.3.1 Méthodes de synthèse	123
III.3.2 Réactivité chimique	125
III.3.3 Les dérivés de l'anhydride isatoïque	126
III.4 Notre approche: synthèse de nouveaux réactifs chimiques	127
III.5 MATÉRIELS ET MÉTHODES	129
III.5.1 Préparation du matériel, techniques de purification et techniques de caractérisation	129
III.5.2 Synthèses	130
III.5.3 Modification de l'ARN par l'anhydride 1-(N-propyn-2-yl)isatoïque:	136
III.5.4 Click chemistry entre l'ARN porteur de la fonction alcyne et la biotine-PEG ₃ -azoture :	137
III.5.5 Protocole général des cinétiques d'hydrolyse	138
III.6 RÉSULTATS	140
III.7 DISCUSSION	147
III.7.1 Choix des anhydrides isatoïques	147
CONCLUSIONS ET PERSPECTIVES	149

Avant propos

L'ARN joue un rôle essentiel en biologie en tant qu'intermédiaire entre le génome et le protéome. Mais en plus de cette activité, l'ARN a des rôles multiples dans d'autres processus. Ainsi, il participe à la défense cellulaire, à la catalyse biologique ou à la régulation de l'expression d'un gène. Certains ARN constituent également le génome de virus.

La fonction des ARN est fortement liée à leur capacité à former des structures complexes hautement organisées, en se repliant sur eux-mêmes. L'ARN peut se lier à des protéines virales ou cellulaires au cours de leur réplication. Il peut interagir avec de petites molécules telles que les acides aminés ou avec diverses macromolécules biologiques ou ligands tels que les protéines ou un autre ARN. Ces interactions permettent par exemple de réguler la virulence de *S.Aureus* au travers d'un réseau dont l'ARN III est le centre. L'ARN au sens large est le thème de recherche central de notre laboratoire, notre équipe s'intéresse plus particulièrement aux ARN de deux virus, le virus de l'immunodéficience humaine de type I (VIH-1), responsable du syndrome d'immunodéficience acquise (SIDA) et le virus Influenza A, responsable de la grippe.

Toutes ces fonctions de l'ARN peuvent être interrogées structuralement en utilisant différentes méthodes enzymatiques ou chimiques. Ces méthodes présentent des avantages aussi bien que des inconvénients. Dans ce cadre nous avons entamé des recherches qui contribuent à améliorer des techniques de cartographie existantes: la première partie du travail concerne l'amélioration de la technique de hSHAPE par la préparation de nouvelles sondes fluorescentes à transfert d'énergie et la deuxième partie a pour objectif le développement de nouveaux réactifs afin de pouvoir réaliser la cartographie d'un génome viral entier *in virio*.

Depuis quelques années, une nouvelle technique appelée hSHAPE (high throughput Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) permet de déterminer la réactivité de chaque nucléotide de l'ARN. La limitation de cette technique réside dans le fait qu'elle nécessite de grandes quantités d'ARN liée à l'utilisation d'un séquenceur d'ADN pour séparer et détecter des ADNc fluorescents. Afin d'améliorer cette technique et de pouvoir réaliser l'étude structurale de faibles quantités d'ARN, nous avons choisi de synthétiser un jeu de sondes fluorescentes à transfert d'énergie, différentes des sondes fluorescentes existantes dans le commerce qui, elles, sont formées d'une seule entité fluorescente.

Toujours dans l'objectif d'améliorer les techniques de cartographie des ARN, nous avons développé des réactifs qui devraient permettre de coupler l'utilisation de sondes chimiques aux techniques de séquençage à haut débit comme cela a été fait récemment pour les sondes enzymatiques. Ces nouveaux composés dérivés de l'anhydride isotoïque sont capables de réagir avec l'ARN *in vivo*, avec des temps de demi-vie variables. De plus, ils sont modifiés par une fonction alcyne afin de pouvoir réagir avec la biotine-PEG₃-azoture et ainsi permettre ultérieurement la sélection des ARN modifiés par des billes recouvertes de streptavidine. Ces réactifs chimiques pourront être utilisés pour l'étude *in vivo*, à haut débit, d'ARN de grande taille. Dans notre équipe nous envisageons par exemple d'étudier les différentes conformations adoptées par l'ARN génomique du VIH-1 durant la phase de maturation des particules virales et de vérifier s'il existe un lien entre ces changements de la structure secondaire de l'ARN et la maturation de la protéine Gag.

En effet, les résistances développées par le VIH-1 et causées par son taux de mutation et de recombinaison élevés sont un problème majeur de santé publique. Dans ce contexte et afin de développer de nouvelles armes contre ce virus, il est nécessaire de mieux comprendre toutes les étapes de son cycle de réplication. La maturation des particules virales est un processus critique du cycle viral au cours duquel un phénomène encore méconnu de changement conformationnel du génome se produit. C'est pour cette raison que notre équipe souhaite déchiffrer toutes les relations qui existent entre la structure du génome du VIH-1, son export, son assemblage, le bourgeonnement et la maturation des particules virales.

Ce manuscrit comporte trois parties : Une introduction générale sur le VIH-1, sa morphologie, son cycle réplcatif et l'organisation de son génome. Quelques notions sur les ARN, leur importance biologique et leur capacité à former des architectures complexes ainsi qu'une description des méthodes de cartographie classiques utilisées pour étudier leur structure secondaire et les améliorations récentes apportées à ces techniques.

Dans le deuxième chapitre, je présente tout d'abord la problématique liée aux sondes fluorescentes existantes et utilisées dans la technique de « hSHAPE ». Je développe ensuite notre choix de sondes à transfert d'énergie, la synthèse de notre fluorophore « BigDye », ainsi que les résultats de spectroscopie obtenus pour cette molécule.

Dans le troisième chapitre, je présente la nouvelle technique que nous souhaitons de développer afin de pouvoir étudier la structure d'ARN de grande taille *in vivo* ainsi que les réactifs chimiques qui nous permettront d'atteindre cet objectif. Après une présentation des

méthodes de synthèse existantes, je développe celle que nous avons choisie pour la préparation de nos dérivés d'anhydrides isotoïques ainsi que leur caractérisation. Pour finir je présente les résultats préliminaires de modification de l'ARN avec nos réactifs et la formation d'un complexe ARN-biotine/streptavidine qui nous permettra de sélectionner les ARN modifiés.

CHAPITRE I. INTRODUCTION GÉNÉRALE

I.1 Le Virus de l'Immunodéficience Humaine de type 1 (VIH-1)

Les virus sont des agents infectieux parasites capables d'infecter tout type de cellules vivantes. Pour classer les virus, différentes propriétés sont prises en compte: la nature de leur génome (ADN ou ARN), la polarité du génome (positive ou négative), la présence ou l'absence d'une enveloppe entourant la capsid virale, la symétrie de cette enveloppe protéique, qui varie d'une simple hélice à une structure plus complexe d'un icosaèdre. Les virus se multiplient en transmettant leur matériel génétique à la cellule hôte et en formant dans une phase extracellulaire les particules virales ou virions.

Le VIH-1 fait partie de la famille des rétrovirus et plus précisément de la sous-famille des lentivirus. Le nom « retroviridea » vient du fait que le génome (ARNg) du virus subit une étape de rétro-transcription pour conduire à l'ADN proviral double brin qui est ensuite intégré dans le génome de la cellule hôte.

I.1.1 Structure et morphologie des virions matures infectieux

Les virions matures et infectieux possèdent une forme sphérique de diamètre de 110 à 130 nm. Ils sont composés d'une membrane phospholipidique qui renferme un « core » en forme de cône qui contient l'ARN génomique viral, des ARN d'origine cellulaire ainsi que des protéines cellulaires et virales (Fig.1). Les virions contiennent :

- Deux copies de l'ARN génomique monocaténaire (+) linéaire, non segmentées qui possèdent une coiffe à l'extrémité 5' et une queue polyA à l'extrémité 3'. La région codante de leur génome est entourée par deux longues séquences répétitives (LTR) qui codent pour la synthèse des principales protéines virales. L'ARNg se condense dans un complexe ARN-protéine durant la phase de bourgeonnement, qui sera enfermé dans une enveloppe lipidique provenant de la membrane plasmique de la cellule hôte infectée. Il se présente sous forme d'un dimère instable dans les particules immatures et sera converti durant la phase de maturation en un dimère plus compact et plus stable.
- Une matrice (MA) qui est obtenue après le clivage protéolytique du précurseur Gag. Elle est myristylée à son extrémité N-terminale.

- Une capside (CA) qui est obtenue au cours de la maturation des particules virales. Elle forme des oligomères pour donner le « core ». Dans le « core » environ 1500 molécules de nucléocapside (NC)¹ enrobent l'ARN génomique en formant des oligomères^{2,3}. On y retrouve également toutes les protéines enzymatiques nécessaires à la réplication ainsi que des protéines d'origine cellulaire et l'ARNt₃^{Lys}.
- Des protéines glycosylées à la surface, incorporées dans la membrane lipidique et qui permettent la reconnaissance avec les récepteurs cellulaires de la cellule hôte pour l'entrée du virus et l'infection : SUgp120 et TMgp41. Ces deux protéines forment des trimères associés de façon non-covalente à la surface du virion.
- Des protéines enzymatiques obtenues principalement après maturation du précurseur Gag-Pol : l'intégrase, la rétro-transcriptase et la protéase.
- Des protéines auxiliaires Nef, Vif et Vpr, contrairement aux autres protéines auxiliaires Vpu et les protéines régulatrices Tat et Rev qui ne sont pas encapsidées dans les virions matures.

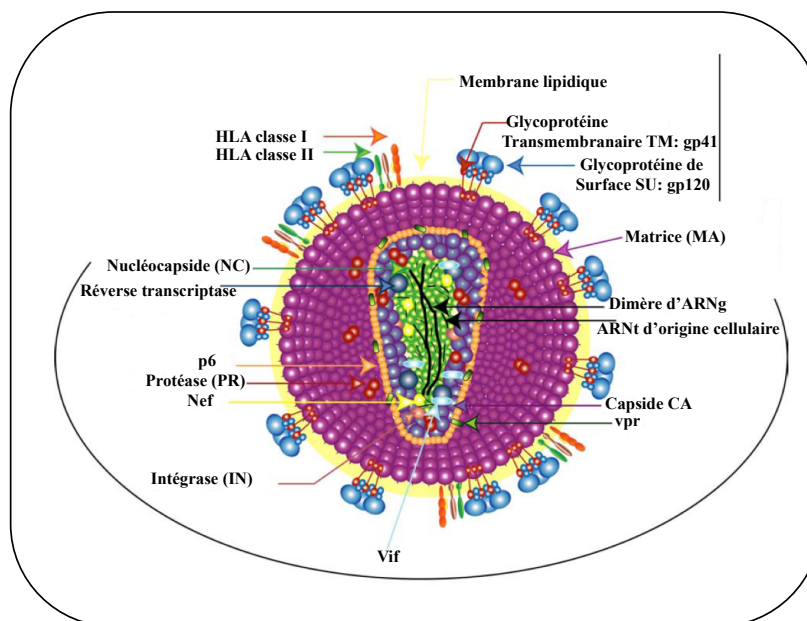


Figure 1: Représentation schématique des différents constituants de la particule virale mature du VIH-1. (Adaptée⁴)

I.1.2 Le cycle viral du VIH-1

La réplication de l'ARN viral du VIH-1 est une succession d'événements qui se caractérisent par deux phases :

La phase précoce ou phase pré-intégrative débute par la reconnaissance de la cellule hôte, la fusion des membranes cellulaires et virales, la décapsulation puis la rétro-transcription du génome viral et enfin l'import et l'intégration dans les chromosomes des cellules hôtes. La phase tardive se caractérise par la transcription de l'ADN en ARNm et la synthèse des différentes protéines structurales, régulatrices, enzymatiques et accessoires. Ensuite, l'ARNg et les protéines s'assemblent pour former la particule virale immature, qui va bourgeonner au niveau de la membrane plasmique cellulaire et la dernière étape consiste en la maturation des virions⁴. (Fig.2)

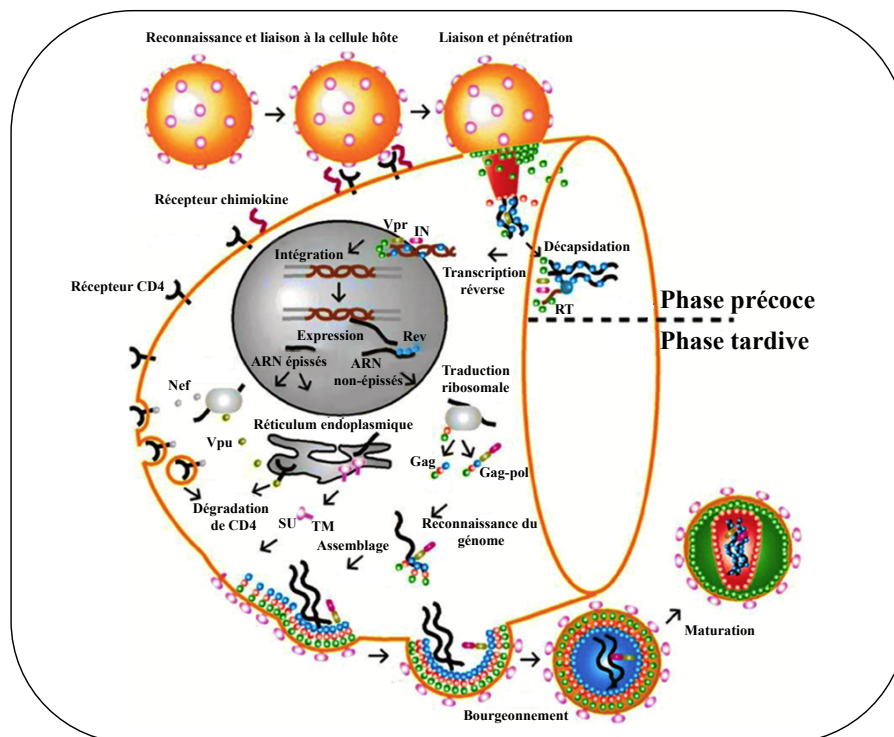


Figure 2: Représentation schématique des deux phases du cycle viral du VIH-1: la phase précoce et la phase tardive. (Adaptée⁴)

I.1.2.1 La reconnaissance et l'entrée dans la cellule hôte

Le VIH-1 entre dans la cellule hôte par fusion de sa membrane avec celle de la cellule hôte. L'enveloppe glycoprotéique du virus consiste en un complexe de deux protéines gp120 et gp41 générées par clivage protéolytique du précurseur gp160. La fusion est réalisée par l'intermédiaire de la protéine gp120 de l'enveloppe virale glycoprotéique qui se lie avec une grande affinité au domaine immunoglobuline amino-terminal du récepteur membranaire CD4⁵.

Ces récepteurs sont présents principalement à la surface des lymphocytes T et sur les cellules présentatrices d'antigènes (APC: Antigen Presenting Cells) telles que les cellules de Langerhans (cellules dendritiques de la peau et de la muqueuse) et les cellules microgliales présentes dans le système nerveux central. Ces interactions sont suffisantes pour la fusion avec la cellule mais pas pour l'infection.

Pour l'infection, le VIH-1 a besoin de l'un des deux co-récepteurs chimiokines CXCR4 ou CCR5 qui sont des récepteurs transmembranaires à sept domaines couplés à la protéine G^{6,7,8}. Des changements conformationnels au niveau de la protéine gp120 auront lieu, suite à la reconnaissance cellulaire, et qui mènent à l'insertion du domaine hydrophobe amino-terminal de la protéine gp41 dans la membrane de la cellule cible, puis à la fusion des membranes cellulaire et virale avec libération de la capside virale dans le cytoplasme de la cellule infectée⁹.

I.1.2.2 La rétro-transcription de l'ARN génomique du VIH-1

La rétro-transcription est un processus catalysé par l'enzyme rétro-transcriptase qui se déroule principalement dans le cytoplasme^{10,11,12}. Le produit de la rétro-transcription est un ADN linéaire double brin obtenu en plusieurs étapes (Fig.3) :

- La synthèse du brin d'ADN (-) en direction de l'extrémité 5' de l'ARN débute par l'hybridation de l'extrémité 3' de l'ARNt₃^{Lys} à sa séquence complémentaire PBS (Primer Binding Site) de l'ARN génomique.
- Une fois arrivé à l'extrémité 5' de la matrice d'ARN, le fragment d'ARN hybridé au nouveau fragment d'ADN est dégradé grâce à l'activité RNase H de la rétro-transcriptase. Ce processus est suivi par le premier transfert du brin d'ADN vers l'extrémité 3' de la matrice d'ARN. Ce transfert est facilité par la protéine virale NCp7. La synthèse du brin (-) peut alors se poursuivre.
- La rétro-transcriptase clive la majeure partie de l'ARN hybridé à l'ADN (-) synthétisé et génère deux fragments riches en purines PPT (PolyPurine Tract), l'un situé au centre du génome et l'autre situé en amont de la région U3 et servant d'amorces pour la synthèse du brin (+).
- Les fragments d'ARN restants sont ensuite éliminés et un second transfert s'effectue permettant l'hybridation de l'extrémité 3' du brin (+) avec l'extrémité du brin (-). Ces deux brins vont servir de matrice pour achever leur synthèse¹³.

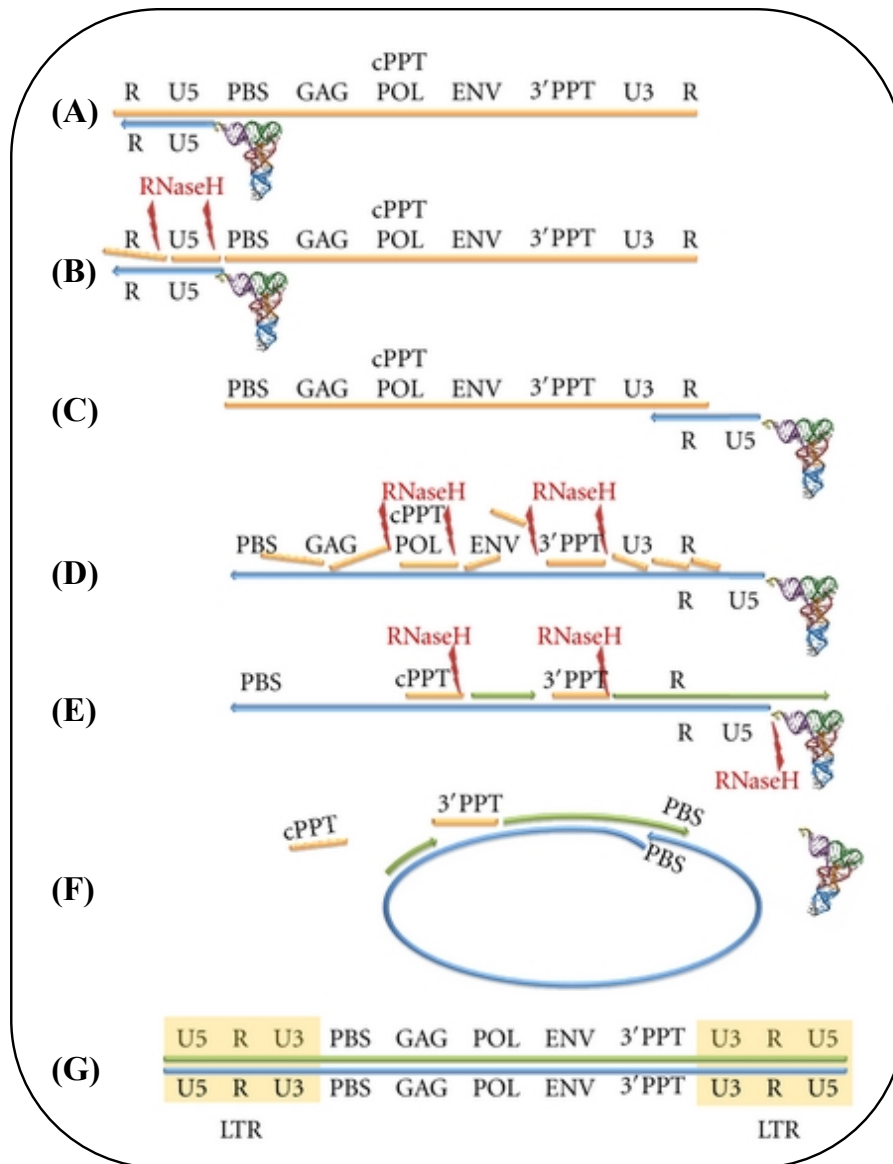


Figure 3: Représentation schématique de la rétro-transcription de l'ARN viral en ADN proviral double brin. (A) L'ARN génomique du VIH (représenté en orange) possède une séquence de bases complémentaires à l'extrémité 5' permettant l'hybridation de l'ARNt_{3^{Lys}}. (B) La RT a commencé la synthèse du brin d'ADN (-) (représenté en bleu), et la RNase H a dégradé la séquence d'ARN hybridée au nouveau fragment d'ADN. (C) Le premier transfert du brin d'ADN (-) vers l'extrémité 3' du ssARN se produit permettant ainsi de continuer la synthèse du brin d'ADN (-). (D) Dégradation de l'ARN : Deux séquences riches en purine ppt, l'une centrale et l'autre en amont de la région U3 sont résistantes à l'activité RNase H, et vont servir d'amorce pour la synthèse du brin d'ADN (+). (E) La synthèse du brin d'ADN (+) (représenté en vert) continue afin de copier une séquence de 18 nucléotides de l'ARNt qui permet ainsi à la RNase H de dégrader et d'enlever l'amorce ARNt. (F) L'élongation de la synthèse des deux brins d'ADN (-) et (+). (G) La synthèse de l'ADN proviral double brin avec une séquence répétitive LTR sur chaque extrémité. (Adaptée¹³)

I.1.2.3 L'import et l'intégration de l'ADN proviral dans le génome de la cellule hôte

L'ADN pro-viral obtenu par la rétro-transcription est transporté dans le noyau de la cellule hôte en présence du complexe d'intégration comprenant l'intégrase, la rétro-transcriptase, les protéines de matrice et de nucléocapside, Vpr et aussi des protéines cellulaires¹⁴. La localisation nucléaire du complexe de pré-intégration est dirigée par la protéine accessoire Vpr, qui ne contient pas un signal de localisation nucléaire, mais semble fonctionner en connectant le complexe de pré-intégration à la machinerie nucléaire d'import de la cellule avec l'aide de l'importin- α et des nucléoporines^{15,16}.

La protéine de matrice contient un signal de localisation nucléaire qui permet au complexe d'intégration de cibler le noyau de la cellule hôte. L'intégration de l'ADN proviral se fait dans les chromosomes de la cellule hôte dans des sites aléatoires¹⁷, mais elle cible en préférence les régions euchromatine non-repliées et accessibles qui se trouvent proches des gènes transcriptionnellement actifs.

Le processus d'intégration est catalysé par l'intégrase qui débute par un clivage de l'extrémité 3' de l'ADN viral dans le cytoplasme, juste après sa synthèse^{18,19}. Elle clive les deux nucléotides terminaux (GT) hautement conservés des extrémités 3' de chaque brin d'ADN et libère une extrémité formée par le dinucléotide CA possédant un groupement hydroxyle libre en position 3'. Les groupements 3'-OH résultants des deux extrémités CA offrent des sites d'attachement de ce provirus au génome de la cellule hôte.

L'intégrase clive l'ADN de la cellule hôte en laissant des extrémités 5' phosphate libres. Lorsque le complexe ADN proviral-intégrase se lie à l'ADN de la cellule hôte, l'intégrase catalyse une trans-estérification directe dans laquelle les groupements 3'-OH terminaux sont ligués aux extrémités 5' phosphate de l'ADN génomique de la cellule hôte.

La dissociation du complexe d'intégration libère un intermédiaire d'intégration qui possède des lacunes dans la jonction entre le génome de la cellule hôte et celui de l'ADN proviral. Ces lacunes sont comblées par la machinerie de réparation cellulaire (polymérase bêta, polymérase delta, et le cofacteur PCNA)^{20,21,22}. (Fig.4)

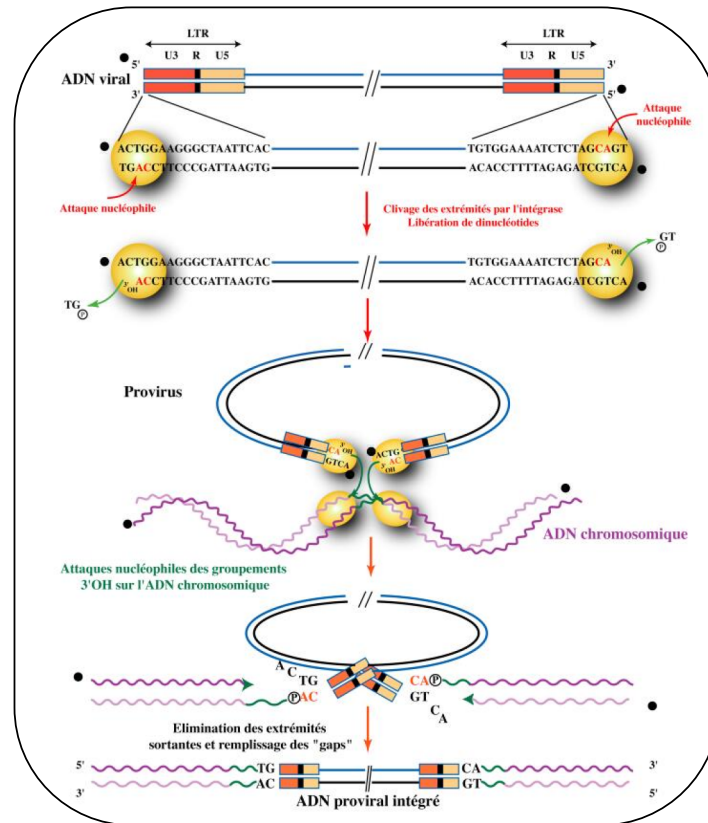


Figure 4: Le processus d'intégration de l'ADN proviral dans le génome de la cellule hôte. Les extrémités 5' sont représentées par des sphères noires. Les deux extrémités 3'OH libres sur les deux nucléotides CA sont obtenues par attaque nucléophile d'une molécule H_2O .

Il existe deux types d'ADN viraux circulaires en addition de l'ADN linéaire dans les cellules infectées : un ADN circulaire possédant une seule séquence LTR, générée par la recombinaison entre les deux séquences LTR de l'ADN linéaire. Le deuxième type d'ADN circulaire possède deux séquences LTR et est généré par ligation des deux extrémités du précurseur linéaire. Dans cet ADN, les séquences LTR sont distribuées d'une façon aléatoire suite à une auto-intégration de l'ADN viral sur lui-même par l'intermédiaire de l'intégrase virale²³. Ces ADN circulaires ne sont pas infectieux mais sont transcriptionnellement actifs²⁴.

I.1.2.4 La transcription de l'ADN proviral en ARNm

L'hétérochromatine joue un rôle très important dans l'expression de l'ADN proviral. En effet, c'est un répresseur transcriptionnel qui permet à l'ADN de rester sous une forme latente et d'échapper aux traitements antirétroviraux. L'ADN intégré possède des promoteurs de transcription qui favorisent le positionnement de l'ARN polymérase II cellulaire sur le site d'initiation de la transcription. La faible efficacité d'élongation de l'ARN polymérase II

hypophosphorylée dans son domaine C-terminal (CTD) et associée à des facteurs négatifs d'élongation de la transcription (NELF et DSIF) entraîne la synthèse de transcrits courts, non polyadénylés. L'activation de la cellule (par la présentation d'un antigène) permet la relocalisation du facteur nucléaire NF- κ B qui se fixe sur le domaine activateur présent dans la région 5'LTR de l'ADN proviral. Le résultat est un remodelage de la chromatine, par l'activation d'une histone acétylase ainsi qu'une augmentation du taux d'initiation et d'élongation de la transcription²⁵. L'ARN viral est traduit pour synthétiser les protéines régulatrices Tat et Rev. La protéine Tat est le facteur positif de l'élongation de la transcription. Elle possède trois domaines fonctionnels : le domaine N-terminal pour l'activation et la fixation des co-facteurs, le domaine basique de fixation à l'ARN et le transport nucléaire, et le domaine C-terminal impliqué dans la stimulation du « capping (phénomène d'ajout de la coiffe sur l'extrémité 5' de l'ARNm) » des ARN viraux²⁶.

La protéine Tat reconnaît le renflement existant dans la tige-boucle TAR en 5' de l'ARN viral et recrute P-TEFb (contenant la cycline T1 et la kinase Cdk9) qui est responsable de l'hyperphosphorylation du domaine CTD (Domaine C-terminal) de la polymérase II par la kinase Cdk9 afin de stimuler l'élongation de la transcription. L'acétylation de Tat et son interaction avec une histone acétylase sont nécessaires pour dissocier le complexe Tat/P-TEFb/TAR et stimuler l'élongation de la transcription de l'ADN proviral en ARNm. (Fig.5)

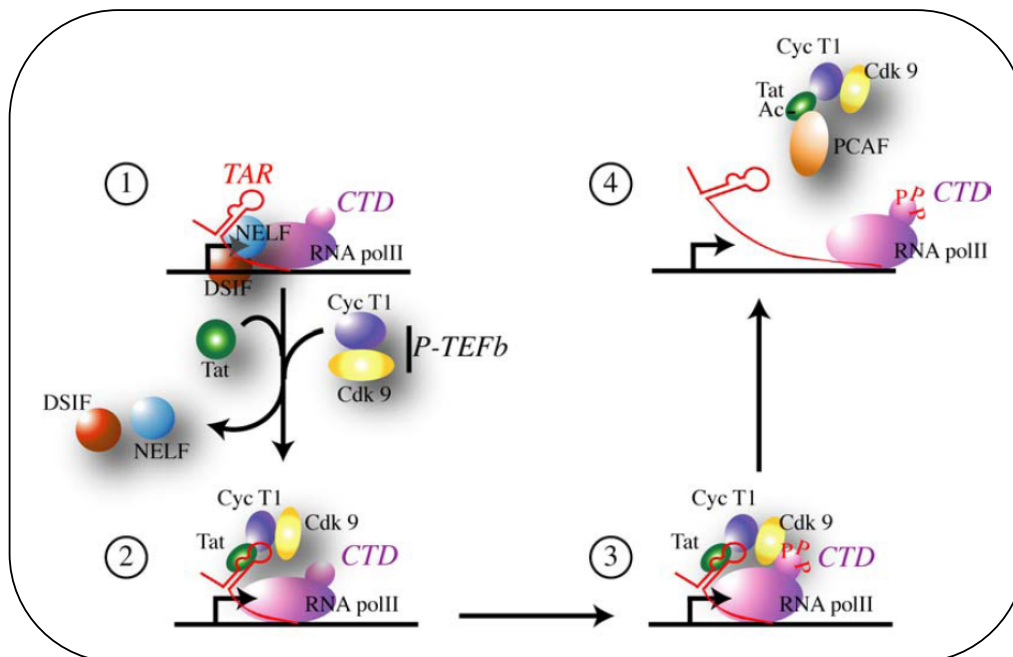


Figure 5: Modèle d'activation de la protéine Tat. (Adaptée²⁶)

I.1.2.5 L'épissage et l'export des ARN

Le VIH-1 a la capacité de générer un très grand nombre d'ARNm viraux de taille variable à partir d'un seul transcrit, suite à un phénomène d'épissage alternatif. La présence de plusieurs sites donneurs et accepteurs d'épissage permet l'obtention de trois classes d'ARN²⁷ :

- Les ARN mono-épissés (4-5kb) qui codent pour les protéines Env, Vif, Vpr et Vpu.
- Les ARN multi-épissés (2kb) qui codent pour les protéines Tat, Rev et Nef.
- Les ARN non-épissés (9kb) qui codent pour les protéines Gag et Gag-Pol.

L'export des ARN mono et non-épissés du noyau vers le cytoplasme est dépendant de la protéine Rev. La protéine Rev possède un domaine NLS (Nuclear Localization Signal) ainsi qu'un domaine NES (Nuclear Export Signal). Grâce à son domaine NLS, la protéine Rev se lie à l'importine β pour former un hétérodimère, qui est transporté via le pore nucléaire. Elle contient aussi un domaine riche en arginines qui lie spécifiquement une structure en tiges-boucles nommée RRE (Rev Responsive element)²⁸ présente dans les ARN du VIH-1 non-épissés et mono-épissés. RRE est une région structurée et localisée dans une séquence appelée ORF (Open Reading Frame). Elle permet l'interaction avec les facteurs d'export nucléaire (CRM1 : Chromosome Maintenance gene1) qui interagit avec des constituants des pores nucléaires. Dans ce processus, RanGTP/GDP fournit l'énergie nécessaire à l'export²⁹. La protéine Rev contient un second domaine proche de l'extrémité C-terminale riche en résidus leucines et nécessaire à l'activité de la protéine Rev, mais qui n'entre pas dans la formation du complexe Rev/RRE. Ce domaine se lie à une protéine de la famille de nucléoporine appelée Rab (Rev activation domain-activation protein)³⁰. Un second site de liaison de la protéine Rev sur l'ARN, distinct de la séquence RRE, a été identifié dans l'extrémité 5'UTR. La protéine Rev se lie à une boucle interne de la tige boucle SL1. Il a été démontré que cette boucle (nommée boucle A) est conservée dans des isolats de VIH-1 et que des mutations de cette boucle réduisent la liaison de la protéine Rev au SL1³¹.

Les ARN multi-épissés utilisent une voie indépendante de la protéine Rev pour leur export du noyau vers le cytoplasme³².

I.1.2.6 La traduction

Dans le cytoplasme, la machinerie cellulaire est détournée par le virus pour permettre la synthèse des protéines auxiliaires, structurales et enzymatiques. La synthèse des précurseurs protéiques Gag et Gag-Pol se fait à partir d'un même transcrit (ARNm non-épissé), ainsi que la synthèse des protéines d'enveloppe et Vpu (ARN mono-épissé). Cependant les mécanismes de régulation de la traduction utilisés sont différents dans les deux cas :

- La synthèse du précurseur protéique Pr160^{Gag-Pol} est obtenue grâce à un décalage du cadre de lecture des ribosomes qui rend silencieux le premier codon stop à la fin de la traduction du précurseur Pr55^{Gag}. Il en résulte la production d'une grande quantité de protéines Gag par rapport à Gag-pol, qui vont générer les protéines structurales de la protéine virale³³. La myristilation N-terminale du précurseur Pr55^{Gag} est nécessaire pour permettre la fixation de ces protéines à la face interne de la membrane plasmique.
- La synthèse des protéines Env et Vpu est obtenue à partir d'un ARNm bi-cistronique et dépendante de la protéine Rev³⁴. Il contient deux ORF, Vpu et Env, qui se suivent et se chevauchent. La protéine Vpu favorise la dégradation, par le protéasome, des CD4 piégés dans un complexe intracellulaire avec gp160 au niveau du réticulum endoplasmique par une déstabilisation intracellulaire des complexes CD4-gp160³⁵. Cette dégradation permet aux protéines gp160 de rejoindre la membrane plasmique. En plus, Vpu favorise le bourgeonnement des virions.

I.1.2.7 L'encapsidation de l'ARNg et l'incorporation des protéines de l'enveloppe

I.1.2.7.1 Encapsidation de l'ARNg

La voie générale de l'assemblage des particules virales est réalisée par l'intermédiaire du précurseur Pr55^{Gag} à travers un réseau d'interactions protéine-ARN, protéine-protéine et protéine-lipide puisqu'elle comprend des domaines requis pour l'attachement à la membrane, la multimérisation, la liaison à des acides nucléiques et l'interaction avec les composants de la machinerie de bourgeonnement de la cellule hôte.

Les ARN non-épissés sont traduits par les ribosomes pour produire les protéines Gag et Gag-Pol. Lorsque la synthèse de la protéine Gag (55 KDa) est achevée, il semblerait que des monomères de Gag se rassemblent pour accompagner l'ARN génomique au site de bourgeonnement sur la membrane plasmique et favorisent son interaction avec les signaux d'encapsidation^{36,37}. Ce processus est différent de l'interaction avec le domaine NC de la protéine Gag, ce dernier facilitant la formation du dimère d'ARN génomique et la rétro-transcription. Bien que le transport de l'ARN soit dépendant de la production de la protéine Gag, la traduction de l'ARN génomique n'est pas une condition préalable à son transport, comme c'est le cas pour le VIH-2^{38,39,40}.

La production de particules sphériques est optimisée par les acides nucléiques⁴¹. Actuellement il y a un manque relatif de connaissances sur les composants cellulaires qui sont responsables du transport du complexe Gag/ARN. Le transport de la protéine Gag dans le cytoplasme engage les microtubules et les microfilaments d'actine du cytosquelette de la cellule hôte et il est facilité par d'autres facteurs cellulaires impliquant les complexes protéiques adaptateurs clathrine-associés (AP-2 et AP-3). Les facteurs UPF1⁴², UPF3b et Staufen sont responsables de la stabilisation de l'ARN génomique. La protéine cellulaire ABCE1⁴³ (une protéine de la famille ATP-Binding Cassette) en association avec la DDX6 et d'autres protéines PBP (Processing Body Proteins) sont impliquées aussi dans l'assemblage des particules virales d'une manière indépendante de l'ARN et facilitent la multimérisation de la protéine Gag⁴⁴. La protéine KIF-4 se lie aussi à l'ARN, mais son rôle dans le recrutement du complexe Gag/ARN à la membrane plasmique n'est pas clair⁴⁵. L'implication séquentielle de toutes ces protéines dans le transport du complexe Gag/ARN est toujours obscure⁴⁶. (Fig.6)

La protéine Gag interagit avec les phospholipides membranaires en impliquant l'extrémité N-terminale (acide myristique) du domaine MA et une région basique conservée HBR « Highly Basic Region ». Cette deuxième interaction réduit les liaisons non-spécifiques et favorise la liaison du domaine MA à des lipides membranaires contenant des résidus phosphatidylinositol diphosphate (PIP2).

Les signaux majeurs de multimérisation se trouvent en aval de l'extrémité C-terminale du domaine CA, le domaine voisin sp1 et le domaine I de la NC. La multimérisation de la protéine Gag est renforcée après la liaison du domaine NC à l'ARN génomique. Le domaine NC forme une sorte de pont entre les monomères de Gag⁴⁷.

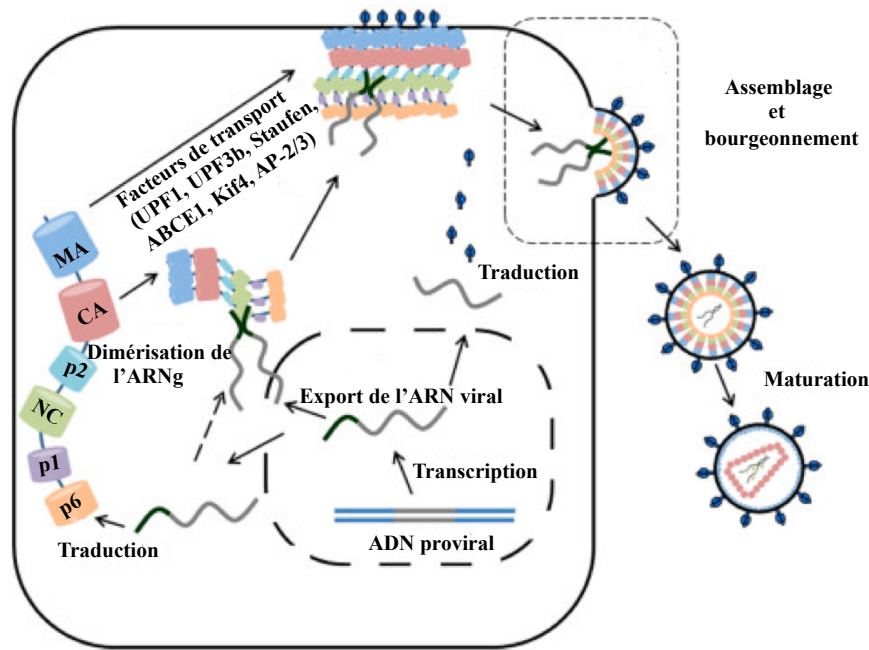


Figure 6: Représentation schématique des stades tardifs du cycle viral. Après transcription, l'ARN viral est exporté du noyau vers le cytoplasme pour la synthèse de la protéine Gag (et Gag-pol). Les ARN épissés sont aussi exportés pour produire l'enveloppe glycoprotéique (bleu foncé). Gag se lie au dimère d'ARN avec son domaine NC pour former un complexe ribonucléoprotéique qui est transporté vers la membrane plasmique où se rassemblent encore plus de monomères de Gag pour que le phénomène de bourgeonnement se produise. (Adaptée⁴⁶)

Les monomères de Gag se rassemblent par l'intermédiaire de l'ARN génomique, et en présence de plusieurs protéines l'une à côté de l'autre (Fig.7). Un changement conformationnel se produit au niveau de la protéine qui permet l'exposition de sa surface à des interactions avec d'autres protéines Gag et ainsi à l'assemblage des particules virales⁴⁸.

Les mécanismes de transport de l'ARN génomique du noyau jusqu'au point d'assemblage et de bourgeonnement ne sont pas bien connus. Deux mécanismes de transport cytoplasmique de l'ARNm ont été décrits : soit par diffusion suivie par piégeage local, soit par transport actif le long du cytosquelette. L'ARN viral est transporté via des voies endosomales en même temps que la protéine Gag et Env sur des vésicules endosomales Rab-11 pour arriver au niveau de la membrane plasmique de la cellule hôte.

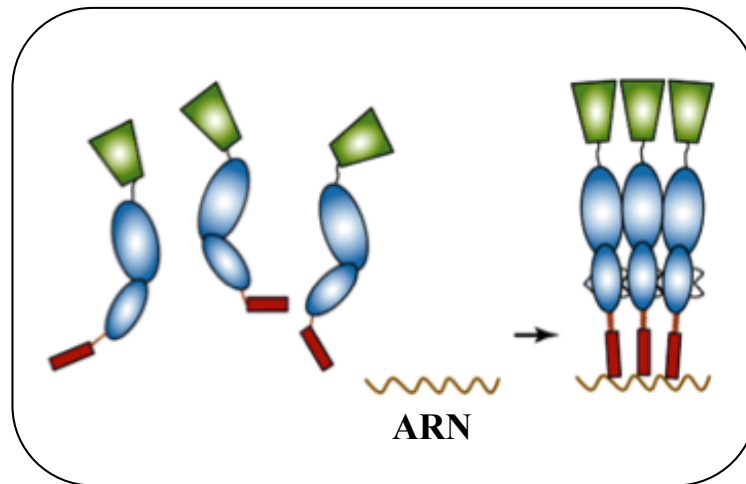


Figure 7: Illustration montrant le phénomène d'assemblage des différents monomères de la protéine Gag et la formation du complexe Gag/ARN nécessaire pour la maturation des virions. (Adaptée⁴⁸)

L'encapsulation de l'ARNg requiert des interactions entre le domaine nucléocapside (NC) du précurseur Gag et une région de l'ARN viral appelée Psi (ψ) localisée dans la région 5'UTR de l'ARN génomique ainsi que le début du gène gag. Une région correspondant aux premiers 1015 nucléotides de l'ARN génomique favorise l'encapsulation des ARN hétérologues, à condition qu'ils soient exportés du noyau indiquant ainsi que tous les facteurs d'encapsulation sont localisés dans la région 5'UTR et ou dans gag. Cependant la région Psi n'est toujours pas très bien définie. Plusieurs études ont identifié la région SL3 comme étant le signal majeur de l'encapsulation du VIH-1. On sait maintenant que TAR, la tige-boucle poly-A, le domaine PBS et SL1 contribuent aussi à l'encapsulation. Ces résultats étaient surprenants du fait que ces domaines sont localisés en amont du site d'épissage (SD) et ainsi présents dans l'ARN génomique mais aussi dans les ARN épissés.

Des travaux récents dans notre laboratoire en collaboration avec le Dr. Marylène Mougel (CPBS, CNRS, Montpellier) ont montrés que le domaine SL1 est le signal majeur responsable de l'encapsulation et non pas le domaine SL3. Bien que le domaine SL1 est présent dans les ARN viraux épissés et non-épissés, il favorise l'encapsulation des ARN non-épissés seulement. Par conséquent, SL1 requiert des séquences localisées en 3' du domaine SD pour exercer sa fonction. Deux interactions tertiaires impliquant des régions en 3' du domaine SD permettraient d'expliquer le phénomène d'encapsulation sélectif de l'ARN génomique. La première interaction se produit entre la boucle SL4 et une région dans U5 formant ainsi l'interaction U5/AUG. Dans l'ARN du VIH-1, la substitution de l'heptanucléotide

GAGAUGG par CUCUAGA cause une diminution de l'encapsidation. Une hypothèse propose qu'un changement conformationnel dans la région 5'UTR est à l'origine de la régulation de la traduction et de l'encapsidation de l'ARN génomique. La deuxième interaction se produit entre la boucle interne de SL1 et la boucle apicale de SL4. Ces deux interactions sont possibles seulement dans l'ARN génomique et pourraient contribuer à la liaison Gag-ARN.

I.1.2.7.2 Incorporation des protéines dans la particule virale

Outre l'ARN viral, d'autres constituants cellulaires et viraux sont indispensables pour l'assemblage des particules virales et nécessaires pour l'infectivité. On peut citer l'ARN_{t3}^{Lys} cellulaire (50-100 molécules/virion)⁴⁹, d'autres protéines virales telles que Vpr (100-200 molécules/virion), Vif (10-100 molécules/virion), et Nef (5-70 molécules/virion). Des protéines cellulaires sont aussi incorporées dans la particule virale, telles que Cyp-A (250 molécules/virion), Hsp70 présent avec un rapport de 1:25 par rapport aux protéines Gag, la LysRS, Tsg101, l'Ubiquitine, APOBEC3G, INI1/HSNF5, Staufen, UNG2, HLA-II, ICAM-1 (CD54), des annexines et des tétraspanines^{50,51}. Les virions immatures sont constitués à ~50% de la protéine Gag par rapport à la masse entière du virion, tandis que la membrane lipidique constitue ~30% de la masse totale⁵². Les autres protéines cellulaires et virales constituent ~20% de la masse du virion, tandis que l'ARNg avec les autres petits ARN constituent ~2,5% de la masse du virion⁵³.

I.1.2.8 Le bourgeonnement et la libération des particules virales

Après l'encapsidation et l'assemblage de tous les composants viraux, les particules virales sont libérées de la cellule. Ce phénomène consiste en une fission membranaire.

Des délétions et des mutations du domaine p6 (situé en C-terminal du précurseur Gag) et plus spécifiquement dans les domaines conservés PTAP et YPXnL résultent en un phénomène de bourgeonnement incomplet.

Le système de protéines ESCRT (ESCRT-0, I, II, et III) est impliqué dans le processus de bourgeonnement des particules virales, ces protéines sont recrutées de façon ordonnée au niveau des membranes endosomales⁵⁴ (Fig.8). ESCRT-I est recruté vers le site de bourgeonnement via des interactions entre le domaine PTAP dans p6 et la protéine TSG101

faisant partie du complexe ESCRT-I. Ensuite l'interaction entre ESCRT-I et les protéines CHMPs du complexe ESCRT-III est un mécanisme méconnu jusqu'à maintenant, mais faisant apparemment intervenir une interaction avec ESCRT-II. En plus, la protéine ALIX (apoptosis-linked gene 2-interacting protein X) se lie au motif YPXnL qui se trouve dans les deux domaines p6 et NC de la protéine Gag respectivement via les domaines V et Bro1. Le rôle du motif YPXnL est de favoriser le bourgeonnement en se liant à la protéine ALIX. L'implication de l'ARN dans la liaison NC-ALIX a été suggérée, mais l'identité de l'ARN n'est pas claire. La famille des ubiquitines ligases (NEDD4) joue également un rôle dans le bourgeonnement, soit dans l'ubiquitination et l'activation du complexe ESCRT-I par l'intermédiaire de la ligase NEDD4-2⁵⁵, ou dans l'interaction avec la protéine ALIX pour libérer les particules virales par l'intermédiaire de la ligase NEDD4-1⁵⁶. Pour que le phénomène de fission membranaire ait lieu, le complexe ESCRT-III polymérise pour former une structure en forme de dôme. Des monomères de la protéine VPS4 (Vacuolar Protein Sorting-associated Protein 4A) régulés par un binôme de complexes protéiques LIP5-CHMP5 et CMHP1-ISTI, sont recrutés au site de bourgeonnement pour désassembler les filaments de ESCRT-III afin d'être recyclés. Le processus de fission de la membrane cellulaire reste aussi méconnu⁴⁶.

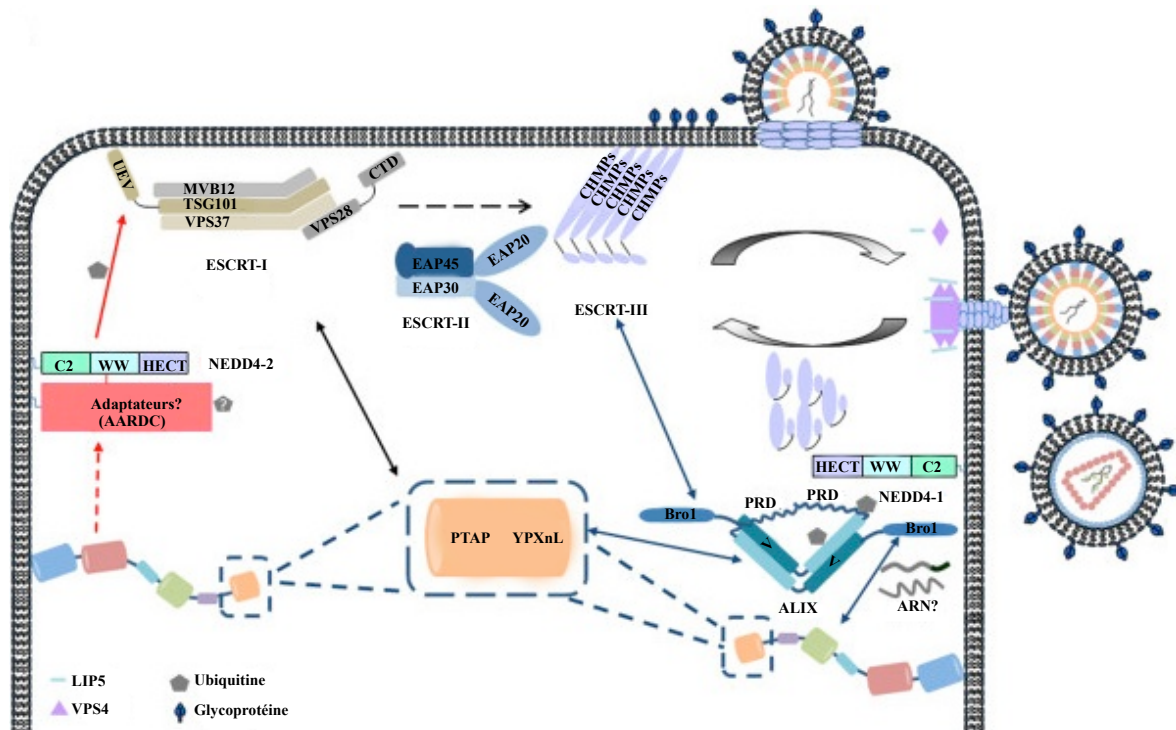


Figure 8: Le phénomène de bourgeonnement des particules virales et l'intervention des différentes protéines cellulaires permettant de libérer les virions immatures de la membrane plasmique. (Adaptée⁴⁶)

I.1.2.9 La maturation des particules virales

Le virus VIH-1 bourgeonne de la cellule hôte en tant que particule immature non-infectieuse. Le processus de maturation des particules virales est en relation avec des clivages au niveau de la protéine Gag. La protéine Gag possède différents domaines fonctionnels avec une extrémité N-terminale (acide myristique ou acide tétradécanoïque) dans le domaine MA et une extrémité C-terminale dans le domaine p6⁵⁷. La formation des particules immatures nécessite des interactions faibles mais coopératives entre les protéines Gag. Ces interactions impliquent principalement l'extrémité C-terminale du domaine CA et le peptide adjacent SP1 de la protéine Gag^{58,59,60,61}. Les domaines MA et p6 semblent être sans aucun rôle au moins dans l'assemblage *in vitro*, par contre le domaine MA est nécessaire pour le ciblage et l'alignement dans les cellules infectées. Le domaine NC est responsable de la liaison de l'ARN avec le génome de la cellule hôte^{62,63,64,65,66}. La maturation des particules virales commence durant ou juste après le bourgeonnement des particules virales par le clivage des précurseurs Gag Pr55^{Gag} et Pr160^{Gag-Pol} en leurs domaines protéiques respectifs par l'intermédiaire de la protéase.

De nombreuses études ont montré qu'un clivage protéolytique correct des protéines Gag et Gag-pol est nécessaire pour générer un dimère d'ARN mature et stable^{67,68,69,70}. Le processus de clivage de la protéine Gag est bien ordonné dans cinq séquences précises⁷¹. Des études *in vitro* ont montré que le clivage de Gag est réalisé par la protéase virale et successivement dans les régions NC/SP1, SP2/p6, MA/CA, NC/SP2, et CA/SP1. Le premier clivage divise la protéine Gag en deux domaines : le domaine p41 contenant MA, CA et SP1 et le domaine p15 (contenant NC, SP2 et p6). Ce premier clivage est associé avec la condensation initiale du génome viral⁷². La protéine chaperonne NC est très importante pour la formation d'un dimère mature, et apparemment le premier clivage SP1/NC est nécessaire pour libérer le domaine p15 dans le virion qui initie la maturation du dimère^{73,72,68,74}. Le second clivage divise p41 en MA et p25 (p25 contient CA et SP1) et divise p15 en p6 et p9 (p9 contient NC et SP2). Le produit intermédiaire obtenu p9 n'a pas d'effet sur la maturation de l'ARN⁷⁵ mais la libération de la MA par un clivage MA/CA et un clivage entre SP2 et p6 séparent des domaines fonctionnels et permettent la condensation du complexe ribonucléoprotéique (RNP). Un clivage ultérieur de CA/SP1 permet la formation des particules matures⁷². Le domaine CA forme une enveloppe en forme de cône autour du complexe ribonucléoprotéique (RNP) condensé. Le

dernier clivage du domaine p9 en NC et SP2 est nécessaire pour une maturation complète du dimère d'ARN⁷⁰.

Curieusement, et immédiatement après le premier clivage de la protéine Gag, les virus contiennent une activité rétro-transcriptase, mais font preuve d'un déficit d'infectivité qui n'est pas en relation avec la fusion cellule-virus, mais corrélée avec un échec d'accumulation des produits finaux de la rétro-transcription^{76,75}. De plus, les changements morphologiques du virion ne correspondent pas nécessairement à une maturation de l'ARN génomique. Des mutations au niveau du site de clivage NC/SP2 ont montré une morphologie sauvage, malgré des défauts au niveau de la formation du dimère d'ARN, et ces virions sont ainsi non infectieux. Au contraire le blocage du site de clivage SP1/CA conduit à une altération au niveau de la morphologie du virion mais n'inhibe pas la réplication virale. Il est très important de noter que la structure native de l'ARN du VIH-1 est nécessaire pour le premier transfert de brin durant la rétro-transcription. (Fig.9)

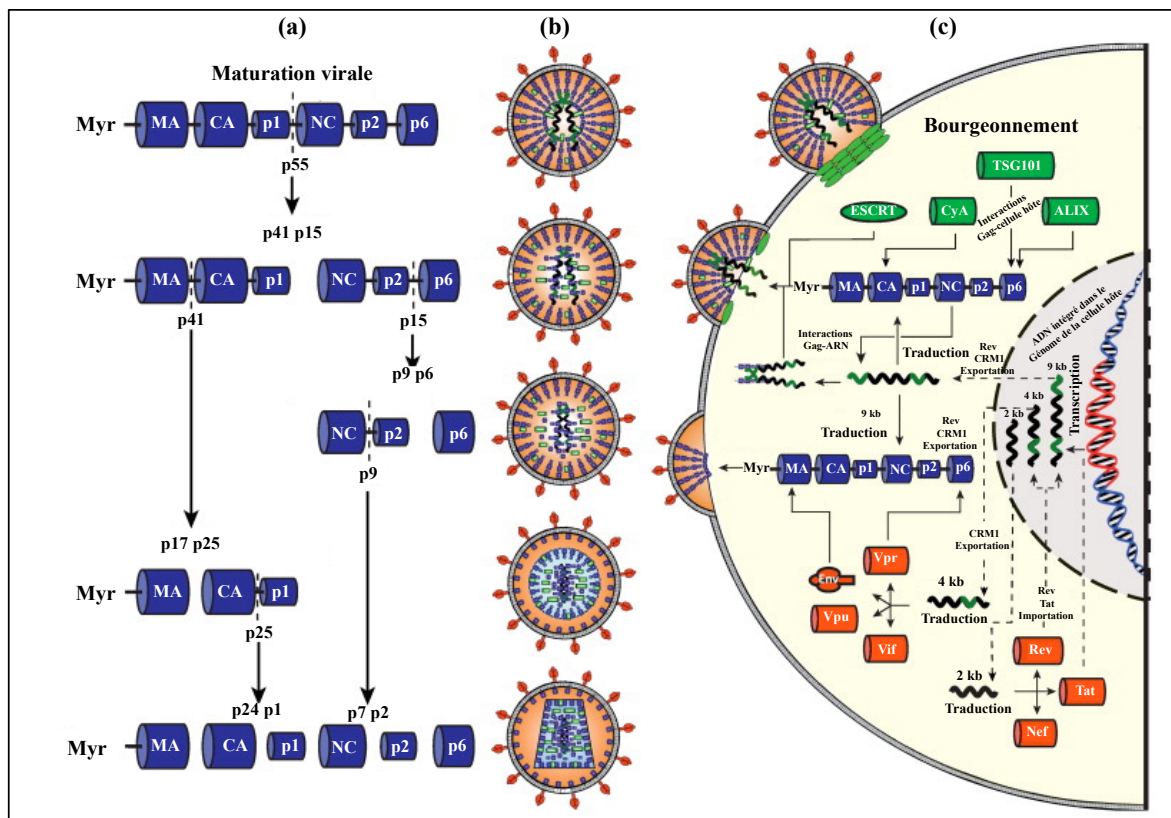


Figure 9: (a) Illustration montrant les intermédiaires obtenus durant les différentes étapes de clivage de la protéine Gag en relation avec la maturation des particules virales (Myr représente l'acide myristique). (b) Représentation schématique des changements structuraux au niveau de l'ARN génomique durant la phase de maturation des virions. (c) Les différentes étapes du cycle viral de l'intégration de l'ADN proviral dans le génome de la cellule hôte jusqu'au bourgeonnement. (Adaptée⁷¹)

I.1.2.9.1 Le génome du VIH-1 est encapsidé sous-forme de dimère

Au moment où les particules du VIH-1 s'assemblent dans les cellules infectées, deux copies non-épissées du génome sont sélectionnées sous la forme d'un dimère non-covalent parmi une large collection d'ARN cellulaires et d'ARN viraux épissés. Des analyses de microscopie électronique ont montré que les deux monomères sont liés d'une façon globalement parallèle au niveau de leurs extrémités 5'.

Des questions restent à résoudre : Pourquoi le génome viral est-il encapsidé sous forme de dimère d'ARN ? Existe-il d'autres formes d'ARN encapsidés?

La présence de dimères d'ARN dans les virions n'indique pas s'ils ont été capturés par la machinerie d'encapsidation sous forme de dimère ou sous la forme de monomères qui auraient dimérisé juste après l'assemblage des particules virales. Il a été généralement trouvé que les rétrovirus mutés qui sont défectifs au niveau de la dimérisation de l'ARN le sont aussi pour l'encapsidation. Dans les rétrovirus au phénotype sauvage récoltés juste après le bourgeonnement, l'ARN est sous forme de dimère. Par ailleurs *Fu et al.*⁶⁷ ont rapporté que l'ARN génomique isolé des virions (PR⁻) est faiblement dimérique sur gel (Fig.10) puisque ce dimère n'est pas stable et présente des différences au niveau structural. En effet, il se dissocie en monomères à température ambiante ou en chauffant à de faibles températures (65°C, pendant 10 minutes) plus rapidement que l'ARN génomique provenant des virions au phénotype sauvage (Fig.10). Ceci indique que l'ARN est encapsidé sous forme d'un dimère immature probablement lié au niveau du DIS et NC-dépendent pour la maturation.

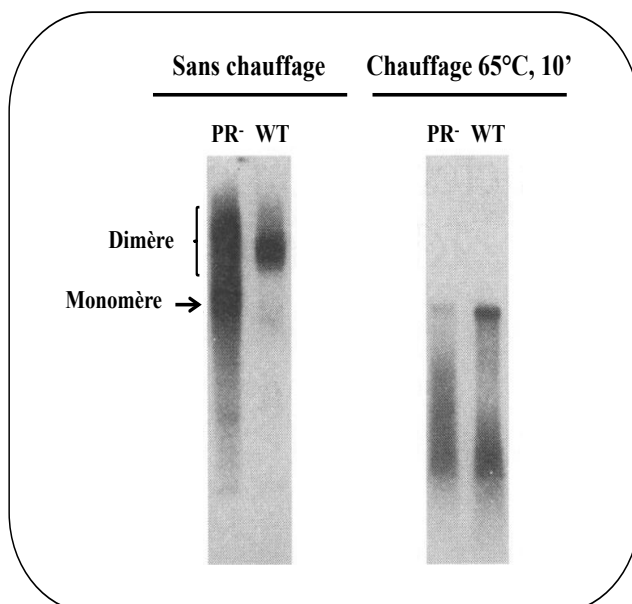


Figure 10: Analyse sur gel d'électrophorèse de l'ARN génomique extrait des virions (WT) et des virions de type (PR⁻). Ce gel montre que la structure de l'ARN est protéase-dépendante. Le gel à gauche montre que dans les virions (PR⁻), le dimère d'ARN est moins stable que dans le WT et qu'il se dissocie même à température ambiante. Après chauffage à 65°C (gel à droite) le dimère n'est quasiment pas observé pour les virions PR⁻ alors qu'il est encore visible pour les virions sauvages (bande du haut). (Adaptée⁶⁷)

Des résultats similaires (Fig.11) ont été obtenus par *Jalalirad et al.*⁷⁴. Leurs travaux avec des virions mutants bloqués à différents stades de la maturation de la protéine Gag, montrent que l'ARN est encapsidé sous-forme de dimère immature et la maturation se fait au fur et à mesure du clivage de la protéine Gag par la protéase. En clivant le domaine p15 pour donner les domaines p9 et p7, le dimère d'ARN se stabilise et devient plus compact⁷⁴.

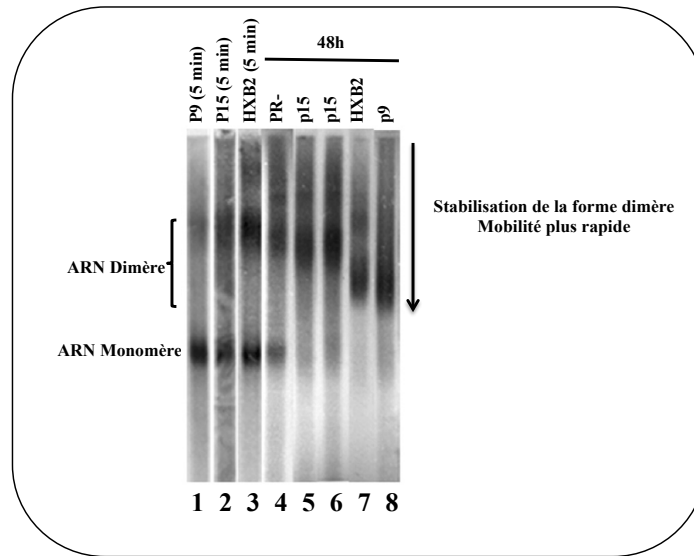


Figure 11: Gel d'électrophorèse montrant la présence d'ARN viral sous forme monomère en majorité dans les particules virales de phénotype sauvage libérées après bourgeonnement (les pistes 1 à 3, 5 minutes après le bourgeonnement). La formation du dimère d'ARN se stabilise en avançant dans le phénomène de maturation (les pistes 5 à 8, des virions après 48h), le dimère d'ARN est plus compact et sa mobilité augmente. (Adaptée⁷⁴)

I.1.3 Organisation génétique et structurale du génome

Le génome du VIH-1 est encapsidé sous forme de deux copies monocaténares d'ARN de 10,000 nucléotides chacune et de polarité positive (Fig.12). Les deux copies d'ARN sont associées d'une façon non-covalente proche de l'extrémité 5' dans une région appelée « DLS » (Dimer Linkage Structure)⁷⁷. La nature et les sites des liaisons existantes entre les deux monomères n'est pas complètement élucidé, mais ce sont essentiellement des liaisons hydrogènes et d'autres liaisons faibles, puisque le dimère peut se dissocier dans des conditions douces^{78,79,80}.

1) Les régions non-codantes :

La région 5'UTR est la région la plus structurée de l'ARN du VIH-1. Elle contient un nombre de séquences reconnues par des protéines ou ARN qui contrôlent des étapes cruciales du cycle viral, comme l'export du noyau vers le cytoplasme et l'encapsidation. Elle peut adopter des structures secondaires ou tertiaires qui peuvent jouer un rôle important dans la réplication virale. Différentes études *in-vitro* et *ex-vivo* y-compris des travaux effectués dans notre laboratoire^{81,82,83,84,85,86,87,88} ont permis l'élucidation de la structure secondaire de cette région. Elle est composée des domaines suivants :

- Le début de la région 5'UTR contient la région R, présente aussi dans la région 3'UTR et comprend la séquence TAR (Tat Responsive element), une structure de 58 nucléotides qui permet l'initiation de la transcription de l'ARN viral non épissé, suivi par la tige-boucle polyA qui contient un signal de polyadénylation qui est inhibé à l'extrémité 5' mais complètement fonctionnel à l'extrémité 3' de l'ARN viral.
- La région U5, formée de 83 nucléotides contient un des deux sites cruciaux pour l'intégration de l'ADN viral dans le génome de la cellule hôte.
- En aval de cette région on trouve le site PBS (Primer Binding Site), une séquence de 18 nucléotides complémentaires à l'extrémité 3' de l'ARNt₃^{Lys}, utilisé pour se fixer à l'ARNt cellulaire et créer ainsi un site d'initiation de la rétro-transcription. La région Psi (putative packaging domain) sous forme de quatre tiges boucles nommées par ordre SL1, SL2, SL3 et SL4 (Fig.12C). SL1 est le site d'initiation de la dimérisation (DIS ou Dimerization Initiation Site)^{89,90}. SL2 ou SD1 (major Splice Donor site) est responsable de la formation des différents ARNm épissés, SL3 est impliqué dans l'encapsidation, SL4 contient le codon d'initiation AUG de la traduction du gène codant pour la protéine Gag⁹¹. SL4 est une tige-boucle instable et peut former un appariement avec la région U5^{92,93}.

2) La région-codante :

Elle Contient les gènes codant pour les protéines Gag, Pol, Env, Tat, Rev, Nef, Vif, Vpr, et Vpu.

- **Gag** : Le gène codant pour la protéine Gag qui est clivée par la protéase pour donner les protéines de la matrice (MA), de la capside (CA) et de la nucléocapside (NC) qui est capable de se lier à l'ARN viral.
- **Pol** : Code pour les protéines enzymatiques telles que la rétro-transcriptase qui va permettre la rétro-transcription de l'ARN viral monocaténaire en ADN double brin. L'intégrase qui va permettre d'intégrer le génome viral dans le génome de la cellule hôte grâce à la formation d'un complexe d'intégration ADN viral-intégrase et la protéase qui permet la maturation des précurseurs Gag et Gag-pol.
- **Env** : Code pour les glycoprotéines d'enveloppe, la gp160 qui va être clivée par des protéases cellulaires en deux protéines la gp120 (protéine de surface) et la gp41 (protéine transmembranaire). Ces protéines gp120 et gp41 sont nécessaires pour la reconnaissance et l'entrée dans la cellule hôte.

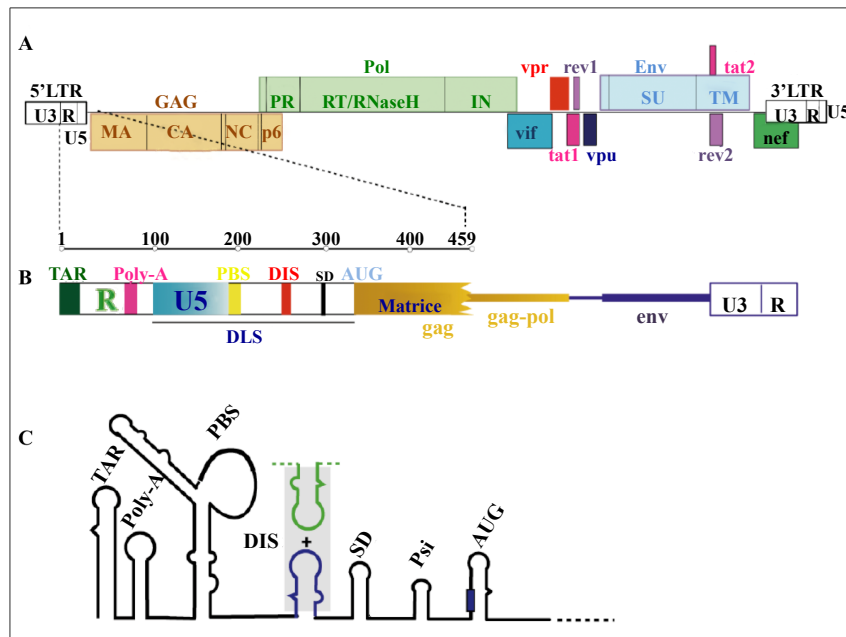


Figure 12: (A) L'organisation génétique de l'ARN proviral du VIH-1 (B) Représentation schématique de l'ARN génomique du VIH-1 (C) Modèle de structure secondaire des 500 premiers nucléotides de l'ARN génomique du VIH-1. (Adaptée⁹⁴)

I.2 Les Acides RiboNucléiques « ARN »

L'ARN ou Acide RiboNucléique est un polymère linéaire constitué d'un enchaînement de nucléotides : un groupement phosphate, un ribose sur lequel on trouve un groupement 2'-OH et non pas un groupement 2'-H comme pour l'ADN, et une base azotée (la cytosine, l'uracile, l'adénine et la guanine) (Fig.13). La présence d'un groupement 2'-OH rend la molécule d'ARN chimiquement labile, l'ARN s'autolyse en conditions alcalines (Fig.13). Il est synthétisé par transcription d'une matrice ADN dont il est la copie. L'ARN est présent dans les cellules sous forme d'un simple brin plus ou moins structuré. L'information génétique est segmentée en gènes dont l'expression est à l'origine de la synthèse des protéines.

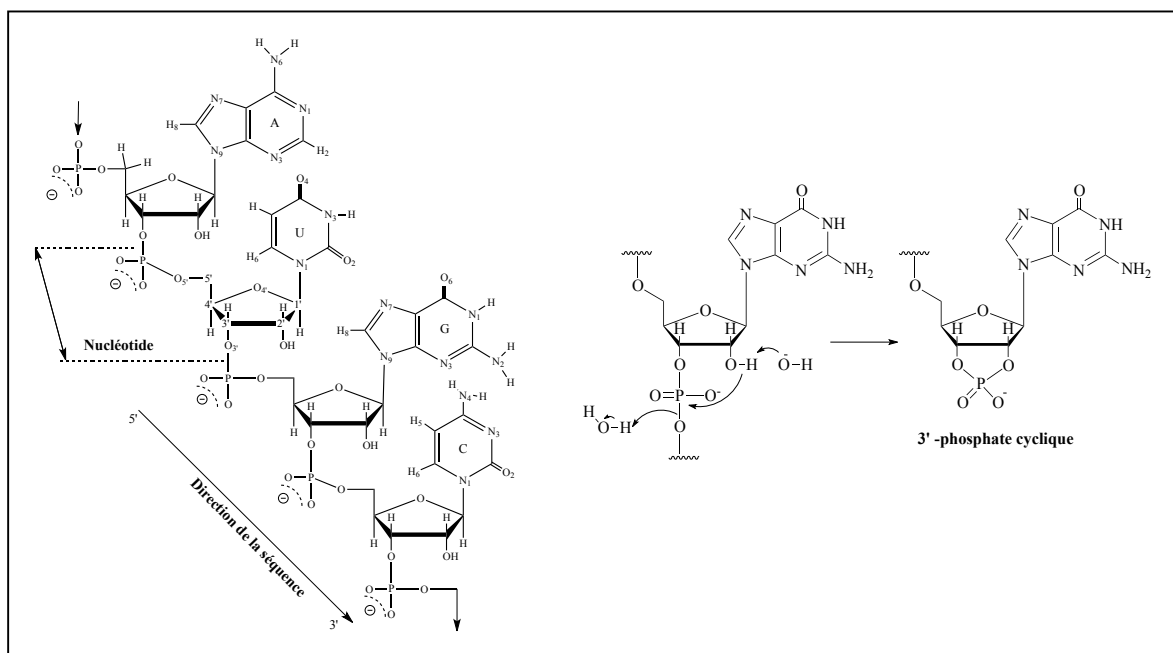


Figure 13: Structure chimique de l'ARN et modèle d'autolyse.

I.2.1 Différentes familles d'ARN et leurs rôles dans les différents processus cellulaires

Il existe deux grandes familles d'ARN qui possèdent chacune une structure et une fonction particulière. Les ARN codants tels que l'ARN messager, et les ARN non-codants⁹⁵ tels que les ARN transcriptionnels (ARNt et ARNr) et les petits ARN de 20 à 30 nucléotides et qui participent à différents mécanismes métaboliques dont la maturation de l'ARN (ARNsi, ARNmi, ARNsno, ARNsn). (Fig.14)

Les ARN messagers sont la matrice pour la synthèse des protéines. On distingue cinq régions principales dans un ARNm : la coiffe à l'extrémité 5' par addition d'une 7-méthylguanosine (N7) sur le premier nucléotide par une liaison 5'-5' triphosphate (7mGpppN), la région 5' non traduite (5'UTR), une séquence codante, la région non traduite (3'UTR) et une queue polyA à l'extrémité 3'.

Les ARN ribosomiques rentrent dans la composition des ribosomes, la machinerie responsable de la synthèse des protéines. En effet 65% à 70% de la masse du ribosome est composée d'ARN. La petite sous-unité du ribosome des eucaryotes 40S contient 1 ARNr avec 35 protéines différentes. La grande sous-unité 60S contient 3 ARNr avec 50 protéines.

Les ARN de transfert sont des ARN courts de 70 à 100 nucléotides. Ils portent les acides aminés et permettent leur incorporation dans les protéines. Les ARNt présentent une structure tertiaire spécifique en feuille de trèfle. Le bras accepteur est une tige supérieure qui porte les deux extrémités 5' et 3'. Le bras anticodon est la tige inférieure, et les deux bras (D) et (T) du milieu contiennent des ribonucléotides modifiés : la ribothymidine (T) et la dihydrouridine (D)⁹⁶.

Les ARN interférents (ARNsi et ARNmi) : régulent l'expression des gènes en ciblant la dégradation des ARNm, ou en inhibant la traduction des protéines^{97,98,99}.

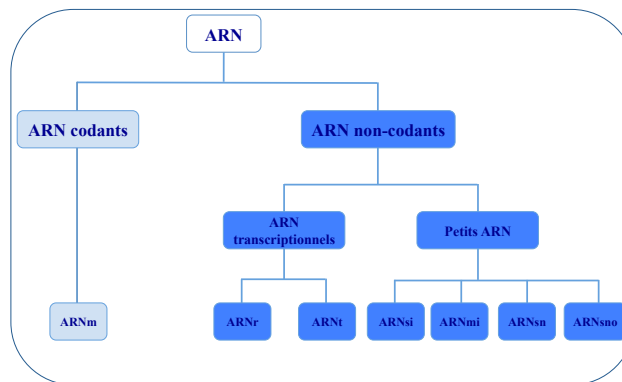


Figure 14: Classification des ARN

I.2.2 Structure des ARN : capacité de l'ARN à former des structures complexes

L'ARN adopte une large variété de structures secondaires et tertiaires en raison des interactions entre les quatre bases qui le composent ou des interactions avec d'autres ARN, des protéines ou des ligands.

I.2.2.1 Diversité d'appariement entre les bases

L'appariement de paires de bases peut varier entre des liaisons canoniques de type Watson-Crick à des liaisons non-canoniques tels les liaisons Wobble, Hoogsteen, Hoogsteen réverse, Watson-Crick réverse et des liaisons de paires de bases cisailées¹⁰⁰ (Fig.15A). Ceux-ci peuvent être classés en fonction des faces d'interaction sur les bases, de l'orientation de la base par rapport au sucre (Syn/Anti), et du sens du squelette ribose-phosphate.

Les liaisons hydrogène dans les appariements Hoogsteen impliquent une face de la purine différente de la face Watson-Crick et en particulier les positions N7 et le groupement oxo ou amino en position 6. En fonction de l'orientation de la base purine par rapport à la pyrimidine on peut distinguer les deux types de liaisons Hoogsteen. Si les deux riboses sont du même côté de l'axe de la paire de base alors la liaison est de type Hoogsteen et si les deux riboses sont de part et d'autre de la paire de base alors la liaison est de type Hoogsteen-réverse¹⁰¹. L'appariement de paires de base impliquant le petit sillon d'une seule base est un appariement cisailé. Dans ce type d'appariement, une base expose ses sites Watson-Crick dans le grand sillon et l'autre base expose ses sites Watson-Crick dans le petit sillon. On peut observer des appariements de bases purine-purine et pyrimidine-pyrimidine de ce type^{102,103}.

Dans un second temps, les bases peuvent se rapprocher de façon à ce que leur ribose soient du même côté ou du côté opposé de la ligne médiane de la liaison hydrogène entre les deux bases (appelé aussi axe d'interaction). Si les deux riboses sont du même côté, alors l'appariement est appelé « cis » et si les deux riboses sont de part et d'autre de la ligne médiane alors l'appariement est appelé « trans ». (Fig.15B)

Enfin certaines paires de bases adoptent des structures avec des liaisons hydrogène non-standards, ou d'une façon plus appropriée sont appelées des liaisons hydrogène chélatées. Ce type de liaisons est observé dans la structure de la boucle E de l'ARNr.

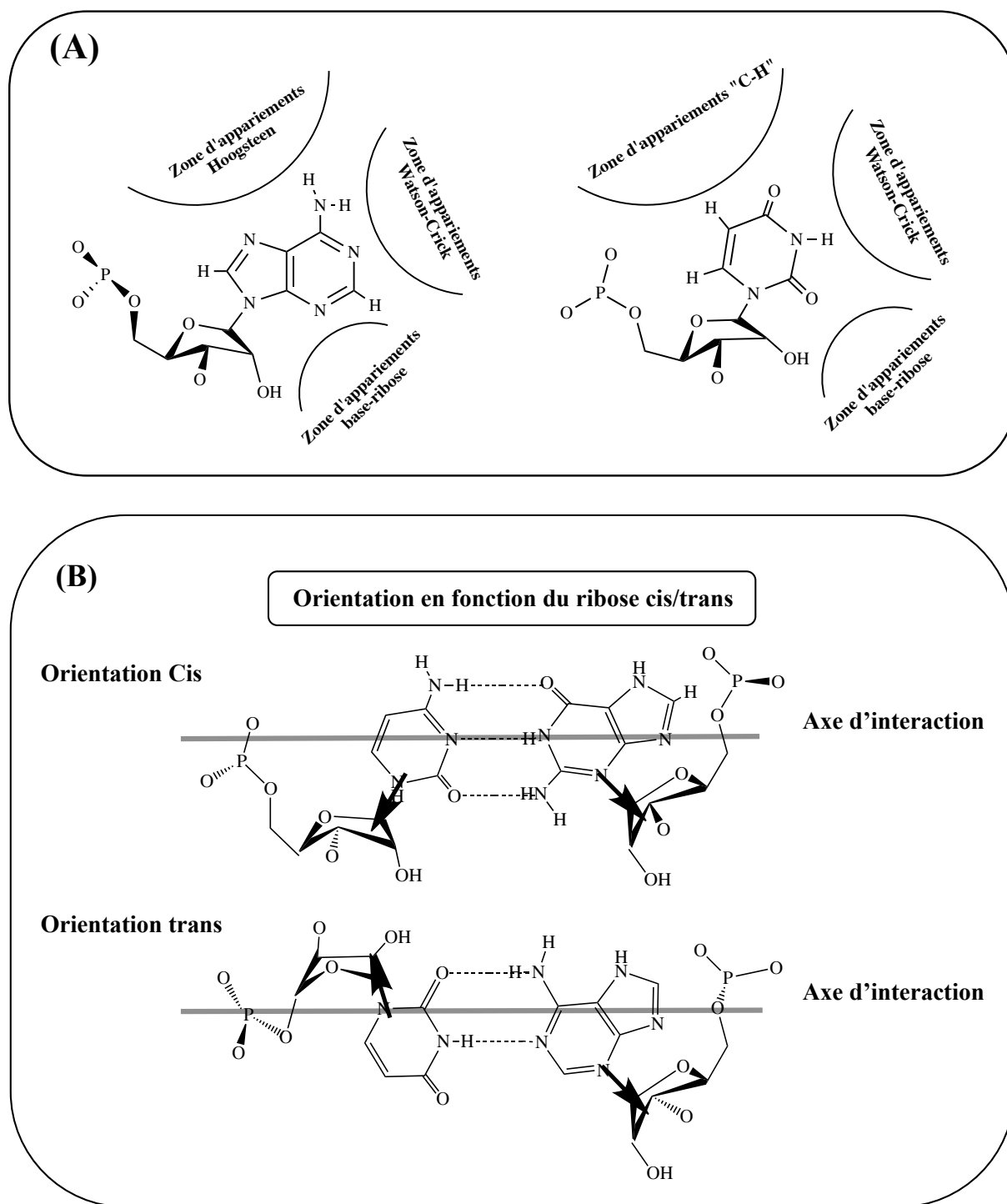


Figure 15: (A) Représentation des différents types d'appariements de paire de base. (B) Orientations cis et trans possibles entre deux paires de base en fonction de l'orientation du sucre : si les deux riboses sont du même côté de l'axe d'interaction alors la conformation est « cis », si les deux riboses sont de part et d'autre de l'axe d'interaction alors la conformation est « trans ». (Adaptée¹⁰⁰)

En plus des appariements Watson-Crick, il existe des appariements non Watson-Crick de type bancals ($G^{\circ}U$) ou Wobble. Ces appariements se trouvent dans les structures d'ARN qui

servent comme des sites de reconnaissance et de liaison à des protéines, acides nucléiques, ligands ou ions. Ces appariements de paires de base sont au moins liés par une liaison hydrogène. Les appariements A-U et G=C sont isostères, tandis que l'appariement G°U provoque une distorsion du squelette ribose-phosphate¹⁰⁴. (Fig.16)

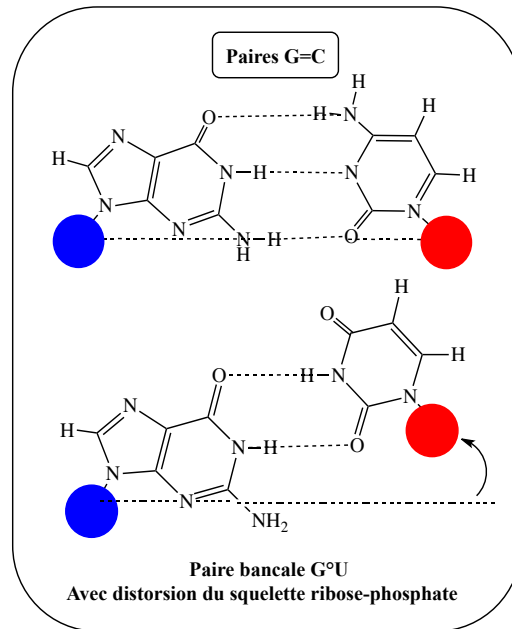


Figure 16: Un modèle d'appariement de type Watson-Crick G=C et un modèle d'appariement de type bancale G°U avec la représentation de la distorsion du squelette ribose-phosphate.

I.2.2.2 Structures secondaires

L'ARN est capable de former des architectures complexes comme des structures secondaires. Les structures secondaires sont des successions de motifs élémentaires formés par des appariements de type Watson-Crick (A-U et G=C) ou des appariements dits bancals (G°U) ou Wobble¹⁰⁵.

On peut distinguer cinq motifs de structure secondaire¹⁰⁶ (Fig.17):

- La tige ou hélice qui est une succession d'appariement de paires de base consécutives.
- La boucle terminale qui est fermée par une paire de base appariée.
- La boucle interne qui est symétrique ou pas fermée par deux paires de bases.
- les boucles multiples qui relient les hélices entre elles.

- La boucle latérale (bulge) qui se trouve sur un seul brin de l'hélice avec une continuité d'appariements de part et d'autre de la boucle latérale et qui restent empilés d'une façon coaxiale.

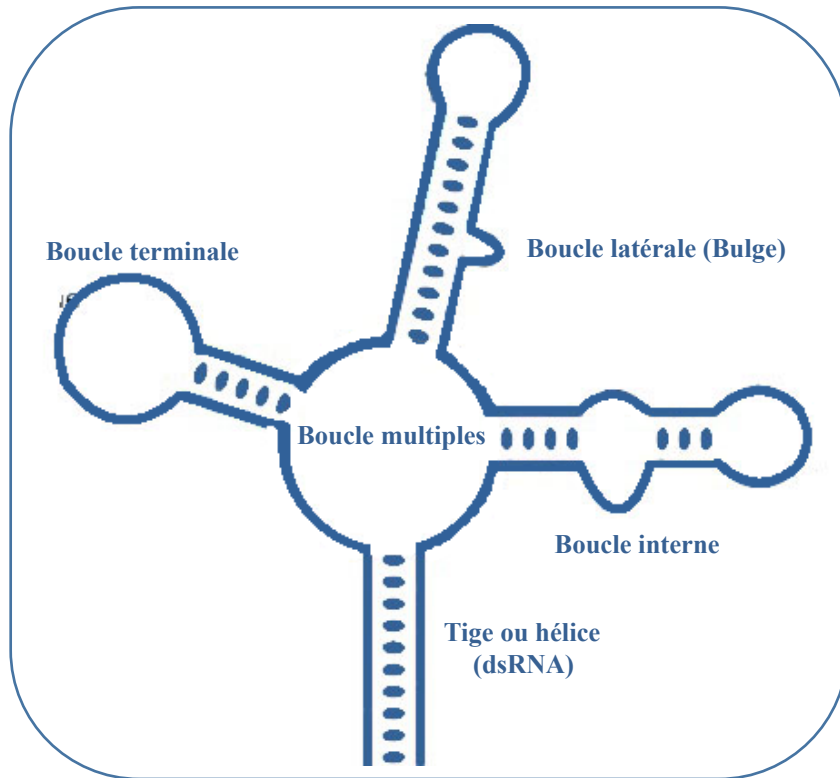


Figure 17: Les différents types de structures secondaires d'ARN. (Adaptée¹⁰⁷)

La structure secondaire des ARN n'est pas forcément liée à leur séquence. En effet, des ARN appartenant à la famille des Introns du groupe I possèdent des structures et des fonctions identiques alors que leur séquences sont très différentes et d'un autre côté, des ARN avec une séquence identique peuvent adopter des formes et des fonctions différentes. De la même façon, un même ARN peut changer de conformation en fonction de son environnement et de petits métabolites qui se fixent à lui, c'est le cas des Riboswitches

I.2.2.3 Structures tertiaires

Les structures tertiaires sont des structures plus complexes dans lesquelles les hélices et les régions non-appariées sont bien organisées dans l'espace^{108,109,110}. Les molécules d'ARN sont capables d'un phénomène d'auto-assemblage pour former des architectures plus complexes parce qu'elles contiennent en plus des liaisons Watson-Crick entre les paires de base qui

forment les structures secondaires, des liaisons non-Watson-Crick et différents types de contacts entre les segments de la chaîne poly-nucléotidique.

Les éléments des structures secondaires sont associés par de nombreuses liaisons Van der Waals, des liaisons hydrogènes spécifiques par la formation d'un petit nombre de liaisons Watson-Crick additionnels ou des appariements de bases inhabituels impliquant les boucles. Ces phénomènes de repliement dépendent d'habitude de la présence d'ions divalent^{111,112} comme par exemple les ions magnésium Mg^{2+} et sont influencés par le changement de température. La réduction de la charge nette de la molécule d'ARN en associant des cations, induit un changement conformationnel dans la chaîne d'ARN qui est à l'origine de la formation d'une structure plus compacte ou structure tertiaire^{113,114,115}.

Les structures tertiaires impliquent des interactions entre deux hélices, deux régions non appariées, ou une région non appariée avec une double hélice^{109,107,116}.

Les interactions entre les deux hélices sont de deux types: soit par empilement des brins proches les uns des autres, soit entre deux hélices distantes l'une de l'autre mais qui se positionnent de façon à ce que leurs petits sillons s'accordent.

Quatre motifs de structure tertiaires peuvent se présenter (Fig.18) :

- Une jonction coaxiale formée par deux tiges qui se co-linéarisent.
- Le pseudonoeud qui provient des interactions entre les régions non-appariées présentes dans les boucles internes ou les boucles terminales avec un brin libre complémentaire en dehors de la boucle.
- « Kissing Loop » est le résultat d'interactions entre deux tiges boucles qui aboutissent à une structure de boucles appariées ou « Kissing Loop ».
- Les interactions entre une région non-appariée et une hélice double brin aboutissent à plusieurs motifs tels que les triplets et les quadruplets. Cette interaction a lieu dans le grand sillon ou à la surface de la double hélice. Ces motifs possèdent un potentiel élevé à former des structures tertiaires parce qu'ils conduisent à une co-axialité dans la double hélice. Ils sont à l'origine d'un ARN compact surtout en présence d'ions divalent ou de polyamines.

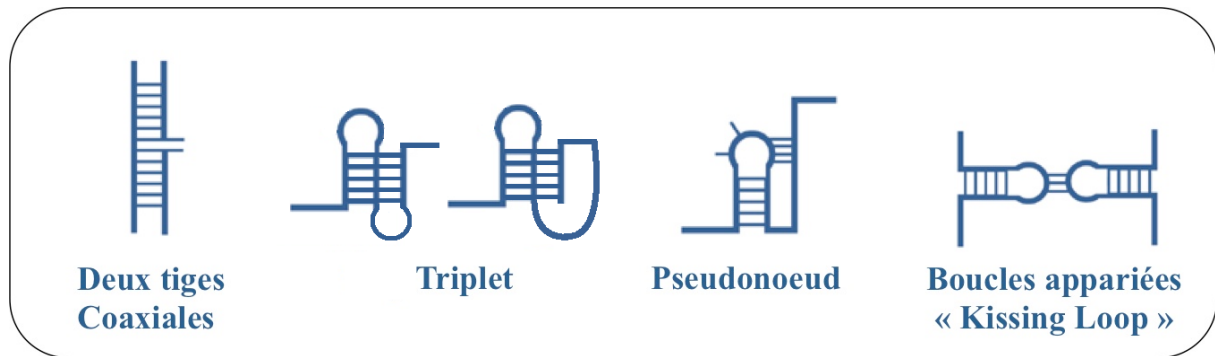


Figure 18: Les différents types de structures tertiaires d'ARN. (Adaptée¹⁰⁷)

I.3 L'étude de la structure secondaire des ARN

Comme nous venons de le voir, une bonne compréhension du fonctionnement d'un ARN nécessite la connaissance de la structure secondaire de cette macromolécule. A l'heure actuelle, il n'est pas possible de prédire correctement la structure secondaire d'un ARN à partir de sa seule séquence primaire. Mais deux méthodes non expérimentales permettent de s'en rapprocher.

Ainsi, la méthode phylogénétique est basée sur le fait que des ARN qui remplissent la même fonction chez des espèces différentes possèdent souvent une structure secondaire proche. En général les régions en hélice et leur topologie sont globalement conservées même si les appariements qui les constituent ne sont pas forcément les mêmes. Ainsi, à partir d'un ARN dont la structure secondaire est établie, on peut prédire la structure de nombreux autres. Les modèles sont construits en effectuant des analyses statistiques dans les séquences et en identifiant les co-variations dans les régions appariées des séquences d'ARN homologues. Il est également possible de valider des modèles de structures secondaires en vérifiant leur compatibilité avec les co-variations observées dans l'évolution, cette approche a été utilisée pour valider les structures secondaires des ARN ribosomiques.

Il est également possible de réaliser des prédictions bio-informatiques. Cette méthode est basée sur le calcul de l'enthalpie libre de toutes les conformations possibles pour une séquence donnée, la conformation d'énergie minimale correspondant à l'état le plus stable et donc le plus probable. Un des premiers algorithmes développés pour réaliser ces prédictions est celui de « Nussinov »¹¹⁷. Cet algorithme maximise le nombre d'appariements entre les bases mais ne tient pas compte des empilements entre les paires de bases successives. En

1981, « M. Zuker »¹¹⁸ a amélioré l'algorithme de Nussinov en prenant en compte ces empilements. Zuker a apporté une autre amélioration en créant un algorithme qui ne donne pas uniquement la structure d'énergie minimale mais également les structures sous-optimales dont l'énergie est légèrement supérieure. Ce dernier algorithme est utilisé par le programme mFold et des variantes ont été créées afin de pouvoir incorporer des données de phylogénétique ou des données expérimentales qui affinent les résultats.

Les méthodes expérimentales sont le meilleur moyen pour obtenir des informations solides sur la structure des ARN. Il est possible d'utiliser la cristallographie suivie par l'analyse aux rayons X, la RMN ou encore la microscopie cryo-électronique. Pour les longs ARN en solution il existe une méthode développée dans les années 1980 et très largement utilisée qui consiste à modifier l'ARN par des sondes chimiques ou enzymatiques qui réagissent spécifiquement avec certaines structures et certaines positions (Fig.19). Dans un deuxième temps les positions réactives de l'ARN sont analysées par migration sur gel par électrophorèse.

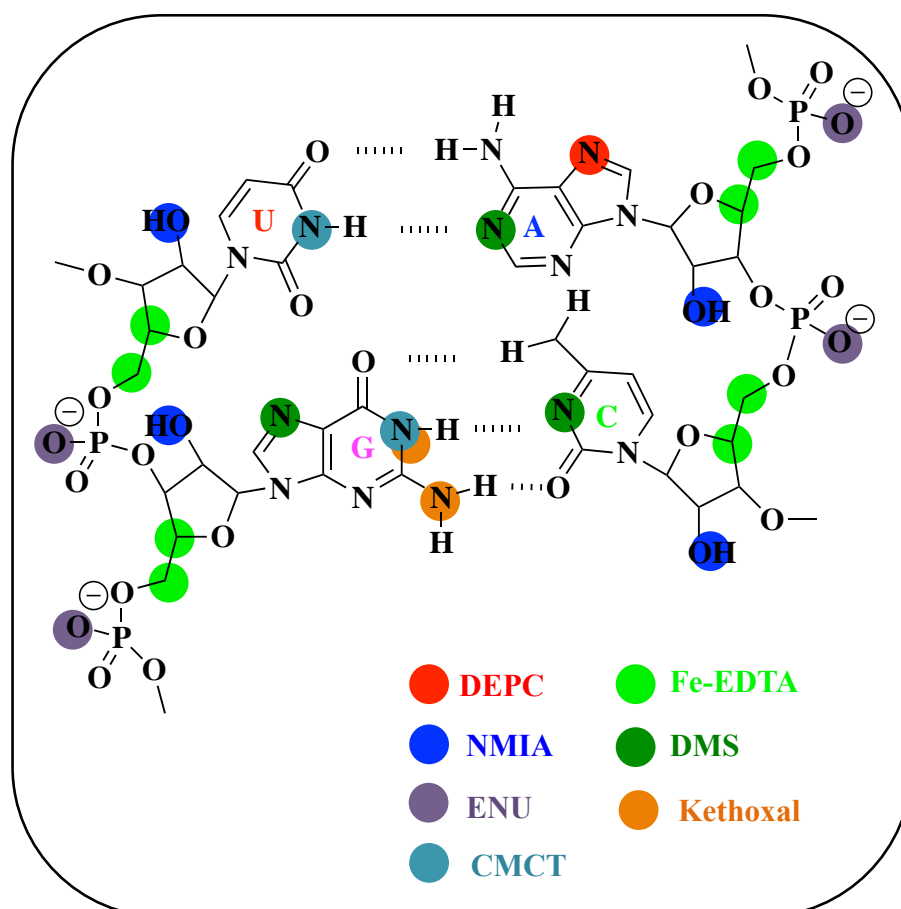


Figure 19: Les différents sites de modification avec des sondes enzymatiques (A) et des sondes chimiques (B).

Des caractéristiques propres à chaque sonde régissent la qualité ultime et l'utilité de l'information structurale déduite. Dans tous les cas, l'ARN doit être modifié une seule fois avec l'enzyme puisque une deuxième modification ne donne pas des informations sur l'état natif ou naturel de cet ARN. À la fin de la réaction, les ARN présents durant le marquage contiennent une empreinte exacte du phénomène de marquage. Le défi est alors d'extraire les informations aussi précises que possible.

I.3.1 Les sondes enzymatiques

1. **RNase T1** : de « *Aspergillus oryzae* » avec un poids moléculaire de 11,000 Da, est une endoribonucléase des ARN simple brins et plus spécifiquement au niveau des bases G (Guanosine). Elle clive au niveau des liaisons internucléotides adjacentes à un 3'-phosphate des résidus guanosine non-appariés, avec la formation d'un intermédiaire guanosine 2'-3'-phosphate cyclique. L'hydrolyse génère des fragments avec une extrémité 3'-phosphate^{119,120,121,122}. Deux valeurs de pH optimales ont été déterminées, la première à 7.5 et la deuxième à 4.5¹²³. À pH= 4.5, la présence d'une concentration 7M en urée augmente l'activité de l'enzyme¹²⁴. Il a été montré que les positions N1 et O6 de la guanosine interagissent par des liaisons hydrogènes avec la nucléase. Les positions naturellement modifiées dans l'ARN, m¹G et m⁷G ne sont pas reconnues par l'enzyme. Elle a été utilisée pour étudier la structure secondaire d'un fragment d'ARN de 345 bases couvrant le domaine de liaison protéique S8/S15 de l'ARNr 16S de *E.Coli*^{122,125} et pour étudier en détails la conformation du domaine III de l'ARNr 16S (nucléotides entre 913-1408)¹²⁶.
2. **RNase U2** : de « *Ustilago sphaerogena* » avec un poids moléculaire de 12,490 Da, clive préférentiellement les liaisons 3'-5' phosphodiester des adénosines non-appariées, mais sa spécificité n'est pas absolue (A>G>>C>U)¹²⁷. Le pH optimal est de 4.5 et l'enzyme est active en présence d'une concentration 7M en urée. Le mécanisme de clivage est similaire à celui de la RNase T1. Les conditions d'hydrolyse ont été adaptées à pH neutre pour marquer les adénosines des ARN simples brins^{125,126}.

3. **RNase CL3** : possède un poids moléculaire de 16,800 Da et clive principalement les cytidines non-appariées en laissant une extrémité 3'-phosphate^{128,129}. L'enzyme clive aussi au niveau de l'adénosine et de l'uridine mais nécessite un temps d'incubation et des concentrations d'enzyme plus élevés. La valeur de pH optimale dépend du type de tampon utilisé dans la réaction (dans le cas du tris-HCl, le pH optimal est de 7.5). L'activité de la RNase augmente en présence d'ions Mg^{2+} et de spermine¹²⁹. Elle a été utilisée pour étudier l'organisation structurale des 159 nucléotides terminaux de l'extrémité 3'-OH de l'ARNt du virus TYMV en combinaison avec des digestions par les nucléases S1 et T1¹²⁰.
4. **RNase T2** : de « *Aspergillus orizae* » avec un poids moléculaire de 36,000 Da, est une endonucléase spécifique des ARN simples brins avec une préférence pour les résidus adénosines¹²². Le mécanisme de clivage est similaire à celui proposé pour la RNase T1. L'activité de cette enzyme ne requiert pas la présence de cations divalents, donc elle a été utilisée pour étudier le rôle des ions Mg^{2+} dans la stabilisation de la structure d'ARN. La valeur de pH optimal est de 4.5, mais l'enzyme est stable à pH neutre¹²⁷. La RNase T2 est inhibée par les ions métalliques lourds tels que le Cu^{2+} . Elle a été utilisée pour analyser et détecter les effets de pH et d'ions métalliques sur les changements conformationnels de l'ARNt^{Phe} des levures¹³⁰.
5. **RNase A** : clive les ARN simple brins et plus spécifiquement au niveau des bases pyrimidines (C et U). Ces enzymes clivent au niveau de la liaison phosphodiester en 3' au niveau d'une base spécifique non impliquée dans une liaison Watson-Crick et l'extrémité 5'-OH au niveau de la base adjacente. Le résultat est un phosphate 2', 3' cyclique¹³¹.
6. **Nucléase S1** : de « *Aspergillus orizae* » avec un poids moléculaire de 32,000 Da, elle clive aussi les ARN simples brins au niveau des bases non-appariées. Elle est stimulée par les ions Zn^{2+} et laisse des fragments avec des extrémités 5'-phosphate. La valeur de pH optimale est de 4.5 et l'hydrolyse est effectuée à pH neutre avec des concentrations élevées de l'enzyme. La nucléase S1 a été largement utilisée à pH= 4.5 pour modifier des ARN dans leurs régions non-appariées. Elle a été utilisée pour étudier les changements conformationnels de l'ARNt^{Phe} des levures et de l'ARNt₂^{Glu} de *E. Coli*¹¹⁹, les propriétés de l'extrémité 3' de l'ARN de TYMV (Turnip Yellow Mosaic Virus)¹²¹ et pour déterminer un modèle de structure tertiaire de l'ARN 5S de *E.*

Coli. La comparaison de ce modèle avec 18 autres de l'ARN 5S appartenant à la famille des procaryotes a permis de généraliser une structure tertiaire de cet ARN, dans lequel la majorité des nucléotides conservés sont dans les régions non-appariées¹³².

7. **Nucléase *Neurospora crassa*** : possède un poids moléculaire de 55,000 Da et clive les ARN ou les ADN simples brins en laissant des fragments avec des extrémités 5'-phosphate. La valeur de pH optimal est entre 7.5 et 8.0 et l'enzyme reste active à 65°C. La présence d'EDTA inhibe son activité. Elle a été utilisée pour étudier les interactions entre l'ARNt^{Trp} et la rétro-transcriptase du myeloblastome aviaire¹³³. Les résultats ont montré que la rétro-transcriptase interagit plus spécifiquement avec le bras D, la boucle anticodon et le domaine psi de la boucle T de l'ARNt^{Trp}.
8. **RNase V1** : du venin de « *cobra Naja naja oxiana venom* » avec un poids moléculaire de 15,900 Da, elle clive d'une façon non-spécifique au niveau des bases appariées dans des doubles hélices. Elle génère des fragments avec une extrémité 5'-phosphate¹³⁴. Il a été montré que la taille minimale d'ARN est de 4 à 6 nucléotides et que 3 à 4 interactions ioniques sont formées entre des groupements chargés positivement de la protéine et les phosphates de l'ARN¹³⁵. La RNase V1 clive aussi les ARN simples brins mais dans des régions empilées^{136,135}. L'enzyme est active à pH entre 4.0 et 9.0. Elle a été utilisée pour étudier les régions appariées et hautement organisées des ARNt^{Phe}, ARNt^{Val}, et ARNt^{Asp} des levures avec une comparaison du modèle d'hydrolyse de chacun de ces ARN durant les différentes étapes de la réaction d'aminacylation¹³⁷, la structure secondaire de l'ARNr 5S de *Bombyx mori* et de *Dictyostelium discoideum* en combinaison avec la nucléase S1¹³⁸ et pour localiser le site de liaison de la protéine S4 sur l'ARNr 16S¹³⁹.
9. **Nucléase P1** : de « *Penicillium Citrinum* » avec un poids moléculaire de 44,000 Da, elle clive les régions non-appariées de l'ARN¹⁴⁰. Elle a été utilisée à pH acide entre 4.0 et 5.0, qui est sa valeur de pH optimale. La température utilisée est dans l'intervalle entre 37°C et 70°C. Les produits obtenus par le clivage de l'ARN sont des fragments avec des extrémités 5-monophosphosphate¹⁴¹. L'extrémité N-terminale de cette enzyme est formée par le résidu tryptophane et l'extrémité C-terminale est formée par le résidu leucine. Cette enzyme est une métalloprotéine (Zn²⁺) contenant trois atomes de zinc par molécule. La nucléase P1 a été utilisée pour étudier la

structure de l'ARNt^{Phe} des levures¹⁴², et pour contrôler l'incorporation de la cystéine et de la sélénocystéine de l'ARNt¹⁴³.

I.3.2 Les sondes chimiques

Les sondes chimiques sont utilisées pour tester des hypothèses déjà émises sur des structures secondaires de l'ARN ou pour déterminer de nouvelles structures d'ARN inconnues. Elles sont aussi utilisées pour déduire des liaisons tertiaires dans l'ARN, pour marquer un ARN lié à une protéine et pour suivre les différentes voies de repliement de l'ARN. C'est une technique de modification de l'ARN *in vivo* ou *in vitro* dépendant du temps de demi-vie du réactif chimique. Selon la nature du réactif chimique utilisé, quatre types d'informations peuvent être déterminées (Fig.20): l'accessibilité des nucléotides au solvant, souvent utilisée pour déterminer la structure secondaire de l'ARN, pour étudier le repliement des molécules d'ARN, et pour identifier les nucléotides protégés d'une attaque chimique en présence d'un ligand (radicaux hydroxyles), informations sur l'empilement des paires de bases, des liaisons hydrogènes et de l'environnement électrostatique (DMS, kethoxal, CMCT, DEPC, Bisulfite), l'environnement local électrostatique adjacent à chaque nucléotide (réactifs bi-fonctionnels), et la flexibilité locale de chaque nucléotide (réactifs de SHAPE : BzCN, NMIA et 1M7). Les réactifs chimiques existants se divisent en deux classes : spécifiques de la base (DMS, kethoxal, CMCT, DEPC), ou non (radicaux hydroxyles, réactifs de SHAPE)¹⁴⁴.

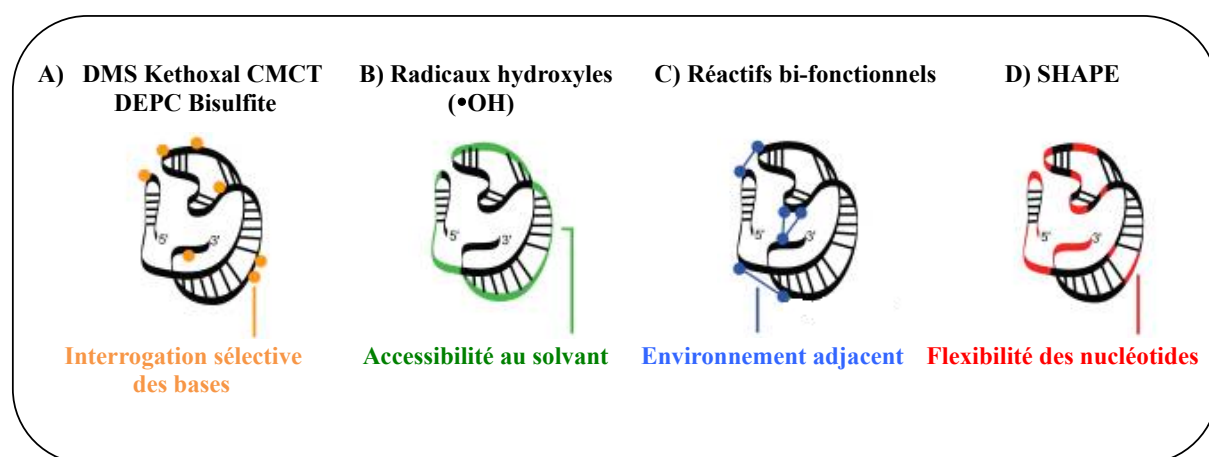


Figure 20: Informations obtenues en interrogeant l'ARN avec les différentes sondes chimiques existantes. (Adaptée¹⁴⁵)

Le 1-cyclohexyl-3'-(2-morpholinoethyl) (CMCT), le diméthyl sulfate (DMS), le kethoxal, le bisulfite et le diéthyl pyrocarbonate (DEPC) modifient les régions non-appariées de l'ARN, les régions appariées de l'ARN A-U et G-C à l'extrémité d'une hélice, et tous les appariements G-U, ou adjacents à ces liaisons G-U. Ils donnent des informations structurales sur l'appariement de paires de base, les liaisons hydrogènes, et sur l'environnement électrostatique adjacent à la base modifiée¹⁴⁶

Le DMS : Le DMS réagit principalement avec les positions N7-G, N1-A et N3-C à pH neutre^{147,148}. La réaction avec la guanosine permet d'ajouter un groupement méthyle en position N7, et la présence d'une charge positive perturbe la résonance électronique du cycle de purine. La perturbation de la double liaison 7,8 du cycle peut être réduite dans une solution diluée de tétrahydruroborate de sodium (NaBH₄) à pH= 8.3. Le résultat m⁷dihydroguanosine, qui s'hydrolyse partiellement ou complètement présente un site pour l'aniline qui induit la scission du double brin d'ARN¹⁴⁹. Cette réaction permet de localiser les positions des N7-G libres (Fig.21). L'absence de réactivité permet d'indiquer que cette position rentre dans un appariement de paires de bases de type non-canonique, ou en coordination avec des ions tels que le Mg²⁺. Cette modification est sensible aux empilements de paires de bases. L'équipe de « *Romby et al.* » a montré d'après des analyses comparatives d'une variété d'ARNt en solution, que les guanosines dans une conformation régulière de type hélice A sont réactives lorsqu'elles sont entourées par des bases pyrimidines¹⁵⁰. Ils ont aussi montré qu'une simple distorsion de l'hélice peut modifier cette règle.

La réaction avec les cytosines non-appariées permet d'ajouter un groupement méthyle en position N3. Le résultat méthyl-3-cytosine peut être détecté directement par extension d'amorce, mais il est également possible d'utiliser des ARN marqués à leurs extrémités et de cliver les ARN au niveau de la modification méthyl-3-cytosine par de l'hydrazine¹⁴⁶.

La réaction avec les adénosines non appariées permet d'ajouter un groupement méthyle en position N1. Cette modification est seulement détectée par extension d'amorce puisqu'il n'existe pas une scission avec d'autres réactifs chimiques du brin d'ARN¹⁴⁶.

Le DMS a été utilisé par exemple pour déterminer le site de liaison de la protéine S4 sur l'ARNr 16S¹³⁹ en marquant les positions N1-A et N3-C en combinaison avec du kethoxal (pour marquer les positions N1-G) et de la ribonucléase V1 (pour cliver les régions appariées de l'ARNr 16S). Il a été utilisé aussi pour déterminer la structure du domaine 5' de l'ARNr 18S des levures, dans des molécules natives dépourvues de protéines et dans les sous-unités

ribosomiques 40S¹⁵¹ et pour déterminer les domaines fonctionnels de l'ARN de la ribonucléase P de *E.Coli* en combinaison avec le CMCT¹⁵².

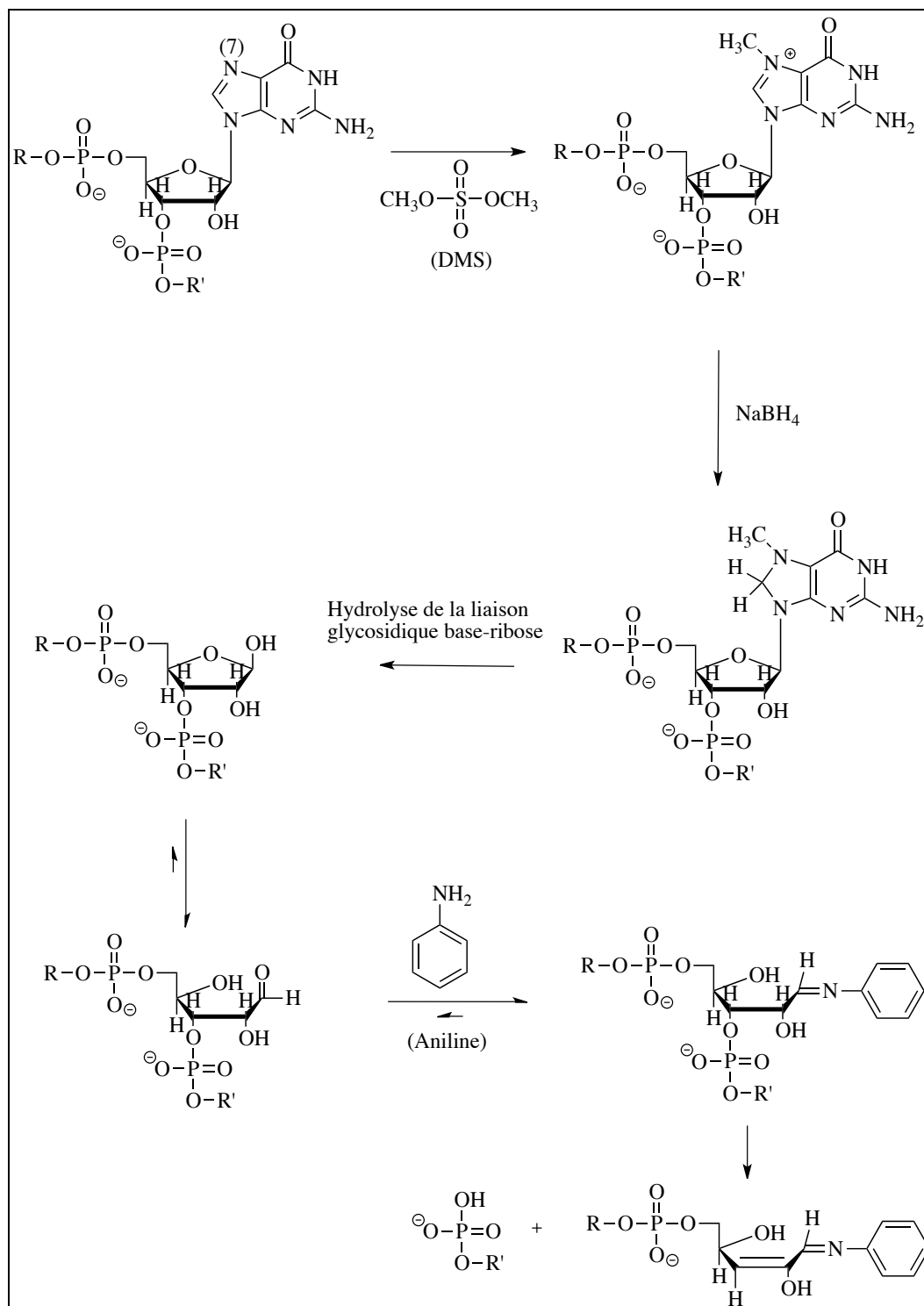


Figure 21: Représentation de la réaction du DMS sur la position N7-G, suivie de la réduction par le NaBH_4 . La réaction de l'aniline après hydrolyse de la liaison glycosidique entre la base modifiée et le ribose. Le cycle ouvert du ribose présente une forme tautomère, dans lequel l'aldéhyde réagit avec l'aniline pour donner l'aldime, et la liaison 3'-phosphoester est clivée par un mécanisme de β -élimination. (Adaptée¹⁴⁶)

Le DEPC : L'atome N7 de l'adénosine est particulièrement susceptible à la carbethoxylation à pH neutre^{153,154,155} (Fig.22). Cette modification détruit la résonance de l'hétérocycle et l'imidazole s'ouvre entre les atomes N7 et C8¹⁵³, permettant ainsi un site de liaison de l'aniline qui va conduire à la scission du double brin d'ARN¹⁵⁶. Cette réaction de marquage met en évidence la participation de la position N7-A dans des interactions tertiaires. En raison de son poids moléculaire élevé, le DEPC est plus sensible à l'empilement de paires de bases que le DMS et toutes les adénosines qui se trouvent dans les hélices ne sont pas réactives.

La réaction de carbethoxylation peut se produire aussi au niveau des positions N7-G en milieu acide et N3-U en milieu légèrement basique¹⁵⁴. Le DEPC a été utilisé pour analyser la structure de l'ARNr 16S de *E.Coli*¹⁵⁷ et pour déterminer la structure de l'ARN 10Sa (ARNtm) de *E.Coli* en combinaison avec les RNase V1, T1, S1, du DMS et du CMCT¹⁵⁸.

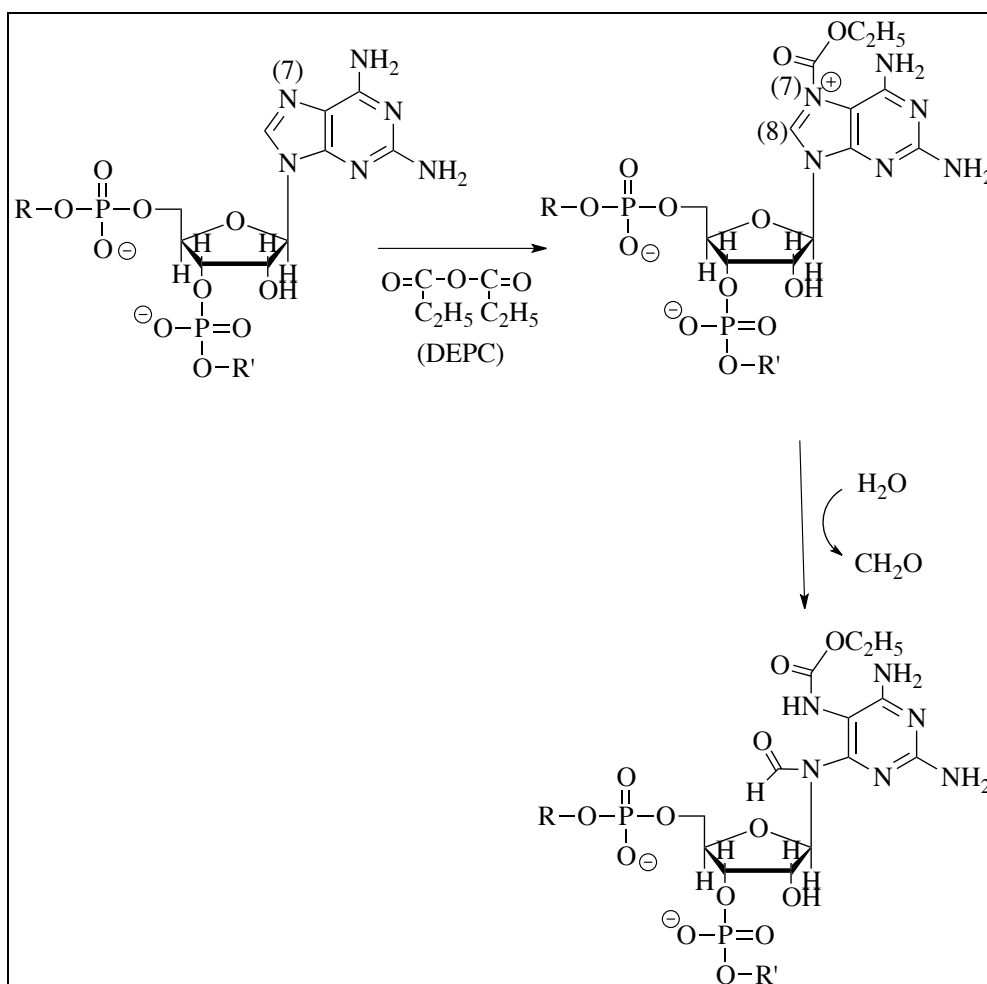


Figure 22: Représentation de la réaction du DEPC sur la position N7-A, et l'ouverture du cycle d'imidazole entre les positions N7 et C8. (Adaptée¹⁴⁶)

Le CMCT : Réagit principalement avec les positions N3-U et N1-G à pH= 8.0, la position N3-U étant plus réactive que la position N1-G (Fig.23). La réaction ne se produit pas à pH neutre ou acide. Le CMCT est utilisé pour localiser les uridines et les guanosines non appariées, les thymidines réagissent aussi mais faiblement. Le CMCT a été utilisé par exemple pour étudier la structure de l'ARNr 16S dans les sous-unités ribosomiques 30S, en combinaison avec du DMS et du kethoxal¹⁵⁹, et pour analyser la structure du domaine 5' de l'ARNr 18S extrait des cellules Hela en combinaison avec du DMS, kethoxal, DEPC et des deux nucléases T1 et V1¹⁶⁰.

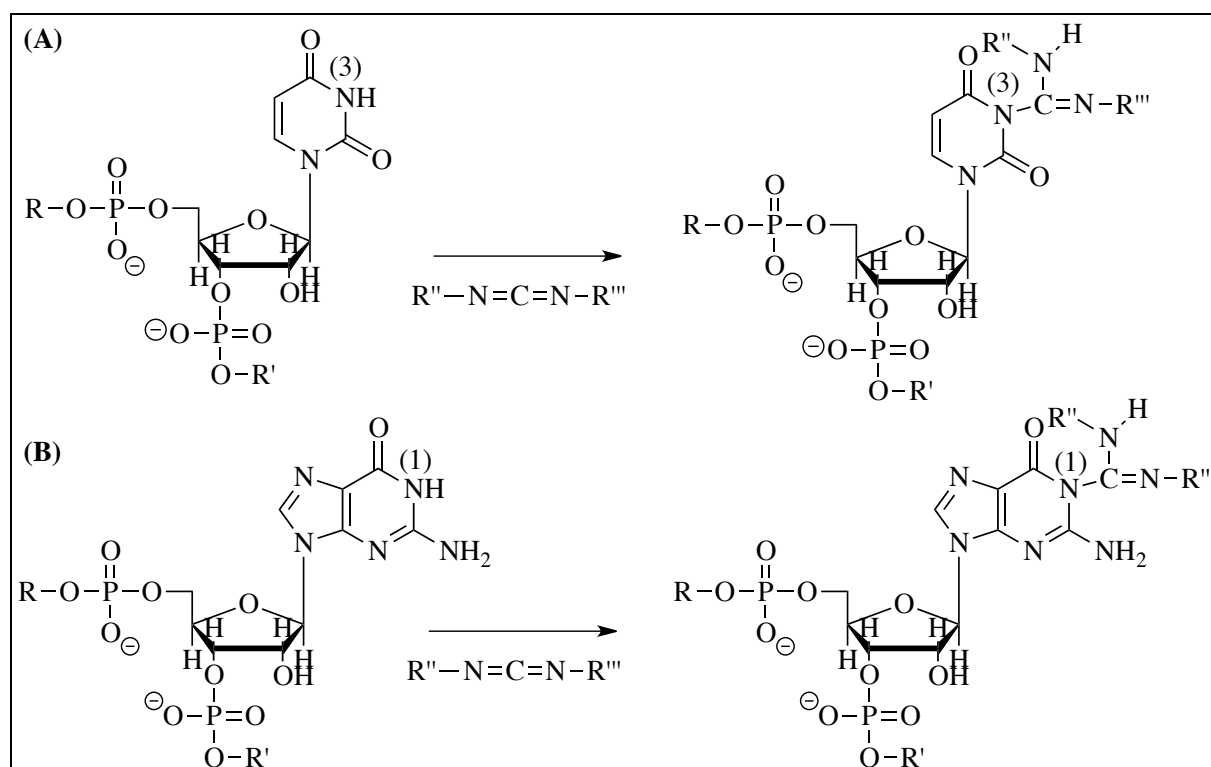


Figure 23: Représentation de la réaction du CMCT sur les deux positions N3-U (A) et N1-G (B). (Adaptée¹⁴⁶)

Le kethoxal : Un dérivé de glyoxal, réagit spécifiquement avec les guanosines non-appariées pour former un nouveau cycle impliquant les positions N1 et N2 du cycle de la guanosine, et les deux fonctions carbonyles du kethoxal (Fig.24). Le complexe formé kethoxal-guanosine est stable dans des conditions légèrement acides, mais se décompose en milieu basique. La présence d'ions borate stabilise le complexe. Il a été utilisé pour étudier la structure de l'ARNr 16S de *E.Coli*¹⁶¹ et pour élaborer un modèle 3D de l'ARNr 5S de *E.Coli* en combinaison avec le DMS, le DEPC, les deux RNase V1 et T1 et de la modélisation¹⁶².

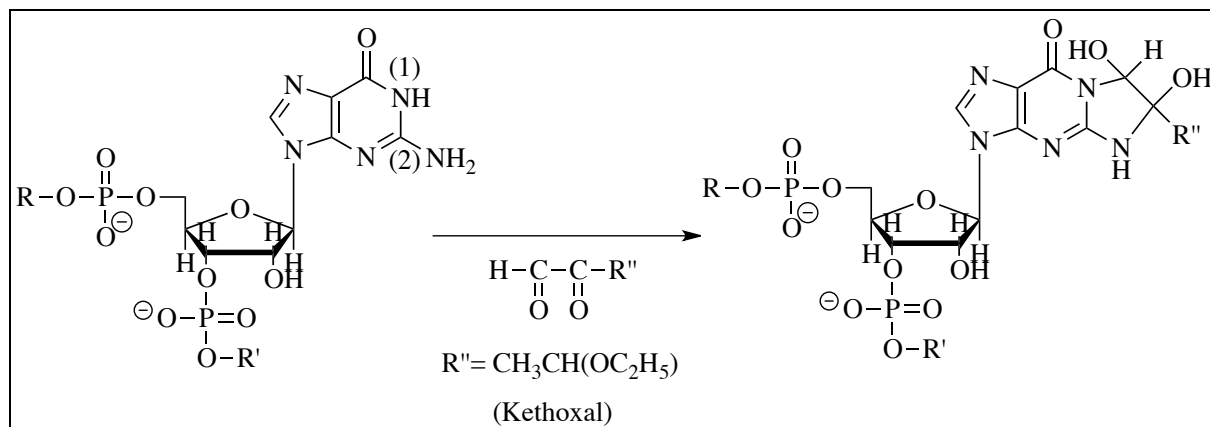


Figure 24: Représentation de la réaction du kethoxal sur les positions N1 et N2 de la guanosine pour former un nouveau cycle. (Adaptée¹⁴⁶)

Le Bisulfite : La modification avec du bisulfite convertit la cytosine en uridine¹⁶³. Le bisulfite réagit avec la cytosine pour former du 5,6-dihydrocytosine-6-sulfonate (Fig.25). À une forte concentration de bisulfite (>1M) et à pH acide (pH 5-6), la formation du dérivé de cytosine facilite la substitution nucléophile sur le groupement amine du cycle, permettant ainsi la conversion du 5,6-dihydrocytosine-6-sulfonate en 5,6-dihydrouracyl-6-sulfonate. Il a été utilisé par exemple pour analyser la structure de l'ARNr 5S de *Spinacea oleracea* en combinaison avec la nucléase S1¹⁶⁴, et pour étudier la dynamique structurale de l'ARNr 5S de *E.Coli*¹⁶⁵.

Le ENU (Ethylnitrosourea) : Est un réactif d'alkylation qui possède une grande affinité pour les oxygènes des groupements phosphates, contrairement aux autres réactifs chimiques tels que le DMS, qui sont des réactifs d'alkylation des groupements amines qui se trouvent sur les cycles aromatiques^{166,167}. Les phosphotriesters qui en résultent sont instables et se dissocient par traitement avec une solution alcaline. Ce réactif chimique réagit plus spécifiquement avec les phosphates qui ne forment pas des liaisons hydrogènes dans les interactions tertiaires et qui ne sont pas appliqués dans une coordination avec des ions cationiques (Fig.26). À noter que les phosphates qui se trouvent dans les hélices sont réactifs. L'ENU a été utilisé par exemple pour localiser les phosphates dans l'ARNr 16S de *E.Coli* appliqués dans des interactions tertiaires par des liaisons hydrogènes ou des coordinations avec des ions. Dans la sous-unité ribosomique 30S l'ENU a permis la localisation des phosphates en contact avec les protéines¹⁶⁸.

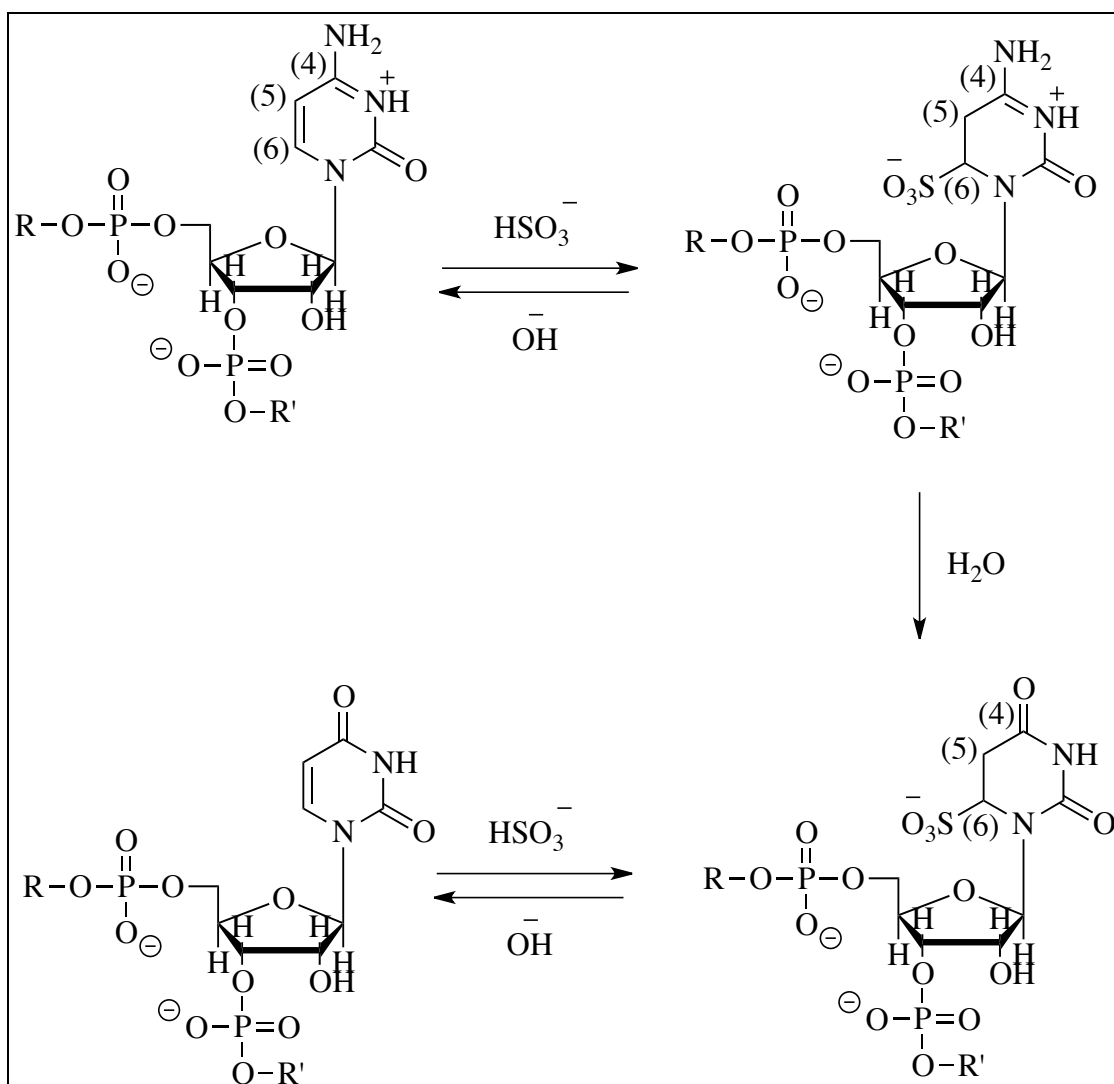


Figure 25: Représentation de la réaction du Bisulfite sur la cytosine qui la convertit en uridine. (Adaptée¹⁴⁶)

Les radicaux hydroxyls: Sont des groupements qui possèdent un électron non-apparié libre, qui peuvent être générés en solution à partir du peroxyde d'hydrogène, ou à partir de l'oxygène moléculaire en utilisant le catalyseur Fe(II)-EDTA ¹⁶⁹ (Eq.1). Ces radicaux hydroxyles peuvent être générés aussi par radiolyse de l'eau en utilisant une haute énergie de radiation allant de 100 eV à l'ordre des MeV, produisant des électrons et des radicaux hydroxyles libres¹⁷⁰ (Eq.2).

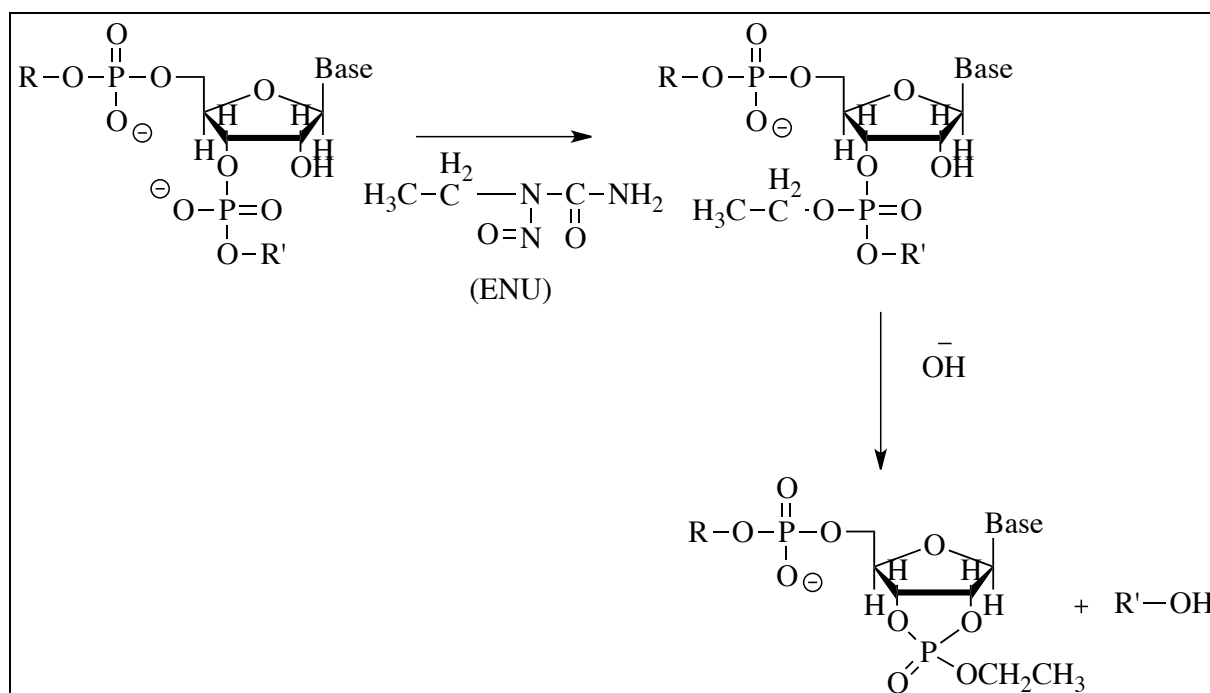
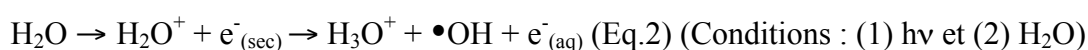
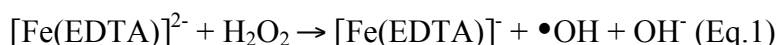


Figure 26: Représentation de la réaction de l'ENU sur les phosphates des riboses pour former des phosphotriesters. (Adaptée¹⁴⁶)

La deuxième méthode de génération de radicaux libres ne nécessite pas l'addition de réactifs chimiques à la solution et donc minimisant ainsi toute perturbation de l'équilibre ou de la transition interrogés.



Le radical libre réagit avec l'ARN en arrachant un proton en position C4' ou C5' du ribose, ou sur les deux positions en même temps, causant une scission de la molécule d'ARN. Les radicaux hydroxyles sont extrêmement réactifs et réagissent avec tous les nucléotides sans discrimination entre les bases, la susceptibilité d'arracher un proton à un nucléotide dépend de son accessibilité au solvant^{169,171,172}. C'est une méthode efficace pour la détermination des structures tertiaires d'ARN et peut être appliquée pendant un temps réduit de quelques millisecondes. Elle a été utilisée par exemple pour étudier le repliement des ARN tels que les ribozymes synthétiques et les ribozymes du groupe I et II des levures, de l'*Azoarcus cyanobacterium* et de la *Tétrahymena thermophila*¹⁶⁹ et pour étudier le repliement du domaine 5' de l'ARNr16S qui est indépendant des protéines¹⁷³.

Les réactifs chimiques bi-fonctionnels: Modifient l'ARN dans une région bien définie et ils sont capables de réagir avec d'autres régions proches sur l'ARN. Le but de cette méthode est d'obtenir des informations à travers l'espace et des contraintes sur des interactions à longue distance. On peut citer la méthode des ribonucléosides modifiés avec une fonction alkylthiol et la formation d'une liaison disulfide intramoléculaire dans l'ARN par oxydation à l'air. La liaison disulfide se forme entre deux fonctions thiols à proximité, dans une structure secondaire ou tertiaire¹⁷⁴. Cette méthode est utilisée pour marquer les ARN en solution, pour stabiliser l'empilement des ARN et pour étudier l'empilement des bases dans des structures tertiaires.

Les réactifs de SHAPE: Les réactifs de SHAPE (Selective 2'-hydroxyl Acylation analyzed by Primer Extension): BzCN, NMIA et 1M7 modifient l'ARN sur l'hydroxyle en position 2' du ribose, et indépendamment de la nature de la base^{175,176,177,178,179,180}. SHAPE est une méthode élaborée par l'équipe de Kevin Weeks. Elle présente l'avantage de réagir avec l'hydroxyle en position 2' du ribose en se basant sur le principe de flexibilité du nucléotide, et détermine l'état de chaque nucléotide pour n'importe quelle séquence d'ARN quelles que soit les conditions^{181,182,183,184,185,186,178,187,188,189}. Pour les nucléotides appariés ou formant des interactions tertiaires, le phosphodiester en position 3' du ribose et d'autres interactions réduisent la réactivité du groupement 2'-hydroxyle vis-à-vis du réactif de SHAPE^{190,191,175}.

Le principe de SHAPE est basé sur des réactifs électrophiles dérivés de la famille des anhydrides isotoïques (NMIA et 1M7). La réactivité de ces composés avec l'ARN peut être modifiée en variant la nature du substituant sur le cycle benzène. En effet, la fonction carbonyle en position 4 du cycle aromatique est sensible aux substituants en position méta et para. Ces réactifs possèdent alors chacun un temps de demi-vie différent l'un de l'autre. Le N-méthylisatoïc anhydride (NMIA) possède un temps de demi-vie élevé (20 minutes) qui permet l'entrée du réactif à l'intérieur de la cellule et la modification de l'ARN *in vivo* et donc l'étude de processus biologiques cellulaires. La réactivité de NMIA est fortement dépendante de la présence d'ions Mg^{2+} en solution¹⁸⁵. Il a été montré que la nucléophilicité du groupement 2'-OH du ribose est influencée à de fortes concentrations d'ions Mg^{2+} . La caractérisation d'un ARN riboswitch « M-Box » a mis en évidence que la présence d'ions Mg^{2+} induit une conformation tertiaire plus compacte puisqu'ils occupent des sites de liaisons spécifiques favorisant ainsi un positionnement régulier des groupements fonctionnels qui stabilise la structure tertiaire et régulant ainsi l'accessibilité des nucléotides¹⁹². Le NMIA réagit avec

l'hydroxyle en position 2' du ribose en libérant du CO_2 et en formant l'ester 2-méthylaminobenzoïque du nucléotide.

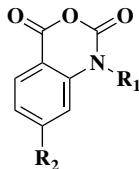
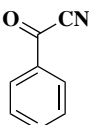
Le 1-méthyl-7-nitroisatoïc anhydride (1M7) est plus réactif que le NMIA avec un temps de demi-vie de $\sim 1 \text{ min}$ ¹⁷⁶. La substitution avec un groupement fortement électro-attracteur (NO_2) en position 7 du cycle aromatique rend le composé chimique plus réactif avec l'ARN, et diminue son temps de demi-vie en favorisant l'hydrolyse par l'augmentation du caractère électrophile du groupement carbonyle et en stabilisant la charge négative dans le composé intermédiaire de la réaction. Une dégradation complète de ce réactif est achevée dans 70 secondes, donc il est facile à utiliser dans les laboratoires et peut être utilisé pour étudier des processus biologiques courts.

La compréhension plus précise des phénomènes de repliement de l'ARN, de conversion entre différents états ou l'étude des fonctions des ARN dans les complexes ribonucléoprotéiques nécessite l'utilisation d'un réactif qui réagit encore plus rapidement que les dérivés de l'anhydride isatoïque. Pour cela, un autre réactif a été recherché pour réagir avec l'ARN dans une échelle de temps de quelques secondes. Le cyanure de benzoyle (BzCN) répond à ces critères puisqu'il réagit avec l'hydroxyle en position 2' du ribose en 0,25 secondes en formant un ester stable et en libérant l'ion cyanure au cours d'une réaction irréversible¹⁷⁷.

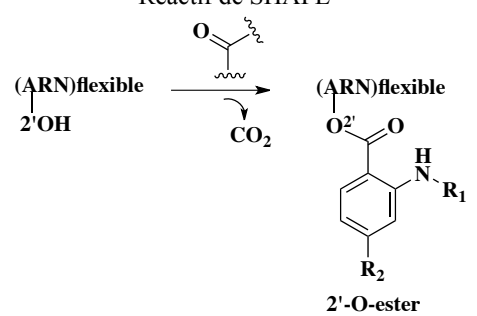
Des études approfondies avec les réactifs NMIA et 1M7 ont montré que la nature du nucléotide influence légèrement la réactivité du groupement 2'-OH du ribose. Ainsi, la réactivité des deux nucléotides G et A sont identiques mais supérieur de 1,4 fois de celle du nucléotide U et de 1,7 fois de celle du nucléotide C. La différence de réactivité entre purine et pyrimidine peut être expliquée du fait d'une différence de pKa entre les OH-2' de ces nucléotides. Le pKa des pyrimidines est plus élevé (pKa=12,8) que celui de l'adénosine (pKa= 12,5) et Guanosine (pKa= 12,7). La formation du 2'-O-ester nécessite d'arracher le proton en position 2'-OH il est donc légèrement plus facile d'arracher ce proton aux purines qui ont une valeur de pKa légèrement plus faibles¹⁹³.

Tous les réactifs de SHAPE subissent une inactivation suite à l'hydrolyse par l'eau présente dans le milieu. Cette réaction secondaire présente deux avantages : i) il n'est pas nécessaire d'ajouter une étape d'inhibition du réactif à la fin de la réaction puisque le composé réagit soit avec l'ARN soit il est inactivé par l'eau ii) le temps de demi-vie du réactif de SHAPE est déterminé par l'hydrolyse avec l'eau dont la concentration (55 M) est beaucoup plus importante que celle de l'ARN. La modification de l'ARN avec les réactifs de SHAPE est

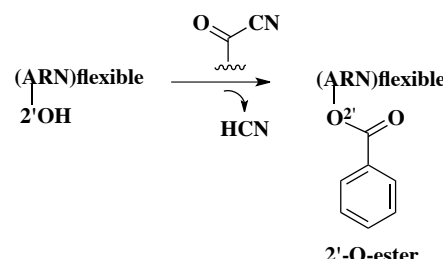
donc très peu sensible à la concentration en ARN, ni aux autres composants de la réaction tels que le tampon, les ligands ou les protéines

Réactifs de SHAPE	R ₁	R ₂	t _{1/2} Hydrolyse (s)	Modification de l'ARN par les réactifs de SHAPE
	IA	H	H	430
	NMIA	CH ₃	H	260
	1M7	CH ₃	NO ₂	14
	4NIA	H	NO ₂	101
	BzCN	—	—	0,25

Réactif de SHAPE



2'-O-ester



2'-O-ester

Tableau 1 : Les différents réactifs de « SHAPE » avec leurs temps de demi-vie. La formation de l'ester en position 2' du ribose par attaque nucléophile de l'hydroxyle en position 2' sur la fonction carbonyle de ces réactifs. (Adapté¹⁹⁴)

La variété des réactifs chimiques disponibles pour étudier la structure d'un ARN est très grande mais si l'on veut étudier un ARN entier, il est nécessaire de réaliser des expériences avec plusieurs réactifs chimiques qui fournissent des informations complémentaires.

I.3.3 Les techniques de détection à haut débit

En principe, les résultats obtenus par la modification chimique de l'ARN et après élongation de l'amorce en présence de la rétro-transcriptase sont étudiés sur des gels d'électrophorèse. Après migration sur gels, les intensités des bandes sont quantifiées individuellement et manuellement, ce qui nécessite beaucoup de temps. Afin de faciliter la quantification le logiciel SAFA¹⁹⁵ « Semi-Automated Footprinting Analysis » a été développé. Il permet d'analyser en 20 à 60 minutes, un gel contenant jusqu'à une centaine de mesure de réactivité.

I.3.3.1 Séquenceur d'ADN et électrophorèse sur capillaire (hSHAPE)

Le séquenceur d'ADN à électrophorèse sur capillaire possède la capacité de réaliser plusieurs expériences en un temps réduit. Ceci est possible suite à la présence d'un faisceau de capillaires qui peut contenir jusqu'à 48 capillaires et le temps d'une électrophorèse qui est relativement court de 30 minutes seulement. Le séquenceur possède un ion laser qui excite les sondes fluorescentes utilisées pour marquer les amorces d'ADN à la longueur d'onde de 488 nm. Ce séquenceur est relié à un ordinateur afin de collecter les informations à analyser. Le principe du séquenceur d'ADN est illustré dans la figure ci-dessous (Fig.27) :

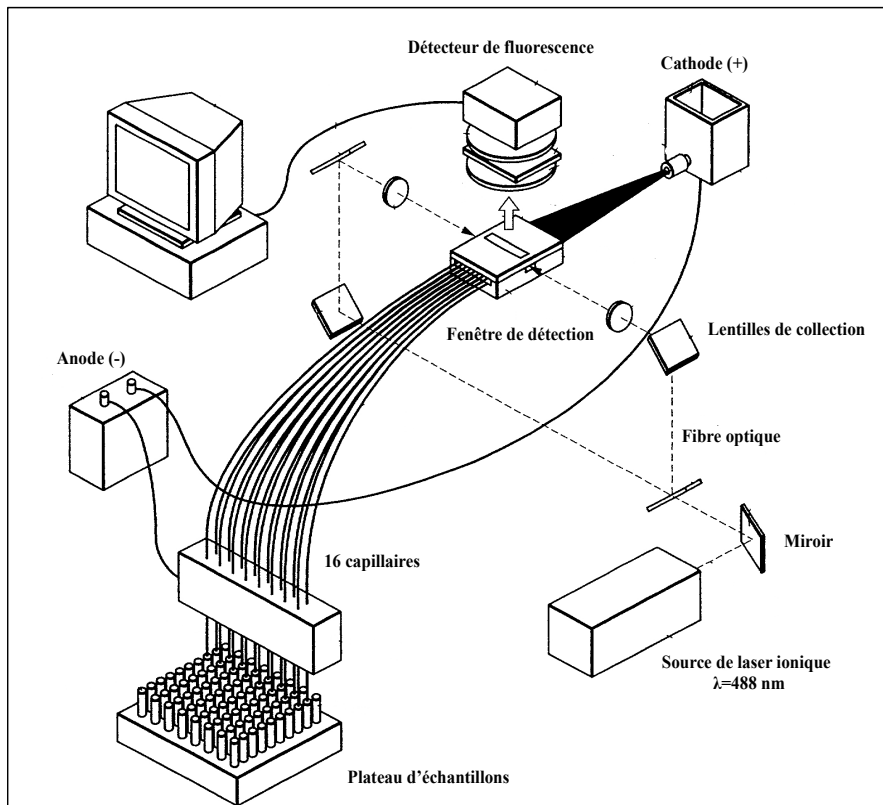


Figure 27: Modèle de séquenceur possédant 16 capillaires reliant la plaque contenant les échantillons marqués par les différentes sondes fluorescentes à un détecteur de fluorescence. Ce séquenceur possède un seul laser excitant les sondes fluorescentes à 488nm.

La méthode de hSHAPE permet d'étudier une séquence d'ARN allant de 400 à 650 nucléotides. La modification de l'ARN par les réactifs de SHAPE est statistique, donc une modification tous les 300 nucléotides. Pour effectuer une expérience de SHAPE, la nécessité d'utiliser une grande quantité d'ARN allant jusqu'à 8 pmoles qui est une quantité considérable dans le cas des ARN viraux.

Suite à la modification chimique de l'ARN, une étape de rétro-transcription est réalisée. Pour cela, les ARN sont hybridés à des amorces marquées sur leurs extrémités 5' par des sondes fluorescentes. Les sondes fluorescentes commerciales utilisées sont de quatre couleurs différentes correspondant chacune à une « sous-réaction » de l'étape enzymatique.

- La première sous-réaction consiste à hybrider une amorce d'ADN marquée par une sonde fluorescente sur l'ARN modifié avec l'un des réactifs de SHAPE, et de réaliser l'extension en présence des quatre dNTP (dATP, dCTP, dGTP et dTTP). La longueur et la quantité des fragments d'ADNc marqués sur leurs extrémités 5' ainsi obtenus, correspondent à la position et au degré de modification de chaque nucléotide.
- La deuxième sous-réaction consiste à hybrider une amorce d'ADN marquée par une autre sonde fluorescente sur l'ARN non modifié, et de réaliser l'extension de l'amorce en présence des quatre dNTP. Cette sous-réaction est un contrôle négatif pour pouvoir identifier les pauses naturelles de la rétro-transcriptase sur l'ARN, ou pour détecter une hydrolyse endogène de l'ARN.
- Les deux dernières sous-réactions sont des étapes de séquençage classiques de l'ARN non modifié en présence de quatre dNTP et d'un ddNTP (différent pour chaque sous-étape) qui permettent de faire correspondre les sites de modification et la séquence primaire de l'ARN.

Ces ADNc sont ensuite collectés, précipités et injectés dans un capillaire pour être séparés sur un séquenceur d'ADN lié à un ordinateur. Ce dernier permet de collecter les données de séquençage^{196,197} qui vont être ensuite analysées par un logiciel informatique (Fig.28).

Deux logiciels ont été développés pour étudier la structure de l'ARN en utilisant l'électrophorèse par capillaire: CAFA « Capillary Automated Footprinting Analysis » et ShapeFinder. CAFA^{198,199} est un logiciel qui est facile à mettre en place, nécessitant l'utilisation d'une seule amorce marquée par une sonde fluorescente pour l'étape de l'élongation par la rétro-transcriptase. Il présente un bon alignement pour des séquences de taille allant jusqu'à 250 nucléotides. Par contre, ShapeFinder²⁰⁰ nécessite l'utilisation d'au moins deux amorces marquées par deux sondes fluorescentes différentes, la mise en place et l'alignement de ce logiciel demandent beaucoup d'effort, mais il permet d'analyser de nouvelles et plus longues séquences. Il peut lire des séquences moyennes allant de 400 à 650 nucléotides par expérience et par capillaire.

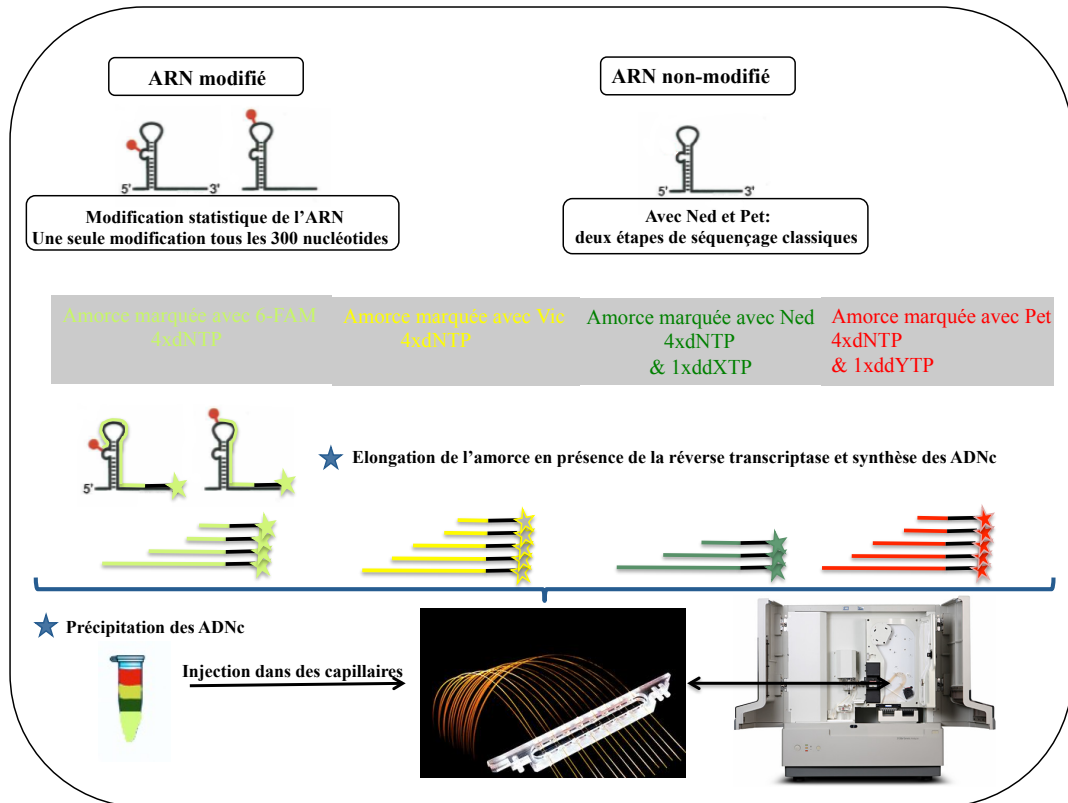


Figure 28: Les différentes étapes de la technique de hSHAPE et la séparation des ADNc par électrophorèse sur capillaire sur un séquenceur d'ADN. L'amorce d'ADN est marquée avec un ensemble de sondes fluorescentes G5 (6-Fam, Vic, Ned et Pet). L'ARN est modifié avec l'un des réactifs de « SHAPE ».

Le résultat d'une expérience de hSHAPE est représenté sous forme de quatre électrophérogrammes qui présentent l'intensité de fluorescence de chaque nucléotide pour les quatre sous-étapes de hSHAPE en fonction du temps d'élution (Fig.29.a). ShapeFinder convertit cette intensité de fluorescence en réactivité pour chaque nucléotide (Fig.29.b). La différence d'intensité pour chaque position entre l'ARN modifié (+) et l'ARN non modifié (-) est importante du fait qu'elle révèle la flexibilité du groupement hydroxyle en position 2' de chaque nucléotide.

Le logiciel ShapeFinder est doté d'une option de correction du signal et du profil de fluorescence. En effet, l'utilisation de quatre sondes fluorescentes possédant des propriétés physico-chimiques différentes l'une de l'autre provoque un décalage de la mobilité et les électrophérogrammes doivent être ré-alignés sur l'axe des x. De plus, l'analyse de toutes les traces dans une expérience de SHAPE indique que les intensités des pics s'affaiblissent en avançant dans la lecture de la séquence d'ARN. Il existe trois sources pour cet affaiblissement : i) La rétro-transcriptase n'est pas parfaitement performante et effectue des pauses même si le nucléotide ne possède pas la modification sur l'hydroxyle en position 2' ii)

La modification sur l'ARN est normalement une modification statistique donc une modification tous les 300 nucléotides, mais cela n'exclut pas la possibilité que l'ARN réagisse deux ou trois fois avec le réactif chimique dans différents sites, dans ce cas juste la première modification est détectée et les ADNc de tailles courtes sont sur-représentés. iii) Pour les étapes de séquençage classique, à chaque fois qu'un didésoxynucléotide est incorporé, il y a observation d'une diminution de l'élongation d'amorce d'une façon exponentielle jusqu'au facteur zéro pour les nucléotides de la zone entre 450 à 600. Il est donc nécessaire de réaliser une correction du signal sur l'axe des y.

Dans un second temps, les pics doivent être alignés avec la séquence de l'ARN et tous les pics correspondant à l'ARN modifié (+) ou à l'ARN non-modifié (-) sont identifiés et quantifiés en utilisant l'intégration Gaussienne. L'absence de pic dans la trace (+) indique que le nucléotide en question est apparié ou entre dans une interaction tertiaire. Donc une identification précise et une analyse quantitative de toutes les positions pour lesquelles les pics sont minimaux ou absents sont importantes.

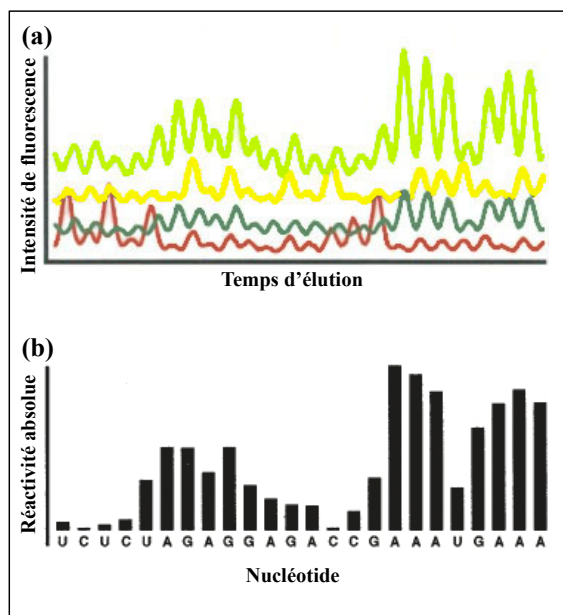


Figure 29: (a) Les 4 électrophérogrammes obtenus après l'expérience d'électrophorèse sur capillaire. Chaque graphe représente l'intensité de fluorescence en fonction du temps d'éluion. (b) Représentation des réactivités absolues obtenues pour chaque position de nucléotide sur la séquence d'ARN après intégration des pics et traitement par ShapeFinder. (Adaptée²⁰⁰)

La technique de hSHAPE présente différents avantages par rapport à d'autres méthodes^{175,201,202} :

- La possibilité de faire plusieurs expériences en un temps réduit suite à la présence de plusieurs capillaires dans le séquenceur d'ADN.
- Possibilité de lire des séquences à taille moyenne allant jusqu'à 650 nucléotides.
- Obtention d'informations quantitatives sur la réactivité de la majorité des nucléotides.

- Les réactivités sont quantifiées automatiquement par le logiciel Shapefinder.
- Pas d'utilisation de la radioactivité puisque les données sont analysées sur un séquenceur d'ADN.
- Méthode applicable aux ARN transcrits *in vitro*, aux ARN cellulaires et aux ARN viraux.
- Les réactifs de SHAPE présentent différents temps de demi-vie permettant ainsi d'explorer différents processus cellulaires.

I.3.3.2 SHAPE-Sequencing ou « SHAPE-Seq »

SHAPE-Seq²⁰³ associe la chimie de la technique de SHAPE « Selective 2'-Hydroxyl Acylation analyzed by Primer Extension » avec le séquençage à haut débit de multiples séquences d'ADN simple brin synthétisées et hybridées aux deux extrémités avec deux amorces. Cette méthode permet d'analyser la structure d'une centaine de molécules d'ARN de séquence arbitraire. La résolution se fait à l'échelle du nucléotide pour la structure secondaire ainsi que tertiaire et la technique donne des informations quantitatives²⁰⁴.

Le SHAPE-Seq consiste à fixer un code barre sur chaque ARN isolé puis à réaliser la modification chimique par un réactif acylant les 2'-OH des nucléotides non-appariés. Comme pour la technique de hSHAPE, un échantillon d'ARN, l'échantillon (+) est traité avec le réactif de SHAPE et le second, échantillon (-), est traité avec le solvant en absence de réactif. Les positions modifiées sont ensuite indirectement détectées grâce à leur capacité à bloquer la réaction de rétro-transcription et au séquençage des ADNc ainsi obtenus. L'amorce hybridée sur une séquence spécifique de l'ARN est taguée une fois avec une séquence adaptatrice avant la rétro-transcription, ensuite l'ADNc obtenu après hydrolyse de l'ARN « matrice » est hybridé avec une deuxième séquence adaptatrice. Ces deux séquences adaptatrices permettent d'augmenter le nombre d'amplification afin d'obtenir plus d'informations (Les plaques Illumina servent à augmenter le nombre d'amplifications et à réduire ainsi la quantité d'ARN nécessaire pour le séquençage) (Fig.30).

Le SHAPE-Seq peut être appliqué à l'étude de mutants multiples d'un ARN ou à celle de plusieurs membres proches issues d'une même famille d'ARN.

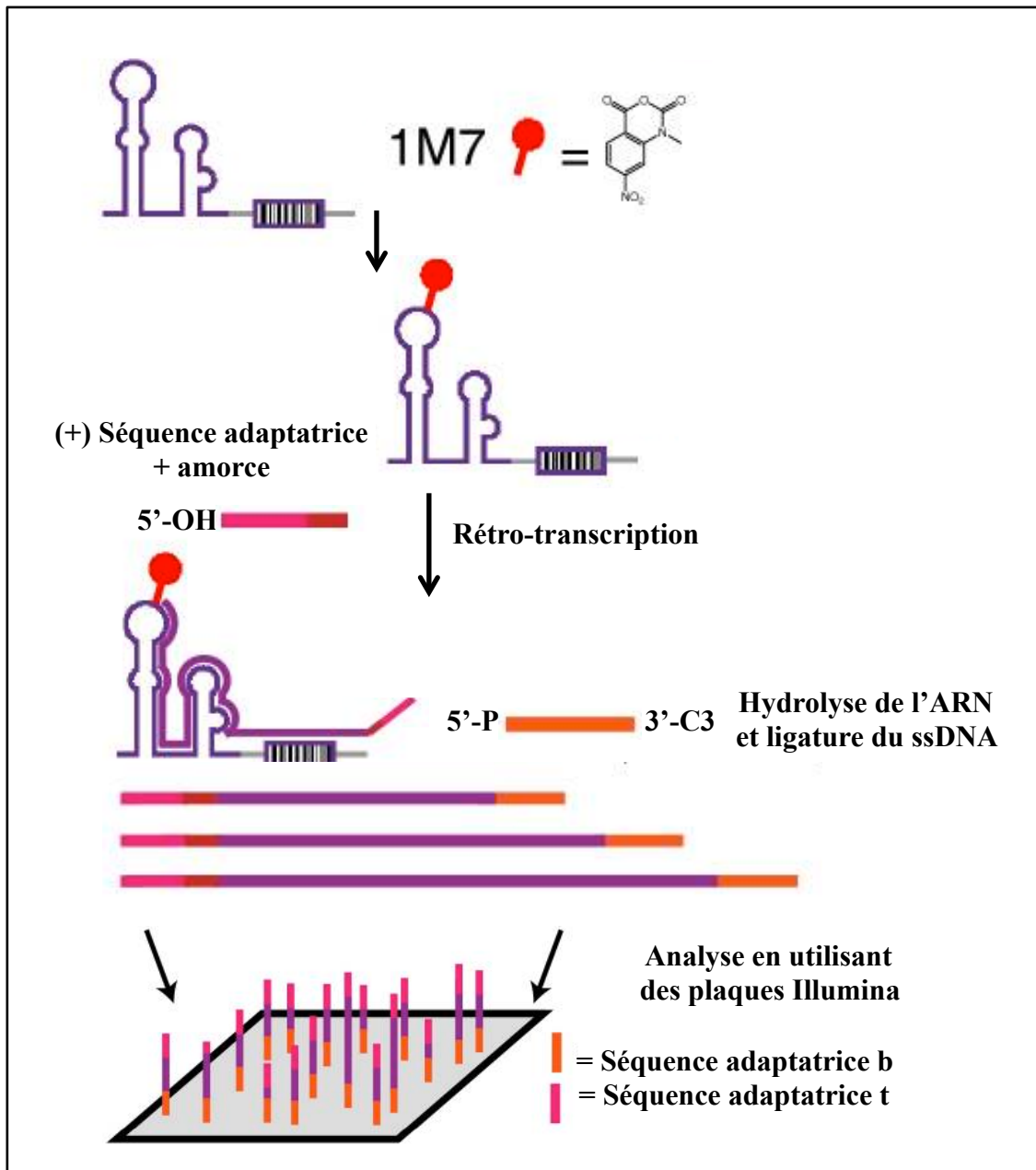


Figure 30: Représentation schématique des différentes étapes de la technique de SHAPE-Seq. Le schéma représente le contrôle positif. (Adaptée²⁰³).

Une fois le séquençage terminé, les données digitales obtenues sont complètement et automatiquement analysées mathématiquement, contrairement à la technique de SHAPE-CE « SHAPE-Capillary Electrophoresis » pour laquelle les données sont quantifiées après une série de traitements manuels (correction de mobilité, d'alignement et d'intégration des réactivités obtenues). Le modèle utilise un nombre maximal d'estimations afin de délivrer un ensemble de réactivités et d'estimer le nombre de modifications par site.

SHAPE-Seq a été développé pour combiner la bonne résolution obtenue avec les réactifs de SHAPE, la quantification, le débit et les capacités multiples d'un séquençage à haut débit d'une séquence doublement taguée sur les deux extrémités. Les avantages de cette technique sont les suivants:

- Le séquençage à haut débit rend cette technique plus sensible en utilisant une petite quantité d'ARN (0,1 pmoles). Cette propriété de la technique de SHAPE-Seq la rend particulièrement utile pour les ARN difficiles à extraire des cellules ou qui sont rares tels les ARN des virus qui se trouvent en petite quantité dans les virions. Par contre, cette technique est limitée à des ARN de taille de 400 nucléotides²⁰⁵.
- En comparaison avec les méthodes enzymatiques récentes et contrairement aux nucléases qui sont encombrantes, de petites molécules chimiques sont utilisées (les réactifs de SHAPE classiques).
- La méthode de SHAPE-Seq est utilisée à des températures et des conditions de tampons variables et n'est pas limitée par les conditions nécessaires pour le fonctionnement des enzymes.
- Le résultat d'une expérience de SHAPE-Seq est un ensemble de réactivités qui peuvent être intégrées directement en utilisant des algorithmes déjà existants pour l'étude de la structure d'ARN afin de guider le repliement de ce dernier.

I.3.3.3 RNA-Sequencing ou « RNA-Seq »

Une méthode utilisée « *in vitro* » pour étudier la composition d'un génome ou découvrir de nouveaux exons ou gènes par séquençage haut débit des ADNc.

Pour cette technique, les ARN sont tout d'abord poly-adénylés à leur extrémité 3' de manière à faire en sorte que l'ARN codant soit séparé de l'ARN non codant en réalisant deux tours de sélection de séquences d'ARN poly(A) par hybridation avec des billes d'oligonucléotides (dT). L'ARNm obtenu subit ensuite une fragmentation contrôlée par hydrolyse catalysée par le magnésium avec chauffage à 94°C pendant quelques minutes. Cette étape est nécessaire pour avoir une uniformité durant le séquençage parmi les transcrits. Elle est aussi nécessaire pour éviter d'avoir des sites fortement favorisés que d'autres pour les sites de liaisons aux amorces²⁰⁶.

Les fragments d'ARN de 200 à 300 nucléotides obtenus après la fragmentation sont ensuite rétro-transcrits en ADNc. À ces ADNc seront ensuite rajoutées des séquences adaptatrices

pour le séquençage. Ensuite, l'alignement des lectures par rapport au génome de référence est réalisé afin de calculer la fréquence d'un transcrit, en utilisant un logiciel bioinformatique de la famille ERANGE (Enhanced Read Analysis of Gene Expression). Le résultat est un graphe qui résume la quantité de lectures (millions de lectures de 25pb) qui couvrent chaque nucléotide²⁰⁶ (Fig.31).

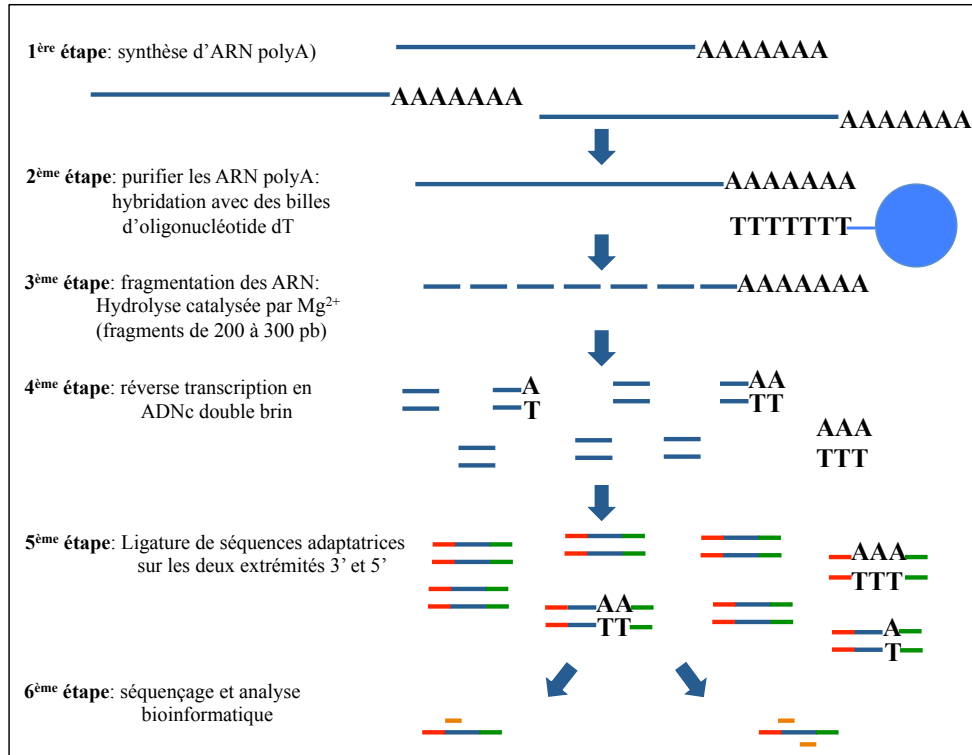


Figure 31: Représentation schématique des différentes étapes de la technique de RNA-Seq. (Adaptée^{206,207,208})

I.3.3.4 Parallel Analysis of RNA Structure ou « PARS »

PARS est une technique utilisée *in vitro* qui consiste à cliver les molécules d'ARN soit par la RNase V1 qui est spécifique des régions où l'ARN forme des structures double brin, soit par la nucléase S1, qui elle clive spécifiquement les structures d'ARN simple brin²⁰⁹. Ces deux enzymes présentent la particularité de couper la liaison phosphate des ARN du côté 3' en laissant un groupement phosphate en position 5'. Après une étape de fragmentation aléatoire (qui forme des fragments d'ARN de 200 nucléotides présentant une extrémité 5'-OH), un adaptateur spécifique des ARN terminés par un 5'-phosphate est hybridé aux fragments générés par la RNase V1 ou la nucléase S1, ce qui permet de les différencier des fragments d'ARN qui sont issus de la fragmentation ou de la dégradation. Ensuite, une étape de rétro-

transcription à partir d'une amorce spécifique de l'adaptateur permet d'obtenir des ADNc qui sont soumis au séquençage haut débit. Le recensement du nombre de séquences qui se terminent à une position donnée, permet de déterminer les nucléotides appariés (expérience avec la RNase V1) ou non-appariés (expérience avec la nucléase S1). Ces données sont traitées par des algorithmes tel que le SeqFold afin de prédire une structure secondaire pour l'ARN d'intérêt²¹⁰.

*Kertesz et al.*²¹¹, ont appliqué la technique de PARS à l'étude de la structure secondaire de l'ARN messager de « *Saccharomyces cerevisiae* ». Ils ont ainsi obtenu le profil de la structure d'environ 3000 transcrits distincts. L'analyse de ces profils a révélée plusieurs propriétés structurales des transcrits de levure.

Une des raisons les plus importantes pour le choix de ces nucléases dans PARS, c'est que ces enzymes clivent l'ARN en laissant des extrémités (5'P) et (3'OH), qui permettent de différencier les ARN clivés par les enzymes des autres ARN présents dans le milieu^{212,135,213}. Afin de confirmer que le (5'P) cloné dans cette procédure est spécifiquement généré par un clivage enzymatique et non pas d'un clivage endogène déjà existant dans la cellule, en plus de la librairie obtenue par la technique de PARS, une autre librairie de fragments d'ARN est obtenue en absence des nucléases²¹¹. Un plus grand nombre de cycle d'amplification par PCR est nécessaire pour générer du matériel d'ADN suffisant en absence des enzymes. Cette évidence montre que le nombre de (5'P) endogènes obtenus par simple hydrolyse est très petit et que la majorité des sites de clivage (5'P) proviennent essentiellement d'un clivage enzymatique.

D'autres méthodes enzymatiques ou chimiques identifient les paires de base par clivage de l'ARN en laissant une extrémité 5'OH²¹⁴. Cette fonction qui nécessite du traitement avec une kinase qui est la T4 PolyNucleotide kinase (PNK) en présence de l'ATP pour générer du (5'P)^{215,216}, avant d'hybrider la séquence adaptatrice. L'addition du phosphate peut introduire un bruit de fond dans les données de séquençage, parce que les fragments d'ARN dégradés présents dans la cellule sont aussi collectés et séquencés. Une façon de distinguer entre le clivage endogène et le clivage enzymatique, est de modifier les groupements (5'OH) en absence du clivage par les nucléases. Cette approche implique l'utilisation de la T4 PNK pour convertir (5'OH) en (5'P) avant la fragmentation et la construction de la librairie d'ADNc. Ceci peut alors servir de contrôle négatif qui permet de soustraire les signaux de bruit de fond de la structure d'ARN modifiée.

La technique de PARS a été modifiée pour étudier aussi les modifications dans la structure d'ARN en utilisant un gradient de température : PARTE (Parallel Analysis of RNA structure with Temperature Elevation) chez les levures. Le but de ce gradient de température est (i) de voir s'il existe des régions avec différents degrés de stabilité, (ii) de déterminer la localisation de ces régions, (iii) et de voir l'effet des différences de stabilité d'une structure sur sa fonction.

En plus des levures, la technique de PARS peut être appliquée *in vitro* sur d'autres ARN isolés provenant d'autres organismes, y-compris ceux qui ont des génomes plus larges et plus complexes. Elle pourra être utilisée aussi pour étudier le changement structural de l'ARN en variant différentes conditions en solution. Contrairement, l'exécution de cette technique sur des organismes dont le génome est répétitif est difficile, à cause de la difficulté de cartographier chaque ARN individuellement et avec précision.

I.3.3.5 Fragmentation Sequencing « FragSeq »

Fragmentation Sequencing (Frag-Seq)²¹⁷ est une technique couplée au séquençage haut débit à partir des fragments d'ARNnc clivés par la nucléase P1 spécifique des bases non-appariées, pour donner des fragments d'ARN avec des extrémités 5'-monophosphate et 3'-hydroxyl libres¹⁴⁰. Les données sont ensuite analysées en utilisant des logiciels bioinformatiques pour déduire les sites de coupure sur l'ARN, donc qui offre un profil sur l'accessibilité de l'ARN. Cette technique est appliquée *in vitro*, pour étudier le génome entier de la souris, elle a été testée sur deux types de cellules soient des ARN provenant de cellules souches non différenciées des embryons de souris, soient des ARN de cellules différenciées en précurseurs neuronaux pour démontrer sa reproductibilité²¹⁷.

Pour cette technique, des ARN de 70 à 300 nucléotides ont été traités par la nucléase P1 pour donner des fragments d'ARN de 20 à 100 nucléotides. Le contrôle est fait sur des ARN en absence de la nucléase pour permettre d'estimer le nombre de fragments d'ARN possédant une extrémité 5'-monophosphate obtenus par une hydrolyse endogène. En parallèle aussi, une masse adéquate d'ARN de 20 à 100 nucléotides est traitée avec une kinase polynucléotide (PNK) et de l'ATP pour permettre la suppression de la phosphorylation en position 5' et la cyclisation en position 3' du phosphate²¹⁸. Cette procédure permet de contrôler les pauses endogènes qui ne libèrent pas un phosphate en position 5' ni un hydroxyle en position 3'.

Après isolation sur gel, les fragments d'ARN obtenus à partir des trois traitements sont hybridés à des séquences adaptatrices spécifiques pour chaque extrémité libre, conservant

ainsi l'orientation de chaque fragment. Donc deux séquences adaptatrices sont hybridées à chaque extrémité de l'ARN, une sur l'extrémité 3'-OH libre et une autre sur l'extrémité 5'-monophosphate. Après l'étape de la rétro-transcription, les librairies d'ADNc obtenues sont couplées à des codes barres puis les produits de la PCR sont rassemblées et séquencées en utilisant la technologie de séquençage SOLiD3 (Sequencing by Oligonucleotide Ligation and Detection) afin de générer des centaines de millions à des billions de séquences à la fois (ABISOLiD3). Ces ADNc sont ensuite comparés au génome de référence et analysés en utilisant un logiciel appelé ABI Small RNA Analysis Pipeline qui permet d'aligner les séquences, de les compter et de donner le résultat pour chaque catégorie d'ARN présente dans la librairie.

Des scores de coupures tout au long de chaque transcrit sont élaborés en utilisant le logiciel de FragSeq, pour les séquences d'ADNc obtenues à partir du contrôle ou de l'ARN modifié avec la nucléase P1. Ces scores correspondent à chaque fois au rapport entre le site de coupure généré en présence de l'enzyme et celui généré en absence de la nucléase P1. Ce score reflète la préférence pour la nucléase à couper dans un site spécifique par rapport à un autre site sur la même séquence d'ARN. C'est le logarithme d'un rapport de probabilité d'observer une pause dans l'ARN traité par la nucléase contre une pause endogène sans nucléase.

L'utilisation de la nucléase P1 permet d'éviter les produits obtenus par une coupure spontanée ou dégradation par la RNase, et permet aussi de fragmenter l'ARN d'une façon spécifique par rapport à la susceptibilité de la nucléase pour chaque site de coupure. Dans la technique de FragSeq, pas d'utilisation de deux nucléases comme c'est le cas de la technique de PARS par exemple, qui élabore un rapport de susceptibilité de chaque site de coupure sur l'ARN par rapport aux deux nucléases. Donc l'utilisation d'une seule nucléase permet d'élaborer un rapport de susceptibilité de la nucléase pour le site de coupure sur l'ARN contre un ARN en absence de nucléase, donc contre un site de coupure spontané ou hydrolyse endogène. La nucléase P1 coupe à l'extrémité 3' d'une base non-appariée, donc un score élevé obtenu dans un site spécifique indique que la base en amont de ce dernier est probablement peu impliquée dans un appariement de paires de bases, ou même dans des interactions tertiaires.

En plus, la nucléase P1 permet l'identification des régions protégées par une liaison spécifique ARN-protéine, et ce type de digestion pourrait être appliqué à des extraits nucléaires ou cellulaires liés à des protéines. La nucléase P1 est un bon candidat pour ce type

de digestion puisque le tampon utilisé pour l'extraction possède les mêmes caractéristiques (pH et composition en sels) que le tampon utilisé dans cette étude.

De plus, la nucléase P1 est stable à de hautes températures, d'où la possibilité de contrôler les domaines de dénaturation de l'ARN.

I.3.3.6 PARS contre Fragseq : comparaison entre les deux méthodes

- Deux méthodes enzymatiques de séquençage à haut débit utilisées *in vitro*, donc l'ARN doit être extrait des cellules et re-naturé.
- Les deux méthodes sont appliquées sur un mélange d'ARN. Les enzymes clivent l'ARN soit dans les régions appariées, soit dans les régions non-appariées.
- Dans la technique de PARS utilisation de deux nucléases S1 qui clive les ARN simples brins et V1 qui clive les ARN double brins. Tandis que dans la technique de Fragseq utilisation d'une seule nucléase P1 qui clive seulement les ARN simples brins.
- Dans les deux méthodes les enzymes clivent l'ARN en laissant une extrémité 5'-phosphate libre au lieu de l'extrémité 5'-OH obtenues par clivage endogène. L'extrémité 5' phosphate libre permet l'hybridation avec des séquences adaptatrices qui permettent de sélectionner les fragments d'ARN modifiés.
- Pour la technique de FragSeq, il n'existe pas d'étape de fragmentation aléatoire de l'ARN.
- Les étapes de rétro-transcription et de PCR sont nécessaires pour les deux techniques après la modification par les enzymes.
- Séquençage à haut débit des bibliothèques d'ADNc obtenues. L'analyse des résultats diffère d'une méthode à l'autre à cause d'une différence d'enzymes. Dans la méthode de PARS, le résultat obtenu est un logarithme d'un rapport entre le nombre de lectures des fragments obtenus après clivage avec l'enzyme V1 et S1. Dans la technique de Fragseq, utilisation d'une seule nucléase avec deux contrôles: le premier contrôle pour estimer le nombre de fragments laissant une extrémité 5'-phosphate obtenus par une hydrolyse endogène et le deuxième contrôle en présence de la kinase T4 pour estimer le nombre de fragments ne laissant pas une extrémité 5'-phosphate libre. Alors le résultat est un logarithme d'un rapport entre le nombre de lectures obtenues par clivage avec l'enzyme P1 et les deux contrôles.

- Dans la technique de PARS, le contrôle des hydrolyses endogènes est indirect en supposant qu'il résulte d'un traitement alcalin qui échappe alors à l'analyse du séquençage. Tandis que pour la technique de FragSeq, les hydrolyses endogènes sont considérées dans l'analyse directement ce qui est plus solide comme technique.
- L'avantage clair de la technique de PARS, c'est le développement d'un logiciel capable de convertir ces séquences obtenues en un format qui peut servir de données pour des logiciels de prédiction de structure secondaire et compatible avec les données expérimentales et les calculs de minimisation d'énergie²¹⁹. (Fig.32)

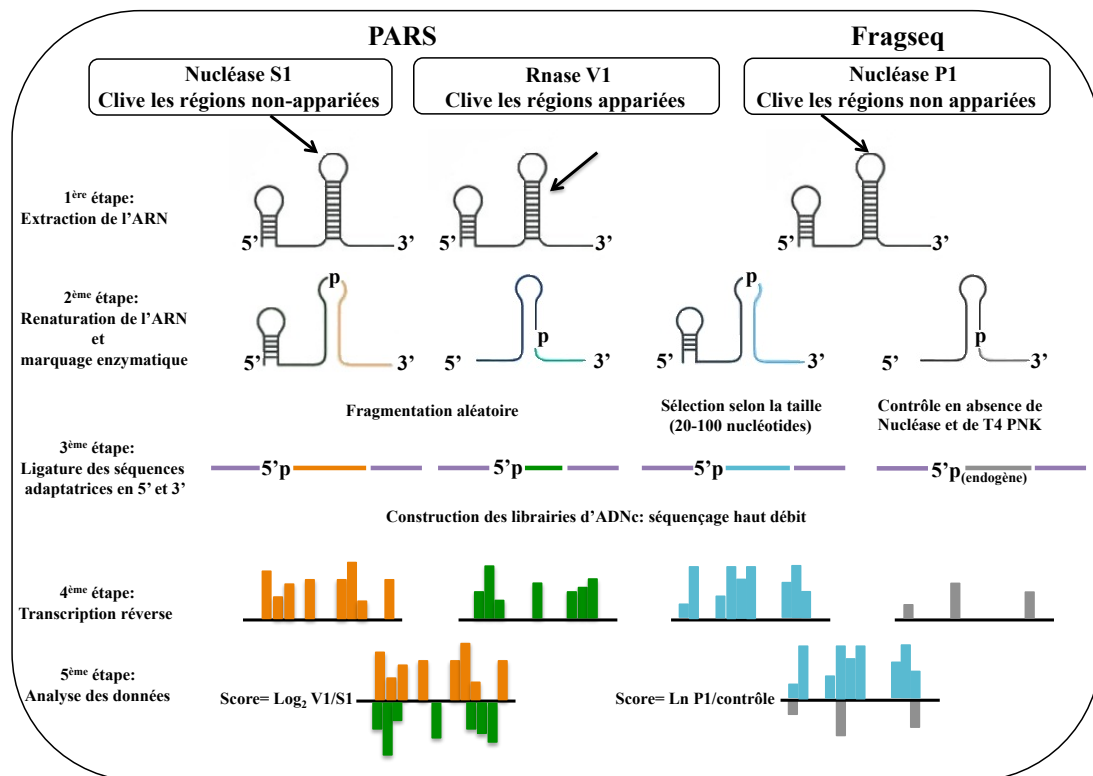


Figure 32: Représentation schématique des deux techniques enzymatiques couplées au séquençage à haut débit d'un génome entier: (A gauche) la technique de PARS développée par Kertesz et al. (A droite) la technique de FragSeq développée par Underwood et al. (Adaptée²¹⁹)

I.4 État des lieux

Nous venons de voir dans ce chapitre que la technique de hSHAPE présente de nombreux avantages (paragraphe I.3.3.1), mais la grande quantité d'ARN nécessaire pour chaque expérience de hSHAPE constitue une limitation majeure lorsque l'on veut étudier des ARN provenant de virus, comme c'est le cas dans notre laboratoire. Afin d'étudier la structure d'ARN viraux, nous avons décidé d'augmenter la sensibilité de détection au niveau du séquenceur d'ADN afin d'augmenter la fluorescence des cDNA et ainsi réduire la quantité d'ARN nécessaire pour réaliser une expérience de hSHAPE. Pour cela, nous avons conçu de nouvelles molécules fluorescentes basées sur le principe de transfert d'énergie et formées de deux entités fluorescentes, un donneur de fluorescence et un accepteur de fluorescence. Dans le deuxième chapitre de mon manuscrit, je présenterai la synthèse et la caractérisation des deux blocs de départ donneur et accepteur de fluorescence ainsi que d'une sonde à transfert d'énergie (tête-queue) formée des deux entités fluorescentes liées entre elles par un bras espaceur.

Nous venons de voir aussi dans ce chapitre (paragraphe I.3.3.2) que diverses techniques de cartographie ont été couplées au séquençage à haut débit afin d'améliorer l'étude structurale des ARN. Malheureusement, ces techniques ne peuvent pas être appliquées pour réaliser en même temps une étude *in virio* et sur un génome entier. Étant donné que notre équipe souhaite étudier la structure de l'ARN génomique du VIH-1 *in virio* au cours du processus de maturation, nous avons choisi de développer une nouvelle technique de cartographie chimique couplée au séquençage à haut débit. Pour cela, nous avons conçu des nouvelles molécules qui permettent dans un premier temps de modifier l'ARN, pour réaliser la cartographie, et dans un deuxième temps de sélectionner les fragments d'ARN modifiés par l'intermédiaire d'une interaction biotine/streptavidine. Cette étape d'enrichissement en ARN modifié devrait nous permettre de réaliser exclusivement le séquençage à haut débit des fragments d'intérêt. Dans le troisième chapitre de mon manuscrit, je présenterai la synthèse et la caractérisation de ces nouvelles molécules ainsi que des résultats préliminaires sur la réaction de ces composés avec des ARN, la fixation d'une molécule de biotine au niveau de la modification et la formation d'un complexe ARN-biotine/streptavidine.

CHAPITRE II : SYNTHÈSE ET CARACTÉRISATION DE NOUVELLES SONDES FLUORESCENTES

II.1 Problématique et objectif

II.1.1 La principale limitation de la technique de hSHAPE

Nous avons vu dans le premier chapitre que la technique de « hSHAPE » est très prometteuse pour l'étude de la structure des ARN, puisque le réactif chimique permet d'obtenir des informations sur toutes les positions de la molécule en une seule expérience. De plus l'utilisation du séquenceur d'ADN permet la quantification des résultats et donc l'obtention d'un grand volume d'informations relatives à la structure de la macromolécule.

Toutefois, une des principales limitations de cette technique réside dans la nature des sondes fluorescentes utilisées pour marquer les amorces et donc les ADNc. En effet, les sondes fluorescentes commerciales (FAM, Vic, Ned et Pet) utilisées dans la technique de « hSHAPE » présentent chacune sa propre longueur d'onde d'excitation et sa propre longueur d'onde d'émission (Fig.33).

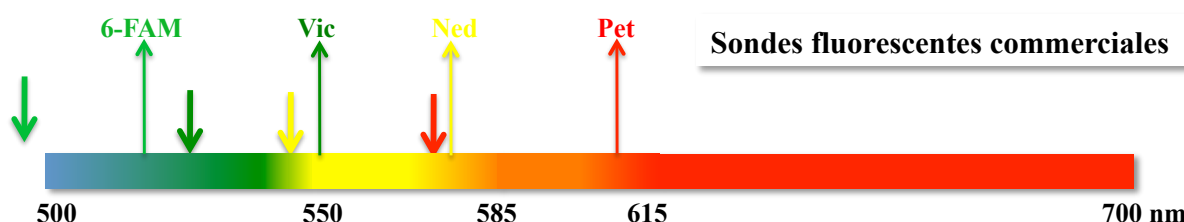


Figure 33: Caractéristiques de chacune des sondes fluorescentes commerciales utilisées dans la technique de hSHAPE.

Ces sondes fluorescentes commerciales sont diversement sensibles au laser ionique du séquenceur. En effet, celui-ci excite ces sondes fluorescentes à la longueur d'onde unique de 488 nm et l'émission de la fluorescence est d'autant plus faible que la longueur d'onde d'excitation de la sonde sera éloignée de celle du laser. Ainsi à quantité égale, une sonde fluorescente telle que la 6-FAM, dont la longueur d'onde d'excitation est à 497 nm, émettra un signal de fluorescence plus important que la sonde Pet dont la longueur d'onde d'excitation est à 595 nm. Face à ce problème, nous avons décidé de synthétiser de nouvelles sondes fluorescentes « à transfert d'énergie » qui seront toutes excitées à la même longueur

d'onde « 488 nm » qui est la longueur d'onde du laser du séquenceur et qui émettront à des longueurs d'onde différentes avec la même intensité de fluorescence pour les quatre sondes (Fig.34).

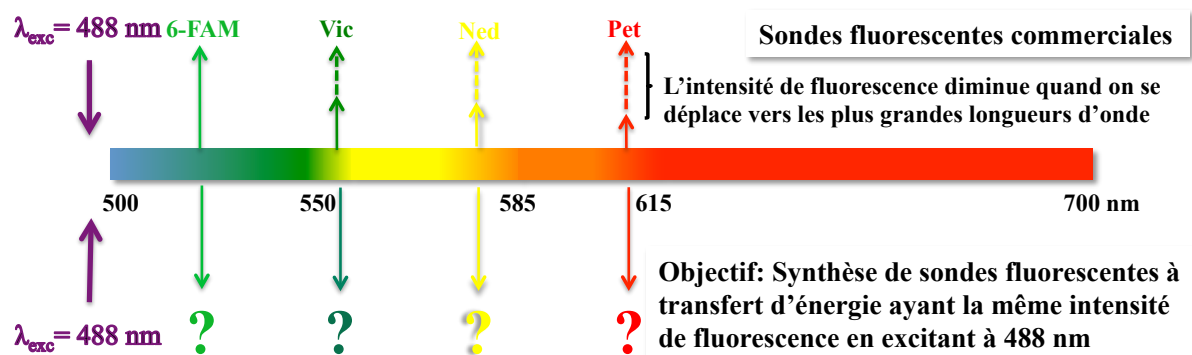


Figure 34: La principale limitation de la technique de hSHAPE et l'objectif de la première partie de mon projet de thèse.

Afin de dépasser cette limitation, une solution consiste à analyser une quantité plus importante de matériel fluorescent et donc de produire plus d'ADNc à partir d'une quantité d'ARN plus importante. Si cette approche est envisageable pour des ARN transcrits *in vitro* ou accessibles en grande quantité, cela est difficilement réalisable pour les ARN qui ne se trouvent pas en grande quantité dans la nature, ou qui sont difficiles à extraire tels que les ARN viraux (VIH-1 et influenza A) que nous étudions dans notre équipe. Il est intéressant de noter que l'équipe de K.Weeks a réussi à étudier le génome entier du VIH-1 par hSHAPE, et que pour cela ils ont du extraire 100 pmoles d'ARN viral à partir de 10 litres de culture²²².

II.1.2 L'optimisation de Kevin Weeks

L'approche de Kevin Weeks pour l'optimisation de la technique de « hSHAPE » consiste dans le développement d'un séquenceur avec un double ion laser, qui permet d'exciter les sondes fluorescentes à deux longueurs d'onde différentes (488 et 514 nm)²²². L'utilisation d'un tel dispositif permet d'augmenter alors les limites de détection moléculaire de l'ordre du zeptomolaire (10^{-21} moles) et d'augmenter la sensibilité d'une expérience de hSHAPE de l'ordre du femtomolaire (10^{-15} moles).

Brièvement, la source laser est reliée à deux tubes photomultiplicateurs, possédant des filtres spectraux. Un seul câble fibre optique possédant des lentilles intégrées est utilisé pour concentrer la lumière provenant de la source d'ion laser. La source laser est placée d'une façon qui permet l'alignement du faisceau lumineux. Le séquenceur est doté d'un objectif

grand angle (40X) et orienté dans le bon angle par rapport au faisceau lumineux, afin de collecter la lumière qui sort des capillaires et de minimiser la perte de lumière. L'objectif est couplé à un cube optique possédant un filtre dichroïque qui reflète la longueur d'onde verte tout en transmettant la lumière jaune. Des lentilles de collecte sont fixées dans le cube optique afin de concentrer la lumière dans une fibre optique qui réagit comme étant un filtre spatial qui enlève la lumière perdue et transmet la lumière fluorescente au détecteur de fluorescence (Fig.35). (Adaptée²²²).

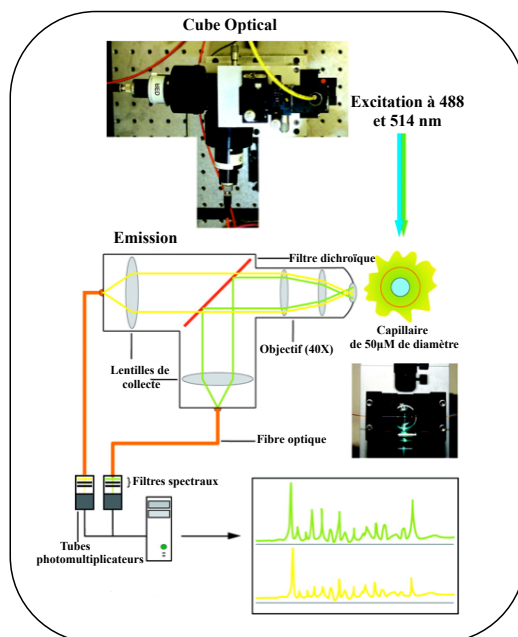


Figure 35: Séquenceur à double ion laser excitant les sondes fluorescentes à deux longueurs d'onde 488 et 514 nm. (Adaptée²²²)

Ce nouveau séquenceur a été utilisé pour étudier l'ARN du virus XMRV (Xenotropic Murine leukemia Related-Virus) *in virio* et *ex virio*²²². La quantité d'ARN utilisée dans cette expérience était de 1fmoles ce qui est 1000 fois inférieure à la quantité d'ARN nécessaire pour du hSHAPE conventionnel et le rapport de corrélation entre les deux approches est correct puisqu'il est de $R=0,89$. L'étude a permis de déterminer des domaines fonctionnels dans la structure qui sont critiques pour l'encapsidation du génome, elle a également montré que le virus XMRV encapside son génome sous forme de dimère comme le virus MuLV (Moloney murine Leukemia Virus), un autre gammarétrovirus.

II.1.3 Objectif du travail de thèse

L'objectif de la première partie de ce travail de thèse est la synthèse d'un jeu de sondes fluorescentes destiné à améliorer la détection des ADNc sur les séquenceurs d'ADN à un seul laser afin de permettre l'étude structurale de faibles quantités d'ARN. Ces sondes fluorescentes, contrairement aux sondes fluorescentes commerciales pour marquer les amorces, sont constituées de deux fluorophores, un accepteur de fluorescence et un donneur de fluorescence reliés entre eux par un bras espaceur (Fig.36). Elles appartiennent à la famille des « BigDyes » développée en littérature²²³.

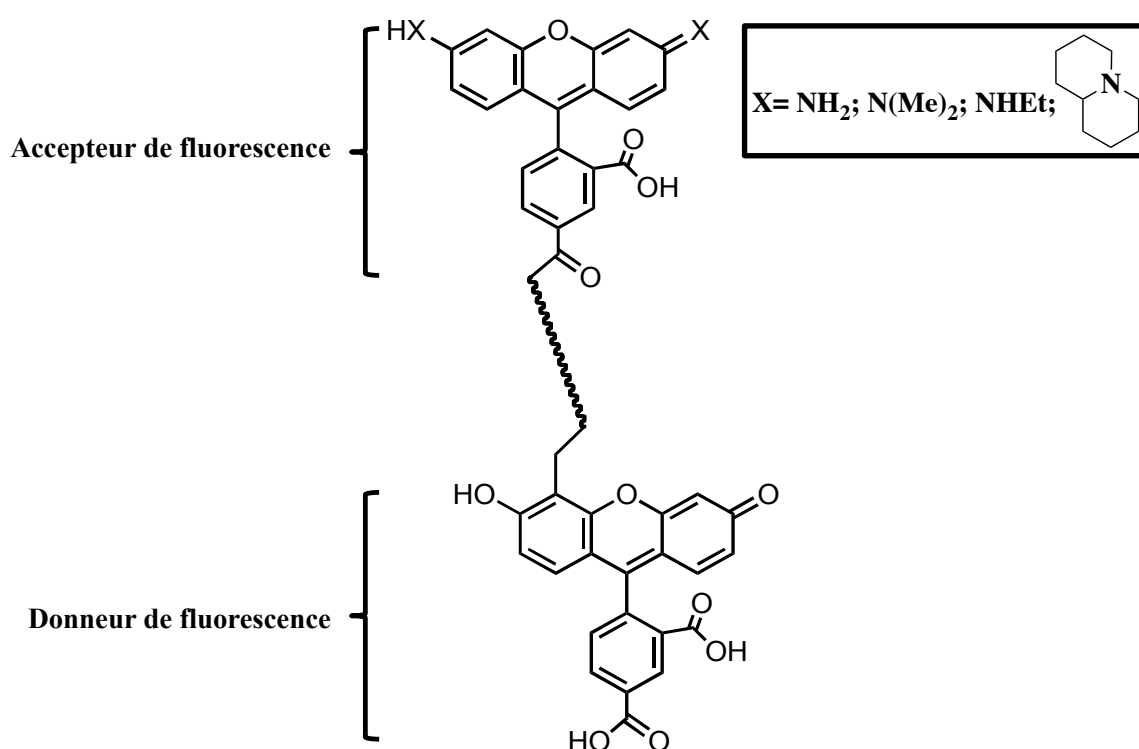


Figure 36: Le modèle des sondes fluorescentes à transfert d'énergie.

Pour ces sondes fluorescentes à transfert d'énergie, la FAM a été choisie en tant que donneur de fluorescence du fait qu'elle présente le maximum d'émission de fluorescence en comparaison avec les trois autres sondes fluorescentes commerciales (Vic, Ned et Pet), en excitant à 488 nm (Fig.34). Pour les entités accepteur de fluorescence, une des conditions nécessaires pour que le transfert d'énergie prenne place entre le donneur et l'accepteur, est que le spectre d'émission du donneur recouvre le spectre d'absorption de l'accepteur. Pour cette raison, nous avons choisi les quatre dérivés de la rhodamine qui présentent en plus, une large gamme de fluorescence et des valeurs de rendement quantique plus au moins élevées (Tableau 2).

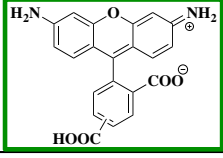
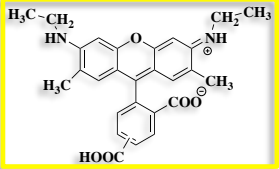
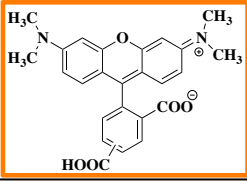
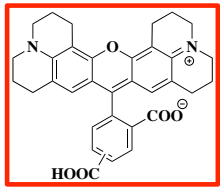
Dérivés de rhodamine	λ_{exc} (nm)	λ_{em} (nm)	Rendement quantique ϕ
 5-, 6- carboxy-rhodamine	485	534	0.91 (Tampon)
 5-, 6- carboxy-Rh6G	524	553	0.94 (EtOH)
 5-, 6- carboxy-TAMRA	555	576	0.71 (Tampon)
 5-, 6- carboxy-ROX	585	602	0.95 (Tampon)

Tableau 2 : Caractéristiques des dérivés de la rhodamine. (Adapté^{220,221})

II.2 Les sondes fluorescentes

Il existe plusieurs méthodes de détection des ADNc obtenus lors d'une expérience de séquençage, mais la méthode de préférence est l'utilisation des sondes fluorescentes qui permet de détecter les fragments d'ADNc par une méthode simple, sécurisée et hautement précise. En fluorescence, quatre sondes fluorescentes avec des propriétés spectrales différentes sont utilisées. Ces sondes fluorescentes sont soit hybridées à l'extrémité 5' de l'amorce (amorce marquée), soit hybridées à un didésoxynucléoside-5'-triphosphate (ddNTP) donc réagissant en tant que terminateurs de chaîne²²⁴.

Différents types de sondes fluorescentes formées d'une seule ou de deux entités fluorescentes ont été décrits dans la littérature. Les sondes fluorescentes formées de deux entités fluorescentes, un donneur de fluorescence et un accepteur de fluorescence, sont basées sur le principe de transfert d'énergie, elles ont été développées par plusieurs équipes en variant toutefois la nature du bras espaceur qui sépare ces deux fluorophores. La méthode de séquençage utilisant ces sondes fluorescentes se différencie pour chaque équipe, entre

amorces marquées ou terminateurs de chaîne, et chaque type de sondes fluorescentes synthétisé présente aussi bien des avantages que des inconvénients pour le séquençage d'ADN.

II.2.1 Une seule entité fluorescente

Les techniques classiques consistent à marquer les amorces avec une entité fluorescente comme par exemple la carboxy-fluorescéine (FAM), carboxy-4',5'-dichloro-2',7'-diméthoxyfluorescéine (JOE), carboxy-tétraméthyl-rhodamine (TAMRA) et carboxy-X-rhodamine (ROX) (Fig.37), qui sont excitées dans les méthodes de séquençage d'ADN avec un laser ionique à une longueur d'onde commune de 488 nm^{225,226,227}.

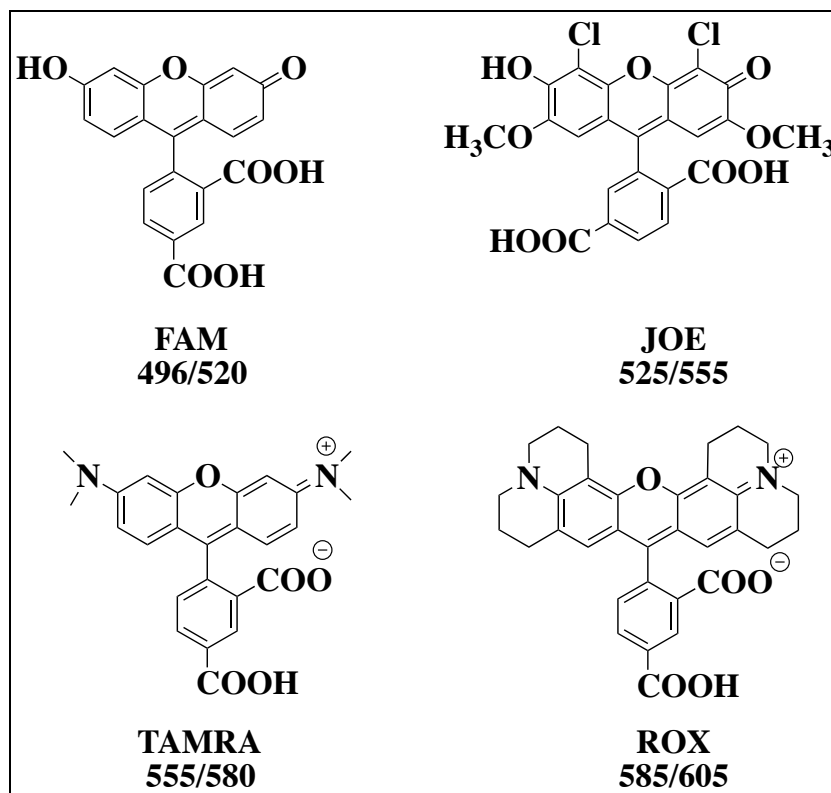


Figure 37: Les quatre sondes fluorescentes FAM, JOE, TAMRA et ROX formées d'une seule entité fluorescente, utilisées dans les méthodes de séquençage d'ADN.

En excitant à une longueur d'onde précise, ces différentes sondes fluorescentes possédant chacune des propriétés spectrales différentes, des longueurs d'onde d'excitation et d'émission spécifique à chacune, et des coefficients d'extinctions différents, émettent différemment. Le set de fluorophores idéal doit posséder une absorbance moléculaire élevée en excitant à une longueur d'onde commune tous les membres du set, et doit présenter une forte émission de fluorescence et bien séparée avec un rendement quantique élevé. La fluorescence émise sera

d'autant plus faible que la longueur d'onde d'excitation de la sonde est loin de la longueur d'onde d'excitation du laser. Ce critère n'est pas évident pour tous les dérivés de la rhodamine tels que la TAMRA et la ROX, qui présentent respectivement une absorbance de 1/8 et de 1/10 par rapport à la fluorescéine en excitant à 488 nm. Ceci provoque une diminution de la réponse et de la lecture des séquences, puisqu'il y aura des séquences marquées mais pas détectées. L'inconvénient de ce type de sondes fluorescentes est alors la nécessité de deux longueurs d'onde d'excitation différentes : 488 nm pour FAM et JOE et 543 nm pour TAMRA et ROX²²⁸.

Des dérivés de la famille de la dichloro-rhodamine ont été développés dans le but d'améliorer l'émission de la fluorescence de ces dérivés de la rhodamine à une seule longueur d'onde d'excitation et d'améliorer la résolution des pics (électrophérogrammes) obtenus après électrophorèse. Un problème commun lié à l'utilisation des dérivés de la rhodamine en tant que terminateurs de chaîne, se produit lorsqu'une base G suit une base A car la taille du pic de fluorescence correspondant à la base G est réduite voir supprimée. Ce problème est fréquent aussi lorsque le pic qui suit celui de la base G a une taille qui dépasse la moyenne. En remplaçant la rhodamine par la d-rhodamine, les spectres d'émission de ces dérivés sont beaucoup mieux résolus par rapport aux composés sans les atomes de chlore. Le remplacement des deux positions A1 et B2 (Fig.38.A) est très avantageux. Une amélioration des propriétés (spectres d'émission plus fins Fig.38.C) de ces dérivés est remarquée si au moins un des quatre groupements A1, A2, A3 et B1 (Fig.38.A) représente un halogénure et plus particulièrement un fluor (F) ou un chlore (Cl)²²⁹.

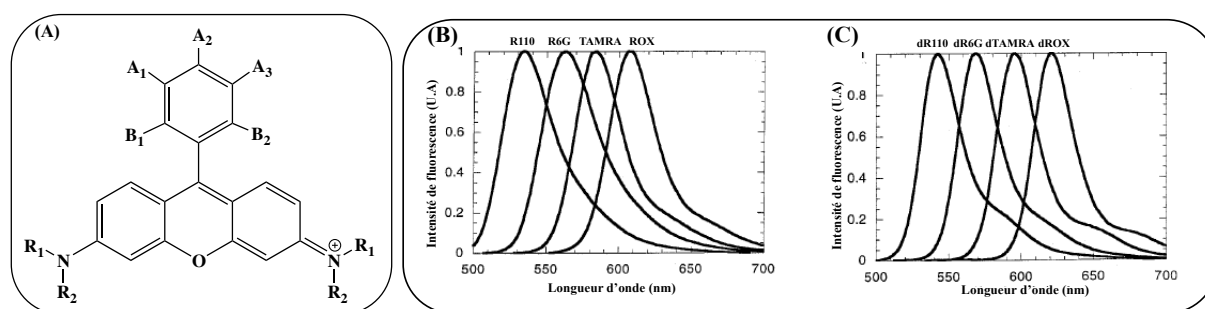


Figure 38: (A) Les dérivés de la rhodamine avec les différentes possibilités de substitution. (B et C) Comparaison entre les spectres d'émission des dérivés de la rhodamine (B) et des dérivés de la d-rhodamine (C). Les spectres des dérivés d-rhodamine sont mieux résolus que ceux de la famille des rhodamines correspondantes. (Adaptée²²³)

Des expériences ont été faites sur ces dérivés de la d-rhodamine en changeant la nature du bras espaceur qui lie ces dérivés de la rhodamine au nucléotide 2',3'-didéoxyribose-triphosphate. Ces composés agissent en tant que terminateurs de chaîne et sont appelés « Dichlororhodamine terminators ». Une série de composés a été synthétisée soit en utilisant le propargylamine (PA), soit le propargyl éthoxy-amine (EO) comme bras espaceur²²⁴ (Fig.39). En marquant les ddNTP avec ces dérivés, les pics de fluorescence observés sont beaucoup plus uniformes et on observe moins de suppression du pic correspondant à la base G^{230,231}.

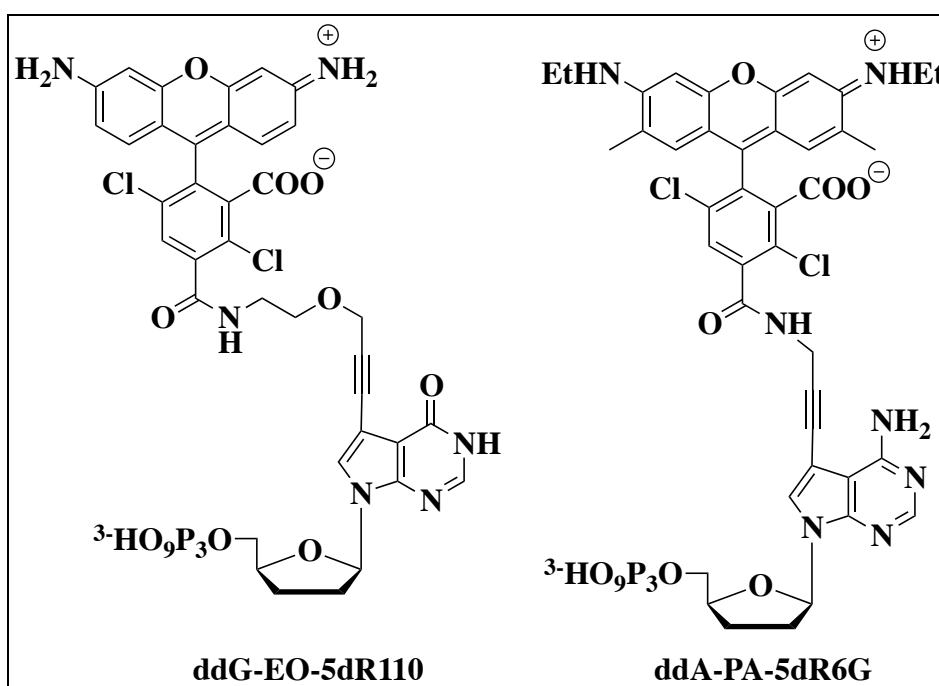


Figure 39: Structure des terminateurs de chaîne d-rhodamine. Le bras espaceur qui lie ces dérivés au ddNTP est soit le groupement propargylamine (PA) soit le groupement propargyl éthoxy-amine (EO).

II.2.2 Deux entités fluorescentes

II.2.2.1 Amorces à transfert d'énergie

Une autre classe de sondes fluorescentes a été développée et correspond aux amorces à transfert d'énergie ou « Energy Transfer Primers ». Dans ce cas, le bras espaceur qui sépare le donneur de l'accepteur de fluorescence est composé d'une série de nucléotides sur la même séquence d'ADN^{232,233} (Fig.40). Ces amorces sont appelées D-N-A en accord avec le « D »

pour Donneur qui sera toujours commun appartenant à la famille de la fluorescéine, le « N » pour le nombre de nucléotides qui séparent les deux fluorophores, et « A » pour accepteur de fluorescence qui sera un des dérivés de la rhodamine. Dans ce cas, les amorces seront marquées par deux sondes fluorescentes : un donneur de fluorescence fixé à l'extrémité 5' de l'amorce et un accepteur de fluorescence lié à une thymine modifiée (T^*) située à une certaine distance du donneur. Pour que cette approche soit fructueuse il faut que les deux sondes fluorescentes soient suffisamment éloignées l'une de l'autre, pour éviter le phénomène de quenching mais pas trop pour avoir une efficacité de transfert suffisante. Différents types de sondes ont été testés en variant la longueur du bras espaceur : F3T, F3R, F10F F10J (J, R, T et F correspondent à JOE, ROX, TAMRA et FAM). En excitant à une longueur d'onde commune de 488 nm, l'absorbance est forte pour les quatre et la fluorescence émise par ces sondes à transfert d'énergie est supérieure à celle émise par chaque accepteur pris séparément. Les amorces à transfert d'énergie F3T et F3R ont montré un phénomène de « quenching » entre les deux moments dipolaires des deux entités fluorescentes²³². La distance optimale entre les deux entités fluorescentes est donc entre 8 et 10 nucléotides, pour avoir un transfert d'énergie efficace. L'inconvénient de ce type de sonde à squelette oligonucléotidique est qu'elle ne peut pas être appliquée à une petite séquence oligomère²³³.

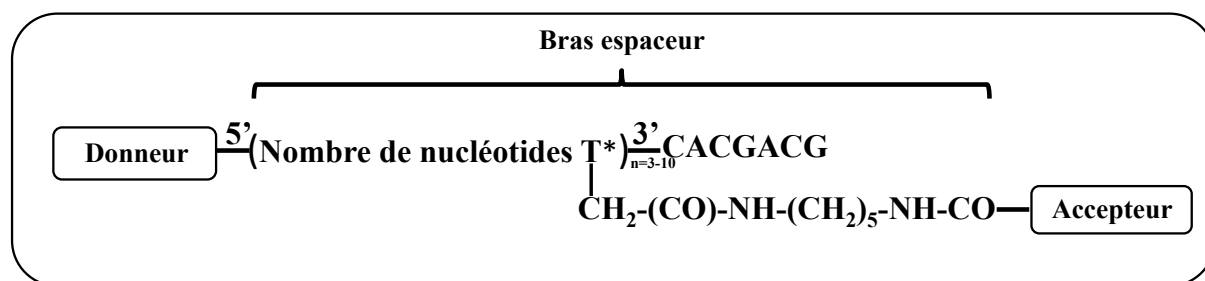


Figure 40: Représentation de l'amorce à transfert d'énergie. Le bras espaceur qui lie les deux entités fluorescentes : donneur et accepteur de fluorescence est un squelette oligonucléotidique. (La séquence d'amorce illustrée est aléatoire). (Adaptée²³²)

Une nouvelle classe d'amorces à transfert d'énergie a été développée indépendante de la séquence d'ARN étudiée, et possédant comme bras espaceur une séquence polymère (S6 : ou séquence de « Spacer » formée de 6 monomères de 1',2'-didéoxyribose-phosphate (Fig.41). Ces nouvelles amorces sont nommées FS₆F, FS₆J, FS₆T, et FS₆R (F, J, T et R correspondent respectivement à FAM, JOE, TAMRA et ROX), elles possèdent toujours la fluorescéine en tant que donneur de fluorescence et un des dérivés de la rhodamine en tant qu'accepteur de fluorescence. La séquence polymère contient une base (T^*) modifiée et possède un groupement protecteur sur l'amine primaire en position C-5. La base T^* est attachée à

l'extrémité 5' de la séquence de l'oligomère et à l'extrémité 3' du bras espaceur. Le donneur est attaché à l'extrémité 5' de l'oligomère et l'accepteur est attaché à l'amine primaire de la base T*. Les avantages de ce type de bras espaceur sont : (i) la séquence du bras espaceur ne s'hybride pas à une séquence aléatoire sur l'ARN (ii) la présence du bras espaceur maintient la fonction du squelette acide nucléique-phosphate (iii) l'élimination des bases azotées sur les groupements désoxyribose dans la séquence du bras espaceur réduit la probabilité d'un phénomène de quenching entre les deux entités fluorescentes suite à un encombrement stérique. Ce nouveau type de sondes a montré une amélioration des propriétés spectrales par rapport aux amorces marquées par une seule sonde fluorescente en excitant à 488 nm²³⁴.

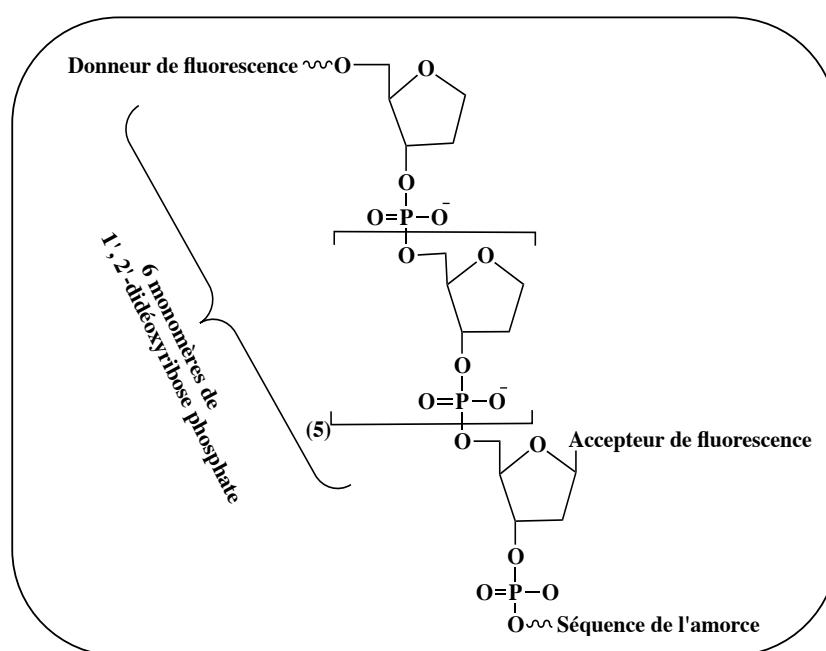


Figure 41: Modèle de l'amorce marquée avec deux entités fluorescentes avec un bras espaceur formé par six monomères de 1',2'-didésoxyribose phosphate intégré dans la séquence d'ADN. (Adaptée²³⁴)

II.2.2.2 Sondes fluorescentes à transfert d'énergie

II.2.2.2.1 Bras espaceur trifonctionnel

Une approche synthétique différente a été utilisée pour produire des fluorophores à transfert d'énergie en tant que terminateurs de chaîne. Elle consiste en l'utilisation comme bras espaceur d'une molécule rigide, linéaire et trifonctionnelle²³⁵. Dans ce cas de sondes fluorescentes à transfert d'énergie, les deux fluorophores sont dans une orientation opposée.

Le t-Boc-L-4-iodophénylalanine est mis en réaction par « couplage de Heck » avec du propargylamine protégé avec du TFA en présence du $(\text{Ph}_3\text{P})_4\text{Pd}(0)/\text{CuI}/\text{Et}_3\text{N}/\text{DMF}$ afin de produire un bras espaceur qui est adapté pour se lier au donneur et à l'accepteur sur les positions protégées d'amines orthogonales et de se lier aussi à l'alkynylpropargylaminodidéoxynucléoside-5'-triphosphate (Fig.42). La liaison du donneur de fluorescence commun (FAM) à l'un des groupements amine et l'accepteur (R110 ou rhodamine, REG ou R6G, TAMRA, ou ROX) au second groupement amine donne quatre cassettes de fluorophores à transfert d'énergie avec des rendements élevés. Ces fluorophores sont adaptés pour se lier aux amorces ou aux ddNTP. Parmi les 64 combinaisons possibles, un nombre de terminateurs de chaîne a été synthétisé et testé dans le séquençage d'ADN avec une variété de séquences d'ADN et d'ADN polymérases. Une série de terminateurs de chaîne a été optimisée et FAM-R110-ddGTP, FAM-REG-ddUTP, FAM-TAMRA-ddATP et FAM-ROX-ddCTP ont été sélectionnés. Ils ont montré une meilleur brillance par rapport aux sondes fluorescentes à une seule entité fluorescente en excitant à 488 nm et plus particulièrement dans le cas de TAMRA et ROX. En effet, le FAM-ROX en tant que terminateur de chaîne est 18 fois plus brillant que le ROX tout seul.

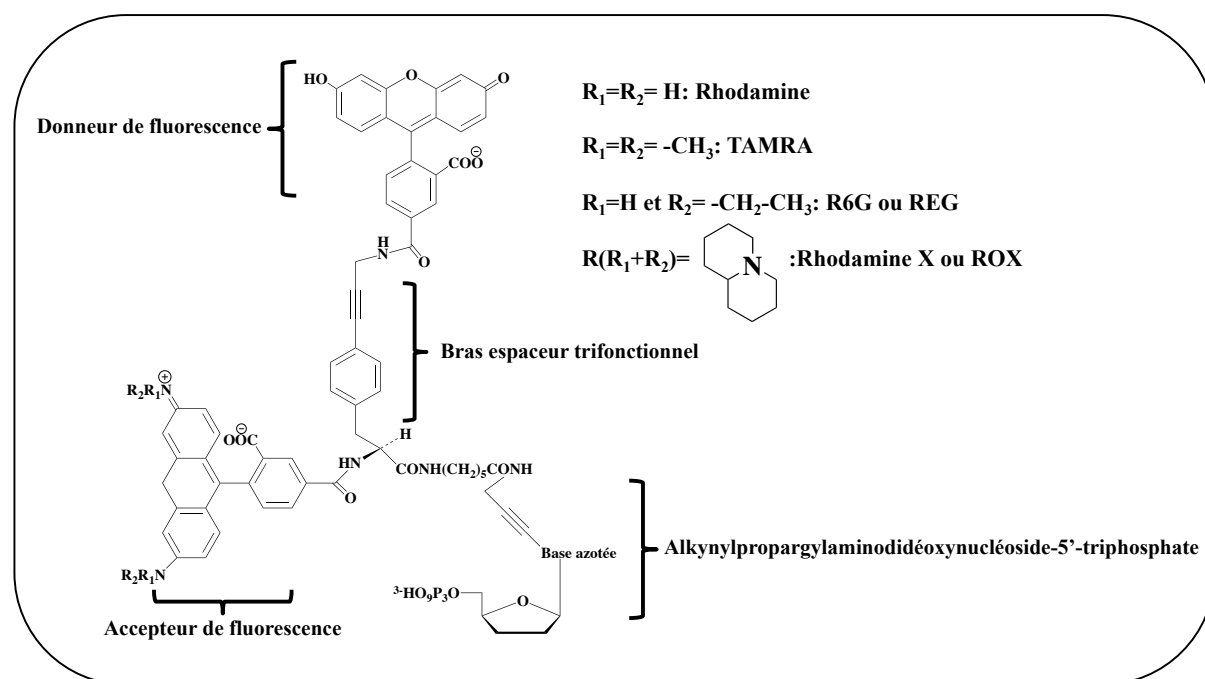


Figure 42: Structure des sondes fluorescentes à transfert d'énergie possédant un bras espaceur trifonctionnel. (Adaptée²³⁵)

Une amélioration de ces sondes fluorescentes a été réalisée en utilisant des dérivés d'acides aminés (phénylalanine ou propargylaminophénylalanine, et glycine ou aminoéthylglycine) comme squelette pour le bras espaceur. Les groupements amines sont utilisés pour lier les deux entités fluorescentes alors que la fonction acide carboxylique est utilisée pour lier la molécule alkynylpropargylaminodidéoxynucléoside-5'-triphosphate²³⁶ (Fig.43). En variant la structure du bras espaceur, une différence significative dans l'efficacité de transfert a été observée et le composé utilisant le para-aminophénylalanine (composé 4, figure 43) comme bras espaceur a montré la meilleure brillance. Par contre, il a été observé qu'en utilisant la phénylalanine ou la propargylaminophénylalanine associée à l'aminéthylglycine comme bras espaceur, l'émission de fluorescence est complètement abolie à cause du phénomène de quenching (composés 5 et 6, figure 43).

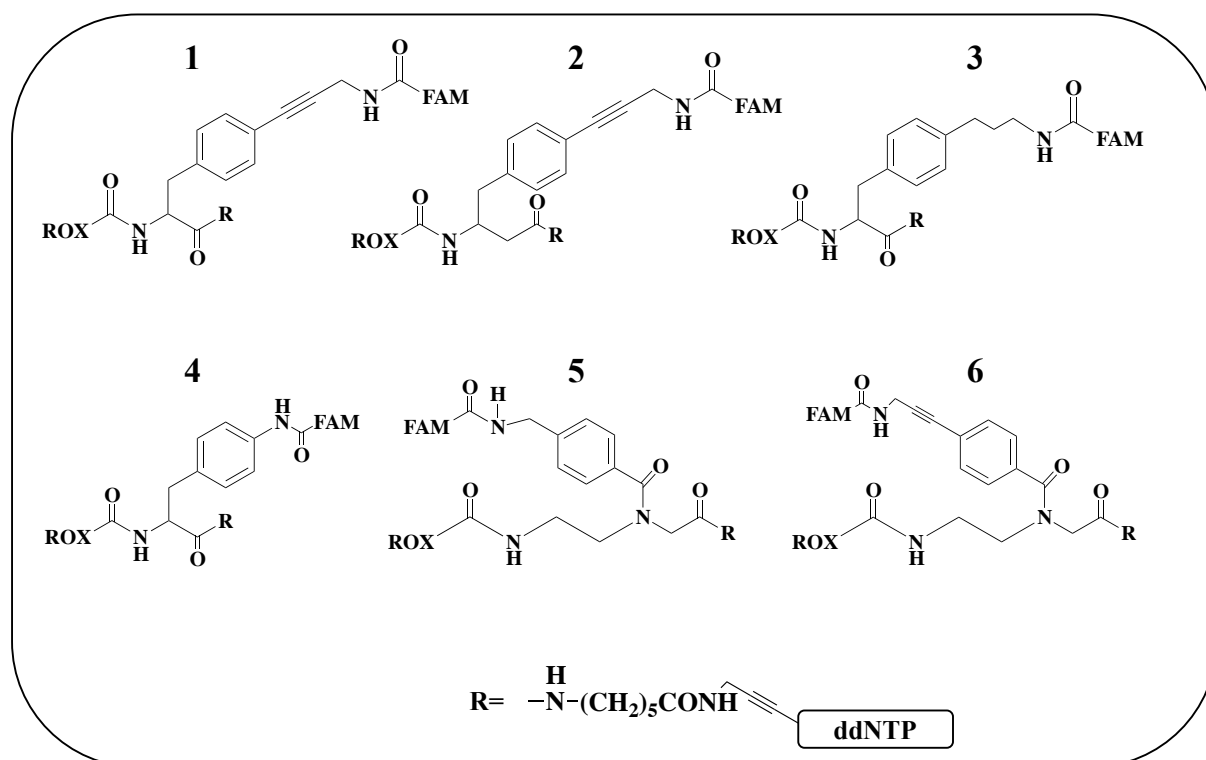


Figure 43: Représentation des différents modèles du bras espaceur trifonctionnel synthétisé à partir des acides aminés. Les modèles (5) et (6) ont montré un phénomène de quenching entre les entités fluorescentes (donneur de fluorescence : « FAM » et accepteur de fluorescence « ROX »). (Adaptée²³⁶)

Il était nécessaire de développer des fluorophores compacts qui fournissent une amélioration dans le domaine du séquençage d'ADN en comparaison avec les sondes fluorescentes formées d'une seule entité fluorescente, ou avec les fluorophores à transfert d'énergie possédant un squelette d'oligonucléotide en tant que bras espaceur. En partant du principe des

amorces à transfert d'énergie, d'autres sondes fluorescentes ont été développées mais en utilisant un autre bras espaceur et une orientation différente des deux entités fluorescentes. Le premier modèle élaboré était le Bifluor-1, dont le donneur de fluorescence est un dérivé de la fluorescéine, la 4'-aminométhyl-5-carboxy-fluorescéine couplée à un accepteur de fluorescence dérivé de la famille des rhodamines, la 5 carboxy-tétraméthylrhodamine. Le bras espaceur du Bifluor-1 est le groupement aminométhyl en position 4' du xanthène de la molécule 5-carboxy-fluorescéine (Fig.44). Dans ce type de fluorophores pour lequel les deux sondes sont dans la même orientation, le donneur de fluorescence joue un double rôle en se fixant à l'accepteur et aussi au nucléotide. Les dérivés de la fluorescéine sont les composés idéaux en étant des molécules bi-fonctionnelles pour réagir en tant que donneur de fluorescence, puisque leur structure est facilement modifiable avant et après la synthèse du chromophore. Le spectre d'absorbance du Bifluor-1 est la somme des absorbances du donneur et de l'accepteur, par contre le spectre d'émission correspond à l'accepteur de fluorescence. La fluorescence du groupement donneur de fluorescence « fluorescéine » est alors complètement inhibé. Le Bifluor-1 semblait être un bon candidat pour le séquençage d'ADN, mais la brillance (ou intensité de fluorescence qui est proportionnelle au coefficient d'extinction ϵ et au rendement quantique ϕ) de ce composé n'était pas meilleure que celle de la tétraméthylrhodamine toute seule en excitant à 488 nm. D'autres composés ont donc été recherchés²²³.

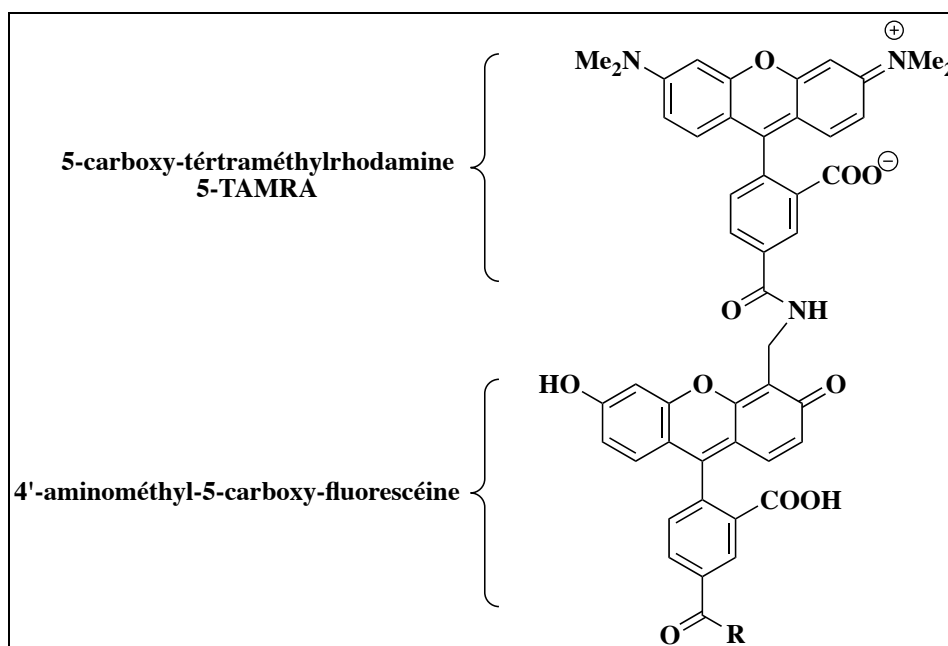


Figure 44: Modèle du Bifluor-1. (Adaptée²²³)

En se basant sur le modèle du Bifluor-1, l'équipe de J.M. Cassel a développée une série de fluorophores formés par les isomères 4'-aminométhyl-5- ou 6- carboxyfluorescéine comme molécule donneur de fluorescence, et un des dérivés de la dichlororhodamine « d-Rhodamine » (R6G, TAMRA, ROX) comme molécule accepteur de fluorescence²²³. Le bras espaceur entre les deux entités est l'acide 4-aminométhylbenzoïque. Le donneur de fluorescence est attaché à la fonction acide carboxylique du bras espaceur, et le groupement amine du bras espaceur est attaché à l'accepteur de fluorescence (Fig.45). Aucune différence de brillance entre les deux isomères 4'-aminométhyl-(5) et (-6) carboxyfluorescéine n'est observée, par contre, couplé à la d-Rhodamine, l'isomère (-5) est beaucoup plus brillant que l'isomère (-6). Ces fluorophores à transfert d'énergie ont montré des meilleures propriétés spectrales par rapport aux fluorophores standards en excitant à 488 nm. En effet, ils apportent une amélioration de la brillance d'un facteur de 2 à 2,5 fois, et une diminution du bruit de fond dû à la fluorescence perdue de 2 fois, ce qui provoque une augmentation du rapport signal/Bruit de 4 à 5 fois. De plus, ces composés peuvent être utilisés pour marquer différents types de composés, tout comme les sondes fluorescentes à une seule entité.

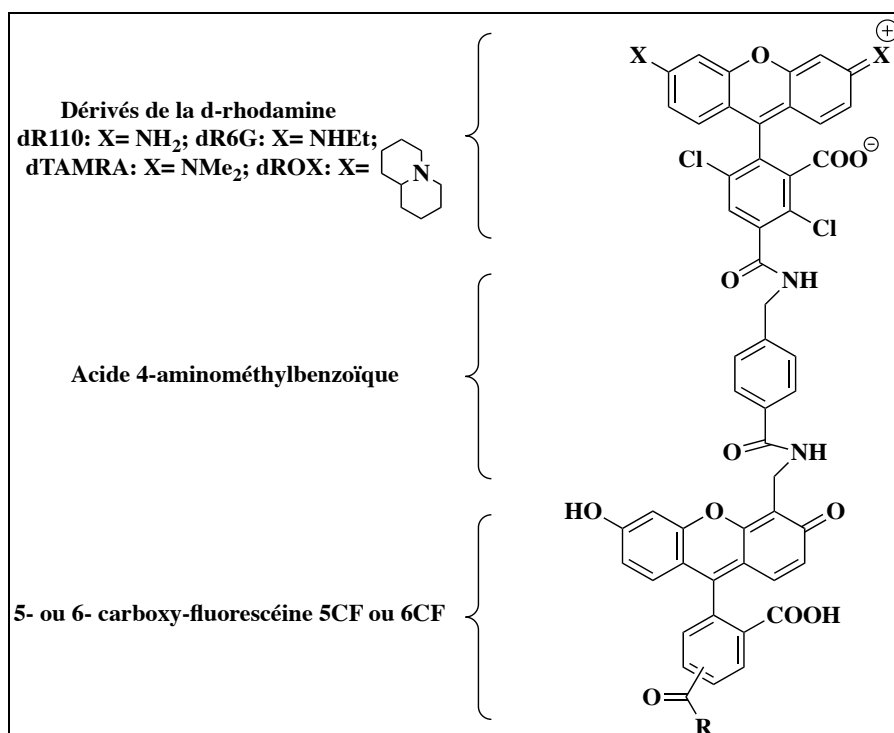


Figure 45: Modèle du BigDye. (Adaptée²²³)

Ces BigDyes sont utilisés commercialement pour marquer les didésoxynucléotides (Fig.46), qui agissent comme terminateurs de chaîne lors des réactions de polymérisation comme la

rétro-transcription. Du fait de leurs propriétés physico-chimiques, ces BigDyes ont été utilisés avec succès pour le séquençage d'ADN de grande masse moléculaire²²⁴.

II.3 Conception et synthèse de nouveaux BigDyes

Nos BigDyes sont aussi constitués de deux entités fluorescentes, un donneur de fluorescence, dérivé de la fluorescéine possédant la plus grande émission de fluorescence en excitant à 488 nm et un accepteur de fluorescence, dérivé de la rhodamine. Ces entités sont liées entre elles par un bras espaceur. Alors que les BigDyes rapportés dans la littérature possèdent un groupement benzoyle comme bras espaceur, dans notre cas, nous avons décidé de lier les deux entités fluorescentes par un cycle triazole qui sera obtenu par une cycloaddition de Huisgen entre une fonction alcyne et une fonction azoture appelée communément réaction de « Click Chemistry » (Fig.47). Le donneur de fluorescence sera l'un des deux isomères (-5) ou (-6) de la carboxyfluorescéine substituée en position 4' du groupement xanthène avec une fonction alcyne. L'accepteur de fluorescence sera l'un des quatre dérivés de la rhodamine.

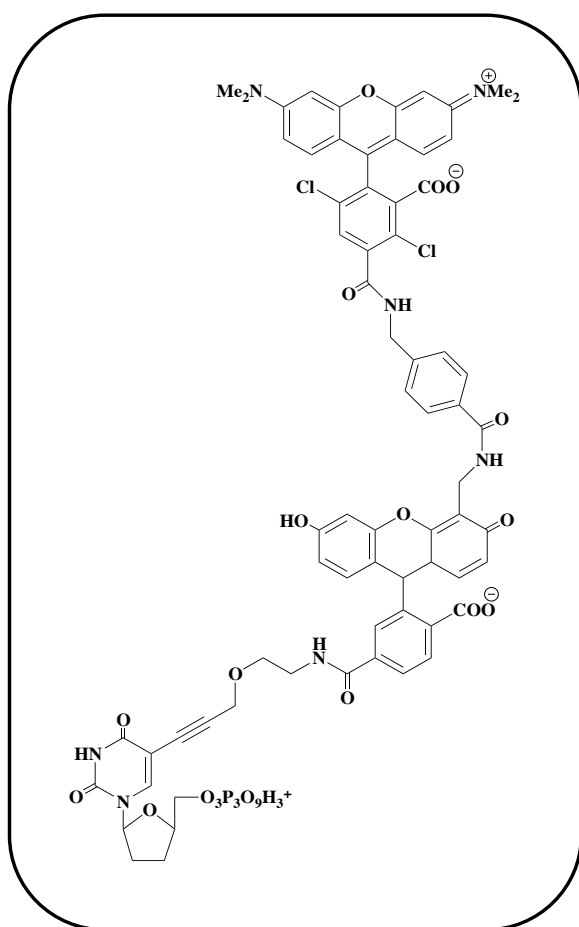


Figure 46: Structure d'un BigDye formé par du 6-FAM lié au 5-TAMRA par l'intermédiaire de l'acide 4-aminométhyl benzoïque, couplé à un didésoxynucléoside triphosphate agissant en tant que terminateur de chaîne. (Adaptée²²⁴)

La modification du substituant X sur le xanthène, permettra d'obtenir des sondes qui émettent de la fluorescence à des longueurs d'ondes différentes : la rhodamine si $X = \text{NH}_2$ ($\lambda_{\text{ém}} = 534$ nm), la tétraméthylrhodamine (TAMRA) si $X = \text{NMe}_2$ ($\lambda_{\text{ém}} = 576$ nm), la rhodamine X (ROX) si X est formée par deux cycles à 6 maillons chacun, qui forment ensemble le cycle de la julolidine ($\lambda_{\text{ém}} = 602$ nm), la rhodamine 6G si $X = \text{NHEt}$ ($\lambda_{\text{ém}} = 553$ nm). Ces dérivés de la rhodamine seront modifiés au niveau de la fonction acide carboxylique afin de pouvoir introduire une fonction azoture et ainsi réaliser le couplage de Huisgen.

La présence de la fonction acide carboxylique sur l'entité « donneur » de fluorescence 5- ou 6- carboxy-fluorescéine permettra l'introduction d'une fonction azoture. Cette fonction azoture pourra réagir avec un nucléotide modifié par une fonction alcyne commerciale, et former le lien avec l'oligomère par réaction de click chemistry.

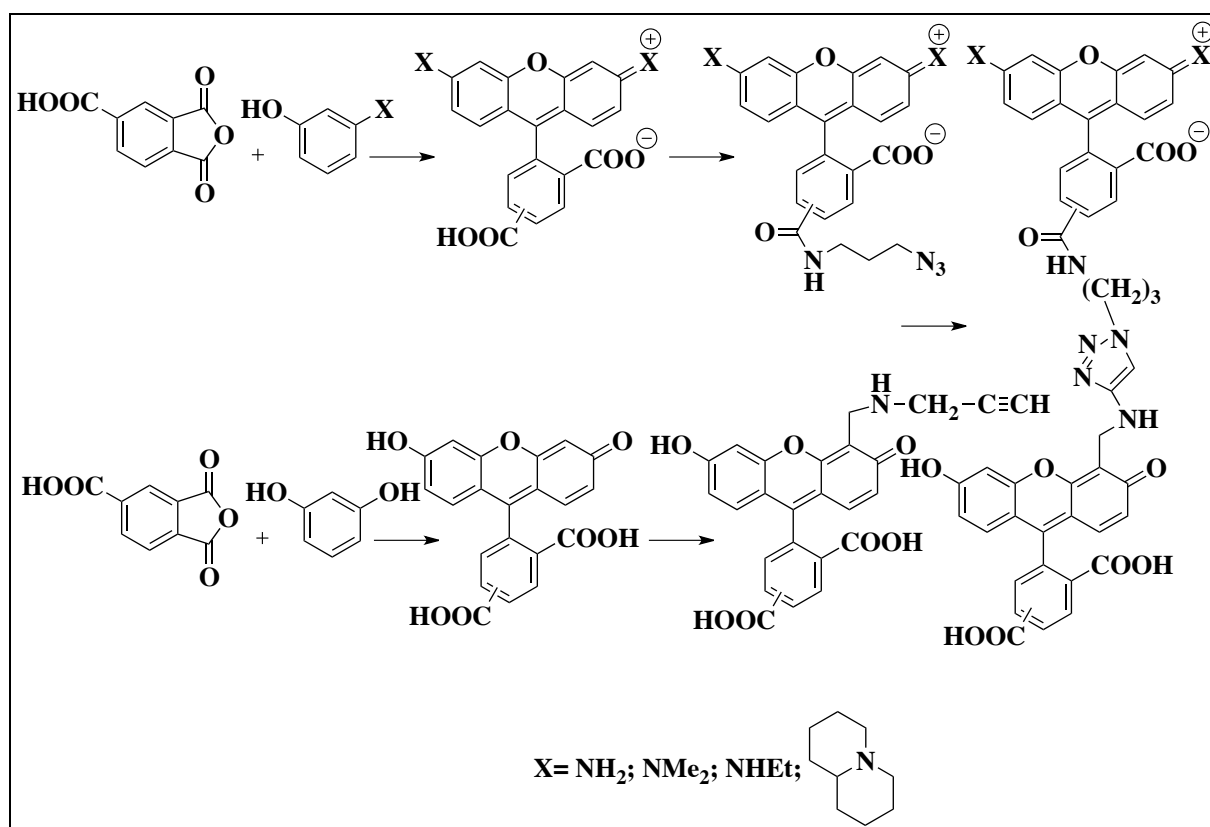


Figure 47: Représentation générale de la synthèse de nos BigDyes et la conception du nouveau bras espaceur par la cycloaddition de « Huisgen » ou réaction de « Click Chemistry ».

Nous pouvons envisager trois sites de fixation différents de ces BigDyes au nucléotide modifié : (i) soit sur la base azotée (ii) soit sur le ribose (iii) soit sur le phosphate (Fig.48).

Ces BigDyes serviront donc à marquer les amorces pour les différentes étapes de hSHAPE, qui nécessitent quatre sondes fluorescentes différentes.

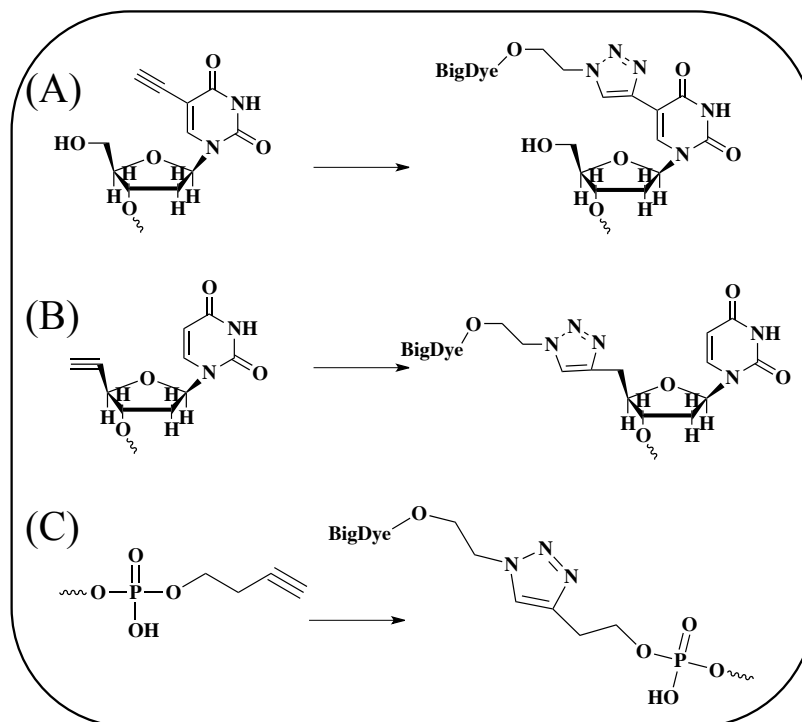


Figure 48: Les trois sites de fixation possibles des BigDyes au nucléotide: (A) sur la base azote (T*), (B) sur le ribose, (C) sur le phosphate.

II.4 MATÉRIELS ET MÉTHODES

II.4.1 Préparation du matériel, techniques de purification et de caractérisation

Préparation du matériel et des solvants

- La verrerie utilisée est séchée à 110°C pendant une nuit au minimum et refroidie sous vide avant utilisation.
- Les cuvettes en quartz pour les analyses en fluorescence et en dichroïsme circulaire sont rincées avec une solution d'Hellmanex (2%) (Hellma), puis avec de l'eau MilliQ et de l'éthanol et séchées avant l'utilisation.
- Pour effectuer les réactions sous atmosphère inerte d'argon, le montage contenant les réactifs est successivement mis sous vide puis sous argon à l'aide d'une rampe à vide. Cette procédure est effectuée 3 fois avant l'ajout du solvant.
- Distillation des solvants : Pyridine sur KOH, méthanol sur sodium, dichlorométhane sur P₂O₅, et Et₃N sur KOH.
- Tampon phosphate (pH=7.0) préparé à partir de deux solutions : Solution (A) : NaH₂PO₄.H₂O à pH= 4,3 (39 ml) et Solution (B) : Na₂HPO₄.12H₂O à pH= 9,1 (61 ml).
- Tampon TEAA (pH=7.0) préparé à partir de CH₃COOH et Et₃N.

Techniques de purification

- Les Chromatographies sur Couche Mince (CCM) sont faites sur des plaques de silice Merck 60F254. Elles sont révélées par une lampe UV (254nm) ou (365nm) pour les sondes fluorescentes, et avec une solution acide de vanilline (1% vanilline, 2% H₂SO₄ dans l'éthanol 95%) ou KMnO₄ pour la réaction de préparation de l'amine (3-azido propylamine)
- Les purifications sur gel de silice sont faites avec la silice Merck Si 60 (40-63µm). Les réactions de purification des fluorophores sont faites en utilisant des éluants composés de MeOH/CH₂Cl₂ ou MeOH/CH₃COOH/CH₂Cl₂.

- Chromatographie en phase liquide à haute performance (HPLC) sur ÄKTA Explorer HPLC (GE life sciences), avec une colonne de silice phase inverse C18. L'éluant est composé du tampon TEAA 1M (0.01% TFA), et une solution TEAA/EtOH (50/50). Le débit est de 0,3 ml/min. Les longueurs d'onde choisies sont entre 260 et 575 nm. Le BigDye obtenu est purifié sur une colonne de silice phase inverse C18.

Techniques de caractérisation

- Les spectres RMN (^1H , ^{13}C) sont réalisés sur des appareils Brüker Avance à 300, 400 et 500 MHz. Les déplacements chimiques (δ) sont exprimés en ppm par rapport au triméthylsilane. Les constantes de couplage sont données en Hz. Les solvants utilisés sont le MeOH-d₄, le DMSO-d₆ et le D₂O
- Les spectres de masse sont effectués sur un appareil Brüker Daltonics MicroTOF LC et un appareil Brüker Daltonics autoflex II TOF/TOF en modes positif et négatif. Les fluorophores sont dissous soit dans du MeOH, soit dans du DMSO.
- Les mesures d'absorbance des fluorophores sont réalisées sur un spectrophotomètre « Nanodrop » en Visible. Les solutions de fluorophores sont dissoutes dans l'eau au préalable.
- Les expériences de fluorescence ont été réalisées dans des cellules en quartz à une température $20 \pm 0,5^\circ\text{C}$, sur un fluorimètre de type Fluoromax-4 (HORIBA Jobin-Yvonne). Tous les spectres d'émission ont été acquis à la longueur d'onde d'excitation de 488 nm. Cette longueur d'onde d'excitation choisie, correspond à la longueur d'onde d'excitation du laser des séquenceurs d'ADN. Les mesures ont été réalisées avec deux solvants différents : MeOH et tampon TEAA pH=6,5-7,0 (Triéthylamonium acétate). La concentration du BigDye choisie pour les mesures de fluorescence été de 10^{-4}M . Les spectres d'excitation et d'émission des deux entités fluorescentes donneur et accepteur ont été aussi réalisés dans les mêmes conditions.
- Les analyses de dichroïsme circulaire ont été effectuées dans des cellules en quartz, avec un spectropolarimètre de type « J-810 » de (Jasco). Les mesures de dichroïsme circulaire ont été acquises entre 190 à 800 nm. La vitesse de balayage a été de 50nm/min et chaque acquisition spectrale correspond à la moyenne de 25 scans.

II.4.2 Synthèses

(-5) et (-6) carboxy-fluorescéine méthanesulfonate

Nous avons utilisé du résorcinol ou « benzène-1,3-diol » (7g, 64 mmol, 2 eq) dans 64 ml d'acide méthanesulfonique. Du 1,2,4 benzène acide tricarboxylique » (6g, 32 mmol, 1 eq) est rajouté à la solution précédente. L'ensemble est chauffé à 85°C pendant 24h. Après 24h, on laisse le mélange réactionnel revenir à température ambiante, puis on le verse sur 30ml d'un mélange H₂O/glace. Ce mélange est placé à « 4°C » toute la nuit. Un précipité se forme, il est recueilli par filtration, puis séché 24h dans une étuve à 100°C. Masse obtenue (mélange 1)= 12g avec un rendement= 92%.

RMN ¹H (MeOH-d₄, 300 MHz) : mélange (1) : 8,56 (1H, s, H₄(isomère 5)) ; 8,35 (1H, d, J=8,1 Hz, H₄(isomère 6)) ; 8,27 (1H, d, J=8,0 Hz, H₆) ; 8,03 (1H, d, J=8,1 Hz, H₅) ; 7,72 (1H, s, H₇(isomère 6)) ; 7,26 (1H, d, J=8,0 Hz, H₇(isomère 5)) ; 2,68 (3H, s, CH₃ du méthanesulfonate).

5-carboxy-fluorescéine-méthanesulfonate

Le mélange (1) (m~12g) est dissous dans 260 ml d'éthanol absolu, puis on rajoute 70 ml d'hexane. La solution est mise au congélateur (-20°C) pendant 48h. Les cristaux sont recueillis par filtration et rincés avec de l'éthanol froid. L'isomère 5-carboxy-fluorescéine méthanesulfonate (composé 2, fig.49), m= 4,8g est obtenu avec un rendement de 40%.

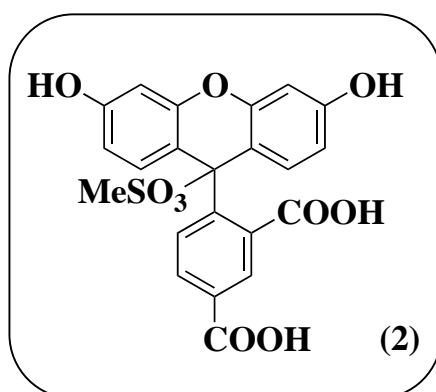


Figure 49: 5-carboxy-fluorescéine-méthanesulfonate/ C₂₂H₁₆O₁₀S/ Masse exacte : 472.05

Le composé 5-carboxy-fluorescéine méthanesulfonate est caractérisé par RMN ^1H dans du MeOH-d_4 et par spectrométrie de masse en mode négatif.

Composé (2): RMN ^1H (MeOH-d_4 , 400 MHz) : 8,90 (1H, s, H4) ; 8,46 (1H, d, $J=8,1$ Hz, H6) ; 7,57 (1H, d, $J=8,1$ Hz, H7) ; 7,39-7,09 (6H, m, H aromatiques du groupement xanthène) ; 2,68 (3H, s, CH_3 du méthanesulfonate). **ESI SM négatif (m/z) :** 471 $[\text{M-H}^+]$.

5-carboxy-fluorescéine

Le composé (2) est traité par une solution de NaOH (3M). Après 16h à température ambiante, une solution HCl (3M) est ajoutée jusqu'à arriver à un pH ~ 3 . Le précipité qui se forme est filtré et séché en réalisant des co-évaporations avec de l'EtOH absolu (Composé 3, fig.50). Le spectre RMN ^1H confirme que le groupement méthanesulfonate a disparu.

RMN ^1H (MeOH-d_4 , 400 MHz) : (Composé 3, fig.50) 8,56 (1H, s, H4) ; 8,35 (1H, d, $J=8,0$ Hz, H6) ; 7,26 (1H, d, $J=8,0$ Hz, H7) ; 6,68-6,52 (6H, m, H aromatiques du groupement xanthène). **ESI-SM positif (m/z) :** 377 $[\text{M+H}^+]$.

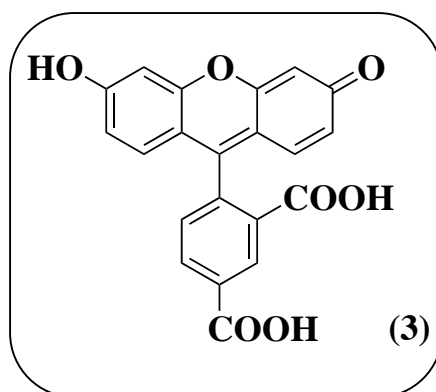


Figure 50: 5-carboxy-fluorescéine/ $\text{C}_{21}\text{H}_{12}\text{O}_7$ / Masse exacte : 376.06

6-carboxy-fluorescéine dipivalate

Le mélange (1) des deux isomères (-5) et (-6) carboxy-fluorescéine méthanesulfonate ($m=5$ g, 13 mmoles), est dissous dans de l'anhydride triméthylacétique (10 ml, 49 mmoles), puis chauffé à 200°C , sous atmosphère inerte. Après 4h on laisse le mélange revenir à température ambiante et (10 ml) de THF et (10 ml) de H_2O sont ajoutés, en laissant l'agitation à température ambiante pendant 2h. Le composé (4) est extrait avec une solution d'éther éthylique (50 ml). La phase organique est lavée (3x) par du tampon phosphate (25 ml, 1,4M, pH= 7.0), puis par une solution HCl (1M) (50 ml) et une solution de NaCl saturée. La phase organique est séchée avec du MgSO_4 , puis concentrée. Le sirop jaune obtenu est dilué dans

(50 ml) d'éthanol absolu, et de la diisopropylamine (5ml, 36 mmoles) est ajoutée. La solution est placée à (-20°C) pour précipiter l'isomère 6-carboxy-fluorescéine dipivalate diisopropylamine. Le précipité est filtré et rincé avec de l'éthanol froid puis avec de l'acétone froid et séché (Composé 4, fig.51). Masse obtenue= 2,39 g avec un rendement= 28%.

RMN ^1H (MeOH- d_4 , 300 MHz): (Composé 4, fig.51) : 8,23 (1H, d, $J=8,1\text{Hz}$, H4); 8,00 (1H, d, $J=8,1\text{Hz}$, H5); 7,72 (1H, s, H7); 7,17-7,14 (2H, m, H aromatiques du xanthène); 6,87-6,83 (4H, m, H aromatiques du xanthène); 3,45-3,39 (2H, m, $2\times\text{CH}_{\text{diisopropylamine}}$); 1,34 (18H, s, $6\times\text{CH}_3(\text{dipivalate})$); 1,24 (12H, d, $J= 6,35\text{ Hz}$, $4\times\text{CH}_3(\text{diisopropylamine})$). **RMN ^{13}C (MeOH- d_4 , 500 MHz) :** 176,7 ($\text{CO}_{(\text{diisopropylamine})}$); 169,2 ($\text{CO}_{(\text{dipivalate})}$); 151,4 (C-3); 145,5 (C-6); 131,1 (CH-4); 129,0 (CH-7); 126,7 (CH-5); 124,8 (CH-8'); 117,6 (CH-2'); 116,4 (CH-7'); 110,5 (CH-4'-5'); 46,3 ($2\times\text{CH}_{(\text{diisopropylamine})}$); 39,1 ($2\times\text{C}_{(\text{dipivalate})}$); 27,0 ($6\times\text{CH}_3(\text{dipivalate})$); 19,0 ($4\times\text{CH}_3(\text{diisopropylamine})$). **ESI SM, mode positif (m/z) :** 648 $[\text{M}+\text{H}^+]$.

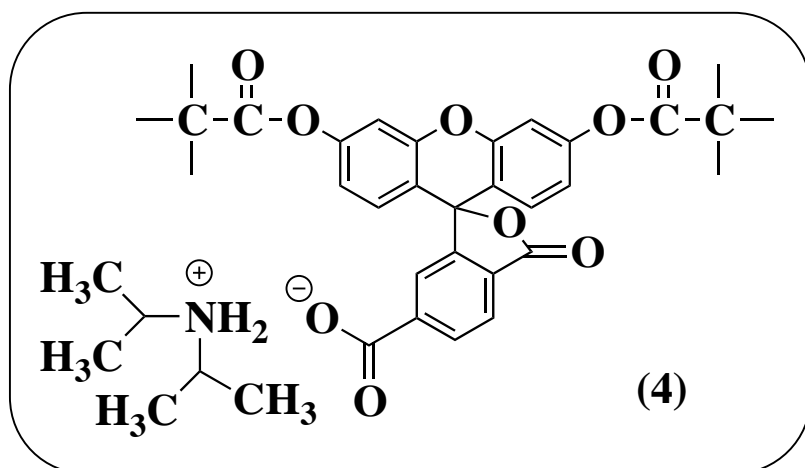


Figure 51: 6-carboxy-fluorescéine dipivalate diisopropylamine/ $\text{C}_{37}\text{H}_{45}\text{NO}_9$ / Masse exacte : 647.31

4'-[(prop-2-yn-1-ylamino)méthyl](5-) ou (6-)carboxy-fluorescéine

Le composé (-5) ou (-6) carboxy-fluorescéine méthanesulfonate (300mg, 0,64 mmoles) est mis en suspension dans l'eau (4 ml). La propargylamine (154 μl , 2.4 eq) est ajoutée au mélange réactionnel qui devient limpide. Le formaldéhyde (56 μl , 2.5 eq) est ajouté goutte à goutte. Le mélange est ensuite porté à 95°C, pendant 3h30'. Le mélange est concentré à sec, et le produit brut est purifié sur gel de silice avec un éluant isocratique (10%MeOH, 2,5% CH_3COOH , 87,5% CH_2Cl_2). Les fractions récoltées de la colonne correspondent aux composés mono- et di- substitués. Le composé monosubstitué est séparé du produit

disubstitué. Masse du composé 4'-[(prop-2-yn-1-ylamino)méthyl]5-carboxy-fluorescéine obtenu = 108 mg avec un rendement de 38%.

RMN ^1H (D₂O, 300 MHz) : Composé 4'-[(prop-2-yn-1-ylamino)méthyl]5-carboxy-fluorescéine: 8,41 (1H, s, H₄) ; 8,29 (1H, d, J= 8,0 Hz, H₆) ; 7,34 (1H, d, J= 8,0 Hz, H₇) ; 6,75 (1H, s, H_{5'}) ; 6,60-6,52 (4H, m, H_{1'}-H_{8'} et H_{2'}-H_{7'}) ; 4,14 (2H, s, CH₂) ; 3,46 (2H, s, CH₂) ; 3,14 (1H, s, CH). **ESI SM positif (m/z) :** 444 [M+H⁺].

RMN ^1H (DMSO-d₆+D₂O, 400 MHz) : Composé 4'-[(prop-2-yn-1-ylamino)méthyl]6-carboxy-fluorescéine: 8,10 (1H, d, J= 8,7 Hz, H₄) ; 7,91 (1H, d, J= 8,7 Hz) ; 7,48 (1H, s, H₇) ; 6,78 (1H, s, H_{5'}) ; 6,60-6,54 (4H, m, H_{1'}-H_{8'} et H_{2'}-H_{7'}) ; 4,29 (2H, s, CH₂) ; 3,73 (2H, s, CH₂) ; 3,35 (1H, s, CH). **RMN ^{13}C (DMSO, 500 MHz) :** 173,0 (CO_(COO⁻)) ; 169,66 (CO_(COOH)) ; 159,67 (C-6') ; 145,71 (CH-1') ; 131,28 (C-6) ; 129,71 (C-3) ; 129,51 (CH-4) ; 129,41 (CH-8') ; 127,25 (CH-5) ; 124,54 (CH-7) ; 113,66 (C-4') ; 112,68 (CH-2') ; 110,03 (CH-7') ; 107,34 (CH-5') ; 103,18 (C_(amine)) ; 77,95 (CH_(amine)) ; 48,95 (CH_{2(amine)}) ; 36,53 (CH_{2(amine)}). **ESI SM positif (m/z) :** 444 [M+H⁺].

Protocole général de synthèse du mélange (-5) et (-6) carboxy-rhodamines

L'anhydride 4-carboxyphtalique (0,062 M, 1 eq) et l'aminophénol (0,125 M, 2 eq) sont traités par de l'acide propionique (40 ml, pK_a= 4,82) avec une trace d'acide sulfurique 2M à reflux (180°C) pendant 10 à 12 h. Lorsque la CCM indique que tout le produit de départ a été consommé, le mélange réactionnel est extrait avec du CH₂Cl₂, et la phase organique est lavée (1x) avec de l'eau. La phase organique est séchée avec du MgSO₄ pour conduire au mélange brut des deux isomères de carboxy-rhodamines.

Un mélange brut de (-5) et (-6) carboxy-rhodamine-6G est purifié sur gel de silice avec un éluant isocratique de 45%EtOH/10%Et₃N/45%CHCl₃²³³. Afin d'éliminer la silice présente dans les fractions d'intérêt, celles-ci sont concentrées à sec, reprises dans du NaOH (1N), et la (-5) ou la (-6) carboxy-rhodamine est précipité avec du HCl (6N). La (-5) carboxy-rhodamine-6G (100 mg), composé (8) fig.52 et la (-6) carboxy-rhodamine-6G (170 mg), composé (9) fig.52 sont obtenus avec un rendement de 33% et 57% respectivement.

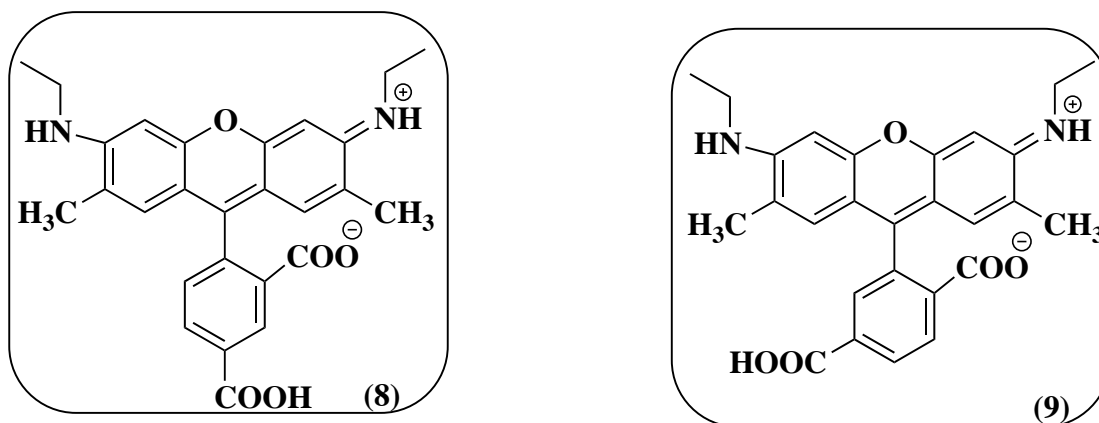
(-5) et (-6) carboxyrhodamine 6G

Figure 52: composé (8) : 5-carboxy-rhodamine-6G et composé (9) : 6-carboxy-rhodamine-6G/ C₂₇H₂₆N₂O₅/ Masse exacte : 458.18

RMN ¹H (DMSO-d₆, 400 MHz) : (Composé 8, fig.52) : 8,57 (1H, s, H4) ; 8,30 (1H, d, J= 8,3 Hz, H6) ; 8,24-8,20 (1H, m, NH) ; 7,43 (1H, d, J= 8,3 Hz, H7) ; 6,66-6,59 (5H, s, H aromatiques du xanthène + NH) ; 3,40 (4H, d, J= 6,6 Hz, 2x(CH₂)_{éthyle}) ; 2,01 (6H, s, 2xCH₃) ; 1,20 (6H, t, J= 6,6 Hz, 2x (CH₃)_{éthyle}). **ESI SM positif (m/z)** 459 [M+H⁺].

RMN ¹H (DMSO-d₆, 400 MHz) : (Composé 9, fig.52) : 8,24 (1H, d, J= 8,0 Hz, H4) ; 8,24-8,20 (2H, m, H5+NH) ; 7,72 (1H, s, H7) ; 6,66-6,59 (5H, 2xs, J= Hz, H aromatiques du xanthène + NH) ; 3,40 (4H, d, J= 6,6 Hz, 2x(CH₂)_{éthyle}) ; 2,16 (6H, s, 2xCH₃) ; 1,20 (6H, t, J= 6,6 Hz, 2x (CH₃)_{éthyle}). **ESI SM positif (m/z)** 459 [M+H⁺].

Un mélange brut de (-5) et (-6) carboxy-X-rhodamine est purifié sur gel de silice (colonne chromatographique flash) avec un gradient de MeOH (10 à 30%) dans le CH₂Cl₂. Les fractions correspondantes à la (-5) carboxy-X-rhodamine ou à la (-6) carboxy-X-rhodamine sont traitées par NaOH (1N) et HCl (6N) afin d'éliminer la silice. La (-5) carboxy-X-rhodamine (14 mg), composé (10) fig.53 et la (-6) carboxy-X-rhodamine (28 mg), composé (11) fig.53 sont obtenus avec un rendement de 10% et 20% respectivement.

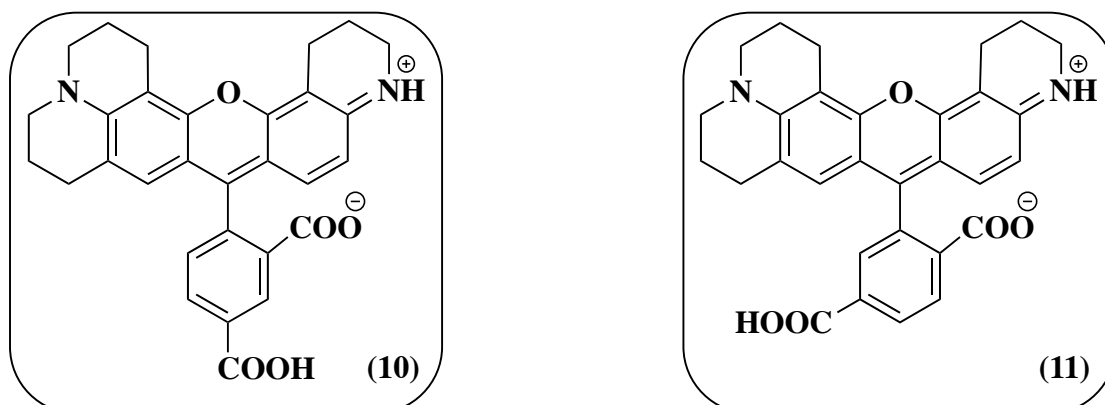
(-5) et (-6) carboxy-X-rhodamine

Figure 53: Composé (10) : 5-carboxy-X-rhodamine et composé (11) : 6-carboxy-X-rhodamine/ $C_{33}H_{30}N_2O_5$ / Masse exacte : 534.22

RMN 1H (DMSO- d_6 , 400 MHz) : (Composé 10, fig.53) : 8,28 (1H, s, H4) ; 8,18 (1H, d, J= 7,9 Hz, H6) ; 7,06 (1H, d, J= 7,9 Hz, H7) ; 6,03 (2H, s, H1'-H8') ; 3,16-3,12 (8H, m, H aromatiques) ; 3,11-3,10 (4H, m, H aromatiques) ; 2,86-2,83 (4H, m, H aromatiques) ; 1,96-1,95 (4H, m, H aromatiques) ; 1,78-1,75 (4H, m, H aromatiques). **ESI SM positif (m/z) :** 535 $[M+H^+]$.

RMN 1H (DMSO- d_6 , 400 MHz) : (Composé 11, fig.53) : 8,03 (1H, d, J= 7,9 Hz, H4) ; 7,77 (1H, d, J= 7,9 Hz, H5) ; 7,43 (1H, s, H7) ; 6,03 (2H, s, H1'-H8') ; 3,16-3,13 (8H, m, H aromatiques) ; 3,12-3,11 (4H, m, H aromatiques) ; 2,88-2,85 (4H, m, H aromatiques) ; 1,97-1,94 (4H, m, H aromatiques) ; 1,79-1,76 (4H, m, H aromatiques). **ESI SM positif (m/z) :** 535 $[M+H^+]$.

Un mélange brut de (-5) et (-6) carboxy-TAMRA est purifié sur gel de silice avec un gradient de MeOH (10 à 30%). Les fractions correspondantes à la (-5) ou (-6) carboxy-TAMRA sont traitées par NaOH (1N) puis HCl (6N) afin d'éliminer la silice. La (-5) carboxy-TAMRA (87 mg) composé (12) fig.54 et la (-6) carboxy-TAMRA (69 mg) composé (13) fig.54 sont obtenus avec un rendement de 19% et 15% respectivement.

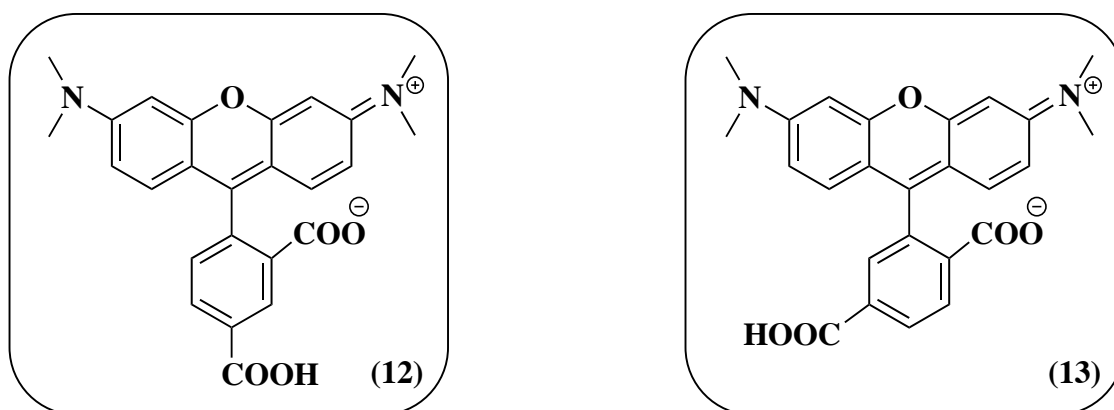
(-5) et (-6) carboxy-tétraméthyl-rhodamine

Figure 54: Composé (12) : 5-carboxy-tétraméthyl-rhodamine et composé (13) : 6-carboxy-tétraméthyl-rhodamine/ $C_{25}H_{22}N_2O_5$ / Masse exacte : 430.15

RMN 1H (DMSO- d_6 , 400 MHz) : (Composé 12, fig.54) : 8,38 (1H, s, H4) ; 8,27 (1H, d, J= 8,1 Hz, H6) ; 7,30 (1H, d, J= 8,1 Hz, H7) ; 6,54-6,47 (6H, m, H aromatiques du xanthène) ; 2,96 (12H, s, 4xCH₃). **ESI SM positif (m/z) :** 431 [M+H⁺].

RMN 1H (DMSO- d_6 , 400 MHz) : (Composé 13, fig.54) : 8,18 (1H, d, J= 8,1 Hz, H4) ; 8,03 (1H, d, J= 8,1 Hz, H5) ; 7,56 (1H, s, H7) ; 6,54-6,47 (6H, m, H aromatiques) ; 2,94 (12H, s, 4xCH₃). **ESI SM positif (m/z) :** 431 [M+H⁺].

3-Azido-propylamine

Le bromure de propargylamine (3,2 g, 14,6 mmoles) est dissous dans l'eau (10 ml) puis l'azoture de sodium (3,2 g, 49,2 mmoles) dissous dans de l'eau (15 ml) est ajouté. Après 48h de réaction à 110°C la réaction est terminée²³⁷. Le mélange réactionnel est concentré et co-évaporé (3x) avec de la pyridine anhydre. Il est nécessaire de faire attention à ne pas laisser sécher l'amine « risque d'explosion ». Le produit final est obtenu sous forme d'un liquide visqueux qui sera utilisé sans autre purification.

Protocole général de synthèse des dérivés N-(3-azido-propyl)carboxamide-rhodamine

La carboxy-rhodamine (0,03M, 1 eq) est mise en solution dans du DMF anhydre (3 ml) puis on ajoute de la 4-diméthylaminopyridine (0,106 M, 2 eq) et du N,N'-di-succinimidyl carbonate (0,103 M, 1,5 eq). La réaction est effectuée à température ambiante sous atmosphère inerte. La deuxième étape consiste à substituer avec l'amine, 3-azido propylamine en présence de la triéthylamine anhydre. Après 6h de réaction, la 3-azido-propylamine (100

mg, 1 mmol) est ajoutée avec de la triéthylamine anhydre (300 μ l), et la réaction est poursuivie à température ambiante pendant 24h supplémentaires. Le brut réactionnel est extrait avec du CH_2Cl_2 (10 ml), la phase organique est lavée avec de l'eau 3x (10 ml), puis séchée avec du MgSO_4 avant d'être purifiée sur colonne de gel de silice.

6-(N-(3-azido-propyl))carboxamide-R6G

Purification sur gel de silice avec un éluant isocratique de 10% MeOH dans 90% de CH_2Cl_2 . Masse obtenue = 65 mg avec un rendement de 70%.

RMN ^1H (DMSO- d_6 , 400 MHz) : (Composé 15, fig.55) : 8,73 (1H, t, $J = 5,9$ Hz, NH) ; 8,13 (1H, d, $J = 8,0$ Hz, H4) ; 8,04 (1H, d, $J = 8,0$ Hz, H5) ; 7,58 (1H, s, H7) ; 6,34-6,25 (4H, 2s, H aromatiques du xanthène) ; 5,33 (2H, t, $J = 5,1$ Hz, 2xNH) ; 3,28-3,14 (8H, m, $(2x\text{CH}_2)_{\text{éthyle}} + 2x\text{CH}_2 (\alpha+\beta)$) ; 1,89 (6H, s, 2x CH_3) ; 1,70 (2H, t, $J = 6,9$ Hz, $\text{CH}_2 (\gamma)$) ; 1,20 (6H, t, $J = 6,6$ Hz, 2x $\text{CH}_3(\text{éthyle})$). **RMN ^1H (DMSO- d_6 + D_2O , 400 MHz) :** 8,04 (2H, s, H4+H5) ; 7,52 (1H, s, H7) ; 6,59-6,52 (4H, 2s, H aromatiques du xanthène) ; 3,33-3,24 (8H, m, $(2x\text{CH}_2)_{\text{éthyle}} + 2x\text{CH}_2 (\alpha+\beta)$) ; 1,94 (6H, s, 2x CH_3) ; 1,73-1,69 (2H, t, $J = 6,9$ Hz, $\text{CH}_2(\gamma)$) ; 1,21- 1,17 (6H, t, $J = 6,6$ Hz, 2x $\text{CH}_3(\text{éthyle})$). **RMN ^{13}C (DMSO- d_6 , 500 MHz) :** 168,64 ($\text{CO}(\text{COO}^-)$) ; 166,07 ($\text{CO}(\text{amide})$) ; 152,67 (C-6') ; 142,57 (C-6) ; 137,40(C-3') ; 137,25 (C-3) ; 128,96 (CH-4) ; 128,12 (CH-5) ; 125,4 (CH-7) ; 122,14 (CH-1'-8') ; 109,31 (C-2'-7') ; 94,74 (CH-4'-5') ; 49,07 ($\text{CH}_2\text{-}\gamma\text{-3-azido propylamine}$) ; 38,05 ($\text{CH}_2\text{-}\alpha\text{-3-azido propylamine}$) ; 37,36 (2x $\text{CH}_3(\text{éthyle})$) ; 28,60 ($\text{CH}_2\text{-}\beta\text{-3-azido propylamine}$) ; 17,64 (2x CH_3) ; 14,26 ($2x\text{CH}_2(\text{éthyle})$). **ESI SM positif (m/z) :** 541 $[\text{M}+\text{H}^+]$.

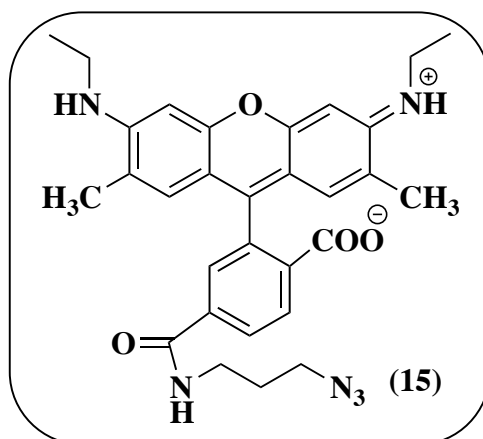


Figure 55: Composé (15): 6-N-(3-azido-propyl)carboxamide-R6G/ $\text{C}_{30}\text{H}_{32}\text{N}_6\text{O}_4$ / Masse exacte : 540.25

5-(N-(3-azido-propyl))carboxamide-ROX

Purification sur gel de silice avec un éluant isocratique de 10% MeOH dans 90% CH₂Cl₂.
Masse obtenue = 100 mg avec un rendement de 35%.

RMN ¹H (DMSO-d₆, 400 MHz) : (Composé 16, fig. 56) 8,85 (1H, t, J= 5,0 Hz, NH) ; 8,45 (1H, s, H4) ; 8,12 (1H, d, J= 8,1 Hz, H6) ; 7,27 (1H, d, J= 8,1 Hz, H7) ; 6,21 (2H, s, H1'-H8') ; 3,5-3,0 (12H, m, 2x(4H_{aromatiques})+ 2xCH₂ (α+β)) ; 2,89 (4H, m, H aromatiques) ; 1,96-1,93 (4H, m, H aromatiques) ; 1,83-1,81 (6H, m, 4H aromatiques + CH₂ (γ)). **RMN ¹H (DMSO-d₆+ D₂O, 400 MHz) :** 8,41 (1H, s, H4) ; 7,93 (1H, d, J= 7,8 Hz, H6) ; 7,17 (1H, d, J= 7,8 Hz, H7) ; 6,52 (2H, s, H1'-H8') ; 3,43-3,34 (12H, m, 2x(4H_{aromatiques}) + 2xCH₂ (α+β)) ; 2,93 (4H, m, H aromatiques) ; 2,56-2,55 (4H, m, H aromatiques) ; 1,96 (4H, m, H aromatiques) ; 1,81-1,78 (6H, t, J= 6,2 Hz, 4H aromatiques + CH₂ (γ)). **RMN ¹³C (DMSO-d₆, 500 MHz) :** 168,19 (CO_(COO⁻)) ; 166,88 (CO_(amide)) ; 156 (C-6') ; 151 (C-10') ; 150,60 (C11') ; 140,10 (C-3) ; 136,73 (C-5) ; 135,38 (CH-6) ; 129,72 (CH-4) ; 129,01 (CH-7) ; 128,53 (C-2) ; 126,34 (CH-1'-8') ; 123,21 (CH-2'-7') ; 112,31 (C-5') ; 105,12 (C-4') ; 50,60 (CH₂-γ-3-azido propylamine) ; 50,13 (2x CH₂₍₉₋₁₀₎ aromatiques) ; 49,07 (2xCH₂₍₃₋₄₎ aromatiques) ; 37,23 (CH₂-α-3-azido propylamine) ; 28,78 (CH₂-β-3-azido propylamine) ; 27,40 (CH₂₍₁₋₁₂₎ aromatiques) ; 21,53 (CH₂₍₂₋₅₎ aromatiques) ; 20,68 (CH₂₍₇₋₈₎) ; 19,93 (CH₂₍₆₎) ; 19,78 (CH₂₍₁₁₎). **ESI SM positif (m/z) :** 617 [M+H⁺].

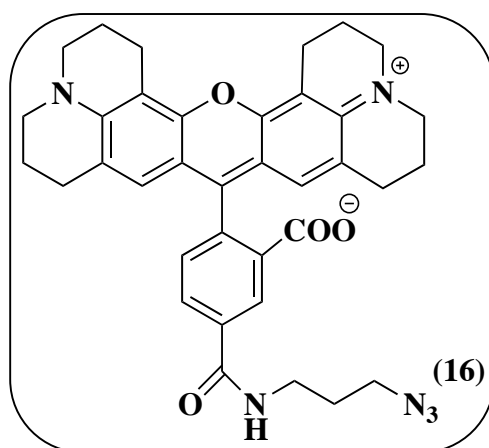


Figure 56: Composé (16): 5-N-(3-azido-propyl)carboxamide-ROX/ C₃₆H₃₆N₆O₄/ Masse exacte: 616.28

5-(N-(3-azido-propyl))carboxamide-TAMRA

Purification sur gel de silice avec un éluant isocratique de 10% MeOH dans 90% de CH₂Cl₂.
Masse obtenue = 81 mg avec un rendement de 44%.

RMN ^1H (DMSO- d_6 , 400 MHz) : (Composé 17, fig.57) 8,89 (1H, t, $J = 5,5$ Hz, NH) ; 8,44 (1H, s, H4) ; 8,22 (1H, d, $J = 8,0$ Hz, H6) ; 7,31 (1H, d, $J = 8,0$ Hz, H7) ; 6,52-6,46 (6H, m, H aromatiques) ; 3,40-3,30 (4H, m, $2 \times \text{CH}_2$ ($\alpha + \beta$)) ; 2,94 (12H, s, $4 \times \text{CH}_3$) ; 1,82 (2H, t, $J = 6,8$ Hz, $\text{CH}_2(\gamma)$). **RMN ^1H (DMSO- $d_6 + \text{D}_2\text{O}$, 400 MHz) :** 8,37 (1H, s, H4) ; 8,16 (1H, d, $J = 7,9$ Hz, H6) ; 7,26 (1H, d, $J = 7,9$ Hz, H7) ; 6,51-6,50 (6H, m, H aromatiques) ; 3,41-3,34 (4H, m, $2 \times \text{CH}_2$ ($\alpha + \beta$)-3-azido propylamine) ; 2,92 (12H, s, $4 \times \text{CH}_3$) ; 1,81-1,78 (2H, t, $J = 6,8$ Hz, CH_2 - γ -3-azido propylamine). **RMN ^{13}C (DMSO- d_6 , 500 MHz) :** 174 ($\text{CO}_{(\text{COO}^-)}$) ; 168,34 ($\text{CO}_{(\text{amide})}$) ; 152,31 (C-6') ; 134,50 (C-3') ; 129,1 (C-3) ; 128,37 (C-5) ; 126,82 (CH-6) ; 124,15 (CH-7) ; 123,14 (CH-4) ; 108,99 (CH-2'-7') ; 105,55 (CH-1'-8') ; 97,98 (CH-5') ; 84,71 (CH-4') ; 48,51 (CH_2 - γ -3-azido propylamine) ; 36,82 (CH_2 - α -3-azido propylamine) ; 28,26 (CH_2 - β -3-azido propylamine). **ESI SM positif (m/z) :** 513 [$\text{M} + \text{H}^+$].

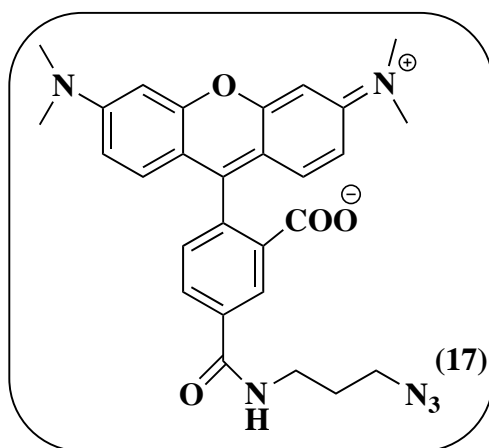


Figure 57: Composé (17) : 5-(N-(3-azido-propyl))carboxamide-TAMRA/ $\text{C}_{28}\text{H}_{28}\text{N}_6\text{O}_4$ /
Masse exacte: 512.22

Protocole général de synthèse du bras espaceur : « click chemistry » ou cycloaddition de Huisgen

La 4'-[(prop-2-yn-1-ylamino)méthyl]6-carboxy-fluorescéine (11 mg, 0,025 mmoles, 1 eq) et la 5-(N-(3-azido-propyl))carboxamide-ROX (20 mg, 0,032 mmoles, 1,3 eq) sont mises en solution dans 1,8 ml ($\text{H}_2\text{O}/\text{tBuOH}/\text{DMSO}$). On ajoute ensuite des solutions fraîchement préparées et dégazées avec de l'argon de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, d'ascorbate de sodium et du TBTA. Le $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (0,3 mg, 0,00125 mmoles, 0,05 eq) correspondant à 3mg dans 50 μl H_2O dont on prélève 5 μl pour la réaction. L'ascorbate de sodium (0,5 mg, 0,0025 mmoles, 0,1 eq), correspondant à 5mg dissout dans 400 μl H_2O dont on prélève 40 μl pour la réaction, et du Tris[(1-benzyl-1H-1,2,3-triazole-4-yl)méthyl]amine (TBTA) (0,66 mg, 0,00125 mmoles, 0,005 équivalent) dans DMSO/tBuOH (3 :1), correspondant à 3,3 mg qu'on dissout dans 25 μl

(DMSO/tBuOH), dont on prélève 5 μ l pour la réaction. Les deux volumes de cuivre et de TBTA sont additionnés et rajoutés en même temps au début de la réaction. On laisse la réaction tourner pendant 24h à température ambiante, en suivant l'évolution avec des plaques CCM. Après 24h, le mélange réactionnel est extrait avec du CH₂Cl₂. La phase organique est lavée (1x) par H₂O, puis purifiée sur HPLC en utilisant une colonne de silice phase inverse (C18). On prélève un volume de cette solution correspondant à une concentration de 25 μ M, et on rajoute un volume de la solution tampon (TEAA-TFA) pour rendre toutes les conditions de la purification adéquates (1 :1). Le volume injecté dans la colonne est de 100 μ l. L'élution se fait avec un gradient d'EtOH (10%-80%) contre TEAA-0,01%TFA (1M) pendant 60 min à 0,3 ml/min. Les fractions contenant le produit attendu sont identifiées par spectrométrie de masse puis concentrées à sec. **ESI mode positif (m/z) : 1061 [M+H⁺]**.

II.5 RÉSULTATS

II.5.1 Synthèse des dérivés de la fluorescéine modifiés

Le protocole se divise en deux étapes, la première étape consiste à synthétiser les deux dérivés de position (5- et 6-) carboxy-fluorescéine par une réaction d'acylation/cyclodéshydratation de type Friedel et Crafts. Cette réaction est une catalyse acide en présence de l'acide méthanesulfonique qui donne un mélange des deux isomères 5- et 6- carboxy-fluorescéine méthanesulfonate. Il était possible d'enlever la fonction méthanesulfonate par traitement avec du NaOH et précipitation par HCl (Fig.58).

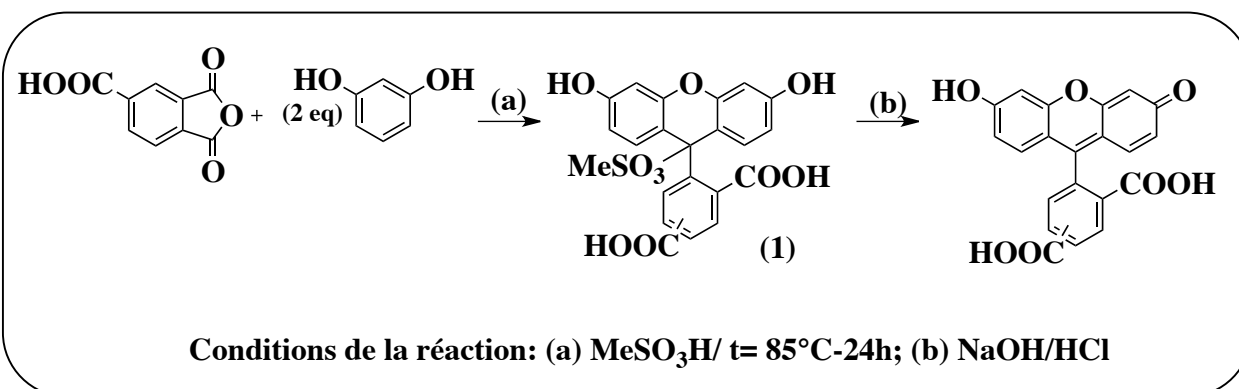


Figure 58: Synthèse du mélange (1) des deux isomères (-5) et (-6) carboxy-fluorescéine.

La deuxième étape de la synthèse consiste à modifier ces dérivés avec la fonction prop-2-yn-ylamino méthyle en position 4' du xanthène par la substitution de Mannich (Fig.59).

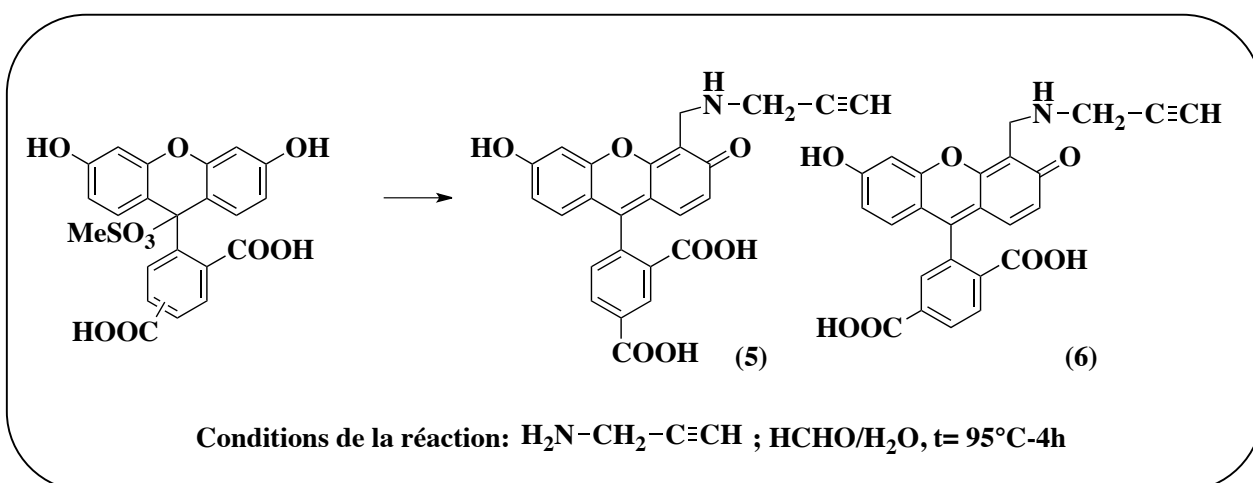


Figure 59: Synthèse du composé 4'-[(prop-2-yn-1-ylamino)méthyl](5-) ou (6-) carboxy-fluorescéine (Composés 5 et 6)

II.5.2 Synthèse de la Carboxy-rhodamine et de ses dérivés modifiés

Le protocole se divise en deux étapes, la première étape consiste à synthétiser les deux isomères de position (5- et 6-) des quatre dérivés de la carboxy-rhodamine. Cette réaction est une catalyse acide en présence de l'acide propionique avec des traces d'acide sulfurique. Le résultat de cette réaction est un mélange des deux isomères de position 5- et 6- carboxy-rhodamine (Fig.60).

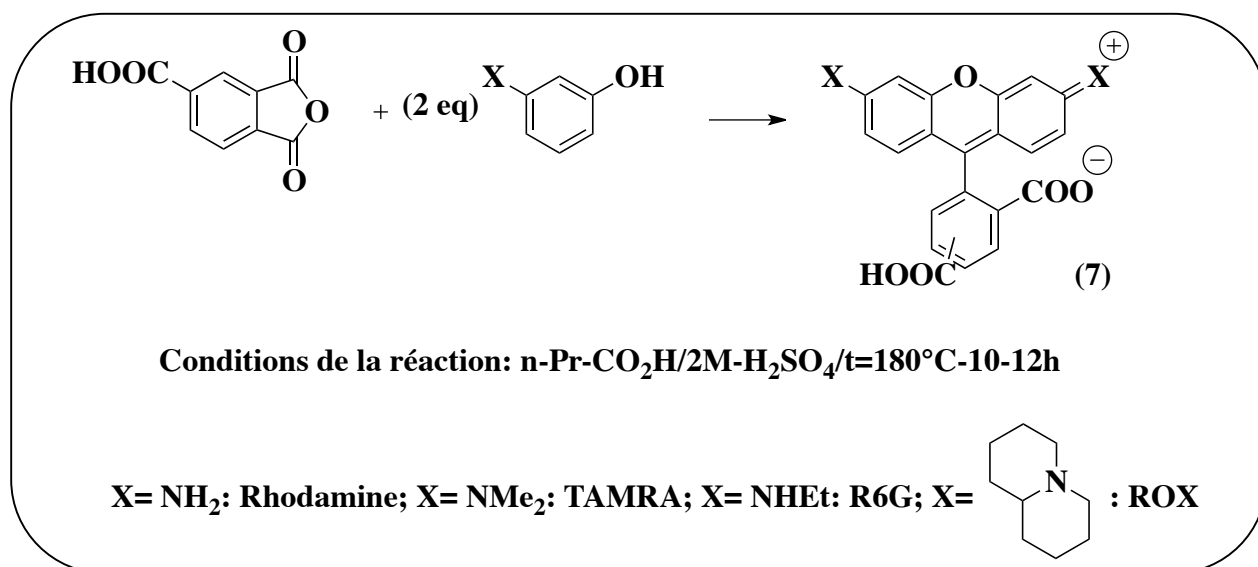
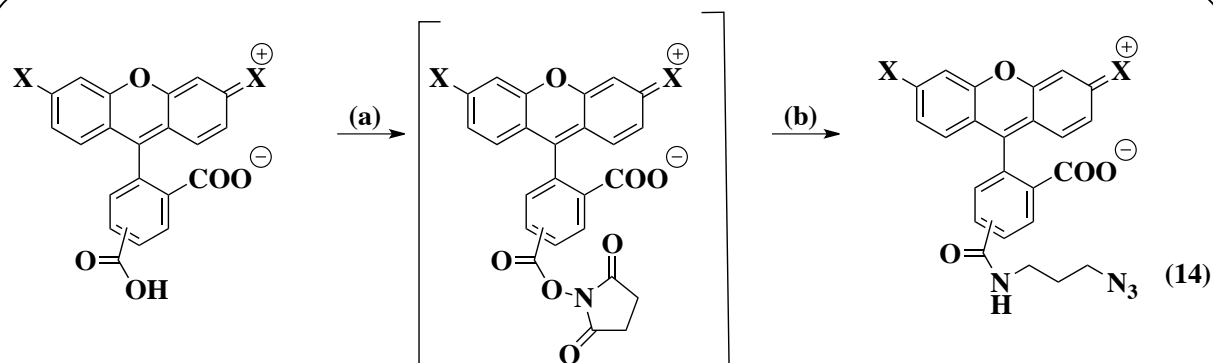


Figure 60: Protocole générale de synthèse du mélange (7) des deux isomères de position de la carboxy-rhodamine et de ces dérivés.

La deuxième étape consiste à modifier les dérivés de la carboxy-rhodamine avec la fonction azoture. Cette étape consiste à activer la fonction acide carboxylique en présence du N,N'-Disuccinimidyl carbonate et de la diméthylaminopyridine pour former l'ester actif. Une fois l'ester actif est formé, la substitution avec le groupement 3-azido propylamine se fait en présence de la triéthylamine (Fig.61).



Conditions de la réaction: (a) DSC, 4-DMAP/DMF_(anh), 6h/température ambiante; (b) 3-azido-propylamine/Et₃N_(anh)/DMF_(anh), 24h/température ambiante

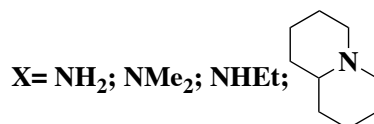


Figure 61: Protocole général de la synthèse des dérivés de la rhodamine modifiés avec la fonction azoture au niveau de la fonction acide carboxylique (mélange 14). Aucune étape de purification intermédiaire entre les deux étapes (a) et (b). Les deux étapes se font sous atmosphère inerte.

II.5.3 Synthèse du bras espaceur : couplage de la 6-FAM-alcyne avec la 5-ROX-azoture

Le premier couplage a été réalisé entre les deux entités donneur « 6-FAM-alcyne » et accepteur 5-ROX-azoture (Fig.62). Cette réaction de couplage est une réaction de click chemistry ou « cycloaddition de Huisgen ». Cette réaction consiste à coupler une fonction azoture avec une fonction alcyne en présence du cuivre sous forme de Cu(I). Le Cu(II) utilisé à partir du CuSO₄.5H₂O est réduit en utilisant l'agent réducteur du cuivre, l'ascorbate de sodium et en présence de l'agent chélatant du cuivre, le TBTA.

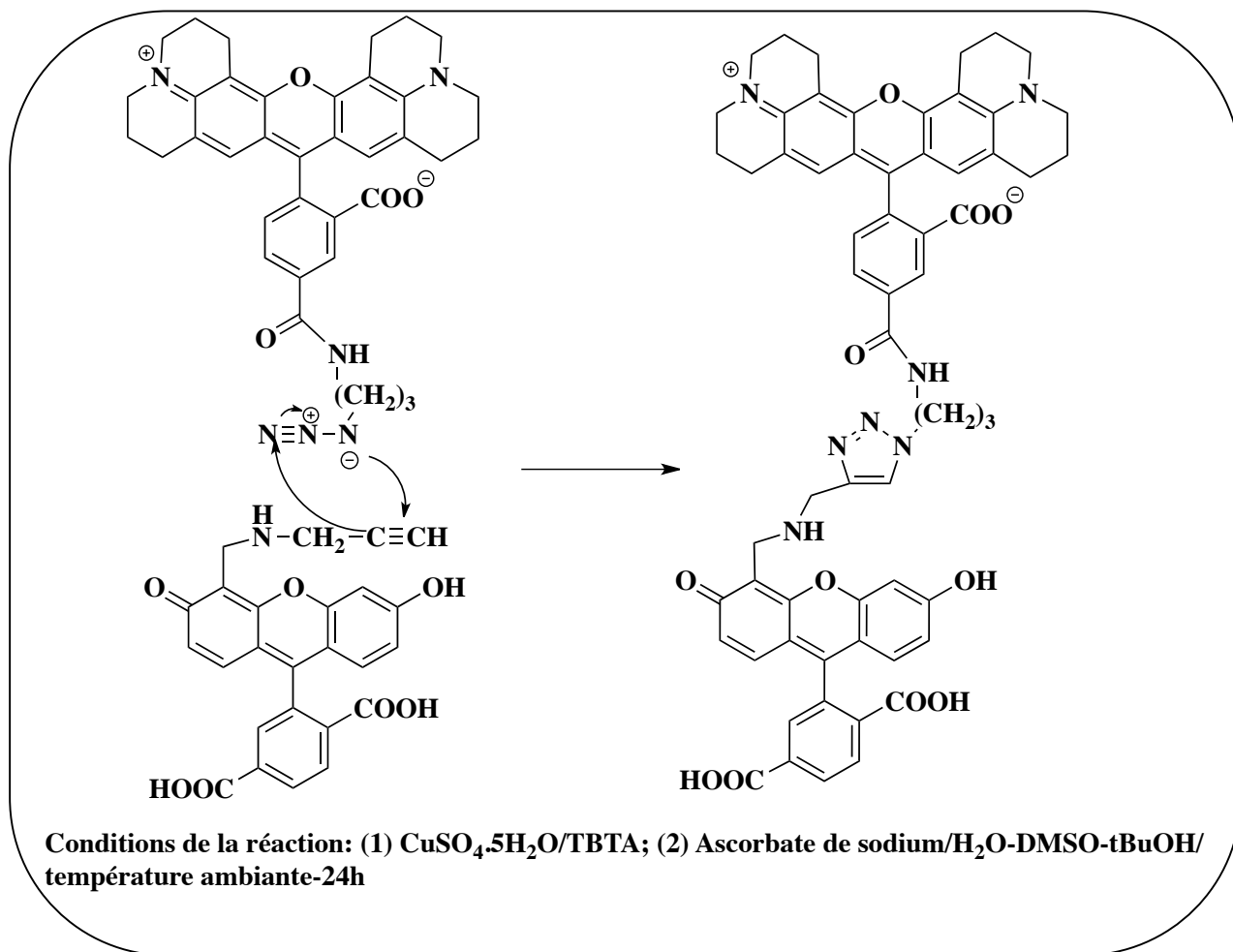


Figure 62: Synthèse du bras espaceur entre les deux entités fluorescentes par la réaction de Click Chemistry.

II.5.4 Analyse du BigDye par spectroscopie de fluorescence

Notre molécule de BigDye est constituée des deux sondes fluorescentes la 4'-[(prop-2-yn-1-ylamino)méthyle]6-carboxy-fluorescéine (6-FAM-alcyne) et la 5-(N-(3-azido-propyl))carboxamide-ROX (5-ROX-azide) séparées par un bras espaceur contenant un cycle de triazole. Deux spectres d'émission de fluorescence du BigDye dans deux solvants différents, du tampon TEAA et du MeOH en excitant à 488 nm ont été pris. Nos analyses par spectroscopie de fluorescence indiquent que le solvant et sa polarité influencent fortement les propriétés physico-chimiques du BigDye, puisqu'il existe une différence dans les spectres d'émission de fluorescence dans les deux solvants utilisés (Fig.63). On observe que dans le MeOH, le BigDye possède une bande à 600nm qui correspond à la longueur d'onde d'émission de l'accepteur de fluorescence (5-ROX-azide) et que dans le tampon TEAA,

l'émission fluorescente du BigDye est caractérisée par deux pics presque de même taille à ~520nm et ~600nm correspondant respectivement aux deux longueurs d'onde d'émission du donneur de fluorescence (6-FAM-alcyne) et de l'accepteur de fluorescence (5-ROX-azide).

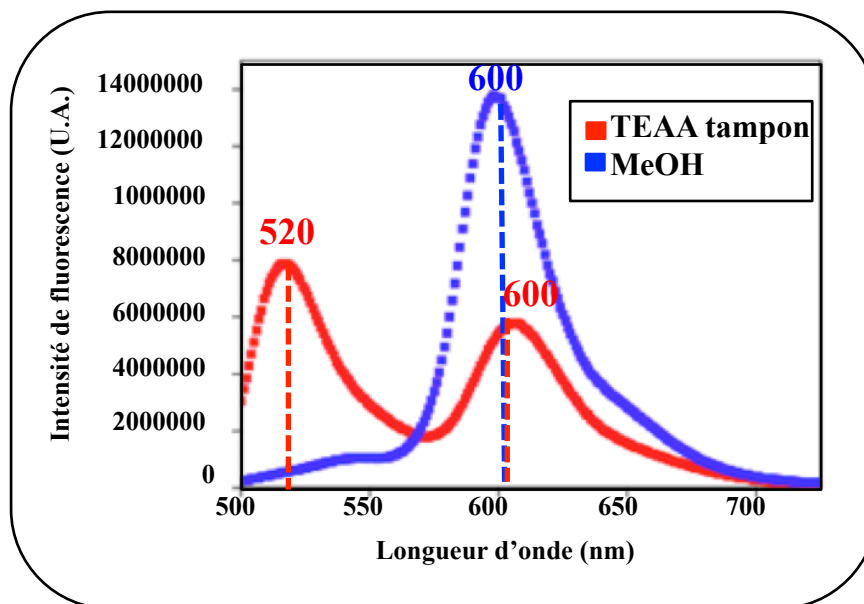


Figure 63: Différence spectrale du BigDye dans les deux solvants, le tampon TEAA (en rouge) et le MeOH (bleu) en excitant à la même longueur d'onde $\lambda_{exc} = 488\text{nm}$.

II.5.5 Étude de modélisation du BigDye

La modélisation du BigDye a été réalisée en utilisant le logiciel « Accelrys / Discovery Studio (3.5) » par Eric ENNIFAR de notre laboratoire, ce qui nous a permis de modéliser les deux configurations extrêmes possibles (Fig.64). Dans le premier modèle déterminé, la flexibilité du bras espaceur permettrait aux sondes de se superposer (stacking) et dans ce cas, la distance intramoléculaire serait de 6,1 Å qui est une distance typique du couplage excitonique (modèle B, fig.64). Dans le deuxième modèle, les deux sondes fluorescentes sont éloignées l'une de l'autre et la distance entre les deux molécules fluorescentes est dans ce cas de 13 Å qui est une distance typique du FRET (modèle A, fig.64), compte tenu de la longueur et de la rigidité du bras espaceur.

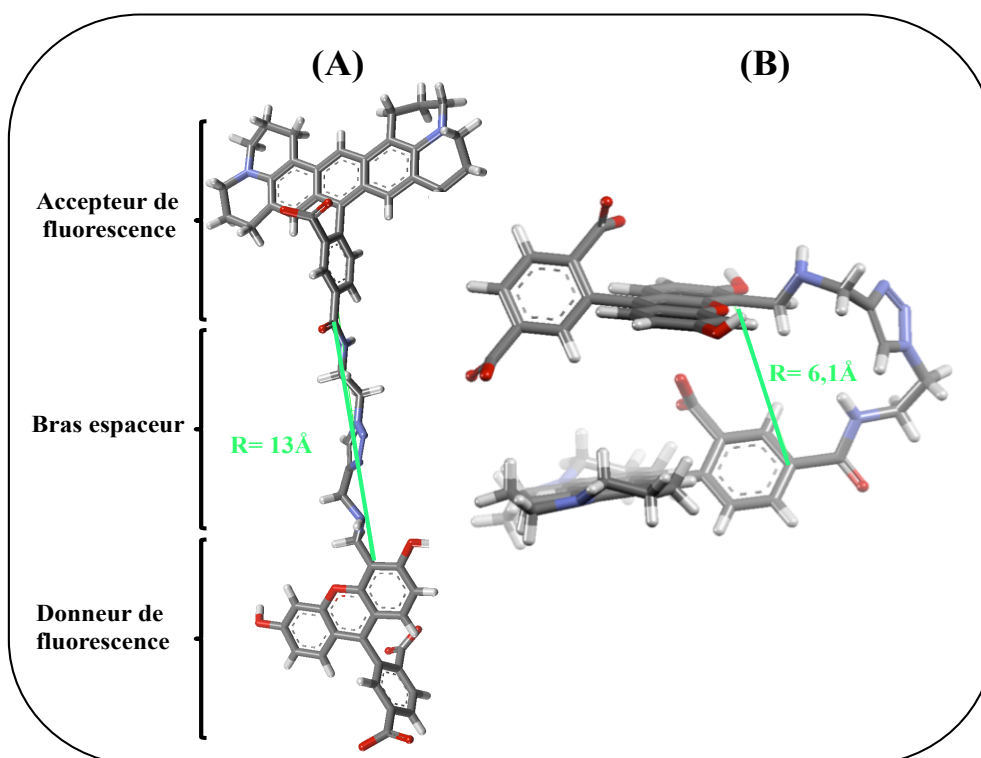


Figure 64 : Étude de modélisation par le logiciel « Accelrys / Discovery Studio (3.5) » du BigDye, en tenant compte des contraintes géométriques (longueurs et angles de liaisons) de la molécule. Deux géométries extrêmes sont prises en compte. Le modèle (A) représente la configuration pour laquelle les deux sondes sont éloignées avec une distance de $R = 13 \text{ \AA}$, et le modèle (B) représente la configuration pour laquelle les deux sondes sont superposées avec une distance $R = 6,1 \text{ \AA}$.

II.5.6 Caractérisation des interactions entre les sondes du BigDye

II.5.6.1 Le couplage excitonique

La théorie de l'exciton se base sur la formation d'un hétérodimère intramoléculaire qui s'établit entre deux chromophores, ce qui permet la formation d'un couplage fort entre les deux moments dipolaires de transition. La formation d'un hétérodimère dépend principalement du solvant et des chromophores (intensité du moment dipolaire de transition, hydrophobicité, et charge électrique du chromophore). Pour un système chromophore/solvant donné, différentes interactions entrent en jeu dans la formation du dimère. Il s'agit notamment des interactions de type hydrophobe, de liaisons hydrogène entre les chromophores ainsi qu'avec le solvant, et de type Van der Waals²³⁸.

La théorie de l'exciton prévoit que l'état excité du monomère subit une dégénérescence suite à la formation du dimère (Fig.65). Dans le cas particulier de deux chromophores identiques

(homodimère), la formation du dimère se manifeste par l'apparition de deux niveaux dégénérés également espacés par rapport à l'état excité du monomère (Fig.65.A), ce qui n'est pas le cas lors d'une dimérisation entre deux chromophores différents (hétérodimère), ce qui est notre cas (Fig.65.B).

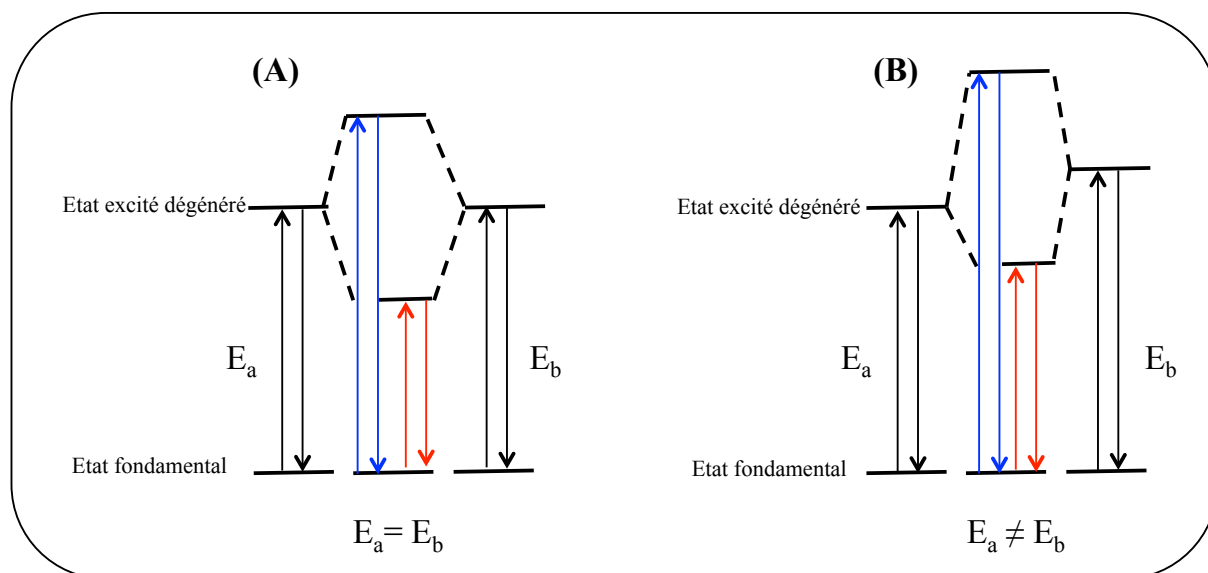


Figure 65 : La théorie du couplage excitonique prévoit que l'état excité du monomère subit une dégénérescence suite à la formation du dimère. (A) représente le diagramme d'énergie dans le cas d'un homodimère. E_a et E_b représentent l'énergie de transition entre l'état fondamental et l'état excité des deux chromophores. Dans le cas d'un hétérodimère, dont le diagramme d'énergie est représenté par (B), les deux états excités dégénérés E_a et E_b ne sont pas également espacés par rapport au niveau excité du monomère.

Suite à la dimérisation des chromophores, les bandes spectrales observées dépendent fortement de la géométrie du système. En effet, la dégénérescence de l'état excité est une conséquence des arrangements possibles des dipôles de chaque chromophore. Ce mécanisme entraîne une délocalisation de l'excitation sur les deux sondes. Par conséquent les propriétés spectrales ne peuvent donc plus être attribuées aux molécules simples, mais au dimère^{239,240}.

Dans le but d'analyser le dimère de chromophores formé par les deux sondes fluorescentes 6-FAM-alcyne et 5-ROX-azide dans le BigDye, nous avons réalisé des expériences de dichroïsme circulaire. Nous avons ainsi comparé les propriétés spectrales du « BigDye » (dimère) et de la « 6-FAM-alcyne » (monomère) dans deux milieux différents, un tampon TEAA (pH=7.0, 1.0 M) et dans du MeOH.

Dans le cas d'un dimère de pyridine six conformations sont possibles, en tenant compte des orientations des moments de dipôle de transition des monomères¹⁶⁵ (Fig.66). Les deux

modèles (A) et (C) correspondent à deux monomères de pyridine orientés dans une conformation colinéaire dans le même sens (A), ou colinéaire dans le sens opposé (C). Ce dimère se manifeste par l'apparition d'une bande dans la zone de fréquence la plus faible par rapport à la bande du monomère (qui équivaut à un décalage de la bande dimérique vers les longueurs d'onde les plus élevées, ou *red-shift*). Les deux configurations (B) et (D) coïncident avec deux monomères de pyridine orientés de manière parallèle (B) ou antiparallèle (D), et ceci correspond à la formation d'une bande du dimère dans la zone de fréquence élevée (longueurs d'onde plus faibles, ou *blue-shift*). Le modèle (E) correspond à une superposition des bandes du dimère et du monomère, et dans ce cas les deux monomères de pyridine sont alors empilés et orientés dans une conformation sécante. Enfin le modèle (F) correspond à la formation de deux bandes espacées de manière équivalente par rapport à la fréquence du monomère, dans cette situation les deux monomères sont orientés dans une conformation sécante dans un sens opposé sans être empilés.

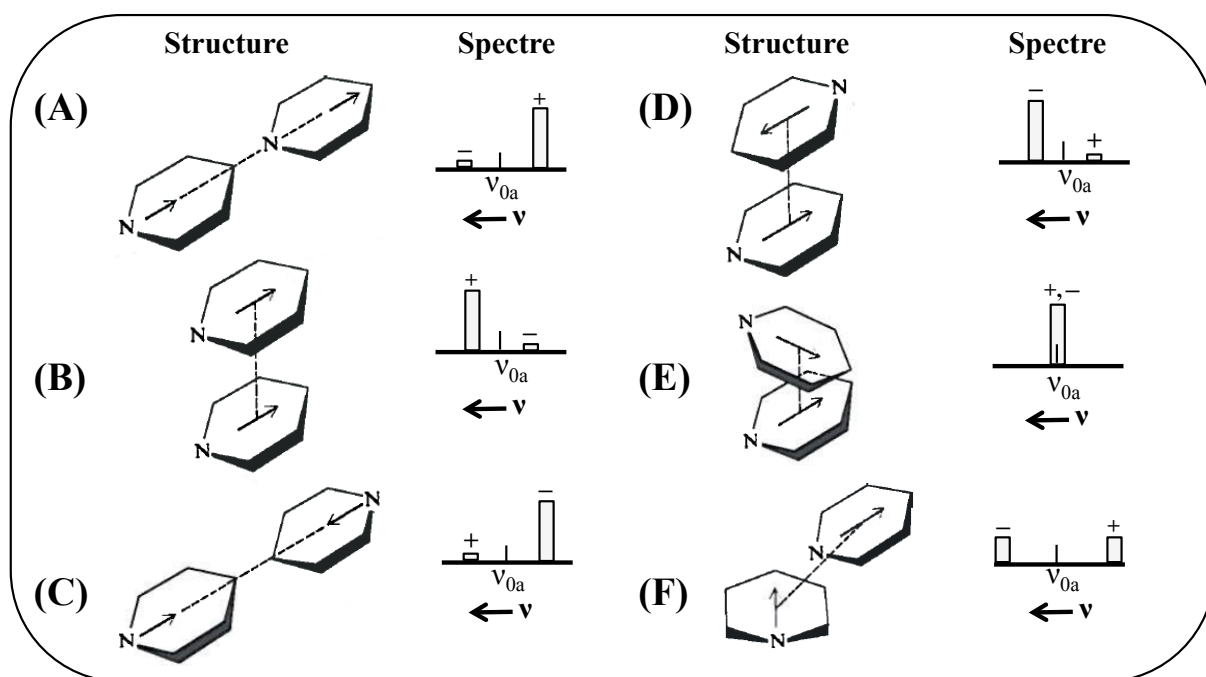


Figure 66: Schéma des six géométries possibles d'un dimère de pyridine, en fonction de la fréquence du dimère. (Adaptée²⁴¹)

Dans notre cas, l'analyse des spectres de dichroïsme circulaire nous a permis de confirmer la présence d'un mécanisme de type couplage excitonique, qui est dû à une interaction forte entre les deux moments dipolaires des deux entités fluorescentes. Nos résultats permettent d'associer à notre dimère une géométrie de type (B) et (D). Les deux sondes fluorescentes sont par conséquent orientées soit dans une conformation parallèle « modèle (B) », soit dans une conformation antiparallèle « modèle (D) » et ceci est vrai dans les deux milieux pris en

compte. Les fluorophores du type xanthène, comme la fluorescéine, la rhodamine ainsi que leurs dérivés, sont connus pour présenter à hautes concentrations, un dimère pour lequel les moments de dipôle de transition sont parallèles ou antiparallèles²⁴². Dans notre cas, puisque les deux sondes fluorescentes sont dans la même orientation, elles ne peuvent pas avoir que la conformation antiparallèle, donc nous pouvons exclure la conformation parallèle.

Les résultats de dichroïsme circulaire sont représentés dans la figure suivante (Fig.67).

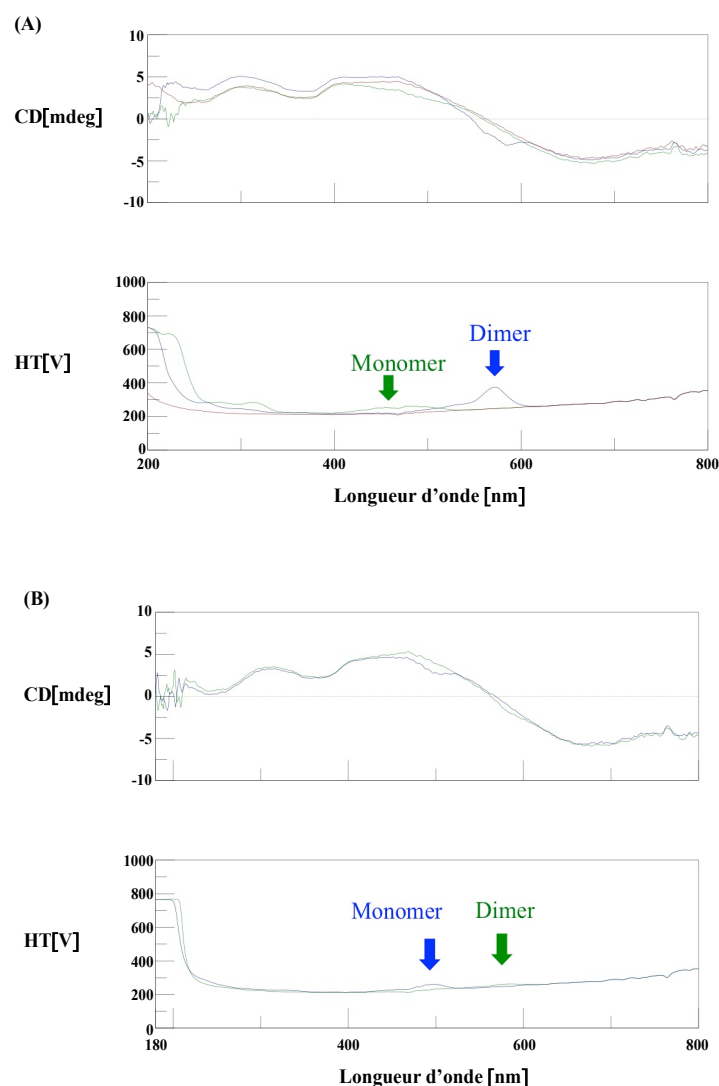


Figure 67: Représentation des spectres de dichroïsme circulaire superposés de la 6-FAM-alcyne (monomère) et du BigDye (dimère) dans le MeOH (A) et dans le TEAA (B).

Pour cette configuration, la formation d'un dimère de molécules où les moments de dipôles sont parallèles ou antiparallèles, au niveau de l'émission fluorescente on observe également une importante inhibition de l'émission de fluorescence. En effet, suite à l'excitation, un

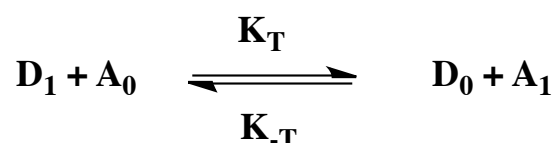
phénomène de conversion interne remplit le niveau dégénéré de plus faible énergie. La transition vers l'état fondamental étant interdite à partir de ce niveau, il en résulte un processus de conversion intersystème et retour à l'état fondamental sans émission de photon.

II.5.7 Caractérisation des interactions entre les sondes du BigDye: transfert d'énergie non radiatif de fluorescence (FRET)

Les fluorophores constituant le BigDye correspondent également à un couple donneur /accepteur du transfert d'énergie de type Förster (FRET, Fluorescence Resonance Energy Transfer). Il est alors possible que dans notre cas une partie des molécules en solution, celles caractérisées par une distance entre les sondes fluorescentes du BigDye supérieure à 10 Å, soient sujettes à un mécanisme d'interaction de type transfert d'énergie.

Dans un système donneur-accepteur, le FRET permet à une molécule fluorescente donnée (donneur) à l'état excité de céder son énergie à une autre molécule (accepteur) de manière non-radiative. Le FRET est un mécanisme qui implique un couplage entre les dipôles de transition du donneur et de l'accepteur de type faible et pour lequel aucune dégénérescence des niveaux énergétiques des deux sondes n'est observable. Les distances typiques du FRET sont généralement comprises entre 10 à 100 Å (Fig.68).

Pour que le FRET puisse avoir lieu, il est nécessaire que le spectre d'émission du donneur D et le spectre d'absorption de l'accepteur A se superposent. Si le donneur à l'état fondamental D_0 est excité à l'état excité D_1 , la réaction du transfert d'énergie peut se réaliser selon le schéma suivant :



La relaxation vibrationnelle de l'accepteur dans son état excité A_1 est assez rapide pour rendre improbable le processus représenté par K_{-T} .

L'équation de la vitesse du transfert d'énergie, K_T dépend de trois facteurs principales : (i) le recouvrement spectral entre l'émission du donneur et l'absorption de l'accepteur, (ii) l'orientation relative des moments de dipôle de transition, (iii) ainsi que la distance entre le donneur et l'accepteur²⁴³:

$$K_T = 8,71 \times 10^{23} K^2 \phi_D / R^6 \tau_D n^{-4} J_{AD} = 1 / \tau_D (R_0/R)^6.$$

ϕ_D et τ_D représentent respectivement le rendement quantique et la durée de vie du donneur, R représente la distance entre les deux fluorophores, et n représente l'indice de réfraction du milieu. K^2 représente le facteur géométrique qui caractérise l'orientation entre les moments de dipôle de transition des deux fluorophores, qui dans le cas d'une orientation aléatoire est égal à 2/3.

L'intégrale J_{AD} de recouvrement entre les spectres d'émission du donneur et l'absorption de l'accepteur, peut être calculée à partir de la formule suivante :

$$J_{AD} = \int \epsilon_A(\lambda) f_D(\lambda) \lambda^4 d\lambda$$

Où $\epsilon_A(\lambda)$ correspond à l'absorption molaire de l'accepteur et $f_D(\lambda)$ correspond à l'intensité de fluorescence normalisée du donneur.

R_0 est la distance de Förster qui peut être calculée à partir de la formule suivante :

$$R_0 = (8.79 \times 10^{23} n^{-4} \phi_D K^2 J_{AD})^{1/6}$$

Le choix du couple donneur-accepteur, ainsi que le milieu dans lequel les sondes se retrouvent, déterminent la valeur de R_0 . Cette valeur varie typiquement entre 20 et 60 Å.

L'efficacité de transfert et le rapport d'émission accepteur/donneur peuvent être influencés par des petits changements d'environnements de molécules en solution, du solvant, ainsi que de la distance entre les deux chromophores. De manière générale, les distances entre les deux entités fluorescentes communes au FRET, sont comprises entre 10 à 100 Å.

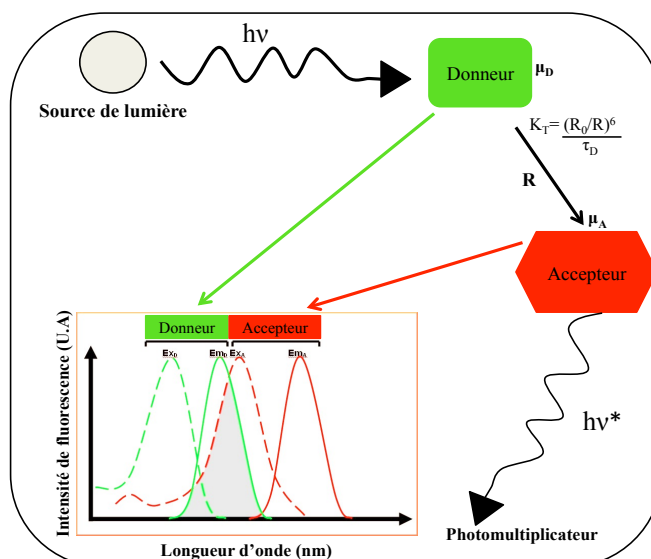


Figure 68: Modèle d'un mécanisme de type FRET. Le schéma de FRET proposé ne tient pas compte d'une émission éventuelle du donneur, ni d'une excitation directe possible de l'accepteur. μ_D et μ_A représentent les moments dipolaires de transition du donneur et de l'accepteur. K_T représente le taux de transfert d'énergie, avec R est la distance entre les deux fluorophores, R_0 est la distance de Förster, et τ_D est le temps de demi-vie du donneur en absence de transfert d'énergie. Un profil typique de recouvrement entre les spectres d'émission du donneur (en vert) et d'absorption de l'accepteur (en rouge) est également représenté par la région hachurée.

L'efficacité de transfert est définie comme étant la fraction de molécules de donneur dans l'état excité qui se désexcitent par transfert d'énergie non radiatif. Elle peut être calculée à partir de la formule suivante :

$$E = (R_0)^6 / (R_0 + R)^6 = 1 / 1 + (R/R_0)^6$$

Où R_0 représente le Rayon de « Förster » qui correspond à la distance entre le donneur et l'accepteur pour laquelle l'efficacité de transfert d'énergie est de 50%. Dans ce cas, si la distance entre le donneur et l'accepteur est égale au rayon de Förster R_0 , la possibilité que le donneur se désexcite en faveur de l'accepteur par transfert d'énergie non radiatif est de 50%. L'efficacité de transfert est une valeur qui varie entre 0 et 100%, alors qu'en pratique la gamme des distances mesurables par FRET est normalement comprise entre $R_0/2$ et $2R_0$ (Fig.57). Le couple de fluorophores choisi lors de la synthèse du BigDye, représente un couple donneur (6-FAM-alcyne) et l'accepteur (5-ROX-azide). Dans notre cas, la distance critique R_0 est d'environ 50 Å. Notre système permet par conséquent d'observer un changement de l'efficacité de transfert si les distances du couple donneur-accepteur varient entre 25 et 100 Å. Ceci n'est pas le cas du BigDye pour lequel la distance maximale calculée par modélisation de 13 Å est imposée par la longueur du bras espaceur (Fig.69).

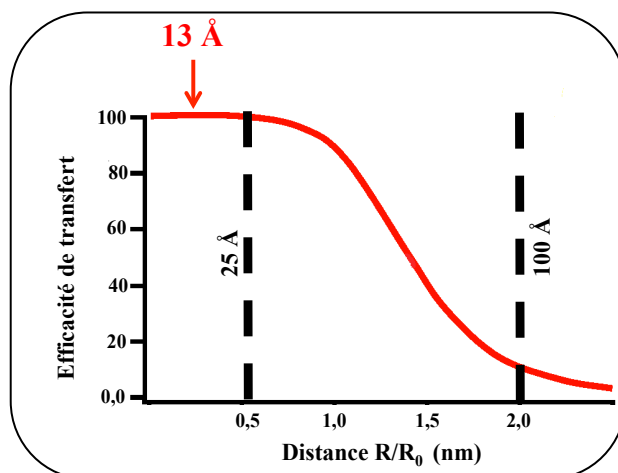


Figure 69: Représentation de la variation de l'efficacité de transfert E en fonction de la distance entre les deux sondes fluorescentes donneur et accepteur.

Expérimentalement, on peut calculer l'efficacité de transfert à partir des intensités de fluorescence du donneur (F_D) en absence ou en présence de l'accepteur (F_{AD}) selon la formule suivante :

$$E = 1 - F_{DA}/F_D$$

II.5.7.1 Efficacité de transfert

Dans notre système, l'efficacité de transfert entre les deux entités fluorescentes, dépend essentiellement de la nature du bras espaceur qui lie les deux sondes, et par conséquent impose la distance maximale entre le donneur et l'accepteur de fluorescence. Dans notre cas, cette distance maximale qui sépare le donneur de fluorescence de l'accepteur est de 13 Å, ce qui correspond à la forme étirée du BigDye (Fig.64). De ce fait, en tenant compte des valeurs de R et R_0 , nous avons pu calculer l'efficacité de transfert qui correspond à 99%.

De même, nous avons calculé l'efficacité de transfert de ce BigDye dans les deux solvants MeOH et tampon TEAA, à partir des spectres d'émission de fluorescence du donneur (6-FAM-alcyne) et du donneur en présence de l'accepteur (BigDye) (Fig.70). Dans le cas du tampon TEAA, l'efficacité de transfert calculée est de 90%, alors que dans le cas du MeOH, l'efficacité calculée est de 99%. Ce résultat est en bon accord avec l'efficacité de transfert calculée à partir du modèle, où les deux sondes se trouvent à la plus grande distance l'une de l'autre ($R=13\text{Å}$, et $E\sim 99\%$).

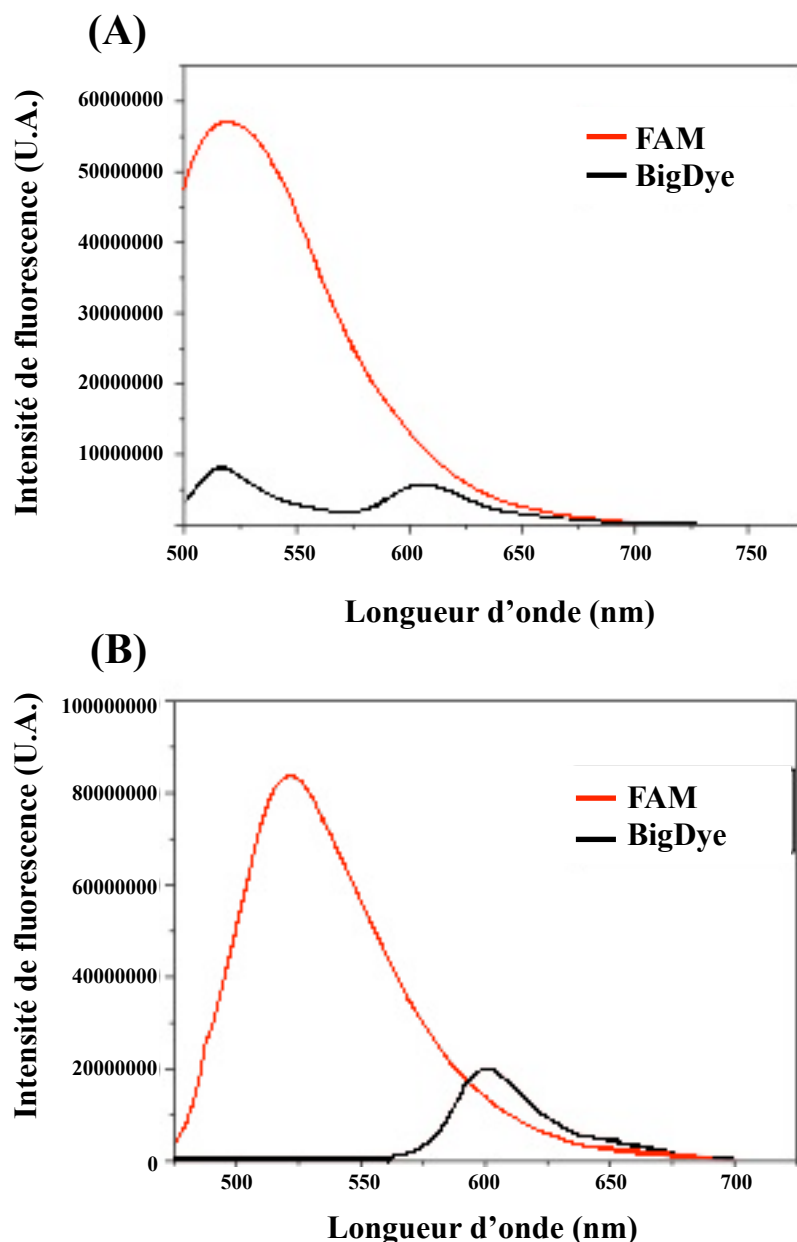


Figure 70: Spectres d'émission de fluorescence ($\lambda_{exc}=488\text{nm}$) du donneur de fluorescence 6-FAM-alcyne et du donneur en présence de l'accepteur (BigDye) à une concentration commune de 10^{-4}M , dans deux solvants TEAA (A) et MeOH (B).

II.5.7.2 Recouvrement spectral

La valeur du rayon de Förster R_0 typique de chaque couple donneur-accepteur dépend entre autre de l'intégral de recouvrement (Paragraphe II.5.7). Comme les propriétés spectrales du donneur et de l'accepteur pourraient différer selon le milieu et sa polarité, nous avons calculé

l'intégral de recouvrement entre le spectre d'émission du donneur et le spectre d'excitation de l'accepteur dans les deux solvants considérés, le MeOH et le tampon aqueux TEAA.

Notre analyse montre que le recouvrement spectral dans les deux solvants, et par conséquent la valeur de l'aire calculée sous la courbe spectrale définie par l'intersection des spectres d'émission de la FAM et d'excitation de la rhodamine est identique dans les deux cas (Fig.71)

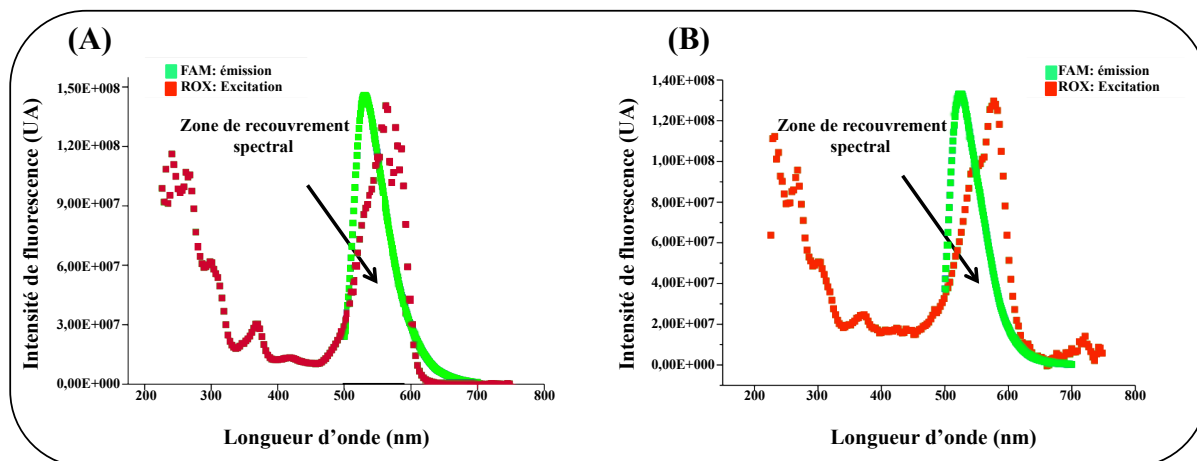


Figure 71: Schéma représentant le recouvrement spectral entre le spectre d'émission du donneur (6-FAM-alcyne) et le spectre d'excitation de l'accepteur (5-ROX-azide) : (A) dans le MeOH et (B) dans le TEAA.

II.5.8 Effets de la température vis-à-vis des interactions entre les sondes fluorescentes

Nous avons ensuite réalisé des expériences de fluorescence à température croissante afin d'évaluer l'impact de la température par rapport aux caractéristiques spectrales du BigDye dans le MeOH et dans le TEAA. La séparation des sondes par effet de la température pourrait provoquer l'éloignement des sondes et par conséquent la rupture du couplage excitonique, ce qui changerait les propriétés photophysiques du BigDye. Dans ce cas, les espèces en solution auraient tendance à adopter la configuration selon laquelle les sondes sont les plus éloignées possibles, avec une distance maximale de 13 Å selon notre modélisation (Fig.64). L'analyse des spectres obtenus (Fig.72) montre que l'augmentation de la température n'a pas d'effet sur les propriétés photophysiques du BigDye dans les deux solvants.

Ce résultat montre en conséquence, l'importance du solvant sur l'émission de fluorescence du BigDye. Dans les deux solvants, les caractéristiques spectrales restent conservées. Ces tests sont représentés dans la figure ci-dessous :

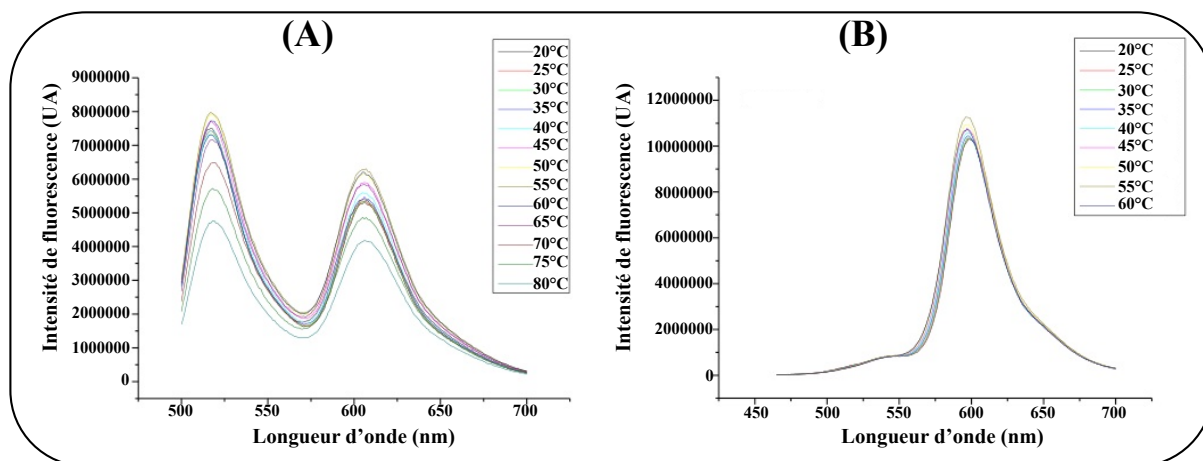


Figure 72: Effet de la température vis-à-vis des propriétés physicochimiques du BigDye dans le tampon TEAA (A), et dans le MeOH (B). La plage de températures choisie varie de 20°C à 80°C pour le BigDye dans le tampon TEAA en excitant à une longueur d'onde λ_{exc} = 488nm, et de 20°C à 60°C pour le BigDye dans le MeOH en excitant à une longueur d'onde λ_{exc} = 450nm.

II.5.9 Stabilité de l'émission fluorescente du BigDye

Nous avons ensuite analysé la stabilité de l'émission fluorescente de la molécule BigDye en observant son signal en fonction du temps. Le « Photobleaching » est le processus durant lequel une source de lumière intense envoyée vers une molécule fluorescente, peut provoquer des changements chimiques au niveau de cette sonde, de façon à inhiber les processus d'émission de fluorescence de cette sonde.

Nous avons étudié la stabilité de l'émission de fluorescence du BigDye en suivant sa fluorescence pour environ 2h (Fig.73).

Les résultats représentés dans la figure ci-dessous, montrent que la fluorescence du BigDye est stable vis-à-vis de la lumière pendant la période d'observation, puisque la variation du signal pendant le temps d'observation (environ 2h) est ~4% et par conséquent, non-significative.

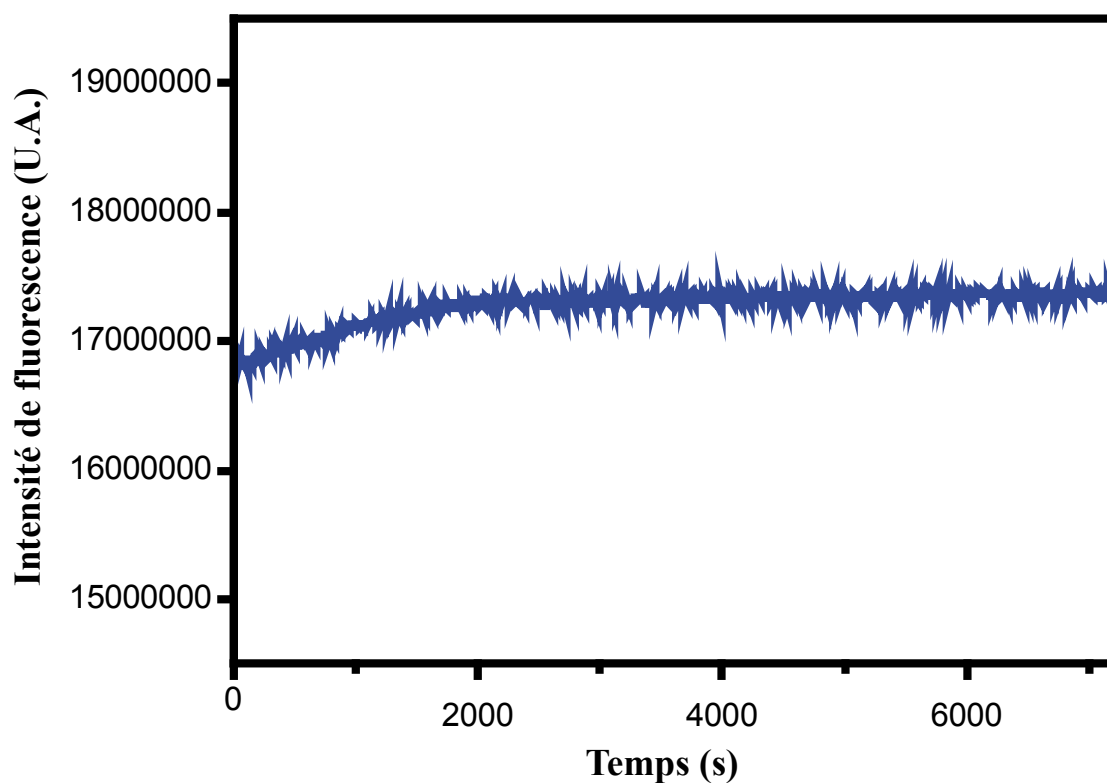


Figure 73: Cinétique du BigDye en fonction du temps ($\lambda_{\text{exc}} = 450 \text{ nm}$ et $\lambda_{\text{emi}} = 600 \text{ nm}$) dans du MeOH. On observe que l'émission de fluorescence de la molécule est stable vis-à-vis de la lumière incidente et que le photobleaching de la molécule est d'environ 4% et par conséquent non-significatif.

II.6 Discussion

II.6.1 Choix et synthèse de la carboxy-fluorescéine

Le choix de la fluorescéine en tant que molécule donneur de fluorescence est basé sur le fait que cette molécule possède une structure tricyclique rigide, une grande absorptivité molaire ($88,000\text{ cm}^{-1}$ à $\text{pH}=9$) en excitant avec un ion laser à une longueur d'onde de 488 nm, un rendement quantique élevé (0.92 à $\text{pH}>8$) et une grande photostabilité^{244,245}. Ces caractéristiques ont d'ailleurs conduit de nombreuses équipes de recherche à modifier la fluorescéine avec diverses fonctions réactives afin de l'attacher à des macromolécules et acides aminés²⁴⁶. En plus, le spectre d'émission de cette molécule recouvre partiellement ou complètement les spectres d'absorption des différents dérivés de la rhodamine, ce qui rend les paires fluorescéine-rhodamine compatibles avec les expériences de transfert d'énergie^{244,245,247}. Par contre, la nature et le pH du solvant utilisé sont très importants pour les expériences réalisées avec la fluorescéine. Dans les solvants aqueux, la fluorescéine se trouve sous différentes formes, cationique, neutre, anionique et dianionique, qui rendent ces propriétés physico-chimiques fortement dépendantes du milieu. Différentes études ont porté sur de nouvelles méthodes pour synthétiser des sondes fluorescentes à partir des fluorescéines modifiées^{223,224,235,248}. La molécule de fluorescéine ne possède pas de groupes fonctionnels pour coupler des macromolécules, il a donc fallu modifier cette molécule, et la carboxy-fluorescéine a été synthétisée, sous la forme de ces deux isomères 5- et 6- carboxy-fluorescéine, pour répondre à ce problème^{249,250}. La synthèse de la carboxy-fluorescéine à partir de l'anhydride et du résorcinol est une réaction d'acylation/ cyclodéshydratation de type Friedel et Crafts. Par la suite, des études ont été réalisées pour simplifier les méthodes de synthèse, pour trouver de nouvelles techniques de purification de chaque isomère seul et introduire de nouveaux groupes fonctionnels dans le but de coupler des biomolécules. Parmi les méthodes de purification, nous pouvons citer celle qui consiste à synthétiser les sels de diesters par réaction avec l'anhydride pivaloïque (Pv_2O) et de la diisopropylamine, puis à précipiter l'isomère 6-carboxyfluorescéine en utilisant de l'éthanol, tandis que l'isomère 5-carboxyfluorescéine est isolé par la suite par précipitation avec du nitrométhane²⁴⁹. Pour mon travail, j'ai préféré la méthode rapportée par *Ueno et al.* qui consiste à utiliser lors de la synthèse, l'acide méthanesulfonique comme agent de déshydratation et de catalyse²⁵⁰. Les produits obtenus constituent donc un mélange des deux isomères 5- et 6- carboxy-fluorescéine méthanesulfonate, plus facile à séparer que les isomères de la carboxy-fluorescéine.

L'isomère 5-carboxy-fluorescéine méthanesulfonate est précipité à partir d'un mélange éthanol/hexane, tandis que l'isomère 6-carboxyfluorescéine est précipité sous forme de sel de dipivalate diisopropylamine²⁴⁹. Ensuite l'isomère 5-carboxy-fluorescéine est obtenu par traitement avec NaOH et précipitation par HCl.

II.6.1.1 Dérivé de la fluorescéine : Substitution de Mannich

A priori, il est possible de modifier le squelette xanthène de la carboxy-fluorescéine à différentes positions, les positions 4' et 5' et les positions 3' et 6'.

En plus des positions 5- et 6- de la carboxy-fluorescéine qui peuvent être toutefois activées avec du dicyclohexylcarbodiimide (DCC) et du N-hydroxysuccinimide (NHS) pour être ensuite couplées à des molécules d'intérêt, les positions 3' et 6' sont aussi couramment modifiées, pour donner les dérivés O-alkylés. On peut citer par exemple les travaux de l'équipe de « *Zhang et Gao* » qui ont employés cette stratégie pour développer un protocole économique, qui permet de lier d'une manière covalente, le composé O-propargylfluorescéine avec des molécules bioactives possédant un groupement azoture par la cycloaddition de Huisgen²⁵¹. L'équipe de « *Wang et al.* » ont synthétisé du N-hydroxy-succinimidyl-fluorescéine-O-acétate pouvant réagir spécifiquement avec les amines primaires. Cette sonde peut être utilisée pour l'analyse des amines aliphatiques dans l'environnement, qui sont difficilement détectées par les méthodes classiques²⁵². Le groupe de « *Li et al.* » ont préparé le composé O-glycidylfluorescéine qui présente une longueur d'onde d'émission plus large et un rendement quantique plus élevé que la fluorescéine lorsqu'il est dissous dans l'éthanol ou dans un solvant aqueux. Son groupement époxy réagit plus spécifiquement avec un cycle imidazole, donc il a été développé avec succès pour détecter l'histidine.

Une nouvelle stratégie pour introduire une substitution sur la fluorescéine est d'utiliser les protons réactifs en positions 4' et 5' sur le squelette du xanthène, pour former les dérivés alkyles de la fluorescéine. L'équipe de « *Swamy et al.* » a développé le composé 2',7'-dichloro-4',5'-bis-([2-(bis-pyridine-2-ylmethylamino)-éthyle]-methylamino-methyl) fluorescéine par la réaction de « Mannich » entre le 2',7'-dichlorofluorescéine et le N-(2-methylaminoethyl)-bis(2-pyridylmethyl)amine en présence du formaldéhyde. Ce composé était développé pour contrôler la présence de pyrophosphate (PPi) et de l'ATP, dans les solvants aqueux en présence d'ions Mn²⁺²⁵³. L'équipe de « *Lippard et Lim* » a développé une

nouvelle classe de fluorescéine modifiée en position 5' par réaction entre le 2'-chloro-5'-formylfluorescéine et la 8-aminoquinoline substituée en position 2. Ce composé était développé pour localiser le monoxyde d'azote dans les cellules. Il a été aussi utilisé pour analyser les métaux toxiques comme le Cu(II)²⁵⁴. L'équipe de « *Ansyln et Wallace* » a développé la fluorescéine-4',5'-dialdoxime, capable de détecter les agents chimiques contenant des groupements oxyfluorure de phosphore²⁵⁵.

Il est intéressant de noter que le cycle phényle a été modifié avec des halogénures ou autres substituants afin d'améliorer les propriétés chimiques de la fluorescéine et de la rhodamine. « *Sun et al.* » ont développé les composés fluoro 5(6)-carboxy-fluorescéine qui présentent une photostabilité plus élevée et un pKa plus faible que la carboxy-fluorescéine²⁵⁶.

Durant ces trois années de thèse, j'ai synthétisé et purifié les dérivés de la fluorescéine puis je les ai modifiés au niveau de la position 4' avec une fonction alcyne en utilisant la substitution de Mannich. Cette méthode qui consiste à former l'aldéhyde en présence du formaldéhyde dans l'H₂O, puis de le substituer avec l'amine qui contient l'alcyne, la propargylamine dans notre cas^{257,258}.

En modifiant le dérivé de la fluorescéine avec la fonction prop-2-yn-1-ylamino-méthyle, nous avons observé la formation de produits secondaires et plus particulièrement du composé disubstitué en position 4' et 5' avec la fonction alcyne. Une étape de purification a été nécessaire afin d'isoler le composé monosubstitué. La carboxy-fluorescéine est une molécule très polaire et la modification avec la fonction prop-2-yn-1-ylamino-méthyle la rendait encore plus polaire et donc plus difficile à être purifiée sur gel de silice. Alors pour pouvoir réaliser la purification sur une colonne de gel de silice, nous avons décidé de neutraliser les fonctions acides carboxyliques portées par la molécule. Pour cela, nous avons réalisé la purification en présence d'acide acétique. Avec ce type d'éluant, nous avons réussi à séparer les dérivés mono- et di-substitués. L'élimination de l'acide acétique en excès a été réalisée par co-évaporations avec du toluène (x3), du MeOH (x3) puis du dichlorométhane (x3).

II.6.2 Choix et synthèse de la carboxy-rhodamine et ses dérivés

Pour la molécule accepteur de fluorescence, notre choix s'est porté sur la rhodamine et ses dérivés. En effet, les différents substituants au niveau de l'amine provoquent des délocalisations d'électrons spécifiques sur le squelette de la rhodamine, ce qui induit les différentes longueurs d'onde d'émission de fluorescence et les couleurs spécifiques des

dérivés de la rhodamine. La rhodamine et ses dérivés sont utilisés dans différents domaines, suite à leur photostabilité élevée et leurs propriétés photo-physiques excellentes. Elles sont appliquées dans la caractérisation des surfaces de nanoparticules de polymères, la fluidité des membranes lipidiques, dans l'étude d'adsorption des oligonucléotides sur latex, dans l'étude de la structure et du dynamique des micelles, dans l'imagerie des cellules vivantes, pour détecter des modifications à la surface des virus et plus particulièrement comme des capteurs chimiques utilisés soit *in vitro* ou *in vivo* pour la détection des ions Hg(II), Cu(II), Fe(III) et Cr(III)²⁵⁹. Comme la fluorescéine, la rhodamine possède plusieurs sites de modifications possibles. On peut citer les amines du groupement xanthène, en positions 3' et 6'. À noter que la modification de ces deux fonctions amines provoque une augmentation de la nucléophilicité de l'oxygène au niveau du benzène et mène à la formation de la lactone. Ces dérivés sont non-fluorescents. Généralement les fonctions amines de la rhodamine sont modifiées par réaction avec du chloroformiate ou avec une fonction acide carboxylique en utilisant un carbodiimide comme agent de couplage. Selon le domaine d'application, différents types de substituants peuvent être introduits. On peut citer les travaux de l'équipe de « *Kim et al.* » qui ont synthétisé des dérivés de rhodamine bis-boronés pour la détection des mono et oligosaccharides²⁶⁰, qui peuvent être utilisés aussi pour détecter des motifs de tétra-sérine dans les protéines²⁶¹. Une autre approche proposée par « *Tang et al.* » est de synthétiser des dérivés de la R6G contenant une liaison Se-N pour détecter les thiols²⁶². Aussi pour la détection des thiols, des dérivés de la rhodamine modifiés avec une fonction sulfonamide ont été préparés par l'équipe de « *Shibata et al.* »²⁶³. D'autres positions modifiées de la rhodamine sont les positions 4 et 5, qui sont extrêmement chers en commerce et souvent vendus comme un mélange d'isomères. Par exemple, des dérivés succinimidyl esters, isothiocyanate et maléimide de la rhodamine (Rho 110), sont vendus chez Sigma-Aldrich ou molecular probes, et sont utilisés pour les expériences de marquage de biomolécules. Ces dérivés modifiés au niveau de leur cycle carboxyphényle sont généralement préparés par condensation entre l'anhydride phtalique fonctionnalisé à l'avance avec des fonctions N-alkyls, ou m-aminophénols en présence d'acide sulfurique concentré. Dans le cas particulier de la préparation des dérivés succinimidyl esters, la méthode la plus utilisée est la préparation du mélange d'isomères (5-) et (6-) carboxy-rhodamine à partir de l'anhydride mellitique, suivie par estérification en utilisant du N-hydroxysuccinimide. L'équipe de « *Menchen et Fung* » a été la première à développer la synthèse des dérivés succinimidyl de la tétraméthylrhodamine et de la rhodamine (101), en utilisant dans la deuxième étape du di-N-succinimidylcarbonate (DSC) et du 4-diméthylaminopyridine

(DMAP) dans du DMF²⁶⁴. Une autre méthode de synthèse de ces dérivés a été élaborée par l'équipe de « *Cruickshank et Bittner* » qui ont remplacé le NHS par du N,N'-diisopropylcarbodiimide (DIPC) comme agent de couplage, lors de la synthèse des dérivés de la rhodamine Rho (110) pour marquer des nucléotides²⁵⁹. La fonction acide carboxylique en position 2 était une cible de modification par des travaux de plusieurs équipes. Citons par exemple la modification de la rhodamine Rho B avec une fonction hydrazide par l'équipe de « *Czarnick et al.* »²⁶⁵, qui a été utilisée comme dosimètre du Cu(II). Ces dérivés de rhodamine-hydrazide sont ensuite modifiés avec des aldéhydes^{266,267}, des cétones²⁶⁸, des isothiocyanates²⁶⁹ pour préparer d'autres capteurs chimiques.

Pour mon projet de recherche, nous avons décidé de synthétiser la carboxy-rhodamine et trois de ces dérivés : la carboxy-TAMRA, la carboxy-R6G et la carboxy-X-rhodamine (ROX). La formation du squelette de la rhodamine peut être effectuée par catalyse acide à haute température. En suivant le protocole décrit dans la littérature, qui utilise l'acide sulfurique concentré à 180°C, les produits désirés n'ont pas pu être obtenus suite à une dégradation des produits de départ par l'acide sulfurique concentré²⁷⁰. De même lorsque ces réactions ont été réalisées dans les mêmes conditions que pour les dérivés de la fluorescéine en utilisant l'acide méthanesulfonique, les dérivés de la rhodamine n'ont pas pu être obtenus. Dans la littérature, plusieurs équipes ont rapporté des conditions de synthèse utilisant la chimie des micro-ondes^{271,272}. Nous avons donc réalisé plusieurs tentatives de synthèse en utilisant la chimie des micro-ondes avec de l'acide propionique et quelques gouttes d'acide sulfurique concentré, en changeant la température et le temps de la réaction, le résultat était toujours négatif, et on n'arrivait pas à obtenir la carboxy-rhodamine elle même.

Pour les dérivés de la rhodamine, un protocole de synthèse classique utilisant l'acide propionique avec des traces d'acide sulfurique a été décrit par l'équipe « *Uddin et al.* » en 2008²²¹. En essayant ce dernier, j'ai réussi à obtenir les dérivés de la carboxy-rhodamine, sans toujours avoir la carboxy-rhodamine elle même. Les rendements de cette réaction varient en fonction du dérivé de la rhodamine mis en réaction. La synthèse conduit à un mélange des deux isomères 5- et 6- carboxy-rhodamine. La séparation de ces deux isomères est rendue difficile par la faible solubilité et une polarité élevée de ces dérivés. Différents types de colonnes chromatographiques ont été utilisés. Le meilleur résultat est obtenu en utilisant le gel de silice et un gradient de MeOH dans CH₂Cl₂ avec un débit rapide. À la fin de la purification, la silice éluée est éliminée par dissolution du composé dans NaOH et précipitation au HCl.

Afin de faciliter la séparation des deux isomères de position, nous avons décidé d'introduire la fonction azoture sur l'acide carboxylique et d'effectuer la purification au niveau des molécules dérivatisées et moins polaires. Dans ce cas, des rendements plus élevés ont été obtenus.

La synthèse des dérivés de la rhodamine comportant la fonction azoture au niveau de l'acide carboxylique est réalisée par formation d'une liaison amide. Dans la littérature, il existe plusieurs méthodes permettant de réaliser ce couplage. La méthode utilisant le BOP-Cl consiste à activer l'acide sous forme d'un anhydride phosphinique et à introduire l'amine en une seule étape avec une catalyse basique²⁴⁹. Lorsque nous avons appliqué cette stratégie à nos molécules, le rendement de cette réaction était très faible. Nous avons donc envisagé une autre méthode. Notre choix s'est porté sur l'activation de l'acide par le groupement N,N'-disuccinimidyl en présence de 4-diméthylaminopyridine suivi d'une étape de purification et finalement substitution avec l'amine^{233,273,270}. Au cours de notre synthèse, nous avons remarqué qu'il y a méthanolyse de l'ester activé, qui remplace l'ester activé avec la fonction succinimidyl par la fonction méthyl et formation de l'ester méthylique. Afin d'éviter cette méthanolyse, nous avons décidé de faire les deux étapes successivement sans réaliser la purification intermédiaire. Ces nouvelles conditions nous ont permis d'obtenir les produits modifiés avec la fonction azoture avec de meilleurs rendements par rapport à la méthode de couplage avec le BOP-Cl. De plus, comme nous l'avons envisagé, la séparation des isomères de position substitués sur la fonction acide carboxylique est beaucoup plus facile que celle des composés de départ.

II.6.3 Choix et synthèse du bras espaceur

La molécule cible que nous souhaitons préparer est constituée de deux molécules fluorescentes. Nous avons modifié l'une de ces deux molécules par une fonction alcyne et l'autre par une fonction azoture afin de pouvoir les lier par une cycloaddition 1,3-dipolaire d'Huisgen, aussi appelée réaction de « click chemistry ».

En effet cette stratégie semble être idéale pour notre projet du fait de sa régiosélectivité, de sa tolérance vis-à-vis de nombreuses fonctions organiques, du large éventail de conditions réactionnelles possibles et le plus important de la formation d'un cycle rigide très important pour réduire le repliement de notre molécule.

Le concept de « click chemistry » a été introduit en 2001 par *Sharpless et al.* afin de développer une large gamme de blocs sélectifs, modulables et facilement assemblables pour un large champ d'applications²⁷⁴. Cette réaction de cycloaddition alcyne/azoture est catalysée par du cuivre (I) et permet la synthèse de composés 1,2,3-triazole substitués en positions 1,4.

Le premier mécanisme de catalyse pour cette CuAAC « CuI-catalyzed Azide-Alkyne 1,3-dipolar Cycloaddition » a été proposé par Sharpless²⁷⁵. Très brièvement, le cycle catalytique débute par la formation d'un complexe Cu-acétylide, suivi d'un complexe Cu-acétylide-azoture qui se cyclise en formant un métal cycle intermédiaire. Une fois le triazole formé, le catalyseur est régénéré par scission de la liaison Cu-C (Fig.74).

Il existe deux méthodes permettant de libérer les ions Cu(I) dans le mélange réactionnel. La plus directe consiste à additionner des sels de Cu(I) au milieu à partir de CuI, CuBr ou Cu(PPh₃)₃Br. Dans ces conditions, l'utilisation d'une base est nécessaire afin de limiter la formation de sous-produits, et d'améliorer le rendement de la réaction en limitant ainsi la dégradation des ions Cu(I) par oxydation. Les bases les plus utilisées dans le cas de la cycloaddition sont la triéthylamine, la 2,6 lutidine, la DIPEA et la PMDETA et les solvants peuvent être aussi divers que le THF, le DMF, le toluène ou le MeCN²⁷⁶. La deuxième voie permettant de libérer le Cu(I), et qui consiste à générer ce dernier *in situ* par réduction du Cu(II). Le Cu(II) est apporté sous forme de CuSO₄ et il est réduit par l'ascorbate de sodium²⁷⁵. Dans ce cas, les solvants utilisés sont constitués d'un mélange d'eau-alcool, le plus souvent eau-tBuOH ou eau-EtOH, mais il est également possible de faire la réaction en présence d'un solvant organique comme le DMSO afin de faciliter la dissolution des blocs à coupler^{277,278}. De plus, pour des réactions avec des molécules sensibles au cuivre, il est nécessaire d'utiliser des ligands du Cu(I) tels que le TBTA qui permettent de le stabiliser et d'utiliser des quantités plus faibles²⁷⁹.

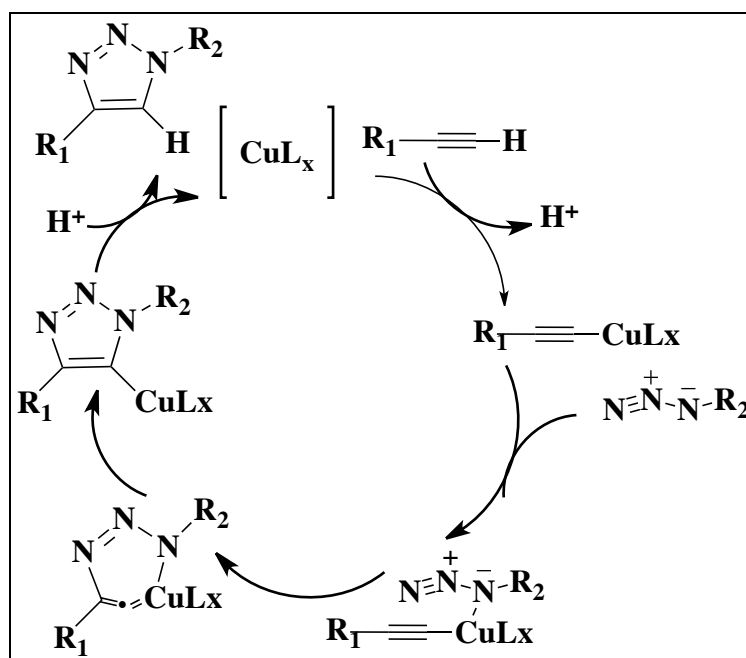


Figure 74: Cycle catalytique du couplage azoture-alcyne catalysé par le Cu(I).
(Adaptée²⁸⁰)

Pour notre synthèse nous avons choisi la deuxième stratégie^{274,281,282,283,284,285}. Celle-ci ne nécessite pas l'utilisation d'une base pour faire la réaction, donc il y a pas besoin de suivre le pH de la solution tout au long de la réaction. En utilisant ces conditions j'ai obtenu le premier BigDye formé par le couplage entre la 4'-[(prop-2-yn-1-ylamino)méthyl]6-carboxy-fluorescéine et la 5-(N-(3-azido-propyl))carboxamide-ROX. Par contre elles ne m'ont pas permis de réaliser le couplage avec les autres dérivés de la rhodamine. J'ai donc modifié les conditions de la réaction. Pour cela, j'ai remplacé le tBuOH par l'éthanol et j'ai augmenté le nombre d'équivalents du Cu(II) et d'ascorbate mais sans résultat positif. Face à cet échec, j'ai choisi d'ajouter le Cu(I) directement dans le milieu réactionnel. L'avantage qu'il présente par rapport à la première méthode est que le cuivre reste en suspension, donc le choix du solvant est moins important dans la réaction et on peut utiliser des solvants organiques. En plus, la présence du cuivre sous forme solide, facilite la purification du produit final. Le CuBr a été utilisé comme source de cuivre et la 2,6 lutidine a permis de rendre le milieu réactionnel plus basique. Malgré ces changements, il ne m'a pas été possible d'obtenir les autres molécules cibles. Cet échec peut être expliqué par le fait que la présence de groupements alcynes à proximité de l'atome de Cu(I), puissent saturer cet atome par action chélatante. Cette complexation du cuivre avec les différentes fonctions alcynes, empêche ensuite la réaction avec l'azoture. Une autre explication de cet échec pourrait être attribué à la formation d'un homo-couplage entre deux fonctions alcynes terminales, qui est aussi catalysé par le cuivre et qui peut réduire la réaction avec l'azoture. Dans les deux possibilités, le cycle catalytique est

perturbé au début, lors de la première étape de formation de Cu-acétylide, puisque c'est le groupement alcyne qui subit les transformations tandis que le groupement azoture reste intact. Ces mécanismes apparaissent clairement dans les spectrométries de masse réalisées en fin de réaction, puisqu'on retrouve à chaque fois la rhodamine porteuse du groupement azoture, ce qui montre qu'elle n'a pas participé à la réaction. D'autre part on ne retrouve pas la carboxy-fluorescéine porteuse du groupement alcyne, ce qui s'explique par le fait qu'elle a réagi pour donner des produits secondaires non désirés.

Nous pouvons aussi envisager que des traces d'acide, provenant de l'acide acétique lors de la purification des dérivés de la carboxy-fluorescéine, soient à l'origine de cet échec de couplage des deux sondes fluorescentes, puisque la réaction de click est réalisée en milieu basique.

II.6.4 Couplage excitonique intramoléculaire: système fluorescéine-rhodamine

Le BigDye obtenu est formé par les deux sondes fluorescentes la 4'-N-méthyl-propargylamine-5-carboxyfluorescéine et la 5-(N-(3-azido-propyl))carboxamide-ROX, séparées par un bras espaceur formé par le cycle de triazole.

Étant donné que les propriétés spectrales d'une sonde fluorescente varient en fonction du solvant utilisé, nous avons testé ce BigDye dans deux solvants différents : tampon TEAA et MeOH. Nous avons par nos expériences de fluorescence, mis en évidence le fait que le solvant détermine de manière importante, les propriétés physico-chimiques du BigDye suite à une différence spectrale dans les deux solvants. L'utilisation de ces BigDyes dans des expériences de séquençage d'ADN, nécessite qu'ils soient fonctionnels dans des solvants aqueux et plus particulièrement dans des tampons.

Nous avons par la suite modélisé le BigDye. Nous avons ainsi pu observer deux configurations extrêmes pour notre système, qui sont essentiellement due à la longueur et la flexibilité du bras espaceur. Une configuration pour laquelle les deux sondes fluorescentes constituant notre BigDye, se superposent avec une distance de 6,1 Å, et une autre correspondante à la forme étirée du BigDye, dans laquelle les deux sondes sont éloignées avec une distance de 13 Å.

L'une des difficultés le plus fréquemment rencontrées dans les études de FRET consiste à éviter l'excitation directe de l'accepteur de fluorescence, dans notre cas la Rhodamine. La longueur d'onde d'excitation (488nm) a été spécifiquement choisie pour éviter l'excitation

directe de la rhodamine dans le BigDye. Nous avons caractérisé les propriétés physico-chimiques des BigDyes. De ce fait nous avons observé que les interactions entre les sondes fluorescentes sont majoritairement dues à un couplage excitonique, à cause de la flexibilité ainsi que des distances imposées par le choix du bras espaceur.

Une étude de dichroïsme nous a permis aussi d'identifier la géométrie possible pour ce BigDye dans les solvants aqueux. Cette analyse nous a permis de valider la présence du couplage excitonique.

Afin de vérifier si notre BigDye est compatible avec les expériences de séquençage, une étude de la stabilité de ce BigDye vis-à-vis de la lumière incidente a été réalisée pendant un temps d'observation de deux heures. Le spectre obtenu a montré une variation non significative du signal et que notre BigDye est stable, donc il n'existe pas de photobleaching.

CHAPITRE III: SYNTHÈSE DE NOUVEAUX RÉACTIFS CHIMIQUES POUR L'ÉTUDE STRUCTURALE D'UN GÉNOME ENTIER *IN VIRIO*

III.1 Objectif

Le Virus de l'Immunodéficience Humaine de type 1 (VIH-1) est un virus qui présente un taux de mutation et de recombinaison élevés, ce qui rend sa prolifération difficile à contrôler. A l'heure actuelle, il existe des résistances à tous les traitements antirétroviraux commercialisés, il est donc important de trouver de nouveaux inhibiteurs. Dans ce contexte, une meilleure connaissance de l'ensemble des processus biologiques qui régissent le cycle réplcatif du VIH-1, pourrait conduire au développement de nouvelles stratégies antivirales et à plus long terme à des inhibiteurs.

La maturation des particules virales est un processus critique du cycle viral nécessaire à l'acquisition de l'infectivité, qui associe la maturation des précurseurs protéiques Gag et Gag-Pol et celle du dimère d'ARN génomique (détails dans le chapitre I). Alors que le mécanisme de formation des protéines de matrice, de capsid et de nucléocapsid par action de la protéase virale sur les précurseurs Gag et Gag-Pol est bien compris, il n'en est pas de même pour la maturation nucléique. Il est connu dans ce cas que la molécule d'ARN n'est pas clivée, et que la maturation est purement un changement structural. Des études⁷⁴ ont montré que les dimères des virus matures sont plus compacts et plus stables que ceux des virus immatures, ce qui laisse penser qu'un réarrangement se produit. Mais à l'heure actuelle, il n'existe pas d'informations sur la nature, sur le lieu de ces changements conformationnels ni sur leur lien éventuel avec le processus de maturation protéique. L'équipe pense que les changements de structure ont lieu tout au long de la molécule d'ARN génomique. Afin de répondre à cette question et comprendre l'ensemble de ces changements conformationnels, il serait idéal de réaliser la cartographie *in virio* de l'ensemble du génome viral à différents stades de la maturation protéique. Etant donné la faible quantité d'ARN présent dans les virions et la taille du génome du VIH-1, cette étude ne peut pas être réalisée par les techniques classiques de cartographie présentées dans le premier chapitre. Il est nécessaire de coupler la cartographie à une technique de séquençage à haut débit, afin d'obtenir des millions de séquences lues en une seule expérience.

III.2 État des lieux

Récemment, la cartographie enzymatique a été couplée au séquençage à haut débit et cela a conduit au développement de deux méthodes, « Parallel Analysis of RNA Structure » (PARS) et « Fragmentation Sequencing » (Frag-Seq)²⁸⁶ (Détails dans le Chapitre I). Il est intéressant de noter que les deux approches fournissent des informations complémentaires, la combinaison des deux peut donc fournir des structures secondaires plus précises.

Comme nous l'avons vu dans le premier chapitre, l'utilisation des sondes enzymatiques telles que S1, V1 ou P1 permet l'étude d'un génome entier grâce à la discrimination entre les coupures endogènes dans l'ARN et les coupures par l'enzyme, malheureusement ces sondes présentent également des inconvénients :

- Toutes les positions de l'ARN ne sont pas interrogées de la même façon, alors la structure secondaire est obtenue avec une faible résolution.
- Les expériences sont réalisées *in vitro*, les ARN doivent donc être synthétisés ou extraits des cellules avant de pouvoir être étudiés, la structure obtenue peut être différente de la conformation *in vivo*.

Afin de surmonter la première limitation, une nouvelle approche qui couple la technique de SHAPE au séquençage à haut débit (SHAPE-seq) a été développée²⁰³ (Détails dans le chapitre I). Cette technique ne peut pas être utilisée pour étudier un génome entier du fait qu'elle utilise une amorce spécifique pour une région donnée sur l'ARN, ainsi pour étudier un génome entier comme celui du VIH-1 qui contient des milliers de séquences, il est nécessaire d'utiliser des milliers d'amorces spécifiques.

Les trois techniques que nous venons de voir ne peuvent pas être appliquées à l'étude d'un génome entier *in virio*. Il faudrait développer de nouveaux réactifs chimiques qui peuvent interroger la structure des ARN dans des cellules vivantes et qui, comme les sondes enzymatiques permettraient de différencier les fragments d'ARN modifiés et ceux non-modifiés, afin de pouvoir adapter cette cartographie au séquençage à haut débit. Dans cet objectif, nous avons donc décidé de développer de nouveaux dérivés d'anhydride isotoïques, puisque ces molécules sont déjà utilisées dans la technique de SHAPE.

III.3 Les anhydrides isatoïques

Les anhydrides isatoïques (Fig.75) sont généralement utilisés comme agents intermédiaires dans la synthèse organique, à l'origine de la formation d'une large variété de composés, possédant des rôles divers dans le domaine de la chimie (chimie des polymères, synthèse hétérocyclique et synthèse de sondes fluorescentes) et dans le domaine médicale et pharmaceutique (marquage d'ARNm et d'ARNt)²⁸⁷.

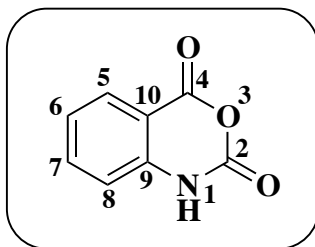


Figure 75: L'anhydride isatoïque avec la numérotation des atomes.

III.3.1 Méthodes de synthèse

Il existe trois voies de synthèse de l'anhydride isatoïque (2H-3,1-benzoxazine-2,4(1H)-dione) couramment utilisées²⁸⁷ (Fig.76) :

- L'oxydation de l'isatine (1H-indole-2,3-dione) qui conduit à l'introduction d'un atome d'oxygène entre les deux groupements carbonyles adjacents, sans décomposition de l'hétérocycle (voie (1), fig.76). Cette synthèse a été réalisée avec plusieurs agents oxydants comme le trioxyde de chrome (CrO_3), l'acide peroxy-carboxylique, le peroxyde d'hydrogène dans l'acide acétique ou acide formique, en présence de quantités catalytiques d'acide sulfurique ou encore le peroxyde d'hydrogène/urée dans un mélange d'acide acétique/anhydride acétique ou acide formique et acide sulfurique²⁸⁸.
- Le réarrangement de dérivés de l'acide phtalique. L'acide phtalamique, le phtalimide, le N-chlorophtalimide ou l'acide phtalique-azide lorsqu'ils sont traités par du NaOCl subissent un réarrangement de type Curtius, qui conduit à la formation de l'anhydride isatoïque (voie (2), fig.76). Ces dérivés de l'acide phtalique peuvent être obtenus par diverses méthodes à partir d'acide ou d'anhydride phtalique²⁷⁹.

- La réaction de cyclisation entre l'acide anthranilique (l'acide 2-aminobenzoïque) et le chloroformiate d'éthyle ou le phosgène est également une méthode intéressante pour la synthèse de l'anhydride isatoïque²⁸⁹ (voie (3), fig. 76).

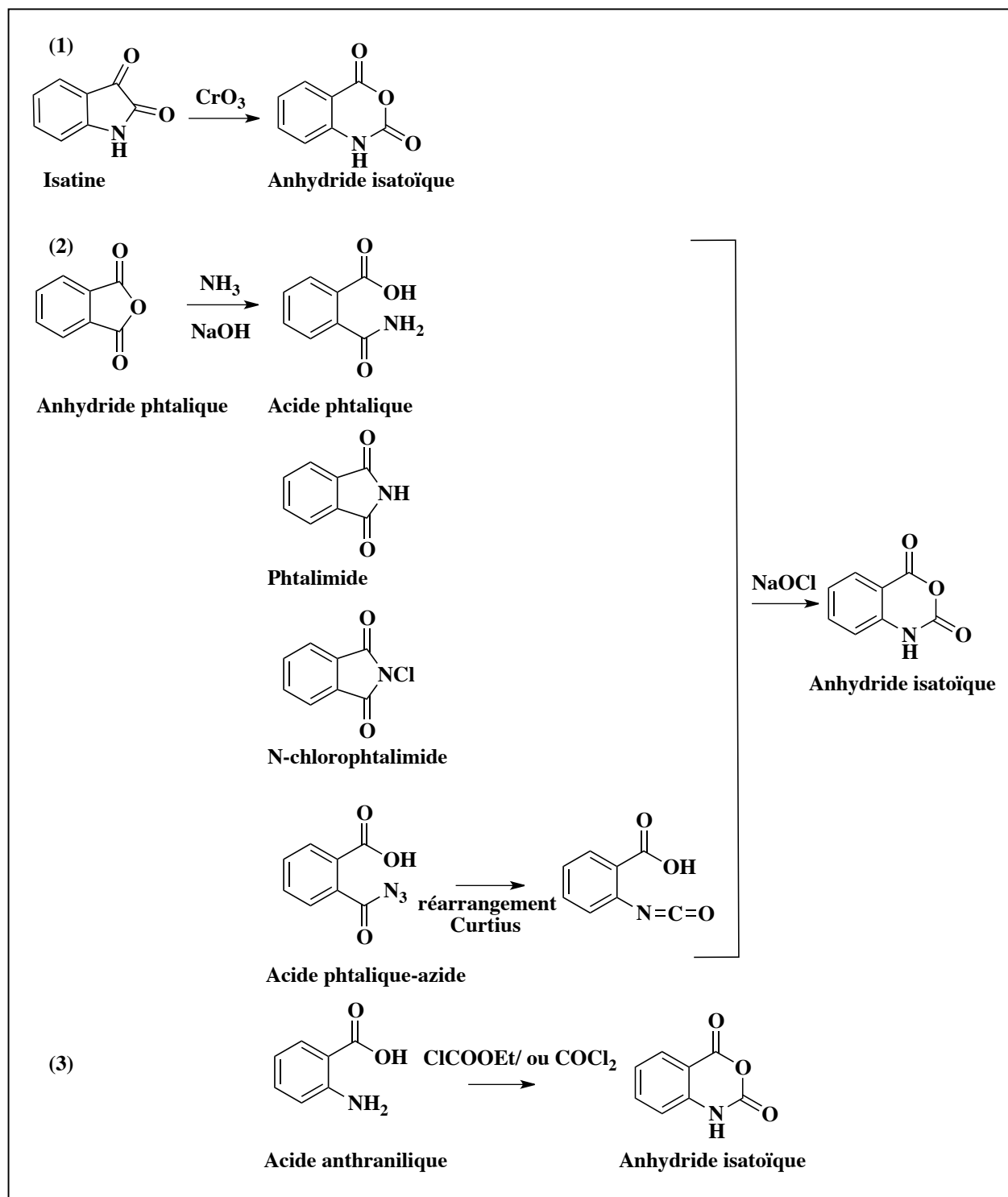


Figure 76: Les trois voies de synthèse de l'anhydride isatoïque. (Adaptée²⁸⁷)

III.3.2 Réactivité chimique

L'anhydride isatoïque est une molécule possédant deux fonctions carbonyles polarophiles, qui peuvent interagir avec une large variété de nucléophiles pour donner les dérivés du 2-aminobenzoyl²⁹⁰.

L'anhydride isatoïque réagit avec l'ammoniac, les amines primaires ou secondaires, et les amides. La réaction se fait selon deux voies: soit par attaque du nucléophile sur le carbonyle en position C4 pour conduire à l'anthranilamide et ses dérivés (voie (a) fig.77), soit par attaque du nucléophile sur le carbonyle en position C2 pour former l'acide O-ureidobenzoïque et ses dérivés (voie (b) fig.77)²⁹¹. Il est intéressant de noter que pour les amines secondaires ou tertiaires encombrées, la réaction évolue uniquement vers la voie (b) car le carbone C2 bien que moins réactif n'est pas affecté par l'encombrement stérique contrairement au carbone C4. Ainsi l'anhydride isatoïque en réaction avec la méthylamine ou l'éthylamine produit l'amide par la voie (a) fig.77, alors que la réaction avec la diéthylamine et la t-butylamine produit majoritairement l'acide uréide par la voie (b) fig.77.

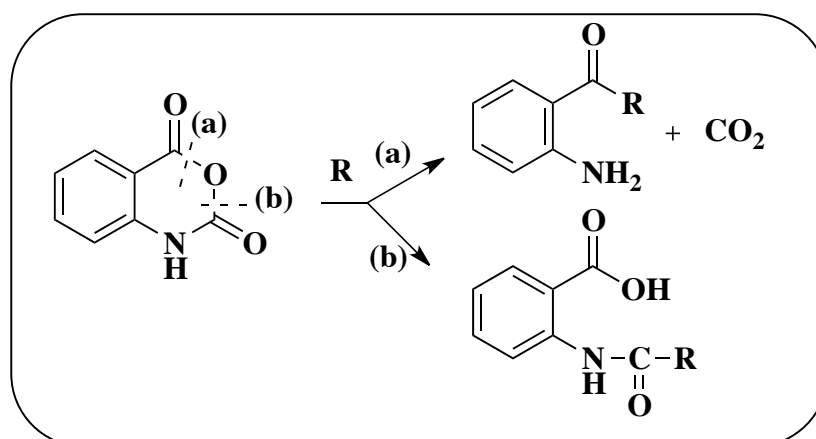


Figure 77: Les deux voies (a) et (b) possibles de la réaction d'un nucléophile (R) sur l'anhydride isatoïque. (Adaptée²⁹²)

L'anhydride isatoïque réagit également avec les alcools primaires et secondaires, les phénols, les thiophénols, les mercaptans (α - ou β - dithiols, et l'éthyle acétoacétate). La réactivité et le déroulement de la réaction dépendent de la nature, de la concentration du nucléophile et de l'encombrement stérique, mais la réaction suit principalement la voie (a) de la fig.77 en produisant des esters et des thioesters de l'acide anthranilique.

III.3.3 Les dérivés de l'anhydride isatoïque

Parmi les dérivés de l'anhydride isatoïque les réactifs de SHAPE, NMIA (N-Methylisatoic anhydride) et 1M7 (1-Methyl-7-Nitroisatoic anhydride) nous intéressent plus particulièrement. Le NMIA est un dérivé de l'anhydride isatoïque substitué au niveau de l'amine secondaire avec une fonction méthyle²⁹³. Le NMIA tout comme l'anhydride isatoïque non substitué, réagit avec une variété de nucléophiles, pour produire des amides, des esters, des thioesters, et de l'acide N-méthylantranilique. En général, la réaction avec le NMIA est plus difficile que l'anhydride non substitué. La réaction avec l'ammoniac et la diméthylamine ne produit pas les dérivés uréide et la réaction avec l'isopropanol ne produit pas les dérivés de carbamate, indiquant que la voie (b) de la « fig.77 » n'est pas applicable. Le mécanisme proposé suivant la voie (a) de la « fig.77 », se fait par une attaque du nucléophile sur le carbonyle C4 de l'anhydride, et le NaOH réagit en tant que catalyseur ionisant l'alcool ou le mercaptan¹⁹⁹. Le carbonyle réactif en position C4 du cycle benzylique est sensible aux substituants en position méta et para. L'anhydride isatoïque possède plusieurs sites de substitution sur le cycle aromatique. La substitution avec des groupements donneurs d'électrons, en position ortho, méta et para. Ainsi la substitution avec des groupements donneurs d'électrons en position ortho ou para par rapport à la position carbonyle « C4 » de l'anhydride isatoïque induit une désactivation de la fonction carbonyle. Par contre, une substitution avec un groupement électro-attracteur, permet de délocaliser les électrons au niveau de la fonction carbonyle et ainsi d'activer cette dernière. L'équipe de « Gelb *et al.* » en 1986 a testé cette hypothèse sur des dérivés d'anhydride isatoïque, afin d'inhiber la trypsine²⁹⁴. Des dérivés d'anhydride isatoïque ont été synthétisés avec une substitution par une fonction aminométhyle (groupement chargé positivement : $-\text{CH}_2\text{NH}_3^+$) en position para par rapport au carbonyle en position C4. Dans les conditions testées, ces composés possédaient des temps de demi-vie inférieurs aux anhydrides non-substitués, par exemple l'anhydride non-substitué possède un $t_{1/2}$ de 3,9 min. tandis que l'anhydride substitué possède un $t_{1/2}$ de 0,7 min. avec une concentration similaire d'inhibiteur. L'équipe de Kevin Weeks¹⁷⁶ a également étudié les réactions d'acylation et d'hydrolyse du composé 1M7. Le composé 1M7 est un dérivé du NMIA substitué avec un groupement nitro « NO_2 » en position para du carbonyle C4. La substitution sur le cycle aromatique par un groupement électro-attracteur fort tel que le NO_2 ($\sigma_p = 0,81$) rend la fonction carbonyle plus réactive vis-à-vis des nucléophiles tels que l'hydroxyle en 2' du ribose des ARN ou les molécules d' H_2O , et stabilise la charge négative

formée dans l'intermédiaire de la réaction. Donc ce composé possède un temps de demi-vie court par rapport aux autres réactifs.

III.4 Notre approche: synthèse de nouveaux réactifs chimiques

Nous venons de voir que les dérivés de l'anhydride isatoïque tels que le NMIA et le 1M7 sont de bons réactifs chimiques pour sonder la structure des ARN *in vivo* et que ces molécules chimiques peuvent être fonctionnalisées facilement sur l'atome d'azote N1. Pour mon projet de thèse, nous avons donc choisi d'introduire un groupement alcyne au niveau de cet atome d'azote afin de pouvoir, après réaction avec l'ARN, lier la biotine-azoture aux ARN modifiés par réaction de click chemistry. Les ARN modifiés pourront par la suite être sélectionnés grâce à la reconnaissance biotine/streptavidine (Fig.78). Cet enrichissement du milieu en ARN modifiés permettra, comme dans les techniques de PARS et Frag-seq, de réduire le bruit de fond en effectuant la rétro-transcription puis le séquençage uniquement sur les séquences d'intérêt.

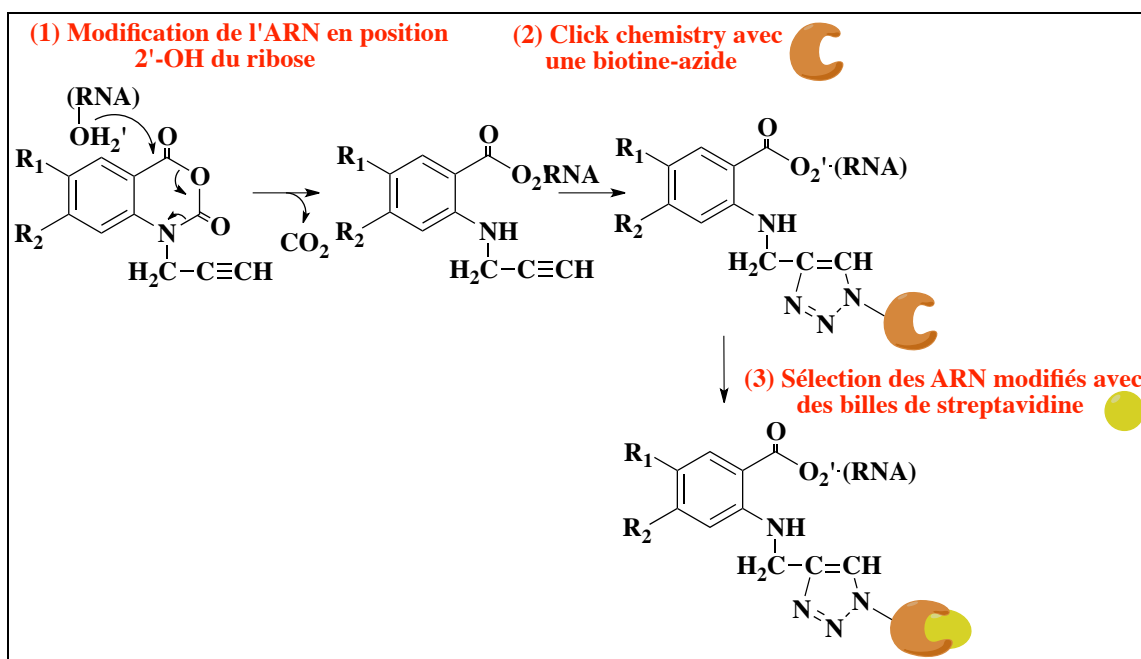


Figure 78: Les différentes étapes de la stratégie de sélection des ARN

Nous avons également fait varier la nature et la position des substituants sur le cycle aromatique afin d'obtenir des composés qui réagiront plus ou moins vite avec l'ARN (Fig.79). De plus, nous avons préparé les anhydrides isatoïques N-méthylés afin d'avoir des molécules de référence.

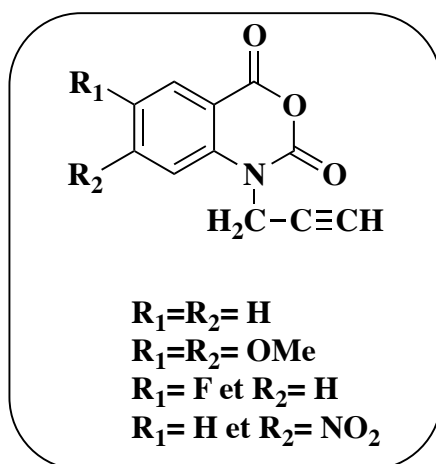


Figure 79: Dérivés alcynes des anhydride isatoïque synthétisés

Parmi les composés qui font l'objet de ce chapitre de thèse, certains ont déjà été synthétisés et rapportés dans la littérature, il s'agit du 1-N-(propyn-2-yl)-6,7-dimethoxyisatoic anhydride²⁹⁵ et le 1-N-propyn-2-yl-isatoic anhydride qui est commercial chez Ukrorgsyntez Building Blocks (CAS N° 50784-22-0).

III.5 MATÉRIELS ET MÉTHODES

III.5.1 Préparation du matériel, techniques de purification et techniques de caractérisation

- Les techniques mises en œuvre lors de la synthèse, la purification et la caractérisation des composés décrits dans ce chapitre sont identiques à celles développées dans le chapitre II.
- Le tampon préparé pour la cinétique d'hydrolyse est composé d'une solution de MgCl_2 (0,1M), NaCl (1M), tampon Hépès pH=8 (1M) dans H_2O .
- Le spectrophotomètre utilisé pour déterminer les temps de demi-vie des réactifs chimiques est de type UVIKON 943/ Double faisceau UV/Vis. Les cuves utilisées pour les mesures de cinétique sont en quartz.
- Gel d'agarose à 1% (TBE 1x) en présence de BET (0,5 $\mu\text{g}/\text{ml}$). La migration se déroule pendant 30 min. sous un champ constant de 120 V (10 V/cm) dans un TBE 1x (89 mM Tris-Borate pH 8,3 ; 2 mM EDTA).
- Gel dénaturant de polyacrylamide 8% (urée 8M, acrylamide/bisacrylamide 19/1, TBE 1x). La polymérisation du gel est initiée par ajout de 0,05% d'APS et 0,002% de TEMED. Après électrophorèse, le gel est fixé dans un tampon de fixation (100 ml acide acétique, 200 ml éthanol 100%, 1,7 l eau) puis séché. Une « Imaging Plate » Fujifilm (lue par un FLA-5000 de Fuji) est utilisée pour révéler les bandes radioactives.
- Tampon RTB 5x (extension d'amorce par la rétro-transcriptase AMV RT) : Tris-HCl 250 mM (pH=8,5), MgCl_2 30 mM, KCl 200 mM (solution commerciale 10x) diluée par 2 avec H_2O .
- RT AMV 2U (extension d'amorce) : Diluer par facteur de 10 le minimum de solution RT AMV 10x par du tampon RTB 1x. Préparer un mix : 100 μl dATP 10 mM + 100 μl dGTP 10 mM + 100 μl dTTP 10 mM + 100 μl dCTP 10 mM.

III.5.2 Synthèses

Protocole de synthèse des dérivés de l'anhydride isatoïque :

A une solution de l'acide 2-amino-benzoïque d'intérêt (0,27 M) dans du THF anhydre (3,8 ml/mmoles) est ajouté le triphosgène (0,35 eq, 0,09 M) et la réaction est effectuée à température ambiante pendant 18h. Le composé qui se forme est précipité en versant le brut réactionnel sur 10 ml d'eau glacée. Le précipité est recueilli, rincé à l'eau froide et séché au dessiccateur, il sera utilisé tel quel sans purification supplémentaire.

6,7-Diméthoxy-anhydride isatoïque (composé 20, fig.80) : masse obtenue= 200mg, rendement= 90%

RMN ^1H (DMSO- d_6 , 400MHz) : 7,24 (1H, s, H5) ; 6,63 (1H, s, H8) ; 3,85 (1H, s, CH_3) ; 3,79 (1H, s, CH_3). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 159,35 ($\text{CO}_{(4)}$) ; 156,52 (C-7) ; 147,39 ($\text{CO}_{(3)}$) ; 145,64 (C-6) ; 137,60 (C-2) ; 108,54 (CH-5) ; 101,18 (C-1) ; 97,67 (CH-8) ; 56,08 et 55,86 ($2\times\text{CH}_3$). **ESI (m/z) mode positif :** 224 [$\text{M}+\text{H}^+$].

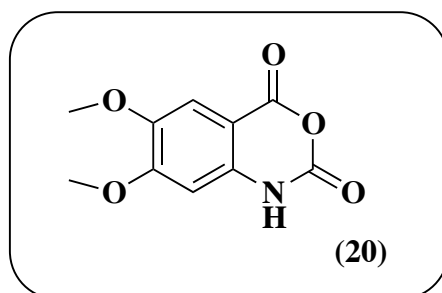


Figure 80: 6,7-diméthoxy-anhydride isatoïque/ $\text{C}_{10}\text{H}_9\text{NO}_5$ / Masse exacte= 223.05

7-Nitro-anhydride isatoïque (composé 21, fig.81) : masse obtenue= 527 mg, rendement= 51%

RMN ^1H (DMSO- d_6 , 400 MHz) : 12,09 (1H, s, NH) ; 8,17-8,14 (1H, d, $J = \text{Hz}$, H6) ; 7,97-7,94 (1H, d, H5) ; 7,86 (1H, s, H8). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 158,75 ($\text{CO}_{(4)}$) ; 151,91 ($\text{CO}_{(3)}$) ; 146,61 (C-7) ; 142,06 (C-2) ; 117,27 (CH-6) ; 115,46 (C-1) ; 110,18 (CH-8). **ESI (m/z) mode positif :** 209 [$\text{M}+\text{H}^+$]

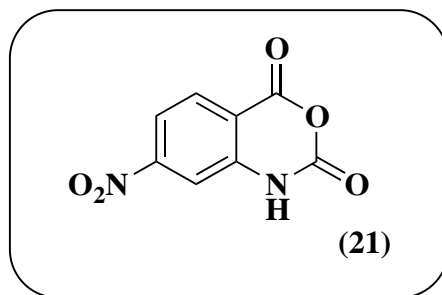


Figure 81: 7-nitro-anhydride isatoïque/ $C_8H_4N_2O_5$ / Masse exacte= 208.01

6-Fluoro-anhydride isatoïque (composé 22, fig.82) : masse obtenue= 579 mg, rendement= 80%

RMN 1H (DMSO- d_6 , 400 MHz) : 11,77 (1H, s, NH) ; 7,62-7,62 (2H, m, H7-H8) ; 7,19-7,16 (1H, m, H5). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 159,16 (C-6) ; 156,30 (CO₍₄₎) ; 146,76 (CO₍₃₎) ; 138,09 (C-2) ; 124,88-124,63 (CH-7) ; 117,59-117,51 (CH-8) ; 114,08-113,84 (CH-5) ; 111,49-111,41 (C-1). **ESI (m/z) mode positif :** 182 [M+H⁺].

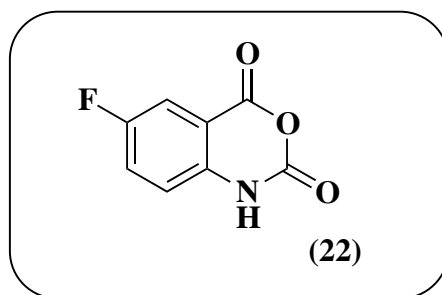


Figure 82: 6-fluoro-anhydride isatoïque/ $C_8H_4FNO_3$ / Masse exacte= 181.02

Protocole général de synthèse des dérivés du 1-N-propyn-2-yl-anhydride isatoïque

L'anhydride isatoïque commercial (SIGMA-ALDRICH) ainsi que les dérivés de l'anhydride isatoïque préparés précédemment (0,2 M) sont mis en suspension dans le DMF anhydre (5 ml/mmoles). L'hydruide de sodium (0,25 M, 1,25 eq) est ajouté par petites fractions. Après 30 min d'agitation, le bromure de propargyle (0,24 M, 1,2 eq) est ajouté et la réaction est poursuivie pendant 20 heures, à température ambiante.

1-N-propyn-2-yl-anhydride isatoïque (composé 23, fig.83) : masse obtenue= 113 mg, rendement= 91%

Le brut réactionnel est concentré sous pression réduite puis purifié sur colonne de gel de silice avec un gradient AcOEt (0 à 40%) dans l'éther de pétrole.

RMN ^1H (DMSO- d_6 , 400MHz) : 8,05-8,02 (1H, d, $J = \text{Hz}$, H5) ; 7,90-7,87 (1H, t, $J = \text{Hz}$, H7) ; 7,52-7,49 (1H, d, $J = \text{Hz}$, H8) ; 7,41-7,35 (1H, t, $J = \text{Hz}$, H6) ; 4,90-4,88 (2H, d, $J = 2,4 \text{ Hz}$, CH_2) ; 3,42 (1H, t, $J = 2,4 \text{ Hz}$, CH). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 158,42 ($\text{CO}_{(4)}$) ; 147,24 ($\text{CO}_{(3)}$) ; 140,41 (C-2) ; 137,11 (CH-7) ; 129,58 (CH-5) ; 124,04 (CH-6) ; 114,98 (CH-8) ; 111,81 (C-1) ; 77,54 ($-\text{C}\equiv$) ; 76,00 ($\equiv\text{CH}$) ; 34,31 (CH_2). **ESI (m/z) mode positif :** 202 $[\text{M}+\text{H}^+]$.

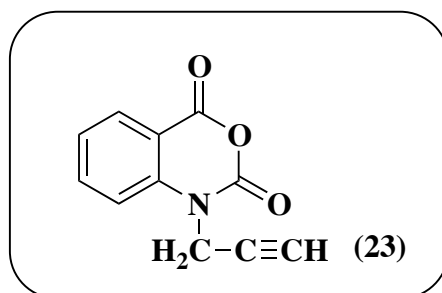


Figure 83: 1-N-propyn-2-yl-anhydride isatoïque/ $\text{C}_{11}\text{H}_7\text{NO}_3$ / Masse exacte= 201.04

1-N-(Propyn-2-yl)-6,7-diméthoxy-anhydride isatoïque (composé 24, fig.84) : masse obtenue= 32 mg, rendement= 33%

Le produit brut est précipité par ajout d'un mélange eau-glace, lavé puis séché sous vide.

RMN ^1H (DMSO- d_6 , 400 MHz) : 7,37 (1H, s, H5) ; 7,00 (1H, s, H8) ; 4,95-4,94 (2H, d, $J = 2,4 \text{ Hz}$, CH_2) ; 3,97 (3H, s, CH_3) ; 3,83 (3H, s, CH_3) ; 3,40 (1H, t, $J = 2,4 \text{ Hz}$, CH). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 157,89 ($\text{CO}_{(4)}$) ; 156,36 (C-7) ; 147,59 ($\text{CO}_{(3)}$) ; 145,69 (C-6) ; 136,66 (C-2) ; 109,54 (CH-5) ; 102,82 (C-1) ; 98,50 (CH-8) ; 77,66 ($-\text{C}\equiv$) ; 76,01 ($\equiv\text{CH}$) ; 56,52 (CH_3) ; 55,97 (CH_3) ; 34,41 (CH_2). **ESI (m/z) mode positif :** 262 $[\text{M}+\text{H}^+]$.

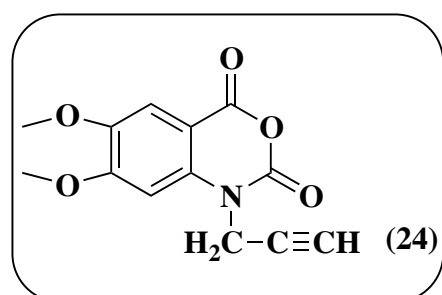


Figure 84: 1-N-(Propyn-2-yl)-6,7-diméthoxy-anhydride isatoïque/ $\text{C}_{13}\text{H}_{11}\text{NO}_5$ / Masse exacte= 261.06

1-N-(propyn-2-yl)-7-nitro-anhydride isatoïque (composé 25, fig.85) : masse obtenue= 81 mg, rendement= 34%

Le produit brut est précipité par ajout d'une solution 1N de HCl (6 ml) puis séché sous vide.

RMN ^1H (DMSO- d_6 , 400 MHz) : 8,03 (1H, s, H8) ; 7,89 (1H, d, $J = 1,2$ Hz, H5) ; 7,86 (1H, d, $J = 1,2$ Hz, H6) ; 4,79 (2H, d, $J = 2,5$, CH_2) ; 3,4 (1H, t, $J = 2,5$ Hz, CH). **ESI (m/z) mode positif :** 247 $[\text{M}+\text{H}^+]$.

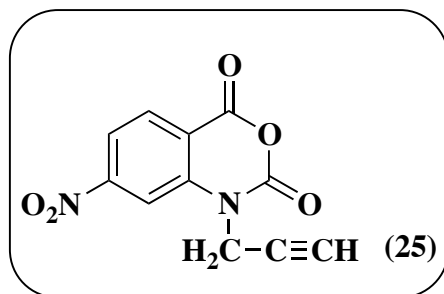


Figure 85: 1-N-(propyn-2-yl)-7-nitro-anhydride isatoïque/ $\text{C}_{11}\text{H}_6\text{N}_2\text{O}_5$ / Masse exacte= 246.03

1-N-(Propyn-2-yl)-6-fluoro-anhydride isatoïque (composé 26, fig.86) : masse obtenue= 138 mg, rendement= 56%

Le brut réactionnel est concentré sous pression réduite puis purifié sur colonne de gel de silice avec un gradient AcOEt (25 à 50%) dans le cyclohexane.

RMN ^1H (DMSO- d_6 , 400 MHz) : 7,57-7,48 (1H, m, H7) ; 7,26-7,22 (1H, m, H8) ; 6,78-6,74 (1H, m, H5) ; 4,90-4,88 (2H, m, CH_2) ; 3,60 (1H, t, $J = \text{Hz}$, CH). **ESI (m/z) mode positif :** 220 $[\text{M}+\text{H}^+]$.

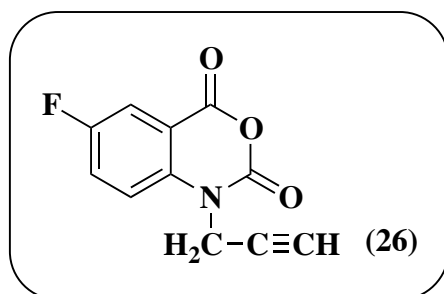


Figure 86: 1-N-(propyn-2-yl)-6-fluoro-anhydride isatoïque : $\text{C}_{11}\text{H}_6\text{FNO}_3$ / Masse exacte= 219.03

Protocole général de synthèse des dérivés du 1-N-méthyl-anhydride isatoïque:

Le protocole est identique à celui décrit précédemment pour la préparation des dérivés 1-N-propyn-2-yl de l'anhydride isatoïque à l'exception du bromure de propargyle qui est remplacé par de l'iodométhane (0,24 M, 1,2 eq.). Les réactions sont effectuées à température ambiante pendant 18 à 24 heures.

1-N-méthyl-anhydride isatoïque (NMIA) (composé 27, fig.87): masse obtenue= 83 mg, rendement= 46%

Le produit brut est précipité par ajout d'une solution 1N de HCl (6 ml) puis séché sous vide.

RMN ^1H (DMSO- d_6 , 400MHz) : 8,02-7,99 (1H, d, J = Hz, H5) ; 7,86-7,83 (1H, t, J = Hz, H7) ; 7,46-7,43 (1H, d, J = Hz, H8) ; 7,37-7,31 (1H, t, J = Hz, H6) ; 3,46 (3H, s, CH_3). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 158,97 ($\text{CO}_{(4)}$) ; 147,71 ($\text{CO}_{(3)}$) ; 142,17 (C-2) ; 137,14 (CH-7) ; 129,27 (CH-5) ; 123,54 (CH-6) ; 114,80 (CH-8) ; 111,50 (C-1) ; 31,64 (N- CH_3). **ESI (m/z) mode positif :** 178 $[\text{M}+\text{H}^+]$.

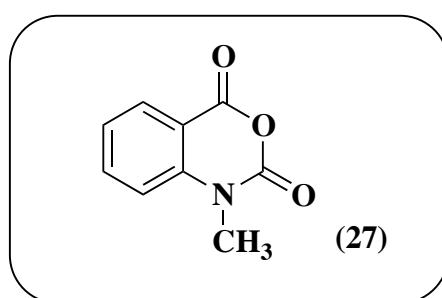


Figure 87: 1-N-méthyl-anhydride isatoïque : $\text{C}_9\text{H}_7\text{NO}_3$ / Masse exacte= 177.04

1-N-méthyl-6,7-diméthoxy-anhydride isatoïque (composé 28, fig.88) : masse obtenue 143 mg, rendement= 50%

Le brut réactionnel est versé sur 10ml de H_2O froide, et le précipité est recueilli, rincé et séché au dessiccateur.

RMN ^1H (DMSO- d_6 , 400 MHz) : 7,34 (1H, s, H5) ; 6,87 (1H, s, H8) ; 3,97 (3H, s, CH_3) ; 3,82 (3H, s, CH_3) ; 3,49 (3H, s, N- CH_3). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 158,39 ($\text{CO}_{(4)}$) ; 156,50 (C7) ; 147,98 ($\text{CO}_{(3)}$) ; 145,41 (C-6) ; 138,51 (C-2) ; 109,25 (CH-5) ; 102,52 (C-1) ; 98,24 (CH-8) ; 56,52 ($\text{CH}_3(\text{OMe})$) ; 55,91 ($\text{CH}_3(\text{OMe})$) ; 31,84 (N- CH_3). **ESI (m/z) mode positif :** 238 $[\text{M}+\text{H}^+]$.

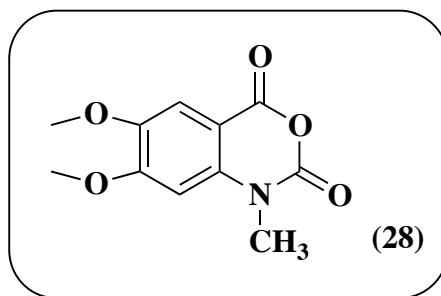


Figure 88: 1-N-méthyl-6,7-diméthoxy-anhydride isatoïque/ $C_{11}H_{11}NO_5$ / Masse exacte= 237.06

1-N-méthyl-7-nitro-anhydride isatoïque (1M7) (composé 29, fig.89): masse obtenue= 127 mg, rendement= 59%

Le produit brut est précipité par ajout d'une solution 1N de HCl (6 ml) puis séché sous vide.

RMN 1H (DMSO- d_6 , 400 MHz) : 8,26-8,23 (1H, d, J = Hz, H6) ; 8,12 (1H, s, H8) ; 8,07-8,04 (1H, d, H5) ; 3,55 (3H, s, CH₃). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 152,41 (CO₍₄₎) ; 147,39 (C-7) ; 143,06 (C-2) ; 131,13 (CH-5) ; 117,55 (CH-6) ; 116,65 (C-1) ; 109,87 (CH-8) ; 32,11 (CH₃). **ESI (m/z) mode positif :** 223 [M+H⁺].

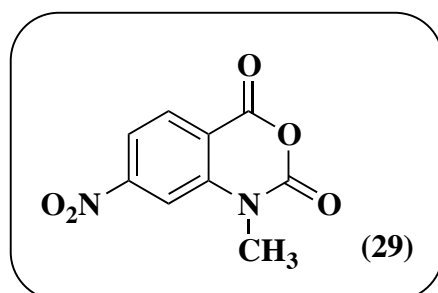


Figure 89: 1-N-méthyl-7-nitro-anhydride isatoïque (1M7) : $C_9H_6N_2O_5$ / Masse exacte= 222.03

1-N-méthyl-6-fluoro-anhydride isatoïque (composé 30, fig.90) : masse obtenue= 120 mg, rendement= 50%

Le brut réactionnel est concentré sous pression réduite puis purifié sur colonne de gel de silice avec un gradient AcOEt (25 à 50 %) dans le cyclohexane.

RMN 1H (DMSO- d_6 , 400 MHz) : 7,50-7,45 (1H, m, H7) ; 7,32-7,26 (1H, m, H8) ; 6,70-6,66 (1H, m, H5) ; 2,81 (3H, s, CH₃). **RMN ^{13}C (DMSO- d_6 , 400 MHz) :** 168,94 (C-6) ; 153,48 (CO₍₄₎) ; 150,43 (CO₍₃₎) ; 148,69 (C-2) ; 122,13 (CH-7) ; 117,30-116,20 (CH-5) ; 112,19-112,10 (CH-8) ; 109,73-109,64 (C-1) 29,52 (CH₃). **ESI (m/z) mode positif :** 196 [M+H⁺].

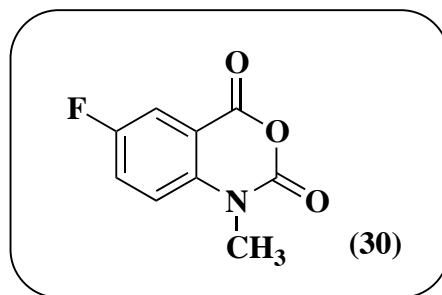


Figure 90: 1-N-méthyl-6-fluoro-anhydride isatoïque/ $C_9H_6FNO_3$ / Masse exacte= 195.03

III.5.3 Modification de l'ARN par l'anhydride 1-(N-propyn-2-yl)isatoïque:

L'ARN 1-311 Mal de HIV-1 (19,5 pmoles) dans de l'eau (2,77 μ l) est dénaturé 1 min. à 95°C, puis mis 1 min. dans la glace avant d'ajouter le tampon 3x (333 mM NaCl, 1 mM Hépès NaOH pH 7,5, 33,3 mM $MgCl_2$) (3 μ l). Après 20 min. d'incubation à 37°C, (3 μ l) de réactif (0, 5 et 15 mM) sont ajoutés et la réaction est continuée pendant 40 min. à 37°C. L'ARN est précipité par ajout d'EtOH 100 (30 μ l), d'AcONa 3M (4 μ l) et de glycogène (1 μ l). Après 30 min. dans un bain éthanol/carboglance, la solution est centrifugée à 4°C et 12000 rpm pendant 20 min. Le surnageant est éliminé, le culot est rincé deux fois avec (20 μ l) d'EtOH 70% et centrifugé pendant 5 min. à 13000 rpm, avant d'être séché 5 min. au speed-vac. Le réactif hydrolysé en excès est éliminé par centrifugation à travers un Centricon 30K à 12000 rpm pendant 5 min. La concentration en ARN est déterminée par mesure de l'absorbance à 260 nm et l'utilisation de la loi de Beer-Lambert. La présence des modifications sur l'ARN est vérifiée par extension d'une amorce fluorescente par une rétro-transcriptase et analyse des ADNc par électrophorèse capillaire. L'amorce fluorescente (fluorophore FAM pour l'ARN modifié et VIC pour l'ARN non modifié (témoin négatif); (1 μ l à 0,9 μ M) est ajoutée à (0,2 μ g) d'ARN dont la structure est dénaturée par incubation 2 min à 90°C puis 2 min dans la glace. (3 μ l) de tampon RT 5x (120 mM TrisHCl, pH 8.3, 250 mM KCl, 10 mM DTT, 25 mM $MgCl_2$) sont ensuite ajoutés ainsi que 2 U de la RT du virus du myéloblastome aviaire (AMV), (0,5 μ l) d'eau et (3 μ l) d'un mélange équimolaire des 4 dNTP (2,5 mM chacun). Après une incubation de 30 min à 42°C puis 15 min à 50°C, le volume réactionnel est ajusté à (50 μ l) avec de l'eau, puis une extraction au phénol/chloroforme (1 volume) est réalisée. En parallèle, deux réactions de séquençage sont effectuées. Pour cela, (0,4 μ g) d'ARN de l'isolat NL4-3 du VIH-1 sont hybridés à 1 μ l (1,8 μ M) d'amorce fluorescente (fluorophores NED et PET). Après hybridation de l'amorce, la rétro-transcription est effectuée comme ci-dessus

hormis pour le tampon de dimérisation qui est 10x concentré et les dNTP : au milieu réactionnel est ajouté 2 μ l de ddXTP 100 μ M (X=A et T) et 3 μ l de X10 (X10 étant un mélange de dXTP à 2,5 mM et des autres dNTP à 10 mM).

L'ARN des 4 conditions (+ et – réactif acylant et les 2 réactions de séquençage) sont mélangés, précipités, lavés et séchés comme décrit précédemment. Le culot est repris dans (10 μ l) de High Difformamide (Applied Biosystems). Après centrifugation (18 000 g, 10 min, 4°C), les échantillons sont incubés 5 min à 90°C puis 5 min dans la glace puis chargés sur la plaque de séquençage.

III.5.4 Click chemistry entre l'ARN porteur de la fonction alcyne et la biotine-PEG₃-azoture :

À une solution d'ARN modifié (1,5 μ g) dans 20 μ l d'eau, on ajoute (7 μ l) d'une solution (0,1 mM) de biotine-PEG₃-azoture (Jena Bioscience). L'oxygène est éliminé de la solution par bullage d'argon puis (5,6 μ l) d'une solution de CuSO₄.5H₂O/TBTA (0,1 mM) dans H₂O/DMSO/tBuOH (50/37,5/12,5) sont ajoutés. Le mélange est à nouveau dégazé avant d'ajouter (5,6 μ l) d'une solution (0,1 mM) d'ascorbate de sodium dans de l'eau. La réaction est réalisée à 37°C sous agitation pendant 3h. Le mélange réactionnel est dégazé toutes les heures. Après réaction, le volume est complété à (100 μ l) avec de l'eau. L'ARN est précipité comme précédemment. Le culot est repris dans (10 μ l) d'eau et la concentration en ARN est contrôlée par mesure de l'absorbance à 260 nm.

Visualisation de l'interaction Biotine/streptavidine :

À une solution d'ARN modifié et couplé à la biotine-PEG₃-azoture (500 ng) est ajouté 0,5 μ l de tampon VIH 10x (Tris-HCl pH 8.3 500 mM, KCl 500 mM, MgCl₂ 60 mM, DTE/DTT 10 mM). Les échantillons sont complétés avec de l'eau afin d'obtenir un volume final de (5 μ l) puis sont incubés à 37°C pendant 15 min en présence ou en absence de streptavidine (1 μ g). Les échantillons sont additionnés de (2 μ l) de bleu de glycérol puis déposés sur gel d'agarose 1% en présence de BET. La migration se déroule pendant 30 min à un voltage constant de 120 V (10 V/cm).

III.5.5 Protocole général des cinétiques d'hydrolyse

La cinétique d'hydrolyse est réalisée par suivi de l'absorbance UV en fonction du temps. Au préalable, un spectre d'absorption du produit de départ (3 mmoles) en fonction de la longueur d'onde est réalisé dans le DMSO (150 μ l) sur le nanodrop. Un deuxième spectre est réalisé dans les mêmes conditions après hydrolyse de l'anhydride isatoïque (Fig.91). Pour cela, l'anhydride isatoïque (30 mmoles) est traité par 150 μ l de tampon 1,1x (6,7 mM $MgCl_2$, 111 mM NaCl, 111 mM Hépès pH 8.0) puis (15 μ l) de cette solution sont dilués 10x par du DMSO. Les cinétiques sont réalisées aux longueurs d'onde décrites dans le tableau 3. Une solution (1,5 mM) d'anhydride isatoïque (150 μ l) dans le DMSO anhydre est ajoutée à (1350 μ l) de tampon 1,1x au temps t_0 . L'absorbance est mesurée toutes les 3s pendant 30 min pour les composés dont l'hydrolyse est rapide ou 50 min pour ceux dont l'hydrolyse est plus lente. Les données sont traitées avec le logiciel Prism. Les temps de demi-vie sont obtenus en appliquant l'équation $Y = Y_0 + (Plateau - Y_0) * (1 - \exp(-K * x))$ qui représente une cinétique d'association de pseudo 1^{er} ordre.

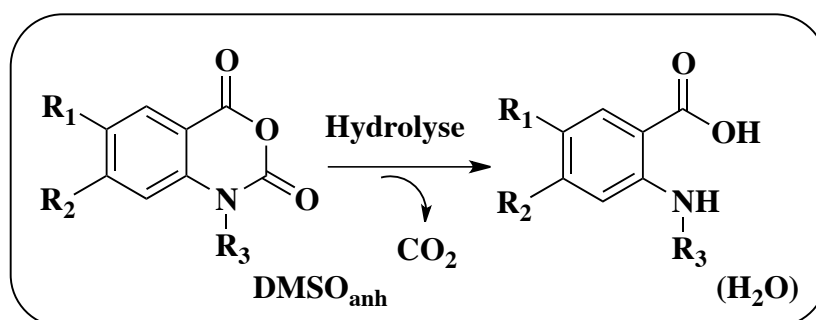


Figure 91 : Hydrolyse de l'anhydride isatoïque

Un exemple qui montre les spectres d'absorbance du composé 1M7 durant l'étape d'hydrolyse (Fig.92). On remarque le changement de la longueur d'onde d'absorption entre le réactif de départ et celui hydrolysé (passage 360 nm à 450 nm).

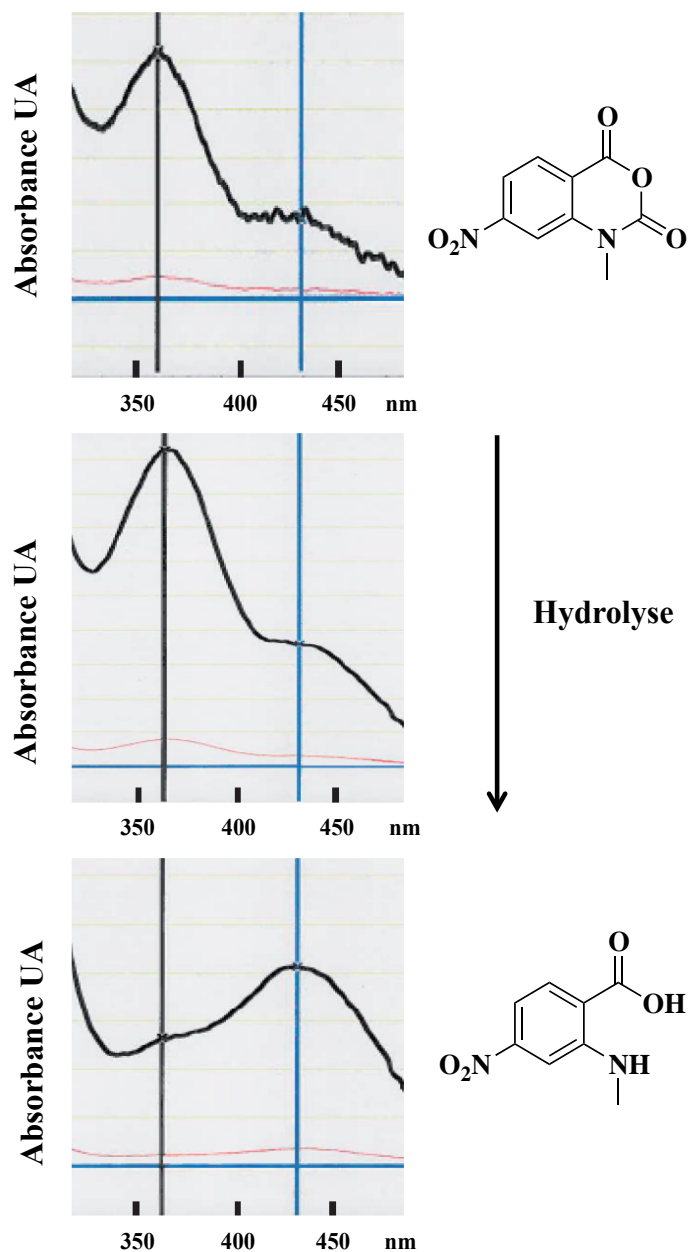


Figure 92: Spectres d'absorbance du composé 1M7 au cours de l'hydrolyse

III.6 RÉSULTATS

Synthèse des dérivés d'anhydride isatoïque

Le protocole se divise en deux étapes, la première étape consiste à synthétiser les différents dérivés de l'anhydride isatoïque substitués sur le cycle benzène, avec les différentes fonctions en position méta et para par rapport au carbonyle en position C4 (F, NO₂, et (OMe)₂) de l'anhydride (composé 18, fig.93). Cette étape est réalisée avec l'acide 4,5-diméthoxy-2-amino-benzoïque, l'acide 5-fluoro-2-amino-benzoïque, l'acide 4-nitro-2-amino-benzoïque^{296,297,298}. La deuxième étape de la synthèse consiste à modifier ces dérivés diversement substitués sur le cycle benzène avec la fonction alcyne, ou méthyle au niveau de l'amine secondaire N1^{299,300,301,293} (composé 19, fig.93).

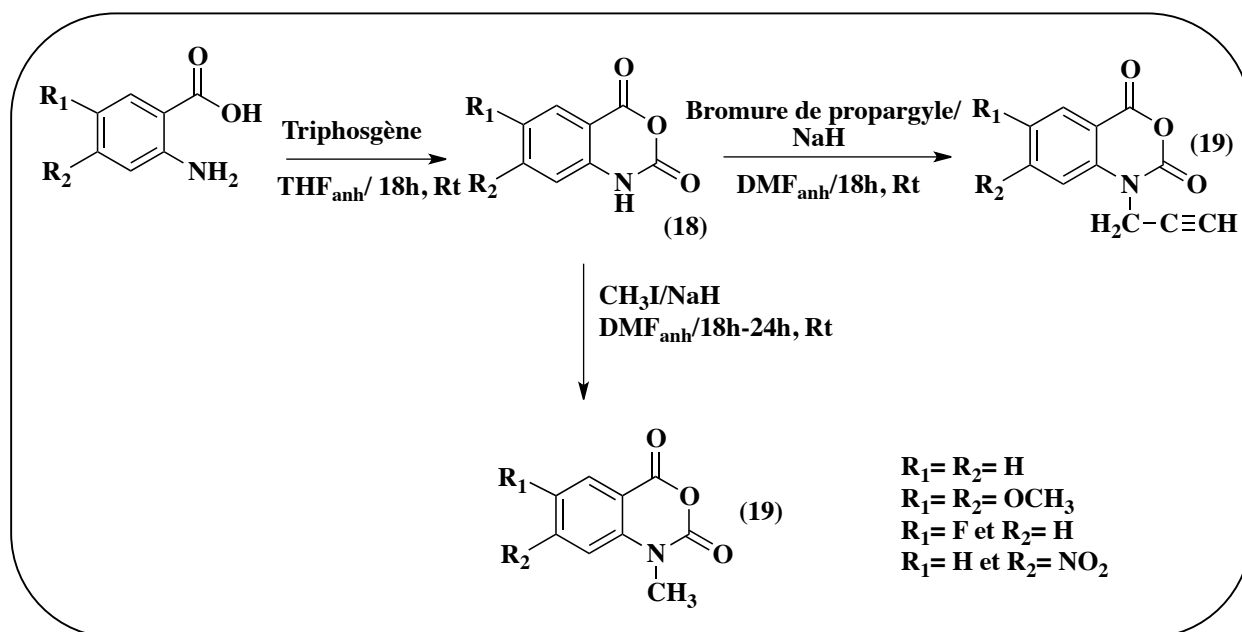
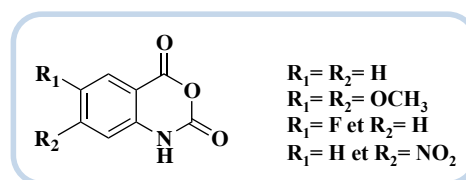
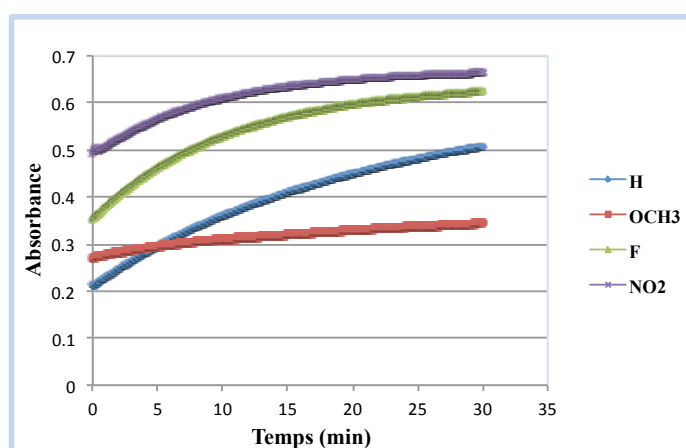


Figure 93: Synthèse des anhydrides substitués sur leurs cycle phényle (18), et de leurs dérivés substitués sur l'amine (19) avec soit une fonction méthyle ou une fonction alcyne.

Etude cinétique : suivie des temps de demi-vie de ces composés dans le tampon d'hydrolyse :

La substitution sur le cycle benzène fait varier le temps de demi-vie de ces dérivés d'anhydride isatoïque (résultats détaillés dans le tableau 3). Les résultats obtenus pour le NMIA et le 1M7 sont identiques à ceux obtenus dans la littérature par l'équipe de Kevin Weeks. Une large gamme de temps de demi-vie a été obtenue, en variant la position du substituant sur le cycle en fonction du carbonyle en C4, que ce soit en position para avec le

groupement nitro, ou méta avec l'atome de fluor, ou à la fois méta et para avec le groupement méthoxy. De manière générale, on peut remarquer que dans une même série (R3 identique) le temps de demi-vie varie dans le sens $\text{NO}_2 < \text{F ou H} < \text{OMe}$. Cette variation peut s'expliquer par les effets mésomères des substituants. En effet, le NO_2 qui est fortement électroattracteur et qui est placé en para du carbonyle renforce fortement le caractère $+\delta$ du carbone, de même le fluor qui a un effet électrodonneur léger mais qui est situé en méta du carbonyle augmente légèrement la charge $+\delta$ du carbonyle. Par contre le groupement méthoxy, électrodonneur, situé en para du carbonyle diminue fortement la charge $+\delta$ portée par le carbone du CO et ceci n'est pas compensé par la présence du deuxième groupement méthoxy situé en méta du carbonyle. Nos résultats indiquent également que la nature du substituant porté par l'amine, N1, joue un rôle sur le temps de demi-vie de nos composés. Dans ce cas, les effets sont plus difficilement explicables. En effet, la substitution de l'amine par un groupement propyn-2-yl diminue le $t_{1/2}$ du dérivé non substitués sur le cycle aromatique (composé 23 : tableau 3) plus fortement que la substitution par un groupement méthyle (composé 27 : tableau 3) alors que l'effet est inverse en présence du NO_2 (composés 25 et 29 : tableau 3). On peut émettre l'hypothèse qu'au niveau de l'amine N1 l'effet inductif des substituants introduits sur l'amine s'ajoutent ou se soustraient aux effets mésomères des substituants présents sur le cycle et que cela change la stabilité des intermédiaires de réaction. La variation des temps de demi-vie de ces dérivés devrait permettre leur utilisation dans les processus cellulaires qui nécessitent, soit une réaction directe et rapide avec l'ARN (composés à durée de vie courte), soit une réaction plus longue (composés à durée de vie élevée).



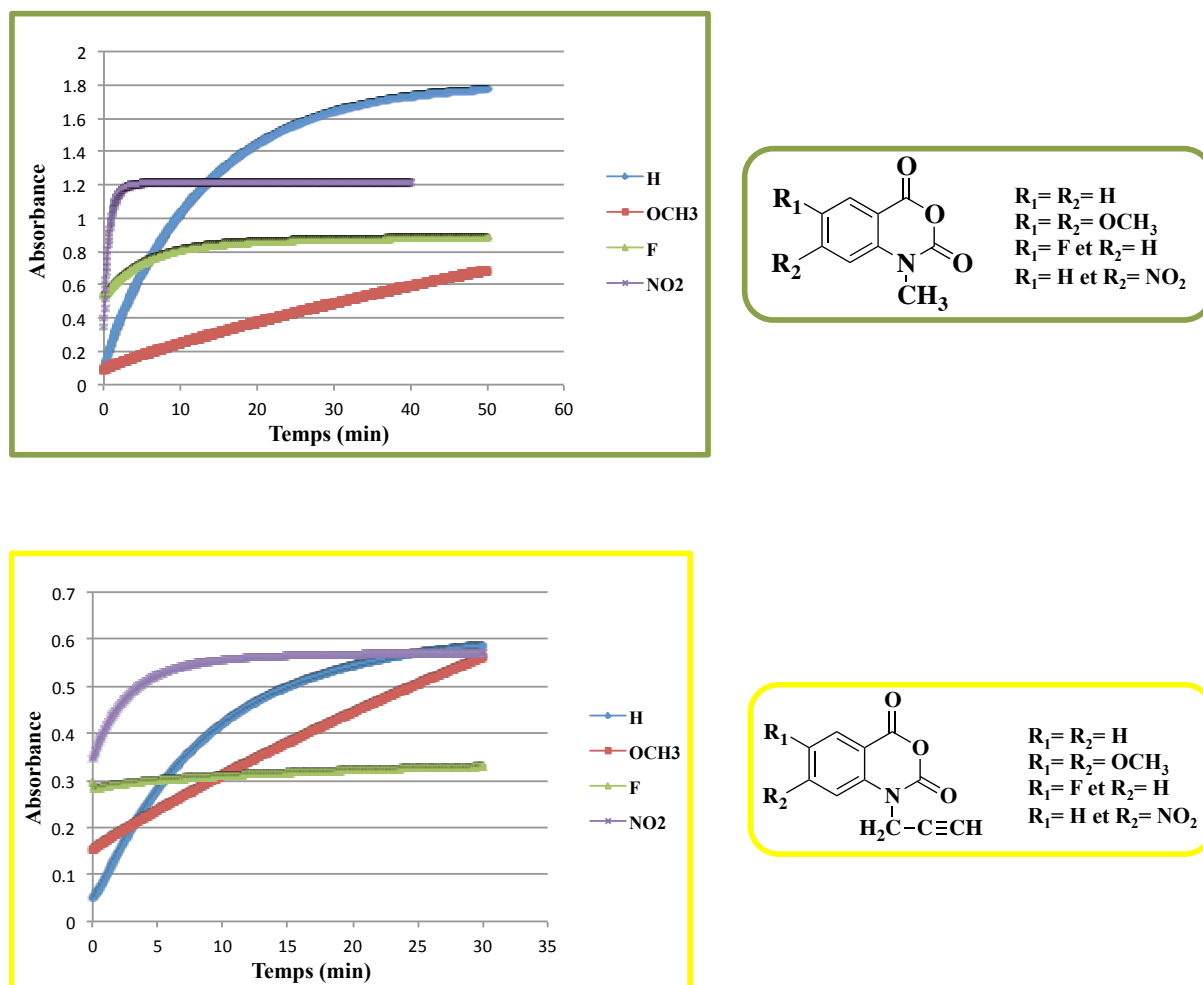


Figure 94: Courbes de cinétique d'hydrolyse (pH=8.0) obtenues pour les différents dérivés d'anhydride isatoïque en mesurant l'absorbance UV en fonction du temps.

N° composé	R ₁	R ₂	R ₃	T°C	Abs nm	T _{1/2} min
Isatoïc anh	H	H	H	20	360	21
20	OMe	OMe	H	20	365	24,5
21	H	NO ₂	H	20	420	6,4
22	F	H	H	20	360	7,5
23	H	H	H ₂ C—C≡CH	20	365	6,3
24	OMe	OMe	H ₂ C—C≡CH	20	365	51
25	H	NO ₂	H ₂ C—C≡CH	20	420	2,2
26	F	H	H ₂ C—C≡CH	20	360	13,4
27	H	H	CH ₃	20	360	13,5
28	OMe	OMe	CH ₃	20	365	62
29	H	NO ₂	CH ₃	22	450	0,8
30	F	H	CH ₃	20	360	4,4

Tableau 3: Résultats de temps de demi-vie calculés pour les différents dérivés d'anhydride isatoïque synthétisés.

Modification/ rétro-transcription de l'ARN 1-311 du VIH-1 avec le composé:

Afin de vérifier que nos composés comportant une fonction alcyne sont capables de modifier l'ARN comme les réactifs SHAPE classiques, j'ai traité l'ARN 1-311 de la souche Mal du VIH-1 par des concentrations croissantes de nos réactifs (0, 5 et 15 mM). À la suite de la réaction d'acylation, l'ARN a été précipité par ajout d'éthanol, d'acétate d'éthyle et de glycogène. L'analyse par UV du précipité indique la présence d'un excès de réactif qui est d'autant plus importante que la concentration en réactif d'acylation était grande. L'ARN a donc été purifié par filtration/centrifugation à travers un centricon. La comparaison des spectres UV de la solution d'ARN et des différents lavages indique que le réactif en excès a été éliminé principalement au cours du premier lavage (Fig.95)

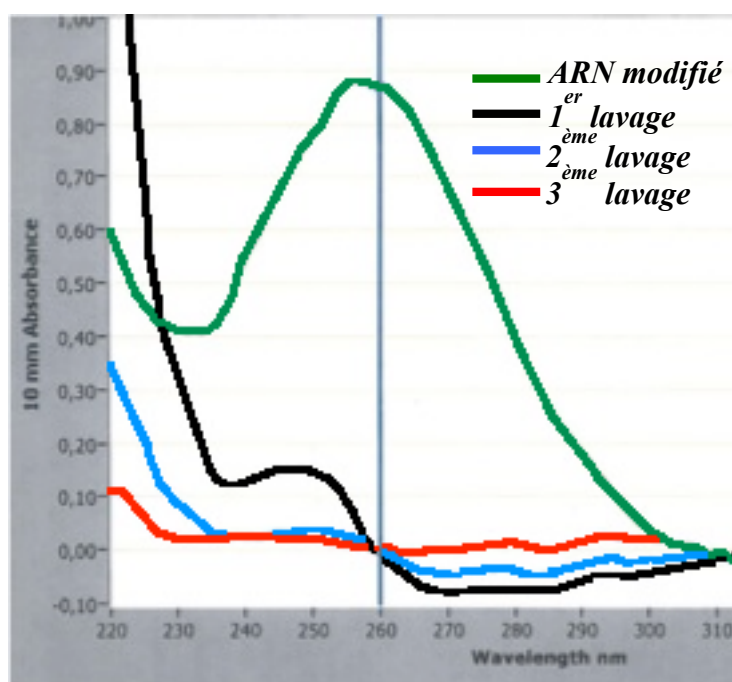


Figure 95: Les spectres de l'absorbance UV de la solution d'ARN modifié par l'anhydride isatoïque-alcyne (courbe verte) et les différents lavages (1^{er} lavage courbe noire, 2^{ème} lavage courbe bleu, et 3^{ème} lavage courbe rouge).

Dans un deuxième temps, nous avons réalisé une rétro-transcription par la RT du virus AMV d'une amorce marquée au p^{32} afin de visualiser les modifications tout au long de l'ARN. En effet, la présence des modifications sur l'ARN doit bloquer la polymérase et créer des ADNc abortifs. Lorsque ces ADNc sont déposés sur un gel d'acrylamide et soumis à une électrophorèse on observe des pauses dont l'intensité augmente en même temps que la fréquence des modifications (Fig.96). Les résultats que nous avons obtenus montrent que nos composés modifiés par la fonction alcyne agissent bien comme les réactifs SHAPE et qu'ils sont capables de modifier l'ARN et de bloquer la rétro-transcription. La fig.89 est un exemple des gels obtenus. La présence de l'ADNc de pleine longueur indique que la plus forte concentration utilisée (15 mM) ne modifie pas tous les ARN et donc que le taux d'une modification statistique par ARN est respecté. Des travaux similaires réalisés par Elodie Mailler lors de son stage de Master 2 avec un ARN plus long (ARN 1-615 de la souche NL4-3 du VIH-1) analysé par électrophorèse capillaire après extension d'amorces fluorescents aboutissent à la même conclusion. Ces études montrent de plus, que les modifications de l'ARN sont localisées dans des régions décrites dans la littérature comme étant non-appariées.

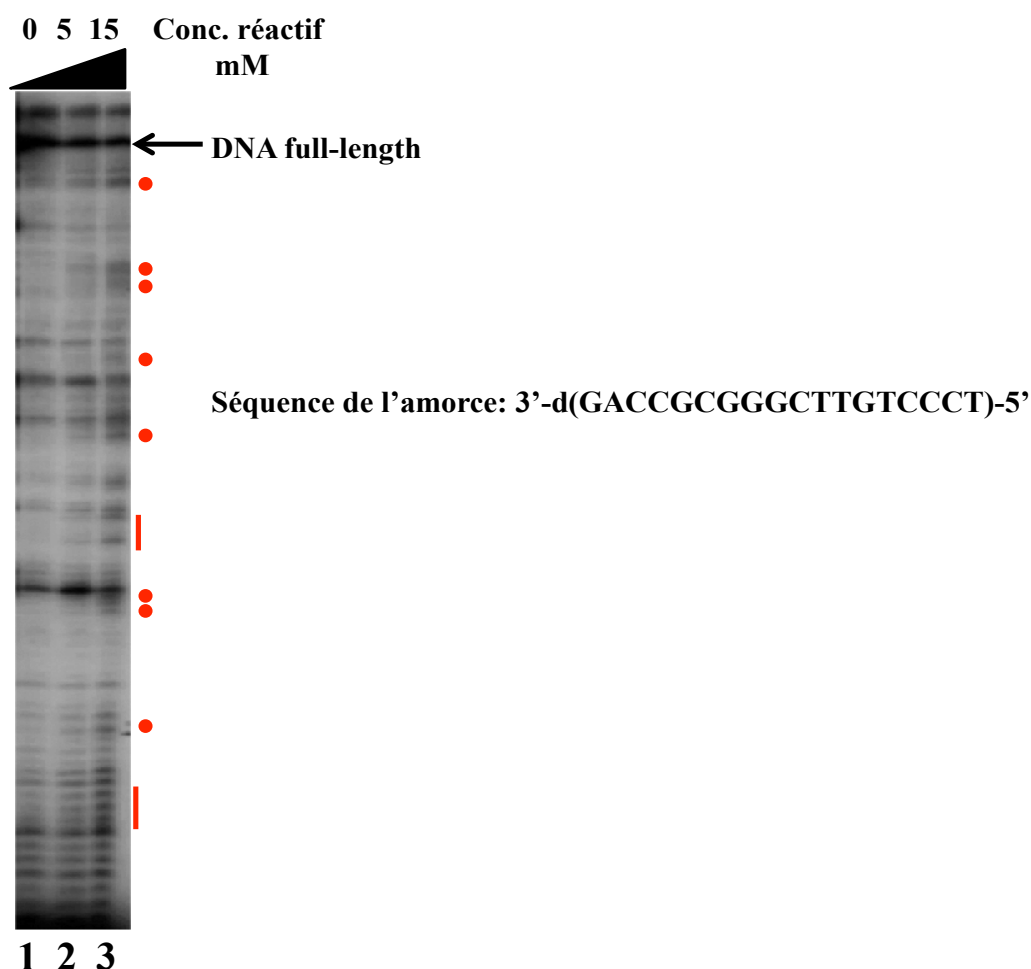


Figure 96: gel d'acrylamide représentant les différentes pauses effectuées par la rétro-transcriptase, suite à la modification de l'ARN avec le réactif chimique (composé 23). Les points rouges représentent les pauses de la transcription réverse. Les trois colonnes représentent la concentration croissante du réactif (0, 5 et 15mM).

Réaction de click chemistry de l'ARN modifié avec une biotine-azoture PEG₃ :

La réaction de click chemistry a été réalisée en présence de TBTA. L'utilisation du TBTA comme ligand est nécessaire pour protéger le cuivre Cu(I) de l'oxydation et également pour protéger l'ARN d'une dégradation. Puisque l'ARN tout comme l'ADN chargé négativement, est sensible à l'oxydation et plus particulièrement sensible à l'attaque des radicaux libres OH•. Ces dommages oxydatifs peuvent donner des bases oxydées par attaque radicalaire, ou causer des cassures de simple brin, lorsque le ribose est attaqué directement, ou même casser des doubles brins. La présence d'ions tels que le Cu, amplifie ces lésions suite à une réaction de type Haber-Weiss Fenton.

Il n'est pas possible de vérifier que le couplage alcyne-azoture a bien eu lieu car la spectrométrie de masse ne permet pas de détecter l'ajout d'environ 200 daltons à une

molécule dont la masse moléculaire est supérieure à 100000 Da, de même il n'est pas possible de visualiser la différence d'élution entre les ARN modifiés ou non modifiés par électrophorèse sur un gel d'acrylamide. Nous avons donc décidé de former un complexe entre les ARN modifiés et biotinylés avec de la streptavidine libre et de visualiser la formation du complexe par retard de migration sur un gel d'agarose. Les résultats obtenus indiquent que l'addition de streptavidine libre à une solution d'ARN modifié par le réactif d'acylation (Composé 23, tableau 3) et soumis au protocole de Click chemistry, conduit bien à la formation d'un complexe de plus haut poids moléculaire alors que le même ARN qui n'a pas subi l'acylation mais qui a été soumis à la Click chemistry ne permet pas la formation du complexe avec la streptavidine (Fig.97).

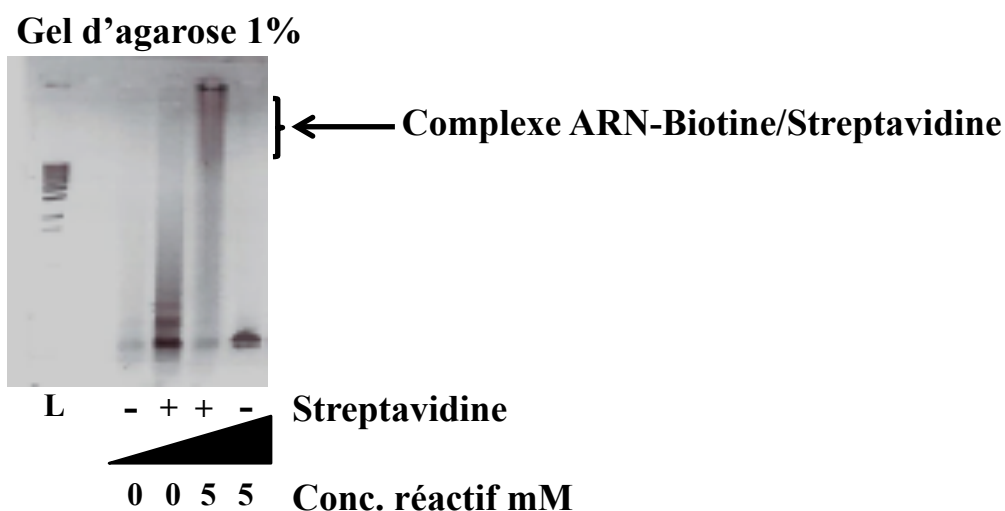


Figure 97: gel d'agarose 1% permettant de visualiser la formation du complexe ARN-Biotine/Streptavidine. Les pistes (+/-) streptavidine correspondent à la présence/absence de la streptavidine, et (+/-) réactif correspondent à la présence/absence de 5mM de réactif lors de la réaction d'acylation.

Ces résultats sont très intéressants puisqu'ils laissent penser qu'il sera possible de sélectionner les ARN modifiés/biotinylés par formation d'un complexe avec des billes magnétiques recouvertes de streptavidine.

III.7 DISCUSSION

III.7.1 Choix des anhydrides isatoïques

Lors du phénomène de bourgeonnement des particules virales du VIH-1 dans les cellules infectées, deux copies d'ARN non épissé sont sélectionnées, sous forme d'un dimère non-covalent, parmi un large excès d'ARN cellulaires et d'ARN viraux épissés. Tout de suite après le bourgeonnement des virions immatures, les protéines virales Gag et Gag-Pol subissent des clivages nécessaires à l'acquisition du pouvoir infectieux. En parallèle, l'ARN génomique subit un réarrangement structural encore mal compris à l'heure actuelle au cours duquel il passe d'une conformation de dimère instable à une conformation de dimère stable et plus compact. Afin d'étudier ces changements structuraux *in virio* et au niveau du génome entier, l'équipe a besoin de développer une nouvelle technique qui allie cartographie chimique et sélection spécifique des ARN modifiés. Dans cet objectif nous avons conçu des réactifs dérivés du NMIA et du 1M7 qui sont des sondes chimiques prometteuses et qui sont capables d'acyler sélectivement les hydroxyles en position 2' des nucléotides non appariés. Le choix s'est porté sur des anhydrides isatoïques puisqu'ils réagissent avec le groupement 2'-OH de l'ARN et subissent une dégradation concomitante (inactivation) par hydrolyse. Afin de pouvoir sélectionner et analyser spécifiquement les ARN modifiés, nos molécules possèdent également une fonction alcyne qui permettra de coupler de façon covalente l'ARN avec une biotine par l'intermédiaire d'une cyclisation de Huisgen. La préparation de ces molécules qui sont pour la plupart inédites a été réalisée en deux étapes à partir des dérivés de l'acide 2-amino-benzoïque, formation de l'anhydride isatoïque puis arrachement du proton de l'amine par de l'hydruure de sodium et finalement alkylation par du bromure de propargyle. Ces réactifs synthétisés ont été caractérisés par RMN ^1H , ^{13}C , SM (en mode positif). Le suivi de la cinétique d'hydrolyse indique que la variation de la nature des substituants en positions 6 et/ou 7 de nos composés a permis de préparer des réactifs avec des temps de demi-vie allant de 2 minutes pour le substituant 7-nitro à 51 minutes pour la double substitution 6 et 7 méthoxy. Nous avons donc à notre disposition une palette de composés avec des temps de demi-vie bien distincts qui permettront d'explorer différents processus cellulaires *in vivo*.

Des études préliminaires ont montré que l'utilisation d'une solution (5 mM) en réactif permet de modifier l'ARN 1-311 du VIH-1 et que cette modification bloque la rétro-transcription par la RT du virus AMV. De plus, nous avons réalisé une réaction de click chemistry catalysée

par le Cu(I) entre la fonction alcyne introduite sur l'ARN, via notre réactif, et la biotine-PEG₃-azoture. Cette réaction n'était pas facile à mettre en œuvre car le cuivre dégrade les acides nucléiques. Dans notre cas, la présence de TBTA qui chélate le cuivre(I) et la faible quantité de CuSO₄/ascorbate de sodium (5 et 20 %, respectivement) a protégé l'ARN et aucune dégradation n'a été observée. Enfin, nous avons réussi à former un complexe entre l'ARN 1-311 modifié par un de nos réactifs puis biotinylé et de la streptavidine libre alors que les mêmes ARN qui ne sont pas modifiés ne permettent pas de former le complexe. Il est important de noter que des études réalisées en parallèle sur un ARN beaucoup plus long (Stage de M2 d'Elodie Mailler) ne conduisent pas exactement au même résultat et que des mises au point sont encore nécessaires. En effet, alors que l'ARN 1-615 du VIH-1 est modifié par nos réactifs et le NMIA de façon identique, il n'a pour l'instant pas été possible après le protocole de click chemistry de former un complexe entre ces ARN et de la streptavidine. Les premiers contrôles laisseraient penser que le problème soit lié au couplage de la biotine-PEG₃-azoture sur l'alcyne.

CONCLUSIONS ET PERSPECTIVES

Durant les trois années de la thèse, j'ai abordé deux sujets bien distincts mais qui visaient tous les deux à améliorer l'étude structurale des ARN. Dans une première partie, j'ai essayé de développer un jeu de sondes fluorescentes qui permettraient suite à une excitation à une longueur d'onde commune d'émettre de la fluorescence à des longueurs d'ondes spécifiques à chaque sonde. Ces molécules ont été conçues sur le modèle des BigDyes qui sont utilisés commercialement pour marquer les ddNTP lors des expériences de séquençage. Au cours de ce travail, j'ai mis au point les conditions de synthèse et de purification de dérivés de la carboxy-fluorescéine porteurs d'une fonction alcyne et de dérivés de la carboxy-rhodamine modifiés par une fonction azoture. J'ai ainsi pu préparer les deux isomères de position (5/6) de la fluorescéine et trois des quatre dérivés cibles de la rhodamine. J'ai également mis au point les conditions de la réaction de click chemistry pour coupler les deux sondes fluorescentes modifiées. Un BigDye « tête-queue » entre la fluorescéine et la ROX a été synthétisé et caractérisé par spectrométrie de masse et fluorescence. Nous avons pu montrer qu'en plus du FRET, il existe un mécanisme de couplage excitonique entre les deux entités fluorescentes qui forment ce BigDye. Des essais pour casser le couplage excitonique ont été effectués mais sans aucun résultat positif.

Cette molécule a montrée une efficacité de transfert dans les deux solvants, mais plus évidente dans le MeOH, tout en sachant bien que les expériences de séquençage d'ADN ne peuvent pas être réalisées dans le MeOH. Donc l'optimisation recherchée pour les sondes fluorescentes commerciales, n'a pas abouti au meilleur résultat désiré. En plus, ce BigDye obtenu a été caractérisé deux fois. Une analyse des spectres d'émission de fluorescence de ce BigDye dans deux solvants différents, MeOH et TEAA a montrée une différence entre les deux spectres, ce qui nous a permis de conclure que le solvant influence fortement les propriétés physico-chimiques de ce BigDye. Une étude de dichroïsme circulaire a montrée la conformation possible pour ce dimère de fluorophores dans un solvant aqueux.

La distance qui sépare les deux entités fluorescentes joue un rôle important dans le mécanisme de FRET. Alors que la distance optimale trouvée entre les deux sondes est de 13 Å, une distance largement inférieure à celle qui permet au FRET de prendre place « 20 Å-100 Å ». Le changement de la nature du bras espaceur qui sépare les deux entités fluorescentes, nécessitait une série de nouvelles réactions chimiques et de purification, alors que la synthèse

des produits de départ était déjà compliquée, et nécessitait du temps pour réaliser et optimiser ces réactions.

Des essais ont été réalisés pour coupler les autres dérivés de la rhodamine modifiés avec une fonction azoture, avec la carboxy-fluorescéine-alcyne, mais les conditions qui avaient permis de former le premier BigDye n'ont pas permis de réaliser la synthèse des autres BigDyes. Le changement du dérivé de la rhodamine, en changeant le substituant sur l'amine, mis en réaction pourrait être la raison pour laquelle les conditions de la réaction changent. Les composés n'étant pas obtenus, il ne nous a pas été possible de vérifier si le mécanisme de FRET existe avec les autres dérivés de la rhodamine ni de le comparer avec le premier BigDye obtenu. Nous n'avons pas pu vérifier si, autre que la distance qui sépare les deux fluorophores et autre que l'effet du solvant qui joue un rôle crucial dans l'émission de fluorescence de ce BigDye, le changement du dérivé de la rhodamine couplé à la fluorescéine va entraîner un changement au niveau de la conformation de ce BigDye, et par la même occasion un changement au niveau de son émission de fluorescence.

Les rendements faibles obtenus lors de la préparation des entités fluorescentes de départ, aussi bien que les conditions difficiles de séparation des deux isomères, n'a pas permis de réaliser la synthèse des BigDyes dans les meilleures conditions expérimentales.

De même pour notre BigDye synthétisé, les quantités obtenues ne nous ont pas permis d'envisager un couplage avec une amorce nucléotidique.

Les conclusions principales de cette première partie du projet sont les suivantes : (i) le solvant influence l'émission de fluorescence dans un système à transfert d'énergie, (ii) la distance entre le donneur et l'accepteur de fluorescence, joue un rôle important dans l'interaction entre les deux dipôles, en favorisant soit une interaction faible alors un mécanisme de FRET, soit une interaction forte alors un mécanisme de couplage excitonique.

Les perspectives de la première partie de ce travail de thèse, consistent alors à optimiser la synthèse de click chemistry, afin d'obtenir les autres composés BigDyes, pour cela il faudrait modifier l'étape de purification des dérivés de la rhodamine car la présence de traces d'acide provenant de l'acide acétique utilisé dans l'étape de purification, pourrait être la raison pour laquelle l'expérience de click chemistry ne donne pas le résultat attendu. Dans le cas où les nouveaux BigDyes obtenus présentent de meilleures propriétés, ces molécules seront alors couplées à des amorces ADN. Dans le cas où les propriétés seraient similaires, il faudrait

essayer alors de modifier la nature du bras espaceur qui sépare les deux entités fluorescentes, afin d'atteindre la distance optimale pour le mécanisme de FRET (entre 20 et 100 Å). On pourrait également envisager un bras espaceur plus rigide et limiter les degrés de liberté en introduisant des doubles liaisons afin de limiter la superposition des cycles xanthènes et diminuer l'interaction entre les deux dipôles de transition des deux fluorophores.

Dans la deuxième partie de mon projet de thèse, j'ai préparé 4 nouveaux réactifs chimiques conçus afin de cartographier l'ARN *in vivo* et également de sélectionner spécifiquement les ARN modifiés afin de pouvoir coupler la cartographie chimique avec une technique de séquençage à haut débit pour l'analyse des ADNc. Ces réactifs chimiques dérivés de l'anhydride isatoïque sont modifiés au niveau de l'amine secondaire, avec une fonction alcyne qui permet le couplage ultérieur avec une biotine modifiée par une fonction azoture par la cycloaddition de Huisgen. L'hydrolyse de ces composés a été suivie par mesure de l'absorbance UV en fonction du temps à une longueur d'onde où le composé non-hydrolysé absorbe peu et où le produit hydrolysé absorbe. Pour cette étude, j'ai également préparé les dérivés de l'anhydride isatoïque non substitués sur l'amine ou substitués par un groupement méthyle (comme les composés utilisés actuellement dans la technique de SHAPE). Nos résultats indiquent que la variation des substituants sur le cycle aromatique (effet mésomère donneur/accepteur) et leur position (méta/para) permet de moduler les temps de demi-vie de ces réactifs chimiques et que la substitution portée par l'amine N1 influence également le $t_{1/2}$. Les composés synthétisés au cours de ce travail pourront donc être utilisés pour cartographier l'ARN cellulaire ou viral *in vivo*.

Un premier test avec l'un de ces réactifs (composé 23 : tableau 3) a permis de mettre au point les conditions de modification de l'ARN, ainsi que la réaction de click avec une biotine-azoture. Ce premier essai a permis de mettre en évidence, la capacité de ces réactifs chimiques à modifier l'ARN et à fixer une biotine-azoture par réaction de click chemistry sans dégrader la molécule d'ARN. Ces résultats doivent toutefois être approfondis puisque la modification d'un ARN plus long n'a pas donné les mêmes résultats. Il semblerait que dans ce cas ce soit le couplage par click chemistry qui soit déficient, il faudra donc optimiser ces conditions. On pourra par exemple, utiliser directement du cuivre(I) pour la catalyse de la réaction et non pas le former *in situ* par réduction du cuivre(II) par de l'ascorbate de sodium.

Lorsque les conditions de réaction seront mises au point il devrait être possible de modifier et étudier la structure d'un ARN génomique dans les virions. Pour cela, l'ARN modifié avec les

réactifs chimiques sera sélectionné avec des billes recouvertes de streptavidine grâce à la présence de la biotine. La rétro-transcription des ARN modifiés permettra d'obtenir des ADNc représentatifs des nucléotides modifiés au niveau de l'ARN. Ces ADNc seront ensuite séquencés en utilisant des séquences adaptatrices Illumina, sur une plateforme HiSeq2000.

Références bibliographiques

1. Chertova, E. *et al.* Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* **80**, 9039–52 (2006).
2. Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H. & Roques, B. P. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**, 523–37 (1995).
3. Roda, R. H. *et al.* Role of the Reverse Transcriptase, Nucleocapsid Protein, and Template Structure in the Two-step Transfer Mechanism in Retroviral Recombination. *J. Biol. Chem.* **278**, 31536–46 (2003).
4. Turner, B. G. & Summers, M. F. Structural biology of HIV. *J. Mol. Biol.* **285**, 1–32 (1999).
5. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell* **89**, 263–273 (1997).
6. Doranz, B. J. *et al.* A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* **85**, 1149–58 (1996).
7. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. a. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–7 (1996).
8. Robin, C. R. P. and W. A. Spoilt for choice of co-receptors. *Nature* **388**, 230–231 (1997).
9. Bour, S., Geleziunas, R. & Wainberg, M. A. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiol. Rev.* **59**, 63–93 (1995).
10. Stephen, G. P. retroviral reverse transcriptase: synthesis, structure, and fusion. *J. Acquir. Immune Defic. Syndr.* 817–831 (1990).
11. Katz, R. a & Skalka, a M. The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133–73 (1994).
12. Mougel, M., Houzet, L. & Darlix, J.-L. When is it time for reverse transcription to start and go? *Retrovirology* **6**, 24 (2009).
13. Esposito, F., Corona, A. & Tramontano, E. HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. *Mol. Biol. Int.* **2012**, 586401 (2012).

14. Miller, M. D., Farnet, C. M. & Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71**, 5382–5390 (1997).
15. Nie, Z. *et al.* The putative alpha helix 2 of human immunodeficiency virus type 1 Vpr contains a determinant which is responsible for the nuclear translocation of proviral DNA in growth-arrested cells. *J. Virol.* **72**, 4104–15 (1998).
16. Popov, S. *et al.* Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *EMBO J.* **17**, 909–17 (1998).
17. Hansen, M. S., Carteau, S., Hoffmann, C., Li, L. & Bushman, F. Retroviral cDNA integration: mechanism, applications and inhibition. *Genet. Eng. (N. Y.)*. **20**, 41–61 (1998).
18. Pauza, C. D. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* **179**, 886–9 (1990).
19. Hong, T., Drlica, K., Pinter, A. & Murphy, E. Circular DNA of human immunodeficiency virus: analysis of circle junction nucleotide sequences. *J. Virol.* **65**, 551–5 (1991).
20. Yoder, K. E. & Bushman, F. D. Repair of Gaps in Retroviral DNA Integration Intermediates Repair of Gaps in Retroviral DNA Integration Intermediates. *J. Virol.* **74**, 11191–11200 (2000).
21. Ciuffi, A. & Bushman, F. D. Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet.* **22**, 388–95 (2006).
22. Cherepanov, P., Maertens, G. N. & Hare, S. Structural insights into the retroviral DNA integration apparatus. *Curr. Opin. Struct. Biol.* **21**, 249–56 (2011).
23. Li, L. *et al.* Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J.* **20**, 3272–81 (2001).
24. Wu, Y. & Marsh, J. W. Selective transcription and modulation of resting T cell activity by preintegrated HIV DNA. *Science* **293**, 1503–6 (2001).
25. Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-ferrat, N. & Peterlin, B. M. NF-KappaB Binds P-TEFb to Stimulate Transcriptional Elongation by RNA Polymerase II. *Mol. Cell* **8**, 327–337 (2001).
26. Strebel, K. Virus – host interactions: role of HIV proteins Vif, Tat, and Rev. *AIDS* **17**, S25–S34 (2003).
27. Schwartz, S., Felber, B. K. & Pavlakis, G. N. Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol. Cell. Biol.* **12**, 207–19 (1992).

28. Wilkinson, T. A., Botuyan, M. V, Kaplan, B. E., Rossi, J. J. & Chen, Y. Arginine side-chain dynamics in the HIV-1 rev-RRE complex. *J. Mol. Biol.* **303**, 515–29 (2000).
29. Askjaer, P., Jensen, T. H., Nilsson, J., Englmeier, L. & Kjems, J. The specificity of the CRM1-Rev nuclear export signal interaction is mediated by RanGTP. *J. Biol. Chem.* **273**, 33414–22 (1998).
30. Fridell, R. A., Bogerd, H. P. & Cullen, B. R. Nuclear export of late HIV-1 mRNAs occurs via a cellular protein export pathway. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 4421–4 (1996).
31. Groom, H. C. T., Anderson, E. C., Dangerfield, J. a & Lever, A. M. L. Rev regulates translation of human immunodeficiency virus type 1 RNAs. *J. Gen. Virol.* **90**, 1141–7 (2009).
32. Kiss, A., Li, L., Gettemeier, T. & Venkatesh, L. . Functional analysis of the interaction of the human immunodeficiency virus type 1 Rev nuclear export signal with its cofactors. *Virology* **314**, 591–600 (2003).
33. Leblanc, J., Weil, J. & Beemon, K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdiscip. Rev. RNA* **4**, 567–80 (2013).
34. Schwartz, S., Felber, B. K., Fenyő, E. M. & Pavlakis, G. N. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J. Virol.* **64**, 5448–56 (1990).
35. Willey, R. L., Maldarelli, F., Martin, M. A. & Strebel, K. Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. *J. Virol.* **66**, 226–34 (1992).
36. Jouvenet, N., Simon, S. M. & Bieniasz, P. D. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19114–9 (2009).
37. Kutluay, S. B. & Bieniasz, P. D. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog.* **6**, e1001200 (2010).
38. Kaye, J. F. & Lever, A. M. Human immunodeficiency virus types 1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *J. Virol.* **73**, 3023–31 (1999).
39. Poon, D. T. K., Chertova, E. N. & Ott, D. E. Human immunodeficiency virus type 1 preferentially encapsidates genomic RNAs that encode Pr55(Gag): functional linkage between translation and RNA packaging. *Virology* **293**, 368–78 (2002).
40. Liang, C., Hu, J., Russell, R. S. & Wainberg, M. A. Translation of Pr55(gag) augments packaging of human immunodeficiency virus type 1 RNA in a cis-acting manner. *AIDS Res. Hum. Retroviruses* **18**, 1117–26 (2002).

41. Datta SA, Heinrich F, Raghunandan S, Krueger S, Curtis JE, Rein A, N. H. HIV-1 Gag extension: conformational changes require simultaneous interaction with membrane and nucleic acid. *J. Mol. Biol.* **406**, 205–214 (2011).
42. Ajamian, L. *et al.* Unexpected roles for UPF1 in HIV-1 RNA metabolism and translation. *RNA* 914–927 (2008). doi:10.1261/rna.829208.recruited
43. Zimmerman, C. *et al.* Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature* **415**, 88–92 (2002).
44. Reed, J. C. *et al.* HIV-1 Gag co-opts a cellular complex containing DDX6, a helicase that facilitates capsid assembly. *J. Cell Biol.* **198**, 439–56 (2012).
45. Tang, Y. *et al.* Cellular motor protein KIF-4 associates with retroviral Gag. *J. Virol.* **73**, 10508–13 (1999).
46. Meng, B. & Lever, A. M. Wrapping up the bad news - HIV assembly and release. *Retrovirology* **10**, 5 (2013).
47. Hogue, I. B., Hoppe, A. & Ono, A. Quantitative fluorescence resonance energy transfer microscopy analysis of the human immunodeficiency virus type 1 Gag-Gag interaction: relative contributions of the CA and NC domains and membrane binding. *J. Virol.* **83**, 7322–36 (2009).
48. Rein, A., Datta, S. a K., Jones, C. P. & Musier-Forsyth, K. Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* **36**, 373–80 (2011).
49. Isel, C., Ehresmann, C. & Marquet, R. Initiation of HIV Reverse Transcription. *Viruses* **2**, 213–43 (2010).
50. Ott, D. E. Cellular proteins detected in HIV-1. *Rev. Med. Virol.* **18**, 159–75
51. Balasubramaniam, M. & Freed, E. O. New insights into HIV assembly and trafficking. *Physiology (Bethesda)*. **26**, 236–51 (2011).
52. Carlson, L.-A. *et al.* Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **4**, 592–9 (2008).
53. Sundquist, W. I. & Kräusslich, H.-G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).
54. Norgan, A. P. *et al.* ESCRT-independent budding of HIV-1 gag virus-like particles from *Saccharomyces cerevisiae* spheroplasts. *PLoS One* **7**, e52603 (2012).
55. Chung, H.-Y. *et al.* NEDD4L overexpression rescues the release and infectivity of human immunodeficiency virus type 1 constructs lacking PTAP and YPYL late domains. *J. Virol.* **82**, 4884–97 (2008).

56. Sette, P., Jadwin, J. a, Dussupt, V., Bello, N. F. & Bouamr, F. The ESCRT-associated protein Alix recruits the ubiquitin ligase Nedd4-1 to facilitate HIV-1 release through the LYPXnL L domain motif. *J. Virol.* **84**, 8181–92 (2010).
57. Wilk, T. *et al.* Organization of Immature Human Immunodeficiency Virus Type 1. *J. Virol.* **75**, 759–771 (2001).
58. Tatyana DORFMAN, Anatoly BUKOVSKY, Asa ÖHAGEN, S. H. and H. G. G. Functional Domains of the Capsid protein of Human Immunodeficiency Virus Type 1. *J. Virol.* **68**, 8180–8187 (1994).
59. Kräusslich, H. G., Fäcke, M., Heuser, a M., Konvalinka, J. & Zentgraf, H. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J. Virol.* **69**, 3407–19 (1995).
60. Liang, C. *et al.* Characterization of a Putative alpha-Helix across the Capsid-SP1 Boundary That Is Critical for the Multimerization of Human Immunodeficiency Virus Type 1 Gag. *J. Virol.* **76**, 11729–11737 (2002).
61. Schwedler, U. K. Von, Stray, K. M., Garrus, J. E. & Sundquist, W. I. Functional Surfaces of the Human Immunodeficiency Virus Type 1 Capsid Protein. *J. Virol.* **77**, 5439–5450 (2003).
62. Reil, H., Bukovsky, a a, Gelderblom, H. R. & Göttlinger, H. G. Efficient HIV-1 replication can occur in the absence of the viral matrix protein. *EMBO J.* **17**, 2699–708 (1998).
63. Burniston, M. T., Cimorelli, a, Colgan, J., Curtis, S. P. & Luban, J. Human immunodeficiency virus type 1 Gag polyprotein multimerization requires the nucleocapsid domain and RNA and is promoted by the capsid-dimer interface and the basic region of matrix protein. *J. Virol.* **73**, 8527–40 (1999).
64. Cimorelli, A., Sandin, S., Höglund, S. & Luban, J. Basic Residues in Human Immunodeficiency Virus Type 1 Nucleocapsid Promote Virion Assembly via Interaction with RNA. *J. Virol.* **74**, 3046–3057 (2000).
65. Accola, M. a, Strack, B. & Göttlinger, H. G. Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J. Virol.* **74**, 5395–402 (2000).
66. Gottlinger, H. . The HIV-1 assembly machine. *AIDS* **15**, S13–S20 (2001).
67. Fu, W., Gorelick, R. J. & Rein, a. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol.* **68**, 5013–8 (1994).
68. M.SHEHU-XHILAGA *et al.* Proteolytic Processing of the P2 / Nucleocapsid Cleavage Site Is Critical for Human Immunodeficiency Virus Type 1 RNA Dimer Maturation. *J. Virol.* **75**, 9156–9164 (2001).

69. M.SHEHU-XHILAGA *et al.* The Conformation of the Mature Dimeric Human Immunodeficiency Virus Type 1 RNA Genome Requires Packaging of Pol Protein. *J. Virol.* **76**, 4331–4340 (2002).
70. Song, R., Kafaie, J., Yang, L. & Laughrea, M. HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. *J. Mol. Biol.* **371**, 1084–98 (2007).
71. Bell, N. M. & Lever, A. M. L. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–44 (2013).
72. Wiegers, K. *et al.* Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J. Virol.* **72**, 2846–54 (1998).
73. Feng, Y. X. *et al.* HIV-1 nucleocapsid protein induces “maturation” of dimeric retroviral RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 7577–81 (1996).
74. Jalalirad, M. & Laughrea, M. Formation of immature and mature genomic RNA dimers in wild-type and protease-inactive HIV-1: differential roles of the Gag polyprotein, nucleocapsid proteins NCp15, NCp9, NCp7, and the dimerization initiation site. *Virology* **407**, 225–36 (2010).
75. Ohishi, M. *et al.* The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acids Res.* **39**, 3404–17 (2011).
76. Müller, B. *et al.* HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *J. Biol. Chem.* **284**, 29692–703 (2009).
77. Bender, W. *et al.* High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J. Virol.* **25**, 888–96 (1978).
78. DUESBERG, P. H. Physical properties of Rous sarcoma virus RNA. *PNAS* 1511–1518 (1968).
79. Bader, J. P. & Steck, T. L. Analysis of the Ribonucleic Acid of Murine Leukemia Virus. *J. Virol.* **4**, 454–459 (1969).
80. Canaani, E. L. I. & Duesberg, P. Role of subunits of 60 to 70S avian tumor virus ribonucleic acid in its template activity for the viral deoxyribonucleic acid polymerase. *J. Virol.* **10**, 23–31 (1972).
81. Clever, J. L., Miranda, D. & Parslow, T. G. RNA Structure and Packaging Signals in the 5J Leader Region of the Human Immunodeficiency Virus Type 1 Genome. *J. Virol.* **76**, 12381–12387 (2002).
82. Clever, J. L. & Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–14 (1997).

83. Damgaard, C. K., Dyhr-Mikkelsen, H. & Kjems, J. Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res.* **26**, 3667–76 (1998).
84. Harrison, G. P. & Lever, a M. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* **66**, 4144–53 (1992).
85. Hayashi, T., Ueno, Y. & Okamoto, T. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS Lett.* **327**, 213–8 (1993).
86. Paillart, J.-C. *et al.* First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* **279**, 48397–403 (2004).
87. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2010).
88. Wilkinson, K. a *et al.* High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **6**, e96 (2008).
89. Paillart, J., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci.* **93**, 5572–5577 (1996).
90. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B. & Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4945–9 (1994).
91. Lu, K., Heng, X. & Summers, M. F. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609–33 (2011).
92. Abbink, T. E. M. & Berkhout, B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem.* **278**, 11601–11 (2003).
93. Damgaard, C. K., Andersen, E. S., Knudsen, B., Gorodkin, J. & Kjems, J. RNA Interactions in the 5' Region of the HIV-1 Genome. *J. Mol. Biol.* **336**, 369–379 (2004).
94. Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2**, 461–72 (2004).
95. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919–29 (2001).
96. Kim, S. H. *et al.* The general structure of transfer RNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 4970–4 (1974).
97. Ambros, V. microRNAs: tiny regulators with great potential. *Cell* **107**, 823–6 (2001).

98. Sharp, P. A. RNA interference--2001. *Genes Dev.* **15**, 485–90 (2001).
99. Grosshans, H. & Slack, F. J. Micro-RNAs: small is plentiful. *J. Cell Biol.* **156**, 17–21 (2002).
100. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30**, 3497–531 (2002).
101. Westhof, E. & Fritsch, V. RNA folding: beyond Watson-Crick pairs. *Structure* **8**, R55–65 (2000).
102. Carl C. Correll, Alexander Munishkin, Yuen-Ling Chan, Zhong Ren, I. G. W. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Biochemistry* **95**, 13436–13441 (1998).
103. Correll, C. C., Wool, I. G. & Munishkin, A. The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J. Mol. Biol.* **292**, 275–87 (1999).
104. Ananth, P., Goldsmith, G. & Yathindra, N. An innate twist between Crick's wobble and Watson-Crick base pairs. *RNA* **19**, 1038–53 (2013).
105. Masquida, B. & Westhof, E. On the wobble GoU and related pairs. *RNA Soc.* **6**, 9–15 (2000).
106. Westhof, E., Masquida, B. & Jaeger, L. RNA tectonics: towards RNA design. *Fold. Des.* **1**, R78–88 (1996).
107. Tian, B., Bevilacqua, P. C., Diegelman-Parente, A. & Mathews, M. B. The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.* **5**, 1013–23 (2004).
108. Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285 (1995).
109. Moore, P. B. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**, 287–300 (1999).
110. Leontis, N. B. & Westhof, E. Analysis of RNA motifs. *Curr. Opin. Struct. Biol.* **13**, 300–308 (2003).
111. Rangan, P., Masquida, B., Westhof, E. & Woodson, S. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1574–9 (2003).
112. Misra, V. K. & Draper, D. E. The Linkage between Magnesium Binding and RNA Folding. *J. Mol. Biol.* **317**, 507–521 (2002).
113. Xingwang Fang, Tao Pan, and T. R. S. A Thermodynamic Framework and Cooperativity in the Tertiary Folding of a Mg²⁺ Dependent Ribozyme. *Biochemistry* **38**, 16840–16846 (1999).

114. Heilman-Miller, S. L., Pan, J., Thirumalai, D. & Woodson, S. Role of counterion condensation in folding of the Tetrahymena ribozyme. II. Counterion-dependence of folding kinetics. *J. Mol. Biol.* **309**, 57–68 (2001).
115. Sosnick, T. R. & Pan, T. RNA folding: models and perspectives. *Curr. Opin. Struct. Biol.* **13**, 309–316 (2003).
116. Lescoute, A. & Westhof, E. Topology of three-way junctions in folded RNAs. *RNA* **12**, 83–93 (2006).
117. Nussinov, R. & Jacobson, A. B. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 6309–13 (1980).
118. Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133–48 (1981).
119. Wrede, P., Wurst, R., Vournakis, J. & Rich, A. Conformational changes of yeast tRNA^{Phe} and E. coli tRNA^{2Glu} as indicated by different nuclease digestion patterns. *J. Biol. Chem.* **254**, 9608–16 (1979).
120. Florentz, C. *et al.* The tRNA-like structure of turnip yellow mosaic virus RNA: structural organization of the last 159 nucleotides from the 3' OH terminus. *EMBO J.* **1**, 269–76 (1982).
121. Rietveld, K., Van Poelgeest, R., Pleij, C. W., Van Boom, J. H. & Bosch, L. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res.* **10**, 1929–46 (1982).
122. Kean, J. M. & Draper, D. E. Secondary structure of a 345-base RNA fragment covering the S8/S15 protein binding domain of Escherichia coli 16S ribosomal RNA. *Biochemistry* **24**, 5052–61 (1985).
123. Warrington, R. C. Ribonuclease T1 catalyzed degradation of transfer RNA: an unusual alteration induced by urea. *Biochim. Biophys. Acta* **353**, 63–8 (1974).
124. Donis-Keller, H., Maxam, A. M. & Gilbert, W. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.* **4**, 2527–38 (1977).
125. Mougél, M. *et al.* Binding of Escherichia coli ribosomal protein S8 to 16 S rRNA. A model for the interaction and the tertiary structure of the RNA binding site. *J. Mol. Biol.* **198**, 91–107 (1987).
126. Baudin, F. *et al.* Higher-order structure of domain III in Escherichia coli 16S ribosomal RNA, 30S subunit and 70S ribosome. *Biochimie* **69**, 1081–96 (1987).
127. Uchida, T., Arima, T. & Egami, F. Specificity of RNase U2. *J. Biochem.* **67**, 91–102 (1970).
128. Boguski, M. S., Hieter, P. A. & Levy, C. C. Identification of a cytidine-specific ribonuclease from chicken liver. *J. Biol. Chem.* **255**, 2160–3 (1980).

129. Levy, C. C. & Karpetsky, T. P. The purification and properties of chicken liver RNase: An enzyme which is useful in distinguishing between cytidylic and uridylic acid residues. *J. Biol. Chem.* **255**, 2153–9 (1980).
130. Vary, C. P. & Vournakis, J. N. RNA structure analysis using T2 ribonuclease: detection of pH and metal ion induced conformational changes in yeast tRNA^{Phe}. *Nucleic Acids Res.* **12**, 6763–78 (1984).
131. Raines, R. T. Ribonuclease A. *chem.reviews* **98**, 1045–1065 (1998).
132. Pieler, T. & Erdmann, V. A. Three-dimensional structural model of eubacterial 5S RNA that has functional implications. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4599–603 (1982).
133. Garret, M., Romby, P., Giegé, R. & Litvak, S. Interactions between avian myeloblastosis reverse transcriptase and tRNA^{Trp}. Mapping of complexed tRNA with chemicals and nucleases. *Nucleic Acids Res.* **12**, 2259–71 (1984).
134. Vasilenko, S. K. & Rytte, V. C. [Isolation of highly purified ribonuclease from cobra (*Naja oxiana*) venom]. *Biokhim(Moscow, Russ.* **40**, 578–83
135. Lowman, H. B. & Draper, D. E. On the recognition of helical RNA by cobra venom V1 nuclease. *J. Biol. Chem.* **261**, 5396–403 (1986).
136. Bolyreva, L. G., Butorin, A. S. & Vasilenko, S. K. [Molecular weight distribution of hydrolysis products of polyuridylic acid by cobra venom ribonuclease. Preparation of polynucleotides of required molecular weight]. *Biokhim(Moscow, Russ.* **43**, 1659–64 (1978).
137. Butorin, A. S., Remy, P., Ebel, J. P. & Vassilenko, S. K. Comparison of the hydrolysis patterns of several tRNAs by cobra venom ribonuclease in different steps of the aminoacylation reaction. *Eur. J. Biochem.* **121**, 587–95 (1982).
138. Troutt, A., Savin, T. J., Curtiss, W. C., Celentano, J. & Vournakis, J. N. Secondary structure of *Bombyx mori* and *Dictyostelium discoideum* 5S rRNA from S1 nuclease and cobra venom ribonuclease susceptibility, and computer assisted analysis. *Nucleic Acids Res.* **10**, 653–64 (1982).
139. Stern, S., Wilson, R. C. & Noller, H. F. Localization of the binding site for protein S4 on 16 S ribosomal RNA by chemical and enzymatic probing and primer extension. *J. Mol. Biol.* **192**, 101–10 (1986).
140. Desai, N. A. & Shankar, V. Single-strand-specific nucleases. *FEMS Microbiol. Rev.* **26**, 457–91 (2003).
141. Gite, S. U. & Shankar, V. Single-strand-specific nucleases. *Crit. Rev. Microbiol.* **21**, 101–22 (1995).
142. Aultman, K. S. & Chang, S. H. Partial P1 nuclease digestion as a probe of tRNA structure. *Eur. J. Biochem.* **124**, 471–6 (1982).

143. Sheppard, K., Akochy, P.-M. & Söll, D. Assays for transfer RNA-dependent amino acid biosynthesis. *Methods* **44**, 139–45 (2008).
144. Brunel, C. & Romby, P. Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol.* **318**, 3–21 (2000).
145. Weeks, K. M. Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* **20**, 295–304 (2010).
146. Ehresmann, C. *et al.* Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–28 (1987).
147. Brookes, P. & Lawley, P. D. The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochem. J.* **80**, 496–503 (1961).
148. LAWLEY, P. D. & BROOKES, P. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.* **89**, 127–38 (1963).
149. Wintermeyer, W. & Zachau, H. G. Tertiary structure interactions of 7-methylguanosine in yeast tRNA Phe as studied by borohydride reduction. *FEBS Lett.* **58**, 306–9 (1975).
150. Romby, P., Moras, D., Dumas, P., Ebel, J. P. & Giegé, R. Comparison of the tertiary structure of yeast tRNA(Asp) and tRNA(Phe) in solution. Chemical modification study of the bases. *J. Mol. Biol.* **195**, 193–204 (1987).
151. Lempereur, L. *et al.* Conformation of yeast 18S rRNA. Direct chemical probing of the 5' domain in ribosomal subunits and in deproteinized RNA by reverse transcriptase mapping of dimethyl sulfate-accessible. *Nucleic Acids Res.* **13**, 8339–57 (1985).
152. Shiraishi, H. & Shimura, Y. Functional domains of the RNA component of ribonuclease P revealed by chemical probing of mutant RNAs. *EMBO J.* **7**, 3817–21 (1988).
153. Leonard, N. J., McDonald, J. J., Henderson, R. E. & Reichmann, M. E. Reaction of diethyl pyrocarbonate with nucleic acid components. Adenosine. *Biochemistry* **10**, 3335–42 (1971).
154. Vincze, A., Henderson, R. E., McDonald, J. J. & Leonard, N. J. Reaction of diethyl pyrocarbonate with nucleic acid components. Bases and nucleosides derived from guanine, cytosine, and uracil. *J. Am. Chem. Soc.* **95**, 2677–82 (1973).
155. Ehrenberg, L., Fedorcsak, I. & Solymosy, F. Diethyl pyrocarbonate in nucleic acid research. *Prog. Nucleic Acid Res. Mol. Biol.* **16**, 189–262 (1976).
156. Peattie, D. A. Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1760–4 (1979).
157. Van Stolk, B. J. & Noller, H. F. Chemical probing of conformation in large RNA molecules. Analysis of 16 S ribosomal RNA using diethylpyrocarbonate. *J. Mol. Biol.* **180**, 151–77 (1984).

158. Felden, B. *et al.* Probing the structure of the Escherichia coli 10Sa RNA (tmRNA). *RNA* **3**, 89–103 (1997).
159. Moazed, D., Stern, S. & Noller, H. F. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.* **187**, 399–416 (1986).
160. Mandiyan, V. & Boublik, M. Structural analysis of the 5' domain of the HeLa 18S ribosomal RNA by chemical and enzymatic probing. *Nucleic Acids Res.* **18**, 7055–62 (1990).
161. Noller, H. F. & Chaires, J. B. Functional modification of 16S ribosomal RNA by kethoxal. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 3115–8 (1972).
162. Brunel, C., Romby, P., Westhof, E., Ehresmann, C. & Ehresmann, B. Three-dimensional model of Escherichia coli ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.* **221**, 293–308 (1991).
163. Hayatsu, H. Bisulfite modification of nucleic acids and their constituents. *Prog. Nucleic Acid Res. Mol. Biol.* **16**, 75–124 (1976).
164. Pieler, T., Digweed, M., Bartsch, M. & Erdmann, V. A. Comparative structural analysis of cytoplasmic and chloroplastic 5S rRNA from spinach. *Nucleic Acids Res.* **11**, 591–604 (1983).
165. Pieler, T., Digweed, M. & Erdmann, V. A. RNA structural dynamics: pre-melting and melting transitions in E. coli 5S rRNA. *J. Biomol. Struct. Dyn.* **3**, 495–514 (1985).
166. Singer, B. All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature* **264**, 333–9 (1976).
167. Jensen, D. E. & Reed, D. J. Reaction of DNA with alkylating agents. Quantitation of alkylation by ethylnitrosourea of oxygen and nitrogen sites on poly[dA-dT] including phosphotriester formation. *Biochemistry* **17**, 5098–107 (1978).
168. Baudin, F. *et al.* Probing the phosphates of the Escherichia coli ribosomal 16S RNA in its naked form, in the 30S subunit, and in the 70S ribosome. *Biochemistry* **28**, 5847–55 (1989).
169. Tullius, T. D. & Greenbaum, J. A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127–34 (2005).
170. Sclavi, B., Woodson, S., Sullivan, M., Chance, M. R. & Brenowitz, M. Time-resolved synchrotron X-ray “footprinting”, a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J. Mol. Biol.* **266**, 144–59 (1997).
171. Celander, D. W. Probing RNA structures with hydroxyl radicals. *Curr. Protoc. nucleic acid Chem.* **Chapter 6**., (2001).

172. Brenowitz, M. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical “footprinting.” *Curr. Opin. Struct. Biol.* **12**, 648–653 (2002).
173. Adilakshmi, T., Ramaswamy, P. & Woodson, S. A. Protein-independent folding pathway of the 16S rRNA 5' domain. *J. Mol. Biol.* **351**, 508–19 (2005).
174. Maglott, E. J. & Glick, G. D. Engineering disulfide cross-links in RNA via air oxidation. *Curr. Protoc. Nucleic Acid Chem.* **Chapter 5**, Unit 5.4 (2001).
175. Merino, E. J., Wilkinson, K. a, Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–31 (2005).
176. Mortimer, S. a & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–5 (2007).
177. Mortimer, S. a & Weeks, K. M. Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.* **130**, 16178–80 (2008).
178. Duncan, C. D. S. & Weeks, K. M. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* **47**, 8504–13 (2008).
179. Mortimer, S. a & Weeks, K. M. Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* **4**, 1413–21 (2009).
180. Low, J. T. & Weeks, K. M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–8 (2010).
181. Badorrek, C. S. & Weeks, K. M. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat. Chem. Biol.* **1**, 104–11 (2005).
182. Badorrek, C. S. & Weeks, K. M. Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* **45**, 12664–72 (2006).
183. Chen, Y. *et al.* Structure of stem-loop IV of Tetrahymena telomerase RNA. *EMBO J.* **25**, 3156–66 (2006).
184. Gherghe, C. & Weeks, K. M. The SL1-SL2 (stem-loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. *J. Biol. Chem.* **281**, 37952–61 (2006).
185. Dann, C. E. *et al.* Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130**, 878–92 (2007).

186. Vicens, Q., Gooding, A. R., Laederach, A. & Cech, T. R. Local RNA structural changes induced by crystallization are revealed by SHAPE. *RNA Soc.* 536–548 (2007). doi:10.1261/rna.400207.)
187. Jones, C. N., Wilkinson, K. A., Hung, K. T., Weeks, K. M. & Spremulli, L. L. Lack of secondary structure characterizes the 5' ends of mammalian mitochondrial mRNAs. *RNA* 862–871 (2008). doi:10.1261/rna.909208.leaderless
188. Stoddard, C. D., Gilbert, S. D. & Batey, R. T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA Soc.* **14**, 675–684 (2008).
189. Wang, B., Wilkinson, K. a & Weeks, K. M. Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry. *Biochemistry* **47**, 3454–61 (2008).
190. Chamberlin, S. I. & Weeks, K. M. Mapping Local Nucleotide Flexibility by Selective Acylation of 2' -Amine Substituted RNA. *J. Am.Chem. Soc* **122**, 216–224 (2000).
191. Chamberlin, S. I., Merino, E. J. & Weeks, K. M. Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14688–93 (2002).
192. Ramesh, A., Wakeman, C. A. & Winkler, W. C. Insights into metalloregulation by M-box riboswitch RNAs via structural analysis of manganese-bound complexes. *J. Mol. Biol.* **407**, 556–70 (2011).
193. Wilkinson, K. A. *et al.* Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA Soc.* **15**, 1314–1321 (2009).
194. Gherghe, C. M., Mortimer, S. a, Krahn, J. M., Thompson, N. L. & Weeks, K. M. Slow conformational dynamics at C2'-endo nucleotides in RNA. *J. Am. Chem. Soc.* **130**, 8884–5 (2008).
195. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. & Altman, R. B. SAFA : Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. 344–354 (2005). doi:10.1261/rna.7214405.interactions
196. Wilkinson, K. A. *et al.* High-Throughput SHAPE Analysis Reveals Structures in HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States. **6**, (2008).
197. McGinnis, J. L., Duncan, C. D. S. & Weeks, K. M. High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol.* **468**, 67–89 (2009).
198. Mitra, S., Shcherbakova, I. V, Altman, R. B., Brenowitz, M. & Laederach, A. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res.* **36**, e63 (2008).

199. Quarrier, S., Martin, J. S., Davis-Neulander, L., Beauregard, A. & Laederach, A. Evaluation of the information content of RNA structure mapping data for secondary structure prediction. *RNA* **16**, 1108–17 (2010).
200. Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M. & Giddings, M. C. ShapeFinder : A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979–1990 (2008).
201. Gherghe, C. M., Shajani, Z., Wilkinson, K. a, Varani, G. & Weeks, K. M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J. Am. Chem. Soc.* **130**, 12244–5 (2008).
202. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–6 (2009).
203. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11063–8 (2011).
204. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11063–8 (2011).
205. Weeks, K. M. RNA structure probing dash seq. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10933–4 (2011).
206. Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 1–8 (2008).
207. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–60 (2008).
208. Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature* **5**, 613–619 (2008).
209. Wan, Y., Qu, K., Ouyang, Z. & Chang, H. Y. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.* **8**, 849–69 (2013).
210. Ouyang, Z., Snyder, M. P. & Chang, H. Y. SeqFold : Genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. 377–387 (2013). doi:10.1101/gr.138545.112.Freely
211. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–7 (2010).
212. Auron, P. E., Weber, L. D. & Rich, A. Comparison of transfer ribonucleic acid structures using cobra venom and S1 endonucleases. *Biochemistry* **21**, 4700–6 (1982).

213. Wurst, R. M., Vournakis, J. N. & Maxam, A. M. Structure mapping of 5'-32P-labeled RNA with S1 nuclease. *Biochemistry* **17**, 4493–9 (1978).
214. Gornicki, P. *et al.* Use of lead(II) to probe the structure of large RNA's. Conformation of the 3' terminal domain of E. coli 16S rRNA and its involvement in building the tRNA binding sites. *J. Biomol. Struct. Dyn.* **6**, 971–84 (1989).
215. Wang, L. K., Lima, C. D. & Shuman, S. Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *EMBO J.* **21**, 3873–80 (2002).
216. Galburt, E. A., Pelletier, J., Wilson, G. & Stoddard, B. L. Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. *Structure* **10**, 1249–60 (2002).
217. Underwood, J. G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
218. Cameron, V. & Uhlenbeck, O. C. 3'-Phosphatase activity in T4 polynucleotide kinase. *Biochemistry* **16**, 5120–6 (1977).
219. Westhof, E. & Romby, P. The RNA structurome: high-throughput probing. *Nat. Methods* **7**, 965–7 (2010).
220. P, L. S., L, M. L. & F, M. W. Rhodamine-based compounds as fluorogenic substrates for serine proteinases. (1983). at <http://www.biochemj.org/bj/209/0299/bj2090299_browse.htm>
221. Uddin, M. J. & Marnett, L. J. Synthesis of 5- and 6-carboxy-X-rhodamines. *Org. Lett.* **10**, 4799–801 (2008).
222. Grohman, J. K., Kottegoda, S., Gorelick, R. J., Allbritton, N. L. & Weeks, K. M. Femtomole SHAPE reveals regulatory structures in the authentic XMRV RNA genome. *J. Am. Chem. Soc.* **133**, 20326–34 (2011).
223. Lee, L. G. *et al.* New energy transfer dyes for DNA sequencing. *Nucleic Acids Res.* **25**, 2816–22 (1997).
224. Rosenblum, B. B. *et al.* New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**, 4500–4 (1997).
225. Swerdlow, H. & Gesteland, R. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res.* **18**, 1415–9 (1990).
226. Karger, a E., Harris, J. M. & Gesteland, R. F. Multiwavelength fluorescence detection for DNA sequencing using capillary electrophoresis. *Nucleic Acids Res.* **19**, 4955–62 (1991).
227. Carson, S. *et al.* DNA sequencing by capillary electrophoresis: use of a two-laser-two-window intensified diode array detection system. *Anal. Chem.* **65**, 3219–26 (1993).

228. França, L. T. C., Carrilho, E. & Kist, T. B. L. A review of DNA sequencing techniques. *Q. Rev. Biophys.* **35**, 169–200 (2002).
229. Josel, et al. rhodamine derivatives and the use thereof. Patent No.: US 6, 184,379 B1 (2001).
230. H.Zakeri, G.Amparo, S.-M.Chen, S. S. and P.-Y. K. Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing. *Biotechniques* **25**, 406–414 (1998).
231. Korch, C. & Drabkin, H. Improved DNA Sequencing Accuracy and Detection of Heterozygous Alleles Using Manganese Citrate and Different Fluorescent Dye Terminators. *Genome Res.* 588–595 (1999). doi:10.1101/gr.9.6.588
232. Ju, J., Ruan, C., Fuller, C. W., Glazer, a N. & Mathies, R. a. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 4347–51 (1995).
233. Hung, S., Ju, J., Mathies, R. A. & Glazer, A. N. Energy Transfer Primers with 5- or 6-Carboxyrhodamine-6G as acceptor chromophores. *Anal. Biochem.* **170**, 165–170 (1996).
234. Ju, J., Glazer, a N. & Mathies, R. a. Cassette labeling for facile construction of energy transfer fluorescent primers. *Nucleic Acids Res.* **24**, 1144–8 (1996).
235. Nampalli, S., Khot, M. & Kumar, S. Fluorescence resonance energy transfer terminators for DNA sequencing. *Tetrahedron Lett.* **41**, 8867–8871 (2000).
236. Kumar, S. *et al.* Synthesis and Energy Transfer Efficiency of FRET Terminators Derived from Different Linkers. *Nucleosides, Nucleotides and Nucleic Acids* **22**, 1595–1598 (2003).
237. Lampkins, A. J., O’Neil, E. J. & Smith, B. D. Bio-orthogonal phosphatidylserine conjugates for delivery and imaging applications. *J. Org. Chem.* **73**, 6053–8 (2008).
238. Valdes-Aguiler, O., & Neckers, D. C. Aggregation Phenomena in Xanthene Dyes. *Accounts Chem. Res.* **22**, 171–177 (1989).
239. Kasha, M. Energy transfer mechanisms and molecular exciton model For molecular aggregates. *Radiat. Res.* **20**, 55–71 (1963).
240. Scholes, G. D. & Ghiggino, K. P. Electronic Interactions and Interchromophore Excitation Transfer. *J. Phys. Chem.* **98**, 4580–4590 (1994).
241. Cantor, Charles.R; Schimmel, P. *Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function.* 395–398 (1980).
242. Packard, B. Z., Toptygin, D. D., Komoriya, a & Brand, L. Profluorescent protease substrates: intramolecular dimers described by the exciton model. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11640–5 (1996).

243. Förster, T. *Compara. Z. Naturforsch. A.* 4:321–327 (1949).
244. Robert Sjöback, Jan Nygern, M. K. Absorption and fluorescence properties of fluorescein. *Spectrochim. Acta* **51**, L7–L21 (1995).
245. Song, A., Wu, T., Chen, S., Zhang, M. & Shen, T. Syntheses and photophysical properties of amphiphilic dyads of fluorescein and carbazole linked with a flexible or semi-rigid bridge. *Dye. Pigment.* **39**, 371–382 (1998).
246. Duan, Y. *et al.* Recent Progress on Synthesis of Fluorescein Probes. *Mini. Rev. Org. Chem.* **6**, 35–43 (2009).
247. Niazi, A. *et al.* Spectrophotometric Investigation of the Acidity Constants of Fluorescein in Various Water-Organic Solvent Media. *Chem. Eng. Commun.* **195**, 1257–1268 (2008).
248. Finn, P. J. *et al.* Synthesis and application of charge-modified dye-labeled dideoxynucleoside-5'-triphosphates to “direct-load” DNA sequencing. *Nucleic Acids Res.* **30**, 2877–85 (2002).
249. Rossi, F. M. & Kao, J. P. Practical method for the multigram separation of the 5- and 6-isomers of carboxyfluorescein. *Bioconjug. Chem.* **8**, 495–7 (1997).
250. Ueno, Y., Jiao, G.-S. & Burgess, K. Preparation of 5- and 6-Carboxyfluorescein. *Synthesis (Stuttg.)* **2004**, 2591–2593 (2004).
251. Zhang, Y.-H. *et al.* An inexpensive fluorescent labeling protocol for bioactive natural products utilizing Cu(I)-catalyzed Huisgen reaction. *Tetrahedron* **63**, 6813–6821 (2007).
252. Wang, H., Li, J., Liu, X. & Zhang, H.-S. N-hydroxysuccinimidyl fluorescein-O-acetate as a highly fluorescent derivatizing reagent for aliphatic amines in liquid chromatography. *Anal. Chim. Acta* **423**, 77–83 (2000).
253. Swamy, K. M. K. *et al.* Fluorescent sensing of pyrophosphate and ATP in 100% aqueous solution using a fluorescein derivative and Mn²⁺. *Tetrahedron Lett.* **48**, 8683–8686 (2007).
254. Lippard, S. J. & Lim, M. H. Fluorescent based sensors for tracking nitric oxide in live cells. WO Patent 2,007,149,105. (2009).
255. Anslyn, E. & Wallace, K. Compositions and methods for the detection of chemical warfare agents. WO Patent 2,008,010,844. (2008).
256. Sun, W.-C., Gee, K. R., Klaubert, D. H. & Haugland, R. P. Synthesis of Fluorinated Fluoresceins. *J. Org. Chem.* **62**, 6469–6475 (1997).
257. Sparano, B. a, Shahi, S. P. & Koide, K. Effect of binding and conformation on fluorescence quenching in new 2',7'-dichlorofluorescein derivatives. *Org. Lett.* **6**, 1947–9 (2004).

258. Mathew, B. P., Kumar, A., Sharma, S., Shukla, P. K. & Nath, M. An eco-friendly synthesis and antimicrobial activities of dihydro-2H-benzo- and naphtho-1,3-oxazine derivatives. *Eur. J. Med. Chem.* **45**, 1502–7 (2010).
259. Beija, M., Afonso, C. a M. & Martinho, J. M. G. Synthesis and applications of Rhodamine derivatives as fluorescent probes. *Chem. Soc. Rev.* **38**, 2410–33 (2009).
260. Kim, K. K. *et al.* Postcolumn HPLC detection of mono- and oligosaccharides with a chemosensor. *Org. Lett.* **5**, 5007–10 (2003).
261. Halo, T. L., Appelbaum, J., Hobert, E. M., Balkin, D. M. & Schepartz, A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc.* **131**, 438–9 (2009).
262. Tang, B. *et al.* A Rhodamine-Based Fluorescent Probe Containing a Se-N Bond for Detecting Thiols and Its Application in Living Cells. *J. Am. Chem. Soc.* 11666–11667 (2007).
263. Shibata, A., Furukawa, K., Abe, H., Tsuneda, S. & Ito, Y. Rhodamine-based fluorogenic probe for imaging biological thiol. *Bioorg. Med. Chem. Lett.* **18**, 2246–9 (2008).
264. Menchen, S.M and Fung, S. 5- and 6- succinimidyl-carboxylate isomers of rhodamine dyes. EP0272007 (1988).
265. Dujols, V., Ford, F. & Czarnik, A. W. A long-wavelength fluorescent chemodosimeter selective for Cu(II) ion in water. *J. Am. Chem. Soc.* **119**, 7386–7387 (1997).
266. DAYU, W., WEI, H., CHUNYING, D., ZHIHUA, L. & QINGJIN, M. Highly sensitive fluorescent probe for selective detection of Hg²⁺ in DMF aqueous media. *Inorg. Chem.* **46**, 1538–1540
267. Xiang, Y., Tong, A., Jin, P. & Ju, Y. New fluorescent rhodamine hydrazone chemosensor for Cu(II) with high selectivity and sensitivity. *Org. Lett.* **8**, 2863–6 (2006).
268. Yang, H. *et al.* Multisignaling optical-electrochemical sensor for Hg²⁺ based on a rhodamine derivative with a ferrocene unit. *Org. Lett.* **9**, 4729–32 (2007).
269. Yang, Y.-K., Yook, K.-J. & Tae, J. A rhodamine-based fluorescent and colorimetric chemodosimeter for the rapid detection of Hg²⁺ ions in aqueous media. *J. Am. Chem. Soc.* **127**, 16760–1 (2005).
270. Gao, Y., Li, J. L. & Shi, Z. Synthesis and application of rhodamine-based fluorescent probe dyes with spacer linker arm. *Chinese Chem. Lett.* **18**, 667–669 (2007).
271. Jiao, G.-S., Castro, J. C., Thoresen, L. H. & Burgess, K. Microwave-assisted syntheses of regioisomerically pure bromorhodamine derivatives. *Org. Lett.* **5**, 3675–7 (2003).

272. Ben Tahar, W. B., Keillor, J. W. & Lubell, W. D. Microwave-assisted synthesis of rhodamine fluorescent tags. *Adv. Exp. Med. Biol.* **611**, 225–6 (2009).
273. Adamczyk, M., Fishpaugh, J. R. & Heuser, K. J. Preparation of succinimidyl and pentafluorophenyl active esters of 5- and 6-carboxyfluorescein. *Bioconjug. Chem.* **8**, 253–5 (1997).
274. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed. Engl.* **40**, 2004–2021 (2001).
275. Himo, F. *et al.* Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *J. Am. Chem. Soc.* **127**, 210–6 (2005).
276. Meldal, M. & Tornøe, C. W. Cu-catalyzed azide-alkyne cycloaddition. *Chem. Rev.* **108**, 2952–3015 (2008).
277. Helms, B., Mynar, J. L., Hawker, C. J. & Fréchet, J. M. J. Dendronized linear polymers via “click chemistry”. *J. Am. Chem. Soc.* **126**, 15020–1 (2004).
278. Lewis, W. G., Magallon, F. G., Fokin, V. V & Finn, M. G. Discovery and characterization of catalysts for azide-alkyne cycloaddition by fluorescence quenching. *J. Am. Chem. Soc.* **126**, 9152–3 (2004).
279. Gramlich, P. M. E., Wirges, C. T., Manetto, A. & Carell, T. Postsynthetic DNA modification through the copper-catalyzed azide-alkyne cycloaddition reaction. *Angew. Chem. Int. Ed. Engl.* **47**, 8350–8 (2008).
280. Hein, J. E. & Fokin, V. V. Copper-catalyzed azide-alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper(I) acetylides. *Chem. Soc. Rev.* **39**, 1302–15 (2010).
281. Rostovtsev, V. V, Green, L. G., Fokin, V. V & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **41**, 2596–9 (2002).
282. Demko, Z. P. & Sharpless, K. B. A Click Chemistry Approach to Tetrazoles by Huisgen 1, 3-Dipolar Cycloaddition: Azides and Acyl Cyanides **. 2113–2116 (2002).
283. Breinbauer, R. & Köhn, M. Azide-alkyne coupling: a powerful reaction for bioconjugate chemistry. *ChemBiochem* **4**, 1147–9 (2003).
284. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128–37 (2003).
285. Thirumurugan, P., Matosiuk, D. & Jozwiak, K. Click Chemistry for Drug Development and Diverse Chemical-Biology Applications. *Chem. Rev.* (2013). doi:10.1021/cr200409f

286. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **12**, 641–55 (2011).
287. W.STADLBAUER, T. K. Isatoic anhydride and their uses in heterocyclic synthesis. *Adv. Heterocycl. Chem.* **28**, 127–178 (1981).
288. Deligeorgiev, T., Vasilev, A., Vaquero, J. J. & Alvarez-Builla, J. A green synthesis of isatoic anhydrides from isatins with urea-hydrogen peroxide complex and ultrasound. *Ultrason. Sonochem.* **14**, 497–501 (2007).
289. Reissenweber, G. PREPARATION OF ISATOIC ANHYDRIDES. Patent US4316020 (1982).
290. Souza, A. M. D. *et al.* 1,3-dipolar cycloaddition-decarboxylation reactions of an Azomethine Ylide with Isatoic Anhydrides : Formation of Novel Benzodiazepinones. *Org. Lett.* **13**, 486–489 (2011).
291. Staiger, P.R; Miller, B. E. Isatoic anhydride .IV. reactions with various nucleophiles. 1214–1219 (1959).
292. STAIGER, P. Roger and WAGNER, E. C. Isatoic anhydride. II. Reactions of isatoic anhydride with amonia. 347–352 (1947).
293. Mortimer, S. a & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–5 (2007).
294. Gelb, M. H. & Abeles, R. H. Substituted isatoic anhydrides: selective inactivators of trypsin-like serine proteases. *J. Med. Chem.* **29**, 585–9 (1986).
295. Gabor, K. *et al.* The chemistry of 2H-3,1-benzoxazine-2,4(1H)dione (isatoic anhydrides) 1. The synthesis of N-substituted 2H-3,1-benzoxazine-2,4(1H)diones. *J. Heterocycl. Chem.* **12**, 565–572 (1975).
296. Carter, M. C. *et al.* 1,4-Benzodiazepines as inhibitors of respiratory syncytial virus. *J. Med. Chem.* **49**, 2311–9 (2006).
297. Güngör, T. *et al.* Synthesis and characterization of 3-arylquinazolinone and 3-arylquinazolinethione derivatives as selective estrogen receptor beta modulators. *J. Med. Chem.* **49**, 2440–55 (2006).
298. Guiles, J. *et al.* Quinazolin-2-ylamino-quinazolin-4-ols as novel non-nucleoside inhibitors of bacterial DNA polymerase III. *Bioorg. Med. Chem. Lett.* **19**, 800–2 (2009).
299. Malamas, M. S. & Millen, J. Quinazolineacetic acids and related analogues as aldose reductase inhibitors. *J. Med. Chem.* **34**, 1492–503 (1991).

300. Wang, T., Lui, A. S. & Cloudsdale, I. S. A Novel Route to Pyrrolo(2,1-C)(1,4) benzodiazepin-5-ones. Formal Total Synthesis of (±) -DC-81. *Org. Lett.* **1**, 1835–1837 (1999).
301. Tedesco, R. *et al.* 3-(1,1-Dioxo-2H-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1H)-quinolinones, potent inhibitors of Hepatitis C Virus RNA-Dependent RNA Polymerase. *J.Med.Chem.* **2**, 971–983 (2006).