HAL
open science

Etude de la dynamique structurale du domaine de liaison au ligand de RXRα et implication de la phosphorylation dans la transcription

Jérôme Eberhardt

To cite this version:

Jérôme Eberhardt. Etude de la dynamique structurale du domaine de liaison au ligand de RXRα et implication de la phosphorylation dans la transcription. Biologie structurale [q-bio.BM]. Université de Strasbourg, 2016. Français. NNT : 2016STRAJ118 . tel-01587807

HAL Id: tel-01587807

https://theses.hal.science/tel-01587807v1

Submitted on 14 Sep 2017

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

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

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE Sciences et de la Vie et de la Santé

Institut de Génétique, Biologie Moléculaire et Cellulaire

THÈSE présentée par :

Jérôme EBERHARDT

soutenue le : 12 décembre 2016

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

Spécialité : **Biophysique et biologie structurale**

Étude de la dynamique structurale du domaine de liaison au ligand de RXRα et implication de la phosphorylation dans la transcription

THÈSE dirigée par :

M STOTE Roland	Directeur de recherches, Université de Strasbourg
Mme DEJAEGERE Annick	Professeur, Université de Strasbourg

RAPPORTEURS :

Mme ETCHEBEST Catherine	Professeur, Université Paris Diderot
M SENET Patrick	Professeur, Université de Bourgogne

AUTRES MEMBRES DU JURY :

Mme KELLENBERGER Esther	Professeur, Université de Strasbourg
M LABESSE Gilles	Directeur de recherches, Université de Montpellier

---

INVITE :

Mme CIANFERANI Sarah	Chargé de recherches, Université de Strasbourg

Table des matières

Glossaire des abréviations **viii**

1 Récepteur X aux Rétinoïdes **1**

1.1 Introduction 2

1.2 Organisation structurale et fonctionnelle 3

1.2.1 Domaine N-terminal 4

1.2.2 Domaine de liaison à l'ADN 5

1.2.3 Région charnière 8

1.2.4 Domaine de liaison au ligand 8

1.3 Partenaires hétérodimériques 12

1.4 Ligands 14

1.5 Dynamique structurale du LBD au sein des NR 18

1.5.1 Étude de la stabilité par protéolyse 20

1.5.2 Étude de la stabilité par Résonance Magnétique Nucléaire (RMN) 20

1.5.3 Étude de la stabilité par Échange Hydrogène/Deutérium (HDX) 23

1.5.4 Étude de la flexibilité de H12 par Anisotropie de Fluorescence 30

1.6 Modification post-traductionnelles 31

1.7 Objectifs 34

2 Dynamique moléculaire accélérée **37**

2.1 Introduction 38

2.2 Dynamique moléculaire accélérée 38

2.2.1 Introduction 38

2.2.2 Théorie de la dynamique moléculaire accélérée 40

2.2.3 Repondération de l'énergie 42
2.2.4 Variantes de la dynamique moléculaire accélérée 45
2.2.5 Applications 47
2.3 Analyse des trajectoires 48
2.3.1 RMSD global 48
2.3.2 Fluctuation RMS 48
2.3.3 Structure secondaire 49

3 Réduction de la dimensionnalité 50
3.1 Introduction 51
3.2 Analyse sur composante principale 52
3.2.1 Introduction 52
3.2.2 Analyse de trajectoires de dynamique moléculaire 53
3.3 MultiDimensional Scaling 54
3.4 ISOmetric MAPping 57
3.5 Stochastic Proximity Embedding 59
3.5.1 Application note 59
3.5.2 Supplementary material 62
3.5.3 Analyse du domaine de liaison à l'immunoglobuline de la protéine G 71

4 Comparaison Modèles-Expériences 75
4.1 Déplacements chimiques RMN 76
4.1.1 Introduction 76
4.1.2 Déplacements chimiques secondaires 80
4.1.3 Prédiction des déplacements chimiques 80
4.1.4 Comparaison avec les données expérimentales 82
4.2 Échange Hydrogène/Deutérium 83
4.2.1 Introduction 83
4.2.2 Mesure expérimentale de l'échange H/D 84
4.2.3 Mécanisme d'échange HX 85
4.2.3.1 Général 85
4.2.3.2 Au sein des protéines 92

4.2.3.3 Modélisation de la vitesse d'échange HX 94
4.2.4 Modélisation du transfert de protons 95
4.2.4.1 Limites des modèles existants 95
4.2.4.2 Modèles utilisés 97
4.2.4.2.1 Modèle simple 97
4.2.4.2.2 Modèle concerté 98
4.2.5 Prédiction des facteurs de protection 99
4.2.6 Comparaison avec les données expérimentales 100

5 Dynamique conformationnelle du LBD de RXRα 101
5.1 Introduction 102
5.2 Matériels et Méthodes 102
5.2.1 Choix et comparaison des structures cristallographiques 102
5.2.2 Protocole de simulation aMD 103
5.2.3 Méthodes d'analyse des trajectoires MD 105
5.2.3.1 RMSD et Fluctuations RMS 105
5.2.3.2 Structure secondaire 105
5.2.3.3 Déplacements chimiques RMN 106
5.2.3.4 Échange Hydrogène/Deutérium 106
5.2.3.5 Analyse sur Composante Principale 106
5.2.3.6 Stochastic Proximity Embedding 107
5.2.3.7 Repondération de l'énergie des trajectoires aMD 108
5.2.3.8 Visualisation des conformations 108
5.3 Résultats 108
5.3.1 Analyse des structures cristallographiques 108
5.3.1.1 Forme inactive du LBD de RXRα 109
5.3.1.1.1 Sélection des structures cristallographiques disponibles 109
5.3.1.1.2 Fluctuations RMS 111
5.3.1.2 Forme active, en complexe avec 9cRA, du LBD de RXRα 112
5.3.1.2.1 Sélection des structures cristallographiques disponibles 112
5.3.1.2.2 Fluctuations RMS 112
5.3.2 Analyse des trajectoires aMD 114

5.3.2.1 Forme libre du LBD de RXRα 114
5.3.2.1.1 RMSD global 114
5.3.2.1.2 Fluctuations RMS 114
5.3.2.1.3 Structure secondaire 116
5.3.2.2 Forme liée à 9cRA du LBD de RXRα 117
5.3.2.2.1 RMSD global 117
5.3.2.2.2 Fluctuations RMS 117
5.3.2.2.3 Structure secondaire 119
5.3.3 Analyse des déplacements chimiques RMN 119
5.3.3.1 Forme libre du LBD de RXRα 119
5.3.3.2 Forme liée à 9cRA du LBD de RXRα 124
5.3.4 Analyse de l'Échange Hydrogène/Deutérium 126
5.3.4.1 Détermination du facteur de protection maximal expérimental 126
5.3.4.2 Détermination des constantes de vitesses d'échanges intrinsèques 127
5.3.4.3 Prédiction du taux d'incorporation en deutérium 128
5.3.4.3.1 Forme libre du LBD de RXRα 128
5.3.4.3.2 Forme liée à 9cRA du LBD de RXRα 138
5.3.5 Analyse des conformations échantillonnées 146
5.3.5.1 Analyse sur composante principale (PCA) 146
5.3.5.1.1 Forme libre du LBD de RXRα 146
5.3.5.1.2 Forme liée à 9cRA du LBD de RXRα 147
5.3.5.2 Stochastic Proximity Embedding (pSPE) 148
5.3.5.2.1 Forme libre du LBD de RXRα 148
5.3.5.2.2 Forme liée à 9cRA du LBD de RXRα 149
5.3.5.3 Description des conformations 151
5.3.5.3.1 Forme libre du LBD de RXRα 151
5.3.5.3.2 Forme liée à 9cRA du LBD de RXRα 158
5.4 Discussion 163

6 Effet de la phosphorylation pSer260 172
6.1 Introduction 173
6.2 Matériels et Méthodes 173

6.2.1 Protocole de simulation aMD 173
6.2.2 Méthodes d'analyse des trajectoires aMD 174
6.2.2.1 RMSD et Fluctuations RMS 174
6.2.2.2 Structure secondaire 174
6.2.2.3 Déplacements chimiques RMN 175
6.2.2.4 Échange Hydrogène/Deutérium 175
6.2.2.5 Stochastic Proximity Embedding 175
6.2.2.6 Repondération de l'énergie des trajectoires aMD 176
6.2.2.7 Visualisation des conformations 177
6.3 Résultats 177
6.3.1 Analyse des trajectoires aMD 177
6.3.1.1 RMSD global 177
6.3.1.2 Fluctuation RMS 178
6.3.1.3 Structure secondaire 179
6.3.2 Analyse des déplacements chimiques RMN 180
6.3.3 Analyse de l'Échange Hydrogène/Deutérium 181
6.3.4 Analyse des conformations échantillonnées 185
6.4 Discussion 188

7 Conclusion et perspectives 191

Bibliographie 196

Annexes 231

Publication 237

Glossaire des abréviations

9cRA	Acide Rétinoïque 9-cis
AF-1	Fonction Activatrice 1
AF-2	Fonction Activatrice 2
aMD	Dynamique Moléculaire accélérée
AtRA	Acide Rétinoïque tout-trans
BPTI	*Bovine Pancreatic Trypsin Inhibitor*
CMAP	*Correction MAP*
cMD	Dynamique Moléculaire classique
CoA	CoAactivateur
CoR	CoRepresseur
cPCA	Analyse sur Composante Principale sur les coordonnées cartésiennes
CV	Variable Collective
DA-SMACOF	*Deterministic Annealing SMACOF*
DBD	Domaine de Liaison à l'ADN
dPCA	Analyse sur Composante Principale sur les angles dièdres
DR	Répétition Directe
DSSP	*Dictionnary of Secondary Structure Protein*
EX1/2	*Exchange mechanism 1/2*
FID	Décroissance Libre d'Induction
FXR	Récepteur X aux Farnesoids
GR	Récepteur aux Glucocorticoïdes
HDX	Échange Hydrogène/Deutérium
HRE	Élément de Réponse aux Hormones
IR	Répétition Inverse
IRM	Imagérie par Résonance Magnétique
ISOMAP	*ISOmetric MAPping*
LBD	Domaine de Liaison au Ligand
LBP	Poche de Liaison au Ligand
LXR	Récepteurs des Oxystérols
MD	Dynamique Moléculaire
MDS	*MultiDimensional Scaling*
NLS	Signal de Localisation Nucléaire
NTD	Domaine N-Terminal
NURR1	*Nuclear Receptor Related 1*
PCoA	*Principal Coordinates Analysis*
pD	potentiel de Deutérium
PDB	*Protein DataBank*
PDLA	Poly-DL-Alanine
pH	potentiel d'Hydrogène
PME	*Particle Mesh Ewald*
PMF	*Potential Mean Force*
PPAR	Récepteur Activé par les Proliférateurs de Peroxysomes
PR	Récepteur à la Progestérone
pSPE	*pivot-based Stochastic Proximity Embedding*
RAR	Récepteur à l'Acide Rétinoïde
REMD	*Replica Exchange Molecular Dynamics*

RE/PT	*Replica Exhange / Parallel Tempering*
RMN	Résonance Magnétique Nucléaire
RMSD	Déviation RMS
RMSF	Fluctuation RMS
RXR	Récepteur X aux Rétinoïdes
SAXS	Diffusion des rayons X aux petits angles
SMACOF	*Scaling by MAjorizing a COmplicated Function*
SPE	*Stochastic Proximity Embedding*
SRC-1	*Steroid Receptor Coactivator 1*
SVD	Décomposition en Valeur Singulière
TR	Récepteurs des hormones Thyroïdes
VDR	Récepteur à la Vitamine D

Chapitre 1

Récepteur X aux Rétinoïdes

1.1 Introduction

La vitamine A et ses dérivés, nommés rétinoïdes, sont des hormones non stéroïdiennes jouant un rôle crucial dans le développement et l'homéostasie dans tous les tissus des vertébrés en agissant sur la différenciation cellulaire, la prolifération et l'apoptose [Ross et al., 2000] [Altucci and Gronemeyer, 2001]. Ce signal est transduit par deux familles de récepteurs nucléaires (NR), le récepteur à l'acide rétinoïde (RAR) et le récepteur X aux rétinoïdes (RXR). Le RXR occupe une place centrale au sein de la superfamille des récepteurs nucléaires en étant le partenaire d'hétérodimérisation de nombreux NR, ainsi qu'en formant des homodimères. Il est le partenaire de la classe II des récepteurs nucléaires [Brelivet et al., 2004], composée du récepteur à l'acide rétinoïque (RAR-α, β et γ), du récepteur des hormones thyroïdiennes (TR-α et β) et du récepteur à la vitamine D (VDR) ainsi que d'autres récepteurs nucléaires « orphelins adoptés » dont le récepteur activé par les proliférateurs de peroxysomes (PPAR-α, β/δ et γ) activé par les lipides, le récepteur X du foie (LXR-α et β) activé par des oxystérols et impliqué dans le métabolisme du cholestérol et des lipides, le récepteur X aux farnésoïdes (FXR) liant l'acide biliaire et le récepteur X aux prégnanes (PXR) impliqué dans la détoxification cellulaire (xénobiotique).

De fait, RXR est un facteur essentiel dans plusieurs processus biologiques et par conséquent joue un rôle prépondérant dans de nombreuses maladies et cancers [Lefebvre et al., 2010] [Gilardi and Desvergne, 2014]. Il existe trois isotypes de RXR, RXRα (NR2B1), RXRβ (NR2B2) et RXRγ (NR2B3), qui sont exprimés différemment selon le tissu, sachant que RXRα est l'isotype prédominant [Germain et al., 2006]. Le premier membre du groupe des RXR a été indépendamment identifié par les laboratoires de Keiko Ozato et Ron Evans. Il a été appelé dans un premier temps, H-2RIIBP pour *H2 region II binding protein* car se liant à un élément régulateur aux gènes de classe I du complexe histocompatibilité chez la souris [Hamada et al., 1989]. Le nom de RXR sera utilisé par la suite du fait qu'il était considéré comme orphelin et qu'il ne répondait spécifiquement qu'aux acides rétinoïques [Mangelsdorf et al., 1990]. Le clone initialement découvert par le laboratoire de Ron Evans sera appelé RXRα, ainsi que RXRβ correspondant à H-2RIIBP, l'homologue chez la souris. Plus tard, un troisième gène sera trouvé chez le poulet ainsi que chez les mammifères et nommé RXRγ [Rowe et al., 1991]. RXR est exprimé dans chaque tissu du corps, RXRα est principalement exprimé dans le foie, les poumons, les muscles, les reins, l'épiderme, l'intestin et la peau, RXRβ est exprimé de façon ubiquitaire tout comme RXRα, tandis que RXRγ est trouvé dans le cerveau, les muscles squelettiques et cardiaques [Germain et al., 2006]. La déficience en RXRα ou RXRβ est léthale au stade de l'embryon, tandis que l'inactivation du gène de RXRγ ne semble pas d'avoir d'effet chez la souris. Parmi l'isotype RXRα, il existe plusieurs isoformes générés par épissage alternatif, dont la forme

dominante RXRα1, qui encode une protéine de 52 kDa. Deux autres isoformes, RXRα2 et RXRα3 avec 28 et 97 acides aminés absents dans le domaine N-terminal ont été identifiés dans les gonades mâles de souris [Brocard et al., 1996].

1.2 Organisation structurale et fonctionnelle

Comme la majorité des récepteurs nucléaires, RXR est composé de 5 domaines fonctionnels et structuraux bien distincts, en excluant le domaine F non présent chez RXR (Figure 1.1 et 1.2). Le premier domaine nommé A/B ou domaine N-terminal (NTD), contient la fonction activatrice indépendante du ligand (AF-1), sur lequel d'autres facteurs de transcription vont se lier, puis le domaine C ou domaine de liaison à l'ADN (DBD), chargé de reconnaitre les éléments de réponse aux hormones spécifiques sur l'ADN, suivi de la région charnière ou domaine D. Finalement, après la région charnière, vient le domaine E qui est le domaine de liaison au ligand (LBD). Ce dernier contient la fonction activatrice dépendante au ligand (AF-2) sur lequel vont se lier des protéines CoActivatrices (CoA) ou CoRépressives (CoR), régulant l'activité transcriptionnelle.

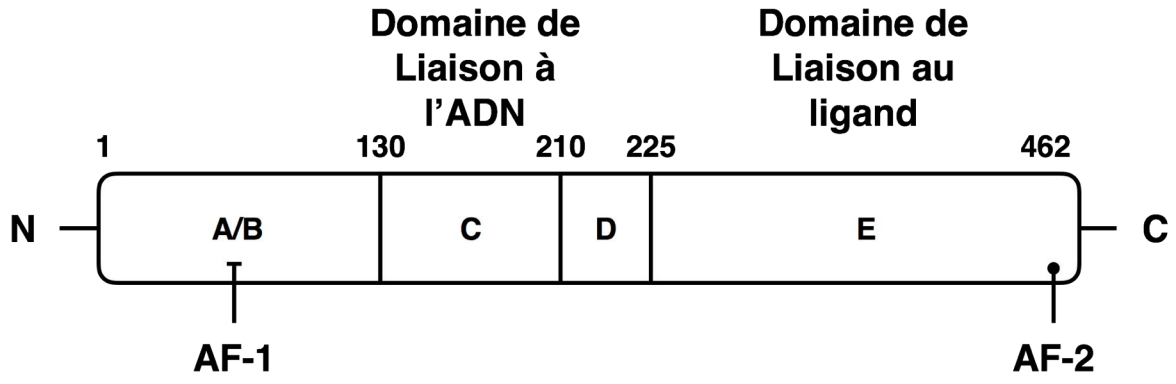


FIGURE 1.1 – Dessin schématique du récepteur nucléaire RXRα. Ce récepteur nucléaire est composé d'un domaine A/B contenant la fonction activatrice AF-1, d'un domaine C qui est le domaine de liaison à l'ADN (DBD), d'un domaine D qui sert de région charnière, et du domaine E qui est le domaine de liaison au ligand (LBD). Les domaines C et E contiennent les interfaces de dimérisation. La fonction activatrice indépendante d'un ligand (AF-1) est localisée dans le domaine A/B, quant à la fonction AF-2 dépendante d'un ligand elle est localisée dans la partie C-terminale du LBD.

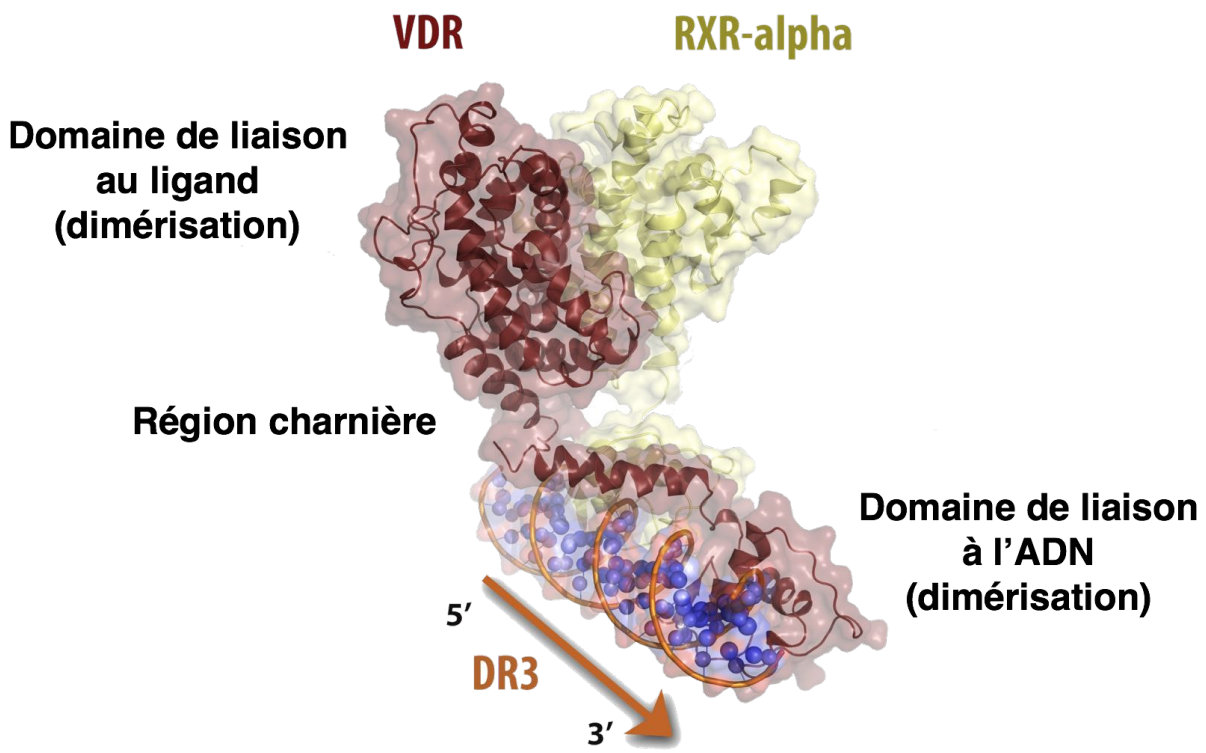


FIGURE 1.2 – Architecture tridimensionnelle suggérée par les études en solution de l'hétérodimère RXRα-VDR lié à un élément de réponse aux hormones (HRE) canonique VDRE [Orlov et al., 2012] [Rochel et al., 2011]. Adapté de [Orlov et al., 2012].

1.2.1 Domaine N-terminal

Le NTD des récepteurs nucléaires varie considérablement en termes de taille et de séquence. Ce domaine est le moins conservé de tous au sein de cette superfamille. De plus, il a été montré que la majorité des isoformes connues diffèrent essentiellement au niveau de ce domaine [Lavery and Mcewan, 2005]. Le domaine NTD possède une fonction AF-1, qui module la transcription des gènes cibles selon le contexte cellulaire et de manière dépendante du promoteur [Kumar and McEwan, 2012]. Le NTD est aussi sujet à des modifications post-traductionnelles, telles que l'ubiquitination, la phosphorylation, la sumoylation ou l'acétylation, régulant l'activité transcriptionnelle du récepteur [Boudjelal et al., 2000] [Choi et al., 2006] [Rochette-Egly, 2003] [Zhao et al., 2007]. Le NTD du RXRα possède 130 acides aminés est caractérisé par une large proportion de résidus Gly et Pro, habituellement trouvés au sein des protéines intrinsèquement désordonnées. Les études d'échange Hydrogène/Deutérium (HDX) ont montré que le NTD de RXRα était

très accessible au solvant [Zhang et al., 2011c]. Ce domaine n'est pas observable au sein des structures cristallographiques des complexes hétérodimériques [Rastinejad et al., 1995] [Lou et al., 2014] [Chandra et al., 2008]. Des études complémentaires du domaine en solution, par SAXS et RMN [Belorusova et al., 2016], ont montré qu'il était principalement désordonné et restait flexible même après liaison à l'ADN, contrairement au NTD des récepteurs aux stéroïdes qui adoptent un état globulaire préfondu (*premolten globule state*).

1.2.2 Domaine de liaison à l'ADN

Le DBD est très conservé au sein de la superfamille des récepteurs nucléaires, avec une identité de séquence entre 88 et 95 % entre les trois isoformes de RXR. L'architecture générale du DBD est identique à celle des DBD d'autres récepteurs nucléaires tels que le récepteur aux glucocorticoïdes (GR), le récepteur à la progestérone (PR), le récepteur aux androgènes (AR) ou le récepteur à l'oestradiol (ER), avec 2 hélices α orientées perpendiculairement. Cependant, contrairement à ces derniers qui se lient aux éléments de réponse aux hormones (HRE) sous forme d'homodimère de type Répétition Inverse (IR) composé de deux demi-sites séparés par trois paires de bases (IR3) [Beato, 1991], les récepteurs non stéroïdiens se lient préférentiellement à des HRE de type Répétition Directe (DR) où l'espacement entre les deux demi-sites est important et dont la séquence consensus est 5"-(G/T)TCA-N_n-PuG(G/T)TCA-3"[Koenig et al., 1987] [Näär et al., 1991] [Umesono et al., 1991]. En plus de la reconnaissance de HRE, le DBD est impliqué dans la dimérisation avec d'autres membres des récepteurs nucléaires. Ainsi plusieurs régions ont été caractérisées au sein du DBD de RXR, leur fonction est liée à (1) la reconnaissance spécifique du HRE et (2) à la dimérisation avec le partenaire, aussi au niveau du LBD (Figure 1.3).

La première structure du DBD de RXR a été déterminée par RMN [Lee et al., 1993] [Holmbeck et al., 1998]. Ce domaine est composé de deux doigts de zinc, qui interagissent chacun avec un ion Zn (II) par le biais de 4 cystéines. Les motifs de ces deux doigts de zinc sont Cys-N_2-Cys-N_{13}-Cys-N_2-Cys et Cys-N_5-Cys-N_9-Cys-N_2-Cys. L'hélice I de reconnaissance, résidus Cys152 à Arg 164, interagit avec le sillon majeur de l'ADN. Les résidus 187 à 190 du doigt de zinc II ainsi que les résidus 191 à 198 constituent l'hélice II, suivie par la boite T. Dans le cas de l'homodimère RXR sur un HRE de type DR1, la boite T du RXR en 3' interagit avec le doigt de zinc II du RXR en 5', fournissant une interface de dimérisation entre les DBD. Les interactions sont médiées par les résidus Gln183 et Arg186, du doigt de zinc II avec le résidu Glu208 de la boite T. La région contenant la boite T peut être soit sous forme désordonnée soit sous forme hélice α, suivant si le DBD est en complexe avec l'ADN ou non. Lorsque le DBD est en solution, la

boite T forme une 3e hélice du fait de l'interaction entre Glu208 et les résidus Lys160 et Arg164, masquant l'hélice I de reconnaissance. Ce qui n'est pas le cas lorsque le DBD est lié à l'ADN [Zhao et al., 2000]. La région du DBD, contenant les résidus 200 à 204, est nécessaire pour la liaison à l'ADN et de même pour la formation de l'homodimère RXR, à partir du résidu 204 [Lee et al., 1993] et la formation de complexes hétérodimériques [Rastinejad et al., 1995] [Wilson et al., 1992].

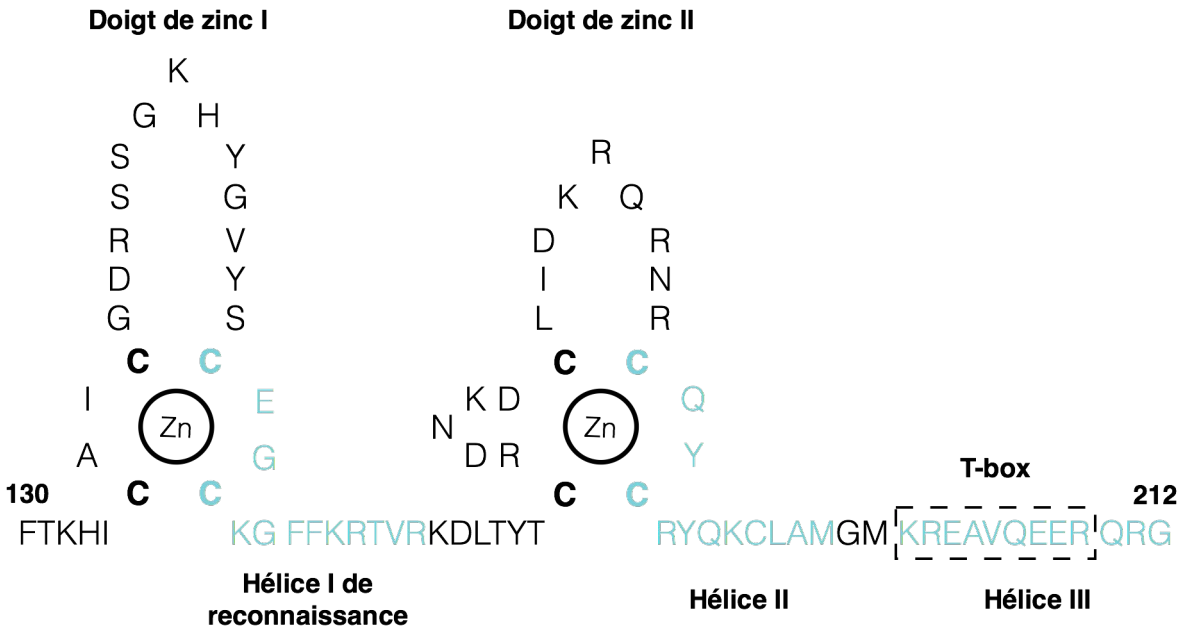


FIGURE 1.3 – Structure du DBD de RXRα montrant les deux doigts de zinc I et II, étant chacun stabilisé par un ion Zinc (II), et avec 4 cystéines (I : C135, C138, C152 et C155 et II : C171, C177, C187 et C190). L'hélice I de reconnaissance à l'ADN, résidus C152 à R164, et l'hélice II, résidus C187 à M198, sont montrées en bleu. La séquence de la boite T, résidus K201 à K209, contenu dans l'hélice III, résidus K201 à G212, en bleu, est encadrée.

En plus de pouvoir former un homodimère sur des HRE de type DR1, RXR peut aussi former des complexes hétérodimériques. Dans le cas d'un hétérodimère, RXR se lie avec une polarité inversée sur un HRE de type DR1, c'est-à-dire du côté 3' du HRE [Rastinejad et al., 2000] [Chandra et al., 2008] [Kurokawa et al., 1994]. Ceci est dû au fait que RXR a une plus forte affinité sur le demi-site en 3' de type DR1 [Osz et al., 2015]. Cependant dans le cas où l'espacement entre les deux demi-sites est supérieur à 1, RXR va se placer préférentiellement du côté 5' du HRE. Par exemple, le HRE optimal pour le complexe

RXR-VDR est de type DR3, pour RXR-TR de type DR4 et RXR-RAR de type DR2 ou DR5 [Perlmann et al., 1993] [Umesono et al., 1991]. De même, la boite T est importante pour l'interaction entre les DBD des partenaires au sein d'un complexe hétérodimérique, son rôle a été confirmé par la résolution de la structure du complexe RXR-TR lié à l'élément de réponse de type DR4, où c'est la boite T du partenaire TR qui interagit avec le doigt de zinc II de RXR, de même qu'elle confirme le rôle discriminatoire du nombre de nucléotides espaceurs dans les DR [Rastinejad et al., 1995] [Zechel et al., 1994] (Figure 1.4).

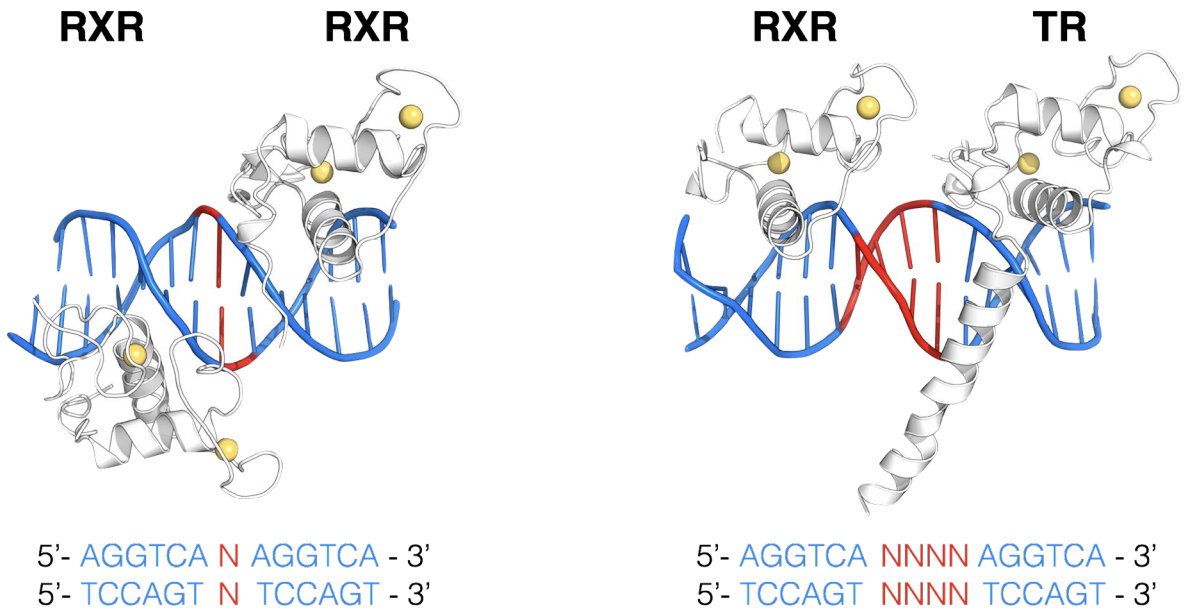


FIGURE 1.4 – Domaines de liaison à l'ADN (DBD) des récepteurs nucléaires. À gauche, l'homodimère RXR-RXR sur un élément de réponse aux hormones (HRE) idéal de type DR1 (PDB id : 1BY4) [Zhao et al., 2000]. Et à droite, le complexe hétérodimérique RXR-TR sur un HRE idéal de type DR4 (PDB id : 2NLL) [Rastinejad et al., 1995]. Les demi-sites sont représentés en bleu, tandis que les paires de bases entre les deux demi-sites sont en rouge. Les ions Zinc (II) sont représentés par des sphères et les séquences des HRE sont représentées en dessous.

Contrairement à d'autres récepteurs nucléaires, la séquence de localisation nucléaire (NLS) de RXRα est située dans le DBD et non la région charnière, entre le DBD et le LBD. La NLS est située entre les deux doigts de zinc, entre les résidus 160 et 165, dont la séquence est KRTVRK [Prüfer and Barsony, 2002].

1.2.3 Région charnière

La région charnière entre le DBD et le LBD est composée d'environ 20 résidus, selon où on positionne le C-terminal du DBD, résidus 200 ou 209. Cette région est très flexible et permet au RXR de se lier à plusieurs HRE selon le partenaire hétérodimérique. Ceci permet, par exemple, au LBD de RXRα de se placer à l'opposé sur l'ADN afin que le LBD de PPARγ ait suffisamment d'espace pour se positionner entre le DBD et le LBD de RXRα [Chandra et al., 2008].

1.2.4 Domaine de liaison au ligand

De tous les domaines du RXR, le LBD a été celui le plus étudié. Il a été étudié soit seul, lié à des ligands de natures diverses, dimérisé avec d'autres récepteurs nucléaires et avec des peptides coactivateurs (CoA) ou corépresseurs (CoR). La structure du LBD, qui est conservée parmi la superfamille des récepteurs nucléaires, est composée de 12 hélices α et un petit feuillet β, entre les hélices H5 et H6. Ces hélices α sont arrangées en sandwich d'hélices α anti parallèles. Les hélices H4, H5, H8, H9 et H11 sont entre deux couches composées de H1 et H3 d'un côté et H6, H7, H10 de l'autre. Ce domaine fonctionnel du RXR inclut une surface d'interaction pour les CoA/CoR, une surface de dimérisation avec les autres récepteurs nucléaires et enfin une poche de liaison au ligand.

La structure du LBD de RXRα en conformation inactive, c'est-à-dire en absence de ligand, fut la première structure d'un LBD de récepteur nucléaire être déterminé par cristallographie aux rayons X [Bourguet et al., 1995]. La comparaison de la structure inactive du LBD de RXRα avec le LBD de RARγ en conformation active en complexe avec AtRA [Renaud et al., 1995], et par la suite avec la conformation active du LBD de RXRα en complexe avec 9cRA [Egea et al., 2000], suggère un mécanisme commun d'activation est activé et décrit selon le modèle nommé « piège à souris » (*mouse trap*) (Figure 1.5).

D'après les structures cristallographiques, les 12 hélices α sont présentes lorsqu'il n'y a pas de ligand. Cependant, la présence d'un ligand agoniste va entrainer le repositionnement de l'hélice H11 dans la continuité de l'hélice H10, permettant ainsi à H12 de libérer la boucle Ω, entre H1 et H3, entrainant le déroulement de l'hélice H2 et provoquant la courbure de la partie N-terminale de H3. Dans la position finale, l'hélice H12 vient sceller la poche de liaison aux ligands et stabilisant davantage le ligand en contribuant à l'environnement hydrophobe (Figure 1.6). De plus, la nouvelle position de H12 va aussi exposer certains résidus, Phe450, Glu453 et Glu456, impliqués dans l'interaction avec les CoA. Ainsi l'implication de l'hélice H12 est critique pour la transactivation et le recrutement des CoA en formant une surface d'interaction permettant au CoA possédant un motif NR-box LxxLL de lier le LBD [Gronemeyer and Laudet,

1995] [Chambon, 1996]. Une fois le CoA lié, celui-ci va recruter d'autres protéines régulatrices, dont le complexe de transcription, connectant le HRE au site d'initiation de la transcription.

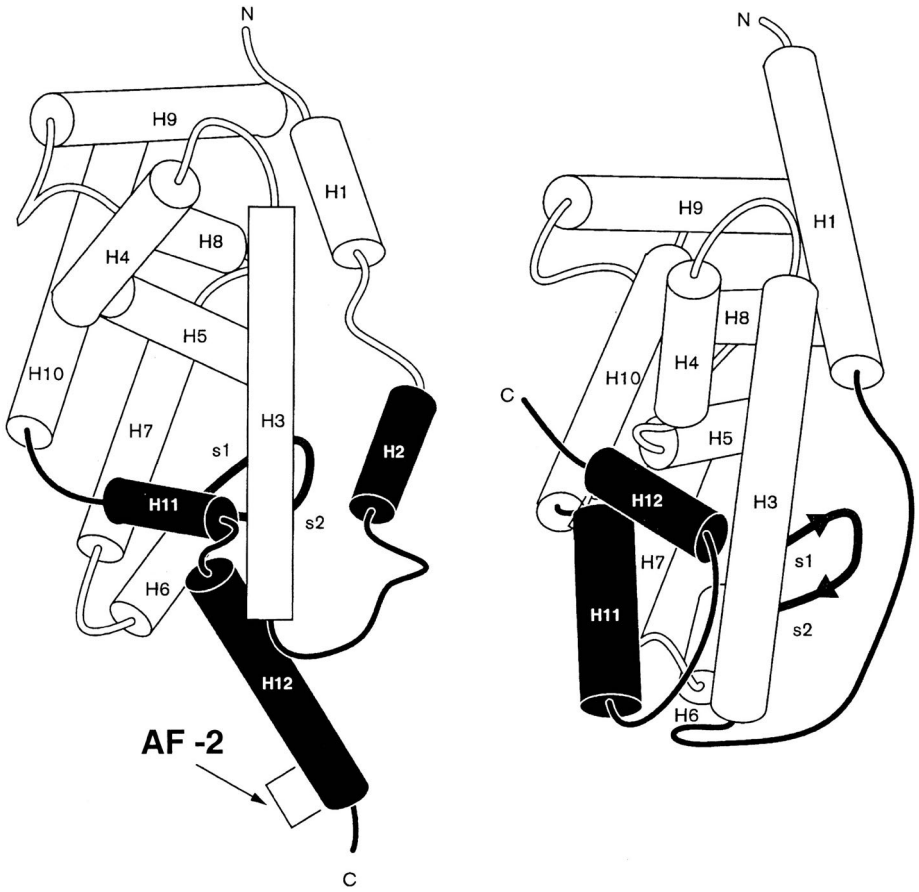


FIGURE 1.5 – Dessin schématique du domaine de liaison au ligand (LBD), adapté de [Wurtz et al., 1996]. Sur la gauche, la structure cristallographique du LBD de RXRα en conformation inactive et sur la droite, le LBD de RARγ en conformation active en complexe avec AtRA [Wurtz et al., 1996]. Ces deux structures cristallographiques sont à l'origine de l'élaboration du mécanisme d'activation de la fonction AF-2. Les hélices α sont représentées par des cylindres numérotés de 1 à 12. La principale différence se situant au niveau de la partie C-terminale avec H12, responsable de la fonction AF-2.

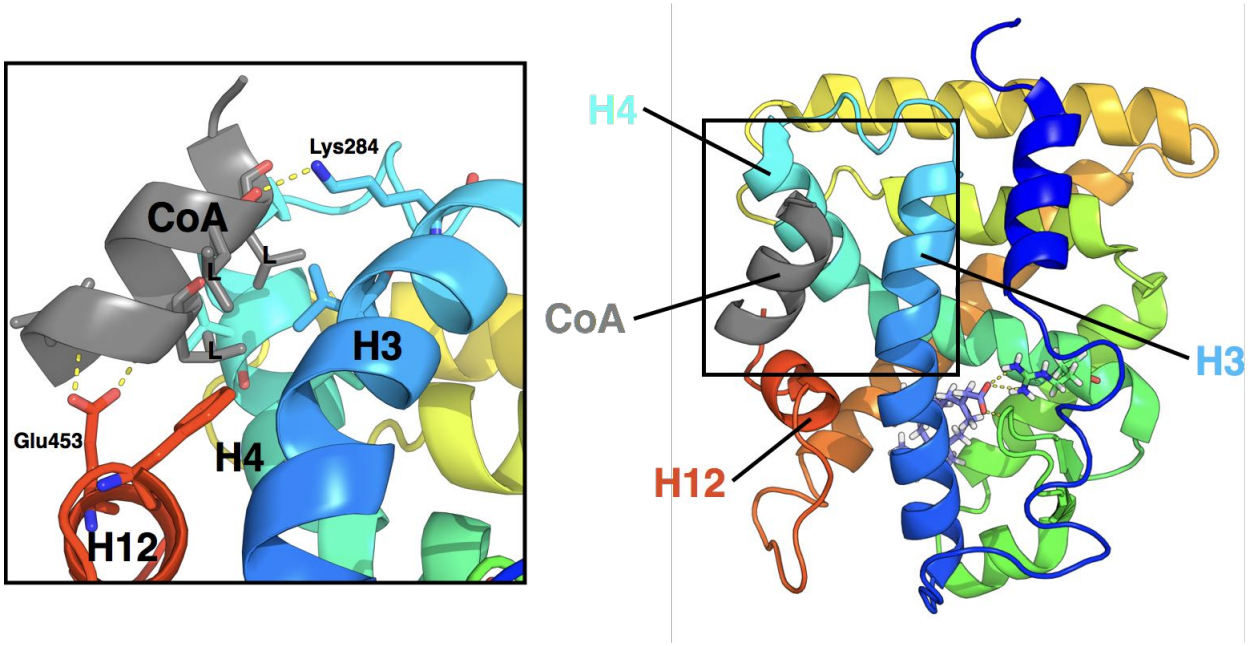


FIGURE 1.6 – Localisation de la surface d'interaction pour les protéines coactivatrices (CoA), composée de H3, H4 et H12, au sein du LBD de RXRα. Après liaison du ligand agoniste, l'acide rétinoïque 9-cis, H12 se repositionne dans une conformation active, d'après le modèle de « piège à souris », permettant la création d'une surface d'interaction spécifique pour les CoA. Cette surface est reconnue par le motif LxxLL contenu dans le peptide CoA. Lys284 (H3) et Glu453 (H12) interagissent avec le peptide CoA en formant une pince chargée.

La poche de liaisons aux ligands (LBP) de RXR est bordée par des résidus principalement hydrophobes. Dans certains récepteurs nucléaires, tels que PPAR, LXR, PXR, la LBP est relativement large et n'est pas totalement occupée par les ligands connus. Dans le cas des récepteurs nucléaires ayant des ligands de plus forte affinité, comme RXR, ceux-ci entretiennent une relation plus étroite avec le récepteur. La LBP est beaucoup plus petite, entre 400 et 500 $Å^3$. Ainsi les ligands qui se lient à RXR doivent s'adapter au volume disponible de la poche par torsion des liaisons simples. C'est le cas par exemple de l'acide rétinoïque 9-*cis*, occupant pratiquement tout le volume de la cavité, et dont une torsion est visible au niveau des carbones C8-C9. Ce qui a pour conséquence d'orienter le groupement $beta$-ionone perpendiculairement à la partie C9-C15. La poche hydrophobe du ligand est formée par les hélices H3, H5, H7, H11 ainsi que le feuillet β et l'hélice H12 (Figure 1.7). De nombreux résidus interagissent directement avec le ligand par des contacts

de van der Waals et par un réseau de liaisons hydrogène et ioniques impliquant leur partie carboxyle et une arginine conservée dans H5 (Arg316) et le squelette peptidique au niveau du feuillet β (Ala327) [Klaholz et al., 1998]. Tous les résidus au sein de la LBP sont conservés dans les trois isoformes α, β et γ de RXR.

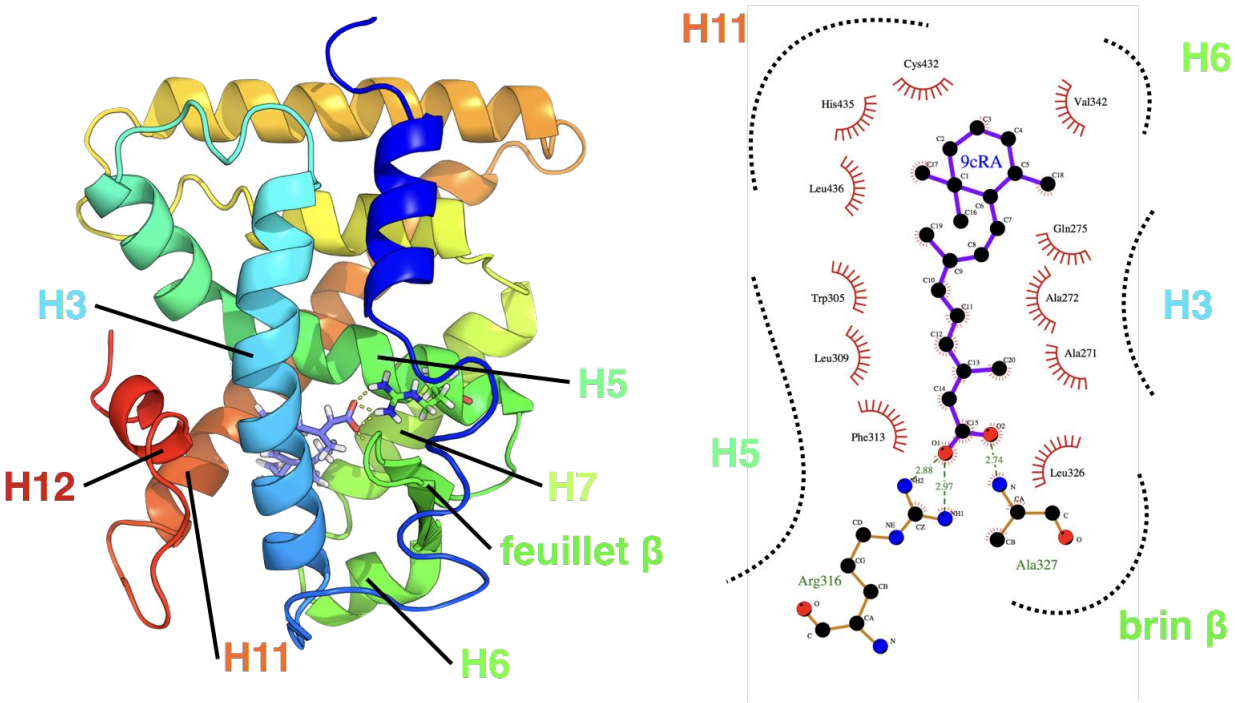


FIGURE 1.7 – Localisation de la poche de liaison au ligand (LBP) au sein du LBD de RXRα. La LBP est composée des éléments structuraux H3, H5, le feuillet β, H6, H7, H11 et H12. L'acide rétinoïque 9-cis (9cRA) interagit principalement avec les résidus Ala271, Ala272 et Gln275 (H3), Trp305, Leu309 et Phe313 (H5), Leu326 (feuillet β), Val342 (H6) et Cys432, His435 et Leu346 (H11) par contacts hydrophobes et avec Arg316 (H5) et Ala327 (feuillet β) par liaisons hydrogène. Cependant, bien que H12 soit un élément essentiel de la LBP, celle-ci n'interagit pas directement avec 9cRA.

L'hélice H10 ainsi que H7, H9 et la boucle H8/H9, dans une moindre mesure, forment l'interface de dimérisation entre les deux monomères (Figure 1.8). Bien que l'hétérodimérisation au niveau du DBD soit nécessaire pour interagir avec le HRE, le LBD de RXR joue aussi un rôle important dans la formation du complexe hétérodimérique au niveau du motif heptad LLLRLPA (résidus 418-424) situé dans l'hélice H10. La mutation de ce motif va affecter leur formation [Au-Fliegner et al., 1993]. De même que pour la LBP, les résidus situés dans l'interface de dimérisation sont conservés entre les différents isoformes de RXR.

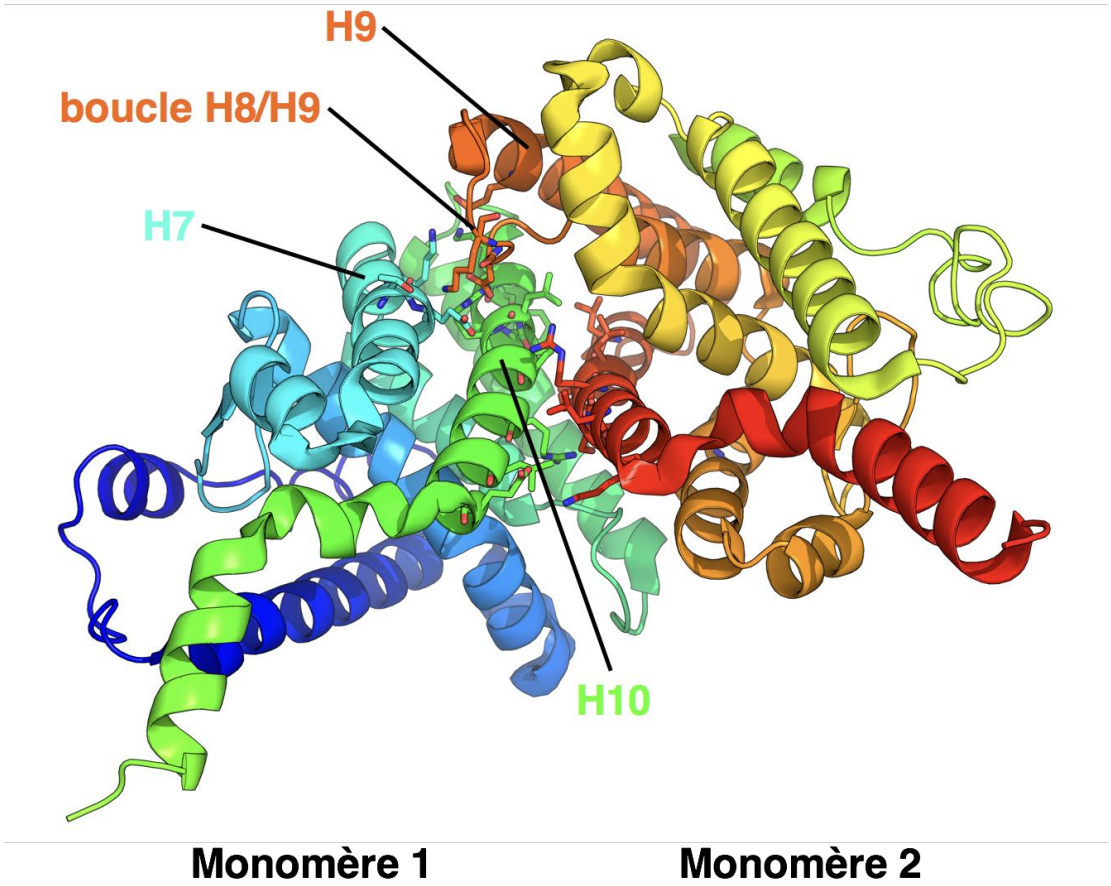


FIGURE 1.8 – Localisation de l'interface de dimérisation au sein de l'homodimère RXRα. L'interface de dimérisation est composée de H7, H9, H10 ainsi que la boucle H8/H9. L'interface entre les monomères 1 et 2 est un ensemble riche en résidus interagissant par contacts hydrophobes et interactions électrostatiques.

1.3 Partenaires hétérodimériques

Une des caractéristiques des récepteurs hétérodimériques est qu'ils peuvent potentiellement être activés, soit par le ligand du RXR ou par celui de son partenaire. On peut ainsi distinguer deux catégories de partenaires hétérodimériques de RXR : les partenaires non permissifs, comme VDR, RAR et TR ou bien les partenaires permissifs, tels que FXR, LXR et PPAR [Forman et al., 1995] [Kurokawa et al., 1993] (Figure 1.9). À cela une troisième catégorie peut être ajoutée dans laquelle le complexe hétérodimérique RXR-RAR est activé sous conditions par un ligand agoniste de RXR. Normalement, dans un complexe non permissif, l'activité de RXR reste silencieuse. Ainsi dans ce contexte, seul le ligand du partenaire peut

activer le récepteur hétérodimérique et induire une activité transcriptionnelle. Dans une telle situation, l'interaction avec le CoR ne peut être interrompue que lorsque le partenaire du RXR est en complexe avec un ligand agoniste [Lammi et al., 2008]. Concernant, les partenaires permissifs, la présence de l'un des deux agonistes, ou des deux, va induire la transcription. La présence des deux agonistes se traduira par une réponse synergique [Leblanc and Stunnenberg, 1995]. De même que les partenaires non permissifs, le CoR sera relaché après liaison d'un ligand agoniste [Lammi et al., 2008] (Figure 1.9).

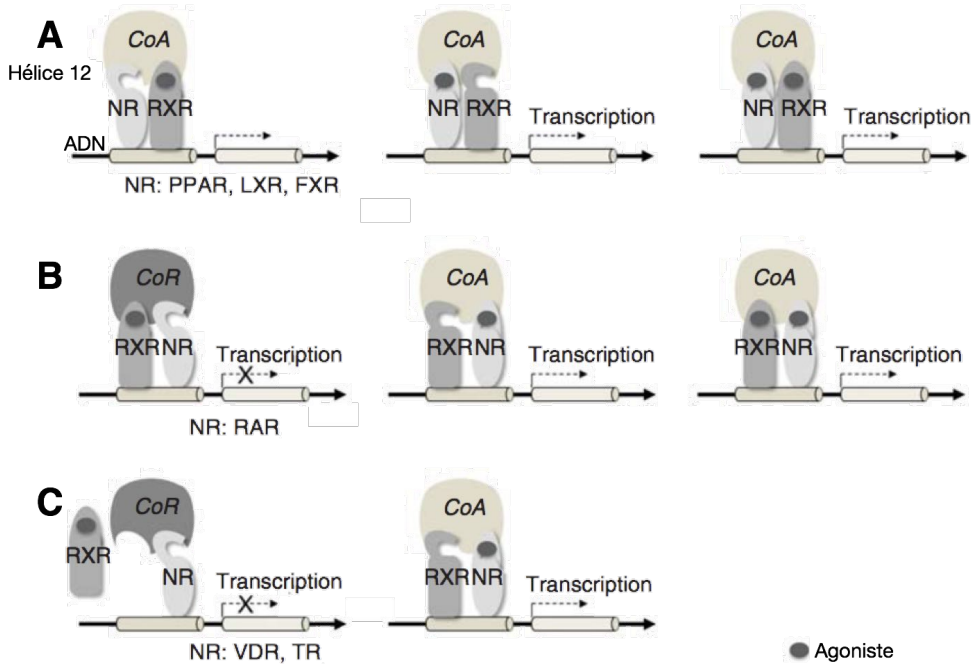


FIGURE 1.9 – Complexes hétérodimériques (A) permissifs, (B) conditionnels et (C) non-permissifs. Dans les complexes non-permissifs, tels que RXR-TR, seule la présence d'un ligand agoniste du TR va permettre d'activer le complexe et également permettre la liaison du ligand et dans le cas de RXR-RAR, sous certaines conditions, le complexe peut être activé par l'un ou par l'autre. La présence des deux ligands va activer le complexe de manière synergique. Dans le cas des complexes permissifs, tel que PPAR-RXR, le complexe peut être activé par un agoniste de RXR ou de PPAR. Adapté de [Yamada and Kakuta, 2014].

Les complexes hétérodimériques permissifs, tels que RXR-PPAR, sont activés aussi bien par un agoniste du RXR que par un agoniste du partenaire. Dans ces complexes, les agonistes de RXR peuvent montrer une activité pharmacologique qui reflète le rôle du partenaire, par exemple la sensibilité à l'insuline de PPARγ, qui est la cible de molécules à action antidiabétique [Lehmann et al., 1995] [Mukherjee et al., 1997]. De même pour le complexe RXR-LXR activé par les oxystéroles et impliqué dans l'homéostasie

du glucose, du cholestérol et des acides gras [Lehmann et al., 1997]. Ainsi les ligands agonistes de RXR peuvent être utilisés pour viser plusieurs cibles thérapeutiques [Altucci et al., 2007] [Desvergne, 2007].

En contraste, les partenaires non permissifs, tels que TR, VDR et RAR, fonctionnent principalement comme des récepteurs aux hormones et ont des ligands endocrines sous contrôle plus restreint. Ainsi en mettant sous silence l'activité de RXR, l'activité transcriptionnelle est directement proportionnelle à la concentration en hormones dans le corps [Shulman and Mangelsdorf, 2005]. Le mécanisme de synergie avec RAR et le rôle de subordination des ligands de RXR dans le complexe RXR-RAR est un exemple bien connu [Germain et al., 2002]. Lorsque RAR ne lie pas de ligand dans le complexe RXR-RAR, RXR peut lier des ligands de manière indépendante et générer une surface d'activation pour recruter des CoA [Széles et al., 2010]. Cependant dans un environnement cellulaire « normal », il ne peut y avoir dissociation du CoR et de fait le CoA ne peut être recruté, car la liaison au complexe hétérodimérique est mutuellement exclusive. Sachant que chaque sous-unité peut lier indépendamment un ligand, ceux-ci peuvent agir en synergie lorsqu'un CoA interagit avec les deux sous-unités en possédant deux motifs LXXLL. La subordination de RXR est due à l'inhabilité des ligands agonistes à induire la liaison d'un CoA dans un environnement cellulaire « normal » en absence de ligands agonistes à RAR. Un ligand agoniste de RXR ne peut activer le complexe RXR-RAR que lorsque le CoR interagit faiblement avec RAR ou que la concentration en CoA est nettement supérieure à celle du CoR. Dans le cas contraire, les agonistes inverses vont renforcer l'interaction du CoR avec RAR et vont totalement empêcher l'activation du complexe RXR-RAR par les ligands agonistes de RXR, alors que les antagonistes vont empêcher l'interaction avec soit les CoA ou bien les CoR [Germain et al., 2002].

Dans la plupart des cas, le partenaire hétérodimérique de RXR ne montre aucune préférence significative pour un des trois isoformes, ils sont tous les trois interchangeables comme partenaire hétérodimérique et comme récepteur pour 9cRA [Mangelsdorf et al., 1992]. Une exception est le complexe RXR-VDR lié au promoteur du gène codant la 24-hydrolase chez le rat, avec une préférence pour RXRα ou RXRγ plutôt que RXRβ [Kephart et al., 1996].

1.4 Ligands

RXR fut considéré comme orphelin à ses débuts, du fait de l'absence de ligand endogène connu, l'acide rétinoïque 9-*cis* (9cRA) fut le premier ligand agoniste à être identifié [Heyman et al., 1992], mais son statut en tant que ligand agoniste endogène reste controversé [Wolf, 2006] [Calléja et al., 2006]. D'autres

ligands agonistes endogènes, tel que l'acide docosahexaénoïque (DHA :2) ont été reportés [Egea et al., 2002]. L'acide phyténique [Kitareewan et al., 1996], l'acide phytanique [Lemotte et al., 1996], l'acide méthoprénique [Harmon et al., 1995], honokiol [Kotani et al., 2010] et le danthron [Zhang et al., 2011b] sont d'autres ligands naturels se liant à RXR (Figure 1.10). Le fait que RXR lie toute une série de ligands de type acide gras suggère fortement que RXR est impliqué dans la régulation des lipides à travers un mécanisme de rétrocontrôle en s'associant à d'autres NR, tels que PPAR et FXR. Ces composés chimiques apparentés à la vitamine A et se liant exclusivement à RXR et non à RAR sont-ce qu'on appelle des réxinoïdes.

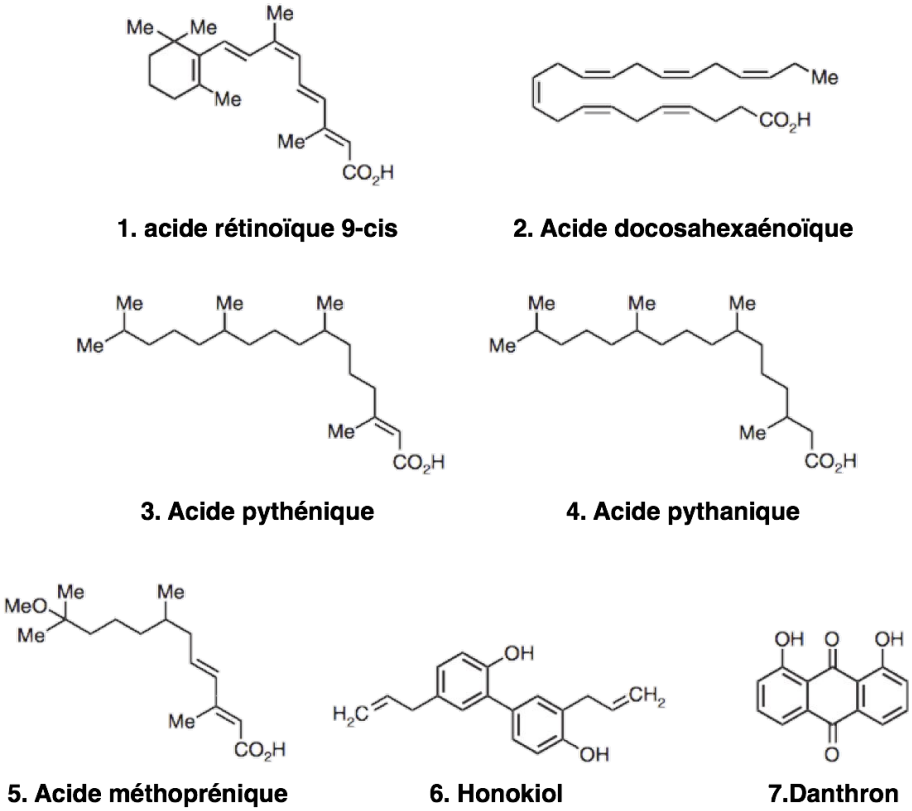
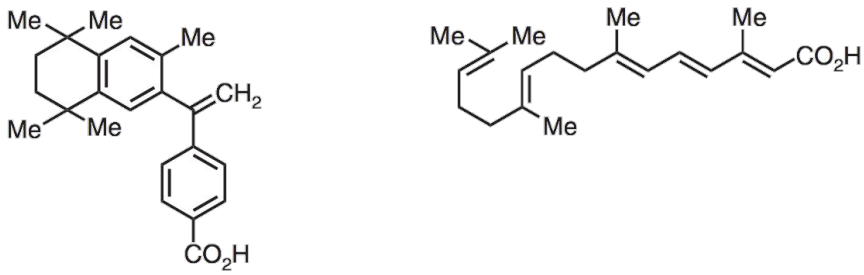


FIGURE 1.10 – Structures chimiques de quelques ligands naturels de RXR. Adapté de [Yamada and Kakuta, 2014].

Ainsi un certain nombre de ligands synthétiques ont été développés ayant des activités agonistes ou antagonistes. Ils présentent divers effets physiologiques qui dépendent de la nature du partenaire dimérique. Par exemple, LGD1069 ou le bexarotène (Targretin ©) (Figure 1.11), un ligand agoniste, est utilisé pour traiter les lymphomes cutanés des cellules T [Boehm et al., 1995]. Son utilisation semble également

prometteuse dans le traitement de la maladie d'Alzheimer, bien que le mécanisme sous-jacent ne soit pas encore clairement établi [Cramer et al., 2012] [Tesseur and De Strooper, 2013] [Pierrot et al., 2015] [Cummings et al., 2016]. L'acide rétinoïque acyclique (Figure 1.11), un ligand agoniste utilisé pour bloquer la carcinogenèse hépatocellulaire [Matsushima-Nishiwaki et al., 2003] [Shimizu et al., 2009]. Dans les cellules de type carcinome hépatocellulaire, RXRα est phosphorylé en position Ser260 et présente une perte de fonction [Matsushima-Nishiwaki et al., 2001] [Yoshimura et al., 2007]. L'acide rétinoïque acyclique va inhiber la phosphorylation de RXRα via l'inhibition de la voie de signalisation MAP-kinase, afin de restaurer sa fonction physiologique. Les essais cliniques sur l'acide rétinoïque acyclique sont encourageants [Honda et al., 2013]. Cependant, bien que RXR soit une cible d'attractive pour le développement de nouveaux ligands agonistes à RXR, l'activation de RXR peut causer un certain nombre d'effets secondaires non négligeables, tels qu'une élévation du taux de triglycérides dans le sang [Li et al., 2005b] [Vu-Dac et al., 1998], une hépatomégalie (augmentation de la taille du foie) [Lenhard et al., 2000] et une hypothyroïdie [Sherman et al., 1999] [Liu et al., 2002]. À cause de ces effets secondaires, les applications cliniques sur les ligands agonistes du RXR, autres que ceux impliqués dans le traitement des lymphomes cutanés des cellules T, ont largement été abandonnées. Ainsi seuls deux composés, le bexarotène et 9cRA, ont été approuvés et sont limités dans leur utilisation.



1. LGD1069 (Bexarotène) **2. Acide rétinoïque acyclique**

FIGURE 1.11 – Structures chimiques des ligands agonistes LGD1069 (Bexarotène) et acide rétinoïque acyclique.

D'autres pistes ont été explorées impliquant le développement de ligands antagonistes à RXR. Le mécanisme d'inactivation n'est pas différent de celui décrit chez d'autres récepteurs nucléaires. La nature antagoniste d'un réxinoïde est due à la présence d'un encombrement stérique provoqué par un des groupements du réxinoïde qui va déstabiliser H12 et affaiblir l'interaction avec les CoA [Nahoum et al., 2007]

[Santin et al., 2009] [Sato et al., 2010]. Cet encombrement stérique peut être médié au travers du résidu du résidu L451 dans H12, tel que LG100754 (Figure 1.12) [Sato et al., 2010]. Ainsi contrairement aux ligands agonistes, les ligands antagonistes ne peuvent pas déplacer l'équilibre conformationnel de RXRα, d'une forme inactive vers une forme plus compacte au niveau de H12. D'un point de vue fonctionnel, le développement d'antagonistes, tel que UVI3003 (Figure 1.12), peut permettre de tester la contribution de RXR au sein d'un complexe hétérodimérique dans un contexte de transactivation [Nahoum et al., 2007]. Ainsi que d'un point de vue pharmacologique, il a été montré que le développement de ligands antagonistes pouvait être utilisé comme antidiabétique, tels que HX531 (Figure 1.12) qui a pour effet une amélioration de la résistance à la leptine, hormone responsable du sentiment de satiété, sans en augmenter sa concentration dans le plasma chez la souris KK-A^y, un animal modèle [Yotsumoto et al., 2005].

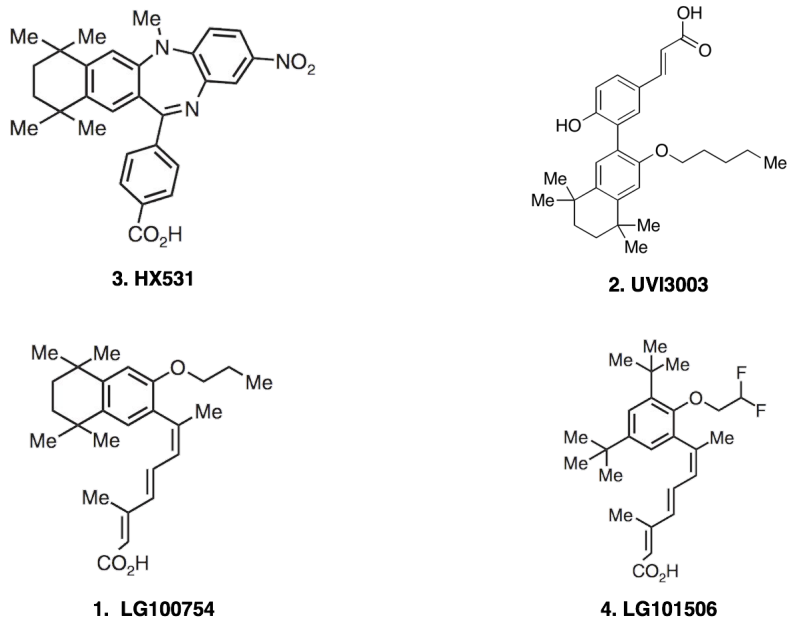


FIGURE 1.12 – Structures chimiques des ligands HX531 et UVI3003 présentant une activité antagoniste, LG100754 exhibant une activité agoniste ou antagoniste selon le partenaire hétérodimérique et LG101506, semblable à LG100754, mais présentant une activité agoniste partielle avec l'homodimère RXR.

Cependant LG100754 ne montre qu'une activité antagoniste envers les complexes RXR-RXR [Canan Koch et al., 1996], RXR-LXR et RXR-FXR [Lala et al., 1996] et non RXR-RAR [Sato et al., 2010] et RXR-PPAR [Lala et al., 1996]. Ce composé appartient à une classe de ligands dont la particularité est d'exhiber une sélectivité en fonction du partenaire hétérodimérique de RXR. Ces modulateurs de RXR

peuvent agir soit de manière agoniste, soit de manière antagoniste avec selon le partenaire. Dans le cas de LG101506 (Figure 1.12), appartenant à la même classe que LG100754 et activant de même le complexe RXR-PPAR, celui-ci exhibe une activité agoniste partielle au sein de l'homodimère RXR [Leibowitz et al., 2006]. Ces deux composés ont aussi une action antidiabétique sans impacter le fonctionnement de la glande thyroïdienne, comme c'est le cas des ligands agonistes complets de RXR, tel que le bexarotène. Ainsi la découverte de tels composés modulant spécifiquement l'activité des NR a permis un regain d'intérêt pour le développement de nouveaux ligands ciblant RXR.

Tous les résidus appartenant à H3, H5, H7, H11 et au feuillet β, constituant la poche de liaison au ligand, sont conservés entre les trois isotypes (α, β et γ), compliquant de fait le développement de réxinoïdes ciblant un isotype spécifiquement. Néanmoins, malgré la conservation des résidus en contact avec le ligand, les résidus de la seconde couche peuvent affecter la flexibilité de la poche. Ainsi, cette différence de flexibilité peut être exploitée afin de développer des composés se liant spécifiquement à un isotype donné [Egea et al., 2002]. De tels ligands ont été reportés dans la littérature [Pérez et al., 2012] [Dawson and Xia, 2012], tel que le tamibarotene présentant une préférence pour les isotypes α et β comparé à l'isotype γ [Fukasawa et al., 2012]. Malheureusement, ces composés ne peuvent être utilisés du fait encore d'un manque de spécificité.

1.5 Dynamique structurale du LBD au sein des NR

La toute première structure d'un LBD de RN obtenue a été celle de la forme libre (apo) du LBD de RXRα [Bourguet et al., 1995]. S'en est suivi les structures des formes liées (holo) des LBD de RARγ en complexe avec l'acide rétinoïque tout-trans (AtRA) ainsi que TRβ lié à T3 [Renaud et al., 1995] [Wagner et al., 1995]. La différence la plus marquante entre la forme libre et la forme liée se situe au niveau de la partie C-terminale du domaine, contenant l'hélice H12. Dans la forme libre, cette dernière se présente dans une conformation étendue, exposant la poche de liaison au ligand au solvant. Cependant dans la forme liée de RARγ et TRβ, H12 est plaquée contre le domaine, enfouissant complètement le ligand [Bourguet et al., 1995] [Renaud et al., 1995] [Wagner et al., 1995].

Cette différence suggère immédiatement un mécanisme d'entrée et de sortie du ligand de la poche de liaison, dans lequel H12 fonctionnerait comme une porte. Ce mécanisme d'activation se comporte comme un piège à souris, où la partie N-terminale de H12 poussée par le ligand va se repositionner dans la continuité de l'hélice H10. Finalement, le restant de H12 viendrait fermer comme un couvercle la poche de

liaison au ligand et former une interface d'interaction pour la fixation de la protéine CoA (Figure 1.13).

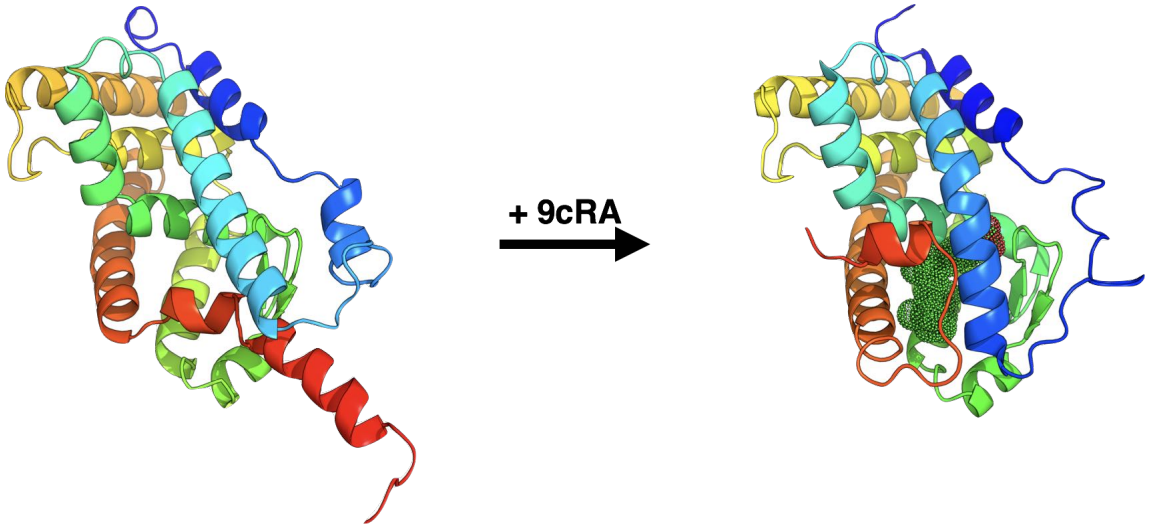


FIGURE 1.13 – Le mécanisme de « piège à souris » est basé sur la comparaison entre la forme inactive et la forme active en complexe avec 9cRA du domaine de liaison au ligand (LBD) de RXRα (PBD ID 1LBD et 1FM9). La présence du ligand agoniste, 9cRA, va induire un large changement conformationnel de la position l'hélice H12. H12 va bouger d'une conformation « inactive », étendue à l'extérieur du domaine, vers une conformation active repliée contre le LBD et formant avec H3 et H4 une surface d'interaction pour les protéines corégulatrices.

Bien que très attirant, ce modèle basé sur le repositionnement de H12 est très probablement inexact : la structure cristallographique de la forme inactive du LBD de RXRα n'était pas un monomère, mais un homotétramère, avec H12 interagissant avec la surface du coactivateur de l'unité voisine [Gampe et al., 2000a]. De nombreuses questions restent en suspens concernant le rôle physiologique du tétramère au sein de la cellule, soit en tant que forme de stockage ou bien en tant que mécanisme d'autorepression [Kersten et al., 1995a] [Kersten et al., 1995c] [Kersten et al., 1995b] [Kersten et al., 1997] [Kersten et al., 1998] [Mangelsdorf et al., 1991]. De plus, d'autres structures cristallographiques de LBD de RN en conformations inactives ont été obtenu dans lesquelles H12 n'est pas conformation étendue. C'est le cas par exemple pour PPARγ, PXR et les RN orphelins comme NURR-1 [Nolte et al., 1998] [Watkins et al., 2003] [Wang et al., 2003].

1.5.1 Étude de la stabilité par protéolyse

De nombreuses études ont montré que de manière générale le LBD des RN est plus flexible en absence de ligand. Mais ces résultats n'impliquent pas nécessairement que cette flexibilité soit liée au détachement de H12 du restant du LBD. Cette flexibilité intrinsèquement liée aux LBD des RN a été mise en évidence, bien avant l'obtention des premières structures cristallographiques, par des expériences de protéolyse en absence et en présence d'un ligand agoniste. Ces expériences ont permis de montrer que la présence d'un ligand agoniste avait un effet protecteur contre la protéolyse en comparaison à la forme non liée des RN telles que PR, ER, PPARγ ainsi que sur RARγ et RXRα [Allan et al., 1992] [McDonnell et al., 1992] [Berger et al., 1996] [Leng et al., 1995] [Leid et al., 1992]. Il a aussi été remarqué que la présence d'un ligand antagoniste rendait plus sensible la partie C-terminale, contenant H12, à la protéolyse chez PR [Allan et al., 1992]. Ainsi tous ces éléments, ont pu laisser supposer que la présence d'un ligand devait entrainer des changements structuraux significatifs en rendant le LBD des RN plus stable et plus compact dans le cas d'un agoniste.

1.5.2 Étude de la stabilité par Résonance Magnétique Nucléaire (RMN)

Des données RMN sur le LBD de PPARγ vont dans le même sens que les données de protéolyse sur l'existence de changements conformationnels survenant en présence d'un ligand agoniste. Sur l'ensemble des 265 pics de corrélation attendus pour la forme libre, seuls, 124 pics peuvent être attribués, correspondant au squelette peptidique, alors que 270 pics sont observés dans le cas où le LBD de PPARγ est lié avec la rosiglitazone, un ligand agoniste complet [Johnson et al., 2000]. L'absence de ces pics dans la forme libre a été attribué à la dynamique conformationnel sur une échelle de temps μ-/ms. De plus, plusieurs pics sont observés pour certains résidus de la forme libre, renforçant l'idée que le LBD de PPARγ échantillonne plusieurs conformations distinctes. Les différences moyennes calculées entre les déplacements chimiques ^{1}H, ^{15}N et ^{13}C entre la forme libre et la forme liée, montrent que les régions les plus stables gardent la même conformation, avec une différence inférieure à 0.05 ppm. Concernant les résidus non détectables au sein de la forme libre, ils sont principalement localisés dans la partie du LBD où se situe la poche de liaison au ligand (LBP). Cette région est délimitée par les hélices H1, H3, H4, H5, H7, H10, H12 et le feuillet β. Ainsi les résultats obtenus montrent que cette région du LBD de PPARγ explore un nombre multiple de conformations et que les interactions formées avec le ligand ont pour effet de stabiliser le domaine, en accord avec les données expérimentales précédentes. Cette observation suggère un déplacement de l'équilibre conformationnel depuis une population de conformations libre vers une population plus restreinte de

conformations dans la forme liée, plutôt qu'un système à deux états comme le suppose le modèle basé uniquement sur les structures cristallographiques [Johnson et al., 2000]. Des résultats similaires ont été reportés pour le LBD de PPARα, ER et de VDR [Cronet et al., 2001] [Luck et al., 2000] [Singarapu et al., 2011]. Des études supplémentaires sur le LBD de PPARγ, cette fois en présence de ligands de natures différentes et aussi en complexe avec le LBD de RXRα en forme libre et liée avec 9cRA [Berger et al., 2003] [Lu et al., 2008] ont confirmé la présence d'équilibres conformationnels. Dans le cas de GW1929, un ligand agoniste complet, le profil RMN obtenu est similaire à celui avec la rosiglitazone. La présence du LBD de RXRα en forme libre et liée à 9cRA n'a qu'un effet modéré sur l'interface de dimérisation et sur l'hélice H12 du LBD de PPARγ [Lu et al., 2008]. Cependant, contrairement au rosiglitazone, nTZDpa, un agoniste partiel, présente un profil RMN à peine différent à la forme libre du LBD de PPARγ, bien que le nTZDpa protège le récepteur de la protéolyse comme la rosiglitazone [Berger et al., 2003]. Ainsi bien que nTZDpa active PPARγ, il maintient le LBD de PPARγ dans un ensemble conformationnel plus vaste que ne le fait la rosiglitazone.

Les données obtenues pour le LBD de RXRα vont dans le même sens que celles observées avec les autres NR étudiés [Lu et al., 2006]. Seuls 164 des 240 résidus au total peuvent être attribués dans la forme libre, avec l'apparition de 47 résidus en présence de 9cRA. La majorité de ces résidus sont localisés dans les éléments structuraux formant la LBP, de même au niveau de H9, suggérant un effet allostérique (Figure 1.14). Cependant, contrairement à PPARγ et PPARα, les hélices H7, H10 et H11 ne sont toujours pas attribuables en présence de 9cRA [Johnson et al., 2000] [Cronet et al., 2001]. De plus et contrairement, à PPARγ qui est majoritairement sous forme monomérique en solution [Johnson et al., 2000], RXRα est principalement sous forme d'homodimère [Lu et al., 2006]. Ainsi du fait qu'ils participent à l'interface de dimérisation au sein du complexe homodimérique, l'absence de pics dans les hélices H7, H10, H11 est très vraisemblablement lié à un phénomène de dissociation/association entre sous-unités.

En plus de fournir des éléments à la compréhension sur la dynamique du LBD de RXRα, les déplacements chimiques RMN apportent des informations plus précises que les expériences de protéolyse quant à la nature des structures secondaires en solution. Ces données montrent l'absence d'hélice H2 dans la boucle H1-H3 des formes libre et liée du LBD de RXRα, par exemple [Lu et al., 2006]. L'étude du temps de relaxation montre aussi aussi que la dynamique de cette région est moins restreinte que le restant de la protéine sur l'échelle du temps allant de la ps à la ns dans la forme libre, et que l'ajout de 9cRA semble la rigidifier [Lu et al., 2006]. Et bien que le modèle basé sur les structures cristallographiques suppose le repositionnement de H12 dans une conformation moins mobile, en présence de 9cRA, les expériences

RMN ne mettent pas cet effet en évidence, contrairement aux observations en solution du LBD de PPARγ où la mobilité de H12 est plus restreinte [Kallenberger et al., 2003].

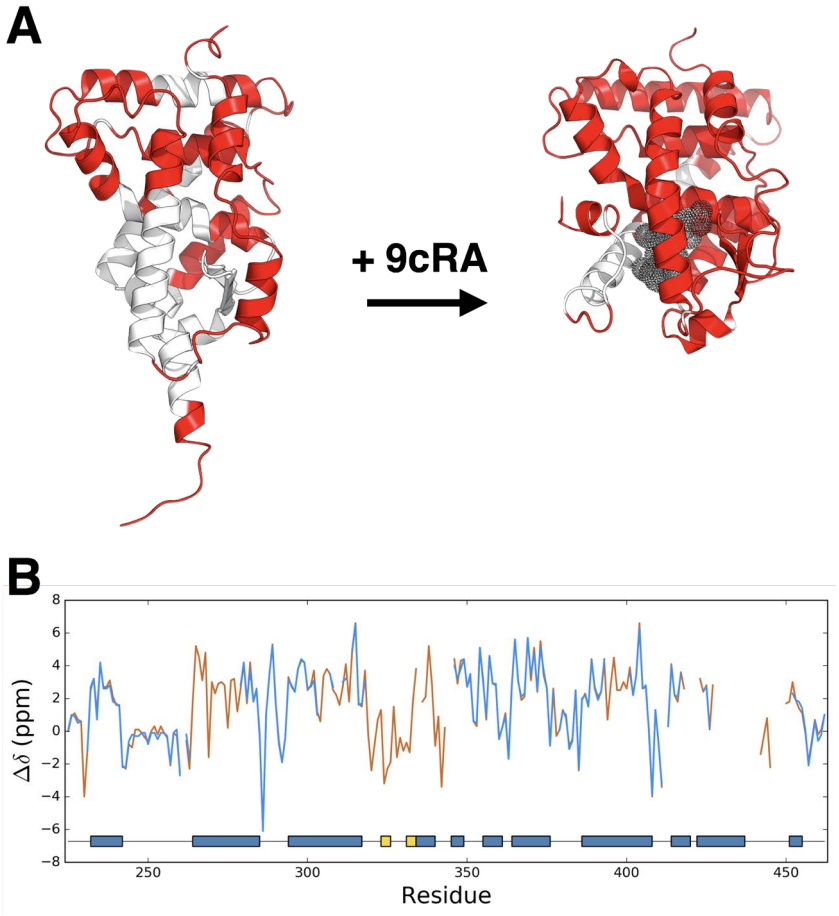


FIGURE 1.14 – (A) Structures cristallographiques de la forme inactive (à gauche) et active, en complexe avec 9cRA (à droite), du domaine de liaison au ligand (LBD) de RXRα avec en rouge les résidus pour lesquels un pic a été observé et assigné pour l'atome ^{13}Cα et en blanc, les résidus pour lesquels aucune information n'est disponible. (B) Déplacements chimiques RMN secondaires $\Delta\delta$ ^{13}Cα par résidu de la forme libre et liée, en complexe avec 9cRA, du LBD de RXRα, obtenus expérimentalement par l'équipe d'Ellen Li [Lu et al., 2006]. En bleu et en orange, les déplacements chimiques RMN secondaires pour la forme libre et liée, en complexe avec 9cRA, du LBD de RXRα, respectivement.

Il faut noter que le rosiglitazone interagit directement avec l'hélice H12, ce qui n'est pas le cas avec 9cRA au sein des structures cristallographiques des formes liées des LBD de PPARγ et RXRα, respectivement. Mais de manière générale, malgré la présence de 9cRA, le LBD de RXRα garde une certaine

dynamique conformationnele d'après les paramètres de relaxation, signifiant que la forme liée existe encore dans un ensemble dynamique de conformations. Globalement, ces résultats indiquent que la liaison d'un ligand ne déplace pas simplement la conformation de la protéine d'un état bien défini à un autre. À la place, le ligand stabilise la protéine existant dans un ensemble large de conformations vers une population plus réduite de conformations différentes.

L'étude de l'effet de ligands antagonistes sur le complexe homodimère LBD RXRα montre que ceux-ci perturbent principalement la dynamique de H3 et de H12, constituant l'interface d'interaction avec les peptides CoA [Lu et al., 2009]. Cependant, la présence d'un ligand antagoniste n'abolit pas complètement l'interaction avec les peptides CoA. Les peptides CoA peuvent toujours interagir avec le LBD de RXRα, mais plus faiblement qu'en présence d'un ligand agoniste. L'interaction est complètement abolie entre les peptides CoA et le LBD de RXRα, que lorsque celui-ci est en complexe avec PPARγ lié à un ligand agoniste [Lu et al., 2009]. Contrairement au LBD d'autres NR, celui de NURR-1, considéré comme orphelin du fait de l'absence de ligand connu se liant à ce dernier, montre un profil RMN différent de ceux observés précédemment [Michiels et al., 2010]. Dans la forme libre des LBD des autres NR, la majorité des résidus situés autour de la poche de liaison du ligand ne sont pas attribuables alors que la forme libre du LBD de NURR-1 présente des résidus non attribuables principalement au niveau de la partie C-terminale de H6 jusqu'au N-terminal de H7, dans la boucle H8/H9 et la région C-terminale de H11. Cette observation tend à montrer que la forme libre du LBD de NURR-1 est stable même en absence de ligand, ce qui explique la fonction activatrice ligand-indépendante de ce récepteur [Wang et al., 2003].

1.5.3 Étude de la stabilité par Échange Hydrogène/Deutérium (HDX)

L'échange Hydrogène/Deutérium par Spectrométrie de masse (HDXMS) est une autre méthode très utilisée afin d'étudier la dynamique conformationnelle des protéines, l'incorporation en deutérium étant directement liée à la stabilité des structures secondaires. Elle permet aussi de suivre directement les changements de stabilité dus à la présence d'un ligand ou d'interaction spécifique entre plusieurs partenaires. Elle fut appliquée pour la première fois au LBD de RXRα [Yan et al., 2004] [Yan et al., 2006] [Yan et al., 2007] [Xia et al., 2010] [Boerma et al., 2014] et par la suite à de nombreux autres récepteurs nucléaires, tels que TR [Figueira et al., 2011], GR [Frego and Davidson, 2006], PPARγ [Chalmers et al., 2006] [Hamuro et al., 2006] [Chalmers et al., 2007] [Bruning et al., 2007] [Hughes et al., 2012] [Amato et al., 2012] [Marciano et al., 2015], VDR [Zhang et al., 2010] [Carson et al., 2014] [Kato et al., 2016], PR [Goswami et al., 2014], FXR [Yang et al., 2014] et ER [Dai et al., 2008] [Dai et al., 2009].

Dans le cas de PPARγ, l'échange a permis de montrer un mécanisme d'activation graduel, où différents modes de liaison peuvent être observés entre les ligands agonistes complets et partiels [Chalmers et al., 2006] [Hamuro et al., 2006] [Chalmers et al., 2007] [Bruning et al., 2007]. En comparaison à la forme non liée, les agonistes complets réduisent très fortement l'échange HX au sein du domaine, au niveau des régions des régions formant la LBP et plus particulièrement au niveau de H12, en interagissant avec cette dernière [Hamuro et al., 2006]. Cette réduction de l'échange HX indique qu'ils stabilisent le LBD de PPARγ. Cependant, contrairement aux ligands agonistes complets, les ligands agonistes partiels, présentant un degré de transactivation moindre, montrent un profil d'échange HX différent. Aucune différence significative n'est observée au niveau de H12, en comparaison avec la forme non liée. Par contre ces ligands induisent une réduction plus importante de l'échange HX au niveau du feuillet β, en comparaison avec la forme liée avec un agoniste complet [Bruning et al., 2007]. Ce résultat souligne l'importance du rôle de la région contenant le feuillet β dans l'activation de PPARγ. Tout comme le LBD de PPARγ, les expériences d'échange HD montrent aussi une stabilisation globale du LBD de TR, et non juste des régions impliquées dans la LBP, que ce soit avec un ligand agoniste (T_3) ou un ligand antagoniste (NH_3). Cependant, les résultats montrent un mécanisme d'activation différent de TR au niveau de la partie C-terminale du domaine, comparé aux autres NR. H12 interagit avec la surface d'interaction aux CoA/CoR, formée par H3, H4 et H5, ce qui stabilise la partie C-terminale de H3. La liaison avec T_3 déstabiliserait la partie C-terminale de H3 forçant H12 à adopter une conformation active et libérant l'interface d'interaction aux CoA/CoR. NH_3 étant incapable de déstabiliser H_3, H12 continuerait à interagir avec le domaine au niveau de la surface d'interaction aux CoA/CoR, empêchant ces derniers d'interagir avec le LBD. De plus, T_3 stabiliserait la surface de dimérisation empêchant la formation d'homodimère TR, ce qui n'est pas le cas de NH_3. De même que PPARγ et TR, on observe aussi une diminution de l'échange HD au sein des LBD de PR, FXR, VDR, ERα/β lorsque ceux-ci sont liés à un ligand agoniste, en comparaison à la forme libre. Dans le cas de GR, aucune comparaison n'est possible, car la forme libre n'est pas soluble [Frego and Davidson, 2006]. De tous les LBD étudiés par HDXMS, les principales différences se situent au niveau de la partie C-terminale des domaines. Aucun changement significatif n'est observé dans la partie C-terminale de ERα/β et de FXR entre la forme libre et la forme liée à un ligand agoniste alors qu'une plus grande protection contre l'échange HD est observée au sein de PPARγ, TR, VDR comparé à la forme libre. De même que l'on observe le même phénomène chez GR avec une plus grande protection au sein de la forme liée à un ligand agoniste comparé à celle liée à un ligand antagoniste.

Concernant le LBD de RXRα, les expériences d'échange HD montrent que les éléments structuraux

impliqués dans la LBP incorporent plus rapidement que les éléments formant le coeur du sandwich d'hélices [Yan et al., 2004]. Ainsi la partie N-terminale (résidus 219-233), la boucle H1-H3 (241-262), une part importante de H3 (263-279), la région comprenant le feuillet β, H6 et la partie N-terminale de H7 (320-346) et enfin toute la partie C-terminale avec H11 et H12 (433-449) échangent très rapidement, avec plus de 85 % des protons amides HN échangés au bout de 15 secondes. Dans le cas du coeur du sandwich d'hélices, composé de H1, la partie C-terminale de H3, H4, H5, H7 (excepté la partie N-terminale), H8, H9 et H10, l'échange est nettement plus lent, avec moins de 35 % des protons amides HN échangés au bout de 15 secondes. Ces éléments de structure secondaire ne sont pas impliqués dans la liaison au ligand au sein du LBD de RXRα, contrairement à la plupart des premiers éléments structuraux qui échangent très rapidement, signe d'une dynamique conformationnelle très importante au niveau de la LBP au sein de la forme libre (Figures 1.15, 1.16 et 1.17). De même que pour les autres NR, la présence de 9cRA va induire une stabilisation du domaine. En effet, la vitesse d'incorporation en deutérium est significativement plus lente en présence du ligand au niveau de H1, de la moitié N-terminale de H3, la région du feuillet β avec la partie C-terminale de H5, H6 et la partie N-terminale de H7 et enfin H11. Cependant, certaines régions ne montrent pas de différences significatives entre la forme libre et la forme liée. Ces régions incluent la partie N-terminale du domaine, la partie C-terminale de H3, H4, la partie N-terminale de H5, H8, H9, la partie N-terminale de H10 et la partie C-terminale de domaine comprenant H12. Ainsi ce sont les régions impliquées dans la liaison au ligand qui montrent une réduction significative de l'échange HD, signifiant que celles-ci deviennent plus rigides en présence de 9cRA (Figures 1.15 et 1.16). Cependant, il est important de remarquer que 9cRA n'a aucune influence sur la stabilité de la partie C-terminale du domaine, incluant H12, que ce soit au sein de la forme libre ou liée, on observe un échange HD très rapide. Ce comportement est aussi observé au sein de ERα/β et FXR.

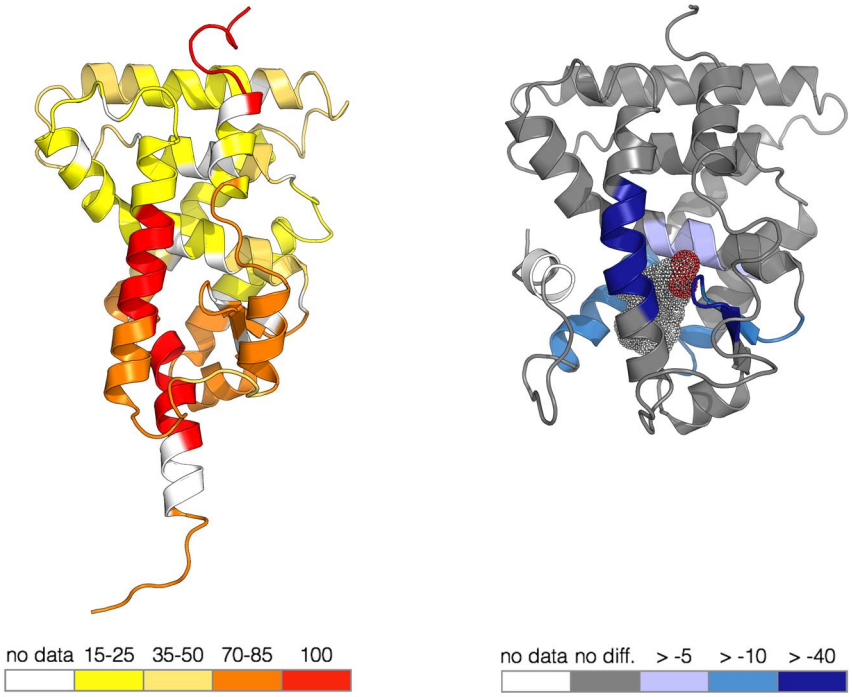


FIGURE 1.15 – Taux d'incorporation en deutérium (exprimé en %) pour la forme libre du LBD de RXRα et taux de protection (exprimé en % par rapport à la forme libre), moyenné à partir de plusieurs peptides, pour la forme liée à 9cRA, observés expérimentalement à 15 secondes par l'équipe de Max L. Deinzer [Yan et al., 2004]. À gauche, le taux d'incorporation en deutérium est représenté par un gradient de couleur allant du jaune pour les régions ayant échangé entre 15 et 25 % au rouge pour les régions ayant totalement échangé. En blanc, les régions pour lesquelles aucune information n'est disponible. À droite, le taux de protection, comparé à la forme libre, de la forme liée du LBD de RXRα en complexe avec 9cRA à 15 secondes. Le taux de protection est représenté par un gradient de couleur allant du bleu clair, pour les régions présentant une légère protection contre l'incorporation en deutérium (entre 5 et 10%), au bleu foncé, pour les régions présentant une forte protection (plus de 40%). En gris, les régions qui ne présentent aucune différence significative (< 5%) entre la forme libre et la forme liée, et en blanc les régions pour lesquelles aucune information n'est disponible.

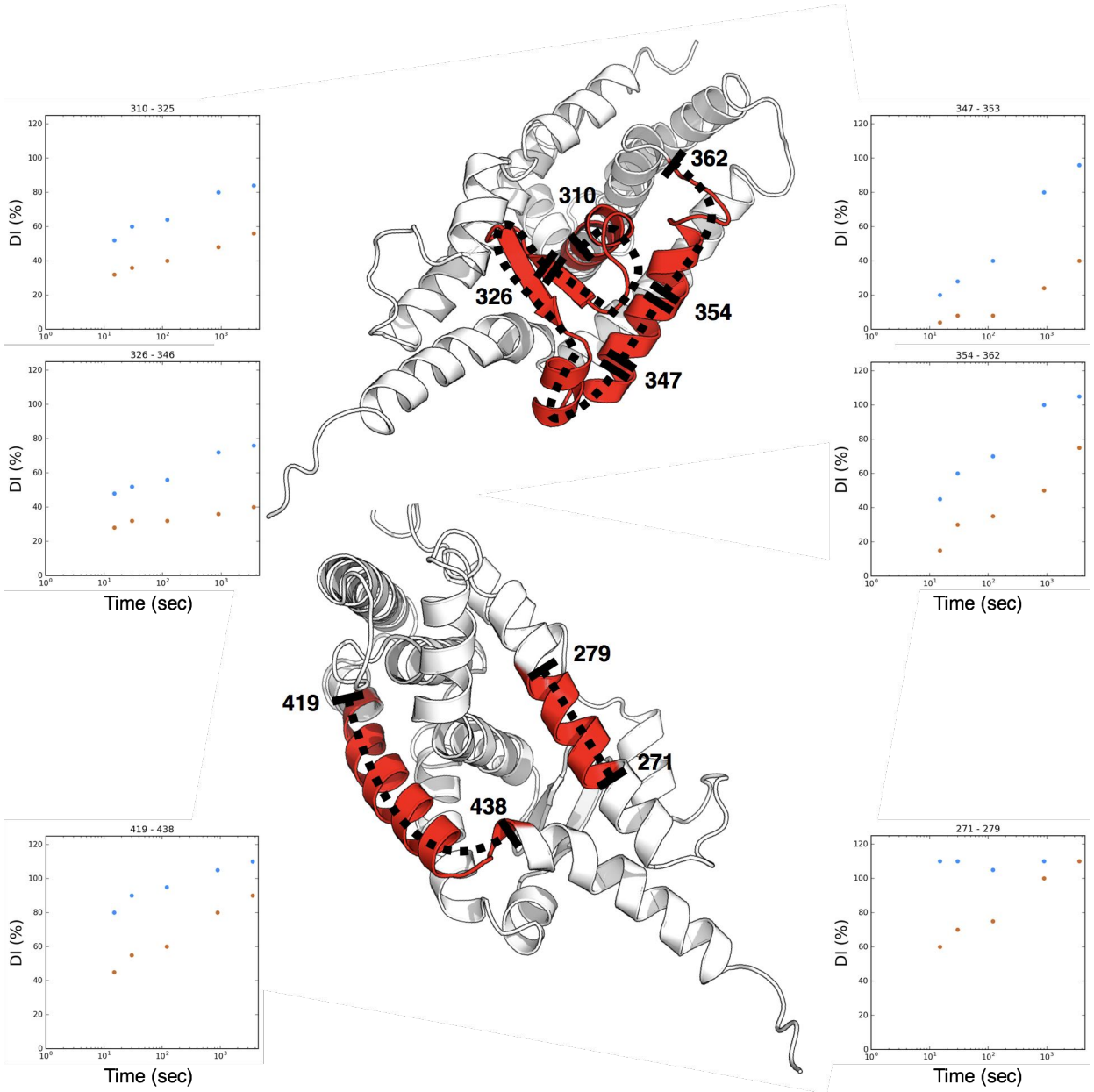


FIGURE 1.16 – Principales différences en termes d'incorporation en deutérium par peptide (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) au sein de la forme libre et de la forme liée à 9cRA du LBD de RXRα, observées expérimentalement par l'équipe de Donald D. Muccio [Xia et al., 2010]. Les régions représentées en rouge, sur la structure cristallographique de la forme libre, correspondent aux régions où il y a une différence significative entre la forme libre et liée en termes d'incorporation en deutérium au cours du temps, avec pour chaque région, le ou les peptides correspondants. En bleu, le taux d'incorporation en deutérium pour la forme libre et en orange pour la forme liée du LBD de RXRα à 15, 30, 120, 900 et 3600 secondes.

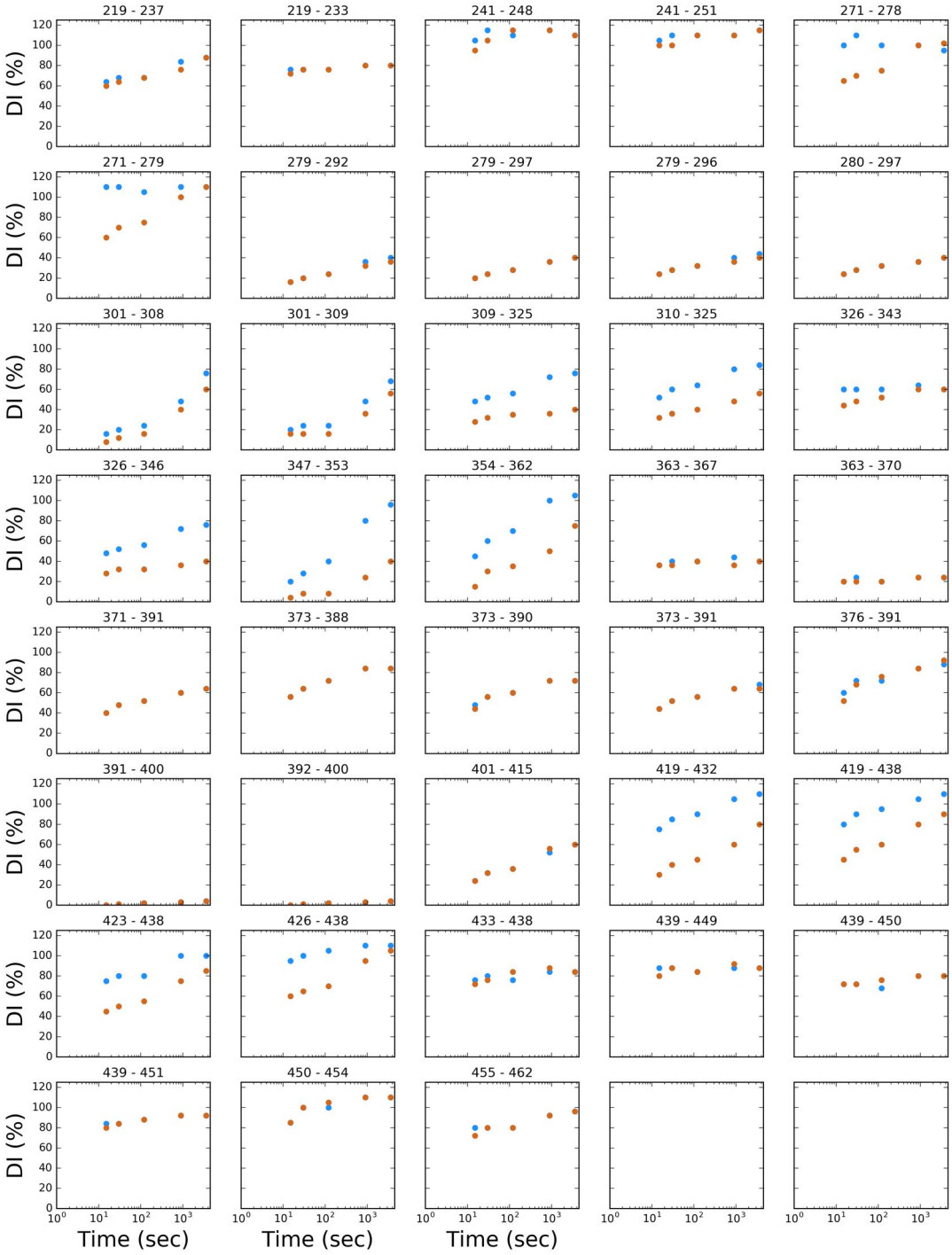


FIGURE 1.17 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme libre et la forme liée à 9cRA du LBD de RXRα, observé expérimentalement par l'équipe de Donald D. Muccio [Xia et al., 2010]. En bleu et en orange, le taux d'incorporation en deutérium à 15, 30, 90, 120 et 3600 secondes pour la forme libre et la forme liée à 9cRA du LBD de RXRα, respectivement.

D'autres études ont été menées afin d'examiner les effets de différents ligands agonistes et antagonistes sur la dynamique conformationnelle du LBD de RXRα. Les observations faites à partir des ligands agonistes montrent que l'affinité est corrélée à la vitesse d'échange HD des régions impliquées dans la LBP, les ligands les plus affins sont ceux qui réduisent le plus la vitesse d'échange HD [Yan et al., 2006]. De même, peu d'effets sont visibles sur la vitesse d'échange HD au niveau de H12, celle-ci échangeant toujours autant que dans la forme libre. Dans le cas des ligands antagonistes, il a été montré que ceux-ci offraient une plus grande protection contre l'incorporation en deutérium, comparé à 9cRA [Yan et al., 2007]. Ce phénomène est largement visible dans les éléments structuraux composant la LBP, du fait d"interactions hydrophobes plus importantes entre les ligands antagonistes et le LBD de RXRα. De même qu'aucun effet n'est visible sur H12 (< 5 %), qui continue à échanger à la même vitesse que dans la forme libre. Sur des temps très longs (> 60 min), le taux d'incorporation devient similaire entre les formes liées aux ligands antagonistes et aux ligands agonistes, et également similaires à l'échange de la forme libre. Ceci suggère que la dynamique conformationnelle du LBD de RXRα est un élément clé dans le mécanisme d'activation de la transcription. Une trop forte rigidité affecterait l'interaction avec d'autres protéines corégulatrices. La diminution de l'échange HD au niveau de la partie C-terminale du domaine n'intervient que lorsqu'un peptide CoA est présent [Xia et al., 2010]. En plus de stabiliser H12 et H11, la présence du peptide CoA va induire une réduction de l'échange HD au niveau de H7, H8 et H10, et aussi de manière évidente au niveau de la surface d'interaction avec les CoA, formée par H3 et H4. Cette réduction de la vitesse d'incorporation en deutérium est à la fois observée au sein de la forme libre et de la forme liée à 9cRA, en apportant une protection contre l'échange supplémentaire dans le dernier cas. D'autres analyses complémentaires en présence d'autres ligands agonistes avec un peptide CoA montrent le même profil d'échange HD, avec une réduction constatée au niveau de H3, de la région du feuillet β incluant H5, H6 et H7, ainsi que de H10 et H12 [Boerma et al., 2014]. Cette étude quant à elle montre que H12 est légèrement moins stable (-10 % de protection) dans les formes liées comparées à la forme libre.

La méthode HXMS a permis aussi de mettre en évidence l'action antagoniste du bexarotène chez PPARγ, jusqu'alors identifié comme ligand agoniste de RXR [Marciano et al., 2015]. Cette méthode peut être également utile dans une approche de « drug design » basée sur des fragments. Par exemple, dans le cas de VDR où il a pu être possible d'identifier des fragments, qui similairement à la vitamine D_3, stabilisaient de manières similaires les hélices H3, H7 et H8 [Carson et al., 2014].

1.5.4 Étude de la flexibilité de H12 par Anisotropie de Fluorescence

Parce que la dynamique de H12 a été reconnue comme étant importante dans la régulation de l'activité des RN, elle a été étudiée par de nombreuses techniques expérimentales. L'anisotropie par fluorescence est la méthode la plus couramment utilisée pour étudier la flexibilité de la partie C-terminale en fonction de la nature du ligand ou de son absence et aussi l'interaction entre le LBD et les peptides coactivateurs. Elle a été principalement utilisée dans le cas de ER, mais aussi de PPARγ et RXRα.

Dans le contexte de ER, des expériences d'anisotropie de fluorescence montrent que H12 est intrinsèquement mobile et que l'ajout de ligands agonistes et de peptides CoA va induire une rigidification du domaine, plus résistant à la protéolyse [Tamrazi et al., 2003]. Dans le cas du peptide CoA, et plus particulièrement de sa séquence, il a aussi été montré qu'il avait une influence sur la vitesse de dissociation du ligand chez ER, mais que son influence était négligeable sur les ligands de type antagonistes [Gee et al., 1999]. Ce qui montre une réciprocité entre le peptide CoA et le ligand, l'un favorisant l'interaction de l'autre avec le LBD, sachant que le peptide CoA peut à lui seul induire une conformation stable du LBD de ER en absence de ligand [Tamrazi et al., 2005]. Des expériences complémentaires de dénaturation suivies par anisotropie de fluorescence montrent que le LBD de ER se dénature en deux phases bien distinctes [Gee and Katzenellenbogen, 2001]. Une phase de dénaturation partielle du LBD de ER, qui peut être retardée lorsque ER est en présence d'un ligand agoniste, et une autre phase de dénaturation totale où tous les résidus TRP sont accessibles au solvant [Gee and Katzenellenbogen, 2001]. Utilisée dans le cadre de l'étude de la mobilité de H12 chez PPARγ, l'anisotropie de fluorescence a montré que la partie C-terminale du LBD était significativement plus mobile que le restant du domaine et qu'en présence d'un ligand agoniste, tel que la rosiglitazone, sa mobilité était réduite et que la partie C-terminale devenant solidaire du reste du domaine [Kallenberger et al., 2003]. Cette étude a permis aussi montrer le lien entre les multiples mutations impliquées dans le syndrome de résistance à l'insuline à la stabilité de H12, renforçant ainsi son rôle dans l'activation de la transcription.

De même dans le contexte du LBD de RXRα, des expériences similaires montrent une stabilisation de H12 en présence d'un ligand agoniste et au contraire une déstabilisation de cette dernière en présence de ligands antagonistes [Nahoum et al., 2007]. L'ajout de peptides CoA stabilise davantage H12, mais échoue en présence d'antagonistes. À noter que l'ajout d'une forte concentration de peptides CoA stabilise H12 en l'absence de ligand activateur [Nahoum et al., 2007]. Ceci montre que ces derniers peuvent à eux seuls déplacer l'équilibre conformationnel vers un état actif du LBD de RXRα, en absence de ligands. De manière générale, cela suppose que l'équilibre dynamique de H12 soit à l'origine de la différence observée

au niveau de l'activité transcriptionnelle basale des récepteurs nucléaires.

1.6 Modification post-traductionnelles

En plus d'être régulés par les ligands, les récepteurs nucléaires sont régulés par des modifications post-traductionnelles (PTM), qui vont impacter leur fonction. Ces PTMs incluent la phosphorylation, l'acétylation, la SUMOylation, l'ubiquitination, la méthylation, la myristylation, l'ADP-ribosylation et l'isoprénylation. Parmi ces PTMs, il est possible de distinguer deux catégories : (1) les modifications réversibles qui fonctionnent par l'ajout ou le retrait d'un groupement chimique (phosphate, acétyle) sur des acides aminés spécifiques (sérine, tyrosine, thréonine, lysine) et (2) des modifications qui fonctionnent par l'ajout d'un polypeptide (SUMO, ubiquitine) sur les résidus lysines. De nombreuses études ont montré que ces PTMs étaient liées au développement de pathologies incluant entre autres des cancers, le diabète et l'obésité [Anbalagan et al., 2012].

RXRα peut être acétylé au niveau de la Lys145 localisée dans le DBD par le complexe CBP/p300 (Figure 1.18). Cette acétylation favorise la liaison du complexe hétérodimérique RXR-TR sur l'ADN, induisant la transcription de gènes impliqués dans la prolifération cellulaire [Zhao et al., 2007]. RXRα peut également être SUMOylé au niveau de la Lys108 (Figure 1.18), localisée dans la partie C-terminal du domaine NTD, dont le motif consensus est IKPP [Choi et al., 2006]. Cette modification régule de manière négative l'activité transcriptionnelle de RXRα, et peut-être également celle des isoformes RXRβ et RXRγ, qui possèdent des motifs proches de celui trouvé dans RXRα.

Cependant, parmi toutes les PTMs existantes, la phosphorylation est celle qui a été la plus étudiée chez RXR et les autres NR [Rochette-Egly, 2003]. De manière générale, il a été montré que RXRα pouvait être phosphorylé par plusieurs membres de la famille Ras-MAPK (*Ras-Raf-Mitogen-Activated Protein Kinase*), comme ERK1/2, MKK4/SEK1 et JNK (*c-Jun NH_2-terminal Kinase*), à plusieurs endroits du domaine N-terminal, aux positions Ser22 (uniquement chez la souris), Ser27 et Ser32 (uniquement chez l'humain), Ser56/61 (humain/souris), Ser70/75 et Thr82/87, et dans le LBD, aux positions Tyr249/254 et Ser260/265 (Figure 1.18). La sérine 22 de RXRα chez la souris est phosphorylée par Cdk1, impliquée dans la régulation du cycle cellulaire [Adam-Stitah et al., 1999]. En utilisant une lignée cellulaire bien particulière pour l'étude de la voie de signalisation de l'acide rétinoïque, nommée F9 (carcinome embryonnaire) [Chambon and Rochette-Egly, 2001], il a été montré que cette phosphorylation était nécessaire chez RXRα pour l'activation de gènes cibles sous la dépendance de l'acide rétinoïque et pour l'action anti-proliférative de ce

dernier [Bastien et al., 2002]. Dans le cas de la sérine 27, cette dernière est phosphorylée par PKA (*Protein Kinase A*) dans la lignée cellulaire COS-1 et sa phosphorylation diminue l'activité transcriptionnelle de RXRα [Harish et al., 2000]. En réponse à l'atrioxyde d'arsenic (As_2O_3), on observe la phosphorylation de Ser32 par JNK [Mann et al., 2005].

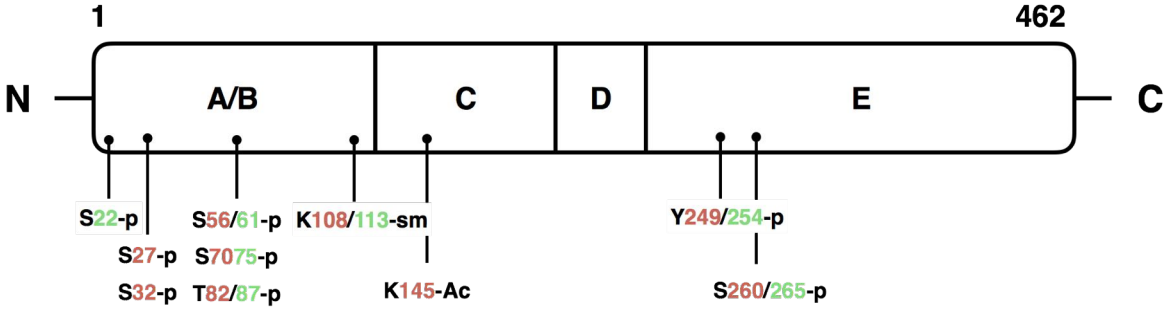


FIGURE 1.18 – Localisation des sites de modifications post-traductionnelles (PTM) dans le RXRα ayant un effet sur l'activité transcriptionnelle. En vert et rouge, les numéros des résidus subissant une modification PTM chez la souris et l'humain, respectivement. RXRα peut être phosphorylée dans la région N-terminale en position 22, chez la souris. Il peut être aussi hyperphosphorylé par des kinases à 4 positions différentes, 3 étant localisées dans la région N-terminale en position 56/61, 70/75 et 82/87, et dans le domaine de liaison au ligand (LBD) en position 260/265 chez l'humain ou la souris. De plus, une tyrosine localisée dans le LBD peut être aussi phosphorylée en position 249/254. RXRα est aussi la cible d'autres modifications PTMs telle que la SUMOylation en position 108/113 dans la région N-terminale et l'acétylation en position 145 chez l'homme dans le domaine de liaison à l'ADN.

De plus, on observe aussi une hyperphosphorylation chez RXRα au niveau de trois autres résidus situés dans le domaine N-terminal, aux positions Ser61, Ser75 et Thr87 chez la souris (positions Ser56, Ser70 et Thr82 chez l'humain) par JNK dans les lignées cellulaires COS-1 et F9. Ceux-ci sont phosphorylés en réponse à l'acide rétinoïque [Gianní et al., 2003] et aussi à différents stress, tels les radiations UV, l'arsenic et l'anisomycine [Adam-Stitah et al., 1999] [Tarrade et al., 2005] [Bruck et al., 2005]. Cette hyperphosphorylation est nécessaire pour la coopération entre RXRα et RARγ [Gianní et al., 2003] [Tarrade et al., 2005] [Bruck et al., 2005]. Cette hyperphosphorylation du domaine N-terminale est aussi reliée à la phosphorylation du LBD au niveau de la position Ser260/265 par JNK, impliquée dans l'inactivation de certains gènes sous le contrôle de l'acide rétinoïque [Adam-Stitah et al., 1999] [Bruck et al., 2005]. De manière indépendante, la position Tyr249/254 peut être aussi phosphorylée [Lee et al., 2000]. Celle-

ci est induite par MKK4/SEK1 directement et non JNK et a pour conséquence d'inhiber la fonction de RXRα [Lee et al., 2000]. De plus, il a été montré que la double mutation Thr82Ala et Ser260Ala inhibait la croissance cellulaire et induisait l'apoptose des cellules hépatique Huh7 humain (hépatocarcinome) en présence de 9cRA. Au contraire, la double mutation Thr82Asp/Ser260Asp, utilisée pour mimer la phosphorylation des sérines, montre un comportement inverse. Les auteurs suggèrent que la phosphorylation empêche l'homodimérisation de RXRα ou son hétérodimérisation avec RARβ [Yoshimura et al., 2007].

L'impact de la phosphorylation de RXRα et de l'action de la vitamine D_3 (1,25-dihydroxyvitamine D_3 [1,25$(OH)_2$ D_3]), modulant l'activité du VDR, a aussi été étudié au sein des lignées cellulaires HPK1A*ras*, surexprimant la protéine *ras* [Sebag et al., 1992] [Solomon et al., 1998] [Solomon et al., 1999]. Ils ont montré que les cellules HPK1A*ras* étaient résistantes à l'effet de la vitamine D_3, contrairement aux cellules HPK1A [Sebag et al., 1992] et que cette résistance serait dû à une perturbation du complexe hétérodimérique RXR-VDR [Solomon et al., 1998]. Cette résistance a de même été observée au sein de lignées cellulaires de carcinomes pancréatiques [Zugmaier et al., 1996] et du sein [Narvaez et al., 1996]. De manière générale, les études sur l'effet de la vitamine D_3 montrent que celle-ci inhibe la croissance cellulaire des cellules cancéreuses et stimule la différenciation cellulaire [Bouillon et al., 2006] [Fleet, 2008]. D'autres travaux menés par la suite indiquent que l'activation de la transcription par VDR est atténuée à travers la phosphorylation de RXRα par les MAPKs [Solomon et al., 1999]. Cette PTM est médiée par la surexpression de la protéine oncogène *ras*, induisant la cascade MAPK et phosphorylant RXRα. Cette phosphorylation a lieu au niveau de la Ser260, situé dans la boucle H1-H3 du LBD de RXRα. L'inhibition de l'activité de MAPK rétablit l'action anti-proliférative de la vitamine D_3, tout comme la mutation Ser260Ala [Solomon et al., 1999].

Par la suite, il a été montré que la phosphorylation en position Ser260 affectait l'interaction avec le complexe hétérodimérique RXR-VDR [Macoritto et al., 2008]. L'utilisation d'inhibiteur MAPK ou la mutation Ser260Ala restaure l'interaction avec les protéines CoA au sein de la lignée cellulaire HPK1A*ras*, soulignant le rôle prépondérant de la phosphorylation de Ser260. Le même comportement est observé avec RARγ, RXRα, TRβ et PPARγ, sans que la dimérisation soit impactée [Macoritto et al., 2008]. Ceci serait dû à la localisation de la Ser260, située dans la boucle Ω entre les hélices H1-H3 du LBD, à proximité de la surface d'interaction avec les protéines CoA (Figure 1.19). La phosphorylation de Ser260 est aussi impliquée dans le développement du carcinome hépatocellulaire, un cancer primitif du foie, en empêchant sa dégradation par le protéasome et favorisant la croissance cellulaire incontrôlée des cellules hépatiques [Matsushima-Nishiwaki et al., 2001] [Adachi et al., 2002].

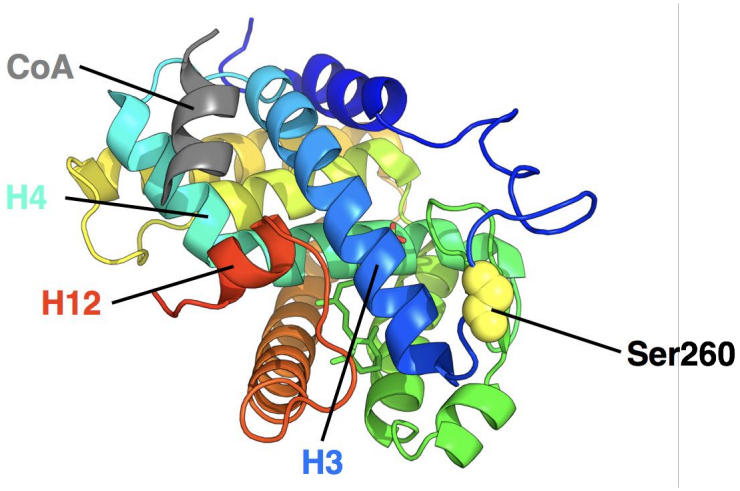


FIGURE 1.19 – Localisation du résidu Ser260 (en jaune) par rapport à la position du peptide CoA (en gris) au sein de la structure cristallographique de la forme liée à 9cRA du LBD de RXRα (PDB id : 1FM9).

1.7 Objectifs

Toutes les observations concernant la dynamique conformationnelle du LBD des NR suggèrent dès lors que celui-ci est capable d'explorer un nombre multiple de conformations et que cette dynamique est nécessaire pour que le ligand puisse entrer et sortir de la LBP. En absence de ligand, le LBD est flexible et les protéines coactivatrices n'interagissent pas correctement. Dans le cas inverse, en présence d'un ligand agoniste, le LBD est plus stable ce qui permet de promouvoir l'activité transcriptionnelle en recrutant les protéines CoA.

Ces études ont mis en évidence que les formes libres des LBD, (excepté pour GR et NURR-1 discutés précédemment), échangeaient plus rapidement dans la partie formant la poche de liaison du ligand que les formes liées à un agoniste. La diminution de la vitesse d'échange H/D suggère fortement que la dynamique locale de la protéine diminue aussi. Sachant que l'échange H/D mesuré est essentiellement influencé par les protons amides du squelette peptidique, impliqués dans le maintien des structures secondaires, les changements observés peuvent être interprétés comme des perturbations conformationnelles dues au ligand ou à son absence. Concernant l'hélice H12, de nombreuses études montrent que celle-ci est plus mobile en absence de ligands. Cependant, cela n'implique pas nécessairement qu'elle soit dans une conformation étendue, comme c'est le cas au sein de la structure cristallographique de la forme libre de LBD de RXRα, ou de manière générale qu'elle soit détachée du domaine. Les structures du LBD de PPARγ sans ligand

[Uppenberg et al., 1998] [Nolte et al., 1998] ou avec un ligand et un peptide CoA [Nolte et al., 1998] sont par exemple très similaires, bien qu'il ait été suggéré que la structure sans ligand obtenue soit un artefact cristallographique dans laquelle un ligand de faible affinité n'a pas été détecté [Liberato et al., 2012].

La présence d'un ligand agoniste va affecter différemment la stabilité des LBD de chaque NR, d'une part par la nature différente des ligands entre chaque NR, et d'autre part par les différences intrinsèques au niveau du LBD même. Par exemple, la présence d'un ligand agoniste au sein de TR, l'hormone thyroïdienne T_3, va non seulement impacter les éléments structuraux qui composent la LBP et plus particulièrement H12, mais aussi d'autres éléments plus éloignés qui ne sont pas en contact direct, tel que H1. Il a été montré à plusieurs reprises par une autre approche que la stabilité de cette dernière était essentielle pour la liaison du ligand et par réciprocité, que le ligand lui aussi agissait de même sur H1 [Pissios et al., 2000] [Huber et al., 2003b] [Huber et al., 2003a]. La dynamique conformationnelle de cet élément est un point clé dans le fonctionnement de TRβ, où l'introduction de mutations accroissant sa dynamique (déstabilisation) va être à l'origine du syndrome de résistance à l'hormone thyroïdienne T_3 [Huber et al., 2003b] [Huber et al., 2003a].

La transcription de gènes spécifiques est contrôlée par les récepteurs nucléaires de manière dépendante à un ligand. La liaison du ligand est en effet liée à l'activation du complexe, ce qui déclenche une cascade d'évènement menant au début de la transcription. Cependant, l'activité du récepteur nucléaire peut être ajustée plus finement selon le contexte cellulaire par l'ajout de PTMs. Dans un contexte cellulaire normal, ces modifications sont nécessaires pour le bon déroulement de processus transcriptionnel et sont sous contrôle pour éviter tout déséquilibre. Cependant, l'exacerbation de certaines voies de signalisation, telles la cascade MAPK ou les kinases dépendantes des cyclines (CDKs) dans le cas de la phosphorylation [Rochette-Egly, 2003], par manque de rétrocontrôle ou sous l'effet d'un stress génotoxique est liée à la dérégulation de la croissance cellulaire pouvant mener à une néoplasie [Anbalagan et al., 2012]. C'est le cas de la phosphorylation du LBD de RXRα au niveau de la sérine 260 qui est liée au développement de carcinome hépatocellulaire. Cette phosphorylation perturbe la dégradation normale de RXRα ainsi que l'interaction avec les protéines CoA [Matsushima-Nishiwaki et al., 2001] [Macoritto et al., 2008].

Ainsi dans un premier temps, l'objectif a été de développer et valider un protocole de dynamique moléculaire accélérée permettant d'étudier la dynamique conformationnelle de la forme libre et liée à 9cRA du LBD de RXRα. Compte tenu de la disponibilité de données expérimentales de diverses sources (HDX, RMN) sur la dynamique du LBD de RXRα, il constitue un cas favorable pour valider le protocole de dynamique moléculaire accélérée. L'objectif de notre étude sera également d'utiliser les conformations

issues de la dynamique pour interpréter les données d'échange H/D. Puis dans un second temps, l'objectif a été d'étudier la dynamique conformationnelle de la forme libre phosphorylée de la sérine 260 (pSer260) du LBD de RXRα par dynamique moléculaire accélérée. Du fait de l'absence de données expérimentales disponibles sur la forme phosphorylée du domaine concernant sa dynamique, les données expérimentales et prédites obtenues précédemment pour la forme libre non phosphorylée ont été utilisées. Ceci nous a permis d'identifier les régions dont la dynamique est perturbée par la phosphorylation pSer260.

Chapitre 2

Dynamique moléculaire accélérée

2.1 Introduction

Nous avons étudié la dynamique structurale du récepteur RXRα par les méthodes de simulations numériques, en particulier la dynamique moléculaire. Le champ de force classique, et la méthodologie de dynamique moléculaire ont été présentés dans de nombreux livres et articles de revue [Leach, 2001] [Karplus and McCammon, 2002] [Adcock and McCammon, 2006] [Schlick, 2010], et nous détaillerons dans cette introduction méthodologique essentiellement la méthode particulière que nous avons utilisée, la dynamique moléculaire accélérée.

2.2 Dynamique moléculaire accélérée

2.2.1 Introduction

Outre l'amélioration continue des capacités de calcul, de nombreux développements méthodologiques ont été proposés afin d'augmenter l'échantillonnage conformationnel au cours des simulations de dynamique moléculaire [Hamelberg and McCammon, 2006]. Historiquement, ceci inclut l'utilisation de l'algorithme SHAKE, par exemple, permettant d'utiliser des pas d'intégration plus longs (de 1 à 2 fs) [Ryckaert et al., 1977] [Hammonds and Ryckaert, 1991] [Weinbach and Elber, 2005], en réduisant les mouvements de hautes fréquences liés aux vibrations impliquant des atomes d'hydrogène. D'autres simplifications, comme l'élimination du solvant explicite par l'introduction d'un modèle implicite de solvatation comme le modèle « *Generalized Born* » [Still et al., 1990] [Bashford and Case, 2000] ou l'introduction de modèles « gros grains », ou un résidu n'est plus représenté que par quelques « billes », [Tozzini, 2005] permettent d'augmenter l'échelle de temps accessible aux simulations, mais introduisent des modifications parfois importantes du paysage énergétique. Cependant, pour les modèles tout atome l'échantillonnage reste limité du à la forte rugosité du paysage énergétique des macromolécules biologiques, avec des minimas multiples souvent séparés par des barrières de plusieurs k_bT (k_bT=0.6 kcal/mol à 298 K) [Onuchic and Wolynes, 2004] [Henzler-Wildman and Kern, 2007]. La méthode classique de dynamique moléculaire à température ambiante ne permet pas de franchir aisément de telles barrières sur l'échelle de temps habituelle des simulations, ce qui limite fortement l'espace exploré. Ainsi beaucoup de mécanismes/processus biologiques se produisant sur des échelles de temps de la milliseconde à la seconde, ne peuvent être observés à l'aide de la méthode classique de dynamique moléculaire. Une façon d'accroitre le nombre d'évènements rares entre plusieurs basins énergétiques est d'élever la température [Bruccoleri and Karplus, 1990], augmentant ainsi la probabilité de passer les barrières énergétiques, comme dans les méthodes de type *replica exchange*

(REMD) avec le *parallel tempering* (RE/PT) [Earl and Deem, 2005]. D'autres stratégies ont été développées, cette fois-ci le paysage énergétique est directement modifié en biaisant l'Hamiltonien $\mathcal{H}$. L'objectif de ces méthodes est de faciliter l'exploration de plusieurs bassins énergétiques pour un temps de calcul réduit.

Parmi ces méthodes, il est possible de différencier deux classes : (1) basées sur le choix de variables collectives (CVs) (ou coordonnées de réaction) dont les méthodes les plus connues sont l'*umbrella sampling*, l'*adaptive biasing force* et la *metadynamics* (2) et d'autres qui ne sont basées sur aucune contrainte comme l'*hyperdynamics*, la dynamique moléculaire accélérée (aMD) ou encore le *parallel tempering*, qui elle est basée sur la température et non sur l'Hamiltonien $\mathcal{H}$. Mais quoi qu'il en soit, il est nécessaire dans les deux cas de repondérer l'énergie biaisée afin de retrouver le vrai paysage énergétique ou le système à la température désirée [Shen and Hamelberg, 2008].

Bien que les méthodes de la première classe permettent de converger rapidement, il est nécessaire de prédéfinir une ou plusieurs variables collectives afin de décrire le système. Ainsi, cela nécessite d'avoir à priori une connaissance accrue du système étudié. De fait, la principale source d'erreur se trouve dans le choix des variables collectives qui peuvent dans certaines situations ne pas représenter fidèlement le système, en omettant une barrière énergétique par exemple [Spiwok et al., 2015]. Les méthodes non restreintes ont l'avantage d'explorer l'espace conformationnel d'une macromolécule biologique sans avoir à priori de connaissances sur le système étudié. Il est possible dès lors d'explorer les chemins de transition entre plusieurs bassins énergétiques, mais aussi pouvoir identifier de possibles états intermédiaires inconnus. Un des principaux désavantages du RE/PT se situe au niveau des ressources de calcul nécessaire pour converger. Le nombre de replicas requis augmente de façon quadratique à mesure que le système étudié augmente en taille [Miao and McCammon, 2016]. En outre, les transitions observées dans les répliques à haute température peuvent ne pas être observées à basse température due à la modification de la balance enthalpie/entropie du système et il peut être nécessaire de filtrer les transitions qui ne pourraient pas se produire à basse température [So et al., 2000]. L'approche menée par Voter, nommée *hyperdynamics*, permet d'accélérer l'échantillonnage conformationnel en diminuant le temps passé dans les minimas locaux, par l'ajout d'un biais $\Delta V(r)$ à la surface d'énergie potentielle $V(r)$ [Voter, 1997b] [Voter, 1997a]. L'ajout du biais $\Delta V(r)$ est fait de telle sorte que la surface d'énergie potentielle ne soit modifiée qu'à proximité des minimas locaux et non près des barrières ou des points-selle (points col). Cependant, ces régions de transitions ne peuvent être identifiées qu'en diagonalisant la matrice Hessienne à chaque pas, ce qui limite la taille du système à environ une centaine d'atomes seulement [Hamelberg and McCammon, 2006]. Une

autre alternative, cette fois-ci peu coûteuse en termes de calcul et pouvant être applicable à des macromolécules biologiques, a été développé par Steiner [Steiner et al., 1998]. Dans cette méthode, un biais $\Delta V(r)$ est ajouté de sorte à rendre la surface d'énergie potentielle « plate ». Ainsi la surface d'énergie potentielle modifiée $V^*(r)$ sera égale à $V(r) + \Delta V(r) = E$ quand $V(r)$ est inférieur ou égal à E et à $V(r)$ quand celle-ci est supérieur à E.

$$V^*(r) = \begin{cases} V(r) & \text{si } V(r) > E \\ E & \text{si } V(r) \leqslant E \end{cases} \tag{2.1}$$

Cependant, la surface d'énergie potentielle est discontinue aux points où la surface d'énergie potentielle $V(r)$ rencontre la surface d'énergie potentielle $V^*(r) = E$. Ce problème a été corrigé par la suite [Rahman and Tully, 2002], mais il reste néanmoins le problème du choix de E. S'il est choisi trop « haut », c'est-à-dire au-dessus de la plupart des barrières énergétiques, le système va bien explorer un large l'espace, mais de manière aléatoire, ce qui n'est pas efficace. Au contraire, dans le cas de la dynamique moléculaire accélérée, la surface potentielle d'énergie modifiée $V^*(r)$ maintient la forme de la surface potentielle d'énergie originale $V(r)$ et fusionne sans problème avec cette dernière lorsque $V(r) = E$.

2.2.2 Théorie de la dynamique moléculaire accélérée

De même que pour les modifications proposées par Steiner *et al.* [Steiner et al., 1998] et Rahman *et al.* [Rahman and Tully, 2002], un biais non négatif $\Delta V(r)$ est ajouté à la surface d'énergie potentielle $V(r)$, quand celui-ci est inférieur à E. Ainsi, la simulation sera réalisée sur la surface d'énergie potentielle modifiée $V^*(r) = V(r) + \Delta V(r)$. Dans le cas où l'énergie est supérieure ou égale à E, la simulation est réalisée sans biais $\Delta V(r)$ sur la vraie surface d'énergie potentielle $V^*(r) = V(r)$.

$$V^*(r) = \begin{cases} V(r) & \text{si } V(r) \geqslant E \\ V(r) + \Delta V(r) & \text{si } V(r) < E \end{cases} \tag{2.2}$$

avec $\Delta V(r)$ qui sera déterminé de sorte que $V^*(r)$ ne présente aucune discontinuité et que la surface modifiée reproduise la forme de la surface d'énergie potentielle originale $V(r)$ même si la valeur de E est élevé. Ainsi le choix de $\Delta V(r)$ sera défini par la relation suivante :

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \tag{2.3}$$

avec α le paramètre permettant de « lisser » (*smoothness*) la surface d'énergie potentielle $V(r)$. Si $\alpha = 0$, alors la surface d'énergie potentielle modifiée $V*(r)$ sera « plate » avec $V^*(r) = E$, tout comme c'est le cas avec les versions alternatives de l'*hyperdynamics* [Steiner et al., 1998] [Rahman and Tully, 2002] présentées plus tôt. Au contraire, à mesure que l'on augmente α, alors la surface d'énergie potentielle modifiée $V^*(r)$ devient de plus en plus « rugueuse » et se rapproche de $V(r)$ (Figure 2.1).

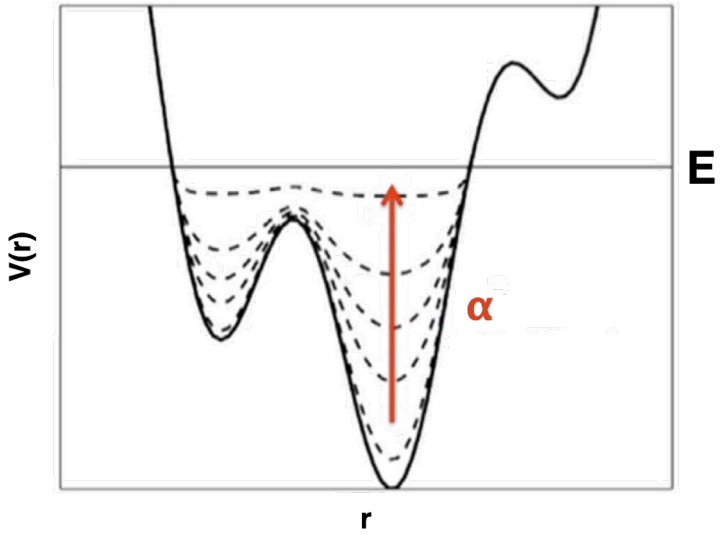


FIGURE 2.1 – Illustration schématique de la dynamique moléculaire accélérée : Un biais non négatif est ajouté à la surface d'énergie potentielle quand l'énergie du système est inférieure à E. Au fûr et à mesure que le paramètre α décroit, la nouvelle surface d'énergie devient de plus en plus « lisse », augmentant le nombre de transitions entre les différents bassins de basses énergies. Adapté de [Miao and McCammon, 2016].

Dès lors le choix des paramètres E et α est très important. De manière générale, E devrait être strictement supérieur au minimum de $V(r)$, V_{min}, correspondant souvent à la structure de départ. Si E est inférieur à V_{min} alors la simulation sera réalisée sur $V(r)$ qui revient au final à réaliser une dynamique moléculaire classique. Sachant que les macromolécules biologiques possèdent plusieurs minimas, il est suggéré de calculer l'énergie potentielle moyenne $< V[r(t)] >$ à partir d'une courte dynamique moléculaire classique en partant de la structure initiale. Si E est trop « bas », c'est-à-dire en dessous de la plupart des barrières énergétiques, alors la probabilité de transiter d'un bassin à l'autre sera la même pour $V(r)$ et $V^*(r)$ et α n'aura que très peu d'importance, du moment que $\alpha > 0$ afin d'éviter toute discontinuité à $V^*(r) = E$. Dans le cas contraire, à mesure que E devient « haut » alors α devient important. Il est nécessaire de choisir

une valeur α beaucoup plus grande que 0, pour (1) éviter que $V^*(r)$ soit isoénergétique et discontinue et (2) faire en sorte que la surface d'énergie potentielle modifiée $V^*(r)$ possède les mêmes bassins énergétiques que la surface non modifiée $V(r)$ et ceci en prenant, par exemple, $\alpha = E - V_{\text{min}}$ [Hamelberg et al., 2004].

2.2.3 Repondération de l'énergie

En dynamique moléculaire accélérée, la repondération de chaque configuration en fonction du biais $\Delta V(r)$ qui lui est associé permet de reproduire correctement la probabilité de distribution canonique [Hamelberg et al., 2004], à condition que la simulation aMD ait convergé. La moyenne d'un ensemble $< A >$ d'une observable quelconque $A(r)$ sur la surface d'énergie potentielle non modifiée $V(r)$ est donnée par la relation :

$$< A >= \frac{\int dr\ A(r)\ e^{-\beta V(r)}}{\int dr\ e^{-\beta V(r)}} \tag{2.4}$$

De manière générale, l'ensemble moyen de n'importe quel observable $A(r)$ sur la surface modifiée $V^*(r)$ peut s'écrire :

$$\begin{aligned} < A^* > &= \frac{\int dr\ A(r)\ e^{-\beta V^*(r)}}{\int dr\ e^{-\beta V^*(r)}} \\ &= \frac{\int dr\ A(r)\ e^{-\beta V(r) - \beta \Delta V(r)}}{\int dr\ e^{-\beta V(r) - \beta \Delta V(r)}} \end{aligned} \tag{2.5}$$

Il est ainsi possible de corriger $< A^* >$ en multipliant chaque configuration par $e^{-\beta \Delta V(r)}$, qui n'est d'autre que le facteur de Boltzmann du biais $\Delta V(r)$ appliqué à chaque configuration, avec $\beta = 1/k_b T$. Dans le cas, où le système a évolué sur la surface non modifiée $V(r)$, alors $\Delta V(r)$ est égal à 0. De cette manière, on obtient l'ensemble moyen corrigé $< A^C >$, qui est tout simplement égal à l'ensemble moyen $< A >$:

$$\begin{aligned} < A^C > &= \frac{\int dr\ A(r)\ e^{-\beta V(r) - \beta \Delta V(r)} e^{\beta \Delta V(r)}}{\int dr\ e^{-\beta V(r) - \beta \Delta V(r)} e^{\beta \Delta V(r)}} \\ &= \frac{\int dr\ A(r)\ e^{-\beta V(r)}}{\int dr\ e^{-\beta V(r)}} \\ &=< A > \end{aligned} \tag{2.6}$$

On peut donc conclure que la méthode de dynamique moléculaire accélérée converge bien vers une distribution canonique et qu'il est possible en théorie d'obtenir un ensemble canonique corrigé en repondérant

chaque configuration par le biais $\Delta V(r)$ appliqué à chacune d'entre elles. Les études de peptides tels que le hepta-alanine ou le dipeptide-alanine ont montré que sur de petits systèmes, la dynamique moléculaire accélérée convergeait bien vers une distribution canonique correcte [Hamelberg et al., 2004]. En pratique, le biais $\Delta V(r)$ appliqué au cours d'une dynamique moléculaire accélérée sera de l'ordre de plusieurs dizaines à quelques centaines de kcal/mol selon la taille de la macromolécule biologique étudiée, contrairement aux méthodes basées sur des variables collectives, de l'ordre de quelques kcal/mol seulement. De fait, il n'est pas possible de repondérer efficacement les simulations aMD et de retrouver la surface d'énergie potentielle d'origine $V(r)$, surtout lorsqu'il s'agit de larges systèmes avec plusieurs centaines de résidus. Ceci est dûe à la méthode de repondération exponentielle, présentée plus tôt, qui souffre de larges erreurs statistiques à cause de facteurs de Boltzmann souvent dominés par quelques configurations possédant un biais $\Delta V(r)$ élevé [Shen and Hamelberg, 2008] [Miao et al., 2014b] [Miao et al., 2014a]. D'autres méthodes ont été ainsi introduites pour supprimer ce bruit en modifiant la repondération exponentielle [Shen and Hamelberg, 2008] [Miao et al., 2014b]. Malgré l'avantage que présente la dynamique moléculaire accélérée pour explorer l'espace conformationnel d'un système sans connaissances à priori et sans devoir définir des variables collectives pour décrire le système, dans les faits, ces variables collectives sont tout de même introduites pour pouvoir repondérer les simulations aMD.

Soit la probabilité de distribution le long d'une coordonnée de réaction $A(r)$, $p^*(A)$. Connaissant $\Delta V(r)$ appliqué à chaque configuration, $p^*(A)$ pourra être repondéré afin de retrouver l'ensemble canonique $p(A)$.

$$p(A_j) = p^*(A_j)\frac{< e^{\beta\Delta V(r)} >_j}{\sum_{j=1}^{M} < e^{\beta\Delta V(r)} >_j} \tag{2.7}$$

avec M le nombre de classes de l'histogramme, $\beta = 1/k_bT$ et $< e^{\beta\Delta V(r)} >_j$ l'ensemble moyen pour la classe j. De cette manière, on obtient une repondération exponentielle, présentée plus tôt. Le potentiel de force moyenne (PMF) final sera obtenu de la façon suivante $F(A_j) = -(1/\beta)\ \ln p(A_j)$. Afin de réduire le bruit, dû aux quelques configurations présentant de larges fluctuations $\Delta V(r)$, il est possible de remplacer le terme exponentiel par une série de MacLaurin.

$$p(A_j) = p^*(A_j)\frac{\sum_{k=0}^{\infty} \frac{\beta^k}{k!} < \Delta V^k >_j}{\sum_{j=1}^{M}\sum_{k=0}^{\infty} \frac{\beta^k}{k!} < \Delta V^k >_j} \tag{2.8}$$

avec k étant l'ordre de la série de Maclaurin allant typiquement de 5 à 10. Il a été montré que les PMF obtenus à l'aide de cette méthode étaient moins bruités que ceux obtenus à l'aide de la méthode

exponentielle [Miao et al., 2014b] [Pierce et al., 2012]. De plus, il est aussi possible d'utiliser la série des cumulants.

$$p(A_j) = p^*(A_j) \frac{< e^{\sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k} >_j}{\sum_{j=1}^{M} < e^{\sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k} >_j} \tag{2.9}$$

avec les 3 premiers cumulants étant égal à

$$\begin{aligned} C_1 &=< \Delta V > \\ C_2 &=< \Delta V^2 > - < \Delta V >^2 = \sigma^2_{\Delta V} \\ C_3 &=< \Delta V^3 > -3 < \Delta V^2 >< \Delta V > +2 < \Delta V >^2 \end{aligned} \tag{2.10}$$

En principe, les deux premiers cumulants permettent de bien mieux supprimer le bruit le long de la coordonnée de réaction choisie que la série de Maclaurin, en sachant que cette dernière est équivalente au premier cumulant, si la distribution du biais $\Delta V(r)$ suit une distribution gaussienne [Shen and Hamelberg, 2008] [Miao et al., 2014b]. De plus, afin de mesurer la qualité de la repondération des simulations aMD, il est possible de calculer l'anharmonicité γ [Miao et al., 2014b].

Soit l'ensemble des biais $\Delta V(r)$ pour chaque configuration dans une classe (*bin*) le long d'une coordonnée de réaction r choisie, la probabilité de distribution normalisée $p(\Delta V(r))$ peut être obtenue et son entropie calculée de la façon suivante :

$$S_{\Delta V} = \int_0^{\infty} p(\Delta V) \ln(p(\Delta V)) \, \mathrm{d}\Delta V \tag{2.11}$$

Si la distribution de $\Delta V(r)$ suit une loi gaussienne, c'est-à-dire si on a un échantillonnage parfait, alors l'entropie maximale est donnée par :

$$S_{\text{max}} = \frac{1}{2} \ln(2\pi e \sigma^2) \tag{2.12}$$

avec σ la déviation standard de $\Delta V(r)$. L'anharmonicité γ au sein d'une classe et ainsi obtenue de la manière suivante :

$$\begin{aligned} \gamma &= S_{\text{max}} - S_{\Delta V} \\ &= \frac{1}{2} \ln(2\pi e \sigma^2) - \int_0^{\infty} p(\Delta V) \ln(p(\Delta V)) \, \mathrm{d}\Delta \end{aligned} \tag{2.13}$$

Quand γ est égal à zéro, alors le biais $\Delta V(r)$ suit une distribution gaussienne et une restitution fidèle

de la surface d'énergie potentielle par repondération est possible. Dans le cas contraire où la distribution $\Delta V(r)$ ne suit pas une distribution gaussienne, alors l'anharmonicité γ sera supérieure à zéro. Plus l'anharnonicité γ est élevée et plus la surface d'énergie potentielle corrigée déviera de la surface d'origine $V(r)$. Ainsi, cette méthode permet de s'assurer de la qualité du paysage énergétique obtenu par l'une des méthodes de repondération.

2.2.4 Variantes de la dynamique moléculaire accélérée

Les champs de force tout atomes sont généralement représentés comme la somme de plusieurs termes d'énergie, dont la forme la plus simple est la suivante :

$$\begin{aligned} V(r) &= \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 \\ &+ \sum_{\text{dièdres}} \sum_{n} \frac{V_n}{2}[1 + cos(n\Phi - \gamma)] \\ &+ \sum_{i<j} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}] + \sum_{i<j} \frac{q_i q_j}{\epsilon R_{ij}} \\ &= V_B(r) + V_\theta(r) + V_D(r) + V_{LJ}(r) + V_{EE}(r) \end{aligned} \tag{2.14}$$

avec $V_B(r)$ et $V_\theta(r)$ les termes harmoniques des énergies des liaisons et des angles, $V_D(r)$ l'énergie des angles de torsions des angles dièdres et $V_{LJ}(r)$ et $V_{EE}(r)$ les termes d'énergies non liées de van der Waals et les interactions électrostatiques, respectivement. En principe, le biais $\Delta V(r)$ peut-être appliqué à l'ensemble des termes d'énergies ou de manière séparée sur l'un ou l'autre des termes. Les principaux changements conformationnels des protéines étant liés à des modifications des angles de torsions, l'accélération ne fut originalement appliquée qu'à l'énergie des angles dièdres et des interactions non liées 1-4. Ainsi les équations précédentes deviennent :

$$V^*(r) = \begin{cases} V_0(r) + V_D(r) & \text{si } V_D(r) \geqslant E_D \\ V_0(r) + V_D(r) + \Delta V_D(r) & \text{si } V_D(r) < E_D \end{cases} \quad \Delta V_D = \frac{(E_D - V_D(r))^2}{\alpha + (E_D - V_D(r))} \tag{2.15}$$

avec $V_0(r)$ l'énergie potentielle sans les termes d'énergies des angles de torsions et des interactions non liées 1-4 et E_D similaire à E dans les équations précédentes. Cette version dénommée *dihedral-boost* est cependant insuffisante pour échantillonner efficacement l'espace conformationnel. Un autre biais fut ajouté à l'énergie potentielle totale, afin d'accélérer la diffusivité du solvant (molécules d'eau) amé-

liorant ainsi l'échantillonnage conformationnel [de Oliveira et al., 2006] [Hamelberg et al., 2007]. Dans cette version, cette fois-ci dénommée *dual-boost*, la surface d'énergie potentielle modifiée $V^*(r)$ devient $V^*(r) = [V_B(r) + V_\theta(r) + [V_D(r) + \Delta V_D(r)] + V_{LJ}(r) + V_{EE}(r)] + \Delta V(r)$. De même l'ensemble canonique corrigé peut-être retrouvé en reponderant chaque configuration par le facteur de Boltzmann du biais $e^{\beta[\Delta V_D(r)+\Delta V(r)]}$. Concernant les paramètres E et α, un jeu initial de paramètres peut-être obtenu à partir des relations empiriques suivantes [Wereszczynski and McCammon, 2012] :

$$\begin{aligned} E_{dihed} &=< V_{\text{dihed}} > +4N_{\text{res}} & E_{total} &=< V_{\text{total}} > + \frac{N_{\text{atomes}}}{5} \\ \alpha_{dihed} &= \frac{4}{5}N_{\text{res}} & \alpha_{total} &= \frac{N_{\text{atomes}}}{5} \end{aligned} \tag{2.16}$$

avec N_{res} et N_{atomes}, le nombre de résidus et d'atomes, respectivement, $< V_{\text{dihed}} >$ et $< V_{\text{total}} >$, les énergies moyennes des angles dièdres et potentielles calculées à partir de trajectoires préliminaires de dynamique moléculaire classique. Si l'échantillonnage est considéré comme insuffisant, il est conseillé d'augmenter E par des multiples de α [Wereszczynski and McCammon, 2012], ce qui devrait modifier de manière notable l'échantillonnage. Dans le cas de protéines membranaires, il est nécessaire de prendre en compte le nombre différent d'angles dièdres entre les lipides et les acides aminés des protéines, ainsi les paramètres E et α seront obtenus à partir des relations suivantes [Miao et al., 2013] [Miao et al., 2014a] :

$$\begin{aligned} E_{dihed} &=< V_{\text{dihed}} > +0.3 < V_{\text{dihed}} > & E_{total} &=< V_{\text{total}} > +0.2N_{\text{atomes}} \\ \alpha_{dihed} &= \frac{0.3}{5} < V_{\text{dihed}} > & \alpha_{total} &= 0.3N_{\text{atomes}} \end{aligned} \tag{2.17}$$

En plus des versions précédentes, d'autres variations de la dynamique moléculaire accélérée ont été testées afin d'améliorer l'échantillonnage, mais aussi pour diminuer les erreurs statistiques et permettre une meilleure repondération des simulations aMD [Doshi and Hamelberg, 2015] [Miao and McCammon, 2016]. Parmi celles-ci, nous avons le *windowed aMD* [Sinko et al., 2012], qui en plus de surélever les minimas, abaisse aussi les barrières énergétiques, puis la *selective aMD* [Wereszczynski and McCammon, 2010], où juste quelques résidus sont accélérés au lieu de l'ensemble du système, réduisant les erreurs statistiques, et de même il existe la version *rotatable aMD* [Doshi and Hamelberg, 2012] dans laquelle le biais $\Delta V(r)$ est ajouté seulement aux angles dièdres qui participent de manière significative aux changements conformationnel, et non à l'ensemble des angles de torsions, par exemple ceux présents dans les cycles des résidus aromatiques [Miao et al., 2016]. Dans la dernière version en date, nommée *gaussian accelerated molecular dynamics*, la surface d'énergie potentielle est modifiée en utilisant une fonction harmonique de sorte que le biais $\Delta V(r)$ appliqué suive une distribution gaussienne, permettant ainsi de

pouvoir repondérer efficacement les simulations aMD de systèmes plus large (> 100 résidus) [Miao et al., 2015b]. La dynamique moléculaire accélérée a aussi été combinée à des méthodes de type *replica exchange* (REXAMD/H-REMD) [Fajer et al., 2008] [Fajer et al., 2009] [Arrar et al., 2012], en faisant varier les niveaux d'accélération, mais aussi couplée à la température pour le *multidimensional REMD* [Roe et al., 2014] ou encore à des protocoles de dynamique moléculaire à pH constant [Williams et al., 2010] et à des calculs *ab initio* [Bucher et al., 2011b] [Pierce et al., 2011] [Ortiz-Sánchez et al., 2012].

2.2.5 Applications

La dynamique moléculaire accélérée a été employée à de nombreuses reprises pour caractériser la dynamique structurale de plusieurs macromolécules biologiques, tel qu'un complexe DNAse E9-Im9 [Baron et al., 2008], l'enzyme TcPR du parasite *Trypanosoma cruzi* [de Oliveira et al., 2011], la protéine-G [Miao et al., 2014a] [Kappel et al., 2015] [Miao et al., 2015a], une bicouche lipidique [Wang et al., 2011] et p38α [Filomia et al., 2010], par exemple. Dans plusieurs exemples (protéines Ras [Grant et al., 2009], GB3 [Markwick et al., 2007], ubiquitine [Markwick et al., 2009] et iκBα [Markwick et al., 2010]), les ensembles des conformations générées étaient en accord avec les données expérimentales obtenues par RMN. Certaines études ont également montré que la dynamique moléculaire accélérée était capable en quelques centaines de nanosecondes de capturer des évènements se produisant sur des échelles de temps beaucoup plus longues, de l'ordre de la micro- à la milliseconde [Pierce et al., 2012] [Miao et al., 2013] [Bucher et al., 2011a] [Markwick et al., 2011] [Tikhonova et al., 2013]. De même, il a été possible récemment d'étudier le repliement de petites protéines en partant d'une chaine polypeptidique complètement dépliée [Doshi and Hamelberg, 2014] [Miao et al., 2015c]. De plus, la dynamique moléculaire accélérée a été utilisée pour étudier l'isomérisation cis/trans. En effet, l'isomérisation *cis-trans* de l'angle ω est un processus très long, de l'ordre de la micro/milliseconde ou de la seconde. Cette isomérisation joue un rôle important dans de nombreux mécanismes biologiques [Craveur et al., 2013] et il est très difficile, voire impossible, de l'observer par dynamique moléculaire classique. Observée d'abord sur de petits peptides possédant soit un motif Ser-Pro [Hamelberg et al., 2005a] [Hamelberg et al., 2005b] ou un motif Gly-Gly [Hamelberg and McCammon, 2005] [Markwick and McCammon, 2011b], l'importance du mécanisme d'isomérisation *cis-trans* sur la dynamique fonctionnelle de la cyclophiline (CypA) [Hamelberg and McCammon, 2009] [McGowan and Hamelberg, 2013] ainsi que l'influence de la phosphorylation sur l'isomérisation *cis-trans* chez Pin-I [Velazquez and Hamelberg, 2013] ont été aussi étudiées. La dynamique moléculaire accélérée a également été utilisée afin d'optimiser les paramètres des angles de torsion de la liaison ω du champ

de force AMBER afin de reproduire l'équilibre *cis-trans* et corriger la barrière énergétique, jusqu'alors sous-estimée [Doshi and Hamelberg, 2009] [Xin et al., 2010] [Doshi and Hamelberg, 2011].

2.3 Analyse des trajectoires

2.3.1 RMSD global

Afin d'étudier les changements conformationnels globaux d'une protéine au cours d'une trajectoire de dynamique moléculaire, la manière la plus simple est de déterminer le RMSD (**R**oot **M**ean **S**quare **D**eviation) entre les conformations issues de cette même trajectoire et une structure de référence, qui peut être soit la structure cristallographique ou bien la première conformation de la trajectoire. Généralement, le RMSD est déterminé en superposant les conformations sur les régions structurées les plus stables de la protéine. Il est calculé selon l'équation suivante :

$$\text{RMSD global} = \sqrt{\frac{\sum_{i=1}^{N}(r_i(t_0) - r_i(t))^2}{N}} \tag{2.18}$$

avec N le nombre d'atomes utilisés et $r_i(t_0)$ et $r_i(t)$ les positions des atomes i aux temps t_0 et t, respectivement. De plus, le calcul du RMSD global au cours permet de voir si le système converge vers un RMSD stable, ou bien si la simulation présente des anormalités, avec des valeurs de RMSD trop élevées.

2.3.2 Fluctuation RMS

En plus de mesurer le RMSD au cours du temps, il est aussi possible d'étudier la flexibilité des résidus ou éléments de structures secondaire de manière individuelle autour d'une position moyenne, en calculant les fluctuations RMS. Ces dernières sont obtenues selon la relation suivante :

$$\text{RMSF}_i = \sqrt{(\frac{1}{N}\sum_{i=1}^{N}(r_i(t) - < r_i >)^2)} \text{ avec } < r_i >= \frac{1}{N}\sum_{k=1}^{N} r_i(t) \tag{2.19}$$

avec $< r_i >$ la position moyenne de l'atome i calculée sur l'ensemble de la trajectoire, $r_i(t)$ la position de l'atome i au temps t et N le nombre total de conformations utilisées. Les fluctuations RMS de chaque atome i sont généralement comparées aux facteurs d'agitation thermique B issus de la structure

cristallographique de référence. L'équation reliant les fluctuations RMS aux facteurs B est la suivante :

$$\mathrm{RMSF}_i = \sqrt{\frac{\mathrm{B}_i \times 3}{8\ \Pi^2}} \tag{2.20}$$

2.3.3 Structure secondaire

Les protéines possèdent habituellement un contenu riche en structure secondaire, telles que des hélices α ou des feuillets β. Le suivi de l'évolution des structures secondaires au cours du temps va nous permettre de juger de la stabilité de ces dernières. Pour cela, nous utiliserons la méthode DSSP (**D**efine **S**econdary **S**tructure of **P**roteins) [Kabsch and Sander, 1983] appliquée à toutes les conformations extraites de la trajectoire, l'aide du module MDAnalysis [Michaud-Agrawal et al., 2011].

DSSP identifie les liaisons hydrogène au sein du squelette peptidique de la protéine en utilisant une définition purement électrostatique. DSSP place des charges partielles sur C,O (+q_1, -q_1) et N,H (-q_2, +q_2) et identifie une liaison hydrogène si E est inférieur à -0.5 kcal/mol dans l'équation suivante :

$$E = q_1 q_2 \left\{ \frac{1}{r(ON)} + \frac{1}{r(CH)} - \frac{1}{r(OH)} - \frac{1}{r(CN)} \right\} \times f \tag{2.21}$$

avec q_1 = 0.42e et q_2 = 0.20e, e étant l'unité de charge électronique et $r(AB)$ la distance interatomique entre A et B. r est en angströms, le facteur de dimension f vaut 332 Å kcal e^{-2} mol^{-1}, et E est en kcal/mol. Une bonne liaison hydrogène possède une énergie d'environ -3 kcal/mol, bien en deçà du paramètre de -0.5 kcal/mol. Ceci permet d'assigner une liaison hydrogène entre C=0 du résidu i et N-H du résidu j même s'il y a des erreurs dans les coordonnées ou des liaisons hydrogène bifurquées.

8 types de structures secondaires différentes sont généralement identifiées : les hélices 3_{10}, α et π dont les symboles sont G, H et I selon si l'espace entre les résidus est de 3, 4 ou 5. Il y a aussi deux types de feuillets β, le pont β avec le symbole B et le bulge β avec E. Nous avons aussi les tournants T et S pour les régions très courbées, quand l'angle entre le résidu i et i+2 est de moins de 70°. Mais il se peut aussi qu'aucune règle ne s'applique. Dans notre cas, ces 8 groupes seront regroupés en 3 groupes distincts avec les hélices (G, H et I) d'un côté, les feuillets β (B et E) de l'autre et le restant (T, S et non identifié) à part.

Chapitre 3

Réduction de la dimensionnalité

3.1 Introduction

L'accumulation importante de données disponibles publiquement, venues de domaines scientifiques divers, motive le développement de méthodes de réduction de la dimensionnalité, qui en permettent l'exploitation. La réduction de la dimensionnalité est la transformation de données de dimensionnalités élevées (généralement $N >> 3$) en une représentation de dimensionnalité plus basse ($n \leq 3$) et significative. On définit la dimension intrinsèque des données comme la dimension minimale (en nombre de paramètres/variables) nécessaires pour représenter fidèlement le contenu des données. Idéalement, la représentation obtenue par réduction de dimensionnalité devrait refléter la dimension intrinsèque des données. Les algorithmes de réduction de la dimensionnalité, tels que ceux utilisant les méthodes de regroupement (*clustering*) sont indispensables pour classifier, visualiser et compresser les informations de haute dimension afin d'en faciliter l'interprétation.

Des méthodes de réduction linéaire sont fréquemment utilisées, telles que l'analyse en composantes principales (*Principal Component Analysis*) (PCA) [Pearson, 1901]. Cependant, cette méthode ne peut pas traiter de manière efficace des données non linéaires. Le positionnement multidimensionnel (*Multidimensional Scaling*) (MDS) est une méthode dont le but est de générer une représentation de basse dimension des données en respectant la distance euclidienne que les sépare. Il existe plusieurs variants qui peuvent être soit linéaires ou non linéaires ou bien encore métriques ou non métriques. Cependant, le terme positionnement multidimensionnel peut aussi être utilisé en tant que classe de méthodes statistiques regroupant les méthodes *ISOmetric MAPping* (ISOMAP)[Tenenbaum et al., 2000] ou encore *Stochastic Proximity Embedding* (SPE) [Agrafiotis, 2003]. Ces dernières sont exclusivement non linéaires, elles ont donc la possibilité de manier des données non linéaires. Ces méthodes présentent un avantage sur les données issues du monde réel, car ces dernières ont une très forte chance d'être non linéaires.

Dans ce chapitre, nous nous intéresserons plus particulièrement à quelques méthodes de réduction de la dimensionnalité qui peuvent être appliquées en biologie structurale. Loin d'être parfaitement exhaustifs, nous ferons un rapide état des lieux de l'utilisation de la PCA, couramment utilisée en modélisation moléculaire. Nous verrons pourquoi il a été nécessaire d'aller plus loin dans le cadre de ce travail pour exploiter au mieux les conformations obtenues par dynamique moléculaire accélérée. Nous décrirons la méthode *Stochastic Proximity Embedding* (SPE) utilisée, ainsi que celles qui l'ont précédé et dont elle s'est inspirée.

3.2 Analyse sur composante principale

3.2.1 Introduction

Pour réduire la dimensionnalité, la méthode la plus connue est la PCA. Initialement introduite en 1901 par Karl Pearson [Pearson, 1901], développée indépendamment et nommée en tant que telle par Harold Hotelling en 1933 [Hotelling, 1933], le but de cette méthode est de construire une représentation de basse dimensionnalité décrivant au mieux la variance au sein des données. Pour cela, il est nécessaire de trouver une configuration linéaire qui diagonalise la matrice de covariance C. Cette dernière est obtenue selon la relation suivante :

$$C(M_{ij}) = \langle (M_i - \langle M_i \rangle)(M_j - \langle M_j \rangle) \rangle \tag{3.1}$$

avec la matrice M étant composée de N dimensions avec K observations pour chaque dimension. Sachant que la matrice de covariance C est symétrique, il est possible alors de trouver des vecteurs propres U la diagonalisant. Durant cette étape, on utilise généralement la méthode de Décomposition en Valeurs Singulières (SVD) :

$$U, S, V^{\mathrm{T}} = \mathrm{SVD}(C) \tag{3.2}$$

avec S la matrice diagonale contenant les valeurs propres et U et V, les matrices unitaires contenant les vecteurs propres. On réarrange ensuite les matrices S et U de telle sorte que les valeurs propres soient rangées dans l'ordre décroissantes. Les valeurs propres représentent la variance que possède chaque vecteur propre. Le vecteur propre avec la plus grande valeur propre est la composante principale des données, puis le second vecteur propre avec la deuxième plus grande valeur est la composante secondaire, et ainsi de suite. Afin de choisir le nombre de vecteurs propres que l'on gardera, on calcule la variance globale associée à l'ensemble des vecteurs propres :

$$s_N = \sum_{i=1}^{N} S[i, i] \tag{3.3}$$

avec N le nombre de dimensions totales. Puis, on choisit le plus petit nombre possible n ($1 \leq n < N$) de vecteurs propres tout en essayant de garder le maximum de variance. On ignorera les composantes de faible importance. Ainsi, en laissant de côté des vecteurs propres, la représentation finale sera de plus basse dimension que les données originales. La dimension finale sera égale au nombre de vecteurs propres rete-

nus. Usuellement, on choisira un nombre n de vecteurs propres représentant au moins 90% de la variance globale :

$$\frac{s_n}{s_N} \geq 0.9 \tag{3.4}$$

Afin d'obtenir la configuration M_n de dimensionnalité n, il suffit de multipler la matrice M contenant les données originales par la matrice U_n contenant les n vecteurs propres retenus, en fonction de la part qu'ils représentent par rapport à la variance globale :

$$M_n = M \cdot U_n \tag{3.5}$$

3.2.2 Analyse de trajectoires de dynamique moléculaire

La PCA est une méthode très populaire, utilisée pour analyser les simulations de dynamiques moléculaires. Elle permet de réduire la dimensionnalité, en séparant les mouvements de grandes amplitudes (lents) des protéines des petits mouvements (rapides). Les mouvements les plus lents sont représentés par les quelques premières composantes principales et les mouvements plus rapides par le reste. Ainsi il a été montré que les larges fluctuations d'un système peut-être décrit par quelques composantes principales, pouvant servir comme coordonnées de réactions [Ichiye and Karplus, 1991]. Généralement, on essaye de se limiter aux deux premières composantes principales pour la visualisation, s'ils sont représentatifs de la variance globale. De cette manière, il est possible d'identifier les conformations métastables et les états de transitions connectant les différents bassins énergétiques.

Afin que la PCA puisse être appliquée aux coordonnées cartésiennes des atomes, nommée cPCA pour *cartesian PCA*, il est nécessaire dans un premier temps, de superposer toutes les conformations afin de supprimer tout mouvement de translation et de rotation au cours de la trajectoire. Soit la structure cristallographique sera utilisée comme référence, soit la première conformation de la trajectoire. Autant pour les protéines « rigides », cette étape est relativement simple, autant pour les peptides « flexibles », cette tâche est plus complexe. Le problème se situant au moment de la phase de rotation pour superposer les conformations. S'il n'est pas possible de supprimer le mouvement global des mouvements internes de la protéine, alors des biais sont introduits, pouvant mener à une représentation faussée du paysage énergétique [Mu et al., 2005].

Pour remédier à ces difficultés, il a été proposé de réaliser la PCA en coordonnées internes, par exemple les angles dièdres ϕ et ψ du squelette peptidique [Mu et al., 2005] [Altis et al., 2007](nommée dPCA pour

dihedral PCA), ou les distances interatomiques [Abseher and Nilges, 1998], par exemple. Ceci permet de séparer sans ambiguïté les mouvements internes des mouvements globaux de la protéine. Les avantages liés à l'utilisation des angles dièdres ont été discutés sur plusieurs exemples [Mu et al., 2005] [Altis et al., 2007]. Cependant, il est nécessaire de transformer les données originales pour éviter les problèmes liés à la circularité des angles. On passera des angles dièdres (θ_n) à des coordonnées linéaires x et y grâce aux fonctions trigonométriques sinus et cosinus (sin θ_n, cos θ_n) :

$$\begin{aligned} M_{2N-1} &= \cos\theta \\ M_{2N} &= \sin\theta \end{aligned} \tag{3.6}$$

avec N le nombre d'angles dièdres utilisés, égal au nombre de dimensions de l'espace original, dans l'analyse. Ainsi la matrice de covariance résultante utilisée pour la dPCA utilisera $2N$ variables. Et en se limitant aux deux premières composantes principales, l'espace énergétique du système peut être obtenu de la manière suivante :

$$\Delta G(U_1, U_2) = -k_\mathrm{B}T(\ln\rho(U_1, U_2) - \ln\rho_\mathrm{max}) \tag{3.7}$$

avec k_B et T la constante de Boltzmann et la température, respectivement, ainsi que ρ la densité de probabilité d'une certaine classe de l'histogramme et ρ_max la densité de probabilité maximale de ce même histogramme utilisée afin de s'assurer que $\Delta G = 0$ pour l'état de plus basse énergie.

3.3 MultiDimensional Scaling

Le Multidimensional Scaling (MDS) classique, aussi connu sous le nom de *Principal Coordinates Analysis* (PCoA) a été introduit de manière indépendante par Torgerson en 1952 [Torgerson, 1952] et Gower en 1966 [Gower, 1966]. Le but de cette méthode est de représenter la similarité (ou la dissimilarité) entre des objets comme distance dans un espace de basse dimension n, généralement égale à 2 ou 3. Ainsi à l'aide de cette méthode, il est possible de directement visualiser les données et d'explorer visuellement leur structure plus facilement, contrairement à une simple matrice de distance (Figure 3.2). La méthode classique part du principe que les distances entre les objets sont euclidiennes, par exemple la distance entre plusieurs villes. L'avantage de cette méthode est qu'elle fournit une solution analytique (non itérative).

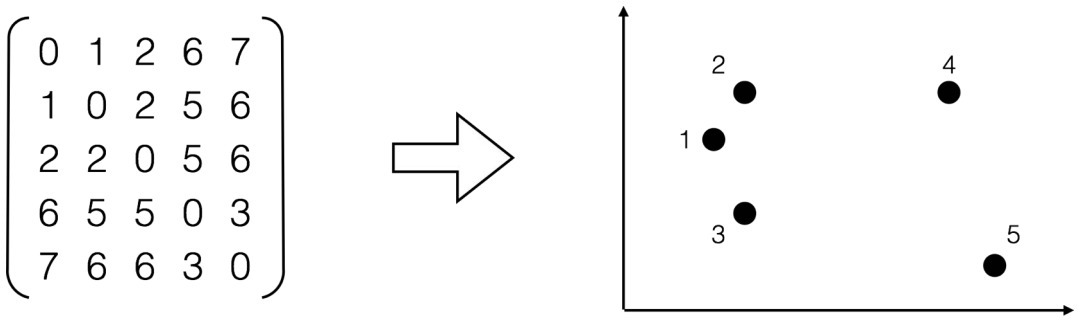


FIGURE 3.1 – Une matrice de distance Δ contenant les distances entre tous les objets est utilisée afin de calculer les coordonnées euclidiennes pour chaque point dans un espace de dimension plus réduite.

La méthode MDS classique implique quelques étapes d'algèbre linéaire, résumée ici par les étapes suivantes [Wickelmaier, 2003] :

1. Calcul de la matrice de distance au carrée $\Delta^{(2)} = [\Delta^2]$
2. Conversion de la matrice $\Delta^{(2)}$ en matrice doublement centrée B, selon la relation suivante :

$$B = -\frac{1}{2} J \Delta^{(2)} J$$

   avec $J = I - \frac{1}{n}$, et I la matrice identité et n le nombre d'objets.
3. Extraction des n plus grandes valeurs propres positives $s_1 ... s_n$ de B avec les n vecteurs propres correspondants $u_1 ... u_n$, par décomposition en valeurs singulières (SVD).
4. On obtient la représentation X de dimension n à partir de :

$$X = u_n s_n^{1/2}$$

   avec u_n la matrice des n vecteurs propres et s_n la matrice diagonale contenant les n valeurs propres de B.

Mais le principal désavantage de la méthode classique réside dans le fait que dans beaucoup d'applications, les mesures ne sont pas quantitatives, mais qualitatives, par exemple en psychométrie lorsqu'on étudie la perception humaine. Ainsi d'autres variantes ont été développées par la suite afin de généraliser la méthode MDS et permettre l'utilisation de métriques quelconques [Kruskal, 1964a] [Kruskal, 1964b] [Shepard, 1962]. Dans ce contexte, on cherchera à minimiser une fonction de coût appelé « stress ». L'algorithme « de la plus forte pente » fut initialement utilisé par Kruskal afin de minimiser le stress [Kruskal,

1964a], mais d'autres stratégies ont été développées comme l'algorithme SMACOF (*Scaling by MAjorizing a COmplicated Function*) par Jan de Leeuw [De Leeuw, 2005] ou encore plus récemment DA-SMACOF (*Deterministic Annealing SMACOF*) afin de s'assurer que l'on converge bien vers le minimum global [Bae et al., 2010]. Ainsi il est possible, suivant la métrique et la fonction de stress utilisées de traiter des données qui peuvent être métriques (quantitatives) ou non-métriques (qualitatives, en conservant le rang seulement) de manière linéaire ou non linéaire. Dans le cas de la version métrique (et linéaire) de la méthode MDS, la fonction de coût est la suivante :

$$\textit{stress} = \sqrt{\frac{\sum_{i<j} (\Delta_{ij} - \delta_{ij})^2}{\sum_{i<j} \Delta_{ij}^2}} \tag{3.8}$$

avec Δ et δ, les matrices de distance entre les objets i et j dans l'espace original et l'espace projeté, respectivement. Une des principales faiblesses des méthodes classique et métrique est qu'elles se concentrent principalement sur le maintien des longues distances. Cependant, projeter correctement les courtes distances a un rôle essentiel dans la réduction de la dimensionnalité des données [Van Der Maaten et al., 2009]. La première méthode à mettre l'accent sur les courtes distances a été introduite par Sammon en 1969 [Sammon, 1969]. La projection de Sammon (*Sammon's non linear mapping*) adapte la fonction de coût de la méthode métrique en pondérant la contribution de chaque paire d'objets i et j par l'inverse de leur distance dans l'espace original. Ainsi, en assignant un poids plus important aux petites distances, la structure locale est bien mieux respectée que dans le cas de la version métrique ou classique. La fonction de coût de la méthode de Sammon est la suivante :

$$\textit{stress} = \frac{1}{\sum_{i<j} \Delta_{ij}} \frac{\sum_{i<j} (\Delta_{ij} - \delta_{ij})^2}{\sum_{i<j} \Delta_{ij}} \tag{3.9}$$

avec Δ et δ, les matrices de distance entre les objets i et j dans l'espace original et l'espace projeté, respectivement. La projection de Sammon est principalement utilisée dans le cadre de la visualisation de données. Cependant, quelque soit la métrique utilisée ou la fonction pour minimiser le stress, la principale limitation de la méthode MDS se situe au niveau de la complexité de calcul qui est de l'ordre de $\mathcal{O}(N^3)$) et de $\mathcal{O}(N^2)$) en ce qui concerne la consommation en mémoire. Ainsi il n'est pas possible d'appliquer cette méthode sur un volume important de données.

3.4 ISOmetric MAPping

La méthode MDS classique ou métrique a été appliquée avec succès dans de nombreux cas. Cependant cette méthode souffre du fait qu'elle essaye de maintenir toutes les distances entre les objets en ne prenant pas en compte la distribution des objets voisins. Si les données se situent sur ou à proximité d'un « espace topologique incurvé », alors cette méthode peut considérer deux points comme proches alors qu'ils sont distants (voir Figure 3.2). C'est dans ce contexte que la méthode ISOmetric MAPping (ISOMAP) a été introduite en 2000 par Tenenbaum [Tenenbaum et al., 2000]. Cette méthode utilise la distance géodésique pour définir la distance entre les objets. La distance géodésique a été au départ définie comme la longueur du chemin le plus court entre deux points sur la surface terrestre. Ce concept peut être généralisé à n'importe quelle surface mathématique, et défini comme « la longueur du chemin le plus court entre deux points qui se situent sur une surface, quand le chemin est contraint de rester sur la surface » (Figure 3.2). Lorsque les objets sont très proches, la distance euclidienne est une bonne approximation de la distance géodésique. Mais si les objets sont très éloignés, alors la distance euclidienne tend à sous-estimer la distance géodésique. Ceci est à l'origine des distorsions qui peuvent apparaitre dans l'espace projeté lorsque la méthode MDS classique ou métrique est utilisée.

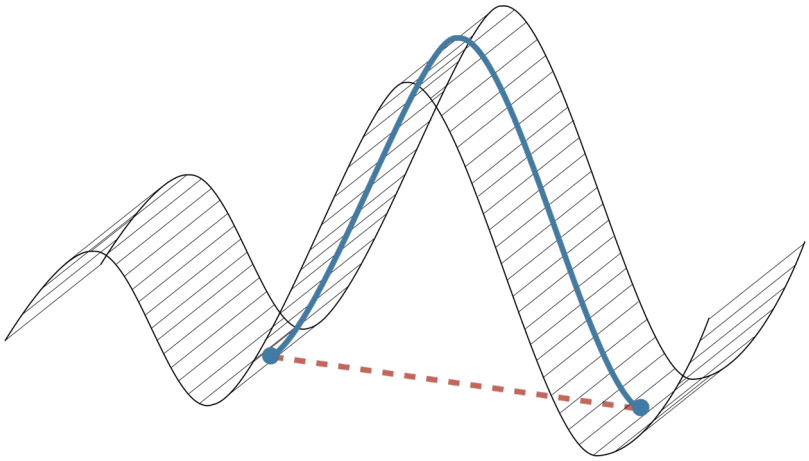


FIGURE 3.2 – La distance géodésique, représentée en bleue, entre les deux points est la longueur du chemin, qui est la distance la plus courte entre les points, se situant sur la surface. Au contraire, la distance euclidienne, représentée en rouge, est le chemin plus court entre les deux points traversant l'espace topologique.

La méthode ISOMAP est une variante de la méthode MDS utilisant la notion de distance géodésique,

afin de pouvoir capturer la non-linéarité des données. Pour cela, la méthode ISOMAP utilise l'algorithme MDS pour projeter les coordonnées des points dans un espace euclidien, mais de telle sorte que les distances mesurées dans cet espace reflètent les distances géodésiques de l'espace original. De cette façon, la méthode ISOMAP va pouvoir « dérouler » la surface non linéaire. Pour calculer la distance géodésique, il est nécessaire de construire un graphe G connectant tous les k voisins les plus proches. Ce graphe des plus proches voisins est alors utilisé pour calculer la distance géodésique entre chaque objet comme le chemin le plus court le long du graphe. Le chemin le plus court entre deux points dans le graphe donne une bonne approximation de la distance géodésique entre ces deux points (Figure 3.3). Naturellement, pour que cela puisse fonctionner, il doit y avoir un nombre suffisant d'objets pour capturer la topologie de l'espace non linéaire. La représentation de plus basse dimension n est enfin obtenue en appliquant la méthode MDS classique qui permet de réduire la dimensionnalité de la matrice de distances géodésiques. Au final, dans l'espace projeté, la distance euclidienne entre les points sera un reflet de leur distance géodésique originale.

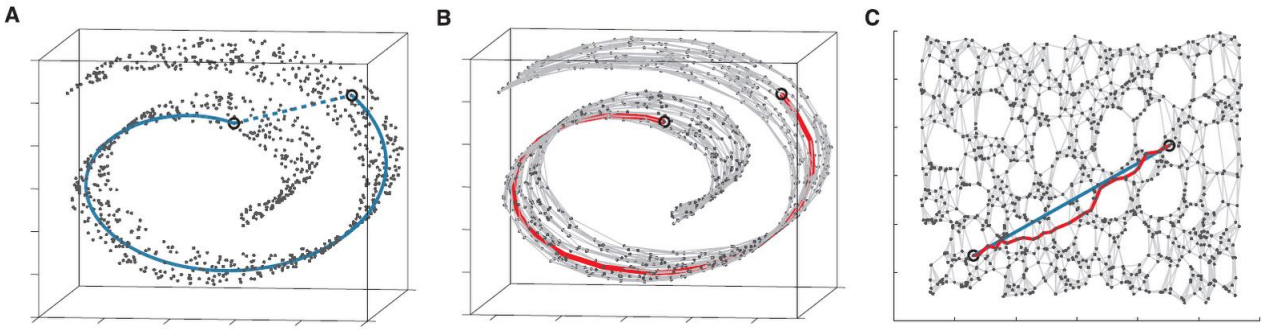


FIGURE 3.3 – Exemple de la méthode ISOMAP sur le « rouleau suisse ». Les données originales possèdent une dimensionnalité $N = 3$, mais la dimensionnalité intrinsèque est égale à $n = 2$. Chaque point est alors connecté aux k voisins les plus proches afin de créer le graphe. Les distances géodésiques entre deux points peuvent alors être estimées comme le chemin le plus court le long du graphe construit (en rouge). Au final, la méthode MDS classique peut être appliquée sur la matrice de distances géodésiques ainsi obtenue afin de projeter les points dans l'espace réduit de dimension $n = 2$. Dans cet espace réduit, les distances euclidiennes (en bleu) approximent bien les distances géodésiques (en rouge).

L'algorithme ISOMAP peut être résumé de la manière suivante :

1. Construction d'un graphe connectant chaque point aux k points voisins les plus proches.

2. Les chemins les plus courts entre les points sont calculés à l'aide de l'algorithme de Dijkstra [Dijkstra, 1959], afin de construire la matrice de distances géodésiques.
3. Application de la méthode MDS classique afin d'obtenir les coordonnées des points dans l'espace de plus basse dimension n.

Au final, la méthode ISOMAP est une méthode dérivée de la méthode MDS métrique en y ajoutant une étape préliminaire où la matrice de distance sera transformée. Cette étape a pour but de capturer la non-linéarité intrinsèque aux données en introduisant une métrique de distance appropriée. Cette étape permet ensuite d'utiliser la méthode MDS sur cette métrique et d'obtenir une représentation projetée fidèle à la distribution originale des données. La méthode ISOMAP retourne une projection dans un espace de plus basse dimension n. Cette projection peut être utilisée afin de mieux comprendre la distribution des données. Cependant, elle ne peut être interprétée de manière intuitive, comme le permet la PCA dont les vecteurs propres correspondent à des coordonnées collectives capturant la variance sous-jacente aux données. De plus, cette méthode est plus coûteuse en termes de calcul que la PCA, en héritant des mêmes problèmes que la méthode MDS, c'est-à-dire une complexité de calcul de l'ordre de $\mathcal{O}(N^3)$. Ceci rend l'analyse très difficile quand le nombre de données devient très important.

Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Structural bioinformatics

Unrolr: structural analysis of protein conformations using Stochastic Proximity Embedding

Jérôme Eberhardt, Roland H. Stote* and Annick Dejaegere*

Department of Integrated Structural Biology, IGBMC (Institute of Genetics and of Molecular and Cellular Biology), Illkirch, France.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Molecular dynamics simulations are widely used to explore conformational space of biological macromolecules. Hardware, as well as methodological, advances currently make the generation of large and complex datasets from simulations much more common. Although different clustering and dimensionality reduction methods, such as Principal Component Analysis, have been applied to MD simulations, there remains a need for improved strategies that can handle non-linear data and/or be applied to very large datasets.

Results: We implemented and tested the pivot-based version of the Stochastic Proximity Embedding method in the context of large MD data sets using the dihedral distance as a metric. The advantages of the algorithm in terms of data storage and computational efficiency will be presented, as well as the particular implementation realized. Application and testing through the analysis of a 200 ns accelerated MD simulation of a 35-residue villin headpiece will be discussed. Analysis of the 800,000 resultant conformations shows the promise of this method to organize and visualize large ensembles of conformations.

Availability and implementation: The python implementation of the pSPE algorithm is freely available on Github: https://github.com/jeeberhardt/unrolr.

Contact: annick@igbmc.fr or rstote@igbmc.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The efficient classification of large datasets of macromolecular structures is essential in order to extract relevant information from molecular dynamics simulations, and more generally, from methods of conformational exploration. Ideally, the classification should capture the heterogeneity in the underlying structures and entail minimal computational and memory requirements. To this end, we present an original combination of stochastic proximity embedding (SPE) together with a metric for structure comparison based on protein main chain dihedral angles. We present here the first application of the method to the study of 3D protein structures, exemplified through the study of the villin headpiece. We demonstrate that our implementation can efficiently organize a large number of conformations sampled in folding/unfolding simulations.

2 Approach

We implemented the pivot-based SPE method (pSPE) to represent N-dimensional protein conformational space in two dimensions. pSPE has been proposed as a way to reduce the dimensionality of large amounts of data without making any assumptions on the linearity of the underlying degrees of freedom (Agrafiotis, 2003; Rassokhin and Agrafiotis 2003). It builds on earlier work for nonlinear dimensionality reduction (Tenenbaum *et al.*, 2000; Roweis et al., 2000). It has the advantage of avoiding the storage of the complete original distance matrix between data points, as well as the storage of the distance matrix between objects in the reduced dimension. Consequently, the computational and memory complexity are linear with the growth of the number of data. To compute the distance

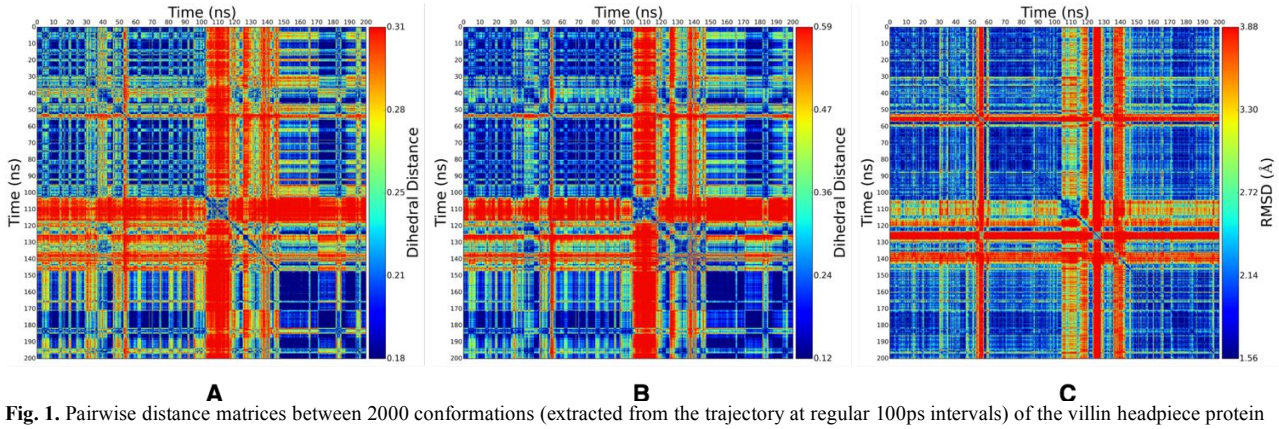


Fig. 1. Pairwise distance matrices between 2000 conformations (extracted from the trajectory at regular 100ps intervals) of the villin headpiece protein from a 200 ns aMD trajectory using as metric (A) the dihedral angle distance, (B) the RMSD and, (C) from the pSPE projection using a neighborhood cutoff $r_c = 0.23$. The color code is devised such that red corresponds to the average plus 1 standard deviation or more, and blue to average minus 1 standard deviation or less. Distances within one standard deviation are color coded as labeled.

between conformations, we used a dihedral distance (as defined by Cossio *et al.*, 2011, see Supplementary Material for details of the method). This metric has the advantage over the classical RMSD metric that it uses internal coordinates and, therefore, is insensitive to the orientation of the molecule. No superposition of structures is needed, which makes the calculation efficient and therefore suitable to compute distances between a large number of conformations. Detailed documentation and source code of Unrolr is accessible under MIT license at https://github.com/jeeberhardt/unrolr.

3 Results

The pSPE algorithm was applied to a 200 ns accelerated MD trajectory of the villin headpiece protein, a 35 residue three helix bundle domain (see Supplementary Material for details). This aMD trajectory samples folding and unfolding events. The full ensemble of conformations extracted from this simulation contains 800K configurations. A reduced set was generated by extracting structures at 100 ps intervals; this reduced set was used to construct the dihedral distance matrix shown in Fig. 1A. An analogous matrix based on the RMSD was also calculated from the same reduced dataset, see Fig. 1B. Comparison of these two show that the dihedral distance performs well as a metric compared to the more conventional RMS distance (Fig. 1B). For example, between 100 and 120 ns, the aMD explores conformations distinct from those sampled elsewhere in the trajectory, and, although this is apparent in both metrics, the dihedral distance metric shows a more pronounced difference than the RMSD. The full ensemble of conformations was subjected to a pSPE analysis, which generated the 2D representation shown in Fig. S3. From this projected data, we reconstructed the dihedral distance matrix. Likewise, the dihedral distance matrix, constructed from the pSPE projected distances (Figure 1C), reflect well the actual dihedral and cartesian (RMS) distances (compare Fig. 1C to 1A & 1B). The advantage of the pSPE projection is that it was done using the full set of 800,000 conformations, which was not practical for the other metrics. For a detailed discussion of the structural clusters, see Supplementary Material.

4 Conclusion

The pivot-based method of stochastic proximity embedding combined with a metric based on dihedral angle is advantageous in terms of memory and computational power requirement and by consequence has the ability to handle large sets of diverse conformations generated by molecular dynamics simulations. The results are comparable to direct measures, but with the advantage that orders of magnitude more conformations can be included with a computational cost that scales linearly. The tool described here is capable of organizing large ensembles of conformations, which make it particularly interesting for use in the analysis of molecular simulations of all sorts.

Acknowledgements

JE acknowledges financial support from the Ministère de l'Enseignement Supérieur et de la Recherche and from la Fondation ARC pour la Recherche sur le Cancer. We thank Dr. L. Chiron and Dr. Y. Chebaro for helpful discussions.

Funding

This work was supported by funds from the Centre National de Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), Université de Strasbourg, as well as by ANR-IAB-2011-BIP:BIP [ANR-10-BINF-0003] and ANR-10-MABX-0030-INRT [ANR-10-IDEX-0002-02]. Computing resources were provided by the Méso-centre de Calcul de l'Université de Strasbourg supported by the project EQUIP@ MESO.

Conflict of Interest: none declared.

References

Agrafiotis, D. K. (2003). Stochastic proximity embedding. Journal of computational chemistry, 24(10), 1215-1221.

Cossio, P., Laio, A., & Pietrucci, F. (2011). Which similarity measure is better for analyzing protein structures in a molecular dynamics trajectory?. Physical Chemistry Chemical Physics, 13(22), 10421-10425.

Rassokhin, D. N., & Agrafiotis, D. K. (2003). A modified update rule for stochastic proximity embedding. Journal of Molecular Graphics and Modelling, 22(2), 133-140.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500), 2323-2326.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319-2323.

Supplementary Material

Unrolr: structural analysis of protein conformations using Stochastic Proximity Embedding

J. Eberhardt, R.H. Stote*, A. Dejaegere*

Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Biocomputing team, INSERM U964 UMR 7104 CNRS, Université de Strasbourg, 1, rue Laurent Fries, BP10142, 67404 Illkirch, France

Details of the method

The pivot-based version of the Stochastic Proximity Embedding method starts with a random configuration X_0 in the low-dimensional m space (generally m = 2). Then, a point x_i in the low-dimensional space representing some given conformation is randomly selected as a pivot. With respect to this pivot point, the positions of all conformations $x_{j\neq i}$ are either modified or left unchanged using as an update criterion the comparison between the distance δ_{ij} between x_j and the pivot point x_i in the low-dimensional space m and the actual distance Δ_{ij} in the original M-dimensional space (with M equal to the number of dihedral angles used). The principle is to have the best possible match between δ_{ij} and Δ_{ij} for neighbour conformations (i.e. conformations where $\Delta_{ij} < r_c$, the neighborhood radius). For more distant conformations ($\Delta_{ij} > rc$), δ_{ij} should not be smaller than Δ_{ij}, but is otherwise not systematically optimized. The detailed algorithm (Rassokhin *et al.*, 2003) is as follows:

Require: λ, r_c and C

Generate a random configuration X_0 of dimension m
    FOR 0 TO C DO
        Select a random point x_i as a pivot
        FOR $j \neq i$ DO
            $\delta_{ij} \leftarrow$ euclidean_distance(x_i, x_j)
            $\Delta_{ij} \leftarrow$ dihedral_distance(x_i, x_j)
            IF ($\Delta_{ij} \leq r_c$) OR ($\Delta_{ij} > r_c$ AND $\delta_{ij} < \Delta_{ij}$) THEN
                $x_j \leftarrow x_j + \lambda \left(\frac{\Delta ij - \delta ij}{\varepsilon + \delta ij}\right)(x_j - x_i)$
            END IF
        END FOR
        Decrease the learning rate λ
    END FOR
RETURN X

with ε a small number to avoid division by zero, Δ_{ij} and δ_{ij}, the distances between conformations i and j in the real and low-dimensional space, respectively, r_c the neighbourhood radius and λ the learning rate. The positions of all points x_j around the pivot are updated or not according to the neighbourhood radius r_c and are corrected according to the learning rate λ. The neighbourhood radius r_c is used to approximate the geodesic distance and

defines which objects are neighbours. The learning rate λ is used to avoid oscillatory behaviour during the optimization process. It also acts as a numerical temperature to avoid local minima traps, starting generally with a value equal to 1 and decreasing linearly to 0.01 at the end. The only parameter of the pSPE method is the neighbourhood radius r_c and its choice influences significantly the representation in the low dimensional space: if r_c is too small, only local distances will be faithfully represented in low dimension, and the final representation will appear as cloud of disconnected clusters. On the contrary, if r_c is too large, the majority of distance Δ_{ij} will fall within r_c and the method will revert to a linear dimensionality reduction method (Borg *et al.*, 2005).

We therefore implemented a systematic testing of r_c, using a subset of conformations sampled at regular intervals from our MD trajectories (see below). On such a reduced set of data, the dimensionality reduction proceeds fast, and it is therefore easy to determine empirically the most adapted neighbourhood radius r_c, i.e. the value of r_c that will simultaneously minimize the stress and maximize the correlation. The stress function is given by (Rassokhin *et al.*, 2003):

$$Stress = \sum_{i<j} \frac{f(\delta_{ij}, \Delta_{ij})}{\Delta_{ij}} / \sum_{i<j} \Delta_{ij}$$

$$f(\delta_{ij}, \Delta_{ij}) = \begin{cases} (\delta_{ij} - \Delta_{ij})^2 & \text{if } (\Delta_{ij} \leq r_c) \vee (\delta_{ij} < \Delta_{ij}) \\ 0 & \text{otherwise} \end{cases}$$

with Δ_{ij} and δ_{ij}, the distances between conformations i and j in the high and low-dimensional space, respectively and r_c the neighbourhood radius. The stress value ranges from 0, if all the distances are respected, to a maximum value of 1. The correlation γ between the distances in low-dimensional space m and the real-dimensional space M is calculated as (Cui *et al.*, 2009):

$$\text{Correlation } \gamma = \frac{X^T Y / |X| - \bar{X}\bar{Y}}{\sigma_X \, \sigma_Y}$$

with $|X|$ the size of the vector X, $\bar{X}$ and σ_X are the mean and standard deviation of X, respectively. X et Y are the vectors containing all the distances in high and low dimensional space between a conformation i and all the other conformations j, respectively. As for the stress, the correlation value ranges from 0 to 1, if the distances are perfectly correlated. Thus, the objective will be to minimise the stress and, on the contrary, maximize the correlation γ while having the smallest possible neighbourhood radius r_c. However, as the number of conformation increases, the computation of the stress and the correlation γ become impracticable because of their quadratic cost. To circumvent this problem, the stress and the correlation γ are determined using randomly sampled data points until their value converge to a tolerance ε equal to 10^{-5}.

To compute the distance between conformations extracted from the aMD simulation, we used a dihedral distance. This metric has the advantage over the classical RMSD metric that it uses internal coordinates and, therefore, is insensitive to the orientation of the molecule. No superposition of structures is needed, which makes the calculation efficient and therefore suitable to compute distances between a large number of conformations. The dihedral distance between two conformations *a* and *b* is define as follows (Cossio *et al.*, 2011):

$$\text{Dihedral distance} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}(1 - \cos(\Psi_i^a - \Psi_i^b))} \tag{2}$$

where M is the number of dihedral angles, Ψ, used; a and b refer to two different conformation. The dihedral distance goes from 0, for identical conformations, to 1 when the difference between all dihedral angles is equal to 180 degrees.

Implementation

The implementation of the stochastic proximity embedding was done in Python (van Rossum, 1991) using the Anaconda distribution (available at https://continuum.io/downloads website). As input, the program requires a Protein Structure File (generated by most modeling programs, CHARMM or XPLOR format) or a Protein Data Bank file (Berman *et al*., 2000) and a trajectory file in CHARMM, GROMACS or AMBER format. The first step consists of extracting the dihedral angles between 4 consecutive Cα atoms or the Φ/Ψ dihedral angles using MDAnalysis package (Michaud-Agrawal *et al.*, 2011). For a maximum efficiency, the dihedral angles are stored in HDF5 format using h5py package (HDF Group, 2014; Collette, 2013). The pSPE method was implemented in OpenCL using pyOpenCL package in order to run on heterogenous computing systems, like CPU or GPU (Stone *et al.*, 2010; Klöckner *et al.*, 2012). Detailed documentation and source code of the pSPE is accessible under MIT licence at https://github.com/jeeberhardt/unrolr.

Details of the accelerated MD simulation

The villin headpiece is a 35 residue domain that folds into a three helix bundle (residues 3-10, 14-19 and 22-32) with a hydrophobic core composed mainly of tryptophan and leucine residues. Its folding pathways and free energy landscape have been extensively studied (Banushkina *et al.*, 2013; Jain *et al.*, 2014, Freddolino *et al.*, 2009). The molecular dynamics simulation of the villin headpiece protein was carried out using the X-ray structure (PDBID: 1YRF) available from the Protein Data Bank (Berman *et al.*, 2000) determined at 1.07 Å resolution (Chiu *et al.*, 2005). Protein residues were set to the standard CHARMM protonation states at neutral pH. The only histidine residue, HIS68, was protonated on the ε-nitrogen. The CHARMM27 force field (Mackerell *et al.*, 1998) was used for the protein. The protein was placed in a cubic box with a edge length of 43.87 Å and solvated by water represented by the TIP3P model (Jorgensen *et al.*, 1983). A conventional molecular dynamics

(cMD) simulation was first performed using the NAMD 2.10 program (Phillips *et al.*, 2005). A cutoff of 12 Å was used for the van der Waals and short-range electrostatics interactions and the long-range electrostatic interactions were computed with the Particle Mesh Ewald (Essmann *et al.*, 1995) summation method using a grid point density of 1 Å. The SHAKE (Ryckaert *et al.*, 1977) algorithm was applied to all hydrogen-heavy atom bonds. The simulations were run using a 1 fs integration time-step. Prior to running the molecular dynamics simulation the water was energy minimized for 1000 steps using the conjugate gradients method keeping the protein atoms constrained at their initial positions. Several cycles of heating of the water molecules to 298.15K, followed by energy minimization were performed, followed by another 2,000 steps of energy minimisation with all atoms free to move. The system was then gradually heated to 298.15K and equilibrated for 85 ps followed by a 10 ns NVE production run. Restarting from the final structure of 10 ns cMD simulation, an accelerated molecular dynamics (aMD) simulation was performed using a "dual-boost" as implemented in the NAMD program (Wang *et al.*, 2011). In dual-boost aMD (Hamelberg *et al.*, 2007), a total boost potential is applied to all atoms in the system in addition to the dihedral boost. The input parameters E_{dihed}, a_{dihed}; E_{total}, a_{total} were calculated from the 10 cMD simulation and set to 293 kcal/mol, 31 kcal/mol, -26,900 kcal/mol and 1812 kcal/mol, respectively. The aMD simulation used a 1 fs time step and was run for $2x10^8$ steps. Configurations were saved every 250 steps for further analysis.

Analysis of the convergence of the SPE method on the villin headpeace trajectory

As described above, the neighbourhood radius r_c is a parameter that can influence the convergence and outcome of the pSPE method. As the optimal value of r_c depends on the data set, we implemented a systematic testing of r_c in our data analysis pipeline, and chose a value of r_c that minimizes the stress and maximizes the correlation between the distances in the M dimensional space and the reduced 2D space. For this, we systematically tested values of r_c from 0.01 to 1.0 by increments of 0.01. However, for sake of efficiency, this test is not performed on the entire data set (i.e. all 800,000 conformations from the aMD), but on a reduced data set (10,000 conformations extracted every 20 ps from the trajectory). For this reduced data set, 5 successive independent runs of the pSPE algorithm are performed, with 5,000 steps of iterative optimization of the distances in the reduced 2D space for each optimisation. The correlation between the actual and projected distances increases as the neighbourhood radius r_c is increased (see Figure S1), and converges to a plateau value (γ = 0.8) for values of r_c larger than 0.23. Further increase in r_c does not improve the correlation but adversely affects the stress (see Figure S1). Likewise, for r_c larger than 0.23, the standard deviation between independent runs of the pSPE significantly decreases, indicating that the pSPE method converges to a similar solution at each independent run despite its stochastic characteristic.

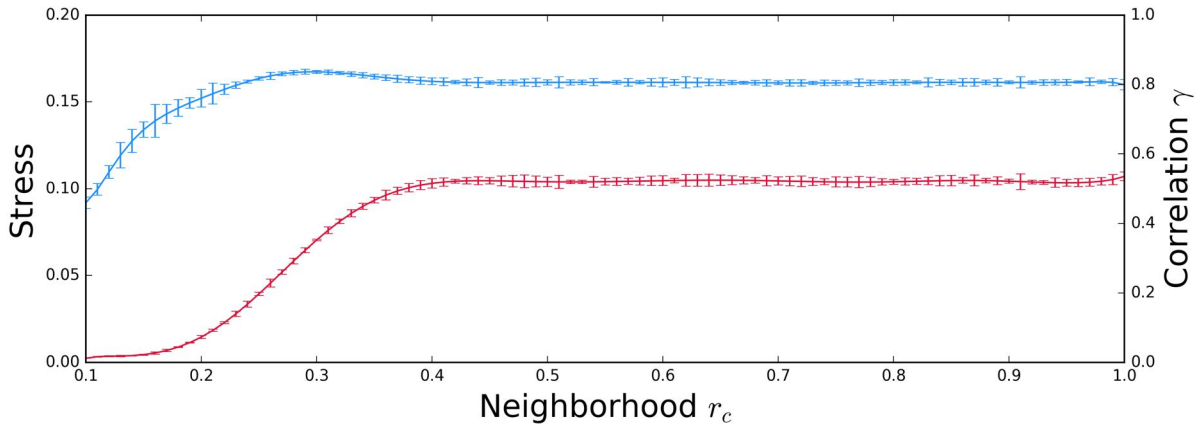


Figure S1: Evolution of the stress (in red) and the correlation γ (in blue) as a function of the neighbourhood radius r_c using 10,000 conformations from the aMD trajectory and 5,000 cycles of optimisation. The standard deviation is calculated from 5 independent runs of pSPE.

We further tested the influence of the number of pSPE optimization cycles and of the number of conformations on the correlation γ and the stress, while keeping the value of r_c fixed at 0.23 (Figure S2). It can be seen that a minimum number of 10^4 cycles of optimization is needed to obtain converged values of the correlation and stress (Figure S2). However, provided that the number of cycles is sufficient, the size of the data set does not affect the convergence, with similar results obtained whether 10^4, 5 10^4 or 10^5 conformations are used. Altogether, these data indicate that the pSPE method, when combined with the dihedral distance metric, yields stable and reproducible dimensionality reduction for large ensembles of conformations.

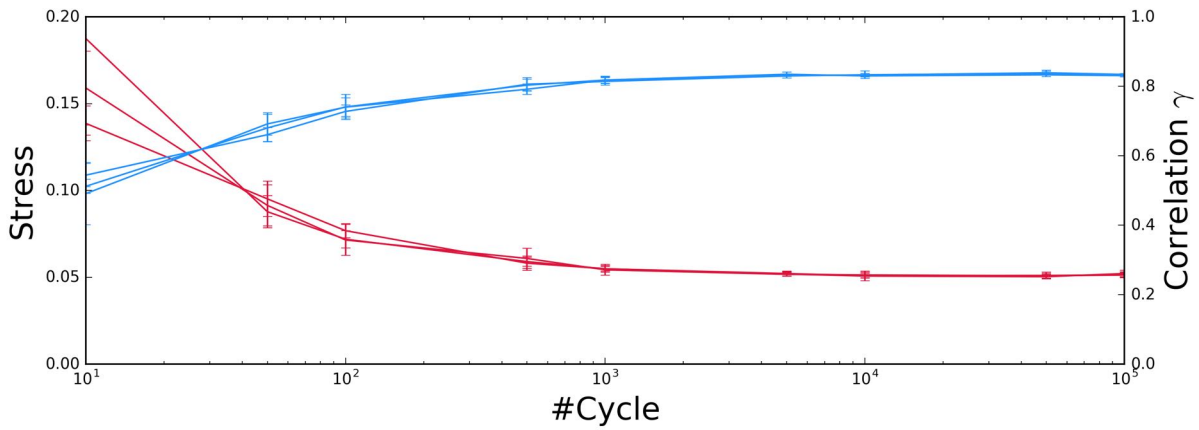


Figure S2: Evolution of the stress (in red) and the correlation γ (in blue) as a function of the number of optimization cycles (using a neighbourhood radius r_c of 0.23). The different curves correspond to different sample sizes (i.e. number of conformations): 10^4, 5 10^4, and 10^5. The standard deviation corresponds to five independent replicates of the pSPE.

Moreover, the pSPE method is computationally efficient, for example in the current implementation, the whole optimisation process takes less than 10 minutes to process 800,000 conformations containing 32 C_α dihedral angles with 50,000 optimisation cycles on a single

AMD Radeon HD 7950 GPU or approximately 45 minutes on an Intel i7-3820 3.6Ghz CPU with 4 cores.

Application to villin headpiece

From the benchmarking of the pSPE method discussed above, we analyzed the full set of conformations (800,000) using a neighbourhood radius $r_c = 0.23$ and 50,000 optimisation cycles. The pSPE method was applied on all the pseudo Cα dihedral angles (32 in total) of the 800,000 conformations extracted from the 200 ns aMD trajectory of the villin headpiece. The correlation γ and the stress of the final configuration from the optimisation process are equal to 0.82 and 0.026, respectively. In order to compare the the pSPE dimensionality reduction to previous representation of the conformational ensemble of the villin headpiece, we used the projected dimensions as reaction coordinates for building a simplified energy landscape (see Figure S3) as a function of the number of conformations populating each region of the 2-dimensional representation.

The folding process of a protein is generally associated with low probability transition events between multiple conformational states separated by energy barriers. Recent studies (Piana *et al.*, 2011; Jain *et al.*, 2014; Ernst *et al.*, 2015) analyzed the free energy landscape of the villin headpiece and identified several conformational substates: besides the native and unfolded state, a near native and different intermediate states were identified. Although our aMD trajectory of 200 ns starting from the native folded structure is not expected to provided exhaustive sampling, the pSPE/aMD analysis identified three conformational basins: the native conformation, from which the simulation was started, and two nearby basins, one where the separation from native is most visible along the 1st projected dimension (X-separated, see Figure 3) and another where the separation is mostly along the 2nd projected dimension (Y-separated, see Figure S3). Interestingly, the structures in the X-separated basin differ from the native in the structure of the C-terminal helix, residues 29 to 33, which is in agreement with the difference observed between the near native and native (Jain *et al.*, 2014), while the Y-separated structures differ mostly in the N-terminal helix, a situation more akin that observed between intermediates and native (Jain *et al.*, 2014). Finally, a range of structures are found in less populated regions of the projected space, including unfolded structures. While it is beyond the scope of the present study to make a full aMD study of the villin headpiece conformational landscape, our data (see Figure S3) clearly show that the pSPE clustering together with dihedral angles metric is a powerful and efficient method to organize and analyze large ensembles of structures generated by such molecular simulations.

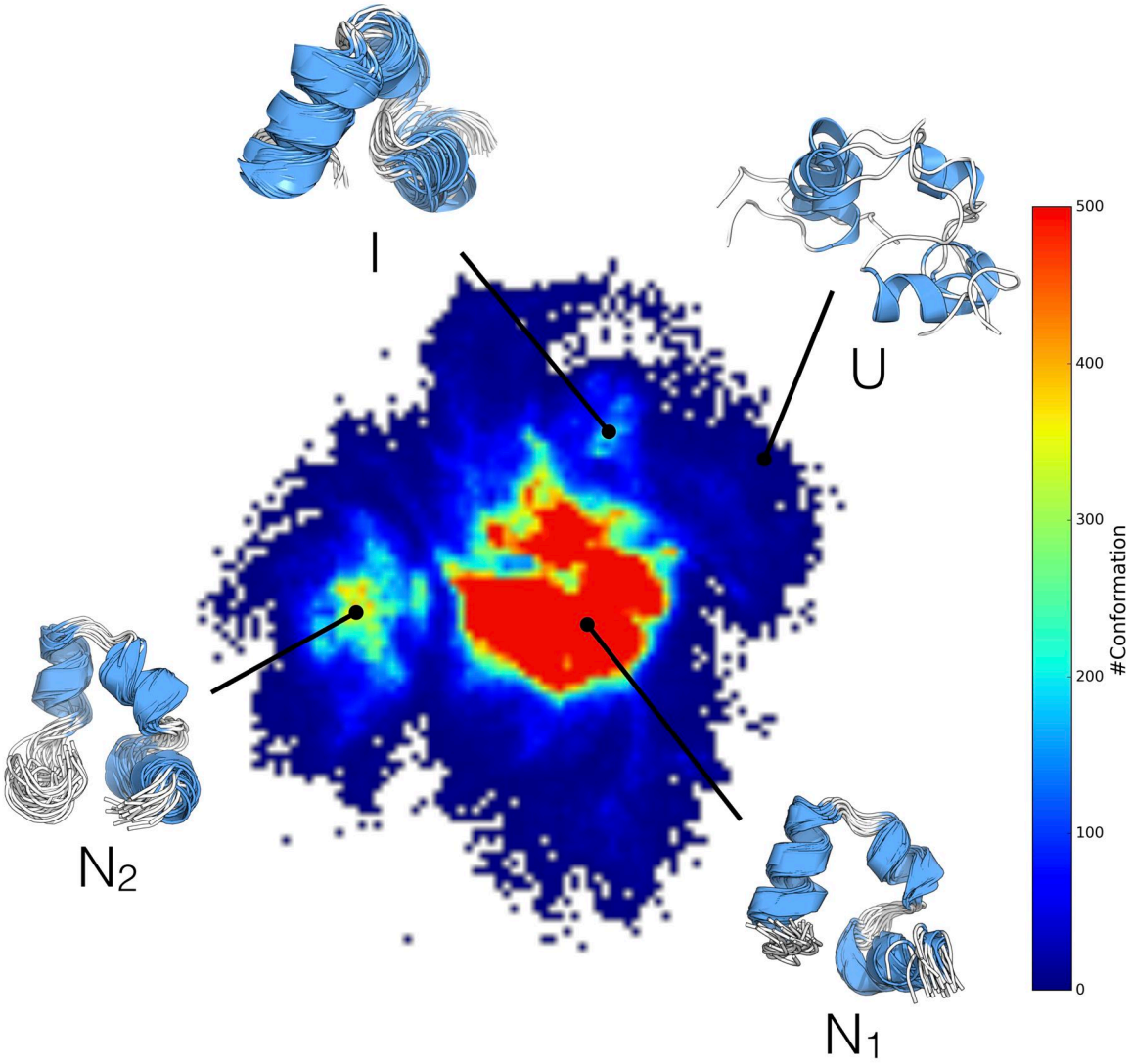


Figure S3. Simplified energy landscape of the villin headpiece, in the 2D dihedral space obtained by pSPE. The color code is displayed using a bin size of 0.025 for each dimension in the projected space. N_1 denotes the native conformation, N_2 conformations that are near-native, I, conformations resembling an Intermediate state, and U, unfolded conformations. See text for details.

Bibliography

Banushkina, P. V., & Krivov, S. V. (2013). High-resolution free-energy landscape analysis of α-helical protein folding: HP35 and its double mutant. *Journal of chemical theory and computation*, *9*(12), 5257-5266.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, *28*(1), 235-242.

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., & Davies, D. R. (2005). High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7517-7522.

Collette, A. (2013). Python and HDF5. " O'Reilly Media, Inc.".

Cossio, P., Laio, A., & Pietrucci, F. (2011). Which similarity measure is better for analyzing protein structures in a molecular dynamics trajectory?. *Physical Chemistry Chemical Physics*, *13*(22), 10421-10425.

Cui, M., Razdan, A., Hu, J., & Wonka, P. (2009). Interactive hyperspectral image visualization using convex optimization. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(6), 1673-1684.

Ernst, M., Sittel, F., & Stock, G. (2015). Contact-and distance-based principal component analysis of protein dynamics. *The Journal of chemical physics*, *143*(24), 244114.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of chemical physics*, *103*(19), 8577-8593.

Freddolino, P. L., & Schulten, K. (2009). Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophysical journal*, *97*(8), 2338-2347.

Hamelberg, D., de Oliveira, C. A. F., & McCammon, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of chemical physics*, *127*(15), 155102.

HDF Group. (2014). {Hierarchical Data Format, version 5}.

Jain, A., & Stock, G. (2014). Hierarchical folding free energy landscape of HP35 revealed by most probable path clustering. *The Journal of Physical Chemistry B*, *118*(28), 7750-7760.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, *79*(2), 926-935.

Klöckner, A., Pinto, N., Lee, Y., Catanzaro, B., Ivanov, P., & Fasih, A. (2012). PyCUDA and PyOpenCL: A scripting-based approach to GPU run-time code generation. Parallel Computing, 38(3), 157-174.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, *102*(18), 3586-3616.

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. Journal of computational chemistry, 32(10), 2319-2327.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, *26*(16), 1781-1802.

Piana, S., Lindorff-Larsen, K., & Shaw, D. E. (2011). How robust are protein folding simulations with respect to force field parameterization?. *Biophysical journal*, *100*(9), L47-L49.

Rassokhin, D. N., & Agrafiotis, D. K. (2003). A modified update rule for stochastic proximity embedding. *Journal of Molecular Graphics and Modelling*, *22*(2), 133-140.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327-341.

Stone, J. E., Gohara, D., & Shi, G. (2010). OpenCL: A parallel programming standard for heterogeneous computing systems. Computing in science & engineering, 12(1-3), 66-73.

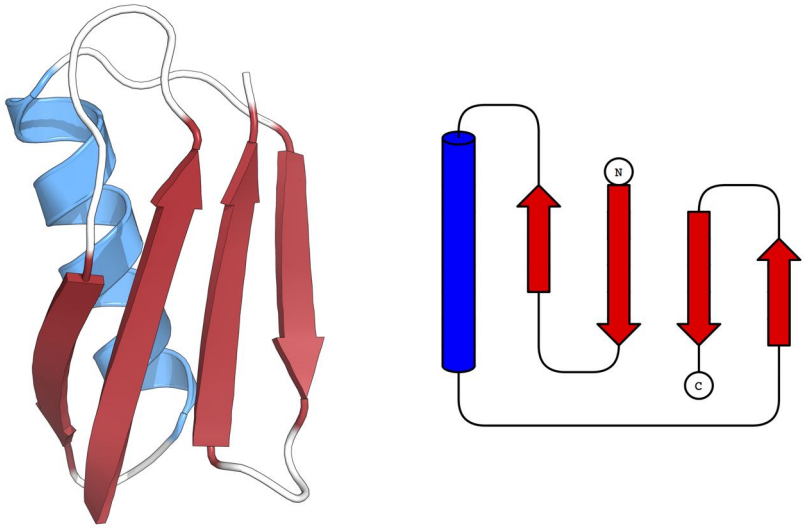
Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319-2323.

van Rossum, G., & de Boer, J. (1991). Interactively testing remote servers using the Python programming language. CWi Quarterly, 4(4), 283-303.

Wang, Y., Harrison, C. B., Schulten, K., & McCammon, J. A. (2011). Implementation of accelerated molecular dynamics in NAMD. *Computational science & discovery*, *4*(1), 015002.

3.5.3 Analyse du domaine de liaison à l'immunoglobuline de la protéine G

Outre l'exemple détaillé ci-dessus l'application de la méthode SPE à la protéine Villin, une protéine au repliement essentiellement en hélice α (ce qui est également le cas de nos protéines d'intérêt, les LBD de récepteurs nucléaires), nous avons appliqué la méthode pSPE utilisant la distance dihédrale comme métrique à une protéine de repliement majoritairement en feuillet β, le domaine de liaison à l'immunoglobuline de la protéine G (ssG) [Guarnera et al., 2009]. Cette protéine avait été étudiée précédemment[Guarnera et al., 2009], et la trajectoire nous a été aimablement fournie par les auteurs. Le domaine est composé de 56 résidus,et la séquence étudiée ne correspond pas à celle de la protéine WT, mais à celle d'une version simplifiée qui contient seulement types de résidus, glycine, alanine et thréonine. Cette simplification a été effectuée dans le but de voir si le mécanisme de repliement est équivalent au domaine sauvage. Les résultats indiquent que la protéine ssG se replie rapidement et conserve la structure secondaire de la version sauvage.[Guarnera et al., 2009]



Séquences	
Protéine G	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE
Protéine ssG	TTTTTTTTTGGTTTTTTTTGGAAAAAAAAAAAAAAGGTTTTTTTTGGTTTTTTT
Structure secondaire	- EEEEEEEE S SEEEEEEEE - S SHHHHHHHHHHHHHHH - - - - EEEETTT - EEEEE -

FIGURE 3.4 – Structure tridimensionnelle de la protéine ssG en représentation cartoon ainsi que sa topologie. Séquences de la protéine G sauvage et de la protéine G simplifiée ssG. Selon la nomenclature DSSP, les lettres E, H, S, T et le symbole - correspondent au feuillet β, l'hélice α, un coude, un tournant et une région non structurée, respectivement.

L'analyse pSPE s'est basée sur la trajectoire cMD de 15 μs à 330 K de la protéine ssG, partant d'une conformation totalement dépliée avec les angles dièdres du squelette peptidique égal à 180 °. Sur un total de 768704 conformations, seules les 750000 premières conformations ont été prises en compte dans notre analyse. De l'ensemble des conformations, 53 angles pseudo-dièdres entre 4 Cα consécutifs ont été extraits. Selon le même protocole appliqué précédemment à la trajectoire aMD de la protéine villine, le cutoff r_c optimal ainsi que le nombre de cycles d'optimisation nécessaires pour la SPE ont été déterminés par tests systématiques, et sont égaux à 0.56 et à 50.000 cycles, respectivement (Figure 3.5). Quant au stress et à la corrélation γ de la configuration finale obtenue, celles-ci sont égales à 0.019 et 0.78, respectivement.

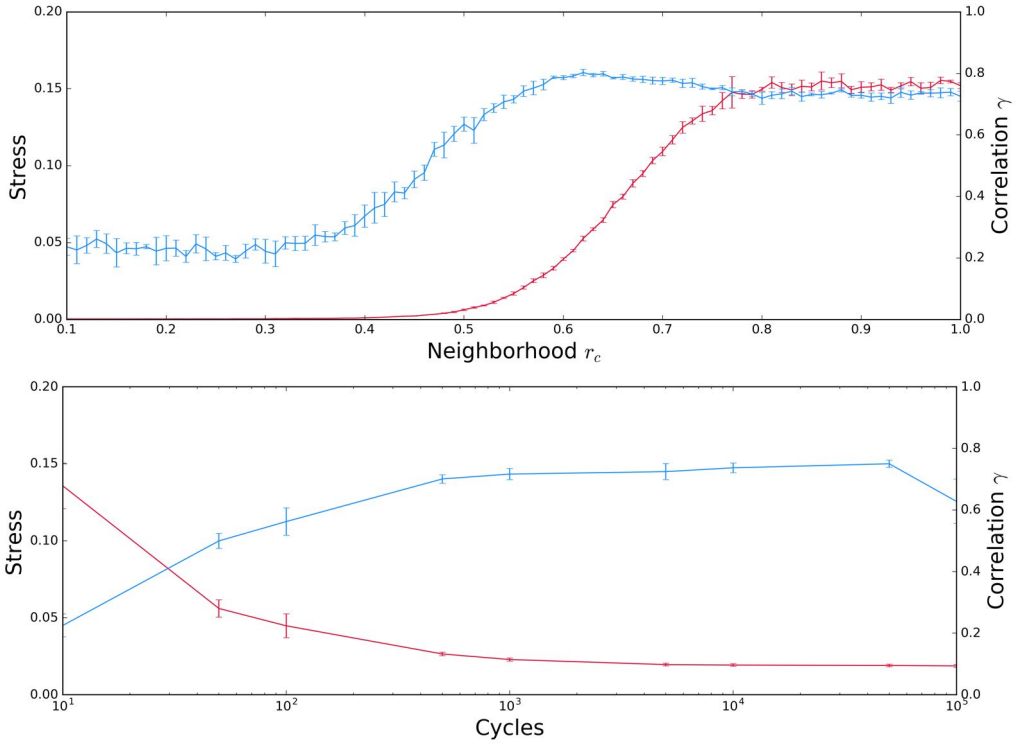


FIGURE 3.5 – Évolution du stress (en rouge) et de la corrélation γ (en bleu) en fonction du cutoff r_c, de 0.1 à 0.01 par incrément de 0.01, et 5.000 cycles d'optimisation (en haut), et aussi en fonction du nombre de cycles d'optimisation, de 10 à 100.000 cycles avec un cutoff r_c égal à 0.56 (en bas). Dans les deux cas, 5.000 conformations au total ont été utilisées à partir de la trajectoire cMD pour la détermination des paramètres

.

L'analyse de la trajectoire de la protéine de séquence simplifiée [Guarnera et al., 2009] montre qu'elle

explore un ensemble de conformations dont le contenu en structure secondaire diffère. La structure native est explorée, mais l'ensemble possède un caractère plus proche d'un « molten globule » que d'une structure native stable [Guarnera et al., 2009]. La méthode pSPE identifie clairement des conformations de structures secondaires différentes (voir Figure 3.6). Parmi celles-ci, la région la plus densément peuplée contient la forme native, c'est-à-dire dont la structure est composée d'un feuillet β de 4 brins β et d'une hélice α (voir Figure 3.4).

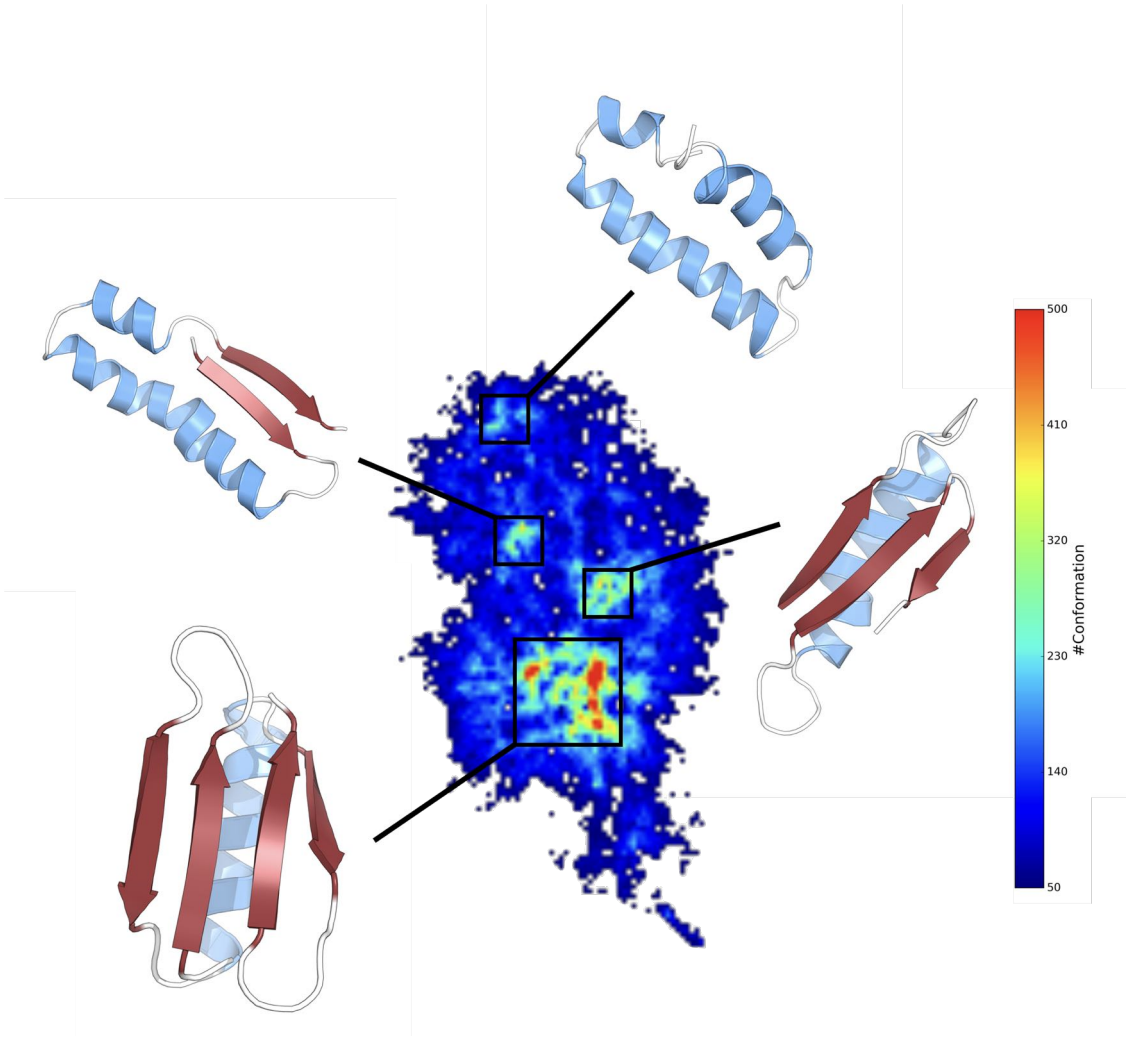


FIGURE 3.6 – Représentation de différentes topologies de la protéine ssG, ayant un contenu en structure secondaire équivalent qui correspond à la région native contenant un feuillet β de 4 brins β et d'une hélice α.

Un examen plus détaillé des conformations ayant un contenu identique en structure secondaire montre

que leur topologie peut différer de manières importe (Figure 3.6). Ce regroupement de structures ayant des topologies différentes est lié à la métrique en angles dièdres utilisée pour calculer la distance entre les conformations. Par exemple, trois conformations du bassin natif, contenant toutes un feuillet β composé de 4 brins β et d'une hélice α, (Figure 3.7) peuvent montrer des différences structurales importantes, de l'ordre de 2 à 4 Å tout en maintenant une distance dihédrale similaire (0.28). Il est à noter cependant la séquence très simplifiée de la protéine ssG lui confère une flexibilité particulièrement importante de la chaîne principale, et que l'analyse des conformations s'en ressent. Dans l'étude originale, l'utilisation d'un cut-off de 5 Å [Guarnera et al., 2009] sur les Cα de la chaîne principale s'est révélé nécessaire pour regrouper les structures contenant des structures secondaires similaires, ce qui est cohérent avec les observations détaillées des structures regroupées par la métrique en angle dièdre (Figure 3.7).

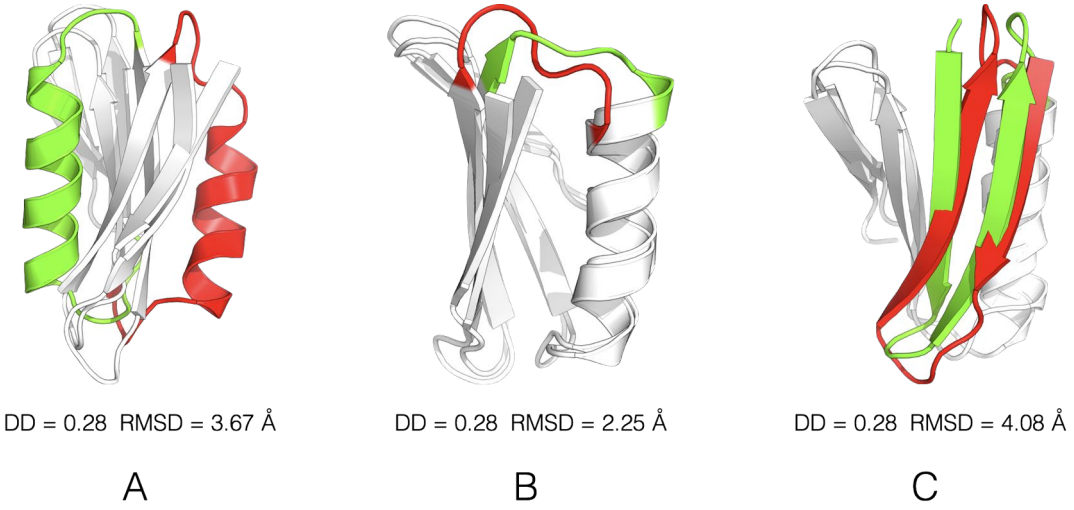


FIGURE 3.7 – Superpositions entre plusieurs conformations de la protéine ssG, possédant la même topologie, extraite de la trajectoire cMD. Les régions qui diffèrent entre les différentes conformations sont représentées en rouge et vert.

En résumé, la méthode pSPE couplée à l'utilisation de la distance dihédrale comme métrique est capable de discriminer efficacement des conformations de la protéine ssG dont le contenu en structure secondaire diffère au cours de la trajectoire cMD de 15 μs. Cependant, lorsque des changements structuraux n'impliquent que quelques angles dièdres (Figure 3.7) cette métrique peut être moins sensible que la distance RMS. Des modifications pourraient être envisagées dans ce type de problèmes, par exemple en appliquant des poids différents à chaque angle dièdre selon leur degré de variabilité entre les conformations.

Chapitre 4

Comparaison Modèles-Expériences

4.1 Déplacements chimiques RMN

4.1.1 Introduction

Le phénomène de Résonance Magnétique Nucléaire (RMN) fut mis en évidence en 1936 par Rabi [Goldman, 2004]. Mais ce n'est qu'à la fin de la seconde guerre mondiale et profitant des derniers progrès en électronique que furent développées de manière indépendante les premières méthodes pour mesurer la RMN, par les équipes de Bloch et de Purcell en 1945. Ainsi jusqu'à la fin des années 50, de nombreuses découvertes fondamentales furent faites, dont celle sur les déplacements chimiques en 1950 par Proctor et Dickinson [Dickinson, 1950] [Proctor and Yu, 1950]. C'est à partir de cette période que la RMN se popularise comme outil au sein de la communauté des chimistes pour l'étude structurale des composés chimiques. Aujourd'hui, la RMN est utilisée dans de nombreux domaines, que ce soit dans le domaine de l'imagerie médicale avec l'IRM ou bien en biologie dans des applications aussi diverses que l'étude des complexes macromoléculaires et le contrôle qualité des denrées alimentaires.

La spectroscopie par RMN est une méthode qui se base sur les propriétés magnétiques des noyaux atomiques [Jacobsen, 2007]. Brièvement, le phénomène RMN peut-être décrit de la manière suivante : si un échantillon est placé dans un champ magnétique et est sujet à des radiations (énergie) à une fréquence particulière, inférieure à 300 GHz, le noyau d'un atome peut absorber cette énergie. La fréquence de résonnance dépend du type de noyau, mais aussi de l'environnement chimique local [Jacobsen, 2007] et peut dépendre de la position du noyau dans l'espace si le champ magnétique n'est pas homogène. Cependant, dans le cadre de la spectrométrie RMN, les conditions habituelles d'expérimentation utilisent un champ magnétique le plus homogène possible (< 10 parties par milliard).

Il existe plusieurs modèles pour expliquer le phénomène de résonance RMN, le modèle classique et le modèle quantique. Dans le cas du modèle classique, les noyaux chargés positivement se comportent comme s'ils tournaient autour d'un axe, couramment appelé spin nucléaire (Figure 4.1). Cette charge en rotation crée un moment magnétique qui s'aligne avec le champ magnétique externe B_0. La rotation du noyau est caractérisée par son moment angulaire ω, et la force du moment (*torque*) générée par le champ magnétique externe B_0 va induire un mouvement de précession autour de l'axe z, ou autrement appelé fréquence de résonnance de Larmor ν_0. Certains spins vont soit tourner autour de l'axe z positif ou au contraire autour de l'axe z négatif. De même qu'ils peuvent passer d'une orientation à l'autre spontanément.

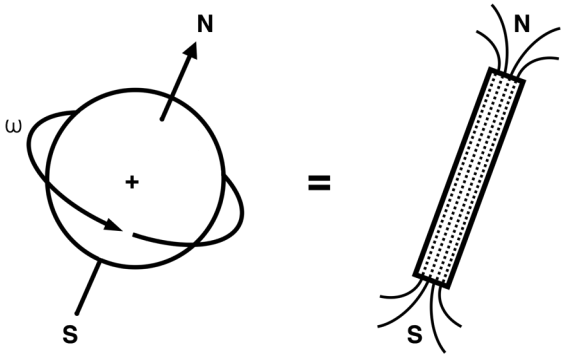


FIGURE 4.1 – Le noyau chargé positivement tournant autour d'un axe avec une vitesse angulaire ω va créer un champ magnétique B, semblable à une barre aimantée dont l'axe va coïncider avec l'axe de rotation du spin. Adapté de [James, 1998].

Selon le modèle quantique, cette description correspond à deux états différents possible, un état de basse énergie, dans le sens du champ magnétique externe B_0, et un de haute énergie, dans le sens contraire (Figure 4.2). Et seuls les noyaux ayant un spin nucléaire $I = 1/2$ (non entier) possèdent deux états différents (sans état intermédiaire).

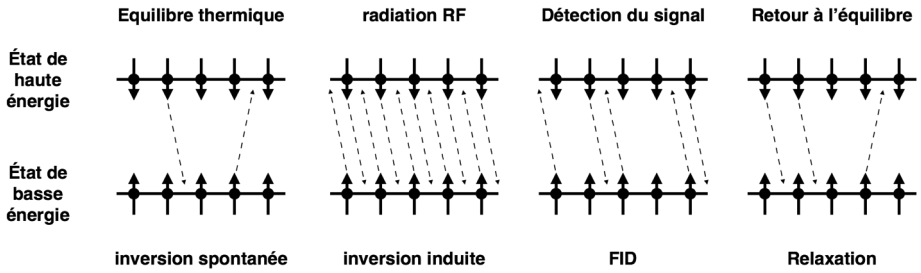


FIGURE 4.2 – À l'équilibre, des inversions de spins peuvent avoir lieu de manière spontanée entre les deux niveaux d'énergie. Lorsque l'échantillon est sujet à des radiations radiofréquences à une fréquence particulière, égalent à l'énergie séparant les deux états, des transitions vont être induites du fait de l'absorption de l'énergie par les spins. On dit alors que le système « résonne ». Après l'impulsion radiofréquence, un signal va être enregistré, correspondant la « décroissance libre de l'induction » (FID). À terme, le système va relaxer et revenir à l'état d'équilibre initial. Adapté de [James, 1998].

Il est important de noter que seuls les isotopes ^{1}H, ^{13}C, ^{15}N, ^{19}F et ^{31}P possèdent un spin nucléaire $I = 1/2$ et que les isotopes les plus communs (^{12}C, ^{14}N et ^{16}O) n'en possèdent pas. Les spins nucléaires

peuvent passer d'un état à l'autre, cependant cet évènement est plutôt rare. Cependant, si l'énergie fournie par irradiation de fréquence radio correspond à la différence d'énergie ΔE entre les états, alors plus d'évènements de transition vont avoir lieu, de l'état de basse énergie vers l'état de haute énergie et vice-versa.

La distribution des spins entre les deux états dépend du type de noyau et de la force du champ magnétique externe B_0. Le signal détectable par spectroscopie RMN est lié à la différence de population entre les deux états de spins. L'envoi d'impulsion radiofréquence sur le système provoque des transitions entre les états de spins, et le signal résultant peut être détecté par une antenne, placée perpendiculairement au champ magnétique externe B_0. Lorsque l'impulsion radiofréquence cesse, s'en suit un phénomène de relaxation où les spins nucléaires vont retourner à leur état d'origine. Le signal obtenu sera visualisé sous la forme d'une « décroissance libre de l'induction » (*free induction decay*) (FID), et la fréquence de résonnance par transformation de Fourrier. L'énergie nécessaire pour induire un changement d'état et obtenir un signal RMN est juste la différence d'énergie entre les deux états, dépendant du champ magnétique externe B_0. Elle est donnée par la relation suivante :

$$\Delta \mathrm{E} = \frac{h\gamma}{2\pi} B_0 \tag{4.1}$$

avec h la constante de Planck (6.63×10^{-27} erg sec), γ le rapport gyromagnétique dépendant du type de noyau et B_0, le champ magnétique externe.

Selon le modèle de Bohr, l'énergie d'une onde électromagnétique est liée à sa fréquence, $E = h\nu$, ainsi l'équation précédente peut être écrite de la manière suivante :

$$\nu_0 = \frac{\gamma}{2\pi} B_0 \tag{4.2}$$

avec ν_0, la fréquence de résonance de Larmor, γ le rapport gyromagnétique et B_0, le champ magnétique externe. La fréquence de résonance de Larmor ν_0 et la fréquence angulaire de résonance de Larmor ω_0 sont reliées par la relation suivante $\omega_0 = 2\pi\nu_0$. Ainsi des noyaux de type atomique différent vont avoir des rapports gyromagnétiques γ différents et résonner à des fréquences différentes. En outre, le champ magnétique ressenti par les spins nucléaires ne sera pas le champ externe B_0, mais un champ effectif B_{local} qui dépendant de l'environnement chimique local (Figure 4.3). Cet effet est lié à la présence des électrons qui soumis au champ magnétique externe B_0 vont induire un léger champ magnétique opposé à B_0, environ 1 million de fois inférieures à ce dernier. On parle alors de phénomène de blindage. En plus des électrons

appartenant au noyau directement, ceux autour participeront aussi à son blindage, ou vont au contraire le déblinder et augmenter le champ magnétique local.

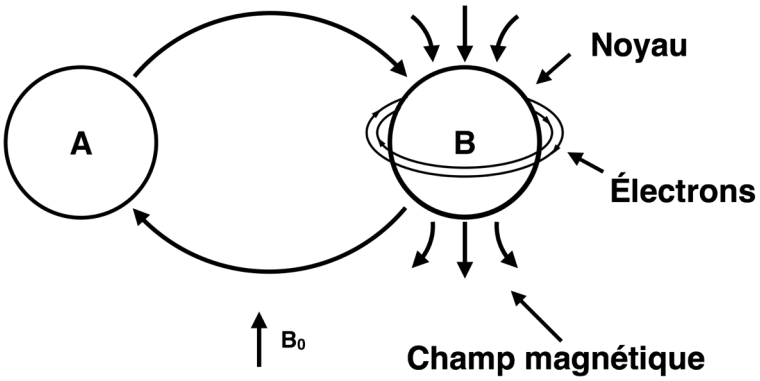


FIGURE 4.3 – Les électrons autour du noyau vont induire un léger champ magnétique opposé au champ magnétique B_0. De même que les noyaux situés à proximité vont avoir un effet sur le champ magnétique induit par les électrons. Adapté de [James, 1998].

En conséquence, le champ magnétique externe B_0 perçu par le noyau atomique sera alterné et caractéristique de l'environnement électronique local. Par exemple, les protons amides HN de deux résidus tryptophane dans une protéine repliée vont résonner (absorber l'énergie) à deux fréquences différentes sachant qu'ils sont dans deux environnements différents. La fréquence de résonance ν_0 sera alors :

$$\nu_0 = \frac{\gamma}{2\pi} B_{\text{local}} = \frac{\gamma}{2\pi} B_0 (1 - \sigma) \tag{4.3}$$

avec B_{local} le champ magnétique local et σ la constante de blindage ou le tenseur d'écrantage. Si la fréquence de résonnance de Larmor ν_0 est de l'ordre de quelques MHz, alors la différence de fréquence de résonnance entre deux atomes de même type sera de quelques Hz. Ainsi la différence entre une fréquence de résonnance de référence et celle obtenue sera ce qu'on appelle le déplacement chimique. Mais au lieu d'être exprimé en Hz, le déplacement chimique est exprimé en partie par million (ppm). Cette mesure est indépendante du champ magnétique externe utilisé et est obtenue selon la relation suivante :

$$\delta = \frac{\nu - \nu_{\text{référence}}}{\nu_{\text{travail}}} \times 10^6 \tag{4.4}$$

avec ν, la fréquence d'intérêt en Hz et $\nu_{\text{référence}}$, la fréquence d'un noyau de référence et ν_{travail}, la fréquence appliquée à l'échantillon.

4.1.2 Déplacements chimiques secondaires

Du fait de la forte corrélation entre les déplacements chimiques et la structure locale, il est possible de localiser et d'identifier le type de structures secondaires présentes au sein d'une protéine. Les déplacements chimiques secondaires $\Delta\delta$ seront calculés à partir de la relation suivante qui relie les déplacements chimiques observés $\delta_{\text{observé}}$ aux valeurs de référence $\delta_{\text{référence}}$ de chaque acide aminé obtenues expérimentalement en mesurant les déplacements chimiques de peptides non structurés (random coil) [Wishart et al., 1992] [Wishart and Sykes, 1994] :

$$\Delta\delta \;=\; \delta_{\text{observé}} - \delta_{\text{référence}} \tag{4.5}$$

Le déplacement chimique secondaire $\Delta\delta$ est dépendant de la structure secondaire de la protéine. Les atomes $^{13}\text{C}\alpha$ dans les hélices α ont tendance à avoir des déplacements chimiques secondaires $\Delta\delta$ positifs et les atomes $^{13}\text{C}\alpha$ dans les feuillets β, des valeurs négatives. Le comportement des atomes $^{13}\text{C}\beta$ est l'exact opposé des $^{13}\text{C}\alpha$. Le déplacement chimique des atomes ^{1}HN, ^{15}N et ^{13}C est aussi influencé par la structure secondaire, mais ce lien n'est pas aussi direct que pour les atomes $^{13}\text{C}\alpha$ et $^{13}\text{C}\beta$. Dans le cas d'une région non structurée, comme les boucles, les déplacements chimiques secondaires $\Delta\delta$ seront proches de zéro.

4.1.3 Prédiction des déplacements chimiques

Sachant que les déplacements chimiques sont basés sur l'environnement dans lequel se trouve l'atome, il serait très intéressant de pouvoir prédire leur valeur à partir d'une structure cristallographique. Des valeurs calculées de déplacements chimiques peuvent être obtenues par des méthodes de chimique quantique, cependant cette approche est gourmande en termes de puissance et de temps de calcul. Elle peut être appliquée à des systèmes de petite taille, mais pas à des complexes protéiques. Une approche alternative est d'établir des corrélations empiriques entre la structure secondaire observée sur des structures expérimentales de protéines obtenues avec une bonne résolution (par exemple par cristallographie à résolution < 2 Angström) et les déplacements chimiques mesurés par RMN sur ces mêmes protéines. À partir de telles corrélations empiriques, il est ensuite possible de prédire les déplacements chimiques à partir de structures tridimensionnelles, ce qui ouvre la voie à diverses applications, telles que la détermination, l'affinement et la validation de structures obtenues à l'aide de méthodes d'échantillonnage conformationnel *in silico*, comme Monte-Carlo ou encore par dynamique moléculaire.

Afin de prédire les déplacements chimiques δ à partir d'une structure, nous utiliserons SHIFTX2 développé par l'équipe du Dr. David S. Wishart [Han et al., 2011]. La prédiction des déplacements chimiques δ sera faite pour les éléments ^{13}Cα, ^{13}Cβ, ^{13}C', ^{1}HN, et ^{15}N. SHIFTX2 est la combinaison de méthodes basées sur la structure avec SHIFTX+ [Neal et al., 2003] et de méthodes basées sur la séquence par SHIFTY+ [Wishart et al., 1997]. Les méthodes basées sur la séquence obtiennent généralement de meilleurs résultats que les méthodes basées sur la structure [Wishart and Nip, 1998], lorsqu'une séquence homologue (+ de 40 % d'identité) est disponible dans la banque de données construite à partir des données disponibles dans REfDB [Zhang et al., 2003], elle-même dérivée de la BioMagResBank (BMRB) [Seavey et al., 1991]. Cependant même si une séquence identique (ou homologue) est trouvée, deux séquences peuvent adopter des conformations totalement différentes (repliée ou non repliée). Dans ce cas la prédiction reposant uniquement sur la séquence est fausse. Il est donc intéressant de recouper les informations obtenues par les méthodes séquences avec celles obtenues par des méthodes basées sur la structure. Les méthodes basées sur la structure utilisent en effet directement les informations structurales disponibles pour prédire les déplacements chimiques, indépendamment de données d'homologie. Dans le programme SHIFTX2, la combinaison de ces deux méthodes est faite de telle sorte que si les prédictions SHIFTY+ et SHIFTX+ sont concordantes, alors les résultats de SHIFTY+ prédomineront. Dans le cas contraire, si une grande différence est observée entre les prédictions SHIFTY+ et SHIFTX+, les résultats seront combinés de sorte à donner un poids plus important aux résultats issus de SHIFTX+ dans les régions où la différence est notable avec les prédictions SHIFTY+ basées sur la séquence seule (régions sans séquence homologue ou bien région dépliée).

Dans notre étude, la situation est un peu particulière, dans la mesure où nous étudions une seule séquence, mais qui explore des ensembles de conformations pouvant être assez différents. Nous avons donc choisi d'utiliser uniquement les déplacements chimiques δ obtenus à partir de la structure (SHIFTX+) et non de la séquence (SHIFTY+), afin que la comparaison entre les déplacements chimiques calculés à partir des simulations de dynamique moléculaire et à partir des structures cristallographiques soit plus significative. Dans le cas SHIFTX+, la prédiction du déplacement chimique δ est obtenue à partir de la somme de plusieurs contributions :

$$\delta_{\text{calc}} = \delta_{\text{coil}} + \delta_{\text{RC}} + \delta_{\text{EF}} + \delta_{\text{HB}} + \delta_{\text{HS}} \tag{4.6}$$

avec δ_{calc}, le déplacement chimique δ prédit, δ_{coil} la valeur de référence du déplacement chimique δ dans un peptide non structuré [Wishart et al., 1992] [Wishart and Sykes, 1994], δ_{RC} la contribution liée à

la présence de résidus aromatique, δ_{EF} l'effet électrostatique dû aux atomes CO, O, OD_n, OE_n et N sur les C_α et les atomes d'hydrogène, δ_{HB} les liaisons hydrogène et enfin la dernière contribution δ_{HS} appelée *Empirical chemical shift hypersurfaces*. Ce denier terme n'est qu'autre qu'une table regroupant toutes les valeurs de déplacements chimiques δ associés à chaque résidu selon plusieurs critères géométriques, tels que les angles de torsion Phi/Psi, la structure secondaire, la taille de certaines liaisons hydrogène et leur présence ou leur absence. Toutes ces contributions sont pondérées selon leur influence afin de correspondre au mieux aux données expérimentales et varient en fonction du type atomique.

4.1.4 Comparaison avec les données expérimentales

SHIFTX+ sera utilisé sur l'ensemble des conformations extraites des trajectoires de dynamique moléculaire, mais aussi sur les structures cristallographiques. Les conformations seront extraites à l'aide du module Python MDAnalysis [Michaud-Agrawal et al., 2011] et REDUCE [Word et al., 1999] sera utilisé afin de placer les atomes d'hydrogène manquants sur les structures cristallographiques. Dans le cas de la forme libre et liée à 9cRA du LBD de RXRα, la prédiction des déplacements chimiques δ des atomes ^{1}HN, ^{13}Cα, ^{13}Cβ, ^{13}C et ^{15}N à l'aide de SHIFTX+ se fera avec les paramètres suivants : un pH égal à 7.4 et une température de 298.15K, en accord avec les conditions expérimentales [Lu et al., 2006]. Les valeurs des déplacements chimiques δ (δMD) obtenus à partir des simulations de dynamique moléculaire seront déterminées en moyennant chaque déplacement chimique δ de chaque conformation de façon linéaire.

Les résultats expérimentaux et prédits seront comparés par le biais des déplacements chimiques secondaires par résidu. Afin de quantifier le lien qui existe entre les données expérimentales et les données prédites, nous calculerons le RMSD, exprimé en ppm.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\Delta\delta_{\text{pred},i} - \Delta\delta_{\text{exp},i})^2} \tag{4.7}$$

avec $\Delta\delta_{\text{pred}}$ et $\Delta\delta_{\text{exp}}$ les déplacements chimiques RMN prédits et expérimentaux, respectivement, et N le nombre de déplacements chimiques RMN total. Ainsi il sera possible d'évaluer si les déplacements chimiques prédits à partir d'une méthode d'échantillonnage sont en meilleur accord avec l'expérience que ceux calculés sur la/les structure(s) cristallographique(s). Sachant que SHIFTX2 est capable de prédire les déplacements chimiques avec un RMSD de 0.4412 ppm, 0.5163 ppm, 0.5330 ppm, 1.1169 ppm et 0.1711 ppm pour les atomes du squelette peptidique ^{13}Cα, ^{13}Cβ, ^{13}C, ^{15}N et ^{1}HN, respectivement, entre les déplacements chimiques prédits et observés [Han et al., 2011].

4.2 Échange Hydrogène/Deutérium

4.2.1 Introduction

Développée à l'origine en 1954 par Linderstorm-Lang afin de localiser les structures secondaires formées au sein des protéines [Hvidt et al., 1954] [Hvidt, 1955], cette méthode devait apporter la preuve de l'existence de structures en hélice α et feuillet β proposées par Pauling et Corey en 1951 [Pauling et al., 1951] [Pauling and Corey, 1951]. Malheureusement, la tâche fut bien plus complexe que prévu et en 1960 arriva la première structure cristallographique déterminée par Xray de la myoglobine et de l'hémoglobine à 5.5 et 2 Angströms de résolution [Perutz et al., 1960] [Kendrew et al., 1960], respectivement, apportant la preuve de l'existence de l'hélice α. Bien que cette méthode ait permis d'apporter quelques réponses au mécanisme de repliement des protéines, notamment que leur stabilité dépend de la température et que le processus de dénaturation est réversible, il s'en est suivi une période de creux pendant laquelle le développement de nouvelles méthodes a cessé [Baldwin, 2011]. L'introduction d'une nouvelle méthode de mesurement de l'échange HX, par SW Englander en 1963, se basant sur l'utilisation du tritium H^3 et d'une colonne de gel filtration de type Sephadex ®a permis un renouveau de la discipline [Englander, 1963]. Le tritium était utilisé pour suivre le « mouvement » des hydrogènes et quant à la colonne Sephadex ®, celle-ci permettait de séparer très facilement la macromolécule triatée du solvant triatée afin de mesurer le tritium échangé. Auparavant, il était nécessaire de lyophiliser (cryodessiccation) la macromolécule afin d'extraire le solvant, puis de la resolubiliser afin de suivre la cinétique d'échange HX [Linderstrom-Lang, 1955]. En plus d'être fastidieuses à mettre en oeuvre, les différentes opérations (cryodessiccation, chauffage (60° C), resolubilisation) pouvaient grandement altérer la structure de la macromolécule. Par la suite, les différents régimes d'échange hydrogène, EX1 et EX2, furent mis en évidence [Hvidt and Nielsen, 1966]. Ainsi dans les années 1970, la compréhension des mécanismes de base de l'échange hydrogène fut complété, principalement par l'utilisation de la spectroscopie RMN, avec la quantification des effets inductifs et le caractère bloquant de certaines chaines latérales voisines sur la vitesse d'échange [Sheinblatt, 1970] [Molday et al., 1972] [Bai et al., 1993]. Aujourd'hui, l'observation de l'échange Hydrogène/Deutérium (HDX) mesuré par spectrométrie de masse (MS) est devenue une méthode de choix afin d'étudier l'aspect dynamique des protéines en solution. Bien que traditionnellement, la spectroscopie 2D par RMN [Krishna et al., 2004] ait été utilisée pour répondre à ces questions, elle a été progressivement supplantée par la méthode HDX/MS du fait, par exemple de l'absence de contraintes quant à la taille du système étudié, virtuellement illimité, mais aussi de la plus grande facilité de mise en oeuvre des méthodes utilisant la MS [Konermann et al.,

2011]. Appliquée de maintes reprises aux complexes macromoléculaires biologiques, l'échange HX a permis d'étudier le repliement ainsi que la stabilité des protéines [Mayo and Baldwin, 1993] [Jennings and Wright, 1993] [Chyan et al., 1993] [Kim et al., 1993] [Bai et al., 1994] [Bai et al., 1995] [Chamberlain et al., 1996], de déterminer leur structure [Lührs et al., 2005] [Lu et al., 2007], d'étudier leur flexibilité et leur dynamique [Závodszky et al., 1998] [Závodszky et al., 1998] [Hernandez et al., 2000] et ainsi que d'étudier l'interaction protéine-protéine et des complexes protéine-ligand [Lee et al., 2004] [Mandell et al., 1998].

4.2.2 Mesure expérimentale de l'échange H/D

L'approche par protéolyse suivie de l'analyse par spectrométrie de masse (HDX/LC-MS ou HDX/MS) est conduite de manière continue. L'incorporation en deutérium au sein d'une protéine est suivie en fonction du temps d'exposition au D_2O. Ainsi, le but de ces expériences est de suivre les changements dus par exemple à la présence d'un ligand. Cette méthode permet d'obtenir des informations sur le comportement de régions spécifiques de la protéine [Konermann et al., 2013].

Ainsi, à un intervalle de temps régulier s'en suit l'étape de désactivation (*quenching*) en abaissant le pH à 2.5 et la température à 0 ° Celsius afin de limiter au maximum le contre-échange D/H durant l'étape de protéolyse par une enzyme de type pepsine. Cependant, on ne conservera que les protons amides HN échangés et non ceux au niveau des chaines latérales, ceux-ci contre-échangeant très rapidement. Les peptides résultants de la protéolyse sont ensuite dessalés et séparés par chromatographie liquide (LC) et analysés par spectrométrie de masse. La pepsine casse les liaisons peptidiques préférentiellement entre les résidus hydrophobes ou au niveau des résidus aromatiques. Néanmoins, il est nécessaire de confirmer l'identité de chaque peptide en parallèle. Le taux de deutération relatif pourra ainsi être obtenu pour chaque région de la protéine au cours du temps de la manière suivante :

$$\text{Taux de deutération } \% = \frac{m(t) - m_0}{m_{100} - m_0} \tag{4.8}$$

avec $m(t)$ la masse de peptide, tandis que m_0 et m_{100} sont les masses de référence pour le peptide non deutéré ou totalement deutéré, respectivement. La figure 4.4 résume une expérience d'échange H/D classique avec une mesure continue du taux d'incorporation au cours du temps.

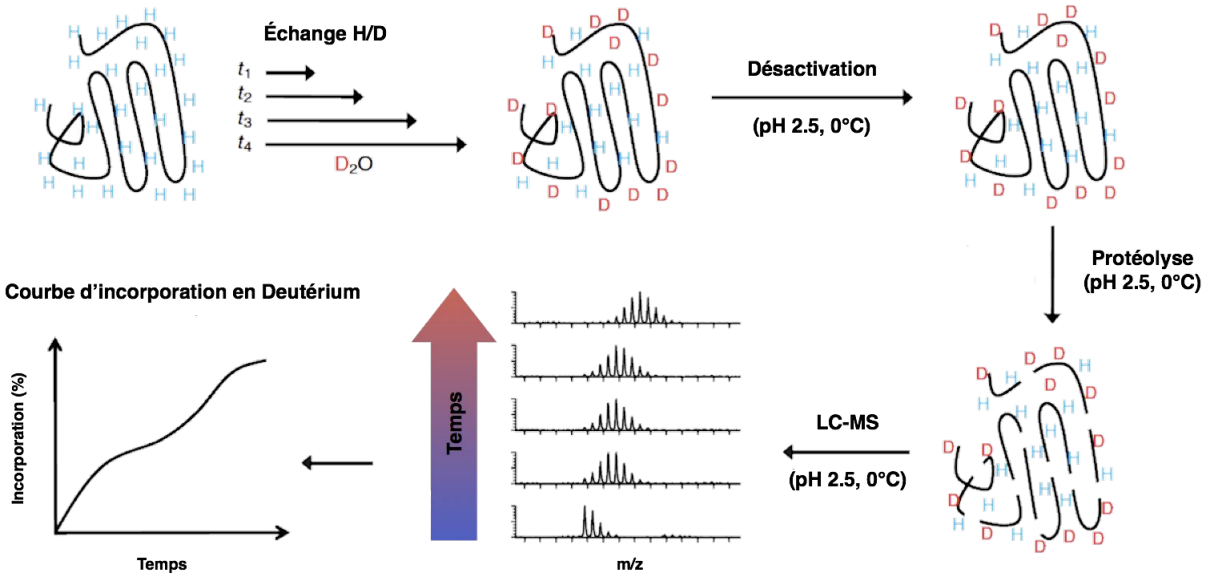


FIGURE 4.4 – Résumé d'une expérience d'échange H/D classique avec mesure continue dans le temps. L'expérience débute par une solution contenant une protéine en solution qui sera mélangée à une solution tampon contenant du D_2O. L'échange H/D sera désactivé (*quenching*) à plusieurs points t. L'échange H/D global peut être soit mesuré directement spectrométrie de masse (MS) à partir de la protéine intacte, ou bien localement après protéolyse de la protéine et analyse par MS des peptides protéolysés. Adapté de [Weis, 2016].

4.2.3 Mécanisme d'échange HX

4.2.3.1 Général

L'échange de protons entre l'eau et des molécules en solution peut procéder de différentes manières selon le catalyseur impliqué, H_3O^+ et OH^- incluant aussi H_2O. Dans le cas de l'étude de l'échange HX appliqué aux macromolécules biologiques, telles que les protéines, ce sont les protons amides HN du squelette peptidique qui sont impliqués, ainsi que les groupements polaires des chaines latérales. Tous ces protons liés à O, N ou S peuvent échanger facilement avec le solvant, car ils forment des liaisons hydrogène avec les catalyseurs H_3O^+, OH^- et H_2O. Ainsi le mécanisme général de transfert de protons peut être décrit de la manière suivante [Hvidt, 1964] :

$$AH + B \underset{k_{-D}}{\overset{k_D}{\rightleftharpoons}} (AH...B \underset{}{\overset{k_c}{\rightleftharpoons}} A...HB) \underset{k_D}{\overset{k_{-D}}{\rightleftharpoons}} A + HB \tag{4.9}$$

avec k_D la constante de vitesse de formation d'une liaison hydrogène entre les espèces AH et B, formant le complexe AHB.

L'équilibre du complexe est donné par $k_c = 10^{\Delta pK}$ avec $\Delta pK = pK_B - pK_A$. Ainsi la fraction de complexe allant dans le sens de la réaction sera :

$$\frac{k_c}{(k_c + 1)} = \frac{10^{\Delta pK}}{(10^{\Delta pK} + 1)} \tag{4.10}$$

Après dissociation du complexe AHB, on peut soit avoir régénération des réactants d'origine AH et B, soit transfert de proton et formation des nouvelles espèces A et BH. Dans le cas où le transfert de proton est énergétiquement favorable, de l'acide le plus fort vers le plus faible, $\Delta pK > 1$, alors la fraction de complexe allant dans le sens de la réaction ($AH \rightarrow BH$) sera proche de 1 et le transfert se fera donc à vitesse maximale, égale à k_D. Concernant les macromolécules biologiques, le pKa des groupements polaires des chaines latérales des protéines est inférieur à 15.7, ainsi ces protons peuvent être échangés très rapidement par OH^-. À chaque fois qu'un complexe se forme, il y a transfert de proton. Cependant, le transfert du proton amide HN du squelette peptidique, avec un pKa de 18.5 (comparé à 15.7 pour OH^-) est plus lent, ainsi on aura seulement un transfert sur 1000 environ qui sera réussi. Dans la situation où le catalyseur est H_3O^+, le transfert de proton est encore plus lent, c'est pour cela que l'échange HX est largement dominé par OH^- jusqu'à un pH égal à 3. En conséquence, les protons qui échangent le plus lentement au sein d'une protéine sont les protons amides HN, excepté certains protons de chaines latérales dans des conditions particulières. De manière générale, la vitesse de transfert de proton est donnée par la relation suivante :

$$k_{\text{ch}} = k_D \frac{10^{\Delta pK}}{(10^{\Delta pK} + 1)} \tag{4.11}$$

Cependant dans le cas du proton amide HN, le transfert du proton est toujours l'étape limitante, car $\Delta pK < 1$. Ainsi la constante de vitesse d'échange chimique k_{ch} sera égale à la vitesse de transfert du proton. En prenant en compte tous les catalyseurs, la vitesse de l'échange HX s'écrit :

$$k_{\text{int}} = \sum_j k_{\text{ch}} [\text{CAT}]_j \tag{4.12}$$

avec $[\text{CAT}]_j$, la concentration du catalyseur j et k_{int} la constante de vitesse d'échange intrinsèque. Pour un peptide non structuré dans l'eau, la catalyse est effectuée par OH^- et H^+ (H_3O^+) principalement, mais aussi par H_2O indépendamment du pH [Gregory et al., 1983] [Englander et al., 1979].

Ainsi la relation précédente devient [Berger and Loewenstein, 1959] :

$$k_{\text{int}} = k_{\text{OH}^-}[\text{OH}^-] + k_{\text{H}^+}[\text{H}^+] + k_{\text{H}_2\text{O}}[\text{H}_2\text{O}] \tag{4.13}$$

La constante de vitesse d'échange intrinsèque k_{int} peut être influencée par une variété de facteurs environnementaux tels que le pH, la température, la nature du solvant, l'induction chimique due aux chaines latérales des résidus flanquants ainsi que leur implication dans une structure secondaire. De manière évidente, le pH est un des principaux facteurs sachant que l'échange HX est catalysé par les ions OH^- et H^+. La vitesse minimale d'échange est atteinte lors le pH est égal à 2.5-3.0, où OH^- et H^+ ont la même activité catalytique (Figure 4.5 à gauche). À un pH au-dessus ou en dessous du pH minimal, l'échange est une réaction du premier ordre ($k_{\text{int}} = k_{\text{OH}^-}[\text{OH}^-]$ ou $k_{\text{int}} = k_{\text{H}^+}[\text{H}^+]$) avec un coefficient directeur égal à -1 et 1 approximativement lorsque le pH est en dessous ou au-dessus du pH minimal. Ainsi lorsque le pH est supérieur à 4, la catalyse d'échange est essentiellement due aux ions OH^-. Typiquement, lors d'une expérience HDX-MS, la phase d'échange H/D est réalisée à pH physiologique, entre 7 et 7.4, où la vitesse d'échange chimique k_{ch} est d'environ 1.7 s^{-1}. La diminution du pH à 2.5-3.0 va mener à une diminution de la vitesse d'échange D/H d'un ordre de grandeur de 10^4 [Weis, 2016]. Par exemple, le temps de demi-vie de l'échange $t_{1/2}$ d'un peptide de référence, le poly-DL-Alanine (PDLA), est d'environ 25 minutes à pH 2.5-3.0, tandis qu'il est de moins de 1 seconde lorsque le pH est égal à 7 [Weis, 2016] (Figure 4.5 à droite). De fait, la diminution du pH est un paramètre important pour la détermination du pourcentage d'incorporation en deutérium dans une expérience de type HDX-MS.

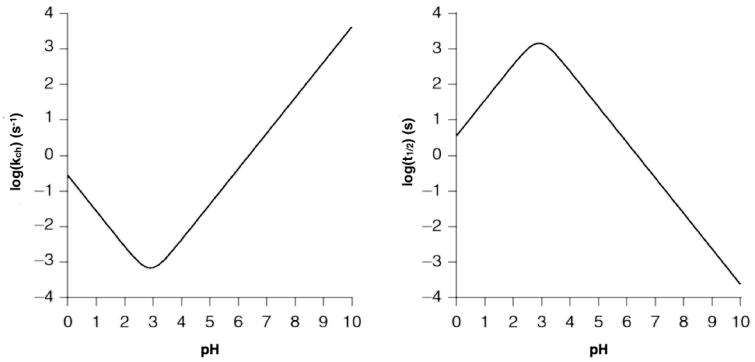


FIGURE 4.5 – Exemple d'un profile pH de la constante de vitesse d'échange chimique k_{ch} d'un proton amide HN (à gauche) et du temps de demi-vie de l'échange $t_{1/2}$ d'un proton amide HN (à droite) localisé dans un peptide de référence PDLA à 20° Celsius. Adapté de [Weis, 2016].

La température influence également l'échange HX de peptides non structurés, principalement par son effet sur la constante d'ionisation de l'eau. Le choix de l'isotope (Hydrogène, Deutérium ou Tritium) influence aussi l'échange HX, mais dans une plus faible mesure. Dans le cas d'un peptide, il est aussi nécessaire de prendre en compte la nature des chaines latérales des résidus voisins [Sheinblatt, 1970] [Molday et al., 1972] [Bai et al., 1993]. En effet, la vitesse d'échange HX des protons varie en fonction de la nature de ces dernières. Par exemple, si le proton amide HN est entouré par deux résidus isoleucine, l'échange H/D sera ralenti, d'environ 10 à 20 fois (selon le pH) [Weis, 2016], du fait d'une accessibilité au solvant moins importante, comparé à un proton amide HN entouré par deux résidus alanine. Dans le cas où les résidus flanquants sont polaires, ces derniers vont avoir un effet inductif sur les électrons, rendant l'environnement plus acide, favorisant la catalyse par OH^- et dans le cas contraire, défavorisant la catalyse par H^+. Ce qui veut dire que l'échange H/D est plus rapide lorsque le proton amide est entouré par des résidus sérine que lorsqu'il est entouré par des résidus alanine (pH > 4). L'effet combiné des chaines latérales aura une incidence sur la vitesse d'échange H/D se reflètera par un décalage de la valeur du pH où l'échange est minimal. Par exemple, le pH d'échange minimal est de 2.9 lorsque le proton amide HN est entouré par des alanines, de 2.2 avec des résidus sérine ou encore de 1.5 avec des résidus histidine. Tous ces effets ont été calibrés par Molday *et al.* [Molday et al., 1972] et de manière plus précise pour chaque résidu par Bai *et al.* [Bai et al., 1993]. Ils ont montré que les chaines latérales voisines agissaient de manière indépendante et additive sur la vitesse d'échange du proton amide HN. Ainsi, il est possible de prédire la vitesse d'échange H/D de n'importe quel proton amide HN au sein d'un polypeptide non structuré si les résidus voisins sont connus, excepté pour le proton amide HN localisé en seconde position due à la charge positive en N-terminal [Molday et al., 1972] [Bai et al., 1993], en partant d'un peptide de référence PDLA, et en y incorporant les différents effets du pH, de la température, de l'isotope et des résidus flanquants.

L'équation suivante permet ainsi de prédire la constante de vitesse d'échange k_{int} d'un proton amide HN en fonction de la nature des chaines latérales dans un polypeptide non structuré (random coil) avec une faible salinité (50 nM) et une température de 293K dans une solution contenant du D_2O en excès [Bai et al., 1993].

$$\begin{aligned} k_{int} &= k_{\mathrm{D^+}} + k_{\mathrm{OD^-}} + k_{\mathrm{H_2O}} \\ &= k_{A,ref}(A_L \times A_R)[D^+] + k_{B,ref}(B_L \times B_R)[OD^-] + k_{W,ref}(B_L \times B_R) \end{aligned} \tag{4.14}$$

avec $k_{A,ref}$, $k_{B,ref}$ et $k_{W,ref}$ égale à 1.62, 10.05 et -1.5 $M^{-1}min^{-1}$ exprimée sur l'échelle logarithmique, respectivement, correspondant aux vitesses d'échanges standards obtenues à partir d'un peptide de

référence PDLA à 20° Celsius. Les termes (A_L, A_R) et (B_L, B_R) correspondent aux facteurs de correction relative au résidu alanine spécifique à chaque chaine latérale et à sa position (L ou R) par rapport au proton amide HN. Toutes ces valeurs sont exprimées sur l'échelle logarithmique et regroupées dans la table 4.1.

Les relations suivantes permettent d'obtenir la contribution de chaque terme k_{D^+}, k_{OD^-} et k_{W}, en prenant l'antilog de la somme de la vitesse d'échange standard, des facteurs de correction correspondant aux chaines latérales L et R ainsi que du $\mathrm{pD_{corr}}$ ou du $\mathrm{pOD_{corr}}$:

$$\begin{aligned} k_{int} &= k_{\mathrm{D}^+} + k_{\mathrm{OD}^-} + k_{\mathrm{H_2O}} \\ &= 10^{log(k_{A,\mathrm{ref}})+log(A_L)+log(A_R)-pD_{\mathrm{corr}}} \\ &+ 10^{log(k_{B,\mathrm{ref}})+log(B_L)+log(B_R)-pOD_{\mathrm{corr}}} \\ &+ 10^{log(k_{W,\mathrm{ref}})+log(B_L)+log(B_R)} \end{aligned} \tag{4.15}$$

avec $\mathrm{pD_{corr}}$ égal à $\mathrm{pH_{read}} + 0.4$ [Glasoe and Long, 1960] et $\mathrm{pOD_{corr}} = \mathrm{pD_{corr}} + \mathrm{pKc}$. pKc correspond à la constante molaire d'ionisation du D_2O déterminée à partir de la relation suivante [Covington et al., 1966] :

$$pKc = A1/T - A2 + A3 \times T \tag{4.16}$$

avec T la température exprimée en Kelvin et $A1$, $A2$ et $A3$ égaux à 5075.91, 8.7357 et 0.0220701, respectivement, et obtenus par ajustement de courbe à l'aide du module Python Scipy [Jones et al., 01]. Afin de prédire la constante de vitesse d'échange k_{int} à différentes températures, chaque vitesse de référence obtenue à l'aide de l'équation suivante devra être modifiée selon la forme modifiée de l'équation d'Arrhenius :

$$\begin{aligned} k_{int}(T) &= k_{\mathrm{D}^+}(T) + k_{\mathrm{OD}^-}(T) + k_{\mathrm{D_2O}}(T) \\ &= k_{\mathrm{D}^+} e^{-Ea(k_A)(\frac{1}{T}-\frac{1.}{293.})/R} \\ &+ k_{\mathrm{OD}^-} e^{-Ea(k_B)(\frac{1}{T}-\frac{1.}{293.})/R} \\ &+ k_{\mathrm{D_2O}} e^{-Ea(k_W)(\frac{1}{T}-\frac{1.}{293.})/R} \end{aligned} \tag{4.17}$$

avec R la constante universelle des gaz parfaits égale à 1.9858775×10^{-3} Kcal.mol^{-1}.K^{-1}, ainsi que $Ea(k_A)$, $Ea(k_B)$ et $Ea(k_W)$ égale à 15, 2.6 et 13 kcal/mol, respectivement, correspondant aux énergies d'activation provenant du programme Sphere (`http://landing.foxchase.org/research/labs/roder/sphere/sphere.html`) [Zhang, 1995]. Dans le cas où l'on choisit une énergie d'acti-

vation $Ea(k_B)$ égale à 17 kcal/mol appliquée à k_{OD^-}, il n'est pas nécessaire de calculer [OD^-] en fonction de la température [Bai et al., 1993].

En résumé, l'échange HX au sein des protéines est généralement dominé par les protons amides HN échangeant lentement. Dans la plupart des conditions, le transfert de protons est catalysé par OH^- et H^+ après collision et formation d'une liaison hydrogène entre le proton amide HN et le catalyseur. Ainsi on peut en déduire qu'une liaison hydrogène intramoléculaire bloquera l'échange HX. Pour finir la vitesse d'échange HX est modulé par plusieurs facteurs, dont les plus importants sont le pH, la température et la nature des chaines latérales des résidus voisins.

	log k_{int}(X) - log k_{int}(Ala)			
	Catalyse acide		Catalyse basique	
Chaine latérale (X)	L	R	L	R
Alanine	0.00	0.00	0.00	0.00
Arginine	-0.59	-0.32	0.08	0.22
Asparagine	-0.58	-0.13	0.49	0.32
Acide aspartique (COO^-)	(0.9)	0.58	-0.30	-0.18
Acide aspartique (COOH)	(-0.9)	-0.12	0.69	(0.6)
Cystéine	-0.54	-0.46	0.62	0.55
$Cystéine_2$	-0.74	-0.58	0.55	0.46
Glycine	-0.22	0.22	0.27	0.17
Glutamine	-0.47	-0.27	0.06	0.20
Acide glutamique (COO^-)	(-0.9)	0.31	-0.51	-0.15
Acide glutamique (COOH)	(-0.6)	-0.27	0.24	0.39
Histidine			-0.10	0.14
$Histidine^+$	(-0.8)	-0.51	(0.8)	0.83
Isoleucine	-0.91	-0.59	-0.73	-0.23
Leucine	-0.57	-0.13	-0.58	-0.21
Lysine	-0.56	-0.29	-0.04	0.12
Méthionine	-0.64	-0.28	-0.01	0.11
Phénylalanine	-0.52	-0.43	-0.24	0.06
Proline (trans)		-0.19		-0.24
Proline (cis)		-0.85		0.60
Sérine	-0.44	-0.39	0.37	0.30
Thréonine	-0.79	-0.47	-0.07	0.20
Tryptophane	-0.40	-0.44	-0.41	-0.11
Tyrosine	-0.41	-0.37	-0.27	0.05
Valine	-0.74	-0.30	-0.70	-0.14
N-term ($NH3^+$)		-1.32		1.62
C-term (COO^-)	0.96		(-1.8)	
C-term (COOH)	(0.05)			

TABLE 4.1 – Facteurs de correction spécifique à chaque chaine latérale en fonction de la leur position par rapport au proton amide HN. L et R désigne les chaines latérales gauche et droite, respectivement. Les valeurs sont en exprimé sur l'échelle logarithmique afin de pouvoir être utilisé directement avec l'équation 4.15. Les valeurs en parenthèse signifient que celles-ci sont moins précises.

4.2.3.2 Au sein des protéines

Le mécanisme général d'échange HX dans des peptides non structurés est donc largement compris [Eigen, 1964] [Perrin and Lollo, 1984] [Baldwin, 2011], et les constantes de vitesses d'échange k_{int} des protons amides HN exposés au solvant dans des polypeptides non structurés peuvent être prédites, quelle que soit la séquence ou les conditions expérimentales (pH, température et isotope) [Molday et al., 1972] [Bai et al., 1993] [Connelly et al., 1993]. Les protons amides HN échangent de manière générale plus lentement au sein des protéines repliées. Cependant d'un proton amide HN à l'autre la vitesse d'échange HX peut varier très nettement sur plusieurs ordres de grandeur. Ainsi, l'échange permet d'obtenir des informations sur la structure, la dynamique et la stabilité de la protéine, et de pouvoir observer l'effet de certaines interactions entre différents partenaires à une résolution de l'ordre du peptide au résidu, selon la méthode utilisée. Cependant, pour exploiter toutes ces informations, il est nécessaire de comprendre comment fonctionne le mécanisme d'échange HX au sein des protéines. Historiquement, plusieurs modèles ont été débattus pour expliquer cette différence, d'une part le modèle de pénétration au solvant soutenu par Woodward [Woodward and Hilton, 1979], et de l'autre le modèle de dépliement local proposé par Linderstrom-Lang et Hvidt [Linderstrøm-Lang, 1958] [Hvidt and Nielsen, 1966]. Le modèle de pénétration/accessibilité au solvant partant du principe que la vitesse d'échange HX est liée au degré d'exposition des protons amides HN, mais aussi liée l'enfouissement, à la densité structurale locale et à la présence (absence) de canaux permettant (empêchant) la pénétration du solvant au sein de la protéine. Ce modèle suppose que les protons amides HN placés à la surface de la protéine vont avoir tendance à échanger plus rapidement que ceux qui sont enfouis. Cependant, il a été très clairement montré pour la double hélice de l'ADN, par exemple, que bien que les protons amides HN des paires de bases soient complètement exposés au solvant, la vitesse d'échange HX est lente [Englander and Downer, 1972], ce qui va à l'encontre du modèle. Ainsi le modèle le plus généralement admis à l'heure actuelle est que la vitesse d'échange HX est liée à l'intégrité du réseau de liaisons hydrogène intramoléculaire et permet de suivre la stabilité et la dynamique des éléments de structures secondaires. Ainsi, le mécanisme général d'échange HX peut être décrit selon la relation de Linderstorm-Lang :

$$NH_{\text{fermé}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} NH_{\text{ouvert}} \xrightarrow[H_2^*O]{k_{int}} NH^*_{\text{ouvert}} \underset{k_{op}}{\overset{k_{cl}}{\rightleftharpoons}} NH^*_{\text{fermé}} \tag{4.18}$$

où k_{cl} et k_{op}, les constantes de vitesses d'ouverture et de fermeture de la structure de la protéine au niveau local et k_{int}, la constante de vitesse d'échange intrinsèque, qui dépend du pH, de la température,

du solvant ainsi que de l'environnement chimique local (nature des chaines latérales des résidus voisins). Ce modèle stipule que le processus d'échange requiert une déprotection du proton amide HN pour qu'il y ait échange HX, déprotection liée à des évènements de dépliements locaux de la protéine [Englander and Downer, 1972]. En partant du principe que $k_{\text{cl}} >> k_{\text{op}}$ et/ou $k_{\text{cl}} >> k_{\text{int}}$, alors la constante de vitesse de l'échange k_{ex} peut être exprimé en fonction des constantes d'ouverture et de fermeture, k_{op} et k_{cl}, et de la constante de vitesse d'échange intrinsèque k_{int} dans un peptide non replié. Ainsi, on obtient la formulation générale de la cinétique de l'échange HX suivante, décrite par Hvidt [Hvidt, 1964] [Hvidt and Nielsen, 1966], dépendant principalement de la rupture d'une liaison hydrogène :

$$k_{\text{ex}} = \frac{k_{\text{op}} k_{\text{int}}}{k_{\text{op}} + k_{\text{cl}} + k_{\text{int}}} \tag{4.19}$$

avec $k_{\text{int}} = k_{\text{ch}}[\text{CAT}]$ et [CAT] la concentration en catalyseur. Si le proton amide HN est principalement dans un état ouvert ($k_{\text{op}} >> k_{\text{cl}}$), alors la constante de vitesse d'échange k_{ex} est égale à k_{op} ou si le proton amide HN est principalement dans un état fermé ($k_{\text{cl}} >> k_{\text{op}}$) alors la constante de vitesse d'échange k_{ex} dépendra de la constante de vitesse intrinsèque k_{int} multipliée par la fraction de temps où le proton amide HN est dans un état ouvert. On obtient alors les relations suivantes :

$$\begin{aligned} k_{\text{ex}} &= k_{\text{op}} && \text{si } k_{\text{int}} >> k_{\text{cl}} \text{ (EX1)} \\ k_{\text{ex}} &= k_{\text{int}} \frac{k_{\text{op}}}{k_{\text{cl}}} && \text{si } k_{\text{cl}} >> k_{\text{int}} \text{ (EX2)} \end{aligned} \tag{4.20}$$

Ces deux mécanismes d'échange HX ont été décrits pour interpréter quantitativement l'échange HX [Linderstrom-Lang, 1955] [Berger and Linderstrøm-Lang, 1957] et communément appelés EX1 et EX2 pour ***EX****change mechanism* [Hvidt and Nielsen, 1966]. On nommera EX1 de type monomoléculaire le mécanisme d'échange HX lorsque $k_{\text{int}} >> k_{\text{cl}}$, où à chaque ouverture un échange HX se produit. Et le mécanisme d'échange HX EX2 de type bimoléculaire lorsque $k_{\text{cl}} >> k_{\text{int}}$, où plusieurs évènements d'ouverture seront nécessaires pour qu'un échange HX se produise. Dans le cas du mécanisme EX1, celui-ci ne dépend pas de la concentration en catalyseur, qui est généralement OH^- dans la gamme de pH (> 4) utilisée pour étudier les macromolécules biologiques, contrairement au mécanisme EX2. Mais de manière générale, en condition physiologique avec un pH neutre et une température basse, le régime EX2 est dominant comparé au régime EX1 caractérisé par un pH extrême (soit très bas ou très élevé), une température élevée, ou une faible stabilité. On peut aussi exprimer l'échange HX en facteur de protection PF, qui est défini comme étant le ratio entre la constante vitesse d'échange intrinsèque k_{int} pour un peptide

non structurée et la vitesse de l'échange k_{ex}mesuré.

$$PF = \frac{k_{\text{int}}}{k_{\text{ex}}} \tag{4.21}$$

De cette relation, on peut alors déduire, à partir du mécanisme d'échange EX2, que le facteur de protection PF est égal à k_{cl}/k_{op}.

4.2.3.3 Modélisation de la vitesse d'échange HX

Connaissant le facteur de protection PF ainsi que la constante de vitesse intrinsèque k_{int}, définie précédemment, la probabilité qu'un proton amide HN ait échangé au sein d'une protéine en fonction du temps suit une loi exponentielle et peut être calculée de la manière suivante [Hvidt and Nielsen, 1966] [Radou et al., 2014] [Park et al., 2015] :

$$DI(t) = 1 - e^{-(k_{int,i}/PF_i)t} \tag{4.22}$$

avec PF_i et $k_{int,i}$, le facteur de protection (sans dimension) et la constante de vitesse intrinsèque en seconde^{-1} ou minute^{-1} du résidu i, respectivement, et t le temps exprimé en seconde ou en minute. Ainsi les protons amides HN tendent à être échangés complètement au cours du temps, quelle que soit la stabilité de la protéine. Cependant en HDX/MS, et contrairement à la méthode HDX/RMN, l'incorporation en deutérium n'est généralement pas résolue à l'échelle du résidu, mais l'est seulement à l'échelle d'un peptide résultant de la protéolyse. Ainsi l'incorporation pour un peptide j est estimée en moyennant la probabilité d'échange HX de chaque résidu i au cours du temps,

$$DI_{\text{peptide}}(t) = \frac{1}{n_j - 2} \sum_{i=m_j+2}^{m_j+n_j-1} (1 - e^{-(k_{int,i}/PF_i)t}) \tag{4.23}$$

n_j and m_j désignent le nombre de résidus, moins les prolines, et l'index du premier résidu dans le peptide j, respectivement. Les deux premiers résidus en N-terminal de chaque peptide ne sont pas pris en compte, car ils contre-échangent très rapidement après protéolyse [Molday et al., 1972] [Bai et al., 1993].

4.2.4 Modélisation du transfert de protons

4.2.4.1 Limites des modèles existants

De nombreux modèles ont été proposés pour prédire l'échange HX basé sur des propriétés calculées à partir de structures connues. Ces modèles s'appuient généralement sur l'accessibilité ou la pénétration au solvant [Lumry and Rosenberg, 1975] [Richards, 1979] [Woodward et al., 1982] [Li and Woodward, 1999], le champ électrostatique [Matthew and Richards, 1983] [Fogolari et al., 1998] [LeMaster et al., 2009] [Anderson et al., 2006] ou encore la dynamique structurale. Parmi les modèles proposés basés sur la dynamique structurale et spécifique à l'échange HX, les plus connus sont ceux de Vendruscolo *et al.* [Vendruscolo et al., 2003] et Liu *et al.* [Liu et al., 2012], l'un basé sur un modèle phénoménologique et l'autre basé sur l'algorithme COREX [Hilser and Freire, 1996], nommé DXCOREX [Liu et al., 2012], respectivement. Le but de ces méthodes est de générer un ensemble de conformations représentatif des données d'échange HX expérimentaux. Cet ensemble est généré par Monte-Carlo, qui dans le cas du modèle phénoménologique est contraint par les données RMN, nécessitant ainsi des données complètes résolues à l'échelle du résidu. Les facteurs de protection PF sont obtenus soit pour DXCOREX en calculant le ratio entre le nombre de protons amides HN accessibles au solvant et ceux qui ne le sont pas, ou pour le modèle Vendruscolo à partir de l'expression phénoménologique $PF = \beta_c N^c + \beta_h N^h$ avec N^c et N^h le nombre de contacts et le nombre de liaisons hydrogène formées, respectivement, et β_c et β_h deux paramètres qui sont ajustés pour améliorer la corrélation entre les données prédites et expérimentales. Cependant, il a été montré que ces modèles s'avéraient insatisfaisants dans leur ensemble [Skinner et al., 2012b] [Park et al., 2015]. Par exemple, quelle que soit la métrique utilisée pour définir si le proton amide est dans un état compétent ou non, des résultats comparables sont obtenus, spécialement dans le cas du modèle de Vendruscolo où les paramètres β sont optimisés [Kieseritzky et al., 2006]. De plus, les méthodes basées sur la surface accessible au solvant, comme DXCOREX, ne permettent pas de séparer clairement les protons amides HN capables d'échanger de ceux qui ne le sont pas.

Ainsi, ces résultats soulignent la difficulté d'interprétation des données HX. Des concepts de base comme la nature des états « ouvert » et « fermé » restent encore assez flous. Ceci est en partie liée au fait que les prédictions HX ont souvent été interprétées dans le contexte de structures cristallographiques, qui par définition sont « statiques ». Or la présence d'échange souligne que l'ensemble conformationnel accessible à une macromolécule biologique est plus large que celui représenté par les structures stables obtenues par cristallographie. D'autres approches ont été proposées basées sur la génération d'ensemble

de conformations, comme celles présentées précédemment [Vendruscolo et al., 2003] [Liu et al., 2012] ou à partir de simulations de dynamiques moléculaires tout atome [Petruk et al., 2012] [Hsu et al., 2013] [Radou et al., 2014] [McAllister and Konermann, 2015] [Park et al., 2015] [Persson and Halle, 2015] afin d'apporter une vue plus dynamique et plus fidèle en y incluant explicitement le solvant pour ces dernières. Malheureusement, l'échelle de temps accessible par les simulations de dynamiques moléculaires classiques tout atome reste limité, au mieux de l'ordre de la micro- à la milliseconde pour les systèmes de petite taille (< 70 résidus) [Piana et al., 2013], ce qui est bien en deçà de l'échelle de temps mesurable par échange HX, de l'ordre de la seconde jusqu'à plusieurs jours [McAllister and Konermann, 2015]. Ce paramètre est aussi l'un des principaux facteurs limitants dans la prédiction de l'échange HX. Connaissant l'importante flexibilité du LBD de certains récepteurs nucléaires en absence de ligand, cette limitation est parfaitement illustrée par la méthode DXCOREX qui en partant de la forme inactive du LBD de PPARγ, n'obtient qu'une corrélation de 0.56 entre les données prédites et expérimentales, contrairement à la forme active en complexe avec le rosiglitazone, beaucoup plus stable, avec une corrélation de 0.80 [Liu et al., 2012]. Des simulations utilisant un modèle « gros-grains » permettent de confirmer l'importance de l'échantillonnage conformationnel [Craig et al., 2011], mais souffrent malheureusement du manque de détails à l'échelle atomique, rendant plus compliquée la comparaison avec les données expérimentales.

En résumé, les limitations dans l'interprétation des données d'échange HX se situent principalement au niveau des modèles utilisés pour identifier le degré de protection d'un proton amide HN, mais aussi au niveau de l'échantillonnage conformationnel. Néanmoins, la plupart des auteurs s'accordent aujourd'hui à dire que le degré de protection des protons amides HN contre l'échange HX ne dépend pas simplement que de leur exposition au solvant, de leur enfouissement ou de l'atome accepteur, provenant soit du squelette peptidique ou d'une chaine latérale [Milne et al., 1998] [Skinner et al., 2012a] [Park et al., 2015] [Persson and Halle, 2015]. Le paramètre déterminant pour expliquer la protection de certains protons amides HN à l'échange HX serait lié à la difficulté de rompre une liaison hydrogène, dans laquelle ils sont impliqués, de sorte qu'ils soient accessibles pour être attaqué par un catalyseur (OH^- ou H^+). Et que cette difficulté à rompre une liaison hydrogène est liée à la stabilité de l'environnement local ou global qui joue un rôle de régulateur dans la vitesse d'échange HX [Craig et al., 2011] [McAllister and Konermann, 2015] [Persson and Halle, 2015]. Ainsi, l'utilisation de simulations de dynamiques moléculaires tout atome offre une meilleure alternative que les structures cristallographiques pour interpréter les données d'échange HX, de fait l'utilisation d'un solvant explicite permettant de suivre directement les interactions entre les protons amides HN et le solvant. De même que pour l'échantillonnage conformationnel, l'introduction de nou-

velles méthodes d'échantillonnage avancées comme la dynamique moléculaire accélérée permettra à terme d'outrepasser cette limite, afin d'explorer plus rapidement toutes les conformations accessibles.

4.2.4.2 Modèles utilisés

Il n'est pas possible d'observer les réactions de transfert de protons par un modèle de mécanique moléculaire. Il est dès lors nécessaire de définir des critères génériques pour tous les protons amides HN, afin d'identifier les états où il pourrait y avoir un échange de protons. À partir des études sur le sujet, il est maintenant admis que les liaisons hydrogène sont un des principaux facteurs limitants dans le mécanisme d'échange HX. La vitesse d'incorporation est affectée par les liaisons hydrogène intramoléculaires qui peuvent bloquer le mécanisme de transfert de protons. De même que la vitesse d'échange HX peut-être ralentit du fait de l'enfouissement du proton amide HN. Ainsi ces observations suggèrent que la tendance qu'a un proton amide HN à échanger devrait être corrélée avec la persistance et la nature des liaisons hydrogène qu'il forme. Une approche basée seulement sur le nombre de liaisons hydrogène où les protons amides HN sont impliqués, obtenue à partir de l'ensemble des conformations générées par dynamique moléculaire, devrait donc permettre de prédire la vitesse d'échange HX.

Par analogie avec le modèle de dépliement local, proposé par Linderstrom-Lang, on définira de manière générale un proton amide HN comme étant dans un état compétent si celui-ci interagit uniquement avec une ou plusieurs molécules d'eau par liaison hydrogène, sans interagir avec le squelette peptidique ou les chaines latérales [Park et al., 2015] [Persson and Halle, 2015]. Dans le cas contraire, si ces critères ne sont pas satisfaits, alors on considéra le proton amide HN comme étant dans un état incompétent. L'identification des liaisons hydrogène se fera selon des critères purement géométriques, en ne prenant en compte que la distance et l'angle entre le proton amide HN et l'atome accepteur. Ainsi, toutes les liaisons hydrogène seront identifiées dans chaque conformation extraite avec un intervalle de 1 ps pour chaque trajectoire aMD et cMD en utilisant le module Python MDAnalysis [Michaud-Agrawal et al., 2011]. Dans le cadre de cette étude, plusieurs modèles ont été explorés.

4.2.4.2.1 Modèle simple

Ce modèle, dérivé directement du modèle cinétique classique, consiste à partir du principe que lorsqu' un proton amide HN sans interaction intramoléculaire établit une liaison hydrogène avec une molécule d'eau, celui-ci est dans un état compétent pour l'échange. Pour ce modèle, les critères géométriques pour détecter une liaison hydrogène seront les suivants :

— Distance NH – $OH_2 \leq 2$ Å

— Angle NH – $OH_2 \geq 150$ °

4.2.4.2.2 Modèle concerté

Ce second modèle dit « concerté », part du principe que deux molécules d'eau, ou plus, à proximité d'un proton amide HN sont nécessaires pour considérer celui-ci comme compétent, contrairement au modèle simple [Persson and Halle, 2015] (Figure 4.6). À l'approche d'un ion hydroxyde OH^-, ce dernier va interagir avec le proton amide HN et l'extraire, convertissant l'ion hydroxyde OH^- en molécule d'eau. Simultanément, une seconde molécule d'eau interagira avec l'anion imidate N^- nouvellement formé, afin que ce dernier puisse acquérir un nouveau proton amide HN. Cette seconde molécule d'eau sera ainsi convertie en ion hydroxyde OH^-.

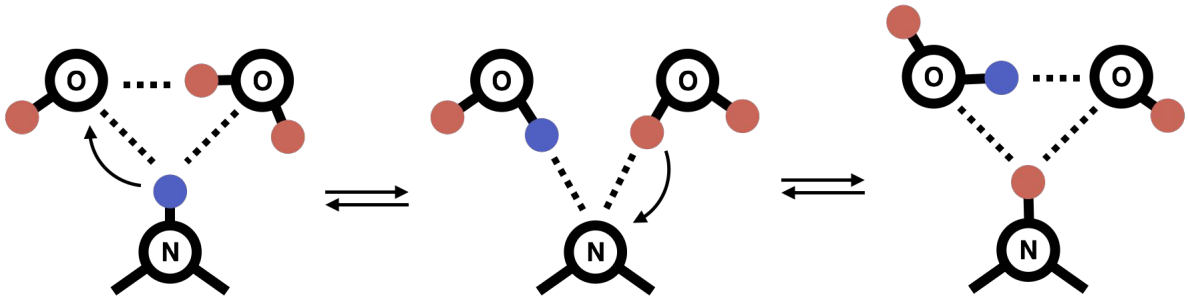


FIGURE 4.6 – Représentation schématique du transfert de proton à l'aide de deux molécules d'eau coordonnées par le proton amide HN. À la fin du processus du transfert d'électron, au moins une des deux molécules d'eau accepte une liaison hydrogène provenant du groupement HN ou DN. Dans l'état intermédiaire, les deux molécules d'eau interagissent avec le groupement imidate N^-.

Une étude récente de l'échange à partir de simulations de la protéine BPTI [Persson and Halle, 2015] indique qu'un critère géométrique impliquant la proximité de deux molécules d'eau, inspiré par le mécanisme ci-dessus, permet une meilleure détermination de l'échange à partir de trajectoires de dynamique moléculaire. Les critères géométriques utilisés pour détecter une liaison hydrogène dans ce cas seront moins stricts que dans le cas du modèle simple. Ainsi, les critères utilisés seront ceux définis par défaut pour la fonction COOR HBONd de CHARMM [De Loof et al., 1992] :

— Distance NH – $OH_2 \leq 2.4$ Å

— Aucune contrainte d'angle entre NH – OH_2

4.2.5 Prédiction des facteurs de protection

Par analogie avec les facteurs de protection expérimentaux, nous calculerons le facteur de protection PF_i associé à chaque résidu i comme étant le ratio entre le nombre de protons amides HN incompétents et le nombre de protons amides HN compétents [Persson and Halle, 2015] :

$$PF_i = \frac{k_{cl}}{k_{op}} = \frac{[\text{Closed}]}{[\text{Open}]} = \frac{[\text{No. (conformation) - No. (compétent)}_i]}{[\text{No. (compétent)}_i]} \tag{4.24}$$

avec No. (compétent)$_i$ le nombre de conformations où le proton amide HN du résidu i est dans un état compétent, et No. (conformation), le nombre total de conformations analysées. Cependant, d'après ce modèle les valeurs des facteurs de protection PF calculées sont bornées entre 0 et l'infini. Elles sont également dépendantes du nombre total de conformations analysées, dans la mesure où l'échantillonnage n'est pas exhaustif. Celles-ci seront ainsi normalisées entre la valeur minimale expérimentale qui est de 1 et une valeur maximale que nous avons estimée par la procédure ci-dessous. Dans une première étape, le peptide pour lequel l'échange est le plus lent est utilisé pour déterminer la valeur maximale du facteur de protection [Hamuro et al., 2006] :

1. Les constantes de vitesse d'échange intrinsèque k_{int} sont calculées pour chaque résidu du peptide échangeant le plus lentement en utilisant pour ces résidus des valeurs des facteurs de protection qui sont successivement 1, 10, 10^2, 10^3, 10^4, 10^5, 10^6 et 10^7 :
2. L'incorporation en deutérium au sein de ce peptide est ensuite calculée pour chacune des valeurs de facteur de protection PF testées.
3. Le RMSD est ensuite calculé entre l'incorporation prédite et expérimentale pour le peptide le plus protégé. Au final, on choisira la valeur du facteur de protection qui minimise la RMSD entre données expérimentales et calculées. Cette valeur finale sera notée b ci-dessous et en pratique valait 10^6 pour nos simulations aMD.

Une fois le facteur de protection maximale obtenue, l'ensemble des données est renormalisé, afin d'obtenir des facteurs de protection $PF_{\text{normalisé}}$ dont la valeur varie entre 1 et la valeur maximale b du facteur de protection obtenue comme décrite ci-dessus. :

$$PF_{i,\text{normalisé}} = a + \frac{(PF_i - PF_{\text{min}})(b-a)}{\max(PF) - PF_{\text{min}}} \tag{4.25}$$

avec a et PF_{min} égalent à 1 et b égale à la valeur maximale des facteurs de protection calibrés sur les

données expérimentales (voir ci-dessus, soit 10^6), et max(PF) la valeur maximale des facteurs de protection PF calculés sur l'ensemble des résidus étudiés. La normalisation permet en particulier la comparaison de données issues de dynamiques moléculaires courtes et de dynamiques aMD plus longues.

4.2.6 Comparaison avec les données expérimentales

Afin de mesurer l'accord entre les données expérimentales et celles prédites, le coefficient de corrélation r ainsi que le RMSD seront calculés à partir de l'incorporation en deutérium aux temps t 15, 30, 120, 900 et 3600 secondes pour chaque peptide. Le coefficient de corrélation r servira à vérifier si la vitesse d'incorporation en deutérium prédite et expérimentale s'effectue au même régime. Le coefficient de corrélation r (sans dimension) sera calculé selon la relation suivante :

$$r_{\text{pep}} = \frac{\sum_t^T (DI_{\text{pep, pred}}(t) - \overline{DI}_{\text{pep, pred}})(DI_{\text{pep, exp}}(t) - \overline{DI}_{\text{pep, exp}})}{\sqrt{\sum_t^T (DI_{\text{pep, pred}}(t) - \overline{DI}_{\text{pep, pred}})^2} \times \sqrt{\sum_t^T (DI_{\text{pep, exp}}(t) - \overline{DI}_{\text{pep, exp}})^2}} \tag{4.26}$$

avec $DI_{\text{pep, pred}}(t)$ et $DI_{\text{pep, exp}}(t)$, le taux d'incorporation en deutérium prédit et expérimental au temps t pour un peptide donné, respectivement, $\overline{DI}_{\text{pep, pred}}$ et $\overline{DI}_{\text{pep, exp}}$, le taux d'incorporation en deutérium moyen prédit et expérimental, respectivement, pour un peptide donné et t égal à 15, 30, 120, 900 et 3600 secondes. Le calcul du RMSD permettra d'apporter une réponse claire et définitive quant à l'accord réel entre les données expérimentales et prédites à partir des trajectoires aMD et cMD, suivant le modèle présenté plus tôt. Les trajectoires cMD seront utilisées afin de pouvoir comparer de manière indirecte les structures cristallographiques de référence, dont les identifiants PBD sont 1LBD et 1FM9 pour la forme inactive et active du LBD de RXRα, respectivement, aux trajectoires aMD. Le RMSD (exprimé en %) sera calculé de la manière suivante :

$$RMSD_{\text{pep}} = \sqrt{\frac{1}{T}\sum_t^T (DI_{\text{pep, pred}}(t) - DI_{\text{pep, exp}}(t))^2} \tag{4.27}$$

avec $DI_{\text{pep, pred}}(t)$ et $DI_{\text{pep, exp}}(t)$, le taux d'incorporation en deutérium prédit et expérimental au temps t pour un peptide donné, respectivement, et t égale à 15, 30, 120, 900 et 3600 secondes.

Chapitre 5

Dynamique conformationnelle du LBD de RXRα

5.1 Introduction

Ce chapitre décrit les résultats obtenus en utilisant l'aMD afin d'étudier la dynamique conformationnelle de la forme libre et de la forme liée à 9cRA du LBD de RXRα. L'objectif fut de développer et valider un protocole de simulation aMD. L'ensemble des conformations générées seront comparées aux données expérimentales (RMN, HDX) en utilisant les modèles théoriques décrits plus tôt. Le très grand nombre de conformations différentes générées a été l'occasion de tester une nouvelle méthode de réduction de la dimensionnalité non linéaire, nommée Stochastic Proximity Embedding, en utilisant la distance dihédrale comme métrique.

5.2 Matériels et Méthodes

5.2.1 Choix et comparaison des structures cristallographiques

Pour notre étude, nous avions à notre disposition plusieurs structures expérimentales des domaines de liaison au ligand (LBD) de RXRα, obtenues par cristallographie aux rayons X et accessibles via la Protein DataBank (PDB) [Berman et al., 2000]. La structure cristallographique aux rayons X du complexe hétérodimérique des domaines de liaison (LBD) de RXRα et PPARγ, en présence de l'acide rétinoïque 9-cis (9cRA) et d'un composé antidiabétique GI262570 ainsi que de deux peptides de 10 et 16 résidus de long provenant de la protéine « Steroid Receptor Coactivator 1 » (SRC-1), a été utilisé (PDB id : 1FM9 [Gampe et al., 2000a]) pour la forme liée. Tandis que pour la forme libre, c'est la structure monomérique de RXRα qui a été utilisée (PDB id : 1LBD [Bourguet et al., 1995]).

	PDB id	Ligand	CoA	Organisation	Résolution (Å)	Références
RXRα	1LBD	N/A	N/A	Monomère (Homotétramère)	2.70	[Bourguet et al. 1995]
	1FM9	9cRA	SRC-1	Hétérodimère (PPARγ)	2.10	[Gampe et al. 2000]

TABLE 5.1 – Récapitulatif des principales caractéristiques des structures cristallographiques de référence pour la forme libre et la forme liée à 9cRA du LBD de RXRα

L'ensemble des structures disponibles correspondant soit à la forme libre ou soit à la forme liée à 9cRA du LBD de RXRα humain ont été analysées à l'aide du programme PSS (**P**rotein **S**tructural **S**tatistics)

(version 1.2) développé au sein de l'équipe par T. Gaillard [Gaillard et al., 2013].

5.2.2 Protocole de simulation aMD

Le protocole de dynamique moléculaire accélérée utilisé dans le cadre de l'étude de la dynamique structurale du LBD de RXRα dans la forme libre et la forme liée à 9cRA est basé sur les travaux publiés par l'équipe de J. Andrew McCammon [Pierce et al., 2012] [Markwick and McCammon, 2011a]. Concernant la forme liée à 9cRA, le LBD du RXRα a été extrait du complexe RXRα/PPARγ et utilisé comme point de départ pour l'ensemble des calculs effectués durant ce projet. Tous les atomes lourds étaient présents dans les fichiers de coordonnées expérimentaux, excepté les chaines latérales des résidus de la boucle Ω (résidus 245 à 262) pour la forme liée du LBD de RXRα. Les chaines latérales de ces résidus ont été reconstruites à l'aide de CHARMM [Brooks et al., 2009]. Le placement des atomes d'hydrogène a été réalisé à l'aide du module HBUILD contenu dans CHARMM. Les correctifs NH_3^+ (NTER) et COO^- (CTER) sont appliqués aux extrémités N et C-terminale.

À pH physiologique (entre 7 et 7.4), les acides aminées basiques, Lysine et Arginine, sont chargés +1 et les acides aminées acides sont chargés -1. En ce qui concerne les histidines, leur l'état de protonation (HSE, HSD ou HSP) a été obtenu à l'aide de la méthode empirique implémentée dans PROPKA [Li et al., 2005a]. Ces résultats ont été comparés à d'autres obtenus antérieurement au sein de l'équipe [Martin, 2008] [Fidelak, 2006] et qui étaient basés sur des méthodes d'électrostatique des milieux continus. Les prédictions des deux méthodes sont généralement en accord, et le choix définitif a été validé en scrutant en détail les interactions établies par les histidines au sein des structures cristallographiques.

Histidine		288	315	331	333	338	406	435
forme libre LBD de RXRα	PROPKA	HSE	HSE	HSE	HSE	HSE	HSE	HSE
Forme liée (+9cRA) LBD de RXRα	[Martin 2008]	HSE	HSE	HSE	HSE	HSE	HSE	HSD
	PROPKA	HSD	HSE	HSE	HSE	HSE	HSE	HSE

TABLE 5.2 – Etat de protonation des histidines de la forme libre et la forme liée à 9cRA du LBD de RXRα obtenu par électrostatique des milieux continus[Martin, 2008] et par la méthode empirique PROPKA.

Les structures cristallographiques de la forme libre et la forme liée à 9cRA du LBD de RXRα ont été soumises au même protocole. Une première étape de minimisation de 700 pas de *Plus Grande Pente* (Stee-

pest Descent) a été réalisée dans le but d'éliminer les mauvais contacts stériques, utilisant une constante diélectrique dépendant de la distance avec un epsilon égale à 4 tout en réduisant les contraintes harmoniques imposées à la protéine au cours de la minimisation.

Une simulation de dynamique moléculaire en solvant explicite du LBD de RXRα a été réalisée en utilisant le programme NAMD (version 2.8/2.9/2.10 CPU et GPU) [Phillips et al., 2005] et le champ de force tout atome CHARMM 27 avec la correction CMAP. Dans le cas de la forme liée du LBD de RXRα, les paramètres et la topologie de 9cRA ont été repris du travail de thèse d'Élyette Martin [Martin, 2008]. Après une courte étape de minimisation d'énergie, présentée précédemment, une boîte rectangulaire contenant des molécules d'eau équilibrées TIP3P, a été construite et centrée sur le centre de masse de la protéine. Des ions K^+/Cl^- ont été ajoutés afin de neutraliser les charges du système dans un premier temps, puis dans un second temps afin d'atteindre une concentration en sels de 0.05 M. La simulation a été réalisée sous conditions périodiques. Les interactions électrostatiques de longue distance ont été traitées à l'aide du maillage particulier d'Ewald (PME) et un cut-off de 12 Å a été appliqué aux interactions van der Waals. Les liaisons entre les atomes d'hydrogène et les atomes lourds ont été contraintes en utilisant l'algorithme SHAKE et un pas d'intégration de 1 fs a été utilisé pour toutes les simulations.

Un protocole de dynamique moléculaire combinée à des étapes de minimisation d'énergie a été utilisé pour préparer la solvatation du système. Premièrement, les molécules d'eau ont été relaxées, avec la protéine et le ligand fixe, en procédant par une minimisation d'énergie de 1000 pas de gradient conjugué (GC) en utilisant une constante diélectrique constante (epsilon de 1). Puis l'ensemble du système est chauffé jusqu'à 600°K pendant 23 ps. Une seconde minimisation de 250 pas de gradient conjugué suivi d'une autre étape de chauffage d'une durée de 23 ps afin d'atteindre une température de 298.15°K. Deuxièmement, les contraintes sur la protéine et le ligand sont enlevées et une étape de minimisation de 2000 pas de gradient conjugué est appliquée suivie d'une étape de chauffage jusqu'à 298.15°K durant 13 ps, puis par une équilibration de 87 ps et une phase de production en maintenant un volume et une énergie constantes (NVE).

Afin de lancer une simulation de dynamique moléculaire accélérée, il est nécessaire de débuter par une simulation de dynamique moléculaire classique. Le temps de simulation varie de 10 à 50 ns de sorte que l'énergie potentielle et celle des angles dièdres se stabilisent. Dans notre cas, 10 trajectoires cMD indépendantes de 10 ns ont été générées au total, pour la forme libre et la forme liée à 9cRA du LBD de RXRα. Ces trajectoires cMD serviront à la fois de point de départ aux simulations de dynamique moléculaire accélérée, mais aussi à calculer les paramètres E et α nécessaire pour déterminer le biais énergétique qui sera appli-

qué à l'énergie potentielle et des angles dièdres à chaque pas d'intégration (voir Chapitre 2). L'ensemble des trajectoires cMD et aMD générés est résumé dans le tableau suivant :

	N_{res}	N_{atomes}	cMD	aMD	ΔV_{avg}	$\sigma_{\Delta V}$	$\delta_{\Delta V}$
forme libre du LBD de RXRα	238	31775 (6) 34537 (4)	10 x 10 ns	1μs (10 x 100 ns)	238	23	195
Forme liée (+9cRA) du LBD de RXRα	232	23147			240	24	195

TABLE 5.3 – Résumé des trajectoires aMD de la forme libre et la forme liée à 9cRA du LBD de RXRα avec N_{res} et N_{atomes}, le nombre de résidu et d'atomes, respectivement, puis ΔV_{avg}, $\sigma_{\Delta V}$ et $\delta_{\Delta V}$, la moyenne, l'écart-type et le rang (maximum - minimum) du biais d'énergie ΔV en kcal/mol, respectivement, sur l'ensemble des trajectoires aMD.

5.2.3 Méthodes d'analyse des trajectoires MD

5.2.3.1 RMSD et Fluctuations RMS

Les déviations RMS au cours du temps ainsi que les fluctuations RMS par résidu ont été obtenues en alignant les atomes du squelette peptidique appartenant aux régions structurales les plus stables d'après les données expérimentales HD, c'est-à-dire le sandwich d'hélices α composé de H1 (résidus 235 à 240), partie C-terminal de H3 (279 à 285), H5 (303 à 313), H7 (349 à 351), H8 (371 à 375), H9 (391 à 400) et H10 (416 à 425) du LBD de RXRα La déviation RMS ainsi que la fluctuations RMS ont été calculées à l'aide du module Python MDAnalysis [Michaud-Agrawal et al., 2011].

5.2.3.2 Structure secondaire

La structure secondaire a été obtenue à l'aide de DSSP [Kabsch and Sander, 1983]. Les conformations ont été extraites toutes les 10 picosecondes à partir des trajectoires aMD de la forme libre et de la forme liée à 9cRA du LBD de RXRα. Ce sont au total 100.000 conformations qui ont été analysées pour chacune des formes du LBD de RXRα.

5.2.3.3 Déplacements chimiques RMN

Les déplacements chimiques RMN ont été prédits en utilisant le programme SHIFTX+ issu de SHIFTX2 en utilisant un pH égal à 7.4 et une température égale de 298.15 K [Han et al., 2011]. Les conformations ont été extraites toutes les 10 picosecondes à partir des trajectoires aMD de la forme libre et de la forme liée à 9cRA du LBD de RXRα. Au total, ce sont 100.000 conformations qui ont été analysées pour chacune des formes du LBD de RXRα. Quant aux structures cristallographiques, les atomes d'hydrogène ont été préalablement ajoutés à l'aide REDUCE [Word et al., 1999].

L'ensemble des scripts Python ayant été utilisés pour calculer les déplacements chimiques à partir des conformations extraites des trajectoires MD, ainsi que pour comparer les déplacements chimiques prédits δ_{pred} aux déplacements chimiques expérimentaux δ_{exp} sont disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/shift`

5.2.3.4 Échange Hydrogène/Deutérium

L'échange Hydrogène/Deutérium a été calculé à partir des constantes de vitesses intrinsèques k_{int} par proton amide HN obtenues à l'aide du modèle empirique développé par Bai *et al.* [Bai et al., 1993], ainsi qu'à partir des facteurs de protection PF de chaque proton amide HN obtenus des trajectoires aMD de la forme libre et liée à 9cRA du LBD de RXRα. Les facteurs de protections PF ont été obtenu à partir du nombre de liaisons hydrogène établies entre chaque proton HN et le solvant selon le modèle simple et le modèle concerté, à l'aide du module python MDAnalysis [Michaud-Agrawal et al., 2011]. Au total, ce sont 1.000.000 conformations qui ont été analysées pour chacune des formes du LBD de RXRα.

L'ensemble des scripts Python pour déterminer les facteurs de protection PF à partir des trajectoires MD ainsi que prédire l'incorporation en deutérium au cours du temps est disponible librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/hdx`

5.2.3.5 Analyse sur Composante Principale

Un très grand nombre de conformations ont été générées durant les simulations aMD de la forme libre et la forme liée à 9cRA du LBD de RXRα, avec au total 10^6 conformations pour chaque forme. Afin de visualiser l'espace conformationnel échantillonné au cours des trajectoires aMD et identifier les conformations les plus stables, une Analyse sur Composante Principale (PCA) a été effectuée dans un premier temps. Pour cela, tous les angles dièdres ϕ et ψ ainsi que les pseudo angles dièdres Cα ont été extraits à partir de 100 000 conformations, avec un intervalle de 10 ps. Au total, 235 et 229 pseudo angles

dièdres ont été extraits pour la forme libre et la forme liée à 9cRA du LBD de RXRα, respectivement. Les composantes principales ont été obtenues selon la méthode décrite dans le chapitre 3. En ce qui concerne le choix des composantes principales, celui-ci a été fait en fonction de leur importance et en prenant un nombre n représentant au moins 90% de la variance globale.

5.2.3.6 Stochastic Proximity Embedding

De même, nous avons utilisé l'algorithme pSPE, qui est une méthode non linéaire de réduction de la dimensionnalité des données, en utilisant la distance dihédrale comme métrique, au lieu du RMSD (voir chapitre 3 pour plus de détails).

Ainsi, tous les pseudo angles dièdres entre 4 Cα consécutifs ont été extraits de chaque conformation présente dans les trajectoires aMD à l'aide du module Python MDAnalysis [Michaud-Agrawal et al., 2011]. Au total, 235 et 229 pseudo angles dièdres ont été extraits pour la forme libre et la forme liée à 9cRA du LBD de RXRα, respectivement. Ainsi que détaillé dans le chapitre 3, la méthode pSPE utilise le paramètre r_c, rayon de proximité utilisé pour déterminer la région où la distance euclidienne est une bonne approximation (*cutoff*) de la distance géodésique. Il n'existe pas de règle permettant de fixer à priori la valeur du paramètre r_c menant à la meilleure réduction de dimensionnalité. L'approche empirique que nous avons implémentée (voir Chapitre 3) est de tester, systématiquement les valeurs de r_c entre 0.1 et 0.5 par intervalles de 0.01. Ces tests sont effectués sur 10 000 structures extraites des trajectoires, pour un nombre fixe (5000) de cycles d'optimisation, et ils sont répétés cinq fois pour chaque valeur du paramètre r_c. L'examen de l'évolution du stress, et de la corrélation γ calculés pour chaque valeur de r_c (voir Chapitre 3) permet de déterminer empiriquement la valeur de r_c qui offre le meilleur compromis entre une valeur maximale de la corrélation γ et minimale du stress. Une fois la valeur optimale de r_c déterminée de cette manière, un certain nombre de tests supplémentaires sont réalisés afin d'assurer de la robustesse de la procédure. Ainsi, le nombre de cycles d'optimisation C est systématiquement modifié, de 10 cycles à 100 000 cycles (valeurs intermédiaires : 50, 100, 500, 1000, 5000, 10 000, 50 000) et ce pour des nombres différents de conformations (10 000, 50 000 et 100 000). Chacun de ces tests est répété cinq fois vu le caractère stochastique de l'algorithme pSPE. Ceci nous permet d'estimer le nombre minimal de cycles d'optimisation C nécessaire pour obtenir la meilleure réduction de dimensionnalité, avec le r_c choisi. L'utilisation de différents échantillons de conformations permet également de tester la stabilité et la vitesse de convergence de la méthode. Une fois l'ensemble de ces tests réalisés, la méthode est appliquée sur l'ensemble des 1 000 000 conformations.

L'ensemble des scripts Python pour extraire les angles dièdres à partir de trajectoires MD, déterminer les valeurs optimales des paramètres r_c et C, ainsi que la méthode pSPE sont disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/unrolr`

5.2.3.7 Repondération de l'énergie des trajectoires aMD

Afin de repondérer l'énergie des conformations issues des trajectoires aMD de forme libre et la forme liée à 9cRA du LBD de RXRα, nous utiliserons comme coordonnées de réaction celles obtenues précédemment à l'aide de la méthode pSPE, puis comme méthode de repondération la série de McLaurin d'ordre 10 avec une température de 298.15 K et un intervalle (*bin size*) de 0.05 pour la forme libre et de 0.075 pour la forme liée à 9cRA.

L'ensemble des scripts Python ayant servi à repondérer l'énergie des trajectoires aMD sont aussi disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/reweight`

5.2.3.8 Visualisation des conformations

Pour permettre de visualiser l'espace conformationnel échantillonné durant une simulation MD, le module Python Bokeh [Bokeh Development Team, 2014] et PyMOL [Schrödinger, LLC, 2015] ont été utilisés. Ainsi, à partir de coordonnées de réaction quelconques, provenant soit de la méthode pSPE ou depuis n'importe quelle méthode d'apprentissage automatique, telle que la PCA, il est possible de visualiser les conformations générées.

L'ensemble des scripts Python sont disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/visualize`

5.3 Résultats

5.3.1 Analyse des structures cristallographiques

Afin de pouvoir comparer ultérieurement les structures cristallographiques du LBD de RXRα aux trajectoires aMD de la forme libre et la forme liée à 9cRA, nous avons sélectionné dans la PDB les structures qui correspondent strictement aux deux formes étudiées. Nous considérons comme inactives les structures cristallographiques, où l'hélice activatrice H12 n'est pas repliée contre le domaine LBD, et où cette absence de repliement ne parait pas induite, ni par un ligand ni par un partenaire avec lequel RXRα forme un

hétérodimère. Dans le cas de la forme liée, on ne considèrera que les structures en complexe avec 9cRA et dont l'hélice H12 est repliée contre le domaine en position active.

5.3.1.1 Forme inactive du LBD de RXRα

5.3.1.1.1 Sélection des structures cristallographiques disponibles

Parmi toutes les structures disponibles dans la PDB du LBD de RXRα humain (68 au total), seules 11 structures pourraient être considérées comme représentant la forme libre. Leurs PDB id sont les suivants : 1G1U [Gampe et al., 2000b], 1G5Y [Gampe et al., 2000b], 1LBD [Bourguet et al., 1995], 2ACL [Jaye et al., 2005], 3FAL [Chao et al., 2008], 3FC6 [Washburn et al., 2009], 3NSP [Zhang et al., 2011b], 3NSQ [Zhang et al., 2011b], 3R2A [Zhang et al., 2011a], 3R29[Zhang et al., 2011a], 4ZO1 [Kojetin et al., 2015] (dernière mise à jour : 2016-08-04).

De toutes ces structures, l'hélice H12 n'est soit pas repliée contre le domaine, soit aucun ligand n'est présent au sein de la structure cristallographique. Cependant, des structures ne présentant pas H12 repliée contre le LBD peuvent contenir un ligand. C'est le cas pour 1G5Y (chaines B et C) en complexe avec atRA, 2ACL (chaines A,C,E et G) en complexe avec 9cRA et formant un hétérodimère avec LXRβ, 3FAL (chaines A et C) en complexe avec 9cRA et LXRβ, de même pour 3FC6 (chaines A et C) en complexe avec 9cRA et mLXRα, 3NSQ (chaine B) en complexe avec Danthron, un antagoniste, et 3R2A (chaines A et C), en complexe avec Rhein, lui aussi antagoniste. D'autres ne possèdent pas de ligand, mais sont très proches structuralement de structures cristallographiques qui en possèdent un. C'est le cas pour 3R2A (chaines B et D) proche de 3R2A (chaines A et C) avec un RMSD de 0.56 Å, 3R29 (chaines A et B) proche aussi de 3R2A (chaines A et C) avec un RMSD de 0.38 Å, 3NSP (chaine B) proche de 3NSQ (chaine B) avec un RMSD de 0.31 Å et de même pour 1G1U (chaines B et C) proche de 1G5Y (chaines B et C) avec un RMSD de 0.17 Å (figure5.1). Ainsi il ne reste plus que 1G1U (chaines A et D), 1G5Y (chaines A et D), 1LBD, 3NSP (chaine A), 3NSQ (chaine A) et 4ZO1 qui ne possèdent aucun ligand et qui ne sont proches structuralement d'aucune structure en ayant un. Sauf que parmi ces structures cristallographiques, certaines sous-unités présentent des cavités assez larges et qui donc pourraient accueillir un ligand. 1G1U (chaines A et D) et 1G5Y (chaines A et D) présentent cette caractéristique avec des volumes semblables à celles retrouvées dans 1G1U (chaines B et C) et 1G5Y (chaines B et C). De plus, toutes ces structures sont très proches avec un RMSD de 0.16 Å, bien que de la densité électronique correspondant à un ligand n'ait été trouvée que dans deux monomères de 1G5Y (chaines B et C). Concernant 1G1U (chaines A et D) et 1G5Y (chaines A et D), il n'y a pas de densité électronique au sein des cavités. Mais au vu de la configuration

et de la nature des résidus qui bordent ces cavités, majoritairement de type hydrophobe, la présence d'un ligand est très fortement suspectée. Dans le cas de 4ZO1, la conformation particulière de l'hélice H12 n'est pas induite pas un ligand, mais par TRβ, en complexe avec un agoniste (T3), avec lequel RXR forme un hétérodimère [Kojetin et al., 2015] (figure5.2).

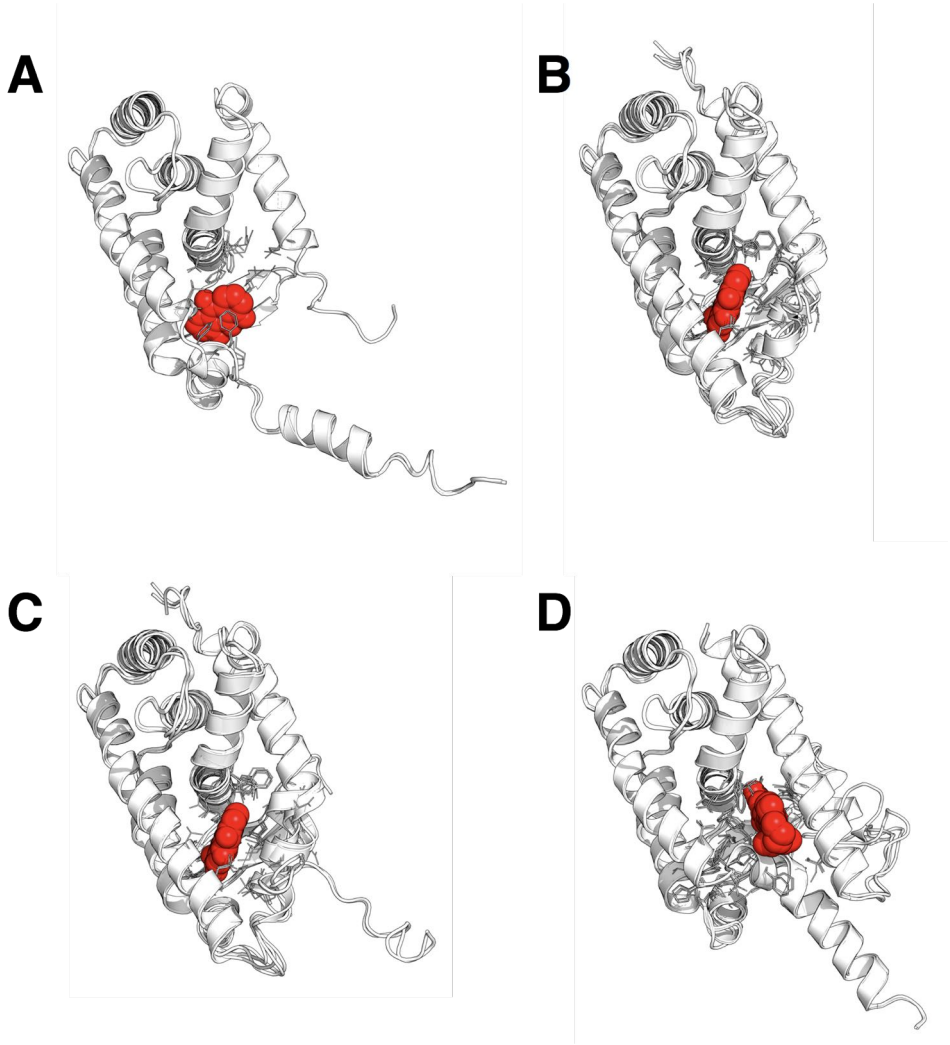


FIGURE 5.1 – Comparaison entre plusieurs structures cristallographiques du LBD de RXRα, en complexe avec un ligand, mais dont l'hélice H12 n'est pas en configuration active, avec d'autres structures cristallographiques très proches structuralement (< 0.6 Å) dans lesquelles aucun ligand n'est visible. Superposition de (A) 3NSP (chaine B) sur 3NSQ (chaine B), (B) 3R29 (chaines A et B) sur 3R2A (chaines A et C), (C) 3R2A (chaines A et B) sur 3R2A (chaine A et C) et (D) 1G1U (chaines B et C) avec 1G5Y (chaines B et C).

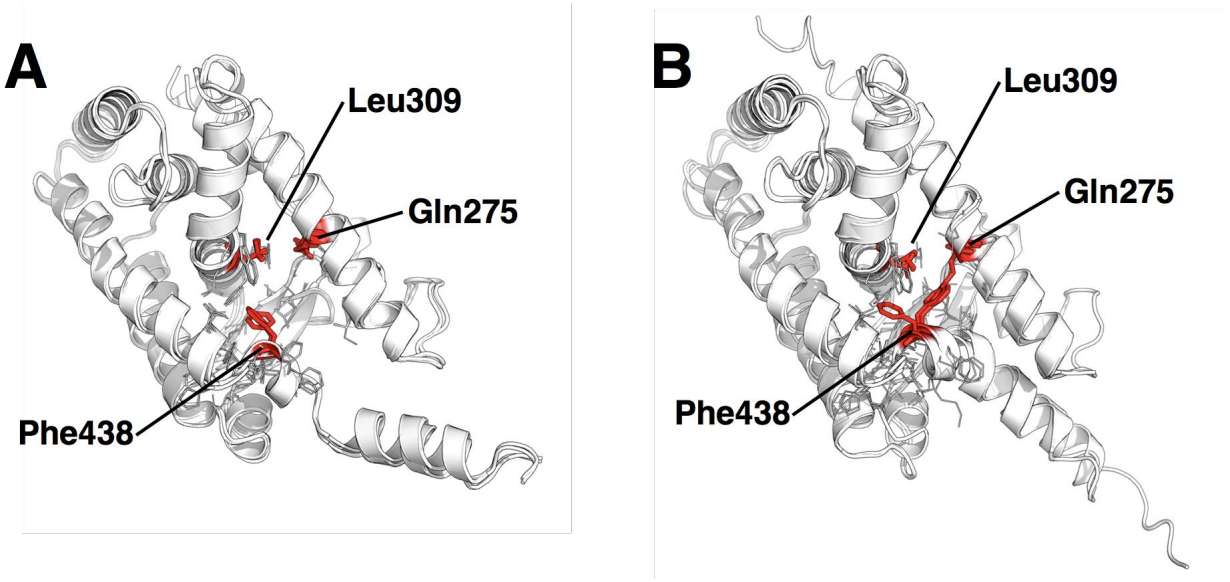


FIGURE 5.2 – Comparaison entres les structures cristallographiques (A) 1G1U (chaines A et D), 1G5Y (chaines A et D) où la présence d'un ligand est fortement suspectée et (B) 1LBD, 3NSP (chaine A) et 3NSQ (chaine A) qui ne possèdent pas de cavités pouvant accueillir un ligand. En rouge sont représentés les résidus Gln275, Leu309 et Phe438.

Au final, il ne reste plus que les structures cristallographiques 1LBD, 3NSP (chaine A), 3NSQ (chaine A) dont le volume des cavités présentes est très réduit, ne laissant pas à la place à un ligand de type acide-gras comme 9cRA. Ce sont les seules structures cristallographiques qui soient indubitablement représentatives de conformations inactives du LBD de RXRα. De ces structures, plusieurs observables expérimentales (HDX, déplacements chimiques RMN, ...) pourront être extraites et comparées à celles obtenues des trajectoires aMD de la forme libre du LBD de RXRα.

5.3.1.1.2 Fluctuations RMS

L'analyse des fluctuations RMS du squelette peptidique par résidu au sein des structures cristallographiques du LBD de RXRα permet d'identifier les régions les plus variables structuralement. Ainsi les régions les plus mobiles sont l'hélice H3, la région comprenant le feuilletβ et H6, résidu 323 à 343, la partie C-terminale de H7 ainsi que la boucle H7/H8, résidus 352 à 362. De même, nous retrouvons la boucle H10/H11, résidus 400 à 415 et enfin la partie C-terminale avec H12, résidus 434 à 456. Concernant toute la région N-terminale, résidus 226 à 263, il n'existe qu'un seul modèle de cette région retrouvée

dans la structure cristallographique que nous utiliserons comme référence (PDB id : 1LBD). C'est le cas aussi pour la partie C-terminale où les résidus 457 à 462 sont aussi manquants, exception faite encore une fois de la structure cristallographique de référence. Ainsi l'absence de modèle dans les autres structures cristallographiques suggère que ces régions sont très flexibles.

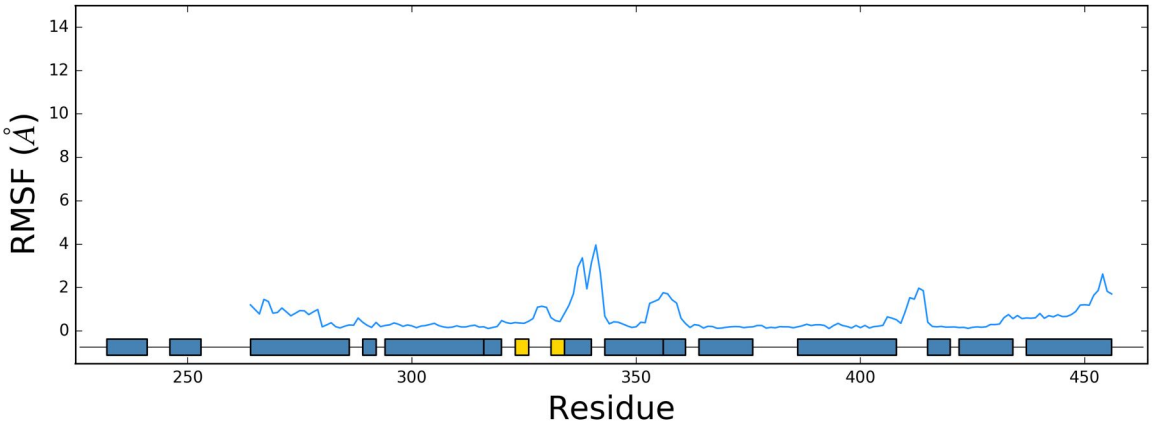


FIGURE 5.3 – Fluctuations RMS du squelette peptidique au sein des structures cristallographiques, calculée à l'aide du programme PSS [Gaillard et al., 2013], du LBD de RXRα en conformation inactive (PDB id : 1LBD, 3NSP (chaine B), 3NSQ (chaine B)).

5.3.1.2 Forme active, en complexe avec 9cRA, du LBD de RXRα

5.3.1.2.1 Sélection des structures cristallographiques disponibles

Parmi toutes les structures cristallographiques du LBD de RXRα, seulement 10 correspondent à une conformation active en complexe avec 9cRA et dont les PDB id sont les suivants : 1FBY [Egea et al., 2000], 1FM6 [Gampe et al., 2000a], 1FM9 [Gampe et al., 2000a], 1K74 [Xu et al., 2001], 1XLS [Suino et al., 2004], 3DZY [Chandra et al., 2008], 3E00 [Chandra et al., 2008], 3OAP [Xia et al., 2010], 3UVV [Putcha et al., 2012] et 4NQA [Lou et al., 2014]). Certaines de ces structures (PDB id : 3DZY, 3E00 et 4NQA) possèdent aussi le domaine de liaison à l'ADN (DBD) relié au LDB par le biais de la région charnière.

5.3.1.2.2 Fluctuations RMS

Concernant cette fois-ci les fluctuations RMS au sein des structures cristallographiques de la forme liée à 9cRA du LBD de RXRα, nous remarquons que le feuillet β et H6 ainsi que H7 et la boucle H10/H11 sont

plus rigides que dans la forme apo. En ce qui concerne la partie N et C-terminale, la boucle H1/H3 et la toute extrémité N-terminale entourant H1 sont très flexibles, tandis qu'H1 reste stable. L'hélice H12 par contre présente une flexibilité plus importante que le feuillet β au sein même des structures cristallographiques (Figure 5.4).

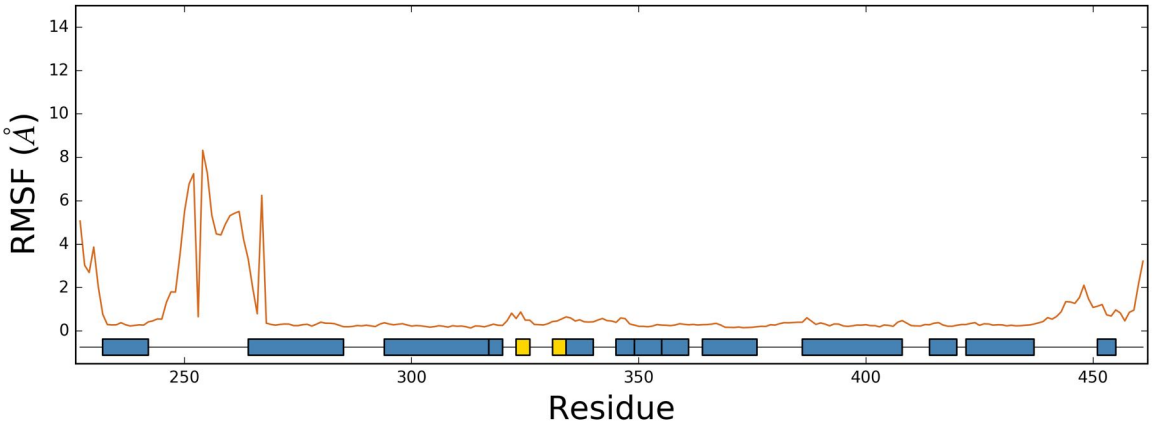


FIGURE 5.4 – Fluctuations RMS du squelette peptidique au sein des structures cristallographiques, calculées à l'aide du programme PSS [Gaillard et al., 2013], du LBD de RXRα en conformation active en complexe avec 9cRA (PDB id : 1FBY, 1FM6, 1FM9, 1K74, 1XLS, 3DZY, 3E00, 3OAP, 3UVV et 4NQA).

Il est important de souligner qu'en ce qui concerne la région entre les hélices H1 et H3, résidus 243 à 263, seules les structures cristallographiques 1FM6, 1FM9, 1K74 et 1XLS possèdent un modèle de cette boucle déposé dans la PDB, mais les données permettant de proposer ce modèle sont manquantes. Dans le cas de 1FM6, 1FM9 et 1XLS, l'examen des cartes de densité électronique ne révèle pas une densité suffisante pour créer un modèle fiable et continu de la boucle entre les hélices H1 et H3. Pour 1K74, les données expérimentales ne sont pas disponibles, ainsi il est impossible de vérifier la légitimité du modèle construit. En résumé, de toutes les structures cristallographiques du LBD de RXRα, toutes formes confondues, inactives et actives en complexe avec 9cRA, seule la cristallographique de référence de la forme libre (PDB id : 1LBD) possède un modèle de la boucle H1/H3 appuyé par la densité déposée dans le PDB.

5.3.2 Analyse des trajectoires aMD

5.3.2.1 Forme libre du LBD de RXRα

5.3.2.1.1 RMSD global

Le RMSD global nous donne une information quant à la stabilité des trajectoires aMD de la forme libre du LBD de RXRα.

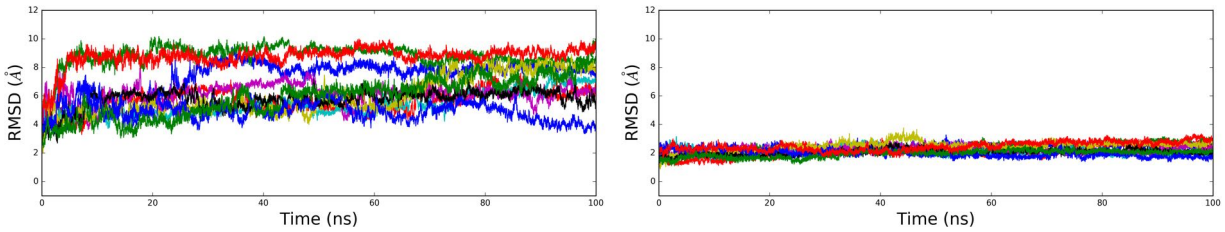


FIGURE 5.5 – RMSD global du squelette peptidique au cours du temps, calculé pour chaque trajectoire aMD de la forme libre du LBD de RXRα. Le RMSD a été calculé en alignant toutes les structures issues des trajectoires aMD sur la structure cristallographique de référence (PDB id : 1LBD) en ne considérant que la partie la plus stable, c'est-à-dire le sandwich d'hélices α composé de H1 (résidus 235 à 240), partie C-terminal de H3 (279 à 285), H5 (303 à 313), H7 (349 à 351), H8 (371 à 375), H9 (391 à 400) et H10 (416 à 425). À gauche, le RMSD de l'ensemble du domaine au cours du temps et à droite, le RMSD du sandwich d'hélices α seul.

On observe une rapide augmentation du RMSD global dès le début (10 premières nanosecondes) de chaque trajectoire aMD, suivit d'une stabilisation entre 4 et 10 Å suivant la trajectoire aMD (Figure 5.4).

Néanmoins, nous n'observons pas de phénomène de dénaturation (augmentation progressive du RMSD sans stabilisation), qui serait du à une trop forte accélération appliquée au système (Figure 5.4). Ceci est confirmé par la très grande stabilité du sandwich d'hélices α avec un RMSD au cours du temps qui oscille autour de 2.5 Å pour chaque trajectoire aMD.

5.3.2.1.2 Fluctuations RMS

Quant aux fluctuations RMS par résidu calculées à partir des trajectoires aMD, ces dernières nous montrent que les régions les plus flexibles sont la région N-terminale, la boucle entre H1 et H3 ainsi que la partie N-terminale de H3, la boucle H3/H4, la région comprenant le feuillet β et H6, la partie C-terminale

de H7 et la partie N-terminale de H8, la boucle H8/H9, les parties N et C-terminales de H9 et la boucle H9/H10 ainsi que la partie C-terminale avec H12 (Figure 5.6).

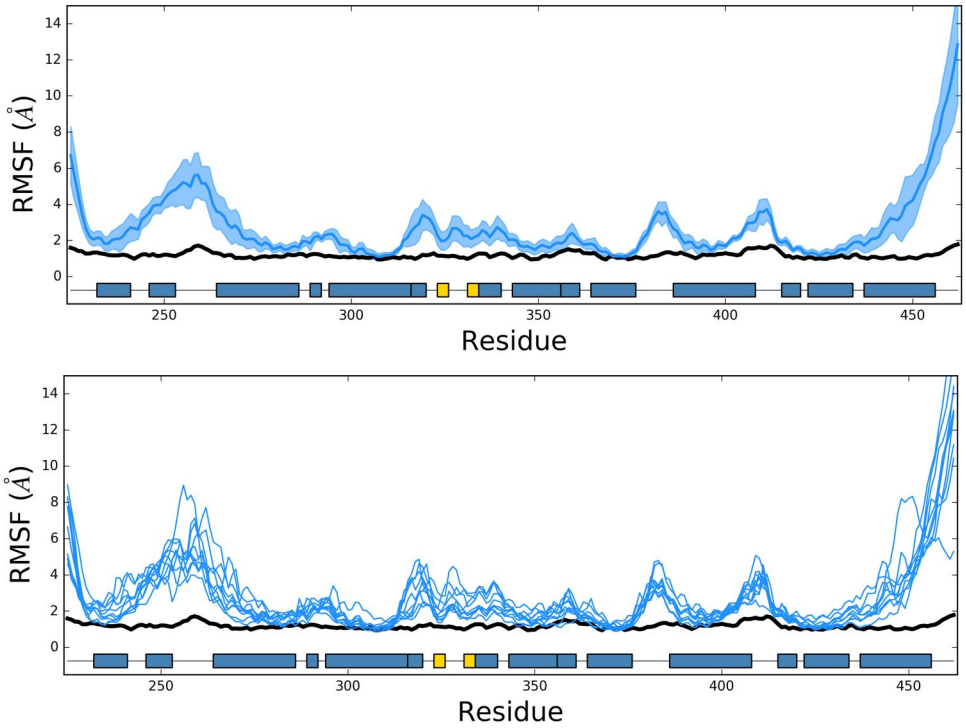


FIGURE 5.6 – Fluctuations RMS par résidu du squelette peptidique, moyennées sur l'ensemble des trajectoires aMD (en haut) ainsi que le détail par trajectoire aMD (en bas) de la forme libre du LBD de RXRα. En bleu, les fluctuations RMS et en noir les fluctuations RMS calculées à partir des facteurs B de la structure cristallographique de référence de la forme libre du LBD de RXRα (PDB id : 1LBD).

Ces fluctuations sont similaires avec celles calculées à partir des structures cristallographiques avec PSS, bien qu'elles soient plus élevées globalement. On constate néanmoins une différence au niveau de la boucle H8/H9. Au sein des structures cristallographiques, les fluctuations RMS moyennes des résidus de la boucle H8-H9 sont proches de zéro. Quant à celles observées dans les trajectoires aMD, celles-ci varient entre 2 et 4 Å. Ceci est dû au fait que la boucle H8-H9 participe à la surface de dimérisation entre les monomères au sein des structures cristallographiques, ce qui n'est pas le cas des trajectoires aMD où seul un monomère est étudié. Concernant la partie N-terminale, aucune comparaison n'est possible.

5.3.2.1.3 Structure secondaire

L'analyse de la structure secondaire, à l'aide de DSSP, de la forme libre du LBD de RXRα montre que globalement celle-ci est conservée au cours des trajectoires aMD par rapport à la structure cristallographique de référence (PDB id : 1LBD) (Figure 5.7). Les hélices qui composent le sandwich d'hélices α, composé de H1, la partie N-terminale de H3, H4 et H5 (excepté en sa partie C-terminale), H7, H8, H9 et H10 restent en hélice plus de 80% du temps au cours des trajectoires aMD, et de même pour le feuillet β.

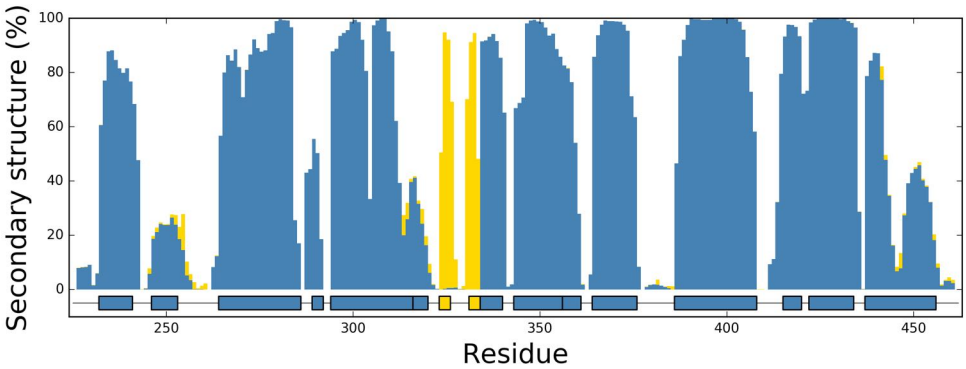


FIGURE 5.7 – Proportion de la structure secondaire (en %) par résidu, calculé à l'aide de DSSP à partir des trajectoires aMD de la forme libre du LBD de RXRα. En bleu sont représentés les résidus en hélice (α, π et 3_{10}) et en jaune les résidus formant un feuillet ou un pont β.

Des différences importantes sont observées dans la boucle H1/H3 où les résidus 246 à 254 formant l'hélice H2 dans la structure cristallographique ne sont en hélice qu'au maximum pendant 20% des trajectoires aMD, de plus ces mêmes résidus peuvent former des contacts β (< 5%). Quant à la partie C-terminale de H5, nous remarquons que les résidus qui la composent ne forment pas une hélice stable (< 40%) et peuvent aussi former des contacts β. Finalement, nous remarquons très clairement que l'hélice H12 ne reste pas entière (un seul bloc), comme c'est le cas dans la structure cristallographique de référence de la forme libre du LBD de RXRα (PDB id : 1LBD), mais qu'elle est scindée en deux parties bien distinctes. La première partie, résidus 437 à 444, est en hélice durant 80% des trajectoires aMD, contrairement à la seconde moitié, résidus 447 à 455, qui est présente moins de 50 % du temps de trajectoire. Et de même pour l'hélice H3, où la propension à rester en hélice est plus faible dans la partie N-terminale qu'en C-terminal.

5.3.2.2 Forme liée à 9cRA du LBD de RXRα

5.3.2.2.1 RMSD global

De même que pour la forme libre, nous n'observons pas de phénomène de dénaturation de la forme liée à 9cRA du LBD de RXRα (Figure 5.5). De manière générale, la forme liée reste plus stable que la forme libre avec un RMSD global oscillant autour de 4 Å. Quant au sandwich d'hélices α, nous observons une très légère augmentation du RMSD, partant de 1.5 Å environ jusqu'à une valeur autour de 2.5 Å à partir de 60 ns (Figure 5.5). En ce qui concerne cette partie en hélices, le comportement de la forme liée à 9cRA et la forme libre est donc très similaire lors des trajectoires aMD.

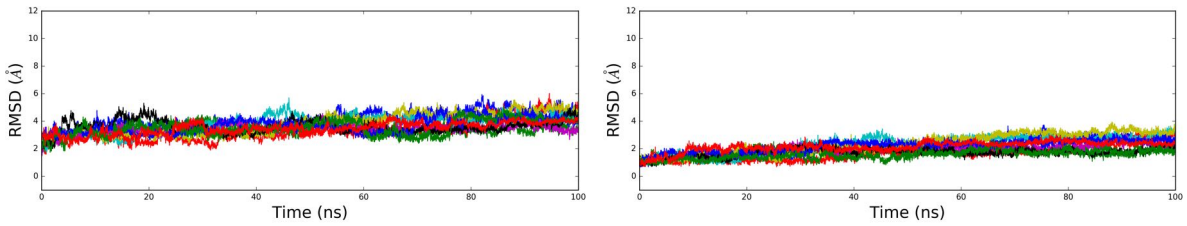


FIGURE 5.8 – RMSD global du squelette peptidique au cours du temps, calculé pour chaque trajectoire aMD de la forme liée à 9cRA du LBD de RXRα. Le RMSD a été calculé en alignant toutes les structures issues des trajectoires aMD sur la structure cristallographique de référence (PDB id : 1FM9) en ne considérant que la partie la plus stable, c'est-à-dire le sandwich d'hélices α composé de H1 (résidus 235 à 240), partie C-terminal de H3 (279 à 285), H5 (303 à 313), H7 (349 à 351), H8 (371 à 375), H9 (391 à 400) et H10 (416 à 425). À gauche, le RMSD de l'ensemble du domaine au cours du temps et à droite, le RMSD du sandwich d'hélices α seul.

5.3.2.2.2 Fluctuations RMS

L'analyse des fluctuations RMS calculées à partir des trajectoires aMD du RXRα montre que les régions les plus dynamiques sont la partie N-terminale du domaine, les boucles H1/H3 et H3/H4, ainsi que la région du feuillet β avec la partie C-terminale de H5 et H6, les boucles H8/H9 et H0/H10 et la partie C-terminale avec H12 (Figure 5.9).

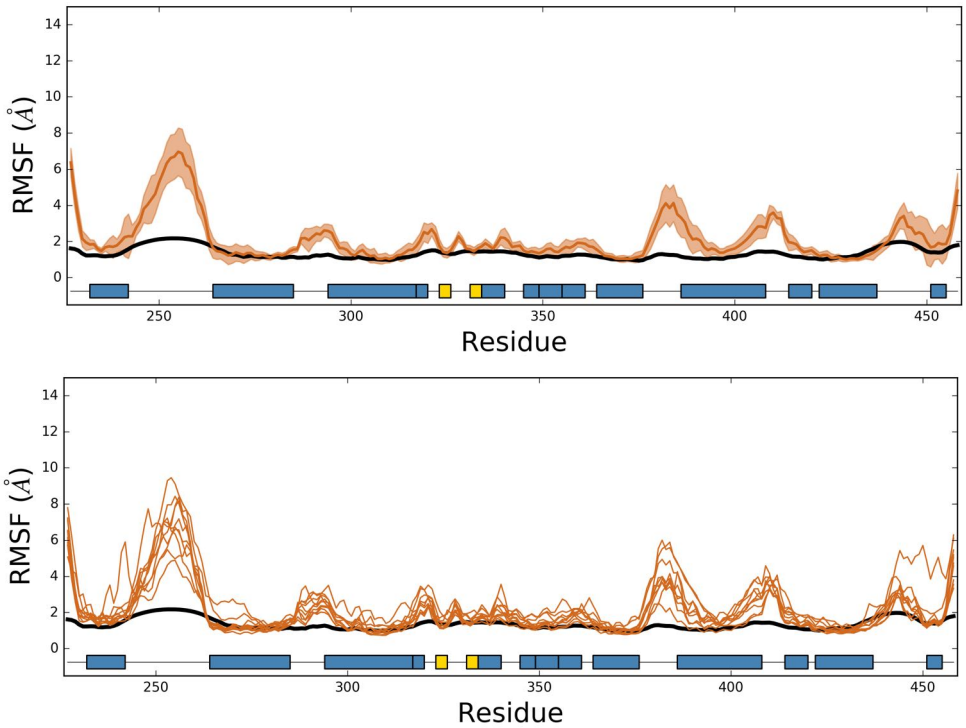


FIGURE 5.9 – Fluctuations RMS par résidu du squelette peptidique, moyennées sur l'ensemble des trajectoires aMD (en haut) ainsi que le detail par trajectoire aMD (en bas) de la forme liée du LBD de RXRα. En orange, les fluctuations RMS et en noir les fluctuations RMS calculées à partir des facteurs B de la structure cristallographique de référence de la forme active, en complexe avec 9cRA, du LBD de RXRα (PDB id : 1FM9).

Ainsi les principales différences ne situent pas tant au niveau de la localisation des régions les plus flexibles, car nous retrouvons les mêmes régions entre la forme libre et la forme liée, mais plutôt au niveau de l'amplitude de leurs fluctuations qui est réduite du à la présence de 9cRA. Dans le cas de la forme liée, H1 et la partie N-terminale de H3 sont plus stables (< 2 Å). L'amplitude des fluctuations est également réduite dans la région du feuillet β et dans la partie C-terminale H12. La comparaison des différentes trajectoires aMD montre en effet qu'H12 reste en général stable, à l'exception d'une trajectoire aMD, ce qui diffère sensiblement des résultats de la forme apo. Cependant, la présence de 9cRA ne semble pas avoir d'effet sur la dynamique de la boucle H1/H3 ni sur les boucles H3/H4 et H9/H10. On observe par contre une plus grande flexibilité de la boucle H8/H9 et de la partie N-terminale de H9, dont le lien avec la présence du ligand devrait être analysé de manière plus approfondie. En résumé, la flexibilité des régions à

proximité de 9cRA est réduite, mais sans complètement les rendre rigides.

5.3.2.2.3 Structure secondaire

De même que les trajectoires aMD de la forme libre du LBD de RXRα, les hélices formant le sandwich d'hélices α de la forme liée à 9cRA du LBD de RXRα restent bien en conformation hélicale tout au long des trajectoires aMD (> 90 %), tout comme le feuillet β (Figure 5.10). Concernant la boucle H1/H3, nous pouvons remarquer cette fois-ci que la propension à former une hélice est plus importante dans le cas de la forme liée (> 30 %) qu'elle ne l'est dans la forme libre (au maximum 20 %). Quant à H12, elle garde également sa conformation en hélice contrairement à ce qui était observé dans la forme libre.

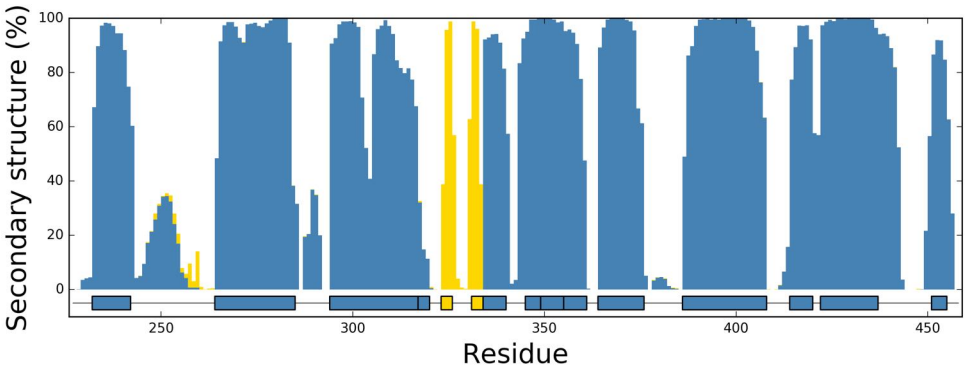


FIGURE 5.10 – Proportion de la structure secondaire (en %) par résidu, calculé à l'aide de DSSP à partir des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα. En bleu sont représentés les résidus en hélice (α, π et 3_{10}) et en jaune les résidus formant un feuillet ou un pont β.

5.3.3 Analyse des déplacements chimiques RMN

5.3.3.1 Forme libre du LBD de RXRα

La comparaison des déplacements chimiques secondaires $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}_\alpha$ prédits à partir des trajectoires aMD avec ceux obtenus expérimentalement $\Delta\delta_{\text{exp}}$ $^{13}\text{C}_\alpha$, montre que ceux-ci correspondent bien de manière générale, avec un RMSD de 1.22 ppm (Figure 5.11). En comparaison, le RMSD entre les $\Delta\delta_{\text{1LBD}}$ $^{13}\text{C}_\alpha$ prédits à partir de la structure de référence (PDB id : 1LBD) et $\Delta\delta_{\text{exp}}$ $^{13}\text{C}_\alpha$ est de 1.618 ppm et de 1.279 ppm lorsqu'on les compare aux $\Delta\delta_{\text{Xray}}$ $^{13}\text{C}_\alpha$ prédits à partir de l'ensemble des structures cristallographiques de la forme inactive (PDB id : 1LBD, 3NSP, 3NSQ).

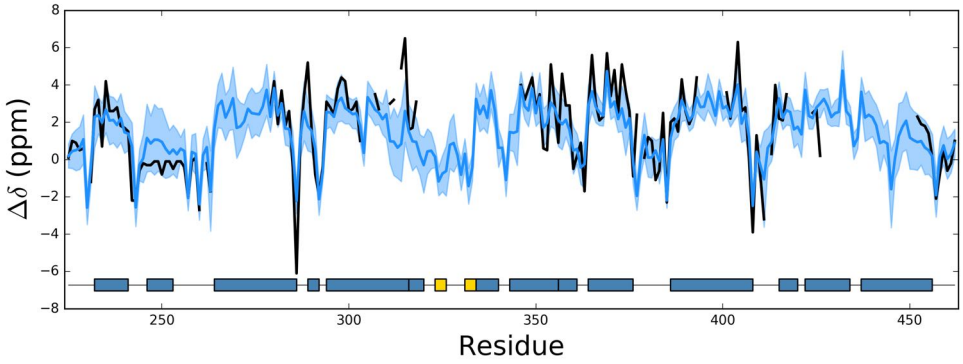


FIGURE 5.11 – Comparaison entre les déplacements chimiques secondaires $\Delta\delta_{\text{exp}}$ ^{13}Cα obtenus expérimentalement (en noir) et ceux prédits $\Delta\delta_{\text{aMD}}$ ^{13}Cα à partir des trajectoires aMD de la forme libre, du LBD de RXRα (en bleu).

Ainsi à partir de ces résultats, on peut conclure que les trajectoires aMD reproduisent globalement mieux les données expérimentales $\Delta\delta_{\text{exp}}$ ^{13}C$_\alpha$ que la structure cristallographique de référence seule. Néanmoins, l'utilisation de plusieurs structures cristallographiques de la forme libre du LBD de RXRα permet également d'améliorer l'accord entre les déplacements calculés et expérimentaux $\Delta\delta_{\text{exp}}$ ^{13}C$_\alpha$ (Table 5.4).

	^{13}C$_\alpha$	^{13}C$_\beta$	^{13}C	^{15}N	^{1}HN
aMD	1.22	1.45	1.27	3.00	0.55
Xray (1LBD)	1.62	1.34	1.32	3.33	0.61
Xray (All)	1.28	1.22	1.24	2.98	0.53

TABLE 5.4 – Détails par élément chimique, ^{13}C$_\alpha$, ^{13}C$_\beta$, ^{13}C, ^{15}N et ^{1}HN, du RMSD obtenu entre les déplacements chimiques RMN secondaires obtenus expérimentalement $\Delta\delta_{\text{exp}}$ et ceux prédits à partir des trajectoires aMD $\Delta\delta_{\text{aMD}}$, ainsi que ceux prédits à partir de la structure cristallographique seule $\Delta\delta_{\text{1LBD}}$ (PDB id : 1LBD) et à partir de toutes les structures cristallographiques de la forme inactive du LBD de RXRα disponibles $\Delta\delta_{\text{Xray}}$ (PDB id : 1LBD, 3NSP et 3NSQ).

Les déplacements chimiques secondaires $\Delta\delta$ ^{13}C, ^{15}N, ^{1}HN, sont également mieux prédit par les trajectoires aMD que par la structure de référence 1LBD, à l'exception des déplacements $\Delta\delta$ ^{13}C$_\beta$. L'utilisation d'un ensemble de structures cristallographiques donne des résultats proches de ceux observés à partir de la trajectoire aMD (Table 5.4).

La comparaison entre les déplacements chimiques δ_{1LBD} $^{13}C_\alpha$ (ou sur l'ensemble des structures δ_{Xray} $^{13}C_\alpha$) et δ_{aMD} $^{13}C_\alpha$ avec δ_{exp} $^{13}C_\alpha$, permet d'identifier les régions où les déplacements sont mal prédits à partir des structures. Les résultats montrent que les régions présentant les divergences les plus larges se situent au niveau de la boucle entre entre les hélices H1 et H3 ainsi que la partie C-terminale de H5, que ça soit au sein de la structure de cristallographique de référence seule ou au sein de l'ensemble des structures cristallographiques (Figure 5.12).

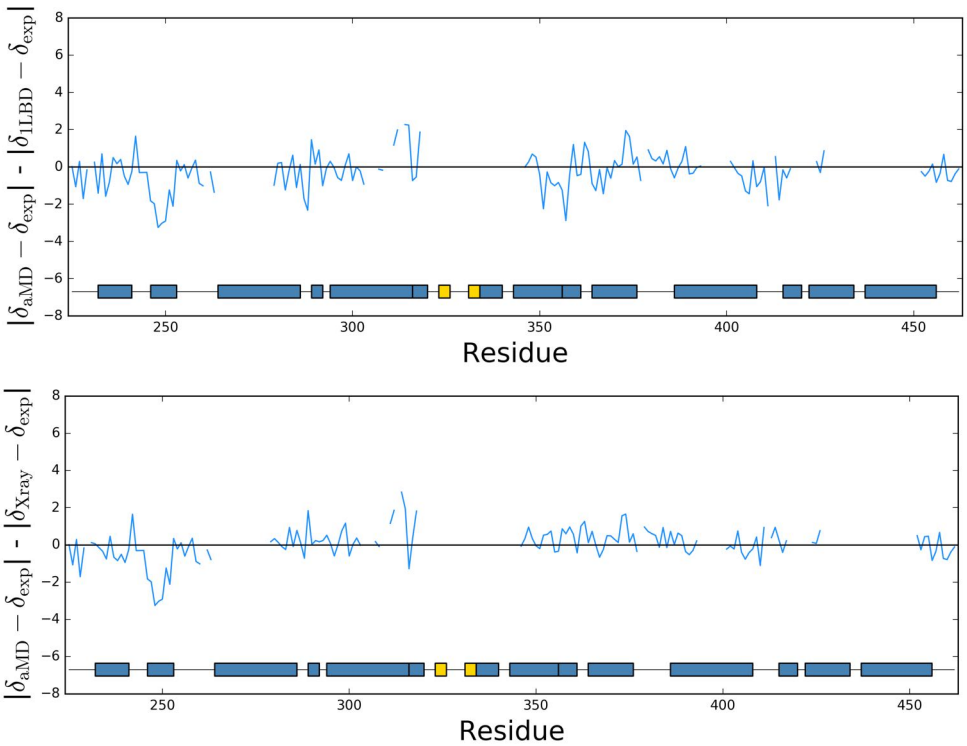


FIGURE 5.12 – Comparaison entre les déplacements chimiques RMN δ $^{13}C\alpha$ mesurés expérimentalement (δ_{exp}), ceux prédits à partir des trajectoires aMD (δ_{aMD}) et ceux prédits à partir de la structure cristallographique seule (δ_{1LBD}) (en haut) ou à partir de toutes les structures cristallographiques de la forme libre du LBD de RXRα disponibles (δ_{Xray}) (en bas). Une valeur positive de $|\delta_{aMD} - \delta_{exp}|$ - $|\delta_{1LBD/Xray} - \delta_{exp}|$ (exprimé en ppm) indique que δ_{1LBD} ou δ_{Xray} est en meilleur accord avec les données expérimentales δ_{exp}, tandis qu'une valeur négative signifie que δ_{aMD} est en meilleur accord.

De manière plus spécifique à la structure cristallographique de référence seule, les résultats montrent aussi que les déplacements chimiques δ_{aMD} $^{13}C_\alpha$ de la boucle H3/H4, la partie C-terminale de H7 et de la boucle H9/H10 sont plus proches des données expérimentales δ_{exp} $^{13}C_\alpha$ (Figure 5.12 A). L'utilisation

de plusieurs structures cristallographiques améliore l'accord avec les δ_{exp} $^{13}\text{C}_\alpha$ au niveau de ces régions (Figure 5.12 B).

Les déplacements chimiques secondaires $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}_\alpha$ diffèrent très nettement de ceux observés au sein de la structure cristallographique de référence au niveau de la région entre les hélices H1 et H3. Sachant que cette région n'est pas présente dans les autres structures cristallographiques (PDB id : 3NSP, 3NSQ) de la forme libre du LBD de RXRα, les déplacements chimiques moyens δ_{1LBD} $^{13}\text{C}_\alpha$ et δ_{Xray} $^{13}\text{C}_\alpha$ sont donc identiques. Au sein de la structure cristallographique de référence (PDB id : 1LBD), cette région révèle la présence d'une hélice α, nommée H2, ce qui est cohérent avec des valeurs $\Delta\delta_{\text{1LBD}}$ $^{13}\text{C}_\alpha$ prédites qui sont positives (> 2 ppm), montrant la présence d'une hélice dans cette région (Figure 5.13).

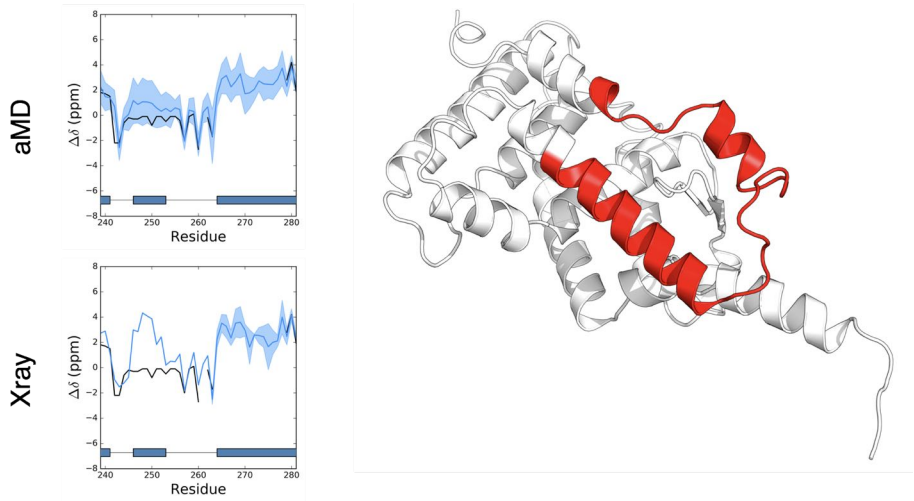


FIGURE 5.13 – Comparaison entre les déplacements chimiques RMN secondaires $\Delta\delta^{13}\text{C}_\alpha$ de la boucle entre les hélices H1 et H3, comprenant l'hélice H2, obtenus expérimentalement (en noir) et ceux prédits à partir des trajectoires aMD et ceux prédits à partir de toutes les structures cristallographiques de la forme libre du LBD de RXRα disponibles (PDB id : 1LBD, 3NSP et 3NSQ).

Cependant, les données expérimentales $\Delta\delta_{\text{exp}}$ $^{13}\text{C}_\alpha$ suggèrent le contraire avec des valeurs proches de zéro ppm, c'est-à-dire que cette région n'est pas structurée et ne présente pas d'hélice α. Au contraire, l'accord est bien meilleur avec les déplacements chimiques secondaires $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}_\alpha$ qui montrent que l'hélice H2 n'est pas présente au sein des trajectoires aMD de la forme libre du LBD de RXRα.

Des divergences sont également observées au niveau de la partie C-terminale de l'hélice H5, résidus 307 à 318, qui est impliquée dans l'interaction avec 9cRA via R316 en formant des liaisons hydrogène avec la partie carboxyle du ligand au sein de la forme liée. D'autres résidus (W305, L309) sont impliqués

dans l'interaction avec le ligand via des contacts hydrophobes. On constate que les valeurs des $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}_\alpha$ sont nettement plus proches de zéro ppm (< 2 ppm), c'est-à-dire que cette région est peu structurée dans les trajectoires aMD, contrairement aux données expérimentales $\Delta\delta_{\text{exp}}$ $^{13}\text{C}_\alpha$ qui montrent une région bien plus structurée en hélice (> 2 ppm) (Figure 5.14).

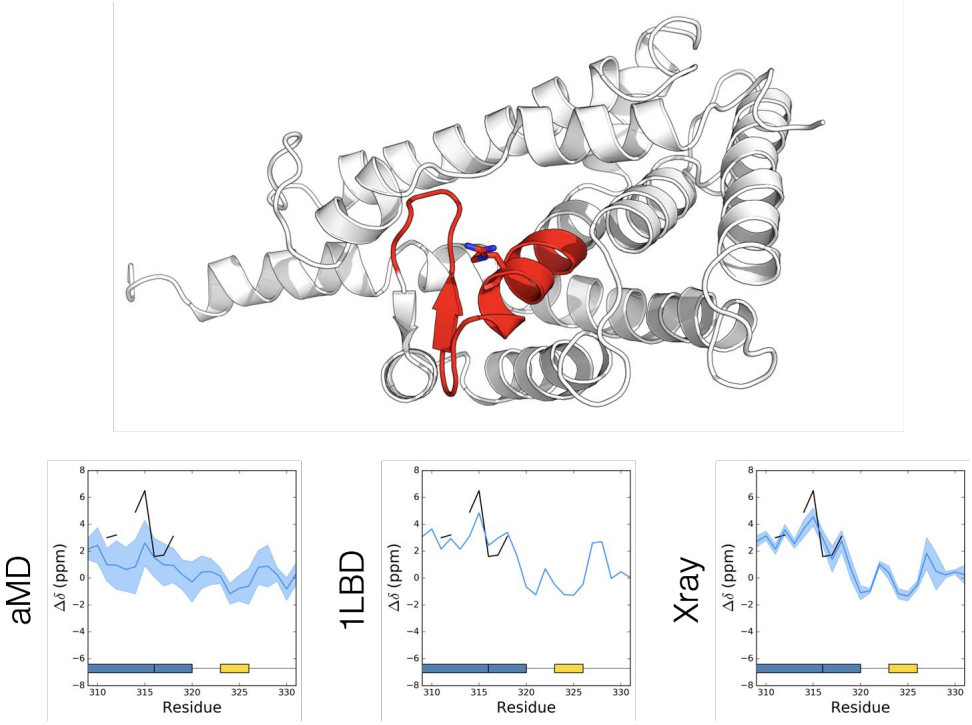


FIGURE 5.14 – Comparaison entre les déplacements chimiques RMN secondaires $\Delta\delta^{13}\text{C}_\alpha$ de l'hélice H5, comprenant le résidu Arg316 (en représentation stick), obtenus expérimentalement (en noir) et ceux prédits à partir des trajectoires aMD, ceux prédits à partir de la structure cristallographique seule (PBD id : 1LBD) et ceux prédits à partir de toutes les structures cristallographiques de la forme libre du LBD de RXRα disponibles (PDB id : 1LBD, 3NSP et 3NSQ).

L'accord avec les déplacements expérimentaux est meilleur dans cette région pour l'ensemble des structures cristallographiques de la forme libre du LBD de RXRα (PDB id : 1LBD, 3NSP, 3NSQ) ou pour la structure cristallographique de référence (PDB id : 1LBD) prise séparément, quoiqu'une différence de 2 ppm reste observée pour le résidu H315. On peut également remarquer que les acides aminés attribués en C-ter de H5 (S313 à I318) sont entourés de régions qui ne sont pas attribuées, et donc qui ont un comportement dynamique. L'origine exacte de la différence observée dans cette région reste à déterminer. De plus, en ne calculant que les déplacements chimiques RMN sur les 50 dernières nanosecondes de chaque

trajectoire, qui correspondent aux conformations de plus basses énergies (voir « Descriptions des conformations »), aucune différence significative n'est observée (Figure 7.1).

En résumé, les résultats montrent que les déplacements chimiques δ_{aMD} prédits à partir des conformations générées par simulations aMD sont globalement en accord avec les données expérimentales δ_{exp} et que suivant la région, l'accord peut être soit bien meilleur que les structures cristallographiques, comme au niveau de la boucle H1/H3, ou soit moins bon comme au niveau de la partie C-terminale de H5. Concernant l'hélice H2 par exemple, il semblerait que sa présence au sein de la structure cristallographique de référence soit liée à une stabilisation lors de la cristallisation, expliquant de fait pourquoi cette région, qui n'est d'autre qu'une boucle d'après les données expérimentales $\Delta\delta_{\text{exp}}$ $^{13}\text{C}_\alpha$, est absente dans les autres structures cristallographiques. Ainsi, l'utilisation de la dynamique moléculaire accélérée a permis, en partant d'une conformation en hélice, de retrouver l'état natif de cette région en accélérant l'échantillonnage conformationnel.

5.3.3.2 Forme liée à 9cRA du LBD de RXRα

Dans le cas de la forme liée à 9cRA, les déplacements chimiques secondaires $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}_\alpha$ prédits à partir des trajectoires aMD, sont en moins bon accord avec les déplacements chimiques expérimentaux que ceux calculés à partir des structures cristallographiques, quoique les différences soient généralement faibles (figure5.15).

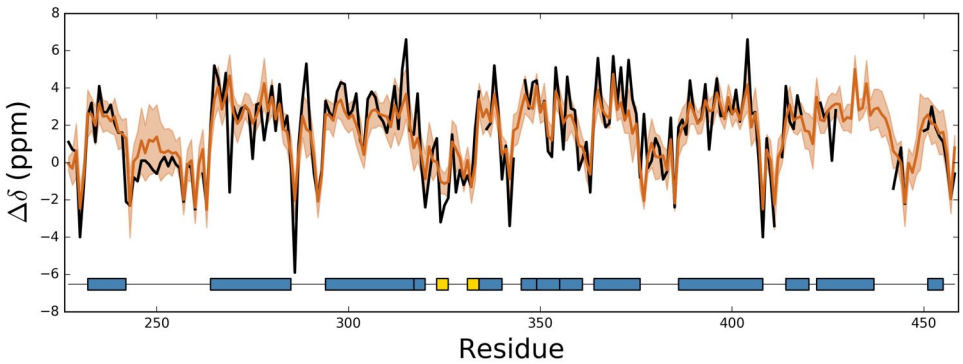


FIGURE 5.15 – Comparaison entre les déplacements chimiques secondaires $\Delta\delta_{\text{exp}}$ $^{13}\text{C}\alpha$ obtenus expérimentalement (en noir) et ceux prédits $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}\alpha$ à partir des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα (en orange).

Le RMSD entre $\Delta\delta_{\text{aMD}}$ $^{13}\text{C}_\alpha$ et $\Delta\delta_{\text{exp}}$ $^{13}\text{C}_\alpha$ est de 1.222 ppm, alors qu'il est de 1.108 ppm et 1.094 ppm pour la structure cristallographique de référence seule (PDB id : 1FM9) et l'ensemble des structures cristallographiques disponibles de la forme liée à 9cRA (PDB id : 1FBY, 1FM6, 1FM9, 1K74, 1XLS, 2ACL, 3DZY, 3E00, 3FAL, 3FC6, 3OAP, 3UVV et 4NQAQ), respectivement (Table 5.5). Ceci est aussi le cas pour les $\Delta\delta$ $^{13}\text{C}_\beta$ et $\Delta\delta$ ^{15}N, où le RMSD est plus élevé entre les $\Delta\delta_{\text{aMD}}$ et $\Delta\delta_{\text{exp}}$ qu'avec les $\Delta\delta_{\text{1FM9}}$ et $\Delta\delta_{\text{Xray}}$. Concernant les $\Delta\delta$ ^{13}C et $\Delta\delta$ ^{1}HN, la structure cristallographique est plus proche des données expérimentales que les trajectoires aMD et l'ensemble des structures cristallographiques (Table 5.5).

	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	^{13}C	^{15}N	^{1}HN
aMD	1.22	1.45	1.38	3.09	0.55
Xray (1FM9)	1.11	1.25	1.32	2.93	0.50
Xray (All)	1.09	1.10	1.49	2.94	0.61

TABLE 5.5 – Détails par élément chimique, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{13}C, ^{15}N et ^{1}HN, du RMSD obtenu entre les déplacements chimiques RMN secondaires obtenus expérimentalement et ceux prédits à partir des trajectoires aMD, ainsi que ceux prédits à partir de la structure cristallographique seule (PDB id : 1FM9) et à partir de toutes les structures cristallographiques de la forme liée à 9cRA du LBD de RXRα disponibles (PDB id : 1FBY, 1FM6, 1FM9, 1K74, 1XLS, 2ACL, 3DZY, 3E00, 3FAL, 3FC6, 3OAP, 3UVV et 4NQAQ).

Cependant, la comparaison entre δ_{1FM9} $^{13}\text{C}_\alpha$ et δ_{aMD} $^{13}\text{C}_\alpha$ d'un coté et δ_{Xray} $^{13}\text{C}_\alpha$ et δ_{aMD} $^{13}\text{C}_\alpha$ de l'autre avec δ_{exp} $^{13}\text{C}_\alpha$ ne révèle pas de régions particulières de désaccord, comme c'est le cas avec la forme libre où il y avait une différence significative au niveau de la région H1/H3 ou de la partie C-terminale de H5. Dans le cas de la forme liée à 9cRA, il semblerait que le meilleur accord observé entre δ_{1FM9} $^{13}\text{C}_\alpha$ et δ_{exp} $^{13}\text{C}_\alpha$ soit global et non limité à quelques régions (Figure 5.16). De même pour l'ensemble des structures cristallographiques, excepté que les différences δ_{aMD} $^{13}\text{C}_\alpha$, δ_{Xray} $^{13}\text{C}_\alpha$ et δ_{exp} $^{13}\text{C}_\alpha$ sont plus importantes que précédemment (Figure 5.16).

En résumé, contrairement à la forme libre du LBD de RXRα, les déplacements chimiques calculés à partir des trajectoires aMD sont globalement en moins bon accord avec les données expérimentales que ceux calculés à partir d'une, ou plusieurs, structures cristallographiques. Cependant, les divergences sont souvent de l'ordre du ppm, ce qui est proche de l'erreur standard de la méthode SHIFTX+ [Han et al., 2011].

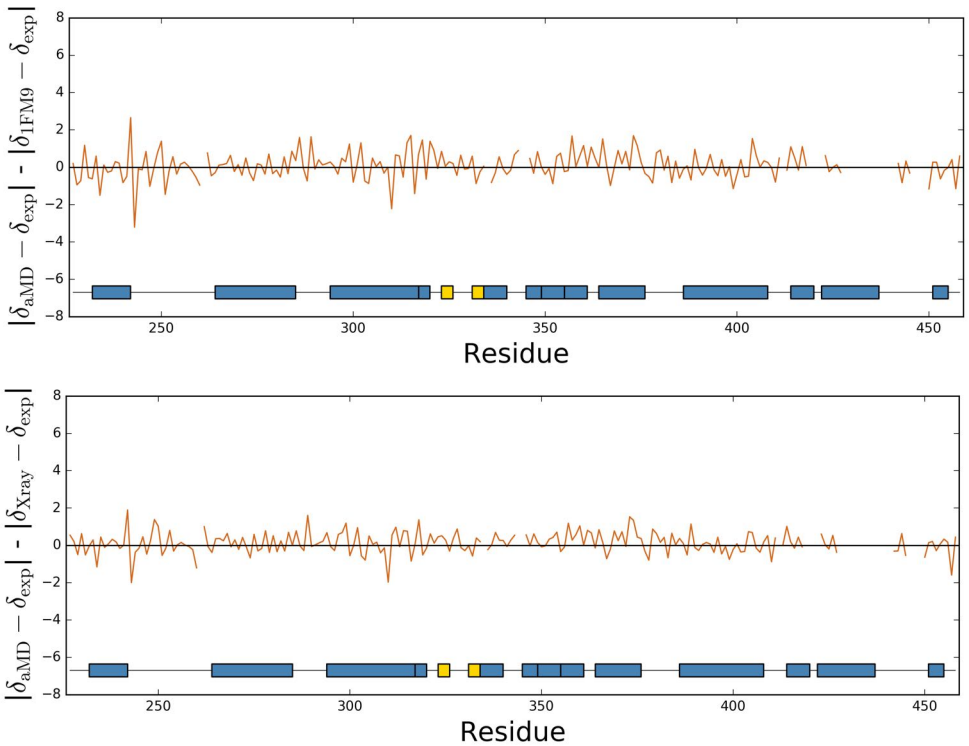


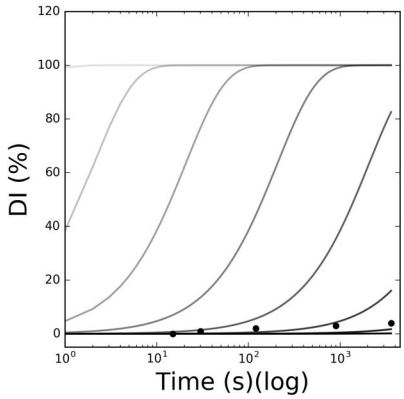
FIGURE 5.16 – Comparaison entre les déplacements chimiques RMN δ ^{13}Cα mesurés expérimentalement (δ_{exp}), ceux prédits à partir des trajectoires aMD (δ_{aMD}) et ceux prédits à partir de la structure cristallographique seule (δ_{1FM9}) (en haut) ou à partir de toutes les structures cristallographiques de la forme liée du LBD de RXRα disponibles (δ_{Xray}) (en bas). Une valeur positive de $|\delta_{aMD} - \delta_{exp}|$ - $|\delta_{1FM9/Xray} - \delta_{exp}|$ (exprimé en ppm) indique que δ_{1FM9} ou δ_{Xray} est en meilleur accord avec les données expérimentales δ_{exp}, tandis qu'une valeur négative signifie que δ_{aMD} est en meilleur accord.

5.3.4 Analyse de l'Échange Hydrogène/Deutérium

5.3.4.1 Détermination du facteur de protection maximal expérimental

Pour déterminer le facteur de protection expérimental maximal propre au LBD de RXRα, le peptide correspondant aux résidus 391 à 400 (hélice H9) a été choisi. Ce peptide présente la plus faible incorporation en deutérium au cours du temps et en conséquence les facteurs de protection les plus élevés (voir Figure 5.18, sa séquence n'étant pas particulièrement résistante à l'incorporation, l'absence d'échange est donc lié au PF). La présence de 9cRA au sein du LBD de RXRα n'a pas d'effet sur la vitesse d'incorporation en deutérium au sein de ce peptide. En effet, sa vitesse d'échange H/D est identique dans la forme libre

et la forme liée à 9cRA du LBD de RXRα. Pour déterminer le facteur de protection à partir des données expérimentales, la procédure décrite dans la partie matériels et méthodes [Hamuro et al., 2006] a été suivie. Comme les données expérimentales ne sont pas résolues au niveau du résidu, mais du peptide, un facteur de protection identique est attribué pour tous les résidus du peptide. Des valeurs successives allant de 1 à 10^7 sont testées. Les résultats montrent un ralentissement de l'incorporation en deutérium prédit au cours du temps au fur et à mesure de l'augmentation des facteurs de protection (Figure 5.17). En partant d'un facteur de protection égal à 1 avec un RMSD de 98.01 % et une corrélation nulle entre les données prédites et expérimentales, le RMSD minimum est atteint à 10^6 avec 1.81 %, quant à la corrélation celle-ci atteint un maximum à 10^3 avec 0.95 et diminue ensuite progressivement jusqu'à 0.82.



Facteur de protection	RMSD	Corrélation
1	98.01	N/A
10^1	97.99	0.70
10^2	85.53	0.88
10^3	63.82	0.95
10^4	38.05	0.88
10^5	5.46	0.83
10^6	1.81	0.82
10^7	2.38	0.82

FIGURE 5.17 – Prédiction de l'évolution de l'incorporation en deutérium au cours du temps au sein du peptide 491-400 du LBD de RXRα de la forme libre et la forme liée à 9cRA, en fonction du facteur de protection allant de 1 à 10^7, ainsi que le RMSD et la corrélation associés à chacune des valeurs testées.

De ces observations, on peut conclure que le facteur de protection maximal est de l'ordre de 10^6, comme c'est le cas par exemple pour le LBD de PPARγ [Hamuro et al., 2006]. Cette valeur sera ainsi utilisée pour normaliser les facteurs de protection obtenus à partir des simulations de dynamiques moléculaires. Par contre, la corrélation globale entre les données d'échange HD prédites et expérimentales ne parait pas un critère suffisamment discriminant pour calibrer le facteur de protection maximal.

5.3.4.2 Détermination des constantes de vitesses d'échanges intrinsèques

Sachant que les constantes de vitesses d'échanges intrinsèques k_{int} ne dépendent en grande partie que de la séquence, ceux-ci seront alors identiques entre la forme libre et la forme liée à 9cRA du LBD de RXRα.

Les séquences sont celles issues des structures cristallographiques de référence (PDB id : 1LBD et 1FM9) et la température et le pD (potentiel Deutérium) égalent à 298.15 K et 7.8 (pH_{lu} + 0.4), respectivement, identiques aux conditions expérimentales. Les résultats montrent que les vitesses d'échange intrinsèques au sein du LBD varient d'un facteur 1 à 40.

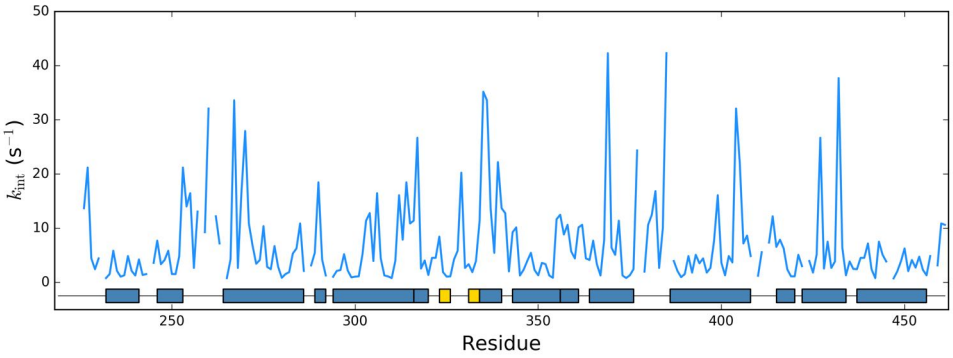


FIGURE 5.18 – Constantes de vitesses d'échanges d'intrinsèques k_{int} (exprimé en sec^{-1}) pour chaque proton amide HN du LBD de RXRα obtenu à l'aide du modèle empirique proposé par l'équipe de Walter S. Englander [Bai Y - 1993].

5.3.4.3 Prédiction du taux d'incorporation en deutérium

5.3.4.3.1 Forme libre du LBD de RXRα

L'analyse des liaisons hydrogène entre les protons amides HN et le solvant, en excluant toutes interactions intramoléculaires, permet d'identifier les régions où les protons sont les plus enclins à échanger. On observe dans le cas du modèle simple que les protons amides HN qui interagissent fréquemment avec l'eau sont localisés principalement dans la partie N-terminale, la boucle entre H1 et H3, comprenant l'hélice H2, la partie N-terminale de H3, le coude formé par H4 et H5, la partie C-terminale de H5, le feuillet β et l'hélice H6, la partie N-terminale de H7, les boucles H7/H8, H8/H9 et H9/H10 ainsi que les hélices H11 et H12 (Figure 5.19). Dans le cas de la trajectoire MD classique, utilisée pour analyser les interactions entre les protons amides HN et le solvant au sein de structures restant proches de la structure cristallographique, on remarque que le nombre de régions où les protons amides HN sont dans un état compétent est fortement restreint. Seuls les protons amides HN localisés dans les boucles interagissent avec le solvant. Cependant, lorsque ceux-ci sont dans un état compétent, ils ont tendance à y demeurer plus longtemps, contrairement aux trajectoires aMD. Ceci est lié à l'échantillonnage plus réduit au cours de la trajectoire MD classique.

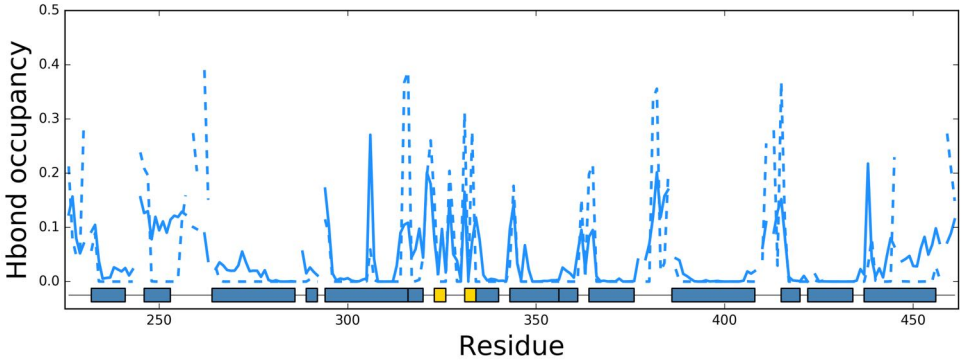


FIGURE 5.19 – Taux de compétence des protons amides HN selon le modèle utilisé simple. En bleu par un trait plein, sont représentées les données obtenues à partir des trajectoires aMD et par un trait en pointillés, celles obtenues à partir de la trajectoire cMD pour la forme libre du LBD de RXRα.

En ce qui concerne l'analyse des liaisons hydrogène en utilisant le modèle concerté, nous retrouvons les mêmes régions susceptibles d'échanger qu'avec le modèle simple. Cependant, les protons amides HN appartenant à ces régions passent nettement moins de temps dans un état compétent (Figure 5.20). Ceci, est dû au fait qu'un proton amide HN n'est considéré comme compétent uniquement si deux molécules d'eau ou plus sont à proximité de ce dernier, malgré le fait que les critères géométriques pour identifier une liaison hydrogène soient beaucoup plus souples que dans le cas du modèle simple. Cependant, vu la longueur des trajectoires analysées, le nombre de liaisons H observées pour le modèle concerté reste suffisant pour permettre le calcul de facteurs de protection (Figure 5.20). De même pour la trajectoire MD classique, seules les boucles sont susceptibles d'échanger, mais dans une moindre mesure qu'avec le modèle simple et seul deux résidus restent dans un état compétent pendant plus de 10% du temps de trajectoire.

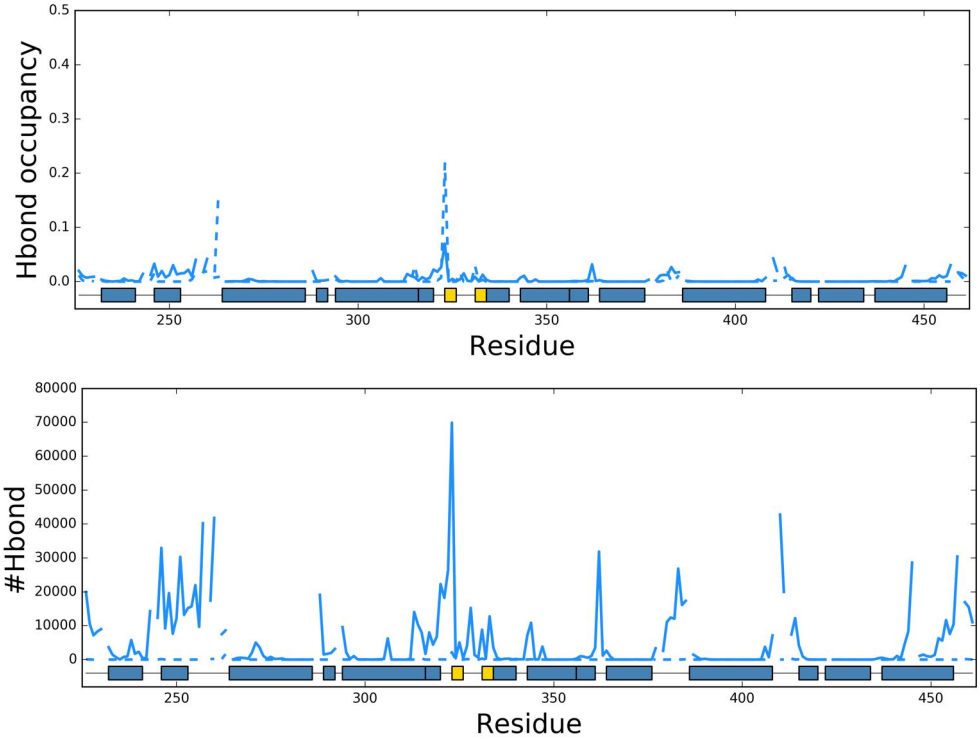


FIGURE 5.20 – Taux de compétence des protons amides HN (en haut) et nombre de liaisons hydrogène compétentes formées entre les protons amides HN et l'eau (en bas) selon le modèle utilisé concerté. En bleu par un trait plein, sont représentées les données obtenues à partir des trajectoires aMD et par un trait en pointillés, celles obtenues à partir de la trajectoire cMD pour la forme libre du LBD de RXRα.

En résumé, la seule différence entre le modèle simple et le modèle concerté se situe au niveau du temps passé par chacun des protons amides HN dans un état compétent, que ce soit pour les trajectoires aMD ou la trajectoire cMD. Malgré les critères géométriques à priori plus restrictifs du modèle simple, concernant l'identification des liaisons hydrogène, les protons amides HN sont considérés comme plus souvent dans un état compétent que lorsque le modèle concerté est utilisé. Dans le cas de la trajectoire MD classique, c'est-à-dire proche de la structure cristallographique, seuls les protons amides HN situés dans les régions flexibles, principalement les boucles, sont dans un état compétent.

Calculés à partir des données précédentes, les facteurs de protection PF sont de fait liés directement au temps passé par chacun des protons amides HN dans un état compétent. Plus la proportion de protons appartenant à un état où ils sont susceptibles d'être échangés est importante, moins le facteur de protection PF sera élevé. Et inversement, plus un proton amide HN est isolé du solvant, plus son facteur de protection

PF sera élevé. Ainsi on observe de manière générale que les éléments de structures secondaires possèdent des facteurs de protection PF plus élevés, compris entre 10^3 et 10^6, excepté pour l'hélice H1, la partie N-terminale de H3, le feuillet β et H11/H12. Quant à la trajectoire MD classique, tous les éléments structuraux secondaires sont complètement protégés avec des facteurs de protection égaux à 10^6, excepté pour les boucles (PF inférieurs ou proches de 10^3). L'utilisation du modèle concerté augmente les facteurs de protection PF d'un facteur 10 à 100 en moyenne pour chaque proton amide HN. Au final, quelque soit le modèle utilisé, les mêmes régions peu protégées sont retrouvées.

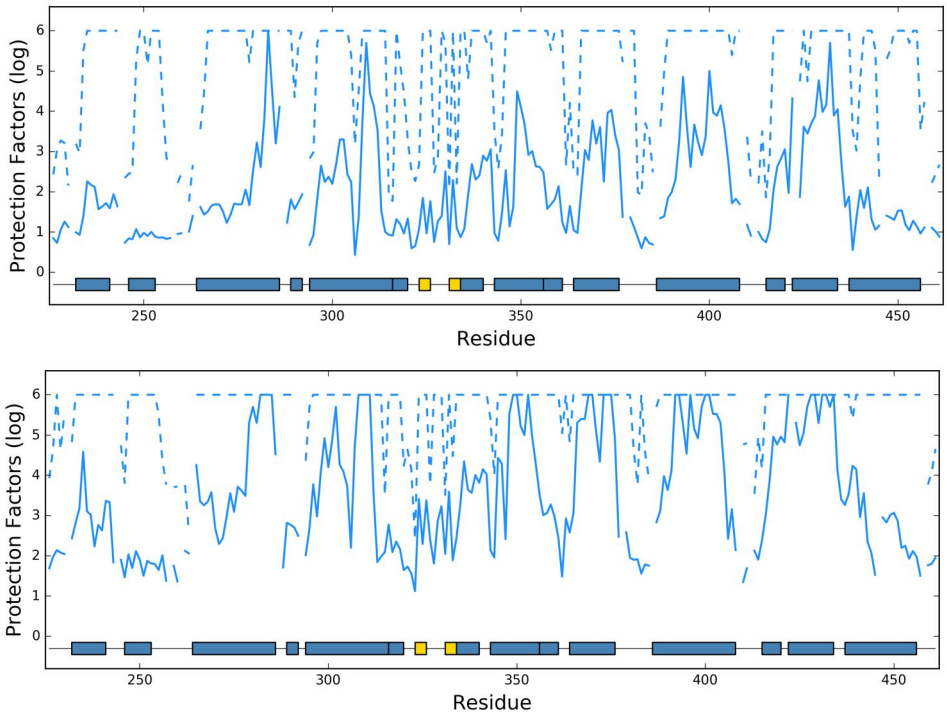


FIGURE 5.21 – Facteurs de protection PF (exprimé sur une échelle logarithmique) pour chaque proton amide HN selon le modèle simple (en haut) et selon le modèle concerté (en bas). En bleu par un trait plein, sont représentées les données obtenues à partir des trajectoires aMD et par un trait en pointillés, celles obtenues à partir de la trajectoire cMD pour la forme libre du LBD de RXRα.

Ainsi, à partir des constantes de vitesses d'échanges intrinsèques k_{int} et des facteurs de protection PF obtenus à partir des trajectoires aMD et cMD, il est possible de prédire l'incorporation en deutérium au sein de la forme libre du LBD de RXRα au cours du temps. Ces résultats pourront alors être comparés directement aux données expérimentales HDX. L'ensemble des résultats concernant l'incorporation en

deutérium au cours du temps pour chaque peptide sont regroupés figures 5.24 et 5.25, pour les modèles simple et concerté, respectivement. De même, dans la table 5.6 sont présentés le RMSD et la corrélation obtenus pour chaque peptide.

Au total, l'incorporation en deutérium au cours du temps a été prédite au sein de 37 peptides et comparée aux données expérimentales, et ceci pour les deux modèles, simple et concerté (Figures 5.24 et 5.25 et Table 5.6). Une première observation importante est que les données prédites à partir de la trajectoire classique et utilisant le modèle simple sous-estiment systématiquement l'incorporation du deutérium au cours du temps dans la forme libre, par rapport aux données expérimentales (Figure 5.22). Les seules exceptions concernent l'hélice H9 (en particulier en son centre, résidus 391 à 400), qui est expérimentalement la plus protégée (Figure 5.22) et la partie N-terminale de H8 (peptides 363-367 et 363-370). Il est à noter que ces deux régions sont en contact dans la structure cristallographique (Figure 5.22).

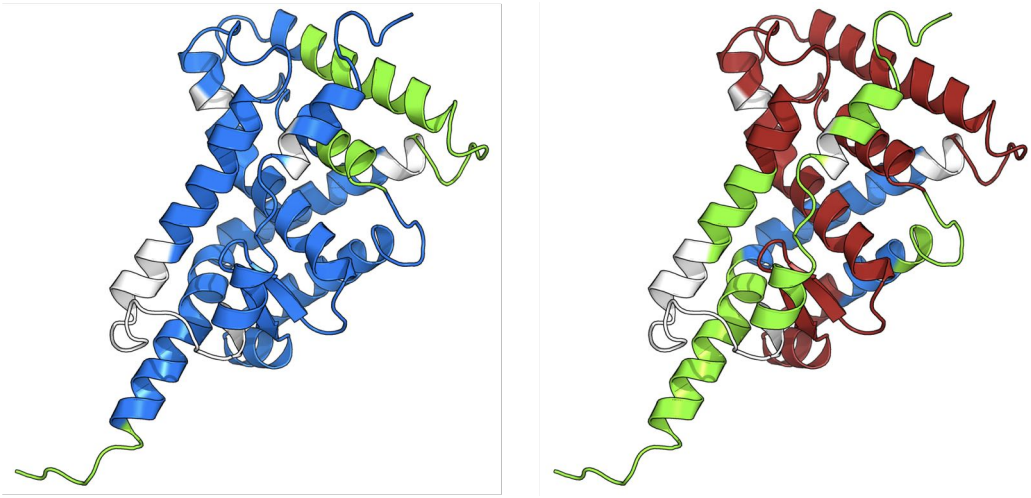


FIGURE 5.22 – Résumé des résultats obtenus à partir de la trajectoire cMD (à gauche) et des trajectoires aMD (à droite) de la forme libre du LBD de RXRα en utilisant le modèle simple. En vert, les régions dont le taux d'incorporation en deutérium prédit est en accord avec les données expérimentales. En rouge et en bleu, les peptides où le taux d'incorporation en deutérium prédit est soit supérieur ou bien inférieur aux données expérimentales, respectivement. En blanc, les résidus pour lesquels aucune donnée expérimentale n'est disponible.

Lorsque les trajectoires aMD sont utilisées, l'accord prédiction-expérience est nettement meilleur, cependant des divergences restent observées entre les valeurs expérimentales et calculées. En outre, les régions où les divergences sont observées ne sont généralement pas les mêmes, suivant que l'on utilise le

modèle concerté ou le modèle simple pour estimer les facteurs de protection. Lorsque le modèle simple est utilisé, un bon accord est observé sur les peptides localisés en N et C-terminal de la forme libre du LBD de RXRα (Figure 5.24), ce qui correspond aux régions où l'incorporation en deutérium est très rapide, avec un taux d'échange H/D minimum de 60 % à partir de 15 secondes. Par exemple, les RMSD pour les peptides 219-237, 219-233, 241-248, 241-251, 271-278 et 271-279, correspondant aux hélices H1, H3 et au milieu de H3, sont de 15 %, 21 %, 15 %, 13 %, 17 % et 27 %, respectivement. Quant aux peptides situés en C-terminal, soit les peptides 433-438, 439-449, 439-450, 439-451, 450-454 et 455-462, correspondant à la partie C-terminale de H10 et aux hélices H11 et H12, les RMSD sont encore moindre, soit 17 %, 12 %, 18 %, 8 %, 8 % et 16 % (Table 5.6). Cependant, dans les autres régions du LBD de RXRα, l'accord entre l'incorporation prédite et observée est moins bon. De manière générale, le modèle simple prédit une incorporation supérieure à ce qui est observé expérimentalement. Parmi les peptides dont la vitesse d'incorporation en deutérium prédite est supérieure à celle obtenue expérimentalement, la différence la plus marquée est observée pour la région 391-400, région centrale de l'hélice H9, qui présente expérimentalement l'échange le plus lent (seuls 4 % des protons amides HN ont échangé au bout d'une heure). Pour ces peptides, la RMS entre prédit et observé est de 42 % et 47 %, respectivement. De manière surprenante, pour certains peptides, l'incorporation expérimentalement observée est plus importante que ce qui est prédit par le modèle concerté, ceci concerne les peptides 347-353 (hélice H7), et les peptides 419-432, 419-438, 423-438 et 426-438 (hélice H10). La différence la plus importante concerne l'hélice H10 avec les peptides 419-432, 419-438, 423-438 et 426-438 qui présentent des RMSD de 58 %, 53 %, 49 % et 66 %, respectivement. Alors que le taux d'incorporation expérimental est de minimum 80 % à partir de 15 secondes, celui prédit est seulement de 80 % au bout d'une heure. Ainsi, il semblerait que l'hélice H10 soit capable de perdre sa structure secondaire et d'échanger, alors qu'elle reste compacte au cours des trajectoires aMD.

Ainsi que discuté ci-dessus, l'utilisation du modèle concerté mène à des facteurs de protection PF calculés plus importants, et donc à une réduction de l'incorporation en deutérium, par rapport au modèle simple. Si les trajectoires classiques sont analysées à l'aide du modèle concerté, l'incorporation en deutérium est extrêmement faible pour l'ensemble des peptides, et en désaccord systématique avec les données expérimentales. L'analyse des trajectoires aMD avec le modèle concerté révèle une réduction systématique de l'échange par rapport au modèle simple. De ce fait, l'ensemble des régions pour lesquelles le modèle simple prédisait une incorporation trop importante, en particulier, les peptides correspondants à la boucle H3/H4, aux hélices H4 et H5, au feuillet β et à l'hélice H6, ainsi qu'à la partie C-terminale de H7, la

boucle H7/H8, l'hélice H8, la boucle H8/H9, et l'hélice H9 sont maintenant prédits de manière correcte, des RMSD entre incorporation prédite et calculée oscillant entre 3 % et 16 % au maximum (Figure 5.23 et Table 5.6). Dans cette région, seul le peptide 401-415, correspondant à la boucle H9/H10 garde une vitesse d'échange H/D prédite supérieure aux données expérimentales. Par contre, pour les régions qui échangent fortement, en N et C-terminal, le modèle simple était plus performant. L'échange pour ces régions (peptides 271-278 et 271-279, correspondant à une partie de l'hélice H3, et peptides 433-438, 439-449, 439-450 et 439-451 correspondant à la partie C-terminale de H10 et aux hélices H11 et H12) est réduit dans le modèle concerté, et le RMS entre prédit et observé plus important (Figure 5.23 et Table 5.6).

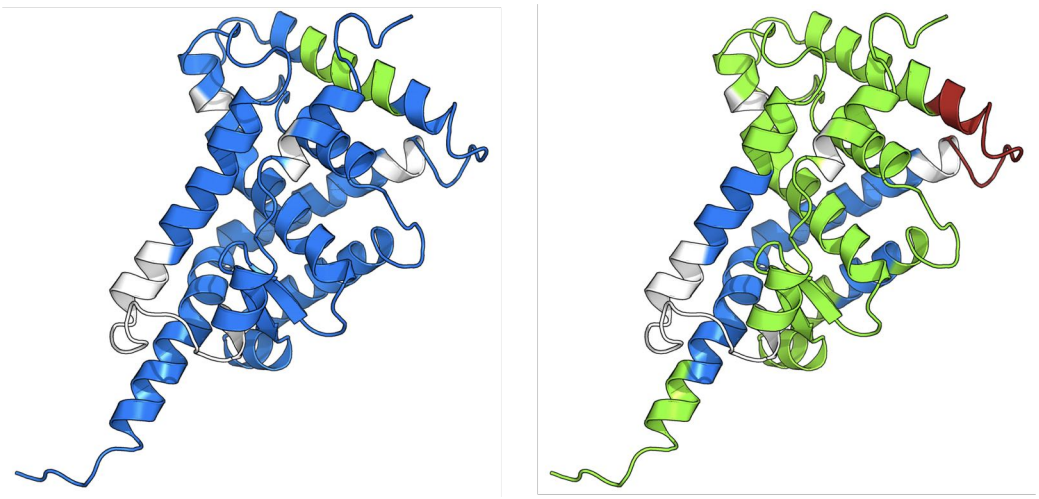


FIGURE 5.23 – Résumé des résultats obtenus à partir de la trajectoire cMD (à gauche) et des trajectoires aMD (à droite) de la forme libre du LBD de RXRα en utilisant le modèle concerté. En vert, les régions dont le taux d'incorporation en deutérium prédit est en accord avec les données expérimentales. En rouge et en bleu, les peptides où le taux d'incorporation en deutérium prédit est soit supérieur ou bien inférieur aux données expérimentales, respectivement. En blanc, les résidus pour lesquels aucune donnée expérimentale n'est disponible.

Finalement, pour les régions (hélices H7 et H10) où l'échange est très rapide et déjà sous-estimé par le modèle simple, le modèle concerté est également en désaccord, et prédit une incorporation significativement plus faible que ce qui est observé. De même que pour les déplacements chimiques RMN calculés auparavant, en ne calculant l'échange HD que sur les 50 dernières nanosecondes de chaque trajectoire, qui correspondent aux conformations de plus basses énergies (voir « Descriptions des conformations »), aucune différence significative n'est observée (Figure 7.2 et 7.3).

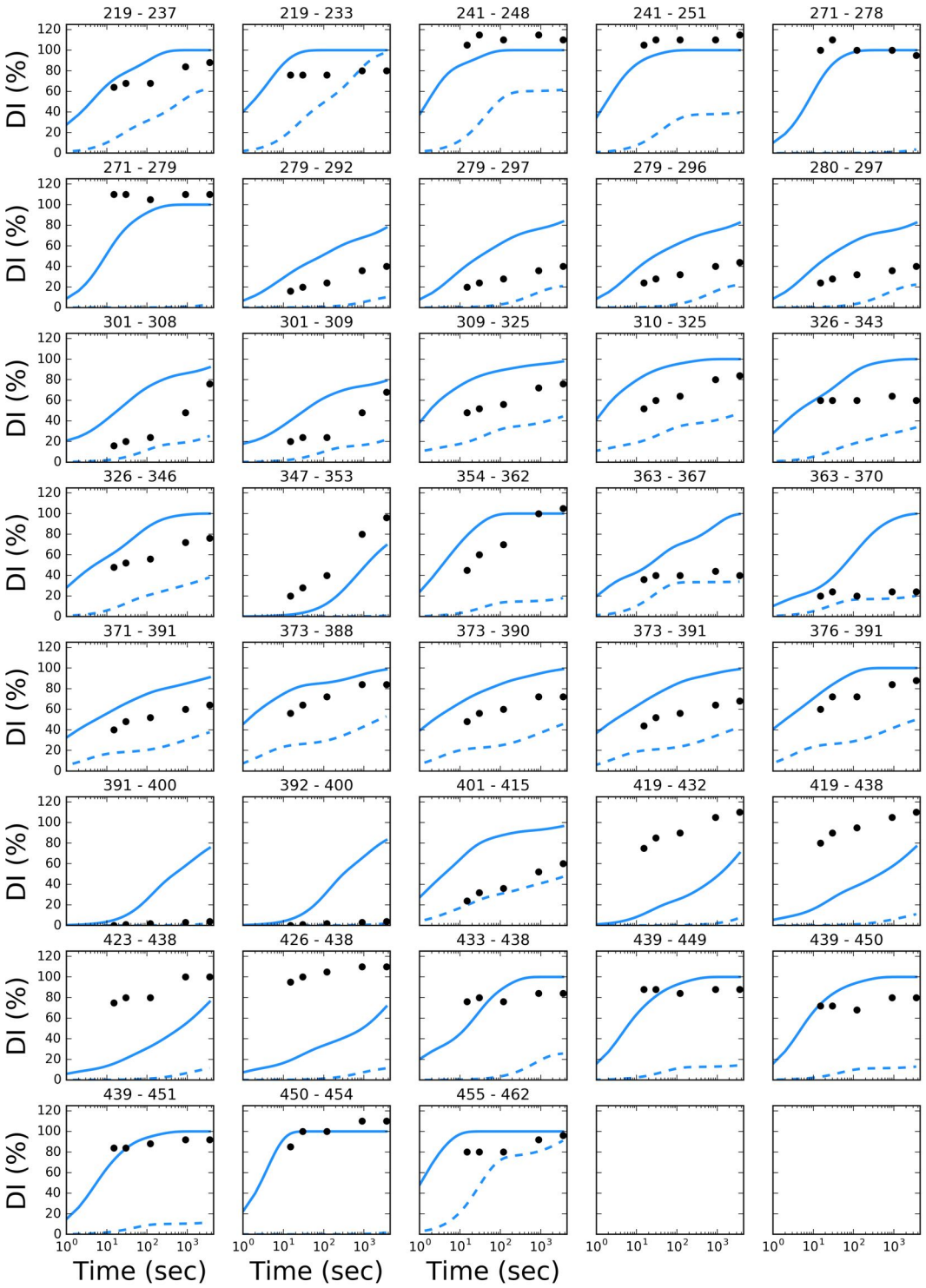


FIGURE 5.24 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme libre du LBD de RXRα (modèle simple). En bleu par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires aMD, en bleu avec des pointillés ceux prédits à partir d'une trajectoire cMD et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

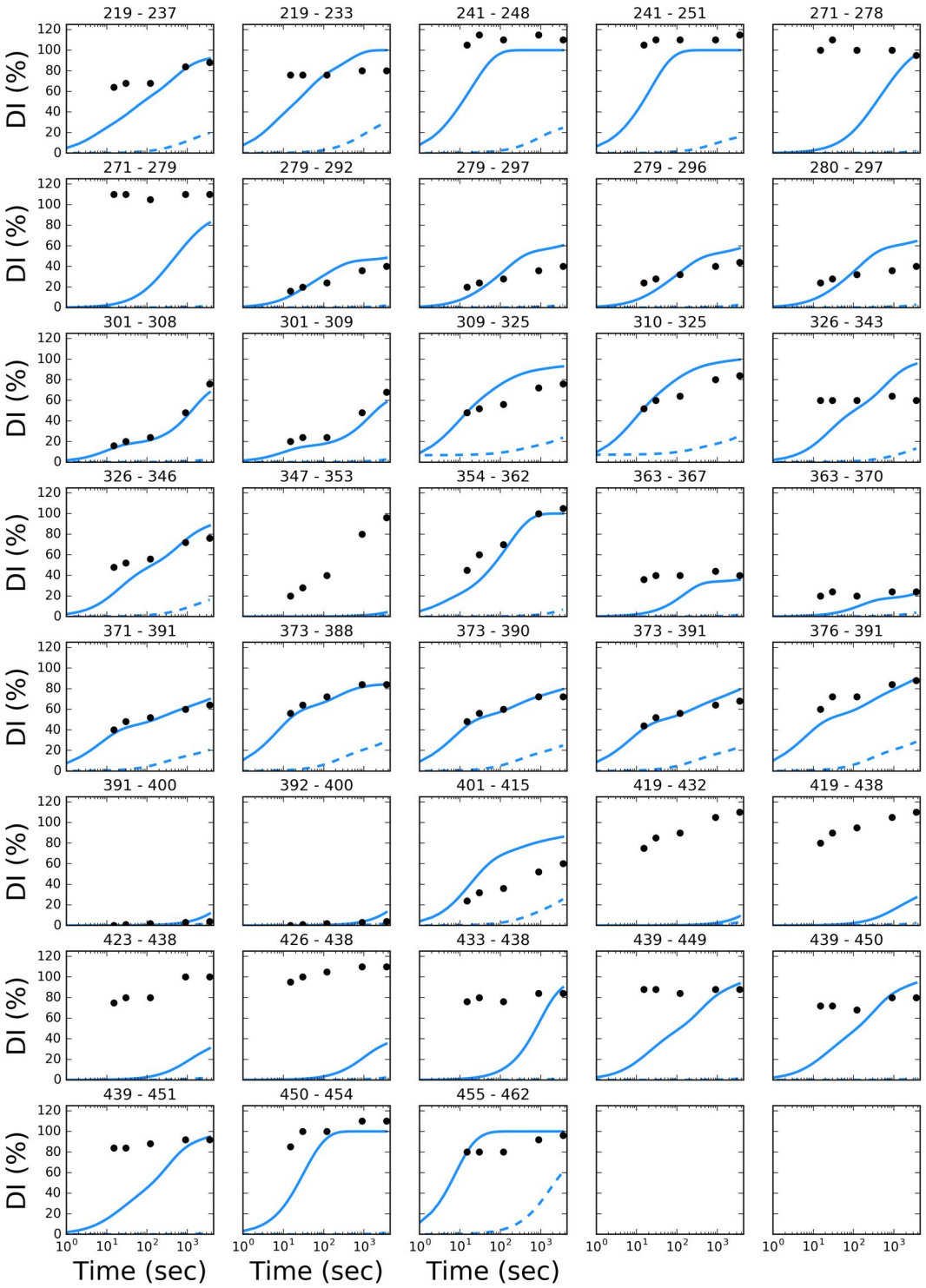


FIGURE 5.25 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme libre du LBD de RXRα (modèle concerté). En bleu par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires aMD, en bleu avec des pointillés ceux prédits à partir d'une trajectoire cMD et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

	Modèle simple				Modèle « concerté »			
	aMD		cMD		aMD		cMD	
	Correlation	RMSD	Correlation	RMSD	Correlation	RMSD	Correlation	RMSD
^{219}GSHMTSSANEDMPVERILE^{237}	0.88	15	0.97	39	0.96	22	0.96	68
^{219}GSHMTSSANEDMPVE^{233}	0.51	21	0.93	34	0.87	30	0.93	68
^{241}AVEPKTET^{248}	0.48	15	0.41	69	0.49	20	0.18	103
^{241}AVEPKTETYVE^{251}	0.79	13	0.78	82	0.78	31	0.80	104
^{271}AADKQLFT^{278}	-0.33	17	-0.66	100	-0.67	72	-0.66	101
^{271}AADKQLFTL^{279}	-0.23	27	0.29	108	0.21	80	0.29	109
^{279}LVEWAKRIPHFSEL^{292}	0.99	28	0.94	25	0.97	7	0.83	28
^{279}LVEWAKRIPHFSELPLDDQ^{297}	0.99	34	0.97	22	0.98	14	0.84	30
^{279}LVEWAKRIPHFSELPLDD^{296}	0.99	30	0.97	25	0.98	11	0.84	33
^{280}VEWAKRIPHFSELPLDDQ^{297}	0.99	33	0.93	24	0.98	17	0.83	32
^{301}LRAGWNEL^{308}	0.87	36	0.94	28	0.99	5	0.96	42
^{301}LRAGWNELL^{309}	0.86	26	0.93	38	0.99	9	0.94	40
^{309}LIASFSHRSIAVKDGIL^{325}	0.95	28	0.95	29	0.94	15	0.95	48
^{310}IASFSHRSIAVKDGIL^{325}	0.94	26	0.96	34	0.96	14	0.94	54
^{326}LATGLHVHRNSAHSAGVG^{343}	0.48	27	0.40	42	0.43	28	0.17	57
^{326}LATGLHVHRNSAHSAGVGAIF^{346}	0.92	24	0.97	39	0.98	16	0.93	56
^{347}DRVLTEL^{353}	0.99	27	0.86	60	0.86	59	0.86	60
^{354}VSKMRDMQM^{362}	0.85	24	0.94	67	0.98	15	0.79	77
^{363}DKTEL^{367}	0.66	37	0.77	14	0.71	24	0.20	40
^{363}DKTELGCL^{370}	0.52	51	0.46	9	0.53	14	0.51	21
^{371}RAIVLFNPDSKGLSNPAEVEA^{391}	0.98	23	0.93	28	0.98	4	0.94	44
^{373}IVLFNPDSKGLSNPAE^{388}	0.96	16	0.89	37	0.99	4	0.92	60
^{373}IVLFNPDSKGLSNPAEVE^{390}	0.98	23	0.90	32	0.97	4	0.92	51
^{373}IVLFNPDSKGLSNPAEVEA^{391}	0.98	28	0.93	29	0.98	6	0.94	47
^{376}FNPDSKGLSNPAEVEA^{391}	0.88	17	0.92	41	0.95	12	0.92	64
^{391}ALREKVYASL^{400}	0.97	42	0.82	2	0.83	3	0.82	2
^{392}LREKVYASL^{400}	0.97	47	0.82	2	0.84	4	0.82	2
^{401}EAYCKHKYPEQPGRF^{415}	0.94	45	0.99	9	0.94	25	0.95	32
^{419}LLRLPALRSIGLKC^{432}	0.95	60	0.81	92	0.81	91	0.80	93
^{419}LLRLPALRSIGLKCLEHLFF^{438}	0.96	53	0.89	93	0.89	87	0.79	96
^{423}PALRSIGLKCLEHLFF^{438}	0.93	49	0.92	83	0.92	77	0.77	87
^{426}RSIGLKCLEHLFF^{438}	0.89	66	0.84	1000	0.83	92	0.67	104
^{433}LEHLFF^{438}	0.68	17	0.86	70	0.84	58	0.71	80
^{439}FKLIGDTPIDT^{449}	-0.19	12	-0.23	78	0.10	41	0.29	87
^{439}FKLIGDTPIDTF^{450}	0.56	18	0.56	66	0.80	31	0.73	74
^{439}FKLIGDTPIDTFL^{451}	0.91	8	0.95	80	0.98	42	0.72	88
^{450}FLMEM^{454}	0.87	9	0.63	101	0.89	30	0.63	101
^{455}LEAPHQMT^{462}	0.41	16	0.78	28	0.54	12	0.96	68

TABLE 5.6 – Détails par peptide de la corrélation et du RMSD obtenus en comparant le taux d'incorporation en deutérium prédit à partir des trajectoires aMD et de la trajectoire cMD au taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes pour la forme libre du LBD de RXRα.

5.3.4.3.2 Forme liée à 9cRA du LBD de RXRα

L'analyse du taux de compétence des protons amides HN au cours des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα montre que les régions où ceux-ci sont le plus souvent compétents sont globalement les mêmes que celles retrouvées au cours des trajectoires aMD de la forme libre.

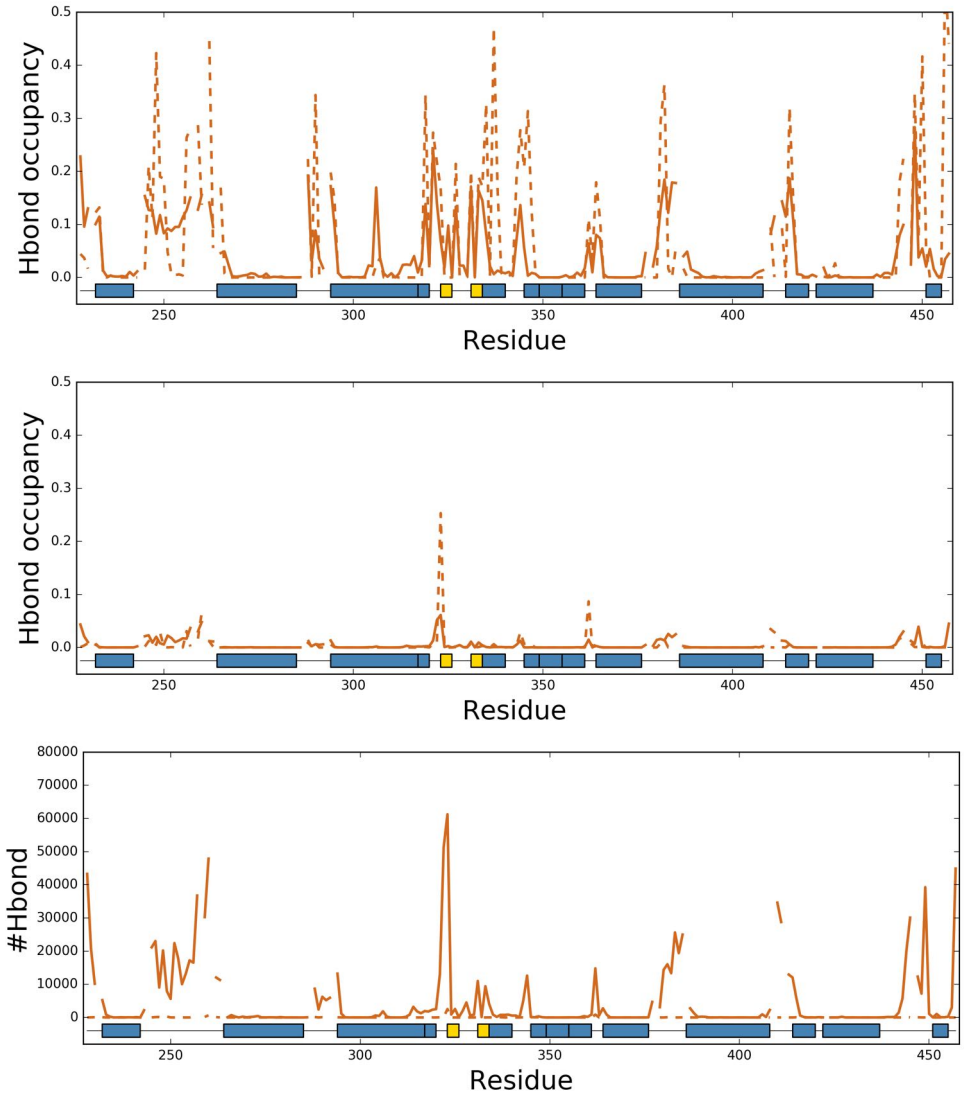


FIGURE 5.26 – Taux de compétence des protons amides HN selon le modèle simple (en haut), selon le modèle concerté (au milieu) et le nombre de liaisons hydrogène compétentes formées entre les protons amides HN et l'eau selon le modèle concerté (en bas). En bleu par un trait plein sont représentées les données obtenues à partir des trajectoires aMD et par un trait en pointillés, celles obtenues à partir de la trajectoire cMD pour la forme liée à 9cRA, du LBD de RXRα.

Ces régions sont les suivantes : la partie N-terminale, la boucle H1/H3, la boucle H3/H4, le coude H4/H5, la partie C-terminale de H5 ainsi que le feuillet β et H6, la partie C-terminale de H7 et les boucles H7/H8, H8/H9, H9/H10 et H11/H12. Les seules différences observées se situent au niveau de la partie N-terminale de H3, avec une forte réduction du taux de compétence des protons amides HN. Une réduction est aussi observée au niveau du coude H4/H5 et de la partie C-terminale de H5 et H7. Ces régions interagissent toutes avec le ligand 9cRA dans la forme liée du LBD de RXRα. Au contraire, le taux de compétence des protons amides HN est légèrement supérieur au niveau de l'hélice H6 (Figure 5.26 - haut). En ce qui concerne la trajectoire MD classique, tout comme la forme libre, seuls les protons amides HN localisés dans les boucles sont dans un état compétent. Pour le modèle concerté, malgré des critères géométriques plus souples que ceux utilisés dans le modèle simple pour identifier les liaisons hydrogène, le fait que de deux molécules d'eau à proximité des protons amides HN soit nécessaire pour les considérer comme dans un état compétent, réduit très fortement leur taux de compétence (Figure 5.26 - bas).

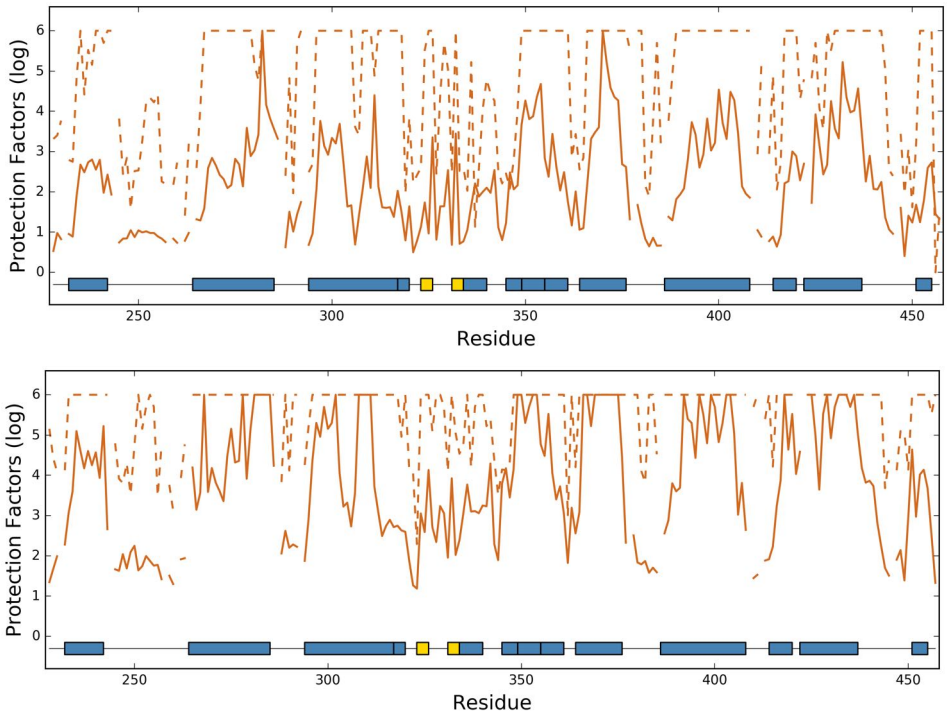


FIGURE 5.27 – Facteurs de protection PF (exprimé sur une échelle logarithmique) pour chaque proton amide HN selon le modèle simple (en haut) et selon le modèle concerté (en bas). En bleu par un trait plein, sont représentées les données obtenues à partir des trajectoires aMD et par un trait en pointillés, celles obtenues à partir de la trajectoire cMD pour la forme liée à 9cRA du LBD de RXRα.

La transformation du taux de compétence en facteur de protection PF permet de mieux distinguer les différences entre les protons amides HN. Ainsi on remarque que globalement la forme liée à 9cRA du LBD de RXRα possède des facteurs de protection PF plus élevés que ceux de la forme libre, exceptée dans la boucle H1/H3 dont la protection ne semble par influencée par la présence de 9cRA. Cohérent avec les observations précédentes, les facteurs de protection PF des régions N-terminale de H3, sont plus élevées, sans toutefois protéger totalement cette région contre l'échange HD, de même pour H4, H7, H8, H11 et H12. Cependant l'impact de 9cRA reste limité sur les facteurs de protection PF de la partie C-terminale de H5 et du feuillet β et a un effet inversé sur H6. Cette observation est valable que ce soit pour le modèle simple et concerté, à la seule différence que ceux obtenus avec ce dernier sont globalement plus élevé.

Au total, l'incorporation en deutérium au cours du temps a été prédite au sein de 37 peptides et comparée aux données expérimentales de la forme holo pour les deux modèles, simple et concerté (Figures 5.30 et 5.31 et Table 5.7).

Contrairement à ce qui était observé pour la forme apo, les trajectoires de dynamiques classiques, analysées avec le modèle simple, prédisent des incorporations en accord avec l'expérience pour plusieurs peptides. Outre les peptides 363-367 et 363-370, correspond au N-terminal de H8, et 391-400 et 392-400, correspondant à H9, dans le cas de la forme liée, un bon accord est aussi observé sur les peptides situés entre les résidus 279 et 343, ce qui couvre la partie C-terminale de H3 jusqu'au N-terminal de H7. Tous ces résidus sont à proximité du site d'interaction avec 9cRA. Il reste cependant des régions de désaccord : les extrémités N et C-terminales, dont l'échange prédit est trop lent.

En ce qui concerne les trajectoires aMD, l'accord avec les données expérimentales au sein des peptides localisés en N et C-terminal (qui correspondent à H1, H2, la partie N-terminale de H3, la boucle H11/H12 et H12) est bon (Figure 5.28 et Table 5.7). Par contre, tout comme pour la forme libre, l'incorporation en deutérium prédit au sein des autres peptides en utilisant le modèle simple est généralement plus rapide que celle mesurée expérimentalement. Dans la forme libre, certains peptides étaient prédits comme échangeant lentement alors que l'expérience indiquait un échange rapide. Ceci est encore le cas, mais dans une moindre mesure. Les peptides 347-353 et 419-432, correspondant à H7 et à la partie N-terminale de H10, sont en effet correctement prédits (RMSD de 9 % et 10 %). Par contre, les données expérimentales prédisent une incorporation significative pour les peptides 419-438, 423-438 et 433-438, qui correspondant à H10 tandis qu'une incorporation plus faible est prédite par le modèle, comme cela avait déjà été observée pour la forme libre (Figure 5.28 et Table 5.7).

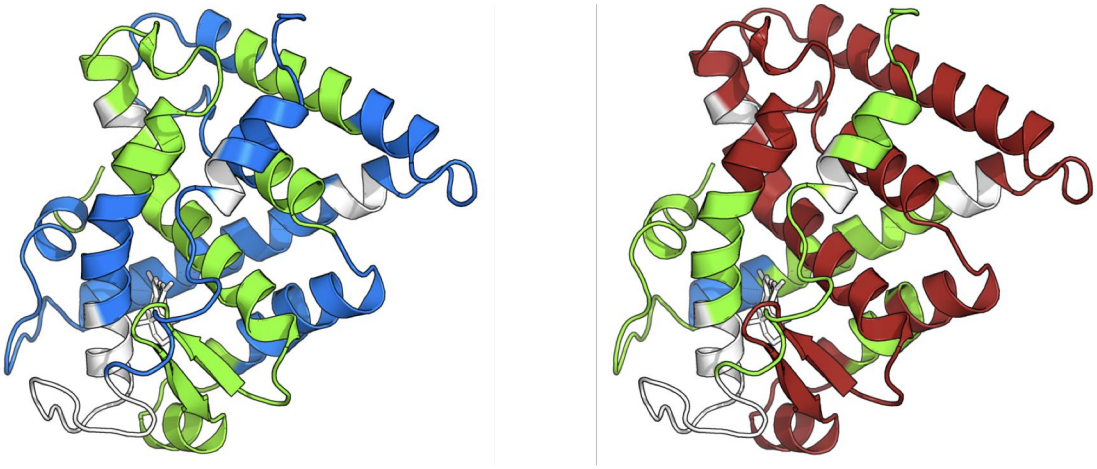


FIGURE 5.28 – Résumé des résultats obtenus à partir de la trajectoire cMD (à gauche) et des trajectoires aMD (à droite) de la forme liée à 9cRA du LBD de RXRα en utilisant le modèle simple. En vert, les régions dont le taux d'incorporation en deutérium prédit est en accord avec les données expérimentales. En rouge et en bleu, les peptides où le taux d'incorporation en deutérium prédit est soit supérieur ou bien inférieur aux données expérimentales, respectivement. En blanc, les résidus pour lesquelles aucune donnée expérimentale n'est disponible.

Lorsque le modèle concerté est utilisé, l'incorporation en deutérium prédite est systématiquement réduite. De ce fait, les prédictions faites à partir des trajectoires classiques sous-estiment systématiquement l'incorporation, sauf au niveau du peptide qui n'échange pas (hélice H9, peptides 391-400 et 392-400). En ce qui concerne les trajectoires aMD, les régions qui échangent de manière importante, et pour lesquels le modèle simple convenait, présentent un moins bon accord avec le modèle concerté. Il s'agit, des extrémités N et C-terminale du domaine, spécialement sur les peptides 271-278, 271-279 (N-terminal de H3) et 439-449, 439-450, 439-451 et 450-454 (C-terminal de H11, boucle H11/H12 et H12) (Figure 5.29 et Table 5.7). De même que pour la forme libre, l'accord est meilleur pour le restant des peptides qui appartiennent à des régions structurées. Une vitesse d'incorporation plus faible est aussi constatée au sein des peptides 347-353 (H7) et 419-432, 419-438, 423-438, 426-438 et 433-438 (H10 et H11). Sauf que dans le cas de la forme liée à 9cRA ce sont les peptides 309-325, 310-325, 326-343 et 326-346 (de H5 au N-terminal de H7, incluant le feuillet β et H6), où la vitesse d'incorporation en deutérium qui est plus élevée (Figure 5.28 et Table 5.7).

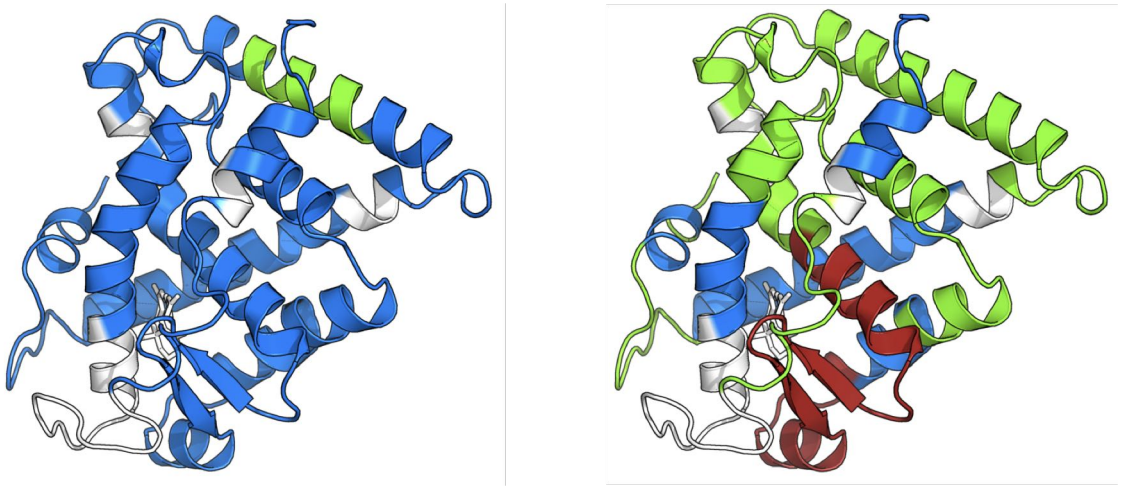


FIGURE 5.29 – Résumé des résultats obtenus à partir de la trajectoire cMD (à gauche) et des trajectoires aMD (à droite) de la forme liée à 9cRA du LBD de RXRα en utilisant le modèle concerté. En vert, les régions dont le taux d'incorporation en deutérium prédit est en accord avec les données expérimentales. En rouge et en bleu, les peptides où le taux d'incorporation en deutérium prédit est soit supérieur ou bien inférieur aux données expérimentales, respectivement. En blanc, les résidus pour lesquelles aucune donnée expérimentale n'est disponible.

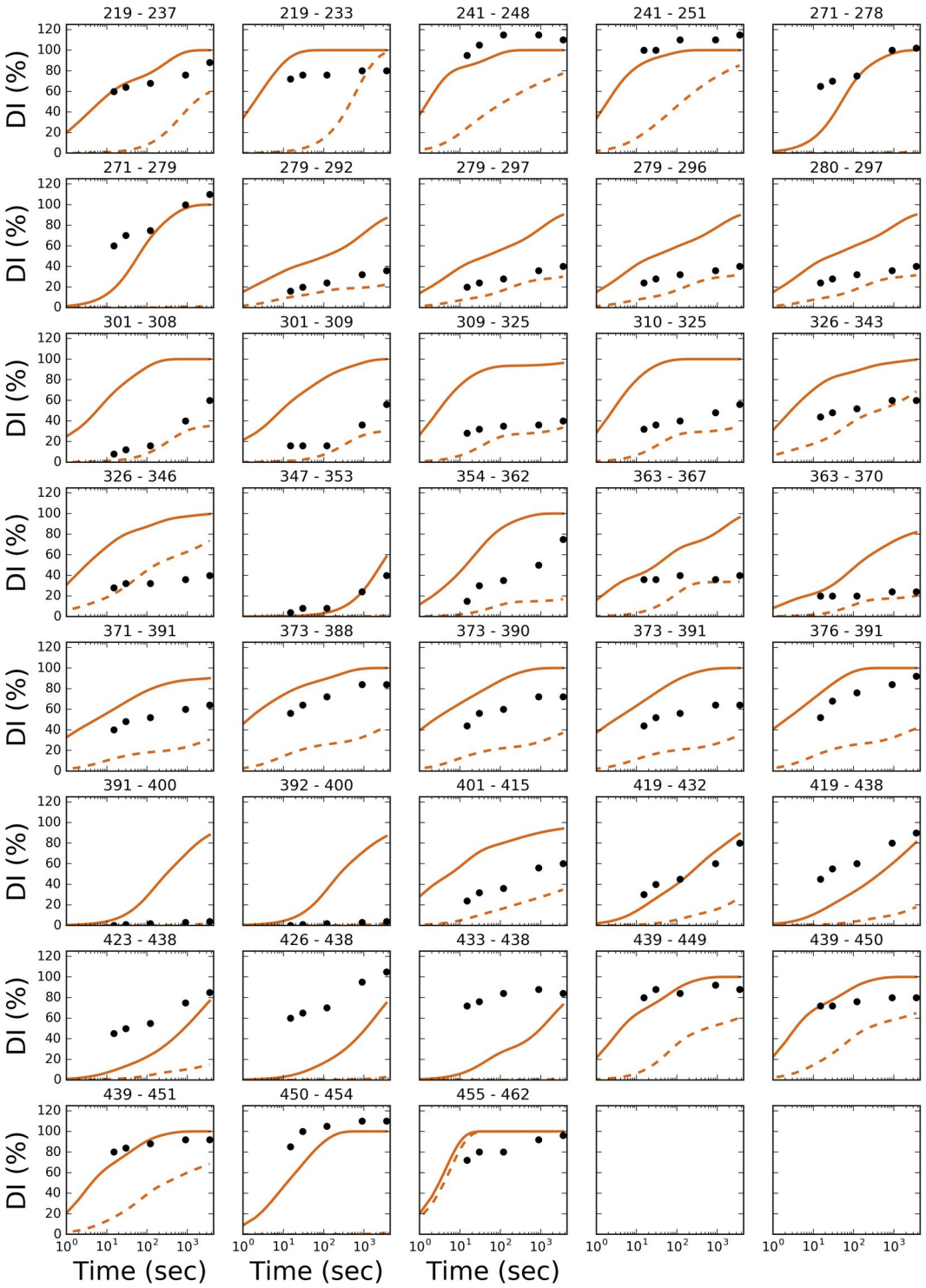


FIGURE 5.30 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme liée à 9cRA du LBD de RXRα. En orange par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires aMD, en orange avec des pointillés ceux prédits à partir d'une trajectoire cMD et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

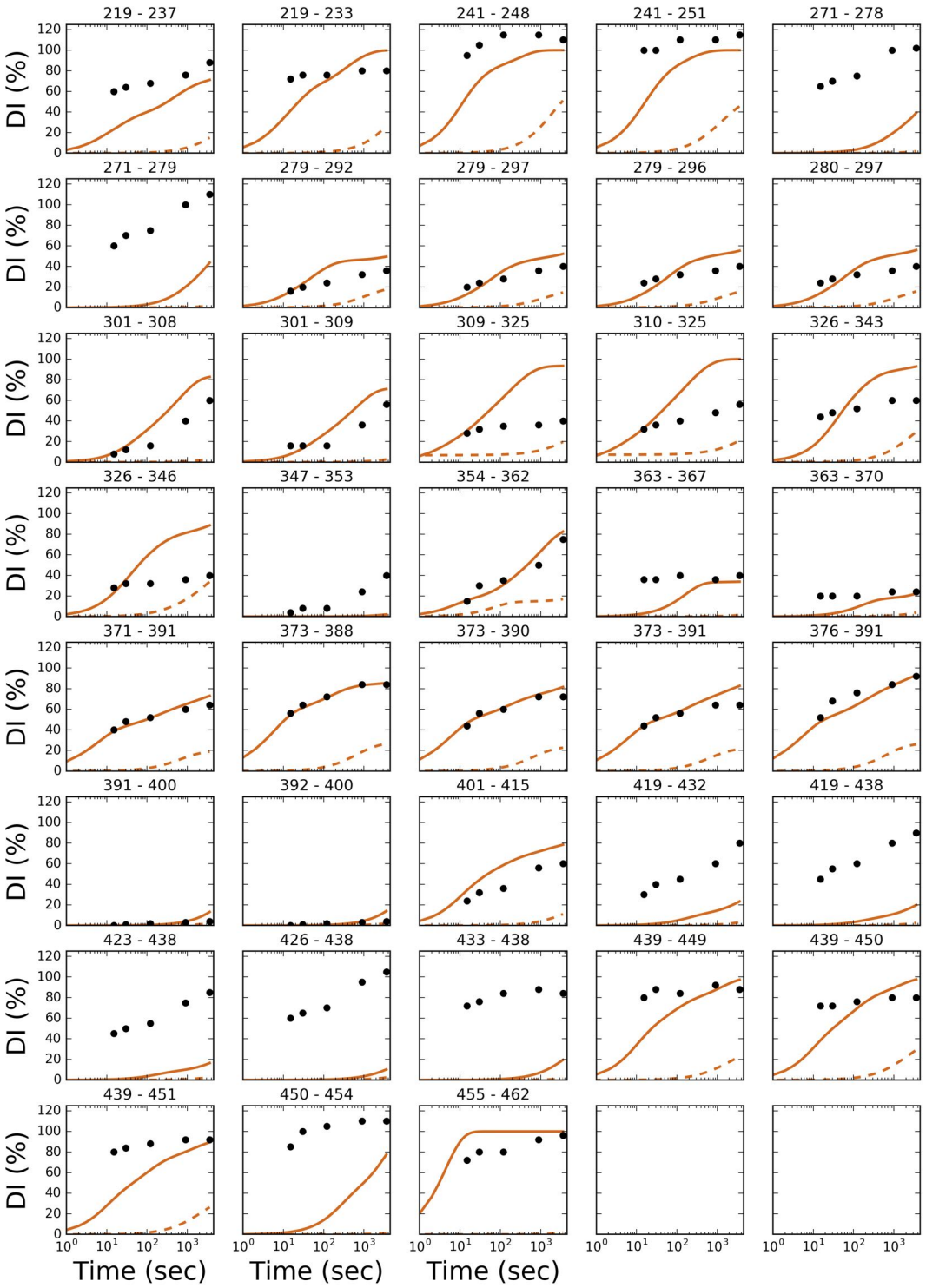


FIGURE 5.31 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme liée à 9cRA du LBD de RXRα. En orange par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires aMD, en orange avec des pointillés ceux prédits à partir d'une trajectoire cMD et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

	Modèle simple				Modèle « concerté »			
	aMD		cMD		aMD		cMD	
	Correlation	RMSD	Correlation	RMSD	Correlation	RMSD	Correlation	RMSD
^{219}GSHMTSSANEDMPVERILE^{237}	0.93	12	0.97	50	0.97	27	0.95	67
^{219}GSHMTSSANEDMPVE^{233}	0.85	22	0.88	53	0.95	20	0.73	70
^{241}AVEPKTET^{248}	0.91	14	0.74	60	0.87	28	0.39	94
^{241}AVEPKTETYVE^{251}	0.91	12	0.94	60	0.93	31	0.84	92
^{271}AADKQLFT^{278}	0.93	25	0.79	83	0.91	70	0.79	83
^{271}AADKQLFTL^{279}	0.95	26	0.85	85	0.93	69	0.85	85
^{279}LVEWAKRIPHFSEL^{292}	0.98	34	0.98	10	0.95	11	0.94	20
^{279}LVEWAKRIPHFSELPLDDQ^{297}	0.99	35	0.99	11	0.98	9	0.94	25
^{279}LVEWAKRIPHFSELPLDD^{296}	0.98	35	0.97	13	0.98	11	0.91	27
^{280}VEWAKRIPHFSELPLDDQ^{297}	0.98	35	0.97	13	0.98	12	0.91	27
^{301}LRAGWNEL^{308}	0.79	62	0.97	13	0.97	18	0.93	32
^{301}LRAGWNELL^{309}	0.82	54	0.94	16	0.93	14	0.96	31
^{309}LIASFSHRSIAVKDGIL^{325}	0.96	56	0.97	13	0.93	37	0.83	24
^{310}IASFSHRSIAVKDGIL^{325}	0.78	54	0.93	20	0.95	32	0.93	32
^{326}LATGLHVHRNSAHSAGVG^{343}	0.99	36	0.96	16	0.98	23	0.81	45
^{326}LATGLHVHRNSAHSAGVGAIF^{346}	0.93	54	0.94	20	0.92	33	0.94	23
^{347}DRVLTEL^{353}	0.98	9	0.95	21	0.95	21	0.95	21
^{354}VSKMRDMQM^{362}	0.97	41	0.90	34	0.96	8	0.91	34
^{363}DKTEL^{367}	0.58	35	0.53	20	0.43	23	0.55	37
^{363}DKTELGCL^{370}	0.91	37	0.87	11	0.93	13	0.78	21
^{371}RAIVLFNPDSKGLSNPAEVEA^{391}	0.97	24	0.96	33	0.98	5	0.92	46
^{373}IVLFNPDSKGLSNPAE^{388}	0.99	18	0.92	44	0.99	2	0.90	62
^{373}IVLFNPDSKGLSNPAEVE^{390}	0.97	26	0.90	37	0.96	5	0.88	52
^{373}IVLFNPDSKGLSNPAEVEA^{391}	0.98	31	0.90	34	0.95	10	0.88	48
^{376}FNPDSKGLSNPAEVEA^{391}	0.94	18	0.95	47	0.94	9	0.88	65
^{391}ALREKVYASL^{400}	0.97	50	0.82	2	0.84	4	0.82	2
^{392}LREKVYASL^{400}	0.97	50	0.82	2	0.84	5	0.82	2
^{401}EAYCKHKYPEQPGRF^{415}	0.97	38	0.98	23	0.97	16	0.85	40
^{419}LLRLPALRSIGLKC^{432}	0.98	10	0.98	41	0.98	43	0.93	53
^{419}LLRLPALRSIGLKCLEHLFF^{438}	0.99	26	0.96	60	0.96	60	0.86	67
^{423}PALRSIGLKCLEHLFF^{438}	0.99	29	0.99	56	0.98	56	0.88	63
^{426}RSIGLKCLEHLFF^{438}	0.98	50	0.87	80	0.90	78	0.87	80
^{433}LEHLFF^{438}	0.75	50	0.44	81	0.51	76	0.44	81
^{439}FKLIGDTPIDT^{449}	0.70	10	0.66	54	0.69	24	0.51	79
^{439}FKLIGDTPIDTF^{450}	0.97	15	0.98	37	0.97	20	0.83	67
^{439}FKLIGDTPIDTFL^{451}	0.99	8	0.97	48	0.97	30	0.78	79
^{450}FLMEM^{454}	0.95	22	0.56	102	0.74	76	0.56	102
^{455}LEAPHQMT^{462}	0.69	17	0.71	16	0.69	17	0.82	84

TABLE 5.7 – Détails par peptide de la corrélation et du RMSD obtenus en comparant le taux d'incorporation en deutérium prédit à partir des trajectoires aMD et de la trajectoire cMD au taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes pour la forme liée à 9cRA du LBD de RXRα.

5.3.5 Analyse des conformations échantillonnées

5.3.5.1 Analyse sur composante principale (PCA)

5.3.5.1.1 Forme libre du LBD de RXRα

On remarque que ce soit en prenant les angles dièdres Φ et Ψ ou les pseudo angles dièdres Cα, que le nombre nécessaire de dimensions à prendre en compte est nettement supérieur à $n >> 2$ pour avoir un ratio $S_n/S_N \geq 0.9$. En effet, dans le cas des angles dièdres Φ et Ψ, 323 dimensions sont nécessaires pour représenter au moins 90 % de la variance globale, sachant que les deux premières composantes principales ne représentent que 9 % (Figure 5.32).

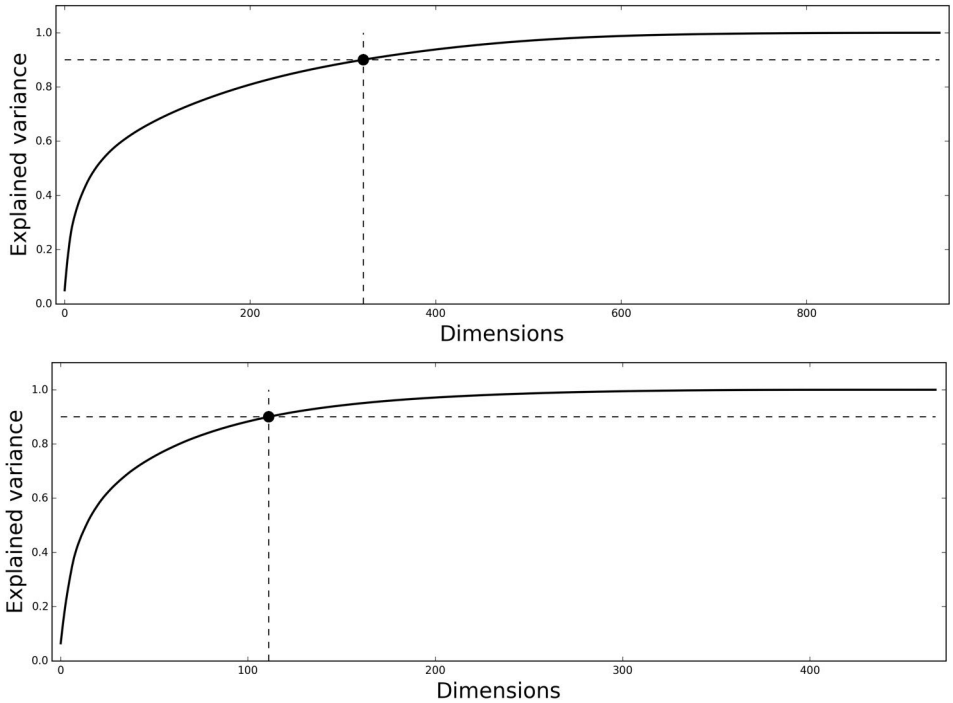


FIGURE 5.32 – Évolution de la variance cumulée en fonction du nombre de dimensions considérées à partir des angles dièdres Φ et Ψ (en haut) et des pseudo angles dièdres Cα (en bas), provenant des trajectoires aMD de la forme libre du LBD de RXRα. Par un point et des traits en pointillés est indiqué à quel moment le nombre de dimensions considérées est représentatif d'au moins 90 % de la variance globale.

Ainsi, l'utilisation de seulement 2 dimensions n'est pas suffisamment représentative et donc ne peut pas être utilisée comme coordonnées de réactions pour visualiser le paysage énergétique. De même avec l'utilisation des pseudo angles dièdres Cα, où 112 dimensions sont nécessaires pour représenter 90 % de

la variance globale avec les deux premières qui ne représentent que 12 % (Figure 5.32). De ces résultats, on peut conclure que l'utilisation de la PCA en prenant les angles Φ et Ψ ou les pseudo angles dièdres Cα n'est pas une méthode adaptée pour analyser les trajectoires aMD de la forme libre du LBD de RXRα.

5.3.5.1.2 Forme liée à 9cRA du LBD de RXRα

De même que pour les trajectoires aMD de la forme libre du LBD de RXRα, on obtient des résultats similaires avec les trajectoires aMD de la forme liée à 9cRA du LBD de RXRα. Pour obtenir une variance globale d'au moins 90 %, 311 et 102 dimensions sont nécessaires, avec les angles dièdres Φ et Ψ ou les pseudo angles dièdres Cα, respectivement (Figure 5.33). Également, les deux premières composantes principales ne sont pas représentatives de l'ensemble des données avec une variance cumulée de seulement 7 et 12 %.

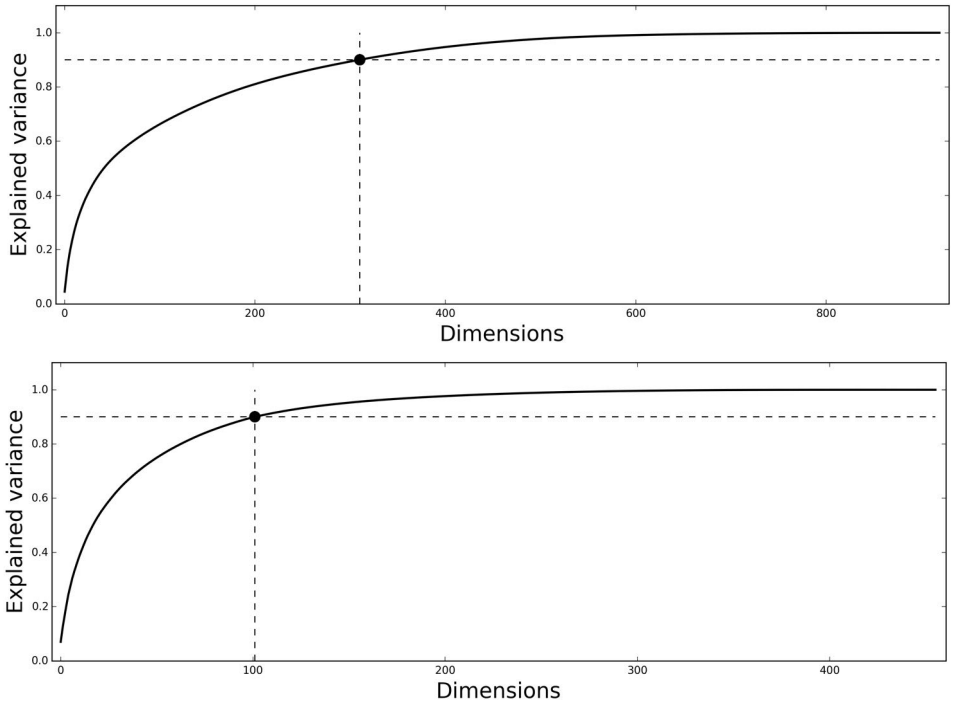


FIGURE 5.33 – Évolution de la variance cumulée en fonction du nombre de dimensions considérées à partir des angles dièdres Φ et Ψ (en haut) et des pseudo angles dièdres Cα (en bas), provenant des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα. Par un point et des traits en pointillés est indiqué à quel moment le nombre de dimensions considérées est représentatif d'au moins % de la variance globale.

Ainsi, on peut conclure que la PCA n'est de même pas adaptée à l'analyse des trajectoires aMD de

la forme liée à 9cRA du LBD de RXRα. Ceci justifie alors le développement d'une nouvelle méthode de réduction de la dimensionnalité, du fait de la nature non linéaire des données et du nombre important de conformations différentes générées.

5.3.5.2 Stochastic Proximity Embedding (pSPE)

5.3.5.2.1 Forme libre du LBD de RXRα

Avant d'analyser les conformations regroupées par l'algorithme pSPE, nous devons déterminer la valeur optimale du cutoff r_c qui maximise la corrélation tout en diminuant le stress (voir méthodes). Dans le cas des trajectoires aMD de la forme libre du LBD de RXRα, les résultats montrent que la corrélation γ (en bleu ci-dessous) se stabilise à partir d'un cutoff r_c égal à une distance dihédrale supérieure à 0.40. À partir de cette valeur, chaque configuration finale obtenue à l'aide de la méthode pSPE possède une corrélation γ supérieure ou égale à 60 %. Quant au stress, on constate aussi une augmentation progressive à partir d'un cutoff r_c égal à 0.30 jusqu'à atteindre aussi un plateau à partir de 0.44, avec un stress supérieur ou égal à 0.17, pour toutes les trajectoires aMD. Cela signifie qu'à partir d'un cutoff r_c de 0.44, toutes les distances entre les conformations sont considérées lors de la procédure d'optimisation pSPE. Ainsi à partir des résultats obtenus, nous avons choisi un cutoff r_c égal à 0.33 qui correspond à un stress faible de moins de 0.01 tout en montrant une corrélation supérieure à 0.6. Ce qui voudra dire que lors du processus d'optimisation par la méthode pSPE, seront considérées comme « voisines » les conformations séparées par une distance dihédrale inférieure ou égale à un cutoff r_c de 0.33.

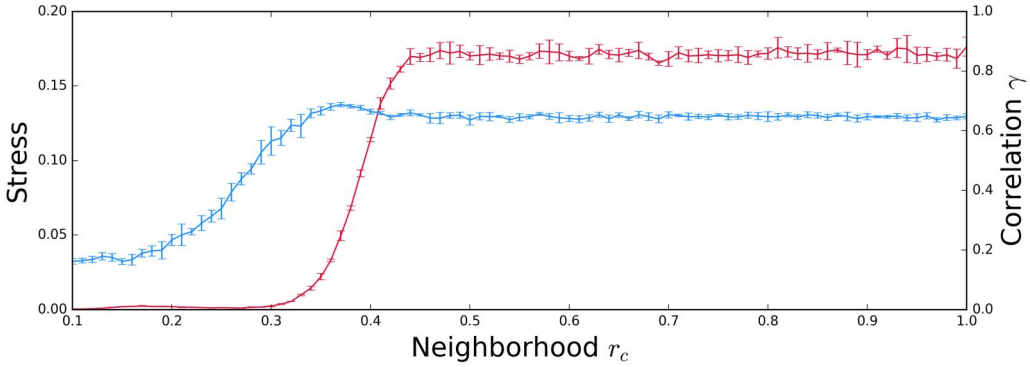


FIGURE 5.34 – Évolution du stress (en rouge) ainsi que de la corrélation γ (en bleu) en fonction du cutoff r_c, allant de 0.1 à 1.0 par intervalle de 0.01, avec 5000 cycles d'optimisation et 10.000 conformations au total, avec un intervalle de 100 ps, provenant des trajectoires aMD de la forme libre du LBD de RXRα.

La qualité de la configuration finale dépend également du nombre de cycles d'optimisation de la méthode pSPE. Plus on augmente le nombre total de cycles d'optimisation, plus fidèle à l'espace réel sera la configuration finale dans l'espace réduit. Néanmoins, on converge entre 5000 et 10000 cycles d'optimisation. On observe une baisse du stress, passant d'environ 0.12 et plus, pour 10 cycles d'optimisation, à environ 0.01 à partir de 5000 cycles d'optimisation, jusqu'à 10 000 cycles d'optimisation où aucune amélioration significative n'est observée. Nous constatons le même comportement en ce qui concerne la corrélation γ en fonction du nombre de cycles d'optimisation. On converge aussi, où à partir de 5000 cycles d'optimisation on atteint plus de 60 % de corrélation.

Cette observation est valable, quelque soit le nombre de conformations utilisées durant le processus d'optimisation pSPE. Le fait d'augmenter le nombre de conformations en passant de 10 000 à 50 000 et enfin 100 000 ne semble pas influencer significativement le nombre minimal nécessaire pour obtenir une configuration finale optimale. Ainsi les résultats montrent que la méthode pSPE est stable, quelque soit le nombre de conformations utilisées lors du processus d'optimisation.

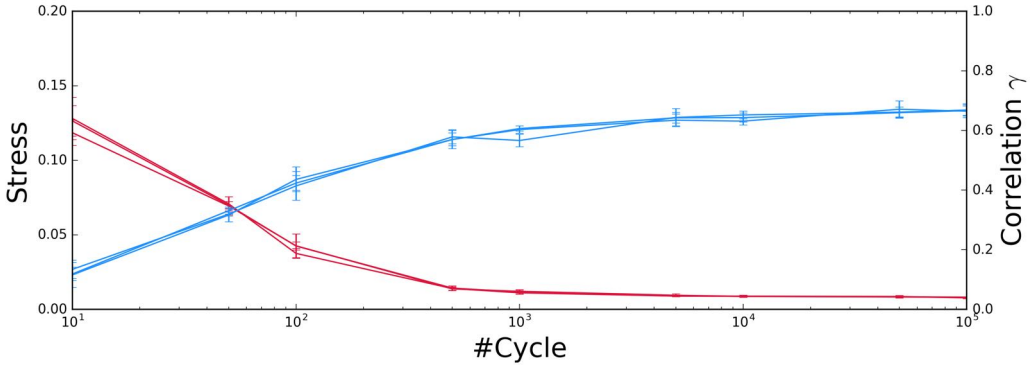


FIGURE 5.35 – Évolution du stress (en rouge) ainsi que de la corrélation γ (en bleu) en fonction du nombre de cycles d'optimisation, 10, 50, 100, 500, 1000, 5000, 10 000, 50 000 et 100 000, avec un cutoff fixe r_c égal à 0.33 et aussi en faisant varier le nombre de conformations utilisées, en passant de 10 000 à 50 000 et 100 000.

5.3.5.2.2 Forme liée à 9cRA du LBD de RXRα

Les trajectoires aMD de la forme liée du LBD de RXRα donnent des résultats similaires à ceux de la forme libre. La corrélation γ se stabilise et plafonne aux alentours de 60 % à partir d'un cutoff r_c égal à 0.40. De même pour le stress avec une augmentation progressive à partir d'un cutoff r_c de 0.26, jusqu'à

atteindre un plateau à 0.40 avec un stress à environ 0.17. Ainsi en nous basant sur l'évolution du stress et de la corrélation γ en fonction du cutoff r_c, nous choisirons un cutoff r_c égal à 0.27 qui correspond à un stress très faible de l'ordre de 0.004 avec une corrélation de 0.6.

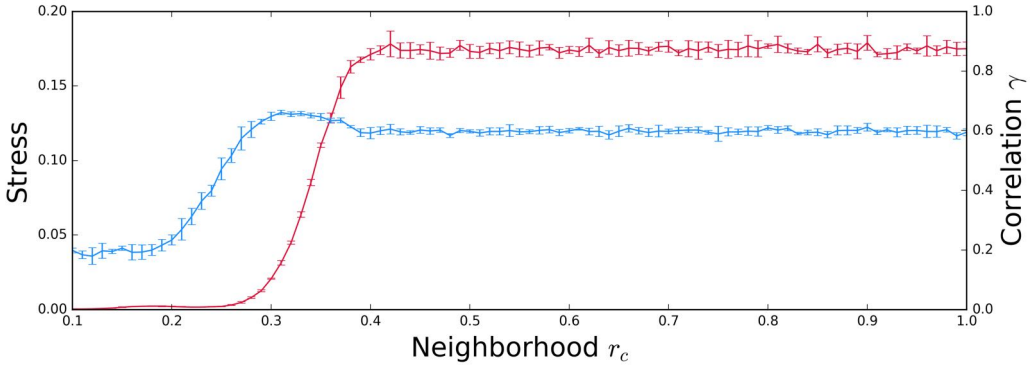


FIGURE 5.36 – Évolution du stress (en rouge) ainsi que de la corrélation γ (en bleu) en fonction du cutoff r_c, allant de 0.1 à 1.0 par intervalle de 0.01, avec 5000 cycles d'optimisation et 10.000 conformations au total, avec un intervalle de 100 ps, provenant des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα.

Nous avons également testé l'influence du nombre de cycles d'optimisation sur la qualité de la configuration finale obtenue. On obtient les mêmes résultats que dans le cas de la forme libre. À mesure que le nombre de cycles d'optimisation augmente, on observe une augmentation de la corrélation γ ainsi qu'une diminution progressive du stress. Dans le cas du stress, un minimum est atteint entre 5000 et 10000 cycles d'optimisation. Cependant, en ce qui concerne la corrélation γ, le plateau n'est réellement atteint qu'à partir de 10.000 cycles d'optimisation avec une corrélation γ de l'ordre de 60 %, contrairement aux trajectoires aMD de la forme libre où l'on observe un plateau à partir de 5000 cycles d'optimisation. De plus, le fait d'augmenter le nombre de conformations, en passant de 10000 à 50000 et 100000 conformations, n'influence en rien le nombre minimal de cycles d'optimisation nécessaire pour obtenir une configuration finale optimale. Même avec 100000 conformations, la corrélation γ et le stress convergent à la même vitesse. À partir de 10000 cycles d'optimisation, on atteint aussi une corrélation aux alentours de 60% et un stress inférieur aussi à 0.01. Au final, on peut à nouveau conclure dans le cas des trajectoires aMD de la forme liée du LBD de RXRα,quelque soit le nombre de conformations utilisées lors du processus d'optimisation, les résultats obtenus sont stables malgré le fait que la méthode pSPE soit de nature stochastique.

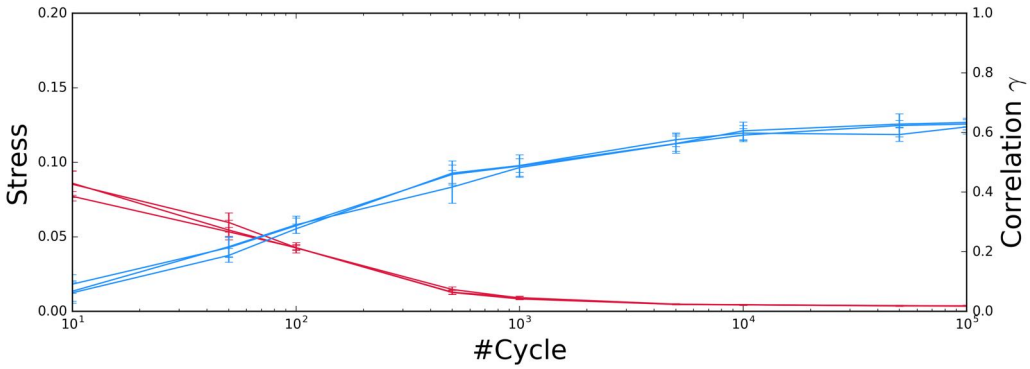


FIGURE 5.37 – Évolution du stress (en rouge) ainsi que de la corrélation γ (en bleu) en fonction du nombre de cycles d'optimisation C, 10, 50, 100, 500, 1000, 5000, 10 000, 50 000 et 100 000, avec un cutoff fixe r_c égal à 0.27 et aussi en faisant varier le nombre de conformations utilisées, en passant de 10 000 à 50 000 et 100 000.

5.3.5.3 Description des conformations

5.3.5.3.1 Forme libre du LBD de RXRα

Le paysage énergétique obtenu à partir de la méthode pSPE ainsi qu'en repondérant l'énergie en fonction du biais (voir Méthodes pour les détails de la repondération) nous montre que ce dernier est composé de plusieurs minimas. Celui-ci est composé principalement d'un bassin de basse énergie au centre, lui-même composé de plusieurs minimas, et dont plusieurs ramifications s'échappent. Ces ramifications correspondent à des trajectoires aMD explorant chacune un espace conformationnel différent de la forme libre du LBD de RXRα. Et selon les états échantillonnés au cours des trajectoires aMD, ces ramifications sont composées soit d' un ou plusieurs minimas de tailles variables selon si la conformation explorée est favorable ou non énergétiquement.

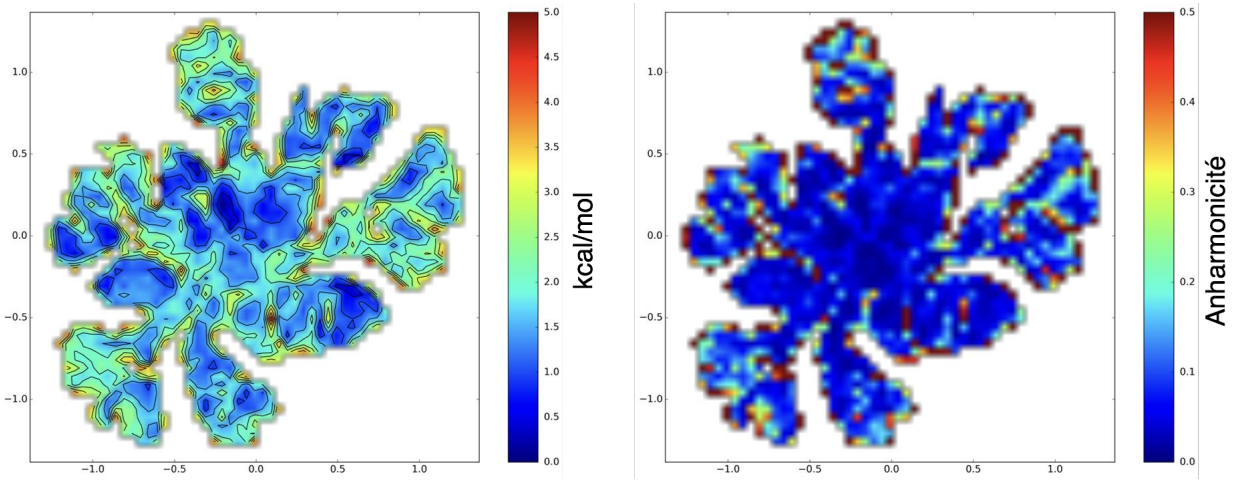


FIGURE 5.38 – À gauche, le paysage énergétique (en kcal/mol) de l'espace conformationnel échantillonné au cours des trajectoires aMD de la forme libre du LBD de RXRα. Celui-ci a été obtenu en repondérant l'énergie à partir du biais appliqué à chaque conformation et en utilisant comme coordonnées de réaction la configuration optimale obtenue à partir de la méthode pSPE. À droite est représenté l'ahnarmonicité pour chaque classe en utilisant la même configuration optimale obtenue à l'aide de la méthode pSPE. L'intervalle choisi est de 0.05 en ne considérant que les classes ayant au moins 10 conformations.

La visualisation des principaux minimas énergétiques montre une très grande variabilité conformationnelle de la forme libre. Les principales différences conformationnelles sont localisées dans la région impliquée dans la liaison au ligand. Cette région comprend essentiellement la boucle H1-H3, l'hélice H3, H5 ainsi que le feuillet β et H6 et puis toute la partie C-terminale avec H11 et H12. Parmi cette grande diversité de conformations retrouvées au cours des trajectoires aMD, il a été possible de les classer en famille selon les particularités conformationnelles observées (Figures 5.39, 5.40, 5.41 et 5.42).

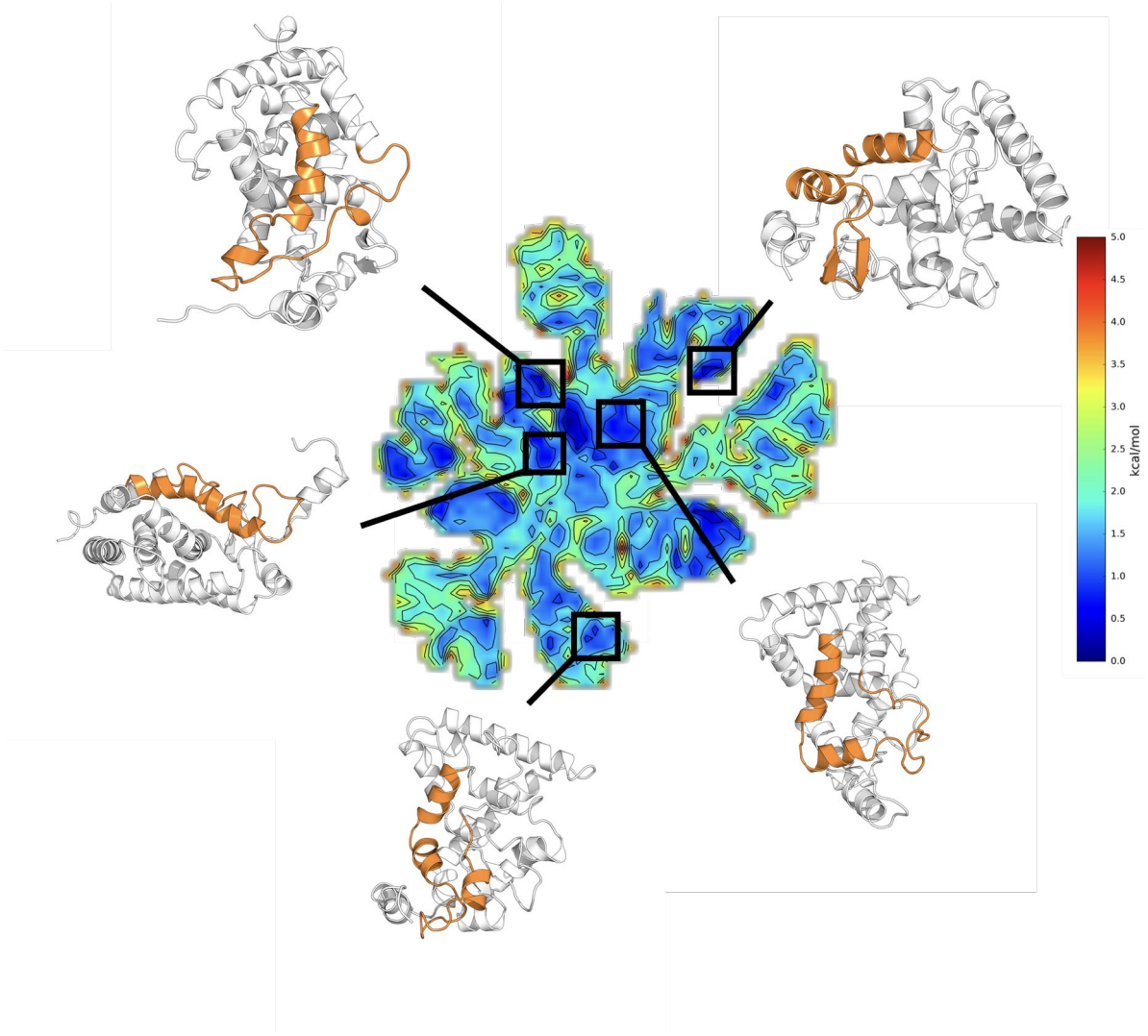


FIGURE 5.39 – Résumé de l'ensemble des conformations pouvant être adoptées au niveau de la boucle H1-H3 et de l'hélice H3 au sein des minimas sélectionnés du paysage énergétique de la forme libre du LBD de RXRα. Bien qu'au sein de la majorité des minimas l'hélice H2 ne soit pas présente dans la boucle H1-H3, il est possible de la retrouver dans certains des minimas sélectionnés. De plus, il est possible d'observer la formation d'un feuillet β au sein de la boucle H1-H3. Quant à l'hélice H3, majoritairement entièrement repliée, elle contribue très fortement à la diversité conformationnelle observée en se déstructurant partiellement au niveau de la moitié N-terminale. Le domaine est représenté en mode cartoon, avec en orange, la boucle H1-H3 et l'hélice H3, et en blanc le restant du domaine.

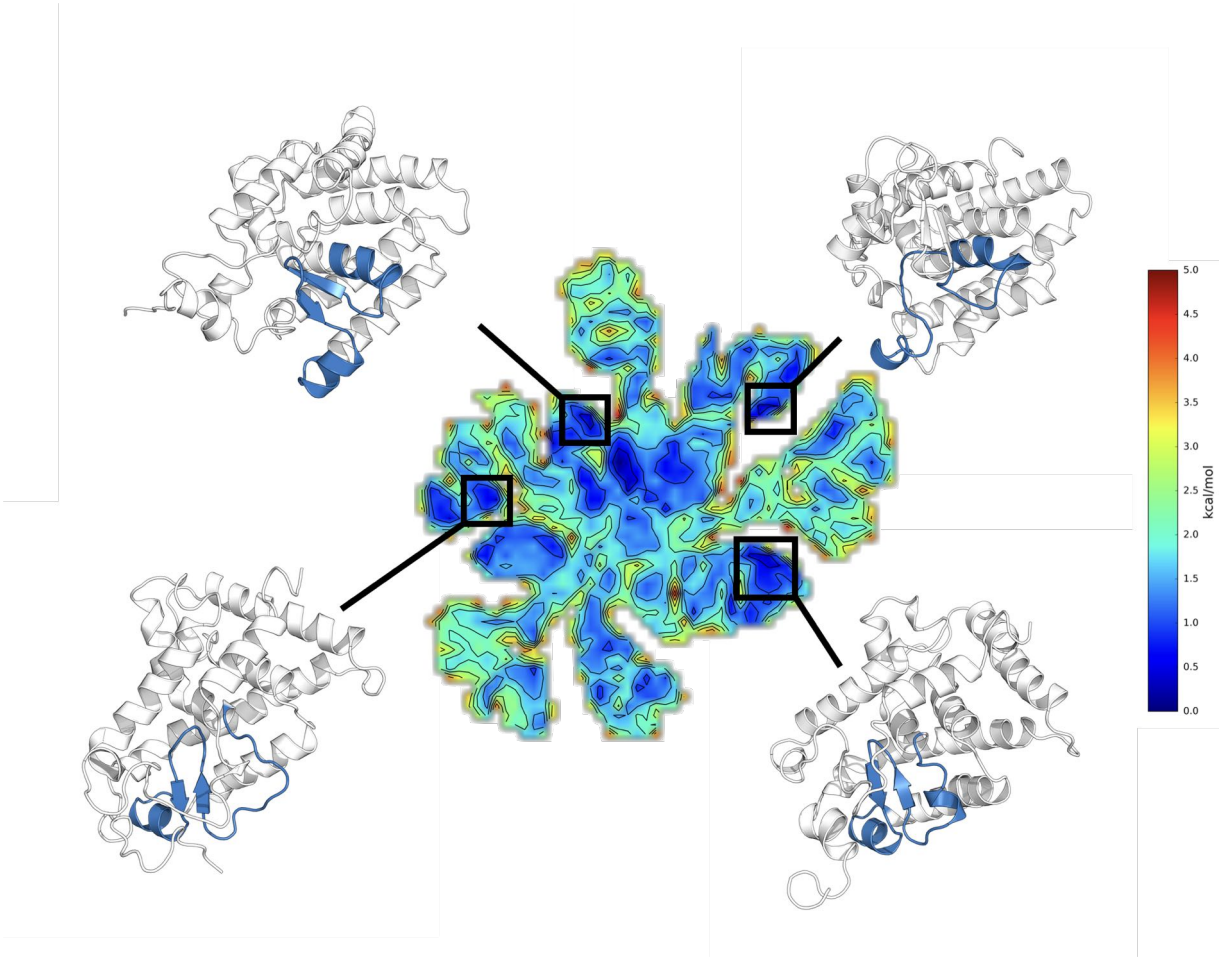


FIGURE 5.40 – Résumé de l'ensemble des conformations pouvant être adoptées au niveau de la région du feuillet β au sein des minimas sélectionnés du paysage énergétique de la forme libre du LBD de RXRα. De manière générale, la partie C-terminale de H5 est en conformations dépliée. Quant au feuillet β, celui-ci est retrouvé formé au sein de la majorité des minimas. Le domaine est représenté en mode cartoon, avec en bleu, l'hélice H5, le feuillet β et H6, et en blanc le restant du domaine.

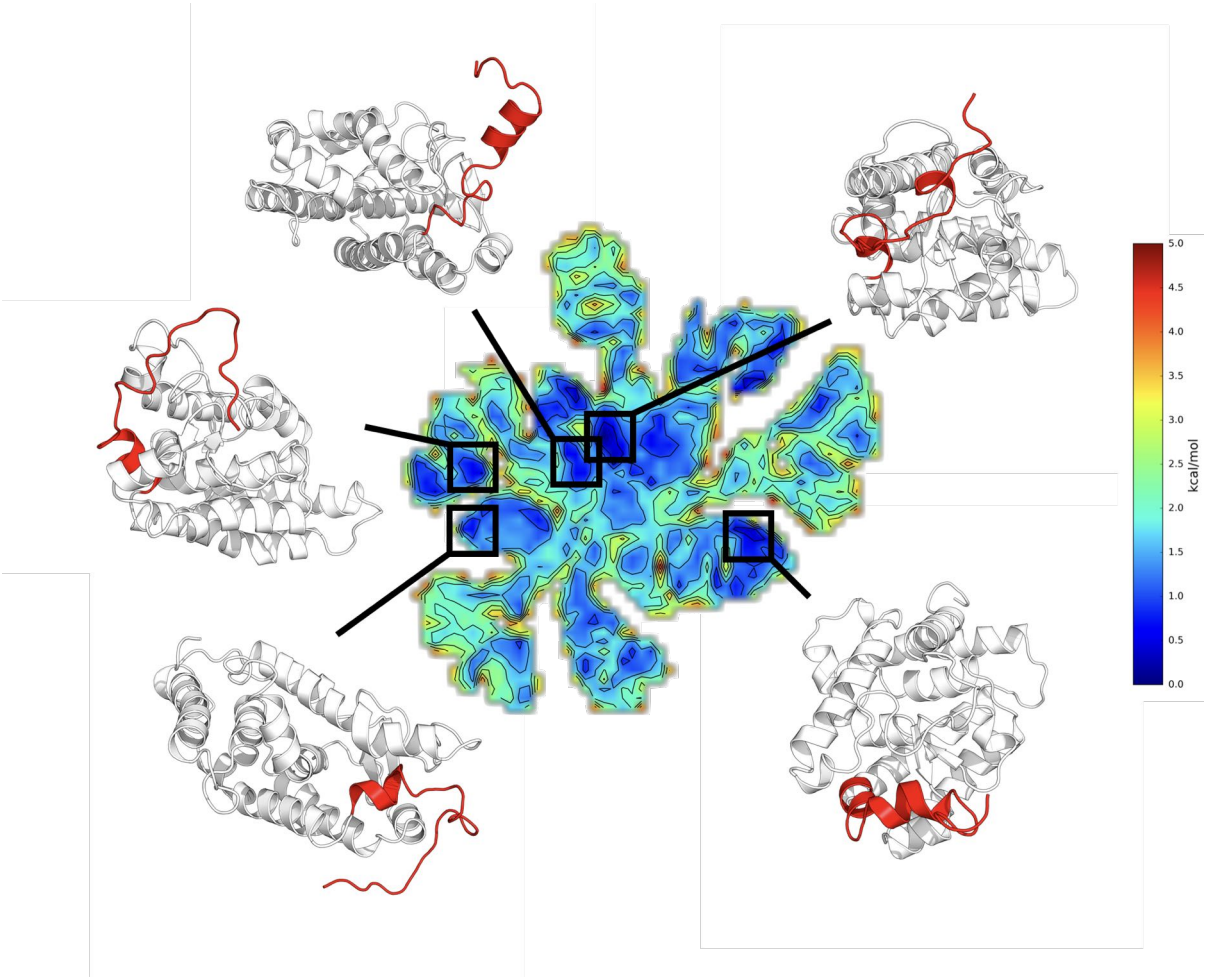


FIGURE 5.41 – Résumé des conformations adoptées par la partie C-terminale du domaine, contenant H11 et H12, au sein des principaux minimas du paysage énergétique de la forme libre du LBD de RXRα. Interagissant principalement avec H3, la boucle H1-H3 et le feuillet β, H12 peut être soit structurée en hélice, de taille variable, ou bien être complètement déstructurée selon les minimas. Concernant H11, celle-ci est majoritairement structurée en hélice et reste positionnée au sein du domaine du fait de la présence de plusieurs résidus TRP, participant au maintien du coeur hydrophobe. Le domaine est représenté en mode cartoon avec en rouge la partie C-terminale et en blanc le restant du domaine.

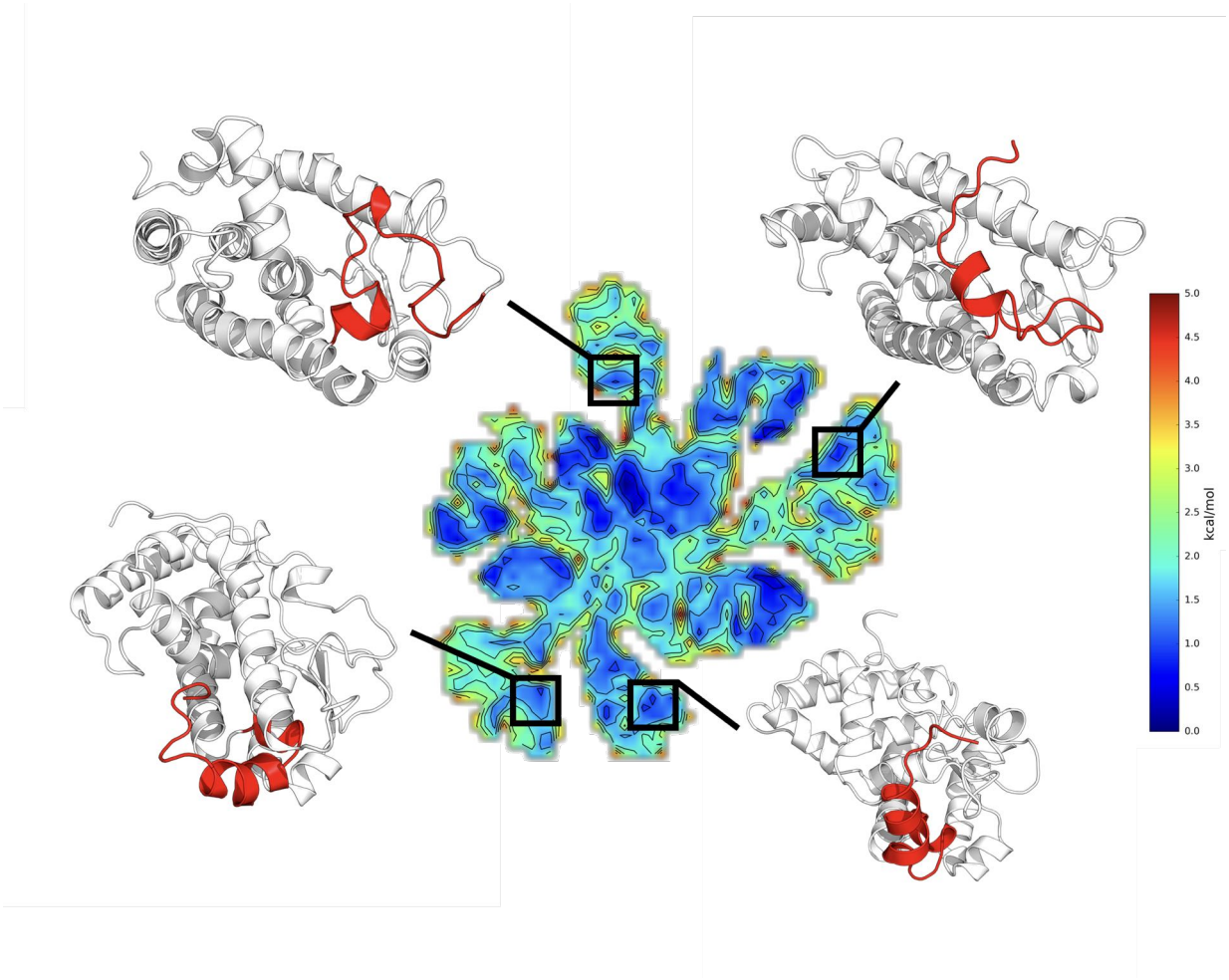


FIGURE 5.42 – Résumé des conformations alternatives pouvant être adoptées pour la partie C-terminale, contenant H11 et H12, au sein des principaux minimas sélectionnés de la forme libre du LBD de RXRα. Parmi les minimas sélectionnés, deux présentent des conformations alternatives intermédiaires avec H12, retenus soit par H11 ou bien dus à une déstructuration partielle de H3. Concernant les autres conformations où H12 est entièrement du côté de l'hélice H4, elles peuvent être soit retrouvées sous forme déstructurée ou bien structurée en hélice dans une position proche de celle retrouvée dans les formes liées en complexe avec un ligand agoniste. Le domaine est représenté en mode cartoon avec en rouge la partie C-terminale et en blanc le restant du domaine.

La première grande particularité conformationnelle se situe au niveau de la boucle H1-H3 et plus particulièrement au niveau de H3. Sans surprise, on remarque que la boucle H1-H3 ne présente pas d'hélice H2 au sein de la majorité des minimas énergétiques, celle-ci étant présente seulement dans 20 % des conforma-

tions explorées d'après l'analyse avec DSSP, et que la boucle est donc principalement sous forme dépliée. La boucle H1-H3 ne semble pas non plus montrer d'interactions préférentielles, interagissant à la fois avec le feuillet β, H12 et H3. On remarquera tout de même la présence d'un feuillet β composé de deux brins antiparallèles au sein d'un minimum, résidus 21-24 et 27-29, à l'emplacement même de l'hélice H2 au sein de la structure cristallographique de la forme libre. L'hélice H3 reste la majorité du temps (environ 60 %) bien structurée en hélice α au cours des trajectoires aMD, mais on remarque qu'au sein de certains minimas elle est partiellement déstructurée dans sa moitié N-terminale. Un examen attentif montre qu'il est possible de distinguer plusieurs parties au sein de H3 qui ont des propensions différentes à rester en hélice : les résidus 264-269 ont une propension bien plus importante à rester en hélice α que les résidus 269-274 tandis que la partie C-terminale, résidus 275-285, reste totalement en hélice α. De fait, cette différence de stabilité au sein de la moitié N-terminale de H3 va induire d'importants changements structuraux très largement visibles au sein du domaine (Figure 5.39).

Une autre région montrant de la diversité structurale comprend l'hélice H5 et la région du feuillet β. Les conformations retrouvées au sein des minima énergétiques tendent à montrer que les résidus 310-318, constituant la partie C-terminale de H5 dans la structure cristallographique, peuvent également adopter une configuration dépliée. De même que l'hélice H3, H5 comprend des sections distinctes qui ont des propensions différentes à rester en hélice. On observe au sein de certains minimas une déstructuration au niveau des résidus 310-314, tout en gardant les résidus 315-319 dans une configuration hélicoïdale. Quant à la moitié N-terminale de H5, résidus 305-309, elle reste stable en hélice dans l'ensemble des minimas. Concernant le feuillet β, il est présent dans tous les minimas, mais de taille variable selon les conformations observées, excepté dans deux minimas où l'on observe sa complète déstructuration (Figure 5.40).

Une autre part importante de la variabilité conformationnelle de la forme libre se situe au niveau de la région C-terminale du domaine. Celle-ci peut adopter un nombre très important de conformations. En particulier, certains minima présentent une région C-terminale contenant une hélice H12 et d'autres une complète déstructuration de cette région. Lorsque l'hélice H12 est présente, sa taille peut être très variable, se limitant parfois à seulement un tour d'hélice. En ce qui concerne les interactions observées, la partie C-terminale interagit préférentiellement avec la boucle H1-H3, la partie N-terminale de H3, le feuillet β et H6, voire la partie N-terminale de H7 (Figure 5.41). Plus rarement, on peut observer un déplacement de la partie C-terminale du domaine (et donc de H12) du côté opposé au feuillet β, vers l'hélice H4 (Figure 5.42). Cette observation est particulièrement intéressante, car dans la forme liée en présence d'un ligand agoniste, H12 se replie contre H4. L'observation de positions de H12 vers H4 en l'absence de ligand indique

donc que ces conformations appartiennent au paysage conformationnel de la séquence apo également. Cependant, ces conformations où H12 se déplace vers H4 ne sont pas strictement identiques à la structure cristallographique de la forme active, en complexe avec 9cRA. Lorsque H12 se place de manière très semblable à la forme active, les interactions hydrophobes sont maintenues entre les résidus Phe437 et Phe438 au niveau de H11.

5.3.5.3.2 Forme liée à 9cRA du LBD de RXRα

De même que pour la forme libre, le paysage énergétique obtenu est composé d'un bassin de basse énergie au centre avec plusieurs ramifications partant de ce dernier. Ces ramifications correspondent aussi à des trajectoires aMD ayant explorées des conformations différentes de la forme liée.

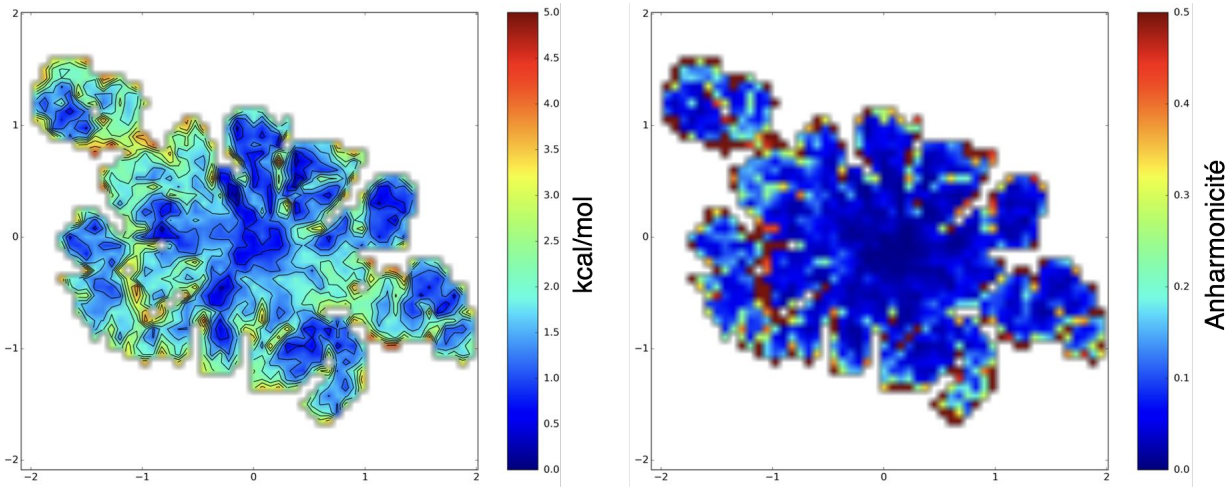


FIGURE 5.43 – À gauche, le paysage énergétique (en kcal/mol) de l'espace conformationnel échantillonné au cours des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα. Celui-ci a été obtenu en repondérant l'énergie à partir du biais appliqué à chaque conformation et en utilisant comme coordonnées collectives la projection optimale obtenue à partir de la méthode pSPE. À droite est représentée l'anharmonicité pour chaque classe en utilisant de même la configuration optimale obtenue à l'aide de la méthode pSPE. L'intervalle choisi est de 0.075 en ne considérant que les classes ayant au moins 10 conformations.

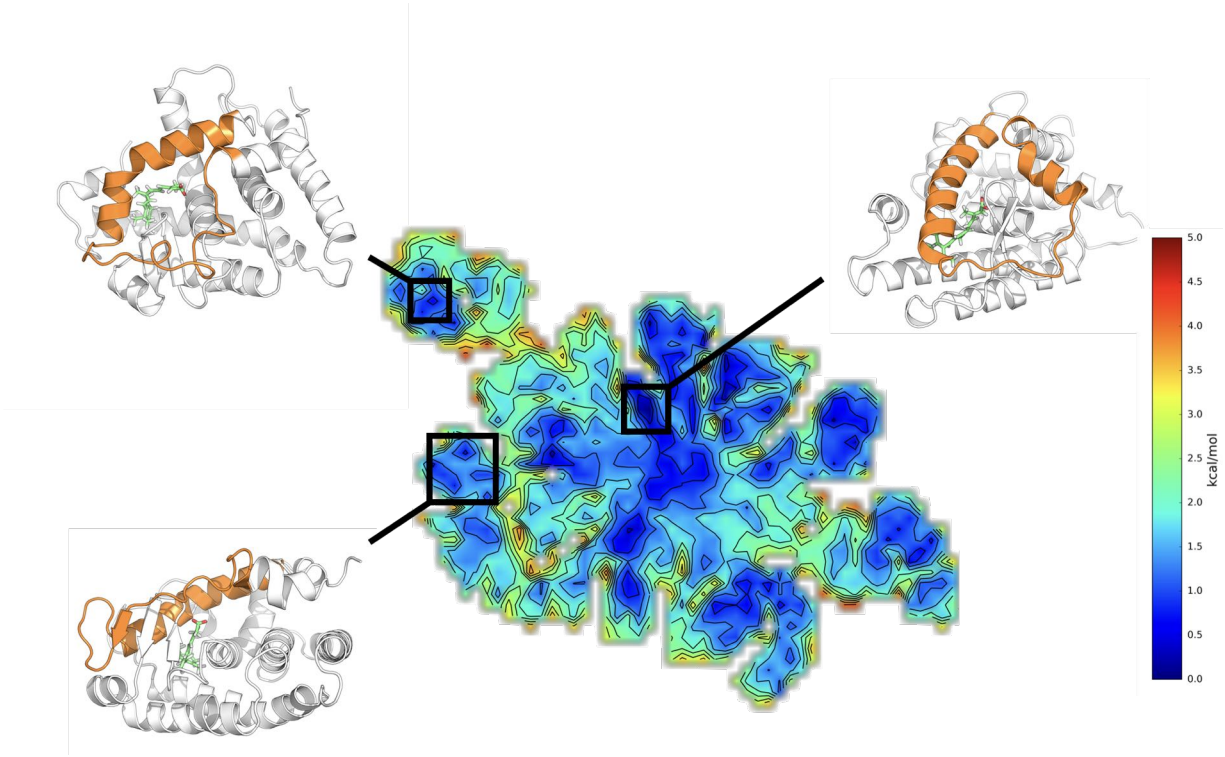


FIGURE 5.44 – Résumé des conformations pouvant être adoptées par la boucle H1-H3 et H3. Ces conformations ont été retrouvées au sein des minimas sélectionnés du paysage énergétique obtenu à partir des trajectoires de la forme liée du LBD de RXRα. Du fait de la présence de 9cRA, la diversité conformationnelle pour cette région est plus restreinte. Cependant, la boucle H1-H3 reste majoritairement dépliée, mais garde une certaine propension à créer soit une hélice H2 ou bien à former un troisième brin β parallèle avec le feuillet β canonique. Concernant H3, cette dernière reste très largement repliée avec des degrés de courbure différents. Le domaine est représenté en cartoon ainsi que le ligand 9cRA en représentation sticks en vert. La boucle H1-H3 et H3 sont en orange avec le restant du domaine en blanc.

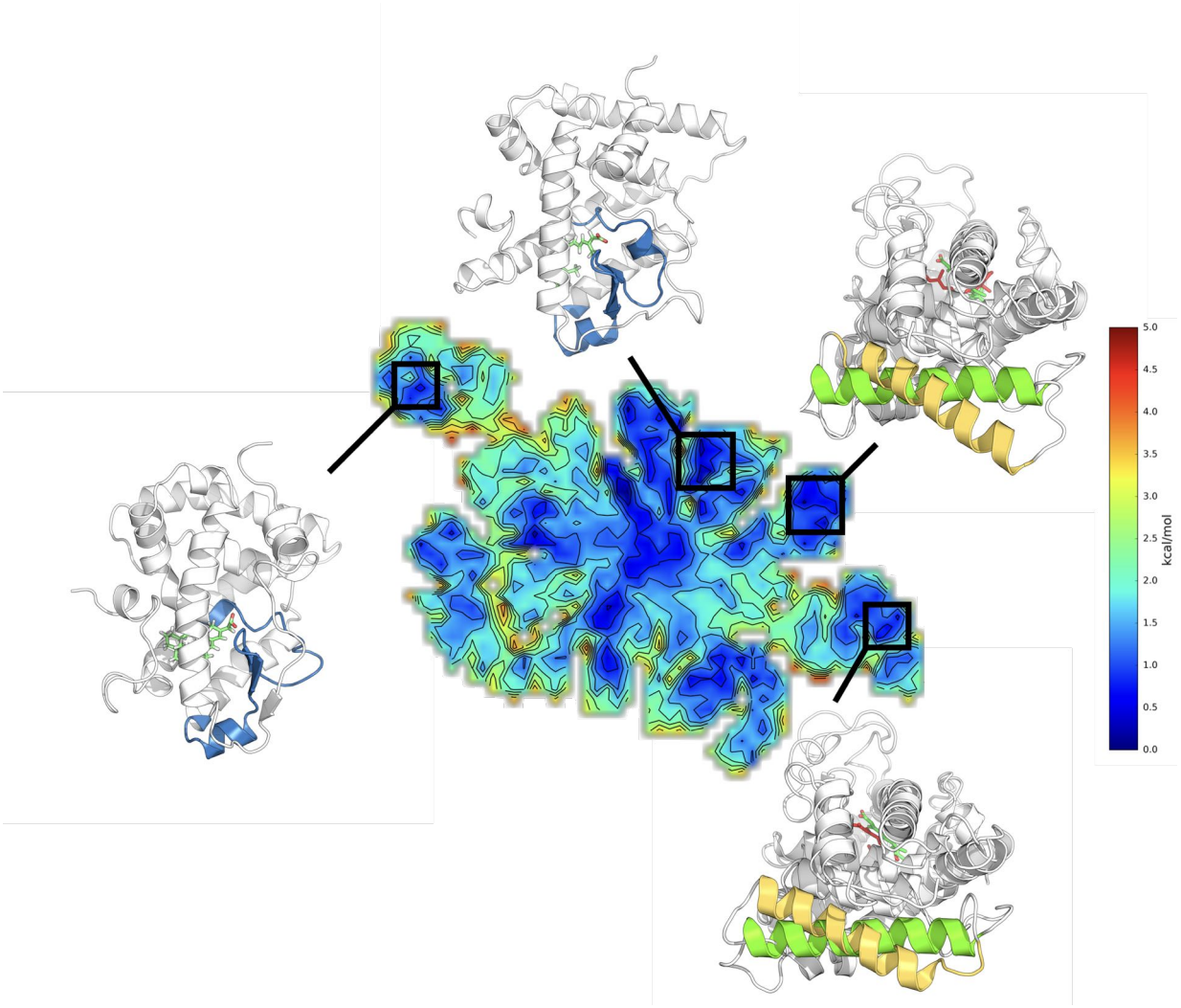


FIGURE 5.45 – Résumé des conformations où l'on observe une déstructuration partielle ou complète de H5 et un détachement de H9. Ces conformations ont été retrouvées au sein de certains minimas du paysage énergétique obtenus à partir des trajectoires aMD de la forme liée à 9cRA du LBD de RXRα. Bien que H5 soit complètement structurée dans le restant des minimas, car stabilisée par l'interaction du 9cRA avec l'Arg316, il est possible d'observer une déstructuration de H5. De même pour H9, qui reste généralement bien positionné dans le sandwich d'hélices, mais qui peut s'en détacher du coté N-terminal. Le domaine est représenté en cartoon avec le ligand 9cRA en représentation stick, avec en bleu H5, le feuillet β et H6, et en jaune H9 avec en vert H9 au sein de la structure cristallographique, le restant du domaine étant en blanc.

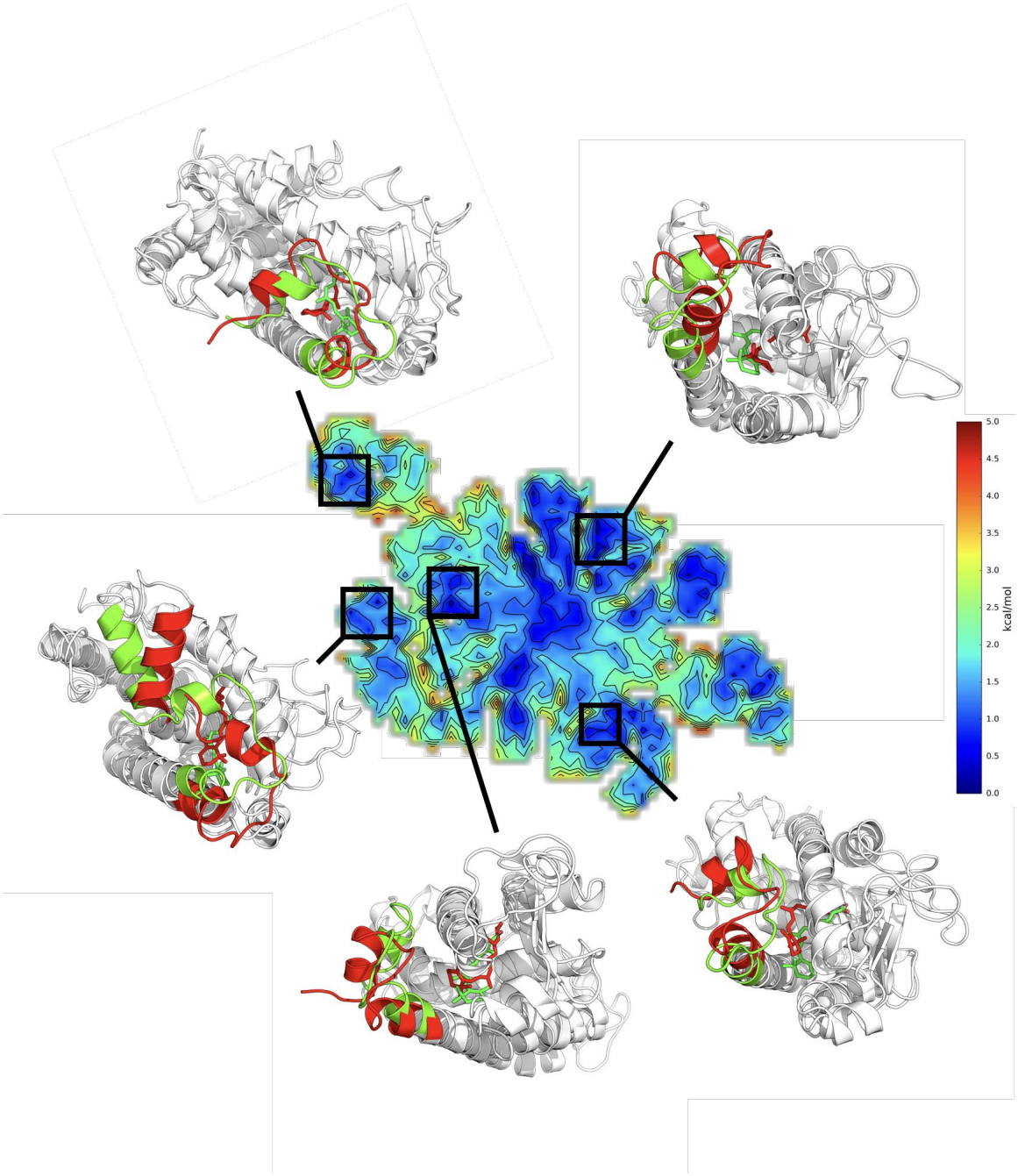


FIGURE 5.46 – Résumé des conformations où l'on peut observer un déplacement de H12 dû à un changement de position de 9cRA au sein de la poche de liaison au ligand. La partie carboxyle du RA interagissant toujours avec Arg316, c'est la position de la partie β-ionone de 9cRA qui fluctue le plus parmi les conformations retrouvées au sein des minimas. Se déplaçant vers H12 le plus souvent, le cycle β-ionone perturbe le positionnement de cette dernière, ainsi que de H4. Le domaine est représenté en cartoon avec le ligand 9cRA en représentation stick, avec en rouge la partie C-terminale du domaine ainsi que 9cRA (pour comparaison, leur position au sein de la structure cristallographique est représentée en vert), le restant du domaine étant en blanc.

Cependant, contrairement à la forme libre, l'ensemble conformationnel échantillonné est plus restreint ce qui est lié à la présence de 9cRA. Ainsi dans le cas de la forme liée, on n'observe pas de dépliement de l'hélice H3 ni du feuillet β au sein des principaux minimas (Figure 5.44). Concernant H2, elle est retrouvée dans certains minimas, mais reste malgré tout absente dans la majorité d'entre eux. Par contre, on peut observer la formation d'un 3e brin β parallèle entre la boucle H1-H3 et le feuillet β canonique, ce qui n'était pas observé dans les trajectoires de la forme libre. Ce type d'interaction est possible du fait du mouvement de bascule de la partie N-terminale de H3 lors de sont interaction avec 9cRA dans la forme liée (Figure 5.44). De plus, du fait de la présence de 9cRA, on constate une rigification de la partie C-terminale de H5. Ceci est dû à la forte interaction entre Arg316 et 9cRA, par le biais de liaisons hydrogène, qui stabilisent cette région du domaine. Cependant, cette interaction n'empêche pas le dépliement de la partie C-terminale de H5 au sein de certains minimas du paysage énergétique (Figure 5.45). Cependant, de manière générale, la forme liée du LBD de RXRα en complexe avec 9cRA est plus compacte que la forme libre.

Les principaux changements conformationnels observés lors de la trajectoire aMD se situent au niveau de H9 et H12. Par exemple, dans le cas de H9, on constate au sein de deux minimas proches dans le paysage énergétique, une rotation de H9 autour du résidu Ser399 d'environ 30° pour l'un et 40° pour l'autre. Dans le dernier cas, la forte rotation de l'hélice H9 entraine un décollement de cette dernière du côté N-terminal. On observe ainsi une perte d'interaction, résidus 386 à 399 avec H8. Néanmoins, les interactions avec H8 sont conservées du côté C-terminal, résidus 400 à 406 (Figure 5.45). En ce qui concerne H12, son positionnement est tributaire des mouvements de 9cRA au sein de la poche de liaison. En effet, le repositionnement du ligand peut entrainer un changement conformationel de H12 similaire à celui observé au cours des trajectoires aMD de la forme libre. De plus, son repositionnement va aussi avoir une influence sur la position de H4 à travers H5 et H12, tous deux en contact avec le ligand (Figure 5.46). Dans le cas où 9cRA est positionné de la même façon que dans la structure cristallographique, H12 et H4 sont de même positionnées dans des positions identiques à celles retrouvées retrouvées au sein de la structure cristallographique (Figure 5.46). De manière générale, les conformations retrouvées au sein des minimas sélectionnés montrent que la partie β-ionone de 9cRA est plus mobile au sein de la poche liaison au ligand que la fonction carboxyle, qui elle reste en contact avec Arg316.

5.4 Discussion

L'analyse des trajectoires aMD de la forme libre ainsi que de la forme liée à 9cRA du LBD de RXRα montre que la présence de 9cRA stabilise le LBD. Elle montre également que les plus grands changements conformationnels sont observés dans les éléments structuraux qui bordent la poche de liaison au ligand (LBP). Cette observation est en accord avec les données expérimentales HDX qui montrent une diminution globale de l'échange HD au niveau de la poche LBP et donc une stabilisation de la structure secondaire de la LBP de RXRα lors de la liaison au 9cRA. Elle est également en accord avec les données de déplacements chimiques, qui montrent l'apparition de pics de corrélation au niveau de cette même région (à l'exception de H10) lors de la liaison au 9cRA, et donc une diminution de l'échange conformationnel.

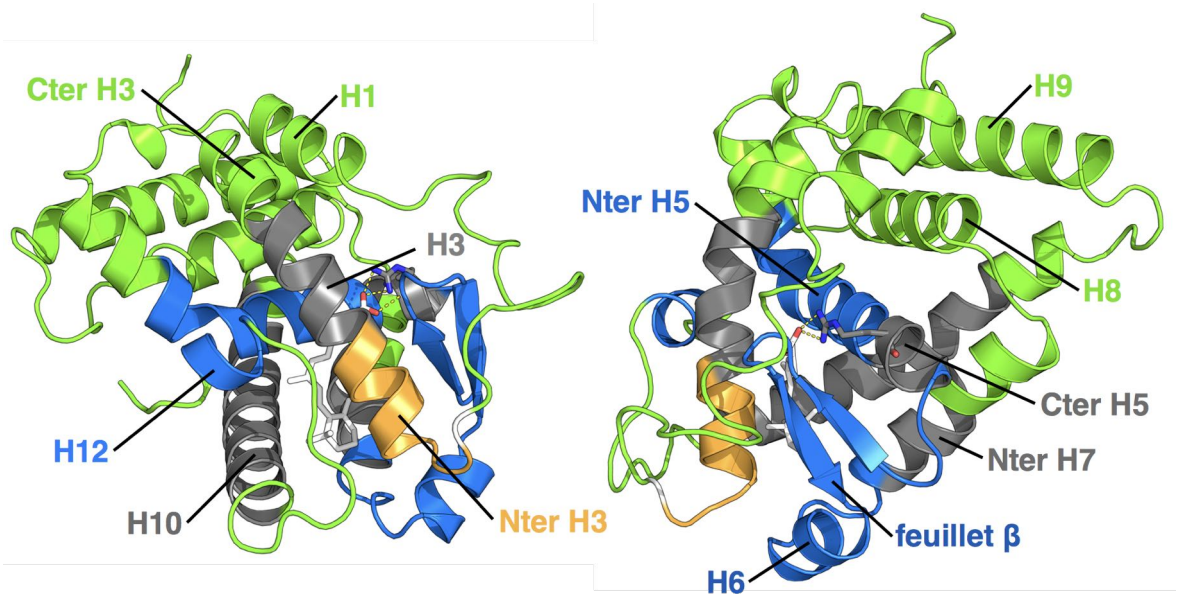


FIGURE 5.47 – Résumé de l'accord entre les données expérimentales HDX et RMN et les données prédites à partir des trajectoires aMD de la forme libre et liée à 9cRA du LBD de RXRα. En vert, les régions où l'on observe un bon accord entre les données expérimentales et les données prédites dans la forme libre et liée à 9cRA. En bleu, uniquement pour la forme libre et en orange, uniquement pour la forme liée à 9cRA. En gris, les régions présentant un désaccord entre les données expérimentales et prédites pour les deux formes et en blanc, les régions pour lesquelles aucune donnée expérimentale n'est disponible.

A notre connaissance, notre travail est le premier à analyser les données d'incorporation HDX d'un récepteur nucléaire et à les comparer aux données expérimentales, à la fois pour un récepteur libre et lié à un

ligand de haute affinité, soit le 9cRA. En outre, contrairement à ce qui a été utilisé dans certaines des études récentes d'incorporation HDX à partir de trajectoires de dynamique moléculaire [Radou et al., 2014] [Park et al., 2015], nous n'avons procédé à aucun ajustement visant à améliorer la corrélation entre les données expérimentales et calculées. La seule procédure de calibration a concerné la valeur maximale des facteurs de protection, où nous avons suivi la même procédure que celle utilisée pour calibrer les données expérimentales [Hamuro et al., 2006]. Nous avons vu que la prédiction de données HDX dépend, non seulement des ensembles conformationnels explorés, mais également du modèle choisi (simple ou concerté) pour déterminer les facteurs de protection. Il n'y a pas encore de consensus établi dans la littérature sur le meilleur modèle d'échange, et de ce fait, notre étude présentait de nombreux challenges. L'examen de l'ensemble des données montre que la procédure utilisant les trajectoires aMD et le modèle concerté permet une bonne prédiction des données HDX pour les formes libre et liée à 9cRA du LBD de RXRα, pour les régions H1, la boucle H1-H3, la partie C-terminale de H3, H4, la partie C-terminale de H7, H8, H9, la partie N-terminale de H10 et la boucle H11-H12. Il y a ensuite un ensemble d'éléments de structure secondaire telle que la partie C-terminale de H4, la partie N-terminale de H5, le feuillet β, H6 et H12 (voir analyse détaillée ci-dessous) où les prédictions pour la forme libre sont correctes, mais par contre où des difficultés sont rencontrées pour la forme liée à 9cRA. Une petite région, la partie N-terminale de H3, résidu 262 à 270, est mieux prédite dans la forme liée à 9cRA que la forme libre. Finalement, la région composée des hélices H7 et H10 qui forment l'interface de dimérisation présente des problèmes à la fois dans la forme libre et la forme liée à 9cRA.

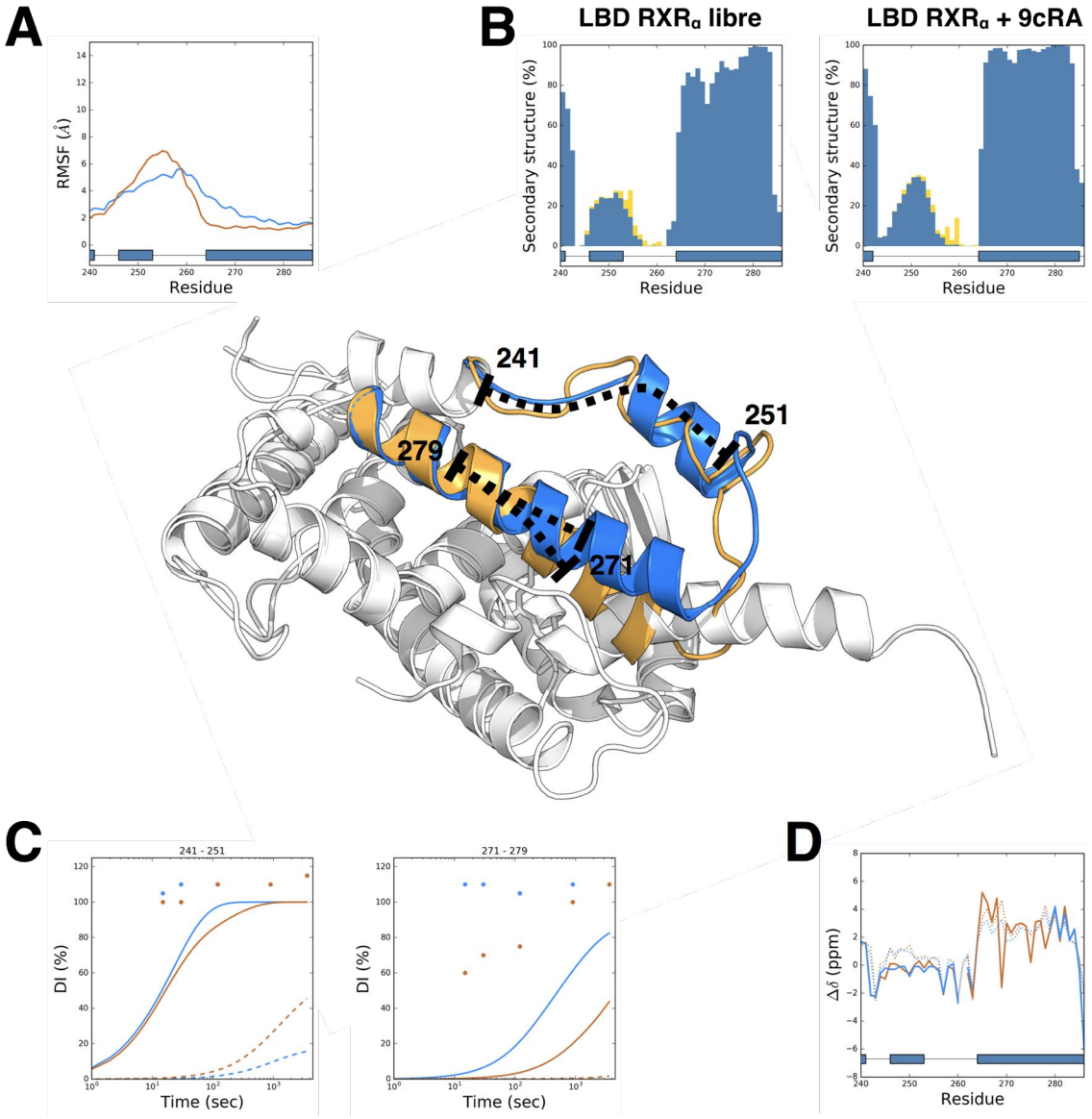


FIGURE 5.48 – Résumé des données expérimentales et prédites à partir des trajectoires de la forme libre et liée à 9cRA du LBD de RXRα pour la boucle H1-H3 et H3. (A) Fluctuations RMS par résidu calculées à partir des trajectoires aMD. (B) Proportion de structures secondaires par résidu calculées à partir des trajectoires aMD. (C) Pourcentage d'incorporation au cours du temps pour les peptides 241-251 et 271-279 prédit à partir des trajectoires aMD (traits pleins), cMD (traits pointillés) et expérimental (points). (D) Déplacements chimiques RMN secondaires $\Delta\delta^{13}C_\alpha$ par résidu expérimentaux (traits pleins) et prédits à partir des trajectoires aMD (traits pointillés). Les données concernant la forme libre et liée à 9cRA sont représentées en bleu et orange, respectivement. La boucle H1-H3 et H3 est représentée en bleu et orange pour la forme libre et liée à 9cRA, respectivement, sur les structures cristallographiques et en pointillés sont délimités les peptides pour lesquels l'échange HD a pu être mesuré.

Cependant, il reste des divergences entre les données calculées et expérimentales. Toutes les informations recueillies au niveau de la région H1-H3, H3 comprise, montrent une stabilisation de la structure secondaire, largement visible au niveau de la moitié N-terminale de H3, une diminution de la fluctuation RMS, une augmentation de la proportion en structure secondaire et de ce fait une augmentation des valeurs de déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα et une diminution de l'échange HD. Toutes ces diminutions sont prédites à partir des trajectoires aMD, mais quantitativement, l'incorporation en deutérium prédite à partir du modèle concerté reste quantitativement inférieure à celle observée expérimentalement. Plusieurs explications peuvent être envisagées : soit le modèle simple est plus approprié, mais il prédit également une incorporation plus faible que celle observée expérimentalement (bien que beaucoup plus proche des valeurs expérimentales dans cette région particulière), soit l'exploration des conformations par aMD n'est pas complète, et plus de conformations compétentes pour l'échange devraient être identifiées, soit finalement les données expérimentales surestiment l'incorporation dans cette région. A ce stade, il reste difficile de trancher entre ces trois hypothèses, et des analyses supplémentaires seront nécessaires. On peut cependant remarquer que les déplacements chimiques prédits à partir des trajectoires aMD, pour l'hélice H2, sont plus importants que les déplacements expérimentaux, qui sont clairement en pelote aléatoire (*random coil*). De même, au niveau du coude dans l'hélice H3, résidu 271 à 279, que l'on distingue lors des analyses de structure secondaire, l'accord entre les déplacements chimiques expérimentaux et calculés est un peu moins bon, quoique ces différences restent faibles, généralement dans la limite (1 ppm) d'erreur des prédictions. Ces observations pourraient indiquer que l'aMD n'a pas encore échantillonné suffisamment de conformations compétentes pour l'échange.

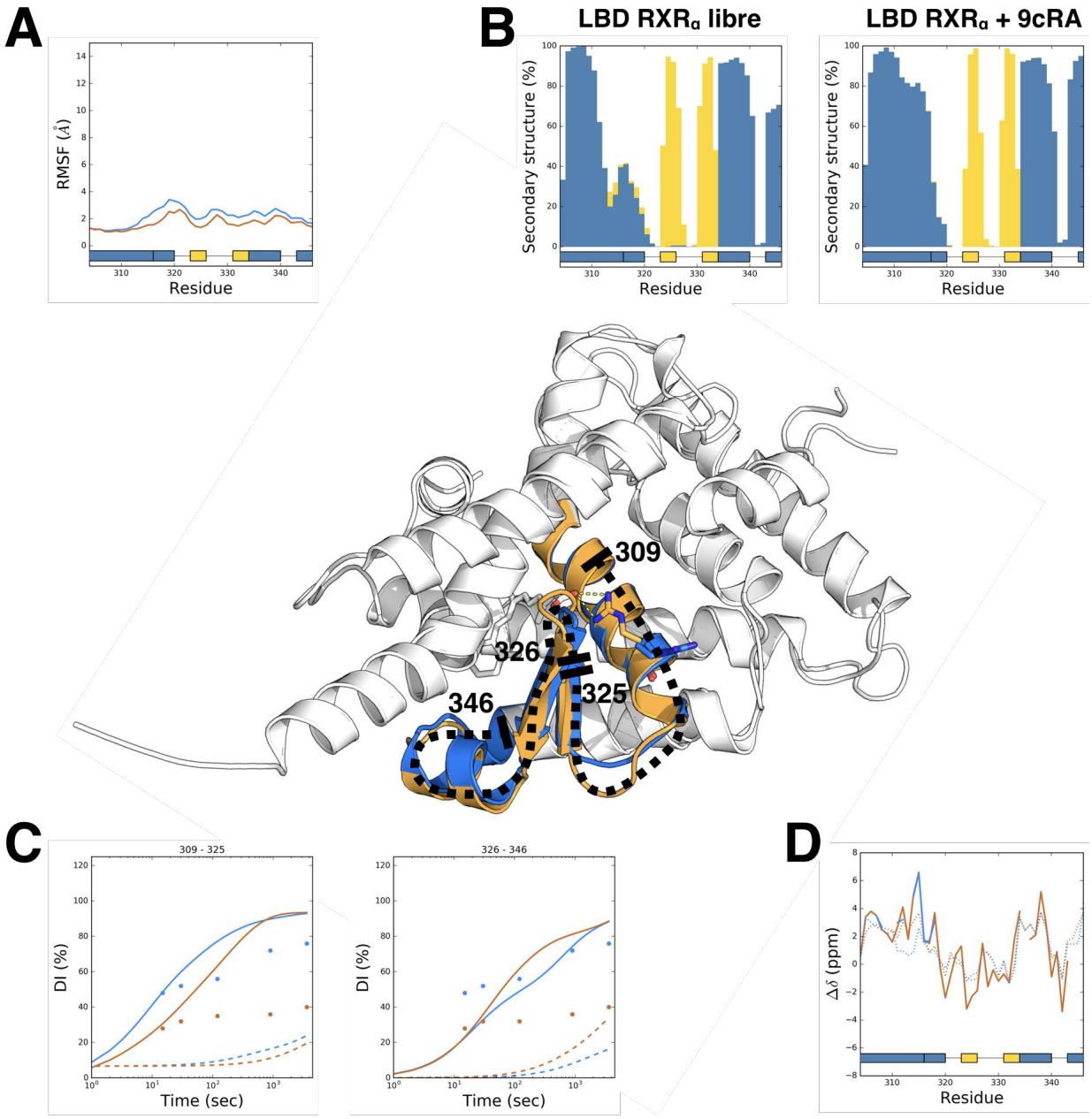


FIGURE 5.49 – Résumé des données expérimentales et prédites à partir des trajectoires de la forme libre et liée à 9cRA du LBD de RXRα pour la région du feuillet β. (A) Fluctuations RMS par résidu calculées à partir des trajectoires aMD. (B) Proportion de structures secondaires par résidu calculées à partir des trajectoires aMD. (C) Pourcentage d'incorporation au cours du temps pour les peptides 309-325 et 326-346 prédit à partir des trajectoires aMD (traits pleins), cMD (traits pointillés) et expérimental (points). (D) Déplacements chimiques RMN secondaires $\Delta\delta^{13}C_\alpha$ par résidu expérimentaux (traits pleins) et prédits à partir des trajectoires aMD (traits pointillés). Les données concernant la forme libre et liée à 9cRA sont représentées en bleu et orange, respectivement. La région du feuillet β est représentée en bleu et orange pour la forme libre et liée à 9cRA, respectivement, sur la structure cristallographique et en pointillés sont délimités les peptides pour lesquels l'échange HD a pu être mesuré.

La région du feuillet β, comprenant aussi H5 et H6 pose des problèmes particuliers. En effet, on observe une différence significative entre les données prédites et observées de l'échange HD, ainsi que des erreurs importantes (localisées sur quelques résidus) dans les déplacements chimiques calculés. L'incorporation en deutérium dans la forme apo est relativement bien prédite, mais, de manière très surprenante, l'incorporation calculée pour la forme liée au 9cRA ne diffère que très peu de celle calculée pour la forme apo. Ceci est en contradiction avec les données expérimentales, et également avec le nombre d'interactions formées entre le ligand et la protéine dans cette région, dont l'interaction électrostatique entre le groupement carboxylate du 9cRA et l'Arg316 de l'hélice H5 (Figure 1.7). Cependant, une stabilisation de cette région est bien observée dans les trajectoires aMD : on observe une diminution globale des fluctuations RMS, une stabilisation de la structure secondaire très visible au niveau de la partie C-terminale de H5 et visible de même au niveau des déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα. Cependant, certains déplacements chimiques sont mal prédits : en C-terminal de H5 (His415), et dans la région du feuillet β, Ile324, Leu325 et Leu326. Dans cette région, la trajectoire aMD semble donc surestimer la perte de structure secondaire. Bien que le champ de force du ligand ait été déterminé avec les procédures recommandées [Martin, 2008], il est possible que les interactions ligand-récepteur ne soient pas parfaitement balancées dans cette région, ce qui pourrait expliquer le désaccord observé (Figures 7.4 et 7.5). Une autre piste pourrait être le choix de la protonation de His315 pour laquelle le désaccord sur les déplacements chimiques est particulièrement important. Cependant, cette histidine ne pointe pas directement vers le ligand.

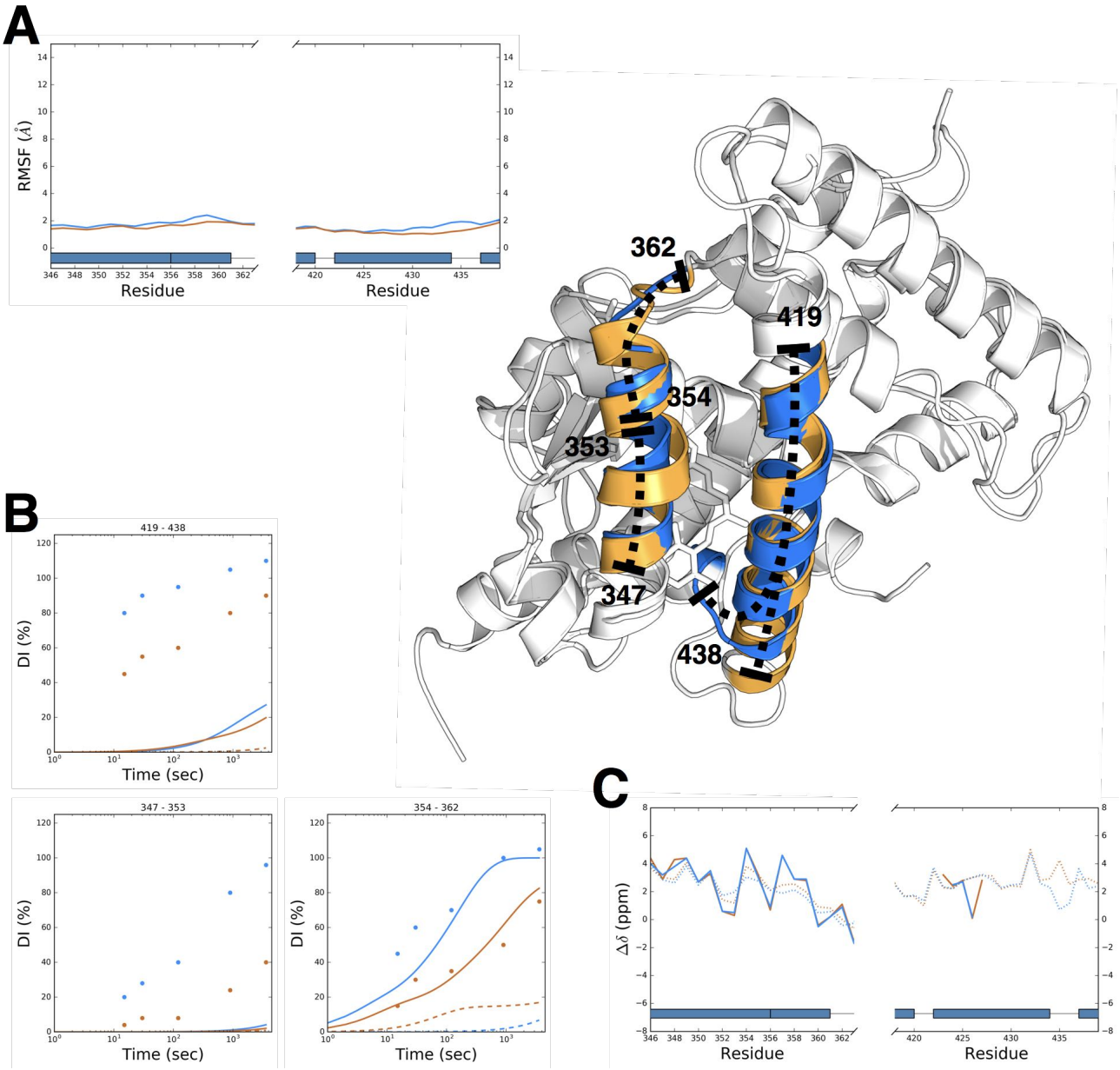


FIGURE 5.50 – Résumé des données expérimentales et prédites à partir des trajectoires de la forme libre et liée à 9cRA du LBD de RXRα pour les régions H7 et H10-H11. (A) Fluctuations RMS par résidu calculées à partir des trajectoires aMD. (B) Pourcentage d'incorporation au cours du temps pour les peptides 347-353, 354-362 et 419-438 prédit à partir des trajectoires aMD (traits pleins), cMD (traits pointillés) et expérimental (points). (C) Déplacements chimiques RMN secondaires $\Delta\delta^{13}C_\alpha$ par résidu expérimentaux (traits pleins) et prédits à partir des trajectoires aMD (traits pointillés). Les données concernant la forme libre et liée à 9cRA sont représentées en bleu et orange, respectivement. Les régions H7 et H10-H11 sont représentées en bleu et orange pour la forme libre et liée à 9cRA, respectivement, sur la structure cristallographique et en pointillés sont délimités les peptides pour lesquelles l'échange HDX a pu être mesuré.

En ce qui concerne l'interface de dimérisation, composée principalement de H10, mais aussi de H7, H9 et la boucle H8-H9, on remarque une faible dynamique des hélices H7 et H10, que ce soit au sein des trajectoires aMD de la forme libre ou liée à 9cRA du LBD de RXRα. On remarque que les parties C-terminales de H7 et H10 sont plus structurées pour la forme liée à 9cRA, avec des valeurs de déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα plus élevées et des fluctuations RMS moyennes moins importantes. Néanmoins, pour le restant des hélices H7 et H10, aucune différence n'est visible entre la forme libre et liée à 9cRA. Dans les deux cas, l'échange H/D est très nettement inférieur à celui observé expérimentalement au niveau des peptides 347-353 et 419-438. L'échange H/D expérimental montre que l'hélice H7 et H10 sont moins stables dans la forme libre que dans la forme liée avec une vitesse d'échange H/D plus rapide. L'existence d'une incorporation expérimentale importante en deutérium dans l'interface de dimérisation, alors que les données HDX ont été collectées dans des conditions où la forme homodimérique du récepteur devrait dominer, montre qu'il y a un échange rapide homodimère/monomère et que la présence de 9cRA déplace l'équilibre vers la forme homodimérique. Ceci suggère que les hélices H7 et H10 pourraient être déstructurées dans la forme monomérique. Par contre, lors des trajectoires aMD, la structure secondaire (Figures 5.7 et 5.10) est conservée dans ces deux régions (voir cependant chapitre « Effet de la phosphorylation pSer260 »). Il se peut donc que les trajectoires aMD n'aient pas suffisamment échantillonné de conformations compétentes pour l'échange dans cette région particulière. En ce qui concerne H7, on peut tout de même remarquer que pour la région (peptide 354-362) qui ne participe pas à l'interface de dimérisation, les prédictions HDX sont correctes et que les déplacements chimiques expérimentaux correspondent à une structure en hélice, et sont bien prédits pour H7. Par contre pour H10, à l'exception des résidus 428 à 437, l'attribution n'a été possible ni dans la forme libre ni dans la forme liée à 9cRA, ce qui tend à confirmer l'hypothèse un équilibre conformationnel.

Enfin, dans le cas de la partie C-terminale du domaine, on observe un bon accord entre les données prédites à partir des trajectoires aMD de la forme libre et de la forme liée à 9cRA. Cependant, on remarque que la vitesse d'échange H/D au niveau de H12 dans la forme liée à 9cRA, peptide 450-454, est bien inférieure à celle observée expérimentalement suggérant que H12 est moins bien structurée que ce qui est observé au cours des trajectoires aMD.

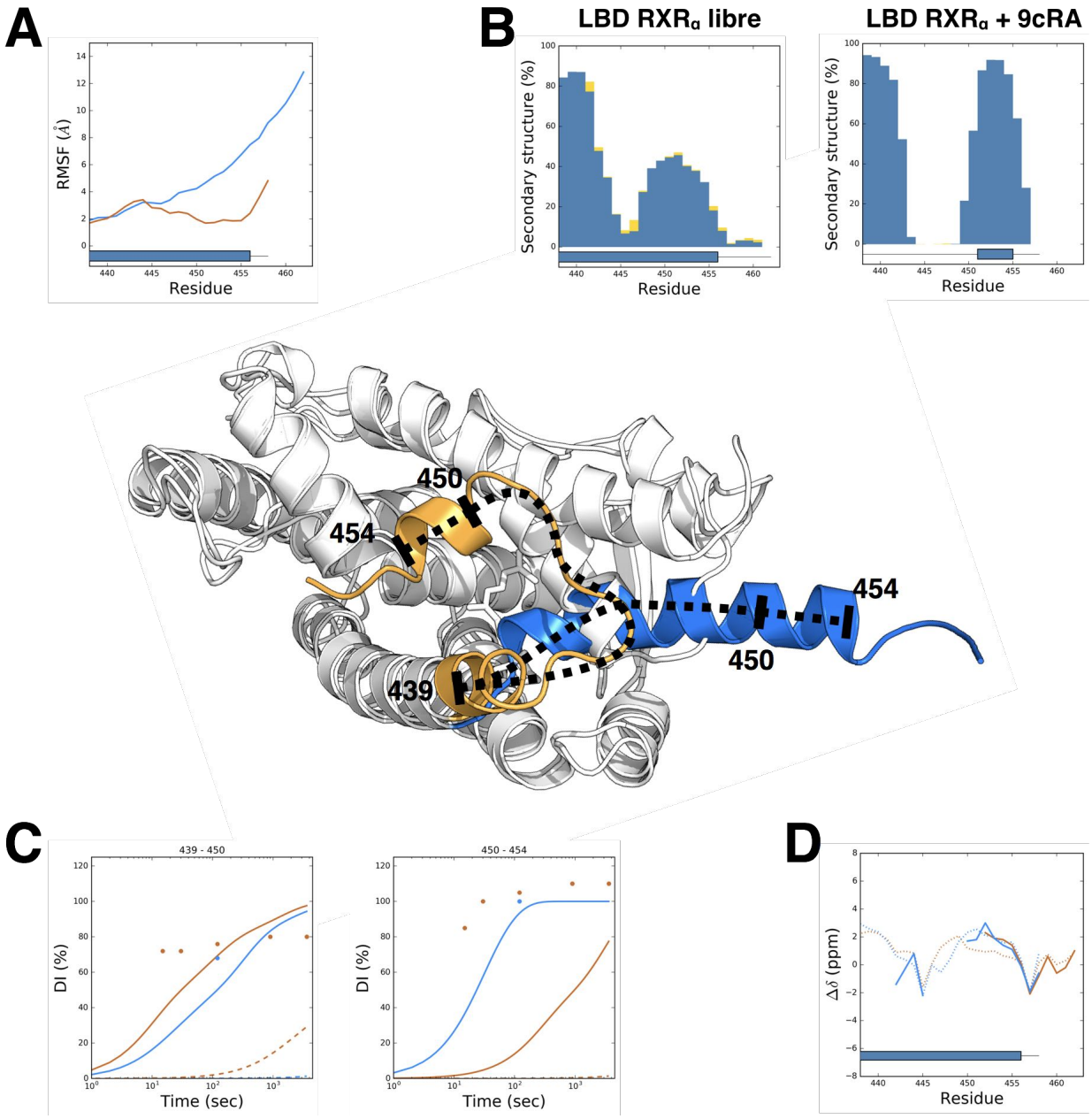


FIGURE 5.51 – Résumé des données expérimentales et prédites à partir des trajectoires de la forme libre et liée à 9cRA du LBD de RXRα pour la partie C-terminale. (A) Fluctuations RMS par résidu calculées à partir des trajectoires aMD. (B) Proportion de structures secondaires par résidu calculées à partir des trajectoires aMD. (C) Pourcentage d'incorporation au cours du temps pour les peptides 438-450 et 450-454 prédit à partir des trajectoires aMD (traits pleins), cMD (traits pointillés) et expérimental (points). (D) Déplacements chimiques RMN secondaires $\Delta\delta^{13}C_\alpha$ par résidu expérimentaux (traits pleins) et prédits à partir des trajectoires aMD (traits pointillés). Les données concernant la forme libre et liée à 9cRA sont représentées en bleu et orange, respectivement. La partie C-terminale sont représentées en bleu et orange pour la forme libre et liée à 9cRA, respectivement, sur la structure cristallographique et en pointillés sont délimités les peptides pour lesquelles l'échange HD a pu être mesuré expérimentalement.

Chapitre 6

Effet de la phosphorylation pSer260

6.1 Introduction

Ce chapitre décrit les résultats obtenus par dynamique moléculaire accélérée afin d'étudier la dynamique conformationnelle de la forme libre du LBD de RXRα phosphorylée au niveau de la sérine 260. L'objectif fut de comprendre l'influence de cette modification sur la dynamique de ce domaine en utilisant le même protocole que précédemment ainsi que les résultats obtenus pour la forme libre non phosphorylée en comparaison.

6.2 Matériels et Méthodes

6.2.1 Protocole de simulation aMD

Le protocole de simulation aMD utilisé dans le cadre de l'étude de l'influence de la phosphorylation de la sérine 260 sur la forme libre du LBD de RXRα est basé sur celui utilisé précédemment pour les formes non phosphorylées. Le placement des atomes d'hydrogène a été réalisé à l'aide du module HBUILD contenu dans le programme CHARMM. Les correctifs NH_3^+ (NTER) et COO^- (CTER) sont appliqués aux extrémités N- et C-terminale. En ce qui concerne l'état de protonation des histidines, ceux-ci ont été repris du protocole de simulation aMD précédent sur la forme non phosphorylée. Dans le cas de l'état de protonation du groupement phosphate sur le résidu SER260, la charge a été fixée à -1, étant donné que la proportion en [$H_2PO_4^{-1}$] et [HPO_4^{-2}] est identique à pH physiologique (entre 7 et 7.4). Pour cette étude, la structure cristallographique de la forme libre du LBD de RXRα a été utilisée, dont la SER260 a été modifiée avec le PATCH SEP, correspondant à une charge -1, dans le champ de force tout atome CHARMM. La topologie et les paramètres sont issus des travaux de thèse menés sur l'étude de l'effet de phosphorylation chez RARγ [Amal, 2013].

Les simulations de dynamique moléculaire accélérée de la forme libre du LBD de RXRα pSer260 ont été réalisées en solvant explicite TIP3P avec le programme NAMD (version 2.10 GPU). Le système a été décrit selon le champ de force tout atome CHARMM 27 avec la correction CMAP. De même, des ions K^+/Cl^- ont été ajoutés afin de neutraliser les charges du système et aussi d'atteindre une concentration en sels de 0.05 M. La simulation a été réalisée sous conditions périodiques et les interactions électrostatiques à longue distance ont été traitées à l'aide du maillage particulier d'Ewald (PME). Tous les liens entre les atomes d'hydrogène et les atomes lourds ont été contraints en utilisant l'algorithme SHAKE, et un paramètre de 12 Å a été appliqué aux interactions van der Waals. Un pas d'intégration de 1 fs a été utilisé pour toutes les simulations. Les paramètres E et α, nécessaires pour déterminer le biais énergétique qui sera

appliqué à l'énergie potentielle et des angles dièdres à chaque pas d'intégration, ont été obtenus préalablement à partir de trajectoires cMD de 10 ns. Ces trajectoires classiques ont été utilisées comme point de départ pour les trajectoires aMD. Celles-ci ont été produites à une température de 298.15 K en condition NVE. L'ensemble des trajectoires cMD et aMD générés est résumé dans le tableau suivant :

	N_{res}	N_{atoms}	cMD	aMD	ΔV_{avg}	$\sigma_{\Delta V}$	$\delta_{\Delta V}$
Forme libre du LBD de RXRα pSer260	238	34587	6 x 10 ns	900 ns (6 x 150 ns)	291.79	32.71	262.25

TABLE 6.1 – Résumé des trajectoires aMD de la forme libre du LBD de RXRα pSer260 avec N_{res} et N_{atomes}, le nombre de résidu et d'atomes, respectivement, puis ΔV_{avg}, $\sigma_{\Delta V}$ et $\delta_{\Delta V}$, la moyenne, l'écart-type et le rang (maximum - minimum) du biais d'énergie ΔV en kcal/mol, respectivement, sur l'ensemble des trajectoires aMD.

6.2.2 Méthodes d'analyse des trajectoires aMD

6.2.2.1 RMSD et Fluctuations RMS

Les déviations RMS au cours du temps ainsi que les fluctuations RMS par résidu ont été obtenues en alignant les atomes du squelette peptidique appartenant aux régions structurales les plus stables d'après les données expérimentales HDX, c'est-à-dire le sandwich d'hélices α composé de H1 (résidus 235 à 240), partie C-terminal de H3 (279 à 285), H5 (303 à 313), H7 (349 à 351), H8 (371 à 375), H9 (391 à 400) et H10 (416 à 425) du LBD de RXRα. La déviation RMS ainsi que la fluctuations RMS ont été calculées à l'aide du module Python MDAnalysis [Michaud-Agrawal et al., 2011].

6.2.2.2 Structure secondaire

La structure secondaire a été obtenue à l'aide de DSSP [Kabsch and Sander, 1983]. Les conformations ont été extraites toutes les 10 picosecondes à partir des trajectoires aMD de la forme libre phosphorylée pSer260 du LBD de RXRα. Ce sont au total 90.000 conformations qui ont été analysées.

6.2.2.3 Déplacements chimiques RMN

Les déplacements chimiques RMN ont été prédits en utilisant le programme SHIFTX+ issu de SHIFTX2 en utilisant un pH égal à 7.4 et une température égale de 298.15 K [Han et al., 2011]. Les conformations ont été extraites toutes les 10 picosecondes à partir des trajectoires aMD de la forme libre phosphorylée pSer260 du LBD de RXRα. Au total, ce sont 90.000 conformations qui ont été analysées.

L'ensemble des scripts Python ayant été utilisés pour calculer les déplacements chimiques à partir des conformations extraites des trajectoires MD, ainsi que pour comparer les déplacements chimiques prédits δ_{pred} aux déplacements chimiques expérimentaux δ_{exp} sont disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/shift`

6.2.2.4 Échange Hydrogène/Deutérium

L'échange Hydrogène/Deutérium a été calculé à partir des constantes de vitesses intrinsèques k_{int} par proton amide HN obtenues à l'aide du modèle empirique développé par Bai *et al.* [Bai et al., 1993], ainsi qu'à partir des facteurs de protection PF de chaque proton amide HN obtenu à partir des trajectoires aMD de la forme libre phosphorylée pSer260 du LBD de RXRα. Les facteurs de protections PF ont été obtenus du nombre de liaisons hydrogène établies entre chaque proton HN et le solvant selon le modèle simple et le modèle concerté, à l'aide du module python MDAnalysis [Michaud-Agrawal et al., 2011]. Au total, ce sont 900.000 conformations qui ont été analysées.

L'ensemble des scripts Python pour déterminer les facteurs de protection PF à partir des trajectoires MD ainsi que prédire l'incorporation en deutérium au cours du temps est disponible librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/hdx`

6.2.2.5 Stochastic Proximity Embedding

Un très grand nombre de conformations ont été générées durant les simulations aMD de la forme libre phosphorylée pSer260 du LBD de RXRα, avec au total 9 10^5 conformations. Il a été ainsi nécessaire de les organiser afin de pouvoir analyser l'espace conformationnel échantillonné. Pour cela, nous avons utilisé l'algorithme pSPE qui est une méthode non linéaire de réduction de la dimensionnalité des données, en utilisant la distance dihédrale comme métrique, au lieu du RMSD (voir chapitre 3 pour plus de détails).

Ainsi, tous les pseudo angles dièdres entre 4 Cα consécutifs ont été extraits de chaque conformation présente dans les trajectoires aMD à l'aide du module Python MDAnalysis [Michaud-Agrawal et al., 2011]. Au total, 235 pseudo angles dièdres ont été extraits pour la forme libre phosphorylée pSer260 du LBD de

RXRα, respectivement. Ainsi que détaillé dans le chapitre 3, la méthode pSPE utilise le paramètre r_c, rayon de proximité utilisé pour déterminer la région où la distance euclidienne est une bonne approximation (*cutoff*) de la distance géodésique. Il n'existe pas de règle permettant de fixer à priori la valeur du paramètre r_c menant à la meilleure réduction de dimensionnalité. L'approche empirique que nous avons implémentée (voir Chapitre 3) est de tester, systématiquement les valeurs de r_c entre 0.1 et 0.5 par intervalles de 0.01. Ces tests sont effectués sur 10 000 structures extraites des trajectoires, pour un nombre fixe (5000) de cycles d'optimisation, et ils sont répétés cinq fois pour chaque valeur du paramètre r_c. L'examen de l'évolution du stress, et de la corrélation γ calculés pour chaque valeur de r_c (voir Chapitre 3) permet de déterminer empiriquement la valeur de r_c qui offre le meilleur compromis entre une valeur maximale de la corrélation γ et minimale du stress. Une fois la valeur optimale de r_c déterminée de cette manière, un certain nombre de tests supplémentaires sont réalisés afin d'assurer de la robustesse de la procédure. Ainsi, le nombre de cycles d'optimisation C est systématiquement modifié, de 10 cycles à 100 000 cycles (valeurs intermédiaires : 50, 100, 500, 1000, 5000, 10 000, 50 000) et ce pour des nombres différents de conformations (10 000, 50 000 et 100 000). Chacun de ces tests est répété cinq fois vu le caractère stochastique de l'algorithme pSPE. Ceci nous permet d'estimer le nombre minimal de cycles d'optimisation C nécessaire pour obtenir la meilleure réduction de dimensionnalité, avec le r_c choisi. L'utilisation de différents échantillons de conformations permet également de tester la stabilité et la vitesse de convergence de la méthode. Une fois l'ensemble de ces tests réalisés, la méthode est appliquée sur l'ensemble des 900 000 conformations.

L'ensemble des scripts Python pour extraire les angles dièdres à partir de trajectoires MD, déterminer les valeurs optimales des paramètres r_c et C, ainsi que la méthode pSPE sont disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/unrolr`

6.2.2.6 Repondération de l'énergie des trajectoires aMD

Afin de repondérer l'énergie des conformations issues des trajectoires aMD de la forme phosphorylée, nous avons effectué une analyse similaire à celle réalisée sur les trajectoires aMD de la forme libre non phosphorylée. Des coordonnées collectives ont été obtenues à l'aide de la méthode pSPE, puis utilisées dans la repondération des conformations, en repondérant avec la série de McLaurin d'ordre 10 avec une température de 298.15 K et un intervalle (*bin size*) de 0.05.

L'ensemble des scripts Python ayant servi à repondérer l'énergie des trajectoires aMD sont aussi disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/reweight`

6.2.2.7 Visualisation des conformations

Pour permettre de visualiser l'espace conformationnel échantillonné durant les simulations, le module Python Bokeh [Bokeh Development Team, 2014] et PyMOL [Schrödinger, LLC, 2015] ont été utilisés. Ainsi, à partir de variables collectives quelconques, provenant soit de la méthode pSPE ou depuis n'importe quelle méthode d'apprentissage automatique, telle que la PCA, il est possible de visualiser les conformations générées. De même, l'ensemble des scripts Python sont disponibles librement sous licence MIT à cette adresse : `https://github.com/jeeberhardt/visualize`

6.3 Résultats

6.3.1 Analyse des trajectoires aMD

6.3.1.1 RMSD global

Le RMSD global de la forme phosphorylée pSer260 augmente rapidement dès le début des simulations aMD, ce qui est cohérent avec le comportement de la forme libre du LBD de RXRα non phosphorylée (Figure 6.1).

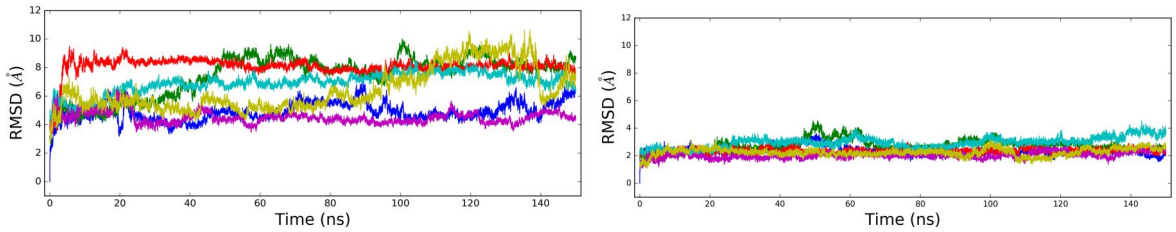


FIGURE 6.1 – RMSD global du squelette peptidique au cours du temps, calculé pour chaque trajectoire aMD de la forme libre phosphorylée pSer260 du LBD de RXRα. Le RMSD a été calculé en alignant toutes les structures issues des trajectoires aMD sur la structure cristallographique de référence (PDB id : 1LBD) en ne considérant que la partie la plus stable, c'est-à-dire le sandwich d'hélices α composé de H1 (résidus 235 à 240), partie C-terminal de H3 (279 à 285), H5 (303 à 313), H7 (349 à 351), H8 (371 à 375), H9 (391 à 400) et H10 (416 à 425). En haut, le RMSD de l'ensemble du domaine au cours du temps et en bas, le RMSD du sandwich d'hélices α seul.

Le RMSD global oscille entre 4 et 10 Å, par rapport à la structure cristallographique non phosphorylée de référence, selon les trajectoires aMD. En ce qui concerne le sandwich d'hélices α, le RMSD global

oscille aussi aux alentours de 2.5 Å au cours du temps, excepté pour quelques trajectoires aMD où l'on observe des pics à 4Å (Figure 6.1).

6.3.1.2 Fluctuation RMS

Quant aux fluctuations RMS par résidu, obtenues à partir des trajectoires aMD de la forme phosphorylée du LBD de RXRα, elles montrent que les régions fluctuant le plus sont les mêmes que dans les trajectoires de la forme non phosphorylée (Figure 6.2).

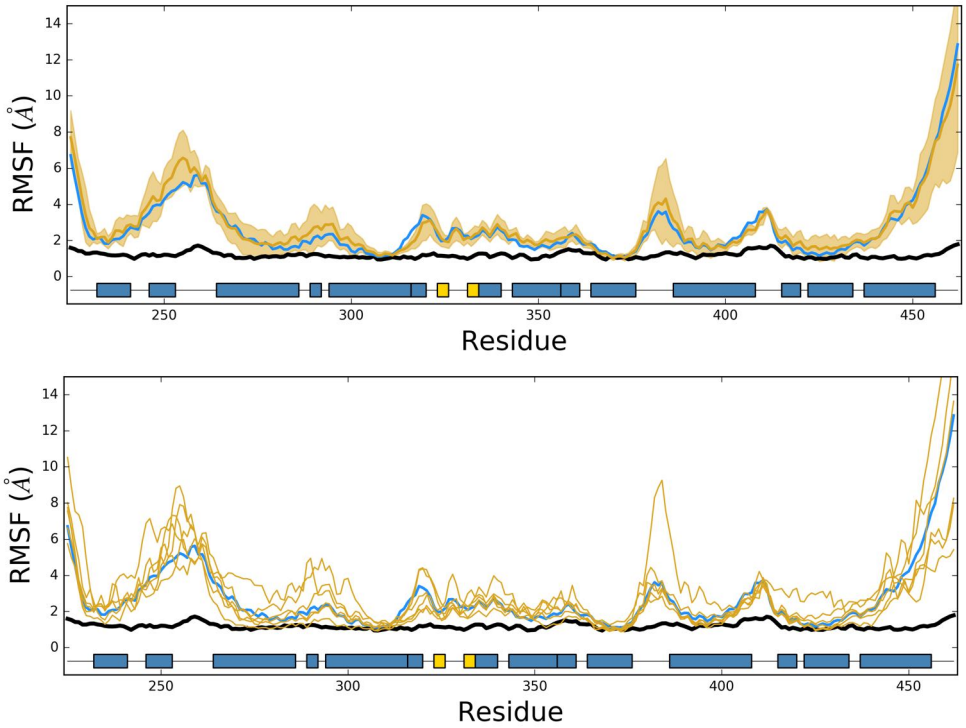


FIGURE 6.2 – Fluctuations RMS par résidu du squelette peptidique, moyennées sur l'ensemble des trajectoires aMD (en haut) ainsi que le détail par trajectoire aMD (en bas) de la forme libre phosphorylée pSer260 du LBD de RXRα. En jaune, les fluctuations RMS par résidu de la forme libre phosphorylée pSer260, en bleu les fluctuations de la forme libre non phosphorylée et en noir les fluctuations RMS calculées à partir des facteurs B de la structure cristallographique de référence de la forme libre du LBD de RXRα (PDB id : 1LBD).

Nous retrouvons la région N-terminale, la boucle entre H1 et H3 ainsi que la moitié de H3 du côté N-terminal, la boucle entre H3 et H4, H4 comprise, la région du feuillet β et H6, les boucles H8 et H9

avec la partie N-terminale de H9, la boucle H9-H10 et pour finir toute la partie C-terminale avec H12 comprise. Globalement, l'examen des fluctuations RMS indique que l'ensemble conformationnel exploré par les trajectoires aMD dans la forme phosphorylée diffère de celui exploré dans la forme libre non phosphorylée. Certaines différences se reproduisent sur un ensemble de trajectoires, et mènent à des fluctuations plus importantes (région allant du C-terminal de H3 au C-terminal de H4) ou moins importants (partie C-terminale de H5 et région du feuillet β) dans la forme phosphorylée que dans la forme libre non phosphorylée. Cependant, vu la nature modifiée de la surface d'énergie potentielle explorée lors des trajectoires aMD, des analyses complémentaires seront nécessaires pour conclure de manière définitive sur l'effet de la phosphorylation sur l'amplitude des fluctuations.

6.3.1.3 Structure secondaire

L'analyse de la structure secondaire à l'aide de DSSP des trajectoires aMD de la forme libre pSer260 du LBD de RXRα montre de même que la structure secondaire est elle aussi globalement conservée. Les hélices qui composent le sandwich d'hélices α restent en hélice plus de 80 % du temps au cours des trajectoires aMD, de même que le feuillet β (Figure 6.3).

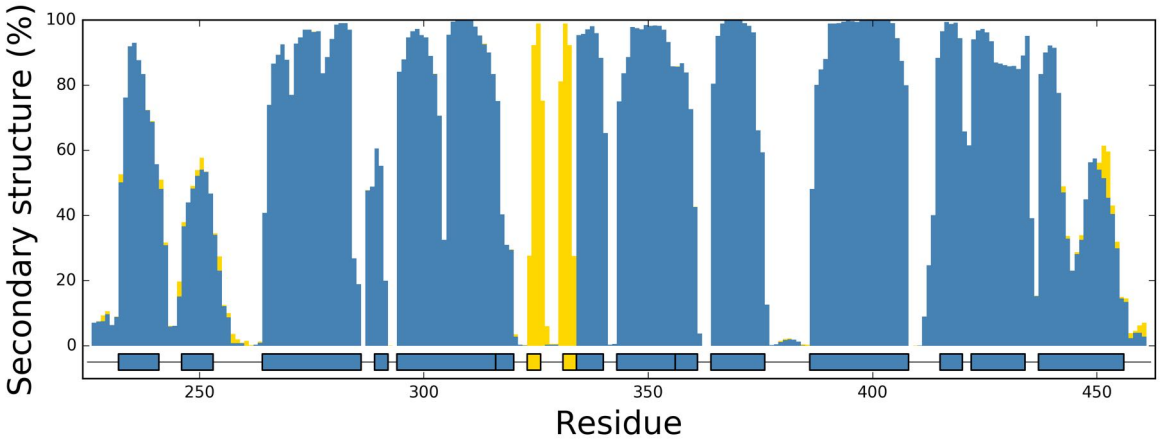


FIGURE 6.3 – Proportion de la structure secondaire (en %) par résidu, calculé à l'aide de DSSP à partir des trajectoires aMD de la forme libre phosphorylée pSer260 du LBD de RXRα. En bleu sont représentés les résidus en hélice (α, π et 3_{10}) et en jaune les résidus formant un feuillet ou un pont β.

Cependant, on remarque tout de même une légère déstructuration au niveau de H1 et de H10. Concernant le reste du domaine, d'importantes différences sont observées dans la boucle H1-H3 et au niveau de

la partie C-terminale de H5 (Figure 6.3). Dans la boucle H1-H3, on remarque une propension à rester en hélice plus importante pour H2 (> 40 %) comparée aux formes non phosphorylées (< 30 %). Quant à la partie C-terminale de H5, celle-ci est tout aussi structurée en hélice qu'au sein de la forme liée à 9cRA non phosphorylée (proche de 80 %), comparé à la forme libre non phosphorylée (< 40 %).

6.3.2 Analyse des déplacements chimiques RMN

Nous n'avons pas de données expérimentales concernant les déplacements chimiques de la forme phosphorylée de RXRα, cependant, la comparaison entre les déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα prédits à partir des trajectoires aMD de la forme phosphorylée et de la forme libre non phosphorylée sont un indicateur des différences entre les ensembles conformationnel non phosphorylés et phosphorylés. D'une manière générale, il y a une bonne correspondance entre les déplacements chimiques des deux formes (Figure 6.4).

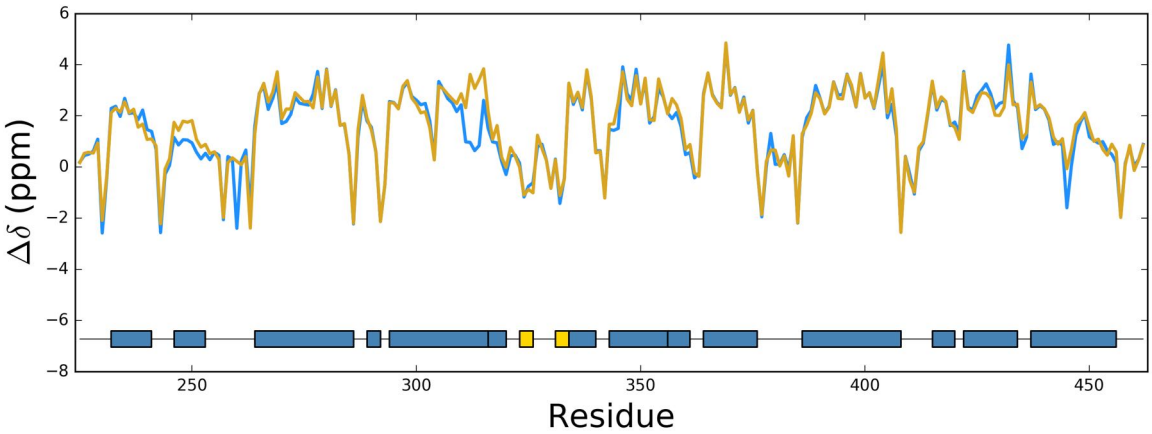


FIGURE 6.4 – Comparaison entre les déplacements chimiques secondaires $\Delta\delta_{\text{exp}}$ ^{13}Cα prédits $\Delta\delta_{\text{aMD}}$ ^{13}Cα, en jaune, à partir des trajectoires aMD de la forme libre phosphorylée pSer260 et, en bleu, de la forme libre non phosphorylée du LBD de RXRα.

Néanmoins, des divergences sont constatées, et particulièrement au niveau de la boucle H1-H3 et au niveau de la partie C-terminale de H5. Concernant la boucle H1-H3, les déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα sont plus élevées que ceux retrouvés au sein de la forme non phosphorylée. Ceci indique que cette région est plus structurée en hélice α dans la forme phosphorylée pSer260. Un comportement similaire est observé dans la région C-terminale de H5, contenant l'arginine 316, qui possède des valeurs

$\Delta\delta^{13}$Cα plus élevées, montrant là aussi une hélicité plus importante. Cette observation est cohérente avec les données des fluctuations RMS.

6.3.3 Analyse de l'Échange Hydrogène/Deutérium

Il n'existe pas de données expérimentales d'échange hydrogène-deutérium pour la forme phosphorylée du LBD de RXRα. Cependant, un mutant Ser260Glu est disponible et des expériences HDX pourraient donc être réalisées à terme. Nous avons donc analysé les liaisons hydrogène formées entre les protons amides HN et l'eau dans les simulations de la forme phosphorylée et calculé des facteurs de protection associés. Toutes les analyses présentées utilisent le modèle concerté de liaison hydrogène (voir Matériel et Méthodes). Ceci nous a permis d'identifier des régions qui pourraient échanger de manière différente entre la forme phosphorylée et la forme libre non phosphorylée (Figure 6.5).

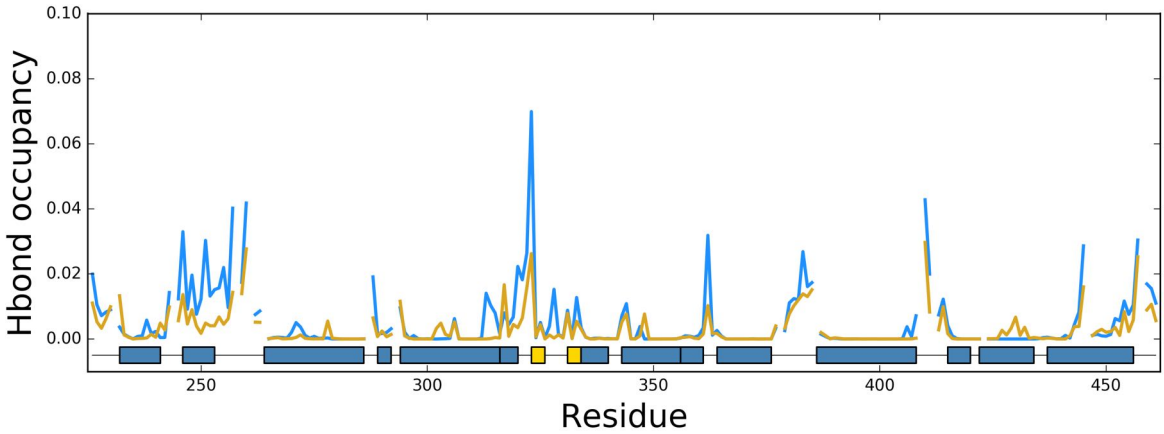


FIGURE 6.5 – Taux de compétence des protons amides HN selon le modèle concerté. En bleu et en jaune, sont représentées les données obtenues à partir des trajectoires aMD de la forme libre du LBD de RXRα non phosphorylée et phosphorylée pSer260, respectivement.

Certaines régions apparaissent plus protégées contre l'échange, elles se situent principalement au niveau de la boucle H1-H3, avec un nombre plus réduit de liaisons hydrogène formées entre les protons amides HN et l'eau. On constate aussi une réduction au niveau de la partie C-terminale de H5 et du feuillet β, de la boucle H7/H8, H8/H9. Au contraire, on remarque de légères augmentations des liaisons formées avec l'eau au niveau de la jointure H4/H5 ainsi qu'au niveau de H9.

Ces observations se traduisent donc par des changements au niveau des facteurs de protection PF

calculés. L'augmentation du nombre de liaisons hydrogène formées dans la forme phosphorylée pSer260 par rapport à la forme non phosphorylée s'accompagnera d'un facteur de protection PF moins élevé, et vice et versa. Ainsi, sont plus élevés les facteurs de protection PF au niveau de la boucle H1-H3, de la partie N-terminale de H3, de H4, ainsi qu'au niveau de la partie C-terminale de H5 et du feuillet β. Par contre, ils sont moins élevés au niveau de la jointure H4/H5, et de manière significative au niveau de H9, en passant de facteurs de protection PF proches de log(6) à moins de log(3) (Figure 6.6).

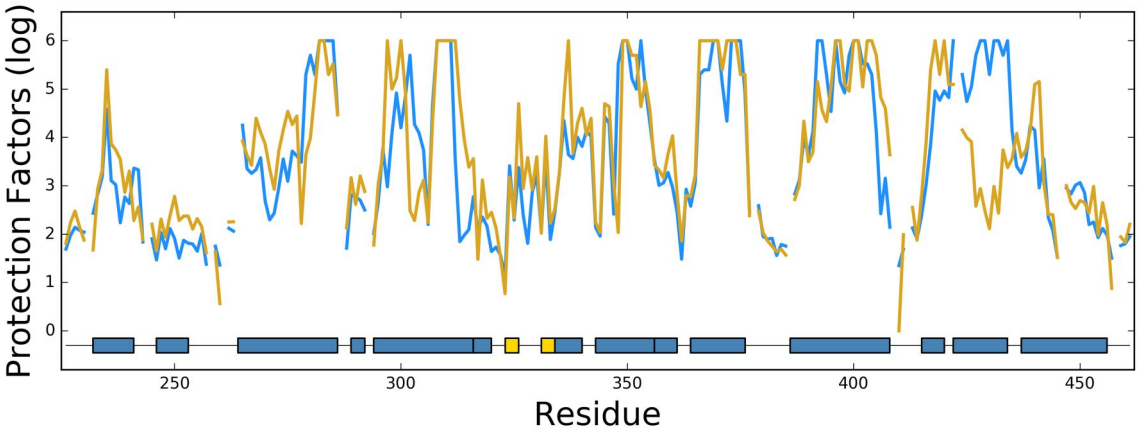


FIGURE 6.6 – Facteurs de protection PF (exprimé sur une échelle logarithmique) pour chaque proton amide HN. En bleu et en jaune, sont représentées les données obtenues à partir des trajectoires aMD de la forme libre du LBD de RXRα non phosphorylée et phosphorylée pSer260, respectivement.

Ainsi la prédiction de l'échange H/D à partir des facteurs de protection PF (Figure 6.6) indique que la phosphorylation a un impact sur la stabilité des structures secondaires au niveau des peptides suivants : résidus 271 à 278 et 271-279 (N-terminale de H3), 301-308 et 301-309 (jointure entre H4 et H5), 309-325 et 310-325 (partie C-terminale de H5 et premier brin du feuillet β), 401-415 (boucle H9/H10) ainsi que les peptides 419-432, 419-438, 423-438 et 426-438 (H10). La présence de la phosphorylation agit de manière différente suivant la région du LBD. Dans un cas, elle semble stabiliser la partie N-terminale de H3, la partie C-terminale de H5 et le feuille β ainsi que la boucle H9/H10. Ces régions sont impliquées dans la fixation du ligand, excepté la boucle H9/H10. Au contraire, la phosphorylation de la sérine 260 semble déstabiliser la région entre H4 et H5 et toute l'hélice H10. La première région forme avec la partie C-terminale de H3 la surface d'interaction pour les protéines CoA et la seconde région constitue la surface d'interaction avec les récepteurs nucléaires. Cependant la rapide incorporation en deutérium prédite observée au sein de l'hélice H10, peptides allant des résidus 419 à 438, est principalement due au dépliement de celle-ci

au cours d'une seule trajectoire aMD. En effet, la prédiction de l'incorporation en deutérium, en excluant cette trajectoire aMD, montre sans ambiguïté que l'échange H/D observé dans H10 est uniquement dû à cette dernière, avec une vitesse d'incorporation semblable à celle prédite à partir des trajectoires aMD de la forme non phosphorylée. Néanmoins, ces résultats montrent que le dépliement complet de H10 n'influence pas la vitesse d'incorporation au sein des autres peptides, et par extension la stabilité des autres éléments de structures secondaires (Figure 6.7).

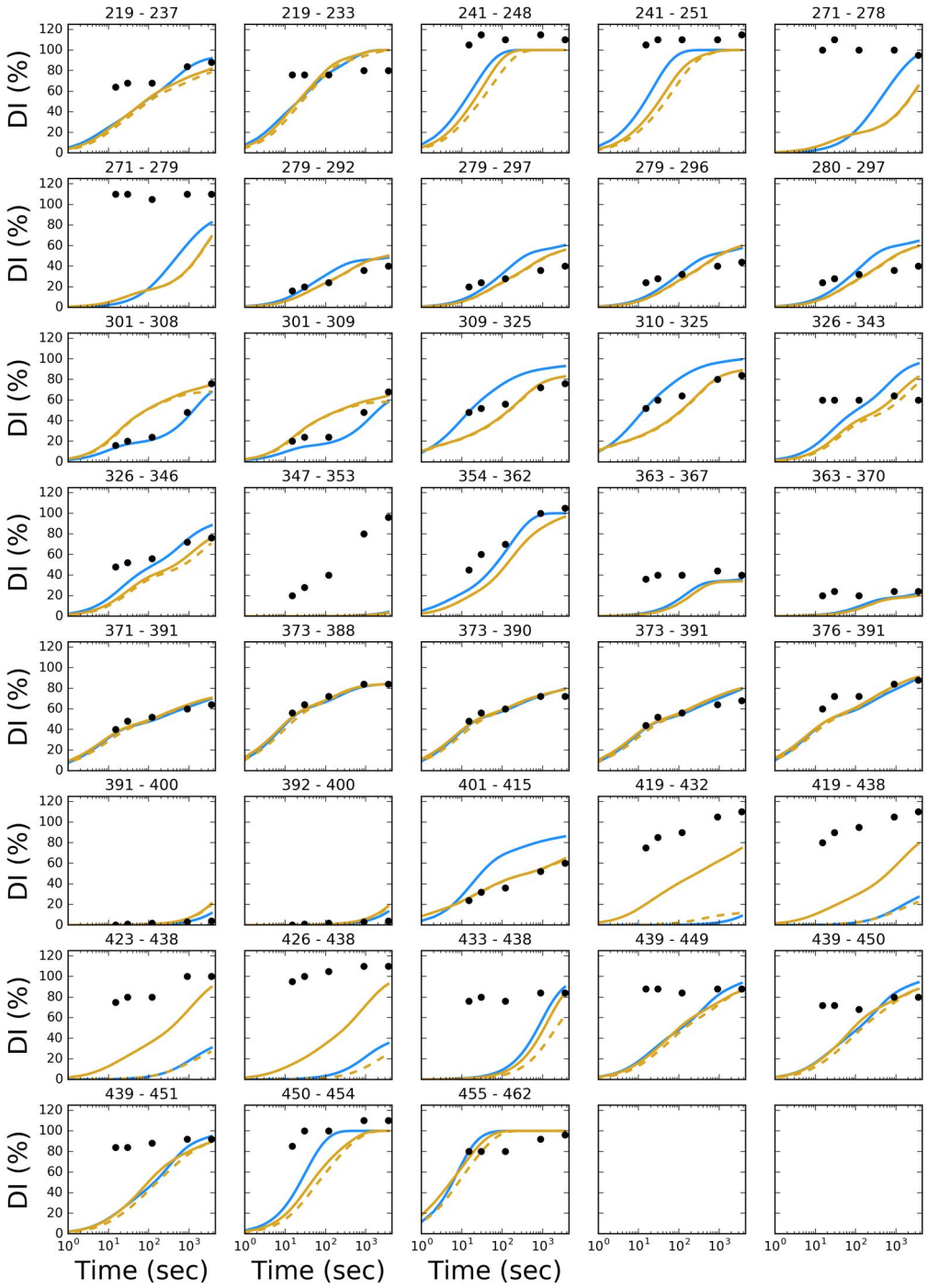


FIGURE 6.7 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique). En bleu et jaune, sont représentés les taux d'incorporation en deutérium prédits à partir des trajectoires aMD des formes libres du LBD du RXRα non phosphorylée et phosphorylée pSer260, respectivement. En jaune en trait pointillé, le taux d'incorporation prédit à partir des trajectoires aMD de la forme libre du LBD de RXRα pSer260, en excluant la trajectoire aMD dans laquelle on observe un dépliement de H10. En noir, le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes pour la forme libre non phosphorylée.

6.3.4 Analyse des conformations échantillonnées

Le paysage énergétique obtenu à partir de la méthode pSPE et en repondérant l'énergie en fonction du biais appliqué à chaque conformation nous permet d'identifier plusieurs régions de basse énergie (Figure 6.8). Les résultats montrent que les trajectoires aMD, en partant de la même structure, explorent des conformations différentes du LBD de RXRα pSer260. Cependant, l'espace échantillonné par chacune des trajectoires aMD ne présente pas la même diversité de conformations ni le même niveau d'énergie. L'anharmonicité calculée pour chaque classe de l'histogramme montre que les régions de basse énergie sont celles qui présentent aussi une anharmonicité proche de zéro ($\gamma < 0.1$), suggérant que ces régions ont été explorées convenablement et que l'énergie obtenue lors de la repondération est fiable. À l'inverse, les régions les moins bien explorées ($\gamma > 0.1$) coïncident avec les régions dont les conformations présentent des énergies plus élevées, signifiant que les conformations sont moins stables (Figure 6.8).

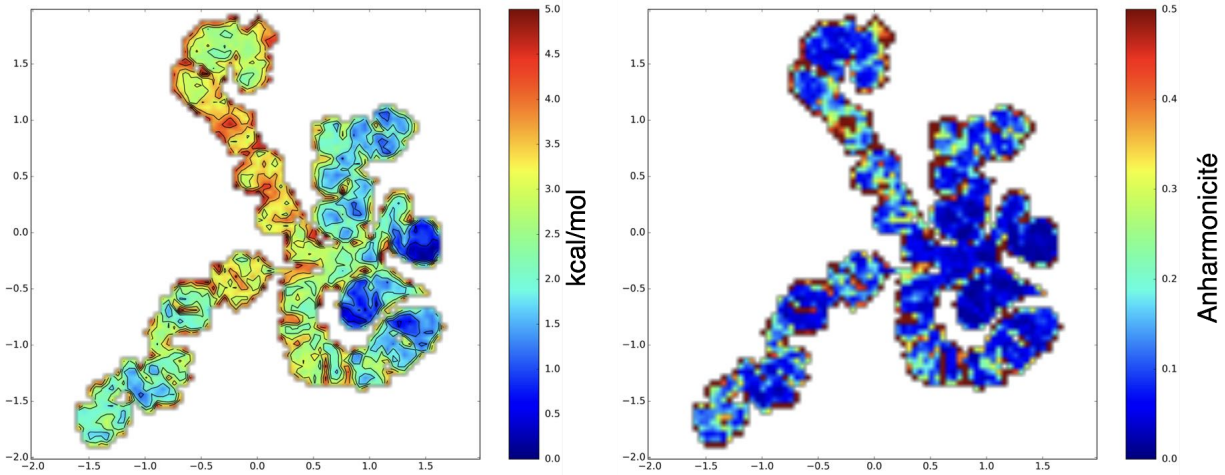


FIGURE 6.8 – À gauche, le paysage énergétique (en kcal/mol) de l'espace conformationnel échantillonné au cours des trajectoires aMD de la forme libre phosphorylée pSer260 du LBD de RXRα. Celui-ci a été obtenu en repondérant l'énergie à partir du biais appliqué à chaque conformation et en utilisant comme coordonnées collectives la projection optimale obtenue à partir de la méthode pSPE. À droite est représentée l'anharmonicité pour chaque classe en utilisant de même la configuration optimale obtenue à l'aide de la méthode pSPE. L'intervalle choisi est de 0.05 en ne considérant que les classes ayant au moins 10 conformations.

La visualisation des conformations échantillonnées au cours de la trajectoire aMD de 900 ns cumulées (6 x 150 ns) montre que les principaux changements structuraux sont localisés au niveau de la boucle H1-

H3, H3, le feuillet β, H12 ainsi qu'au niveau de H4 et H10 suivant les trajectoires visualisées (Figure 6.9). On distingue 3 minimas (labellisés 1, 2, 3) dont les structures diffèrent essentiellement dans l'organisation de la bouche H1-H3 (avec l'hélice H2), de l'hélice H12, et d'une partie de l'hélice H3. La structure (1) ressemble à la structure cristallographique 1LBD, en particulier, l'hélice H12 reste formée et étendue vers l'extérieur du LBD. Par contre, l'hélice H2 est déstructurée, et l'hélice H3 légèrement déplacée. Dans la structure (2), comme déjà observé lors des trajectoires de la forme libre non phosphorylée, H12 se repositionne en configuration « active-like ». Cependant, contrairement à la forme non phosphorylée, dans cette structure « active-like », H1 et H2 forment une longue hélice courbée. Le restant du domaine ne présente par contre pas de changements structuraux significatifs en comparaison avec la structure cristallographique de référence (1LBD). La structure (3) conserve la structure globale de H1, H2, H3, mais par contre H12 est partiellement déstructurée, ce qui avait également été observé sur les trajectoires de la forme libre non phosphorylée. Un quatrième minimum (4) ressemble à celui observé en (3) mis à part au niveau de H3, qui est déstructurée dans son milieu. Finalement, deux trajectoires ont exploré un ensemble de structures, avec peu de régions de basse énergie. Le minimum (5) présente un dépliement de H10, ainsi que la formation d'un troisième brin β antiparallèle entre H12 et le feuillet β. La région (6) ne présente pas de minimum marqué, certaines structures présentent une déstructuration de H4/H5. Ces régions (5) et (6) n'ont pas été observées dans les trajectoires de la forme libre non phosphorylée.

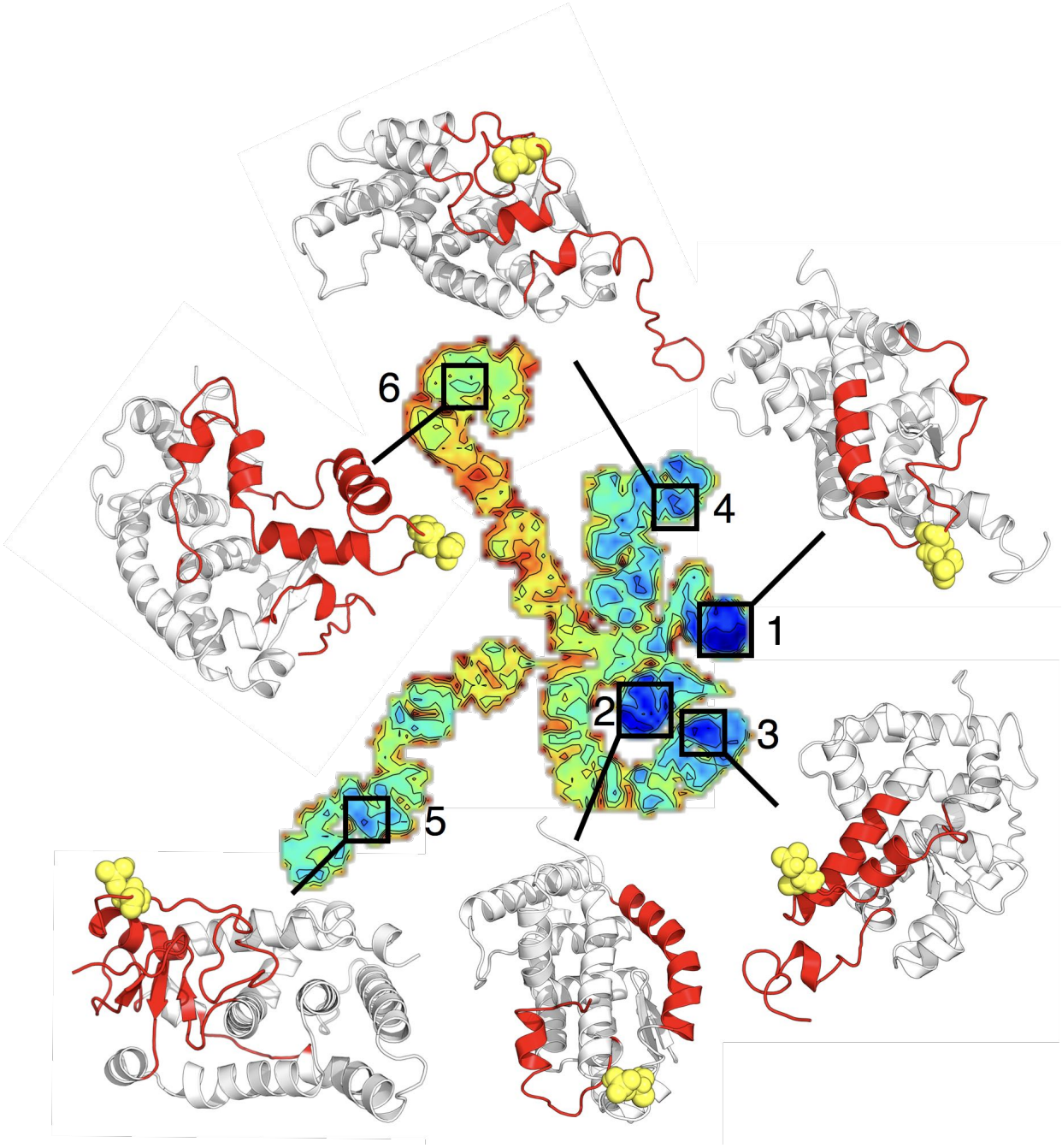


FIGURE 6.9 – Résumé des conformations pouvant être adoptées par la forme libre phosphorylée pSer260 du LBD de RXRα. Ces conformations ont été retrouvées au sein des minimas sélectionnés du paysage énergétique obtenu à partir des trajectoires de la forme libre phosphorylée pSer260 du LBD de RXRα. Les régions présentant un grand changement conformationnel par rapport à la structure cristallographique de la forme inactive du LBD de RXRα sont représentées en rouge avec le résidu Ser260 phosphorylé en jaune et le restant du domaine en blanc.

6.4 Discussion

L'analyse des trajectoires aMD de la forme libre du LBD de RXRα pSer260 montre que la phosphorylation de la sérine 260 semble influencer de manière différente des régions bien distinctes de ce domaine de RXRα. Parmi ces régions, nous retrouvons la boucle H1-H3, la surface d'interaction avec les protéines CoA, composées de H3, H4 et H5, la partie C-terminale de H5 ainsi que le premier brin du feuillet β qui sont impliqués dans l'interaction spécifique avec le ligand et la boucle H9/H10 dont la fonction n'est pas connue au sein du RXRα (Figures 6.10 et 6.11). En ce qui concerne H10, qui constitue pour l'essentiel l'interface de dimérisation avec les autres récepteurs nucléaires, il n'est pas clair si le dépliement observé au cours d'une trajectoire aMD peut être relié à la phosphorylation pSer260 (Figure 6.11). Néanmoins, cette observation nous éclaire quant à sa dynamique conformationnelle globale et que son dépliement n'influence pas la stabilité générale du LBD de RXRα.

Selon la région considérée du LBD de RXRα, la phosphorylation pSer260 peut soit entrainer une stabilisation de la structure secondaire soit une déstabilisation. Cependant, une stabilisation de la structure secondaire n'implique pas nécessairement une diminution de la fluctuation. La position d'une hélice α peut très bien fluctuer au cours du temps sans que l'hélice se déplie pour autant. Nous observons ce comportement pour l'hélice H2, dont la propension à rester en hélice est plus importante dans les trajectoires phosphorylées que dans les trajectoires de la forme libre non phosphorylée, mais dont les fluctuations RMS restent élevées. Dans le cas de la région C-terminale de H5 et du premier brin du feuillet β, en plus d'observer une rigidification de la structure secondaire avec des déplacements chimiques secondaires $\Delta\delta^{13}$Cα moyens prédits plus élevés et une vitesse d'échange H/D prédite plus lente pour la forme phosphorylée pSer260 comparée à la forme non phosphorylée. On observe de même une fluctuation RMS globalement plus basse, excepté dans 2 des 6 trajectoires aMD où la fluctuation RMS est supérieure à la fluctuation RMS moyenne observée pour la forme non phosphorylée. C'est le cas aussi pour la boucle H9/H10 avec une vitesse d'échange H/D prédite plus lente et une fluctuation RMS plus basse. Cependant, aucune différence n'est observée au niveau des déplacements chimiques secondaires $\Delta\delta^{13}$Cα prédits dans cette région.

En résumé, à partir des observables expérimentaux prédits des trajectoires aMD, il semblerait que la phosphorylation de la sérine 260, située dans la boucle H1-H3, stabilise la structure secondaire des éléments structuraux H2, H3, H4, H5 et plus particulièrement la partie C-terminale jusqu'au premier brin du feuillet β et la boucle H9/H10 du LBD de RXRα, mais malgré cette stabilisation, la mobilité des éléments structuraux H3 et H4 formant la surface d'interaction pour les protéines CoA avec H12, reste très importante.

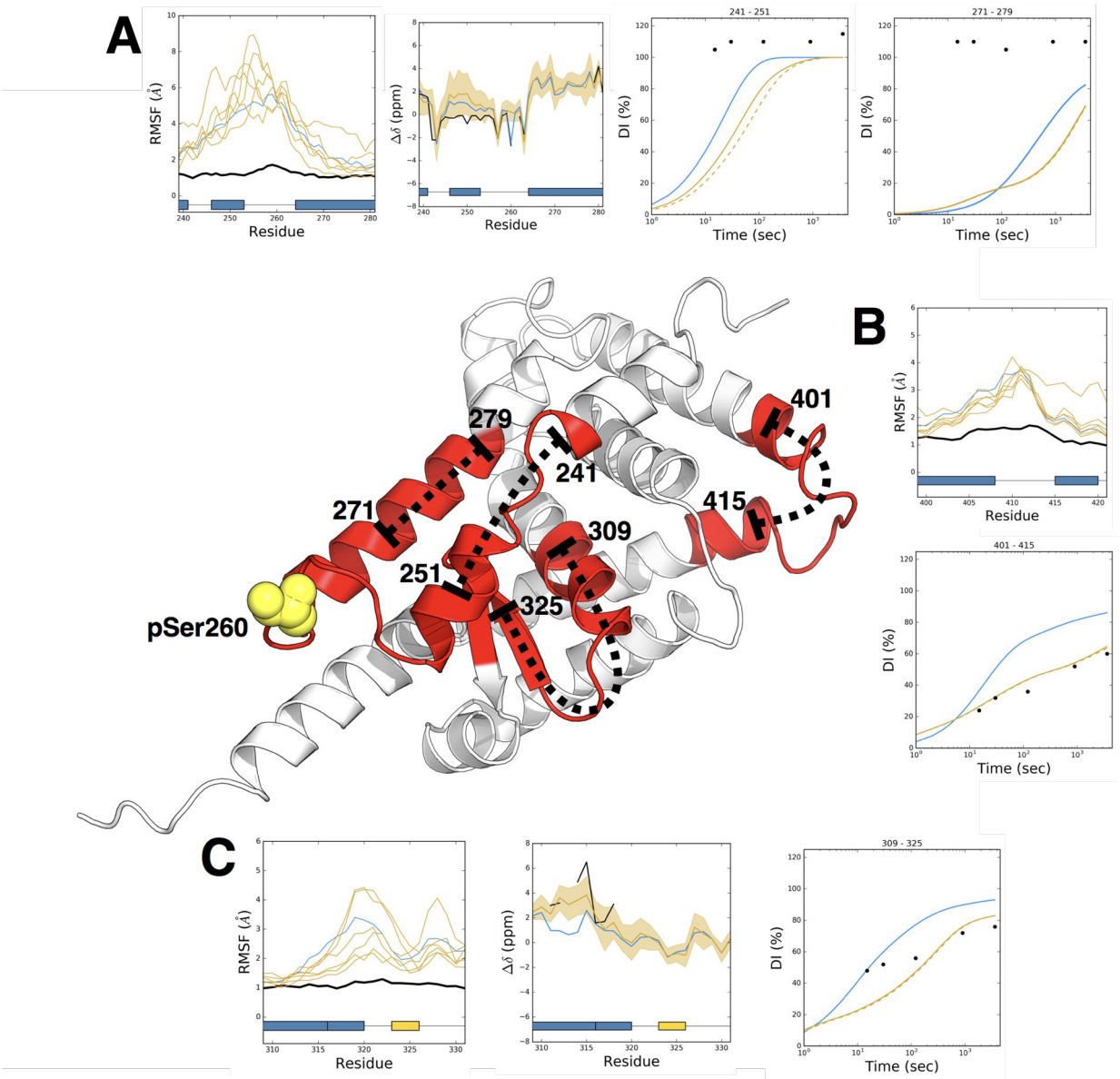


FIGURE 6.10 – Résumé des données expérimentales et prédites à partir des trajectoires aMD de la forme libre pSer260 du LBD de RXRα. (A) Fluctuations RMS par résidu, déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα et échange H/D pour les peptides 241-251 et 271-279 au niveau de la région H1-H3 avec H3 comprise. (B) Fluctuations RMS par résidu et échange H/D pour le peptide 401-415 de la boucle H9/H10. (C) Fluctuations RMS par résidu, déplacements chimiques RMN secondaires $\Delta\delta^{13}$Cα et échange H/D pour le peptide 309-325 de la région C-terminale de H5 et feuillet β. Les données concernant la forme libre pSer260 et la forme libre non phosphorylée sont représentées en jaune et bleues, respectivement. Les régions d'intérêt sont représentées en rouge sur la structure cristallographique de la forme inactive avec le site de phosphorylation en jaune et en pointillés sont délimités les peptides pour lesquelles l'échange H/D a pu être mesuré expérimentalement. Les fluctuations RMS par résidu sont représentées par trajectoire aMD de la forme libre pSer260. Dans le cas de l'échange H/D, le trait plein jaune correspond à la prédiction de l'échange H/D à partir de toutes les trajectoires aMD et le trait en pointillé jaune à partir de toutes les trajectoires aMD, excepté celle où H10 est déstructurée.

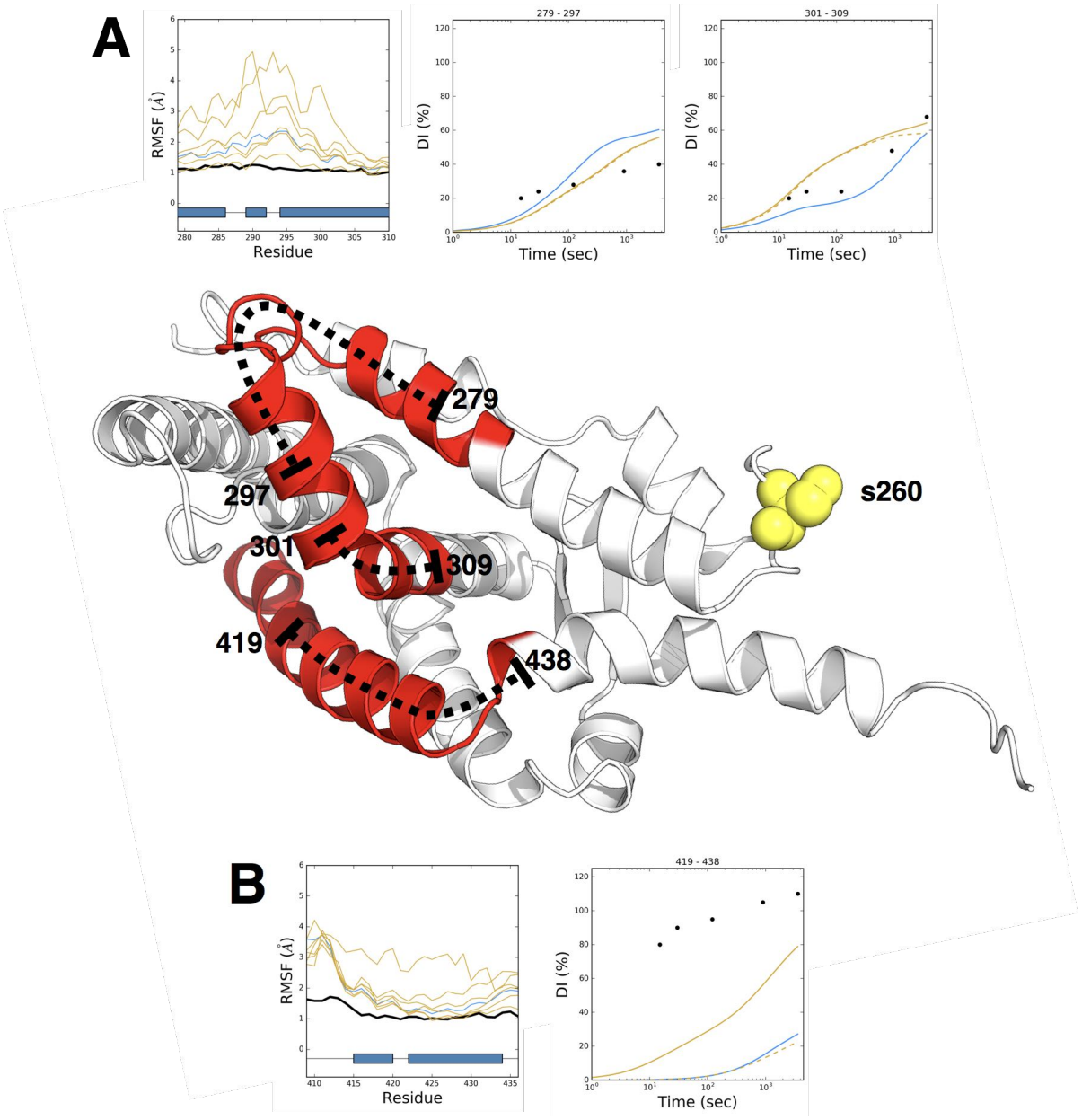


FIGURE 6.11 – Résumé des données expérimentales et prédites à partir des trajectoires aMD de la forme libre pSer260 du LBD de RXRα. (A) Fluctuations RMS par résidu et échange H/D pour les peptides 279-297 et 301-309 pour la région C-terminale de H3 au N-terminal de H5, compris. (B) Fluctuations RMS par résidu et échange H/D pour les peptides 419-438 pour l'hélice H10. Les données concernant la forme libre pSer260 et la forme libre non phosphorylée sont en jaune et bleu, respectivement. Les régions d'intérêt sont représentées en rouge sur la structure cristallographique de la forme inactive avec le site de phosphorylation en jaune et en pointillés sont délimités les peptides pour lesquelles l'échange H/D a pu être mesuré expérimentalement. Les fluctuations RMS par résidu sont représentées par trajectoire aMD de la forme libre pSer260. Dans le cas de l'échange H/D, le trait plein jaune correspond à la prédiction de l'échange H/D à partir de toutes les trajectoires aMD et le trait en pointillé jaune à partir de toutes les trajectoires aMD, excepté celle où H10 est déstructurée.

Chapitre 7

Conclusion et perspectives

Au cours de ce travail de thèse, plusieurs aspects ont été traités concernant soit le développement de méthodes d'analyse de trajectoires MD et de comparaison avec des données expérimentales ou bien l'étude de la dynamique structurale de la forme libre et de la forme liée à 9cRA du LBD de RXRα. L'implication de la phosphorylation pSer260 fut également abordée afin de comprendre son influence sur la dynamique conformationnelle de ce domaine.

Les données expérimentales publiées ces dernières années concernant la dynamique conformationnelle des récepteurs nucléaires, et plus particulièrement le LBD, suggèrent que ce domaine n'existe pas seulement soit dans un état « inactif » ou soit dans un « actif » en présence d'un ligand agoniste, mais qu'il explore un ensemble conformationnel plus vaste que ce qui est observé par cristallographie aux rayons X. Ainsi afin d'explorer au mieux l'espace conformationnel accessible nous nous sommes tournés vers une méthode d'échantillonnage amélioré, la dynamique moléculaire accélérée. Contrairement à la dynamique moléculaire classique, cette méthode permet de passer plus rapidement les barrières énergétiques entre les différents bassins. Étant récente, cette méthode est encore peu utilisée pour étudier la dynamique de systèmes de grandes tailles. De fait une partie de cette thèse a été dédiée au développement d'un ensemble de méthodes permettant d'analyser les trajectoires obtenues à l'aide de la dynamique moléculaire accélérée et de comparer les données qu'elles produisent aux données expérimentales disponibles. Des données de déplacements chimiques par RMN ainsi que des données d'échange H/D sont disponibles dans la littérature sur le LBD de RXRα et ont été mises à profit pour valider le protocole de dynamique moléculaire accélérée, mais aussi pour apporter une interprétation possible de ces données à l'échelle moléculaire. Dans le cas des déplacements chimiques RMN, les modèles théoriques sont matures et testés dans de nombreux exemples. Cependant dans le cas de l'échange H/D, le mécanisme sous-jacent est encore incertain au sein des protéines repliées. Ainsi plusieurs modèles ont été testés pour reproduire le mécanisme d'échange de proton, mais aussi pour en déduire les facteurs de protection nécessaires pour prédire l'échange H/D.

Les résultats montrent que l'espace conformationnel exploré par dynamique moléculaire accélérée des formes libre et liée à 9cRA du LBD de RXRα est en accord avec les données expérimentales. La comparaison entre les données RMN montre que les trajectoires aMD reproduisent les données expérimentales. On observe un meilleur accord entre les trajectoires aMD et les données expérimentales qu'avec les structures cristallographiques en particulier au niveau de la boucle H1-H3. Celles-ci suggèrent la présence d'une hélice (H2) au sein de cette boucle alors que les déplacements chimiques RMN secondaires montrent que cette région est déstructurée et ne présente pas d'hélice. Néanmoins, on observe une différence au niveau de la partie C-terminale de H5 où se situe Arg316, qui forme des liaisons hydrogène avec 9cRA dans la

forme liée. Les déplacements chimiques RMN secondaires montrent que cette région est très structurée en hélice, ce qui n'est pas le cas au cours des trajectoires aMD de la forme libre ainsi que pour la forme liée à 9cRA, qui malgré la présence du ligand ne stabilise pas suffisamment cette région en hélice. Lorsque l'on compare les trajectoires aMD aux données d'échange H/D, nous remarquons un bon accord au niveau du sandwich d'hélices α que ce soit pour la forme libre ou la forme liée à 9cRA. Cependant, on remarque dans la forme liée à 9cRA que les éléments structuraux formant la LBP échangent tout aussi rapidement que dans la forme libre, contrairement aux données expérimentales qui montrent une stabilisation de la LBP en présence d'un ligand. Cette divergence serait liée aux paramètres de champ de force de 9cRA. Dans le cas de la forme libre, nous remarquons que certaines régions telles que H3 n'échangent pas assez. Ceci est particulièrement le cas au niveau de l'interface de dimérisation, formée par H7 et H10 principalement, où l'on observe une forte structuration de ces hélices dans les trajectoires aMD. Ceci tend à montrer qu'un certain nombre de conformations n'ont pas encore été explorées au cours de trajectoires aMD du LBD de RXRα.

Dans le cas de la forme phosphorylée pSer260, l'analyse des trajectoires aMD montre que la phosphorylation semblerait avoir un effet sur la stabilité de l'interface d'interaction avec les protéines CoA et aussi au niveau de la partie C-terminale de H5, où le ligand interagit avec Arg316 par le biais de liaison hydrogène et de manière globale de l'ensemble de la LBP. Dans un cas, on observe une déstabilisation et dans l'autre cas, on observe une rigidification de l'ensemble des éléments structuraux impliqués dans la LBP. En ce qui concerne la surface d'interaction avec les protéines CoA, une étude précédente a montré que la phosphorylation pSer260 perturbait l'interaction entre les complexes hétérodimériques et les protéines CoA [Macoritto et al., 2008].

En plus de l'étude de la dynamique structurale de la forme libre et liée à 9cRA du LBD de RXRα, ainsi que de l'influence de la phosphorylation pSer260 sur la dynamique de ce dernier, il a été nécessaire de développer une méthode afin d'organiser la grande quantité de conformations générées par AMD, mais aussi d'obtenir une projection pouvant être utilisée comme coordonnées collectives afin de pouvoir repondérer l'énergie en fonction du biais appliqué et d'identifier les conformations de plus basses énergies. Les méthodes existantes ne sont pas adaptées pour traiter des données non linéaires, telles que la PCA ou ne sont pas adaptées pour traiter une quantité importante de données, telle que la méthode ISOMAP. De même, plutôt que d'utiliser le RMSD, du fait d'un coût de calcul trop important pour être utilisé pour comparer un nombre important de conformations, de l'ordre du million, nous avons utilisé une métrique basée sur les angles dièdres. Ainsi pour cela, nous avons utilisé l'algorithme pSPE, qui garantit une com-

plexité de calcul augmentant de manière linéaire en fonction du nombre de conformations analysées, avec la distance dihédrale comme métrique. La combinaison des deux a ainsi permis d'analyser nos longues trajectoires, contenant 1 million de conformations chacune très rapidement (< 1 heure), et d'identifier plusieurs conformations alternatives stables qui n'avaient encore jamais été observées pour le LBD d'un récepteur nucléaire. Par exemple, on observe pour la forme libre du LBD de RXRα un ensemble de conformations présentant l'hélice H12, qui porte la fonction activatrice dépendante du ligand (AF-2), dans une conformation proche de celle observée au sein de la structure cristallographique de la forme active, en complexe avec 9cRA, du LBD de RXRα. Cette observation peut expliquer pourquoi une forte concentration en protéine CoA pouvait activer RXRα, en déplaçant l'équilibre vers cette conformation [Nahoum et al., 2007]. Cependant, cette méthode peut poser des soucis suivant le système étudié et plus particulièrement si la protéine contient un nombre important de feuillets β. Dans le cas par exemple de la protéine ssG, bien que les conformations soient bien classées selon leur architecture, leurs topologies peuvent différer. Ceci est dû à un manque de sensibilité de la métrique basée uniquement sur les angles dièdres.

Ces travaux peuvent prendre plusieurs directions, par exemple une prochaine étape concernant l'amélioration de la métrique en y appliquant des poids différents selon le degré de variabilité au cours des trajectoires ou bien en fonction de la différence entre les angles au moment du calcul de la distance entre deux conformations. De même, il serait possible d'étendre aussi l'analyse en y incorporant en plus les angles dièdres des chaines latérales. Ceci pourrait permettre d'affiner encore plus l'analyse. Dans le cas de la prédiction de l'échange H/D à partir des trajectoires MD, de nombreuses modifications sont possibles pour améliorer le modèle. Cependant, cela nécessiterait une trajectoire MD classique ou accélérée dans laquelle toutes les conformations d'une protéine quelconque ont été explorées et pour laquelle des données d'échange H/D sont disponibles à l'échelle du résidu. En ce qui concerne la dynamique structurale de la forme libre et la forme liée à 9cRA du LBD de RXRα, on remarque qu'un certain de nombre de conformations n'ont pas été encore explorées, ainsi il serait intéressant de les continuer afin de voir si on peut améliorer l'accord entre les données prédites et les données expérimentales, et plus particulièrement l'échange H/D au niveau de la surface de dimérisation. Dans le cas de la forme liée à 9cRA uniquement, d'autres analyses complémentaires sont nécessaires afin de comprendre qu'elles sont les paramètres qui sont liés à la déstabilisation de la LBP au cours des trajectoires aMD.

Bibliographie

Bibliographie

[Abseher and Nilges, 1998] Abseher, R. and Nilges, M. (1998). Are there non-trivial dynamic cross-correlations in proteins ? *Journal of molecular biology*, 279(4) :911–920.

[Adachi et al., 2002] Adachi, S., Okuno, M., Matsushima-Nishiwaki, R., Takano, Y., Kojima, S., Friedman, S. L., Moriwaki, H., and Okano, Y. (2002). Phosphorylation of retinoid x receptor suppresses its ubiquitination in human hepatocellular carcinoma. *Hepatology*, 35(2) :332–340.

[Adam-Stitah et al., 1999] Adam-Stitah, S., Penna, L., Chambon, P., and Rochette-Egly, C. (1999). Hyperphosphorylation of the retinoid x receptor α by activated c-jun nh2-terminal kinases. *Journal of Biological Chemistry*, 274(27) :18932–18941.

[Adcock and McCammon, 2006] Adcock, S. A. and McCammon, J. A. (2006). Molecular dynamics : survey of methods for simulating the activity of proteins. *Chemical reviews*, 106(5) :1589–1615.

[Agrafiotis, 2003] Agrafiotis, D. K. (2003). Stochastic proximity embedding. *Journal of computational chemistry*, 24(10) :1215–1221.

[Allan et al., 1992] Allan, G. F., Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W. (1992). Ligand-dependent conformational changes in the progesterone receptor are necessary for events that follow dna binding. *Proceedings of the National Academy of Sciences*, 89(24) :11750–11754.

[Altis et al., 2007] Altis, A., Nguyen, P. H., Hegger, R., and Stock, G. (2007). Dihedral angle principal component analysis of molecular dynamics simulations. *The Journal of chemical physics*, 126(24) :244111.

[Altucci and Gronemeyer, 2001] Altucci, L. and Gronemeyer, H. (2001). Nuclear receptors in cell life and death. *Trends in Endocrinology & Metabolism*, 12(10) :460–468.

[Altucci et al., 2007] Altucci, L., Leibowitz, M. D., Ogilvie, K. M., De Lera, A. R., and Gronemeyer, H. (2007). Rar and rxr modulation in cancer and metabolic disease. *Nature Reviews Drug Discovery*, 6(10) :793–810.

[Amal, 2013] Amal, I. (2013). *Étude in silico de la régulation allostérique du récepteur à l'acide rétinoïque par phosphorylation*. PhD thesis, Strasbourg.

[Amato et al., 2012] Amato, A. A., Rajagopalan, S., Lin, J. Z., Carvalho, B. M., Figueira, A. C., Lu, J., Ayers, S. D., Mottin, M., Silveira, R. L., Souza, P. C., et al. (2012). Gq-16, a novel peroxisome proliferator-activated receptor γ (pparγ) ligand, promotes insulin sensitization without weight gain. *Journal of Biological Chemistry*, 287(33) :28169–28179.

[Anbalagan et al., 2012] Anbalagan, M., Huderson, B., Murphy, L., and Rowan, B. G. (2012). Post-translational modifications of nuclear receptors and human disease. *Nucl Recept Signal*, 10(001).

[Anderson et al., 2006] Anderson, J. S., LeMaster, D. M., and Hernández, G. (2006). Electrostatic potential energy within a protein monitored by metal charge-dependent hydrogen exchange. *Biophysical journal*, 91(11) :L93–L95.

[Arrar et al., 2012] Arrar, M., de Oliveira, C. A. F., Fajer, M., Sinko, W., and McCammon, J. A. (2012). w-rexamd : a hamiltonian replica exchange approach to improve free energy calculations for systems with kinetically trapped conformations. *Journal of chemical theory and computation*, 9(1) :18–23.

[Au-Fliegner et al., 1993] Au-Fliegner, M., Helmer, E., Casanova, J., Raaka, B. M., and Samuels, H. (1993). The conserved ninth c-terminal heptad in thyroid hormone and retinoic acid receptors mediates diverse responses by affecting heterodimer but not homodimer formation. *Molecular and cellular biology*, 13(9) :5725–5737.

[Bae et al., 2010] Bae, S.-H., Qiu, J., and Fox, G. C. (2010). Multidimensional scaling by deterministic annealing with iterative majorization algorithm. In *e-Science (e-Science), 2010 IEEE Sixth International Conference on*, pages 222–229. IEEE.

[Bai et al., 1993] Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins : Structure, Function, and Bioinformatics*, 17(1) :75–86.

[Bai et al., 1994] Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins*, 20(1) :4–14.

[Bai et al., 1995] Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995). Protein folding intermediates : native-state hydrogen exchange. *Science (New York, NY)*, 269(5221) :192.

[Baldwin, 2011] Baldwin, R. L. (2011). Early days of protein hydrogen exchange : 1954–1972. *Proteins : Structure, Function, and Bioinformatics*, 79(7) :2021–2026.

[Baron et al., 2008] Baron, R., Wong, S. E., De Oliveira, C. A., and McCammon, J. A. (2008). E9-im9 colicin dnase- immunity protein biomolecular association in water : A multiple-copy and accelerated molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 112(51) :16802–16814.

[Bashford and Case, 2000] Bashford, D. and Case, D. A. (2000). Generalized born models of macromolecular solvation effects. *Annual review of physical chemistry*, 51(1) :129–152.

[Bastien et al., 2002] Bastien, J., Adam-Stitah, S., Plassat, J.-L., Chambon, P., and Rochette-Egly, C. (2002). The phosphorylation site located in the a region of retinoic x receptor α is required for the antiproliferative effect of retinoic acid (ra) and the activation of ra target genes in f9 cells. *Journal of Biological Chemistry*, 277(32) :28683–28689.

[Beato, 1991] Beato, M. (1991). Transcriptional control by nuclear receptors. *The FASEB Journal*, 5(7) :2044–2051.

[Belorusova et al., 2016] Belorusova, A. Y., Osz, J., Petoukhov, M. V., Peluso-Iltis, C., Kieffer, B., Svergun, D. I., and Rochel, N. (2016). Solution behavior of the intrinsically disordered n-terminal domain of retinoid x receptor α in the context of the full-length protein. *Biochemistry*, 55(12) :1741–1748.

[Berger and Linderstrøm-Lang, 1957] Berger, A. and Linderstrøm-Lang, K. (1957). Deuterium exchange of poly-dl-alanine in aqueous solution. *Archives of biochemistry and biophysics*, 69 :106–118.

[Berger and Loewenstein, 1959] Berger, A. and Loewenstein, A. (1959). Nuclear magnetic resonance study of the protolysis and ionization of N-methylacetamide1. *Journal of the American Chemical Society*, 81(1) :62–67.

[Berger et al., 1996] Berger, J., Bailey, P., Biswas, C., Cullinan, C. A., Doebber, T. W., Hayes, N. S., Saperstein, R., Smith, R., and Leibowitz, M. D. (1996). Thiazolidinediones produce a conformational change in peroxisomal proliferator-activated receptor-gamma : binding and activation correlate with antidiabetic actions in db/db mice. *Endocrinology*, 137(10) :4189–4195.

[Berger et al., 2003] Berger, J. P., Petro, A. E., Macnaul, K. L., Kelly, L. J., Zhang, B. B., Richards, K., Elbrecht, A., Johnson, B. A., Zhou, G., Doebber, T. W., et al. (2003). Distinct properties and advantages of a novel peroxisome proliferator-activated protein γ selective modulator. *Molecular Endocrinology*, 17(4) :662–676.

[Berman et al., 2000] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, 28(1) :235–242.

[Boehm et al., 1995] Boehm, M. F., Zhang, L., Zhi, L., McClurg, M. R., Berger, E., Wagoner, M., Mais, D. E., Suto, C. M., and Davies, P. J. (1995). Design and synthesis of potent retinoid x receptor selective ligands that induce apoptosis in leukemia cells. *Journal of medicinal chemistry*, 38(16) :3146–3155.

[Boerma et al., 2014] Boerma, L. J., Xia, G., Qui, C., Cox, B. D., Chalmers, M. J., Smith, C. D., Lobo-Ruppert, S., Griffin, P. R., Muccio, D. D., and Renfrow, M. B. (2014). Defining the communication between agonist and coactivator binding in the retinoid x receptor α ligand binding domain. *Journal of Biological Chemistry*, 289(2) :814–826.

[Bokeh Development Team, 2014] Bokeh Development Team (2014). *Bokeh : Python library for interactive visualization*.

[Boudjelal et al., 2000] Boudjelal, M., Wang, Z., Voorhees, J. J., and Fisher, G. J. (2000). Ubiquitin/proteasome pathway regulates levels of retinoic acid receptor γ and retinoid x receptor α in human keratinocytes. *Cancer research*, 60(8) :2247–2252.

[Bouillon et al., 2006] Bouillon, R., Eelen, G., Verlinden, L., Mathieu, C., Carmeliet, G., and Verstuyf, A. (2006). Vitamin d and cancer. *The Journal of steroid biochemistry and molecular biology*, 102(1) :156–162.

[Bourguet et al., 1995] Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor rxr-a. *Nature*, 375(6530) :377–382.

[Brelivet et al., 2004] Brelivet, Y., Kammerer, S., Rochel, N., Poch, O., and Moras, D. (2004). Signature of the oligomeric behaviour of nuclear receptors at the sequence and structural level. *EMBO reports*, 5(4) :423–429.

[Brocard et al., 1996] Brocard, J., Kastner, P., and Chambon, P. (1996). Two novel rxrα isoforms from mouse testis. *Biochemical and biophysical research communications*, 229(1) :211–218.

[Brooks et al., 2009] Brooks, B. R., Brooks, C. L., MacKerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). Charmm : the biomolecular simulation program. *Journal of computational chemistry*, 30(10) :1545–1614.

[Bruccoleri and Karplus, 1990] Bruccoleri, R. E. and Karplus, M. (1990). Conformational sampling using high-temperature molecular dynamics. *Biopolymers*, 29(14) :1847–1862.

[Bruck et al., 2005] Bruck, N., Bastien, J., Bour, G., Tarrade, A., Plassat, J.-L., Bauer, A., Adam-Stitah, S., and Rochette-Egly, C. (2005). Phosphorylation of the retinoid x receptor at the omega loop, modulates

the expression of retinoic-acid-target genes with a promoter context specificity. *Cellular signalling*, 17(10) :1229–1239.

[Bruning et al., 2007] Bruning, J. B., Chalmers, M. J., Prasad, S., Busby, S. A., Kamenecka, T. M., He, Y., Nettles, K. W., and Griffin, P. R. (2007). Partial agonists activate pparγ using a helix 12 independent mechanism. *Structure*, 15(10) :1258–1271.

[Bucher et al., 2011a] Bucher, D., Grant, B. J., Markwick, P. R., and McCammon, J. A. (2011a). Accessing a Hidden Conformation of the Maltose Binding Protein Using Accelerated Molecular Dynamics. *PLoS Comput. Biol.*, 7(4) :e1002034–10.

[Bucher et al., 2011b] Bucher, D., Pierce, L. C. T., McCammon, J. A., and Markwick, P. R. L. (2011b). On the Use of Accelerated Molecular Dynamics to Enhance Configurational Sampling in Ab Initio Simulations. *J Chem Theory Comput*, 7(4) :890–897.

[Calléja et al., 2006] Calléja, C., Messaddeq, N., Chapellier, B., Yang, H., Krezel, W., Li, M., Metzger, D., Mascrez, B., Ohta, K., Kagechika, H., et al. (2006). Genetic and pharmacological evidence that a retinoic acid cannot be the rxr-activating ligand in mouse epidermis keratinocytes. *Genes & development*, 20(11) :1525–1538.

[Canan Koch et al., 1996] Canan Koch, S. S., Dardashti, L. J., Hebert, J. J., White, S. K., Croston, G. E., Flatten, K. S., Heyman, R. A., and Nadzan, A. M. (1996). Identification of the first retinoid x receptor homodimer antagonist. *Journal of medicinal chemistry*, 39(17) :3229–3234.

[Carson et al., 2014] Carson, M. W., Zhang, J., Chalmers, M. J., Bocchinfuso, W. P., Holifield, K. D., Masquelin, T., Stites, R. E., Stayrook, K. R., Griffin, P. R., and Dodge, J. A. (2014). Hdx reveals unique fragment ligands for the vitamin d receptor. *Bioorganic & medicinal chemistry letters*, 24(15) :3459–3463.

[Chalmers et al., 2006] Chalmers, M. J., Busby, S. A., Pascal, B. D., He, Y., Hendrickson, C. L., Marshall, A. G., and Griffin, P. R. (2006). Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Analytical chemistry*, 78(4) :1005–1014.

[Chalmers et al., 2007] Chalmers, M. J., Busby, S. A., Pascal, B. D., Southern, M. R., and Griffin, P. R. (2007). A two-stage differential hydrogen deuterium exchange method for the rapid characterization of protein/ligand interactions. *Journal of Biomolecular Techniques*, 18(4) :194.

[Chamberlain et al., 1996] Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Biol.*, 3(9) :782–787.

[Chambon, 1996] Chambon, P. (1996). A decade of molecular biology of retinoic acid receptors. *The FASEB Journal*, 10(9) :940–954.

[Chambon and Rochette-Egly, 2001] Chambon, P. and Rochette-Egly, C. (2001). F9 embryocarcinoma cells, a cell autonomous model to study the functional selectivity of rars and rxrs in retinoid signaling. *Histology and histopathology*.

[Chandra et al., 2008] Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T. P., and Rastinejad, F. (2008). Structure of the intact pparγ-rxrα nuclear receptor complex on dna. *Nature*, 456(7220) :350–356.

[Chao et al., 2008] Chao, E. Y., Caravella, J. A., Watson, M. A., Campobasso, N., Ghisletti, S., Billin, A. N., Galardi, C., Wang, P., Laffitte, B. A., Iannone, M. A., et al. (2008). Structure-guided design of n-phenyl tertiary amines as transrepression-selective liver x receptor modulators with anti-inflammatory activity. *Journal of medicinal chemistry*, 51(18) :5758–5765.

[Choi et al., 2006] Choi, S. J., Chung, S. S., Rho, E. J., Lee, H. W., Lee, M. H., Choi, H.-S., Seol, J. H., Baek, S. H., Bang, O. S., and Chung, C. H. (2006). Negative modulation of rxrα transcriptional activity by small ubiquitin-related modifier (sumo) modification and its reversal by sumo-specific protease susp1. *Journal of Biological Chemistry*, 281(41) :30669–30677.

[Chyan et al., 1993] Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A., and Baum, J. (1993). Structure and stability of the molten globule state of guinea pig. alpha.-lactalbumin : A hydrogen exchange study. *Biochemistry*, 32(21) :5681–5691.

[Connelly et al., 1993] Connelly, G. P., Bai, Y., Jeng, M.-F., and Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins : Structure, Function, and Bioinformatics*, 17(1) :87–92.

[Covington et al., 1966] Covington, A., Robinson, R., and Bates, R. G. (1966). The ionization constant of deuterium oxide from 5 to 50. *The Journal of Physical Chemistry*, 70(12) :3820–3824.

[Craig et al., 2011] Craig, P. O., Lätzer, J., Weinkam, P., Hoffman, R. M., Ferreiro, D. U., Komives, E. A., and Wolynes, P. G. (2011). Prediction of native-state hydrogen exchange from perfectly funneled energy landscapes. *Journal of the American Chemical Society*, 133(43) :17463–17472.

[Cramer et al., 2012] Cramer, P. E., Cirrito, J. R., Wesson, D. W., Lee, C. D., Karlo, J. C., Zinn, A. E., Casali, B. T., Restivo, J. L., Goebel, W. D., James, M. J., et al. (2012). Apoe-directed therapeutics rapidly clear β-amyloid and reverse deficits in ad mouse models. *Science*, 335(6075) :1503–1506.

[Craveur et al., 2013] Craveur, P., Joseph, A. P., Poulain, P., de Brevern, A. G., and Rebehmed, J. (2013). Cis–trans isomerization of omega dihedrals in proteins. *Amino acids*, 45(2) :279–289.

[Cronet et al., 2001] Cronet, P., Petersen, J. F., Folmer, R., Blomberg, N., Sjöblom, K., Karlsson, U., Lindstedt, E.-L., and Bamberg, K. (2001). Structure of the pparα and-γ ligand binding domain in complex with az 242 ; ligand selectivity and agonist activation in the ppar family. *Structure*, 9(8) :699–706.

[Cummings et al., 2016] Cummings, J. L., Zhong, K., Kinney, J. W., Heaney, C., Moll-Tudla, J., Joshi, A., Pontecorvo, M., Devous, M., Tang, A., and Bena, J. (2016). Double-blind, placebo-controlled, proof-of-concept trial of bexarotene in moderate alzheimer's disease. *Alzheimer's research & therapy*, 8(1) :1.

[Dai et al., 2009] Dai, S. Y., Burris, T. P., Dodge, J. A., Montrose-Rafizadeh, C., Wang, Y., Pascal, B. D., Chalmers, M. J., and Griffin, P. R. (2009). Unique ligand binding patterns between estrogen receptor α and β revealed by hydrogen- deuterium exchange. *Biochemistry*, 48(40) :9668–9676.

[Dai et al., 2008] Dai, S. Y., Chalmers, M. J., Bruning, J., Bramlett, K. S., Osborne, H. E., Montrose-Rafizadeh, C., Barr, R. J., Wang, Y., Wang, M., Burris, T. P., et al. (2008). Prediction of the tissue-specificity of selective estrogen receptor modulators by using a single biochemical method. *Proceedings of the National Academy of Sciences*, 105(20) :7171–7176.

[Dawson and Xia, 2012] Dawson, M. I. and Xia, Z. (2012). The retinoid x receptors and their ligands. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1821(1) :21–56.

[De Leeuw, 2005] De Leeuw, J. (2005). Applications of convex analysis to multidimensional scaling. *Department of Statistics, UCLA*.

[De Loof et al., 1992] De Loof, H., Nilsson, L., and Rigler, R. (1992). Molecular dynamics simulation of galanin in aqueous and nonaqueous solution. *Journal of the American Chemical Society*, 114(11) :4028–4035.

[de Oliveira et al., 2011] de Oliveira, C. A. F., Grant, B. J., Zhou, M., and McCammon, J. A. (2011). Large-Scale Conformational Changes of Trypanosoma cruzi Proline Racemase Predicted by Accelerated Molecular Dynamics Simulation. *PLoS Computional Biology*, 7(10) :e1002178–7.

[de Oliveira et al., 2006] de Oliveira, C. A. F., Hamelberg, D., and McCammon, J. A. (2006). On the application of accelerated molecular dynamics to liquid water simulations. *The journal of physical chemistry B*, 110(45) :22695–22701.

[Desvergne, 2007] Desvergne, B. (2007). Rxr : from partnership to leadership in metabolic regulations. *Vitamins & Hormones*, 75 :1–32.

[Dickinson, 1950] Dickinson, W. (1950). Dependence of the f 19 nuclear resonance position on chemical compound. *Physical Review*, 77(5) :736.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1) :269–271.

[Doshi and Hamelberg, 2009] Doshi, U. and Hamelberg, D. (2009). Reoptimization of the amber force field parameters for peptide bond (omega) torsions using accelerated molecular dynamics. *The Journal of Physical Chemistry B*, 113(52) :16590–16595.

[Doshi and Hamelberg, 2011] Doshi, U. and Hamelberg, D. (2011). Extracting Realistic Kinetics of Rare Activated Processes from Accelerated Molecular Dynamics Using Kramers' Theory. *J Chem Theory Comput*, 7(3) :575–581.

[Doshi and Hamelberg, 2012] Doshi, U. and Hamelberg, D. (2012). Improved statistical sampling and accuracy with accelerated molecular dynamics on rotatable torsions. *Journal of chemical theory and computation*, 8(11) :4004–4012.

[Doshi and Hamelberg, 2014] Doshi, U. and Hamelberg, D. (2014). Achieving Rigorous Accelerated Conformational Sampling in Explicit Solvent. *J. Phys. Chem. Lett.*, 5(7) :1217–1224.

[Doshi and Hamelberg, 2015] Doshi, U. and Hamelberg, D. (2015). Towards fast, rigorous and efficient conformational sampling of biomolecules : Advances in accelerated molecular dynamics. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(5) :878–888.

[Earl and Deem, 2005] Earl, D. J. and Deem, M. W. (2005). Parallel tempering : Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23) :3910–3916.

[Egea et al., 2002] Egea, P. F., Mitschler, A., and Moras, D. (2002). Molecular recognition of agonist ligands by rxrs. *Molecular endocrinology*, 16(5) :987–997.

[Egea et al., 2000] Egea, P. F., Mitschler, A., Rochel, N., Ruff, M., Chambon, P., and Moras, D. (2000). Crystal structure of the human rxrα ligand-binding domain bound to its natural ligand : 9-cis retinoic acid. *The EMBO journal*, 19(11) :2592–2601.

[Eigen, 1964] Eigen, M. (1964). Proton transfer, acid-base catalysis, and enzymatic hydrolysis. part i : Elementary processes. *Angewandte Chemie International Edition in English*, 3(1) :1–19.

[Englander et al., 1979] Englander, J. J., Calhoun, D. B., and Englander, S. W. (1979). Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry. *Analytical biochemistry*, 92(2) :517–524.

[Englander, 1963] Englander, S. W. (1963). A hydrogen exchange method using tritium and sephadex : Its application to ribonuclease*. *Biochemistry*, 2(4) :798–807.

[Englander and Downer, 1972] Englander, S. W. and Downer, N. W. (1972). Hydrogen exchange. *Annual Review of Biochemistry*, 41(1) :903–924.

[Fajer et al., 2008] Fajer, M., Hamelberg, D., and McCammon, J. A. (2008). Replica-exchange accelerated molecular dynamics (rexamd) applied to thermodynamic integration. *Journal of chemical theory and computation*, 4(10) :1565–1569.

[Fajer et al., 2009] Fajer, M., Swift, R. V., and McCammon, J. A. (2009). Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics. *Journal of computational chemistry*, 30(11) :1719–1725.

[Fidelak, 2006] Fidelak, J. (2006). *Etude théorique de l'équilibre conformationnel du PPAR-γ*. PhD thesis, Strasbourg 1.

[Figueira et al., 2011] Figueira, A., Saidemberg, D., Souza, P., Martínez, L., Scanlan, T., Baxter, J., Skaf, M., Palma, M. S., Webb, P., and Polikarpov, I. (2011). Analysis of agonist and antagonist effects on thyroid hormone receptor conformation by hydrogen/deuterium exchange. *Molecular Endocrinology*, 25(1) :15–31.

[Filomia et al., 2010] Filomia, F., De Rienzo, F., and Menziani, M. C. (2010). Insights into MAPK p38α DFG flip mechanism by accelerated molecular dynamics. *Bioorganic & Medicinal Chemistry*, 18(18) :6805–6812.

[Fleet, 2008] Fleet, J. C. (2008). Molecular actions of vitamin d contributing to cancer prevention. *Molecular aspects of medicine*, 29(6) :388–396.

[Fogolari et al., 1998] Fogolari, F., Esposito, G., Viglino, P., Briggs, J. M., and McCammon, J. A. (1998). p k a shift effects on backbone amide base-catalyzed hydrogen exchange rates in peptides. *Journal of the American Chemical Society*, 120(15) :3735–3738.

[Forman et al., 1995] Forman, B. M., Umesono, K., Chen, J., and Evans, R. M. (1995). Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell*, 81(4) :541–550.

[Frego and Davidson, 2006] Frego, L. and Davidson, W. (2006). Conformational changes of the glucocorticoid receptor ligand binding domain induced by ligand and cofactor binding, and the location of cofactor binding sites determined by hydrogen/deuterium exchange mass spectrometry. *Protein science*, 15(4) :722–730.

[Fukasawa et al., 2012] Fukasawa, H., Nakagomi, M., Yamagata, N., Katsuki, H., Kawahara, K., Kitaoka, K., Miki, T., and Shudo, K. (2012). Tamibarotene : a candidate retinoid drug for alzheimer's disease. *Biological and Pharmaceutical Bulletin*, 35(8) :1206–1212.

[Gaillard et al., 2013] Gaillard, T., Schwarz, B. B., Chebaro, Y., Stote, R. H., and Dejaegere, A. (2013). Protein structural statistics with pss. *Journal of chemical information and modeling*, 53(9) :2471–2482.

[Gampe et al., 2000a] Gampe, R. T., Montana, V. G., Lambert, M. H., Miller, A. B., Bledsoe, R. K., Milburn, M. V., Kliewer, S. A., Willson, T. M., and Xu, H. E. (2000a). Asymmetry in the pparγ/rxrα crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Molecular cell*, 5(3) :545–555.

[Gampe et al., 2000b] Gampe, R. T., Montana, V. G., Lambert, M. H., Wisely, G. B., Milburn, M. V., and Xu, H. E. (2000b). Structural basis for autorepression of retinoid x receptor by tetramer formation and the af-2 helix. *Genes & development*, 14(17) :2229–2241.

[Gee et al., 1999] Gee, A. C., Carlson, K. E., Martini, P. G., Katzenellenbogen, B. S., and Katzenellenbogen, J. A. (1999). Coactivator peptides have a differential stabilizing effect on the binding of estrogens and antiestrogens with the estrogen receptor. *Molecular endocrinology*, 13(11) :1912–1923.

[Gee and Katzenellenbogen, 2001] Gee, A. C. and Katzenellenbogen, J. A. (2001). Probing conformational changes in the estrogen receptor : evidence for a partially unfolded intermediate facilitating ligand binding and release. *Molecular Endocrinology*, 15(3) :421–428.

[Germain et al., 2006] Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J., and Gronemeyer, H. (2006). International union of pharmacology. lxiii. retinoid x receptors. *Pharmacological reviews*, 58(4) :760–772.

[Germain et al., 2002] Germain, P., Iyer, J., Zechel, C., and Gronemeyer, H. (2002). Co-regulator recruitment and the mechanism of retinoic acid receptor synergy. *Nature*, 415(6868) :187–192.

[Gianní et al., 2003] Gianní, M., Tarrade, A., Nigro, E. A., Garattini, E., and Rochette-Egly, C. (2003). The af-1 and af-2 domains of rarγ2 and rxrα cooperate for triggering the transactivation and the degradation of rarγ2/rxrα heterodimers. *Journal of Biological Chemistry*, 278(36) :34458–34466.

[Gilardi and Desvergne, 2014] Gilardi, F. and Desvergne, B. (2014). Rxrs : collegial partners. In *The Biochemistry of Retinoic Acid Receptors I : Structure, Activation, and Function at the Molecular Level*, pages 75–102. Springer.

[Glasoe and Long, 1960] Glasoe, P. K. and Long, F. (1960). Use of glass electrodes to measure acidities in deuterium oxide1, 2. *The Journal of Physical Chemistry*, 64(1) :188–190.

[Goldman, 2004] Goldman, M. (2004). Histoire de la rmn. *L'actualité chimique*, 273(204) :57–60.

[Goswami et al., 2014] Goswami, D., Callaway, C., Pascal, B. D., Kumar, R., Edwards, D. P., and Griffin, P. R. (2014). Influence of domain interactions on conformational mobility of the progesterone receptor detected by hydrogen/deuterium exchange mass spectrometry. *Structure*, 22(7) :961–973.

[Gower, 1966] Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4) :325–338.

[Grant et al., 2009] Grant, B. J., Gorfe, A. A., and McCammon, J. A. (2009). Ras Conformational Switching : Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics. *PLoS Comput. Biol.*, 5(3) :e1000325.

[Gregory et al., 1983] Gregory, R. B., Crabo, L., Percy, A. J., and Rosenberg, A. (1983). Water catalysis of peptide hydrogen isotope exchange. *Biochemistry*, 22(4) :910–917.

[Gronemeyer and Laudet, 1995] Gronemeyer, H. and Laudet, V. (1995). Transcription factors 3 : nuclear receptors. *Protein profile*, 2(11) :1173.

[Guarnera et al., 2009] Guarnera, E., Pellarin, R., and Caflisch, A. (2009). How does a simplified-sequence protein fold ? *Biophysical journal*, 97(6) :1737–1746.

[Hamada et al., 1989] Hamada, K., Gleason, S. L., Levi, B.-Z., Hirschfeld, S., Appella, E., and Ozato, K. (1989). H-2riibp, a member of the nuclear hormone receptor superfamily that binds to both the regulatory element of major histocompatibility class i genes and the estrogen response element. *Proceedings of the National Academy of Sciences*, 86(21) :8289–8293.

[Hamelberg et al., 2007] Hamelberg, D., de Oliveira, C. A. F., and McCammon, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of Chemical Physics*, 127(15) :155102.

[Hamelberg and McCammon, 2005] Hamelberg, D. and McCammon, J. A. (2005). Fast peptidyl cis-trans isomerization within the flexible gly-rich flaps of hiv-1 protease. *Journal of the American Chemical Society*, 127(40) :13778–13779.

[Hamelberg and McCammon, 2006] Hamelberg, D. and McCammon, J. A. (2006). Accelerating conformational transitions in biomolecular simulations. *Annual Reports in Computational Chemistry*, 2 :221–232.

[Hamelberg and McCammon, 2009] Hamelberg, D. and McCammon, J. A. (2009). Mechanistic Insight into the Role of Transition-State Stabilization in Cyclophilin A. *Journal of the American Chemical Society*, 131(1) :147–152.

[Hamelberg et al., 2004] Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics : a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*, 120(24) :11919–11929.

[Hamelberg et al., 2005a] Hamelberg, D., Shen, T., and Andrew McCammon, J. (2005a). Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations. *The Journal of Chemical Physics*, 122(24) :241103–5.

[Hamelberg et al., 2005b] Hamelberg, D., Shen, T., and McCammon, J. A. (2005b). Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine-proline motifs : accelerated molecular dynamics analysis. *Journal of the American Chemical Society*, 127(6) :1969–1974.

[Hammonds and Ryckaert, 1991] Hammonds, K. D. and Ryckaert, J.-P. (1991). On the convergence of the shake algorithm. *Computer Physics Communications*, 62(2) :336–351.

[Hamuro et al., 2006] Hamuro, Y., Coales, S. J., Morrow, J. A., Molnar, K. S., Tuske, S. J., Southern, M. R., and Griffin, P. R. (2006). Hydrogen/deuterium-exchange (h/d-ex) of pparγ lbd in the presence of various modulators. *Protein science*, 15(8) :1883–1892.

[Han et al., 2011] Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011). Shiftx2 : significantly improved protein chemical shift prediction. *Journal of biomolecular NMR*, 50(1) :43–57.

[Harish et al., 2000] Harish, S., Ashok, M., Khanam, T., and Rangarajan, P. (2000). Serine 27, a human retinoid x receptor α residue, phosphorylated by protein kinase a is essential for cyclicamp-mediated downregulation of rxrα function. *Biochemical and biophysical research communications*, 279(3) :853–857.

[Harmon et al., 1995] Harmon, M. A., Boehm, M. F., Heyman, R. A., and Mangelsdorf, D. J. (1995). Activation of mammalian retinoid x receptors by the insect growth regulator methoprene. *Proceedings of the National Academy of Sciences*, 92(13) :6157–6160.

[Henzler-Wildman and Kern, 2007] Henzler-Wildman, K. and Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172) :964–972.

[Hernandez et al., 2000] Hernandez, G., Jenney, F. E., Adams, M. W., and LeMaster, D. M. (2000). Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *PNAS*, 97(7) :3166–3170.

[Heyman et al., 1992] Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., Stein, R. B., Eichele, G., Evans, R. M., and Thaller, C. (1992). 9-cis retinoic acid is a high affinity ligand for the retinoid x receptor. *Cell*, 68(2) :397–406.

[Hilser and Freire, 1996] Hilser, V. J. and Freire, E. (1996). Structure-based calculation of the equilibrium folding pathway of proteins. correlation with hydrogen exchange protection factors. *Journal of molecular biology*, 262(5) :756–772.

[Holmbeck et al., 1998] Holmbeck, S. M., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J., and Wright, P. E. (1998). High-resolution solution structure of the retinoid x receptor dna-binding domain. *Journal of molecular biology*, 281(2) :271–284.

[Honda et al., 2013] Honda, M., Yamashita, T., Yamashita, T., Arai, K., Sakai, Y., Sakai, A., Nakamura, M., Mizukoshi, E., and Kaneko, S. (2013). Peretinoin, an acyclic retinoid, improves the hepatic gene signature of chronic hepatitis c following curative therapy of hepatocellular carcinoma. *BMC cancer*, 13(1) :1.

[Hotelling, 1933] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6) :417.

[Hsu et al., 2013] Hsu, Y.-H., Bucher, D., Cao, J., Li, S., Yang, S.-W., Kokotos, G., Woods Jr, V. L., McCammon, J. A., and Dennis, E. A. (2013). Fluoroketone inhibition of ca2+-independent phospholipase a2 through binding pocket association defined by hydrogen/deuterium exchange and molecular dynamics. *Journal of the American Chemical Society*, 135(4) :1330–1337.

[Huber et al., 2003a] Huber, B. R., Desclozeaux, M., West, B. L., Cunha-Lima, S. T., Nguyen, H. T., Baxter, J. D., Ingraham, H. A., and Fletterick, R. J. (2003a). Thyroid hormone receptor-β mutations confer-

ring hormone resistance and reduced corepressor release exhibit decreased stability in the n-terminal ligand-binding domain. *Molecular Endocrinology*, 17(1) :107–116.

[Huber et al., 2003b] Huber, B. R., Sandler, B., West, B. L., Cunha Lima, S. T., Nguyen, H. T., Apriletti, J. W., Baxter, J. D., and Fletterick, R. J. (2003b). Two resistance to thyroid hormone mutants with impaired hormone binding. *Molecular Endocrinology*, 17(4) :643–652.

[Hughes et al., 2012] Hughes, T. S., Chalmers, M. J., Novick, S., Kuruvilla, D. S., Chang, M. R., Kamenecka, T. M., Rance, M., Johnson, B. A., Burris, T. P., Griffin, P. R., et al. (2012). Ligand and receptor dynamics contribute to the mechanism of graded pparγ agonism. *Structure*, 20(1) :139–150.

[Hvidt, 1955] Hvidt, A. (1955). Deuterium exchange between ribonuclease and water. *Biochimica et biophysica acta*, 18 :306–308.

[Hvidt, 1964] Hvidt, A. (1964). A discussion of the ph dependence of the hydrogen-deuterium exchange of proteins. *Comptes-rendus des travaux du Laboratoire Carlsberg*, 34 :299.

[Hvidt et al., 1954] Hvidt, A., Johansen, G., LINDERSTRØM, L. K., and Vaslow, F. (1954). Exchange of deuterium and 18o between water and other substances. i. methods. *Comptes rendus des travaux du Laboratoire Carlsberg. Série chimique*, 29(9) :129.

[Hvidt and Nielsen, 1966] Hvidt, A. and Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Advances in protein chemistry*, 21 :287–386.

[Ichiye and Karplus, 1991] Ichiye, T. and Karplus, M. (1991). Collective motions in proteins : a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins : Structure, Function, and Bioinformatics*, 11(3) :205–217.

[Jacobsen, 2007] Jacobsen, N. E. (2007). *NMR spectroscopy explained : simplified theory, applications and examples for organic chemistry and structural biology*. John Wiley & Sons.

[James, 1998] James, T. L. (1998). Fundamentals of nmr. *Online Textbook : Department of Pharmaceutical Chemistry, University of California, San Francisco*, pages 1–31.

[Jaye et al., 2005] Jaye, M. C., Krawiec, J. A., Campobasso, N., Smallwood, A., Qiu, C., Lu, Q., Kerrigan, J. J., De Los Frailes Alvaro, M., Laffitte, B., Liu, W.-S., et al. (2005). Discovery of substituted maleimides as liver x receptor agonists and determination of a ligand-bound crystal structure. *Journal of medicinal chemistry*, 48(17) :5419–5422.

[Jennings and Wright, 1993] Jennings, P. A. and Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, 262(5135) :892–896.

[Johnson et al., 2000] Johnson, B. A., Wilson, E. M., Li, Y., Moller, D. E., Smith, R. G., and Zhou, G. (2000). Ligand-induced stabilization of pparγ monitored by nmr spectroscopy : implications for nuclear receptor activation. *Journal of molecular biology*, 298(2) :187–194.

[Jones et al., 01] Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy : Open source scientific tools for Python. [Online ; accessed 2016-11-17].

[Kabsch and Sander, 1983] Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure : pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12) :2577–2637.

[Kallenberger et al., 2003] Kallenberger, B. C., Love, J. D., Chatterjee, V. K. K., and Schwabe, J. W. (2003). A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nature Structural & Molecular Biology*, 10(2) :136–140.

[Kappel et al., 2015] Kappel, K., Miao, Y., and McCammon, J. A. (2015). Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Q. Rev. Biophys.*, 48(04) :479–487.

[Karplus and McCammon, 2002] Karplus, M. and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology*, 9(9) :646–652.

[Kato et al., 2016] Kato, A., Itoh, T., Anami, Y., Egawa, D., and Yamamoto, K. (2016). Helix12-stabilization antagonist of vitamin d receptor. *Bioconjugate chemistry*, 27(7) :1750–1761.

[Kendrew et al., 1960] Kendrew, J., Dickerson, R., Strandberg, B., Hart, R., Davies, D., Phillips, D., and Shore, V. (1960). Structure of myoglobin : A three-dimensional fourier synthesis at 2 a. resolution. *Nature*, 185(4711) :422–427.

[Kephart et al., 1996] Kephart, D. D., Walfish, P. G., DeLuca, H., and Butt, T. R. (1996). Retinoid x receptor isotype identity directs human vitamin d receptor heterodimer transactivation from the 24-hydroxylase vitamin d response elements in yeast. *Molecular Endocrinology*, 10(4) :408–419.

[Kersten et al., 1998] Kersten, S., Dong, D., Lee, W.-y., Reczek, P. R., and Noy, N. (1998). Auto-silencing by the retinoid x receptor. *Journal of molecular biology*, 284(1) :21–32.

[Kersten et al., 1995a] Kersten, S., Kelleher, D., Chambon, P., Gronemeyer, H., and Noy, N. (1995a). Retinoid x receptor alpha forms tetramers in solution. *Proceedings of the National Academy of Sciences*, 92(19) :8645–8649.

[Kersten et al., 1995b] Kersten, S., Pan, L., Chambon, P., Gronemeyer, H., and Noy, N. (1995b). Role of ligand in retinoid signaling. 9-cis-retinoic acid modulates the oligomeric state of the retinoid x receptor. *Biochemistry*, 34(42) :13717–13721.

[Kersten et al., 1995c] Kersten, S., Pan, L., and Noy, N. (1995c). On the role of ligand in retinoid signaling : positive cooperativity in the interactions of 9-cis retinoic acid with tetramers of the retinoid x receptor. *Biochemistry*, 34(43) :14263–14269.

[Kersten et al., 1997] Kersten, S., Reczek, P. R., and Noy, N. (1997). The tetramerization region of the retinoid x receptor is important for transcriptional activation by the receptor. *Journal of Biological Chemistry*, 272(47) :29759–29768.

[Kieseritzky et al., 2006] Kieseritzky, G., Morra, G., and Knapp, E.-W. (2006). Stability and fluctuations of amide hydrogen bonds in a bacterial cytochrome c : a molecular dynamics study. *JBIC Journal of Biological Inorganic Chemistry*, 11(1) :26–40.

[Kim et al., 1993] Kim, K. S., Fuchs, J. A., and Woodward, C. K. (1993). Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry*, 32(37) :9600–9608.

[Kitareewan et al., 1996] Kitareewan, S., Burka, L., Tomer, K., Parker, C., Deterding, L., Stevens, R., Forman, B., Mais, D., Heyman, R., McMorris, T., et al. (1996). Phytol metabolites are circulating dietary factors that activate the nuclear receptor rxr. *Molecular Biology of the Cell*, 7(8) :1153–1166.

[Klaholz et al., 1998] Klaholz, B., Renaud, J.-P., Mitschler, A., Zusi, C., Chambon, P., Gronemeyer, H., and Moras, D. (1998). Conformational adaptation of agonists to the human nuclear receptor rarγ. *Nature Structural & Molecular Biology*, 5(3) :199–202.

[Koenig et al., 1987] Koenig, R. J., Brent, G. A., Warne, R. L., Larsen, P. R., and Moore, D. D. (1987). Thyroid hormone receptor binds to a site in the rat growth hormone promoter required for induction by thyroid hormone. *Proceedings of the National Academy of Sciences*, 84(16) :5670–5674.

[Kojetin et al., 2015] Kojetin, D. J., Matta-Camacho, E., Hughes, T. S., Srinivasan, S., Nwachukwu, J. C., Cavett, V., Nowak, J., Chalmers, M. J., Marciano, D. P., Kamenecka, T. M., et al. (2015). Structural mechanism for signal transduction in rxr nuclear receptor heterodimers. *Nature communications*, 6.

[Konermann et al., 2011] Konermann, L., Pan, J., and Liu, Y.-H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews*, 40(3) :1224–1234.

[Konermann et al., 2013] Konermann, L., Vahidi, S., and Sowole, M. A. (2013). Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Analytical chemistry*, 86(1) :213–232.

[Kotani et al., 2010] Kotani, H., Tanabe, H., Mizukami, H., Makishima, M., and Inoue, M. (2010). Identification of a naturally occurring rexinoid, honokiol, that activates the retinoid x receptor. *Journal of natural products*, 73(8) :1332–1336.

[Krishna et al., 2004] Krishna, M. M., Hoang, L., Lin, Y., and Englander, S. W. (2004). Hydrogen exchange methods to study protein folding. *Methods*, 34(1) :51–64.

[Kruskal, 1964a] Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1) :1–27.

[Kruskal, 1964b] Kruskal, J. B. (1964b). Nonmetric multidimensional scaling : a numerical method. *Psychometrika*, 29(2) :115–129.

[Kumar and McEwan, 2012] Kumar, R. and McEwan, I. J. (2012). Allosteric modulators of steroid hormone receptors : structural dynamics and gene regulation. *Endocrine reviews*, 33(2) :271–299.

[Kurokawa et al., 1994] Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M. G., Heyman, R. A., Glass, C. K., et al. (1994). Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature*, 371(6497) :528–531.

[Kurokawa et al., 1993] Kurokawa, R., Yu, V., Näär, A., Kyakumoto, S., Han, Z., Silverman, S., Rosenfeld, M., and Glass, C. (1993). Differential orientations of the dna-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. *Genes & development*, 7(7b) :1423–1435.

[Lala et al., 1996] Lala, D. S., Mukherjee, R., Schulman, I. G., Koch, S. S. C., et al. (1996). Activation of specific rxr heterodimers by an antagonist of rxr homodimers. *Nature*, 383(6599) :450.

[Lammi et al., 2008] Lammi, J., Perlmann, T., and Aarnisalo, P. (2008). Corepressor interaction differentiates the permissive and non-permissive retinoid x receptor heterodimers. *Archives of biochemistry and biophysics*, 472(2) :105–114.

[Lavery and Mcewan, 2005] Lavery, D. N. and Mcewan, I. J. (2005). Structure and function of steroid receptor af1 transactivation domains : induction of active conformations. *Biochemical Journal*, 391(3) :449–464.

[Leach, 2001] Leach, A. R. (2001). *Molecular modelling : principles and applications*. Pearson education.

[Leblanc and Stunnenberg, 1995] Leblanc, B. P. and Stunnenberg, H. G. (1995). 9-cis retinoic acid signaling : changing partners causes some excitement. *Genes and Development*, 9(15) :1811–1816.

[Lee et al., 2000] Lee, H.-Y., Suh, Y.-A., Robinson, M. J., Clifford, J. L., Hong, W. K., Woodgett, J. R., Cobb, M. H., Mangelsdorf, D. J., and Kurie, J. M. (2000). Stress pathway activation induces phosphorylation of retinoid x receptor. *Journal of Biological Chemistry*, 275(41) :32193–32199.

[Lee et al., 1993] Lee, M. S., Kliewer, S. A., Provencal, J., Wright, P. E., and Evans, R. M. (1993). Structure of the retinoid x receptor alpha dna binding domain : a helix required for homodimeric dna binding. *Science*, 260(5111) :1117–1121.

[Lee et al., 2004] Lee, T., Hoofnagle, A. N., Kabuyama, Y., Stroud, J., Min, X., Goldsmith, E. J., Chen, L., Resing, K. A., and Ahn, N. G. (2004). Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Mol Cell*, 14(1) :43–55.

[Lefebvre et al., 2010] Lefebvre, P., Benomar, Y., and Staels, B. (2010). Retinoid x receptors : common heterodimerization partners with distinct functions. *Trends in Endocrinology & Metabolism*, 21(11) :676–683.

[Lehmann et al., 1997] Lehmann, J. M., Kliewer, S. A., Moore, L. B., Smith-Oliver, T. A., Oliver, B. B., Su, J.-L., Sundseth, S. S., Winegar, D. A., Blanchard, D. E., Spencer, T. A., et al. (1997). Activation of the nuclear receptor lxr by oxysterols defines a new hormone response pathway. *Journal of Biological Chemistry*, 272(6) :3137–3140.

[Lehmann et al., 1995] Lehmann, J. M., Moore, L. B., Smith-Oliver, T. A., Wilkison, W. O., Willson, T. M., and Kliewer, S. A. (1995). An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor γ (pparγ). *Journal of Biological Chemistry*, 270(22) :12953–12956.

[Leibowitz et al., 2006] Leibowitz, M. D., Ardecky, R. J., Boehm, M. F., Broderick, C. L., Carfagna, M. A., Crombie, D. L., D'Arrigo, J., Etgen, G. J., Faul, M. M., Grese, T. A., et al. (2006). Biological characterization of a heterodimer-selective retinoid x receptor modulator : potential benefits for the treatment of type 2 diabetes. *Endocrinology*, 147(2) :1044–1053.

[Leid et al., 1992] Leid, M., Kastner, P., and Chambon, P. (1992). Multiplicity generates diversity in the retinoic acid signalling pathways. *Trends in biochemical sciences*, 17(10) :427–433.

[LeMaster et al., 2009] LeMaster, D. M., Anderson, J. S., and Hernández, G. (2009). Peptide conformer acidity analysis of protein flexibility monitored by hydrogen exchange. *Biochemistry*, 48(39) :9256–9265.

[Lemotte et al., 1996] Lemotte, P. K., Keidel, S., and Apfel, C. M. (1996). Phytanic acid is a retinoid x receptor ligand. *European Journal of Biochemistry*, 236(1) :328–333.

[Leng et al., 1995] Leng, X., Blanco, J., Tsai, S. Y., Ozato, K., O'Malley, B. W., and Tsai, M.-J. (1995). Mouse retinoid x receptor contains a separable ligand-binding and transactivation domain in its e region. *Molecular and cellular biology*, 15(1) :255–263.

[Lenhard et al., 2000] Lenhard, J. M., Croom, D. K., Weiel, J. E., and Winegar, D. A. (2000). Hiv protease inhibitors stimulate hepatic triglyceride synthesis. *Arteriosclerosis, thrombosis, and vascular biology*, 20(12) :2625–2629.

[Li et al., 2005a] Li, H., Robertson, A. D., and Jensen, J. H. (2005a). Very fast empirical prediction and rationalization of protein pka values. *Proteins : Structure, Function, and Bioinformatics*, 61(4) :704–721.

[Li and Woodward, 1999] Li, R. and Woodward, C. (1999). The hydrogen exchange core and protein folding. *Protein Sci.*, 8(8) :1571–1590.

[Li et al., 2005b] Li, X., Hansen, P. A., Xi, L., Chandraratna, R. A., and Burant, C. F. (2005b). Distinct mechanisms of glucose lowering by specific agonists for peroxisomal proliferator activated receptor γ and retinoic acid x receptors. *Journal of Biological Chemistry*, 280(46) :38317–38327.

[Liberato et al., 2012] Liberato, M. V., Nascimento, A. S., Ayers, S. D., Lin, J. Z., Cvoro, A., Silveira, R. L., Martínez, L., Souza, P. C., Saidemberg, D., Deng, T., et al. (2012). Medium chain fatty acids are selective peroxisome proliferator activated receptor (ppar) γ activators and pan-ppar partial agonists. *PLoS One*, 7(5) :e36297.

[Linderstrom-Lang, 1955] Linderstrom-Lang, K. (1955). Symp. on peptide chemistry. *Chem. Soc.,(London), Spec. Publ*, 2 :1.

[Linderstrøm-Lang, 1958] Linderstrøm-Lang, K. (1958). Deuterium exchange and protein structure. In *Symposium on protein structure*, pages 23–34. Methuen London.

[Liu et al., 2002] Liu, S., Ogilvie, K. M., Klausing, K., Lawson, M. A., Jolley, D., Li, D., Bilakovics, J., Pascual, B., Hein, N., Urcan, M., et al. (2002). Mechanism of selective retinoid x receptor agonist-induced hypothyroidism in the rat. *Endocrinology*, 143(8) :2880–2885.

[Liu et al., 2012] Liu, T., Pantazatos, D., Li, S., Hamuro, Y., Hilser, V. J., and Woods Jr, V. L. (2012). Quantitative assessment of protein structural models by comparison of h/d exchange ms data with exchange

behavior accurately predicted by dxcorex. *Journal of the American Society for Mass Spectrometry*, 23(1) :43–56.

[Lou et al., 2014] Lou, X., Toresson, G., Benod, C., Suh, J. H., Philips, K. J., Webb, P., and Gustafsson, J.-A. (2014). Structure of the retinoid x receptor α–liver x receptor β (rxrα–lxrβ) heterodimer on dna. *Nature structural & molecular biology*, 21(3) :277–281.

[Lu et al., 2008] Lu, J., Chen, M., Stanley, S. E., and Li, E. (2008). Effect of heterodimer partner rxrα on pparγ activation function-2 helix in solution. *Biochemical and biophysical research communications*, 365(1) :42–46.

[Lu et al., 2006] Lu, J., Cistola, D. P., and Li, E. (2006). Analysis of ligand binding and protein dynamics of human retinoid x receptor alpha ligand-binding domain by nuclear magnetic resonance. *Biochemistry*, 45(6) :1629–1639.

[Lu et al., 2009] Lu, J., Dawson, M. I., Hu, Q. Y., Xia, Z., Dambacher, J. D., Ye, M., Zhang, X.-K., and Li, E. (2009). The effect of antagonists on the conformational exchange of the retinoid x receptor alpha ligand-binding domain. *Magnetic Resonance in Chemistry*, 47(12) :1071–1080.

[Lu et al., 2007] Lu, X., Wintrode, P. L., and Surewicz, W. K. (2007). β-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proceedings of the National Academy of Sciences*, 104(5) :1510–1515.

[Luck et al., 2000] Luck, L. A., Barse, J. L., Luck, A. M., and Peck, C. H. (2000). Conformational changes in the human estrogen receptor observed by 19 f nmr. *Biochemical and biophysical research communications*, 270(3) :988–991.

[Lührs et al., 2005] Lührs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Döbeli, H., Schubert, D., and Riek, R. (2005). 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *PNAS*, 102(48) :17342–17347.

[Lumry and Rosenberg, 1975] Lumry, R. and Rosenberg, A. (1975). Mobile-defect hypothesis of protein function. Technical report, Minnesota Univ., Minneapolis (USA).

[Macoritto et al., 2008] Macoritto, M., Nguyen-Yamamoto, L., Huang, D. C., Samuel, S., Yang, X. F., Wang, T. T., White, J. H., and Kremer, R. (2008). Phosphorylation of the human retinoid x receptor α at serine 260 impairs coactivator (s) recruitment and induces hormone resistance to multiple ligands. *Journal of Biological Chemistry*, 283(8) :4943–4956.

[Mandell et al., 1998] Mandell, J. G., Falick, A. M., and Komives, E. A. (1998). Identification of protein-protein interfaces by decreased amide proton solvent accessibility. *PNAS*, 95(25) :14705–14710.

[Mangelsdorf et al., 1992] Mangelsdorf, D. J., Borgmeyer, U., Heyman, R. A., Zhou, J. Y., Ong, E. S., Oro, A. E., Kakizuka, A., and Evans, R. M. (1992). Characterization of three rxr genes that mediate the action of 9-cis retinoic acid. *Genes & development*, 6(3) :329–344.

[Mangelsdorf et al., 1990] Mangelsdorf, D. J., Ong, E. S., Dyck, J. A., and Evans, R. M. (1990). Nuclear receptor that identifies a novel retinoic acid response pathway. *Nature*, 345(6272) :224–229.

[Mangelsdorf et al., 1991] Mangelsdorf, D. J., Umesono, K., Kliewer, S. A., Borgmeyer, U., Ong, E. S., and Evans, R. M. (1991). A direct repeat in the cellular retinol-binding protein type ii gene confers differential regulation by rxr and rar. *Cell*, 66(3) :555–561.

[Mann et al., 2005] Mann, K. K., Padovani, A. M., Guo, Q., Colosimo, A. L., Lee, H.-Y., Kurie, J. M., and Miller Jr, W. H. (2005). Arsenic trioxide inhibits nuclear receptor function via sek1/jnk-mediated rxrα phosphorylation. *The Journal of clinical investigation*, 115(10) :2924–2933.

[Marciano et al., 2015] Marciano, D. P., Kuruvilla, D. S., Pascal, B. D., and Griffin, P. R. (2015). Identification of bexarotene as a ppar antagonist with hdx. *PPAR research*, 2015.

[Markwick et al., 2007] Markwick, P. R., Bouvignies, G., and Blackledge, M. (2007). Exploring multiple timescale motions in protein gb3 using accelerated molecular dynamics and nmr spectroscopy. *Journal of the American Chemical Society*, 129(15) :4724–4730.

[Markwick and McCammon, 2011a] Markwick, P. R. and McCammon, J. A. (2011a). Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Physical Chemistry Chemical Physics*, 13(45) :20053–20065.

[Markwick et al., 2009] Markwick, P. R. L., Bouvignies, G., Salmon, L., McCammon, J. A., Nilges, M., and Blackledge, M. (2009). Toward a Unified Representation of Protein Structural Dynamics in Solution. *Journal of the American Chemical Society*, 131(46) :16968–16975.

[Markwick et al., 2010] Markwick, P. R. L., Cervantes, C. F., Abel, B. L., Komives, E. A., Blackledge, M., and McCammon, J. A. (2010). Enhanced Conformational Space Sampling Improves the Prediction of Chemical Shifts in Proteins. *Journal of the American Chemical Society*, 132(4) :1220–1221.

[Markwick and McCammon, 2011b] Markwick, P. R. L. and McCammon, J. A. (2011b). Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Physical Chemistry Chemical Physics*, 13(45) :20053–20065.

[Markwick et al., 2011] Markwick, P. R. L., Pierce, L. C. T., Goodin, D. B., and McCammon, J. A. (2011). Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam. *J. Phys. Chem. Lett.*, 2(3) :158–164.

[Martin, 2008] Martin, E. (2008). *Etude in silico des interfaces protéines-protéines et protéines-ligands : Application à plusieurs systèmes impliqués dans des cancers*. PhD thesis, Université Louis Pasteur (Strasbourg).

[Matsushima-Nishiwaki et al., 2001] Matsushima-Nishiwaki, R., Okuno, M., Adachi, S., Sano, T., Akita, K., Moriwaki, H., Friedman, S. L., and Kojima, S. (2001). Phosphorylation of retinoid x receptor α at serine 260 impairs its metabolism and function in human hepatocellular carcinoma. *Cancer research*, 61(20) :7675–7682.

[Matsushima-Nishiwaki et al., 2003] Matsushima-Nishiwaki, R., Okuno, M., Takano, Y., Kojima, S., Friedman, S. L., and Moriwaki, H. (2003). Molecular mechanism for growth suppression of human hepatocellular carcinoma cells by acyclic retinoid. *Carcinogenesis*, 24(8) :1353–1359.

[Matthew and Richards, 1983] Matthew, J. B. and Richards, F. M. (1983). The ph dependence of hydrogen exchange in proteins. *Journal of Biological Chemistry*, 258(5) :3039–3044.

[Mayo and Baldwin, 1993] Mayo, S. L. and Baldwin, R. L. (1993). Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science*, 262(5135) :873–876.

[McAllister and Konermann, 2015] McAllister, R. G. and Konermann, L. (2015). Challenges in the interpretation of protein h/d exchange data : a molecular dynamics simulation perspective. *Biochemistry*, 54(16) :2683–2692.

[McDonnell et al., 1992] McDonnell, D. P., Vegeto, E., and O'Malley, B. W. (1992). Identification of a negative regulatory function for steroid receptors. *Proceedings of the National Academy of Sciences*, 89(22) :10563–10567.

[McGowan and Hamelberg, 2013] McGowan, L. C. and Hamelberg, D. (2013). Conformational Plasticity of an Enzyme during Catalysis : Intricate Coupling between Cyclophilin A Dynamics and Substrate Turnover. *Biophys. J.*, 104(1) :216–226.

[Miao et al., 2016] Miao, Y., Baudry, J., Smith, J. C., and McCammon, J. A. (2016). General trends of dihedral conformational transitions in a globular protein. *Proteins : Structure, Function, and Bioinformatics*.

[Miao et al., 2015a] Miao, Y., Caliman, A. D., and McCammon, J. A. (2015a). Allosteric Effects of Sodium Ion Binding on Activation of the M3 Muscarinic G-Protein-Coupled Receptor. *Biophys. J.*, 108(7) :1796–1806.

[Miao et al., 2015b] Miao, Y., Feher, V. A., and McCammon, J. A. (2015b). Gaussian accelerated molecular dynamics : Unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation*, 11(8) :3584–3595.

[Miao et al., 2015c] Miao, Y., Feixas, F., Eun, C., and McCammon, J. A. (2015c). Accelerated molecular dynamics simulations of protein folding. *J Comput Chem*, 36(20) :1536–1549.

[Miao and McCammon, 2016] Miao, Y. and McCammon, J. A. (2016). Unconstrained enhanced sampling for free energy calculations of biomolecules : a review. *Molecular Simulation*, 42(13) :1046–1055.

[Miao et al., 2013] Miao, Y., Nichols, S. E., Gasper, P. M., Metzger, V. T., and McCammon, J. A. (2013). Activation and dynamic network of the m2 muscarinic receptor. *Proceedings of the National Academy of Sciences*, 110(27) :10982–10987.

[Miao et al., 2014a] Miao, Y., Nichols, S. E., and McCammon, J. A. (2014a). Free energy landscape of g-protein coupled receptors, explored by accelerated molecular dynamics. *Physical Chemistry Chemical Physics*, 16(14) :6398–6406.

[Miao et al., 2014b] Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., and McCammon, J. A. (2014b). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of chemical theory and computation*, 10(7) :2677–2689.

[Michaud-Agrawal et al., 2011] Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). Mdanalysis : a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, 32(10) :2319–2327.

[Michiels et al., 2010] Michiels, P., Atkins, K., Ludwig, C., Whittaker, S., van Dongen, M., and Günther, U. (2010). Assignment of the orphan nuclear receptor nurr1 by nmr. *Biomolecular NMR assignments*, 4(1) :101–105.

[Milne et al., 1998] Milne, J. S., Mayne, L., Roder, H., Wand, A. J., and Englander, S. W. (1998). Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Sci.*, 7(3) :739–745.

[Molday et al., 1972] Molday, R., Englander, S., and Kallen, R. (1972). Primary structure effects on peptide group hydrogen exchange. *Biochemistry*, 11(2) :150–158.

[Mu et al., 2005] Mu, Y., Nguyen, P. H., and Stock, G. (2005). Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins : Structure, Function, and Bioinformatics*, 58(1) :45–52.

[Mukherjee et al., 1997] Mukherjee, R., Davies, P. J., Crombie, D. L., Bischoff, E. D., et al. (1997). Sensitization of diabetic and obese mice to insulin by retinoid x receptor agonists. *Nature*, 386(6623) :407.

[Näär et al., 1991] Näär, A. M., Boutin, J.-M., Lipkin, S. M., Victor, C. Y., Holloway, J. M., Glass, C. K., and Rosenfeld, M. G. (1991). The orientation and spacing of core dna-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell*, 65(7) :1267–1279.

[Nahoum et al., 2007] Nahoum, V., Pérez, E., Germain, P., Rodriguez-Barrios, F., Manzo, F., Kammerer, S., Lemaire, G., Hirsch, O., Royer, C. A., Gronemeyer, H., et al. (2007). Modulators of the structural dynamics of the retinoid x receptor to reveal receptor function. *Proceedings of the National Academy of Sciences*, 104(44) :17323–17328.

[Narvaez et al., 1996] Narvaez, C. J., Vanweelden, K., Byrne, I., and Welsh, J. (1996). Characterization of a vitamin d3-resistant mcf-7 cell line. *Endocrinology*, 137(2) :400–409.

[Neal et al., 2003] Neal, S., Nip, A. M., Zhang, H., and Wishart, D. S. (2003). Rapid and accurate calculation of protein 1h, 13c and 15n chemical shifts. *Journal of biomolecular NMR*, 26(3) :215–240.

[Nolte et al., 1998] Nolte, R. T., Wisely, G. B., Westin, S., Cobb, J. E., Lambert, M. H., Kurokawa, R., Rosenfeld, M. G., Willson, T. M., Glass, C. K., and Milburn, M. V. (1998). Ligand binding and coactivator assembly of the peroxisome proliferator-activated receptor-γ. *Nature*, 395(6698) :137–143.

[Onuchic and Wolynes, 2004] Onuchic, J. N. and Wolynes, P. G. (2004). Theory of protein folding. *Current opinion in structural biology*, 14(1) :70–75.

[Orlov et al., 2012] Orlov, I., Rochel, N., Moras, D., and Klaholz, B. P. (2012). Structure of the full human rxr/vdr nuclear receptor heterodimer complex with its dr3 target dna. *The EMBO journal*, 31(2) :291–300.

[Ortiz-Sánchez et al., 2012] Ortiz-Sánchez, J. M., Bucher, D., Pierce, L. C. T., Markwick, P. R. L., and McCammon, J. A. (2012). Exploring the Photophysical Properties of Molecular Systems Using Excited State Accelerated ab InitioMolecular Dynamics. *J Chem Theory Comput*, 8(8) :2752–2761.

[Osz et al., 2015] Osz, J., McEwen, A. G., Poussin-Courmontagne, P., Moutier, E., Birck, C., Davidson, I., Moras, D., and Rochel, N. (2015). Structural basis of natural promoter recognition by the retinoid x nuclear receptor. *Scientific reports*, 5.

[Park et al., 2015] Park, I.-H., Venable, J. D., Steckler, C., Cellitti, S. E., Lesley, S. A., Spraggon, G., and Brock, A. (2015). Estimation of hydrogen-exchange protection factors from md simulation based on amide hydrogen bonding analysis. *Journal of chemical information and modeling*, 55(9) :1914–1925.

[Pauling and Corey, 1951] Pauling, L. and Corey, R. B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences*, 37(5) :251–256.

[Pauling et al., 1951] Pauling, L., Corey, R. B., and Branson, H. R. (1951). The structure of proteins : two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4) :205–211.

[Pearson, 1901] Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11) :559–572.

[Pérez et al., 2012] Pérez, E., Bourguet, W., Gronemeyer, H., and de Lera, A. R. (2012). Modulation of rxr function through ligand design. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1821(1) :57–69.

[Perlmann et al., 1993] Perlmann, T., Rangarajan, P. N., Umesono, K., and Evans, R. M. (1993). Determinants for selective rar and tr recognition of direct repeat hres. *Genes & development*, 7(7b) :1411–1422.

[Perrin and Lollo, 1984] Perrin, C. L. and Lollo, C. P. (1984). Mechanism of NH proton exchange in amides and proteins : solvent effects and solvent accessibility. *Journal of the American Chemical Society*, 106(10) :2754–2757.

[Persson and Halle, 2015] Persson, F. and Halle, B. (2015). How amide hydrogens exchange in native proteins. *Proceedings of the National Academy of Sciences*, 112(33) :10383–10388.

[Perutz et al., 1960] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., and Will, G. (1960). Structure of hæmoglobin : a three-dimensional fourier synthesis at 5.5-å. resolution, obtained by x-ray analysis. *Nature*, 185 :416–422.

[Petruk et al., 2012] Petruk, A. A., Defelipe, L. A., Rodríguez Limardo, R. G., Bucci, H., Marti, M. A., and Turjanski, A. G. (2012). Molecular dynamics simulations provide atomistic insight into hydrogen exchange mass spectrometry experiments. *Journal of chemical theory and computation*, 9(1) :658–669.

[Phillips et al., 2005] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16) :1781–1802.

[Piana et al., 2013] Piana, S., Lindorff-Larsen, K., and Shaw, D. E. (2013). Atomic-level description of ubiquitin folding. *Proceedings of the National Academy of Sciences*, 110(15) :5915–5920.

[Pierce et al., 2012] Pierce, L. C., Salomon-Ferrer, R., Augusto F. de Oliveira, C., McCammon, J. A., and Walker, R. C. (2012). Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of chemical theory and computation*, 8(9) :2997–3002.

[Pierce et al., 2011] Pierce, L. C. T., Markwick, P. R. L., McCammon, J. A., and Doltsinis, N. L. (2011). Accelerating chemical reactions : Exploring reactive free-energy surfaces using accelerated ab initio molecular dynamics. *The Journal of Chemical Physics*, 134(17) :174107–13.

[Pierrot et al., 2015] Pierrot, N., Lhommel, R., Quenon, L., Hanseeuw, B., Dricot, L., Sindic, C., Maloteaux, J.-M., Octavea, J.-N., and Ivanoiu, A. (2015). Targretin improves cognitive and biological markers in a patient with alzheimer's disease. *Journal of Alzheimer's Disease*, 49(2) :271–276.

[Pissios et al., 2000] Pissios, P., Tzameli, I., Kushner, P. J., and Moore, D. D. (2000). Dynamic stabilization of nuclear receptor ligand binding domains by hormone or corepressor binding. *Molecular cell*, 6(2) :245–253.

[Proctor and Yu, 1950] Proctor, W. and Yu, F. (1950). The dependence of a nuclear magnetic resonance frequency upon chemical compound. *Physical Review*, 77(5) :717.

[Prüfer and Barsony, 2002] Prüfer, K. and Barsony, J. (2002). Retinoid x receptor dominates the nuclear import and export of the unliganded vitamin d receptor. *Molecular Endocrinology*, 16(8) :1738–1751.

[Putcha et al., 2012] Putcha, B.-D. K., Wright, E., Brunzelle, J. S., and Fernandez, E. J. (2012). Structural basis for negative cooperativity within agonist-bound tr : Rxr heterodimers. *Proceedings of the National Academy of Sciences*, 109(16) :6084–6087.

[Radou et al., 2014] Radou, G., Dreyer, F. N., Tuma, R., and Paci, E. (2014). Functional dynamics of hexameric helicase probed by hydrogen exchange and simulation. *Biophysical journal*, 107(4) :983–990.

[Rahman and Tully, 2002] Rahman, J. A. and Tully, J. C. (2002). Puddle-skimming : An efficient sampling of multidimensional configuration space. *The Journal of chemical physics*, 116(20) :8750–8760.

[Rastinejad et al., 1995] Rastinejad, F., Perlmann, T., Evans, R. M., and Sigler, P. B. (1995). Structural determinants of nuclear receptor assembly on dna direct repeats. *Nature*, 375(6528) :203–211.

[Rastinejad et al., 2000] Rastinejad, F., Wagner, T., Zhao, Q., and Khorasanizadeh, S. (2000). Structure of the rxr–rar dna-binding complex on the retinoic acid response element dr1. *The EMBO journal*, 19(5) :1045–1054.

[Renaud et al., 1995] Renaud, J.-P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the rar-γ ligand-binding domain bound to all-trans retinoic acid. *Nature*, 378(6558) :681–689.

[Richards, 1979] Richards, F. M. (1979). Packing defects, cavities, volume fluctuations, and access to the interior of proteins. including some general comments on surface area and protein structure. *Carlsberg Research Communications*, 44(2) :47–63.

[Rochel et al., 2011] Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y., et al. (2011). Common architecture of nuclear receptor heterodimers on dna direct repeat elements with different spacings. *Nature structural & molecular biology*, 18(5) :564–570.

[Rochette-Egly, 2003] Rochette-Egly, C. (2003). Nuclear receptors : integration of multiple signalling pathways through phosphorylation. *Cellular signalling*, 15(4) :355–366.

[Roe et al., 2014] Roe, D. R., Bergonzo, C., and Cheatham III, T. E. (2014). Evaluation of enhanced sampling provided by accelerated molecular dynamics with hamiltonian replica exchange methods. *The Journal of Physical Chemistry B*, 118(13) :3543–3552.

[Ross et al., 2000] Ross, S. A., McCaffery, P. J., Drager, U. C., and De Luca, L. M. (2000). Retinoids in embryonal development. *Physiological reviews*, 80(3) :1021–1054.

[Rowe et al., 1991] Rowe, A., Eager, N., and Brickell, P. (1991). A member of the rxr nuclear receptor family is expressed in neural-crest-derived cells of the developing chick peripheral nervous system. *Development*, 111(3) :771–778.

[Ryckaert et al., 1977] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3) :327–341.

[Sammon, 1969] Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on computers*, 18(5) :401–409.

[Santin et al., 2009] Santin, E. P., Germain, P., Quillard, F., Khanwalkar, H., Rodriguez-Barrios, F., Gronemeyer, H., de Lera, A. R., and Bourguet, W. (2009). Modulating retinoid x receptor with a series of

(e)-3-[4-hydroxy-3-(3-alkoxy-5, 5, 8, 8-tetramethyl-5, 6, 7, 8-tetrahydronaphthalen-2-yl) phenyl] acrylic acids and their 4-alkoxy isomers. *Journal of medicinal chemistry*, 52(10) :3150–3158.

[Sato et al., 2010] Sato, Y., Ramalanjaona, N., Huet, T., Potier, N., Osz, J., Antony, P., Peluso-Iltis, C., Poussin-Courmontagne, P., Ennifar, E., Mély, Y., et al. (2010). The "phantom effect" of the rexinoid lg100754 : structural and functional insights. *PloS one*, 5(11) :e15119.

[Schlick, 2010] Schlick, T. (2010). *Molecular modeling and simulation : an interdisciplinary guide : an interdisciplinary guide*, volume 21. Springer Science & Business Media.

[Schrödinger, LLC, 2015] Schrödinger, LLC (2015). The PyMOL molecular graphics system, version 1.8.

[Seavey et al., 1991] Seavey, B. R., Farr, E. A., Westler, W. M., and Markley, J. L. (1991). A relational database for sequence-specific protein nmr data. *Journal of biomolecular NMR*, 1(3) :217–236.

[Sebag et al., 1992] Sebag, M., Henderson, J., Rhim, J., and Kremer, R. (1992). Relative resistance to 1, 25-dihydroxyvitamin d3 in a keratinocyte model of tumor progression. *Journal of Biological Chemistry*, 267(17) :12162–12167.

[Sheinblatt, 1970] Sheinblatt, M. (1970). Determination of an acidity scale for peptide hydrogens from nuclear magnetic resonance kinetic studies. *Journal of the American Chemical Society*, 92(8) :2505–2509.

[Shen and Hamelberg, 2008] Shen, T. and Hamelberg, D. (2008). A statistical analysis of the precision of reweighting-based simulations. *The Journal of chemical physics*, 129(3) :034103.

[Shepard, 1962] Shepard, R. N. (1962). The analysis of proximities : Multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27(2) :125–140.

[Sherman et al., 1999] Sherman, S. I., Gopal, J., Haugen, B. R., Chiu, A. C., Whaley, K., Nowlakha, P., and Duvic, M. (1999). Central hypothyroidism associated with retinoid x receptor–selective ligands. *New England Journal of Medicine*, 340(14) :1075–1079.

[Shimizu et al., 2009] Shimizu, M., Takai, K., and Moriwaki, H. (2009). Strategy and mechanism for the prevention of hepatocellular carcinoma : Phosphorylated retinoid x receptor α is a critical target for hepatocellular carcinoma chemoprevention. *Cancer science*, 100(3) :369–374.

[Shulman and Mangelsdorf, 2005] Shulman, A. I. and Mangelsdorf, D. J. (2005). Retinoid x receptor heterodimers in the metabolic syndrome. *New England Journal of Medicine*, 353(6) :604–615.

[Singarapu et al., 2011] Singarapu, K. K., Zhu, J., Tonelli, M., Rao, H., Assadi-Porter, F. M., Westler, W. M., DeLuca, H. F., and Markley, J. L. (2011). Ligand-specific structural changes in the vitamin d receptor in solution. *Biochemistry*, 50(51) :11025–11033.

[Sinko et al., 2012] Sinko, W., de Oliveira, C. A. F., Pierce, L. C. T., and McCammon, J. A. (2012). Protecting High Energy Barriers : A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations. *J Chem Theory Comput*, 8(1) :17–23.

[Skinner et al., 2012a] Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E., and Englander, S. W. (2012a). Protein dynamics viewed by hydrogen exchange. *Protein Science*, 21(7) :996–1005.

[Skinner et al., 2012b] Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E., and Englander, S. W. (2012b). Protein hydrogen exchange : Testing current models. *Protein Science*, 21(7) :987–995.

[So et al., 2000] So, M. R., Voter, A. F., et al. (2000). Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics*, 112(21) :9599–9606.

[Solomon et al., 1998] Solomon, C., Sebag, M., White, J. H., Rhim, J., and Kremer, R. (1998). Disruption of vitamin d receptor-retinoid x receptor heterodimer formation following ras transformation of human keratinocytes. *Journal of Biological Chemistry*, 273(28) :17573–17578.

[Solomon et al., 1999] Solomon, C., White, J. H., and Kremer, R. (1999). Mitogen-activated protein kinase inhibits 1, 25-dihydroxyvitamin d 3–dependent signal transduction by phosphorylating human retinoid x receptor α. *The Journal of clinical investigation*, 103(12) :1729–1735.

[Spiwok et al., 2015] Spiwok, V., Sucur, Z., and Hosek, P. (2015). Enhanced sampling techniques in biomolecular simulations. *Biotechnology advances*, 33(6) :1130–1140.

[Steiner et al., 1998] Steiner, M., Genilloud, P.-A., and Wilkins, J. (1998). Simple bias potential for boosting molecular dynamics with the hyperdynamics scheme. *Physical Review B*, 57(17) :10236.

[Still et al., 1990] Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, 112(16) :6127–6129.

[Suino et al., 2004] Suino, K., Peng, L., Reynolds, R., Li, Y., Cha, J.-Y., Repa, J. J., Kliewer, S. A., and Xu, H. E. (2004). The nuclear xenobiotic receptor car : structural determinants of constitutive activation and heterodimerization. *Molecular cell*, 16(6) :893–905.

[Széles et al., 2010] Széles, L., Póliska, S., Nagy, G., Szatmari, I., Szanto, A., Pap, A., Lindstedt, M., Santegoets, S. J., Rühl, R., Dezsö, B., et al. (2010). Research resource : transcriptome profiling of

genes regulated by rxr and its permissive and nonpermissive partners in differentiating monocyte-derived dendritic cells. *Molecular endocrinology*, 24(11) :2218–2231.

[Tamrazi et al., 2003] Tamrazi, A., Carlson, K. E., and Katzenellenbogen, J. A. (2003). Molecular sensors of estrogen receptor conformations and dynamics. *Molecular Endocrinology*, 17(12) :2593–2602.

[Tamrazi et al., 2005] Tamrazi, A., Carlson, K. E., Rodriguez, A. L., and Katzenellenbogen, J. A. (2005). Coactivator proteins as determinants of estrogen receptor structure and function : spectroscopic evidence for a novel coactivator-stabilized receptor conformation. *Molecular Endocrinology*, 19(6) :1516–1528.

[Tarrade et al., 2005] Tarrade, A., Bastien, J., Bruck, N., Bauer, A., Gianni, M., and Rochette-Egly, C. (2005). Retinoic acid and arsenic trioxide cooperate for apoptosis through phosphorylated rxr alpha. *Oncogene*, 24(14) :2277–2288.

[Tenenbaum et al., 2000] Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500) :2319–2323.

[Tesseur and De Strooper, 2013] Tesseur, I. and De Strooper, B. (2013). When the dust settles : what did we learn from the bexarotene discussion ? *Alzheimer's research & therapy*, 5(6) :1.

[Tikhonova et al., 2013] Tikhonova, I. G., Selvam, B., Ivetac, A., Wereszczynski, J., and McCammon, J. A. (2013). Simulations of Biased Agonists in the β 2Adrenergic Receptor with Accelerated Molecular Dynamics. *Biochemistry*, 52(33) :5593–5603.

[Torgerson, 1952] Torgerson, W. S. (1952). Multidimensional scaling : I. theory and method. *Psychometrika*, 17(4) :401–419.

[Tozzini, 2005] Tozzini, V. (2005). Coarse-grained models for proteins. *Current opinion in structural biology*, 15(2) :144–150.

[Umesono et al., 1991] Umesono, K., Murakami, K. K., Thompson, C. C., and Evans, R. M. (1991). Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin d 3 receptors. *Cell*, 65(7) :1255–1266.

[Uppenberg et al., 1998] Uppenberg, J., Svensson, C., Jaki, M., Bertilsson, G., Jendeberg, L., and Berkenstam, A. (1998). Crystal structure of the ligand binding domain of the human nuclear receptor pparγ. *Journal of Biological Chemistry*, 273(47) :31108–31112.

[Van Der Maaten et al., 2009] Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). Dimensionality reduction : a comparative. *J Mach Learn Res*, 10 :66–71.

[Velazquez and Hamelberg, 2013] Velazquez, H. A. and Hamelberg, D. (2013). Conformation-Directed Catalysis and Coupled Enzyme–Substrate Dynamics in Pin1 Phosphorylation-Dependent Cis–Trans Isomerase. *J Phys Chem B*, 117(39) :11509–11517.

[Vendruscolo et al., 2003] Vendruscolo, M., Paci, E., Dobson, C. M., and Karplus, M. (2003). Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *Journal of the American Chemical Society*, 125(51) :15686–15687.

[Voter, 1997a] Voter, A. F. (1997a). Hyperdynamics : Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78(20) :3908.

[Voter, 1997b] Voter, A. F. (1997b). A method for accelerating the molecular dynamics simulation of infrequent events. *The Journal of chemical physics*, 106(11) :4665–4677.

[Vu-Dac et al., 1998] Vu-Dac, N., Gervois, P., Torra, I. P., Fruchart, J.-C., Kosykh, V., Kooistra, T., Princen, H., Dallongeville, J., and Staels, B. (1998). Retinoids increase human apo c-iii expression at the transcriptional level via the retinoid x receptor. contribution to the hypertriglyceridemic action of retinoids. *Journal of Clinical Investigation*, 102(3) :625.

[Wagner et al., 1995] Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., et al. (1995). A structural role for hormone in the thyroid hormone receptor. *Nature*, 378(6558) :690.

[Wang et al., 2011] Wang, Y., Markwick, P. R. L., de Oliveira, C. A. F., and McCammon, J. A. (2011). Enhanced Lipid Diffusion and Mixing in Accelerated Molecular Dynamics. *J Chem Theory Comput*, 7(10) :3199–3207.

[Wang et al., 2003] Wang, Z., Benoit, G., Liu, J., Prasad, S., Aarnisalo, P., Liu, X., Xu, H., Walker, N. P., and Perlmann, T. (2003). Structure and function of nurr1 identifies a class of ligand-independent nuclear receptors. *Nature*, 423(6939) :555–560.

[Washburn et al., 2009] Washburn, D. G., Hoang, T. H., Campobasso, N., Smallwood, A., Parks, D. J., Webb, C. L., Frank, K. A., Nord, M., Duraiswami, C., Evans, C., et al. (2009). Synthesis and sar of potent lxr agonists containing an indole pharmacophore. *Bioorganic & medicinal chemistry letters*, 19(4) :1097–1100.

[Watkins et al., 2003] Watkins, R. E., Maglich, J. M., Moore, L. B., Wisely, G. B., Noble, S. M., Davis-Searles, P. R., Lambert, M. H., Kliewer, S. A., and Redinbo, M. R. (2003). 2.1 å crystal structure of human pxr in complex with the st. john's wort compound hyperforin. *Biochemistry*, 42(6) :1430–1438.

[Weinbach and Elber, 2005] Weinbach, Y. and Elber, R. (2005). Revisiting and parallelizing shake. *Journal of Computational Physics*, 209(1) :193–206.

[Weis, 2016] Weis, D. D. (2016). *Hydrogen Exchange Mass Spectrometry of Proteins : Fundamentals, Methods, and Applications*. John Wiley & Sons.

[Wereszczynski and McCammon, 2010] Wereszczynski, J. and McCammon, J. A. (2010). Using selectively applied accelerated molecular dynamics to enhance free energy calculations. *Journal of chemical theory and computation*, 6(11) :3285–3292.

[Wereszczynski and McCammon, 2012] Wereszczynski, J. and McCammon, J. A. (2012). Accelerated molecular dynamics in computational drug design. *Computational Drug Discovery and Design*, pages 515–524.

[Wickelmaier, 2003] Wickelmaier, F. (2003). An introduction to mds. *Sound Quality Research Unit, Aalborg University, Denmark*, 46.

[Williams et al., 2010] Williams, S. L., De Oliveira, C. A. F., and McCammon, J. A. (2010). Coupling constant ph molecular dynamics with accelerated molecular dynamics. *Journal of chemical theory and computation*, 6(2) :560–568.

[Wilson et al., 1992] Wilson, T. E., Paulsen, R. E., et al. (1992). Participation of non-zinc finger residues in dna binding by two nuclear orphan receptors. *Science*, 256(5053) :107.

[Wishart et al., 1992] Wishart, D., Sykes, B., and Richards, F. (1992). The chemical shift index : a fast and simple method for the assignment of protein secondary structure through nmr spectroscopy. *Biochemistry*, 31(6) :1647–1651.

[Wishart and Nip, 1998] Wishart, D. S. and Nip, A. M. (1998). Protein chemical shift analysis : a practical guide. *Biochemistry and cell biology*, 76(2-3) :153–163.

[Wishart and Sykes, 1994] Wishart, D. S. and Sykes, B. D. (1994). The 13c chemical-shift index : a simple method for the identification of protein secondary structure using 13c chemical-shift data. *Journal of biomolecular NMR*, 4(2) :171–180.

[Wishart et al., 1997] Wishart, D. S., Watson, M. S., Boyko, R. F., and Sykes, B. D. (1997). Automated 1h and 13c chemical shift prediction using the biomagresbank. *Journal of biomolecular NMR*, 10(4) :329–336.

[Wolf, 2006] Wolf, G. (2006). Is 9-cis-retinoic acid the endogenous ligand for the retinoic acid-x receptor ? *Nutrition reviews*, 64(12) :532–538.

[Woodward et al., 1982] Woodward, C., Simon, I., and Tüchsen, E. (1982). Hydrogen exchange and the dynamic structure of proteins. *Molecular and cellular biochemistry*, 48(3) :135–160.

[Woodward and Hilton, 1979] Woodward, C. K. and Hilton, B. D. (1979). Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu. Rev. Biophys. Bioeng.*, 8(1) :99–127.

[Word et al., 1999] Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999). Asparagine and glutamine : using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of molecular biology*, 285(4) :1735–1747.

[Wurtz et al., 1996] Wurtz, J.-M., Bourguet, W., Renaud, J.-P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nature Structural & Molecular Biology*, 3(1) :87–94.

[Xia et al., 2010] Xia, G., Boerma, L. J., Cox, B. D., Qiu, C., Kang, S., Smith, C. D., Renfrow, M. B., and Muccio, D. D. (2010). Structure, energetics, and dynamics of binding coactivator peptide to the human retinoid x receptor α ligand binding domain complex with 9-cis-retinoic acid. *Biochemistry*, 50(1) :93–105.

[Xin et al., 2010] Xin, Y., Doshi, U., and Hamelberg, D. (2010). Examining the limits of time reweighting and Kramers' rate theory to obtain correct kinetics from accelerated molecular dynamics. *The Journal of Chemical Physics*, 132(22) :224101–8.

[Xu et al., 2001] Xu, H. E., Lambert, M. H., Montana, V. G., Plunket, K. D., Moore, L. B., Collins, J. L., Oplinger, J. A., Kliewer, S. A., Gampe, R. T., McKee, D. D., et al. (2001). Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. *Proceedings of the National Academy of Sciences*, 98(24) :13919–13924.

[Yamada and Kakuta, 2014] Yamada, S. and Kakuta, H. (2014). Retinoid x receptor ligands : a patent review (2007–2013). *Expert opinion on therapeutic patents*, 24(4) :443–452.

[Yan et al., 2004] Yan, X., Broderick, D., Leid, M. E., Schimerlik, M. I., and Deinzer, M. L. (2004). Dynamics and ligand-induced solvent accessibility changes in human retinoid x receptor homodimer determined by hydrogen deuterium exchange and mass spectrometry. *Biochemistry*, 43(4) :909–917.

[Yan et al., 2006] Yan, X., Deinzer, M. L., Schimerlik, M. I., Broderick, D., Leid, M. E., and Dawson, M. I. (2006). Investigation of ligand interactions with human rxrα by hydrogen/deuterium exchange and mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 17(11) :1510–1517.

[Yan et al., 2007] Yan, X., Pérez, E., Leid, M., Schimerlik, M. I., De Lera, A. R., and Deinzer, M. L. (2007). Deuterium exchange and mass spectrometry reveal the interaction differences of two synthetic modulators of rxrα lbd. *Protein Science*, 16(11) :2491–2501.

[Yang et al., 2014] Yang, L., Broderick, D., Jiang, Y., Hsu, V., and Maier, C. S. (2014). Conformational dynamics of human fxr-lbd ligand interactions studied by hydrogen/deuterium exchange mass spectrometry : Insights into the antagonism of the hypolipidemic agent z-guggulsterone. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844(9) :1684–1693.

[Yoshimura et al., 2007] Yoshimura, K., Muto, Y., Shimizu, M., Matsushima-Nishiwaki, R., Okuno, M., Takano, Y., Tsurumi, H., Kojima, S., Okano, Y., and Moriwaki, H. (2007). Phosphorylated retinoid x receptor α loses its heterodimeric activity with retinoic acid receptor β. *Cancer science*, 98(12) :1868–1874.

[Yotsumoto et al., 2005] Yotsumoto, T., Naitoh, T., Kanaki, T., and Tsuruzoe, N. (2005). A retinoid x receptor antagonist, hx531, improves leptin resistance without increasing plasma leptin level in kk-a y mice under normal dietary conditions. *Metabolism*, 54(5) :573–578.

[Závodszky et al., 1998] Závodszky, P., Kardos, J., Svingor, and Petsko, G. A. (1998). Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *PNAS*, 95(13) :7406–7411.

[Zechel et al., 1994] Zechel, C., Shen, X.-Q., Chambon, P., and Gronemeyer, H. (1994). Dimerization interfaces formed between the dna binding domains determine the cooperative binding of rxr/rar and rxr/tr heterodimers to dr5 and dr4 elements. *The EMBO Journal*, 13(6) :1414.

[Zhang et al., 2011a] Zhang, H., Chen, L., Chen, J., Jiang, H., and Shen, X. (2011a). Structural basis for retinoic x receptor repression on the tetramer. *Journal of Biological Chemistry*, 286(28) :24593–24598.

[Zhang et al., 2003] Zhang, H., Neal, S., and Wishart, D. S. (2003). Refdb : a database of uniformly referenced protein chemical shifts. *Journal of biomolecular NMR*, 25(3) :173–195.

[Zhang et al., 2011b] Zhang, H., Zhou, R., Li, L., Chen, J., Chen, L., Li, C., Ding, H., Yu, L., Hu, L., Jiang, H., et al. (2011b). Danthron functions as a retinoic x receptor antagonist by stabilizing tetramers of the receptor. *Journal of Biological Chemistry*, 286(3) :1868–1875.

[Zhang et al., 2010] Zhang, J., Chalmers, M. J., Stayrook, K. R., Burris, L. L., Garcia-Ordonez, R. D., Pascal, B. D., Burris, T. P., Dodge, J. A., and Griffin, P. R. (2010). Hydrogen/deuterium exchange

reveals distinct agonist/partial agonist receptor dynamics within vitamin d receptor/retinoid x receptor heterodimer. *Structure*, 18(10) :1332–1341.

[Zhang et al., 2011c] Zhang, J., Chalmers, M. J., Stayrook, K. R., Burris, L. L., Wang, Y., Busby, S. A., Pascal, B. D., Garcia-Ordonez, R. D., Bruning, J. B., Istrate, M. A., et al. (2011c). Dna binding alters coactivator interaction surfaces of the intact vdr–rxr complex. *Nature structural & molecular biology*, 18(5) :556–563.

[Zhang, 1995] Zhang, Y.-Z. (1995). Protein and peptide structure and interactions studied by hydrogen exchange and nmr. *PA, USA : University of Pennsylvania*.

[Zhao et al., 2000] Zhao, Q., Chasse, S. A., Devarakonda, S., Sierk, M. L., Ahvazi, B., and Rastinejad, F. (2000). Structural basis of rxr-dna interactions. *Journal of molecular biology*, 296(2) :509–520.

[Zhao et al., 2007] Zhao, W.-x., Tian, M., Zhao, B.-x., Li, G.-d., Liu, B., Zhan, Y.-y., Chen, H.-z., and Wu, Q. (2007). Orphan receptor tr3 attenuates the p300-induced acetylation of retinoid x receptor-α. *Molecular Endocrinology*, 21(12) :2877–2889.

[Zugmaier et al., 1996] Zugmaier, G., Jäger, R., Grage, B., Gottardis, M., Havemann, K., and Knabbe, C. (1996). Growth-inhibitory effects of vitamin d analogues and retinoids on human pancreatic cancer cells. *British journal of cancer*, 73(11) :1341.

Annexes

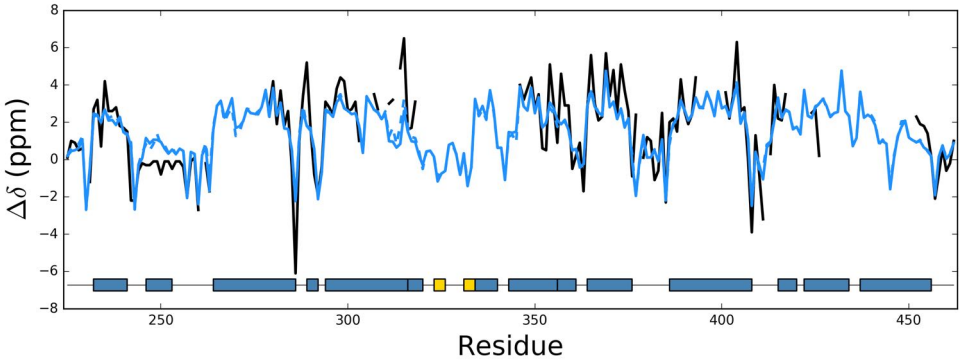


FIGURE 7.1 – Comparaison entre les déplacements chimiques secondaires $\Delta\delta_{\text{exp}}$ ^{13}Cα obtenus expérimentalement (en noir) et ceux prédits $\Delta\delta_{\text{aMD}}$ ^{13}Cα à partir des trajectoires aMD entières (en bleu par un trait plein) et à partir des trajectoires aMD où les 50 premières nanosecondes de chaque trajectoire ont été supprimé (en bleu par un trait en pointillé) de la forme libre du LBD de RXRα.

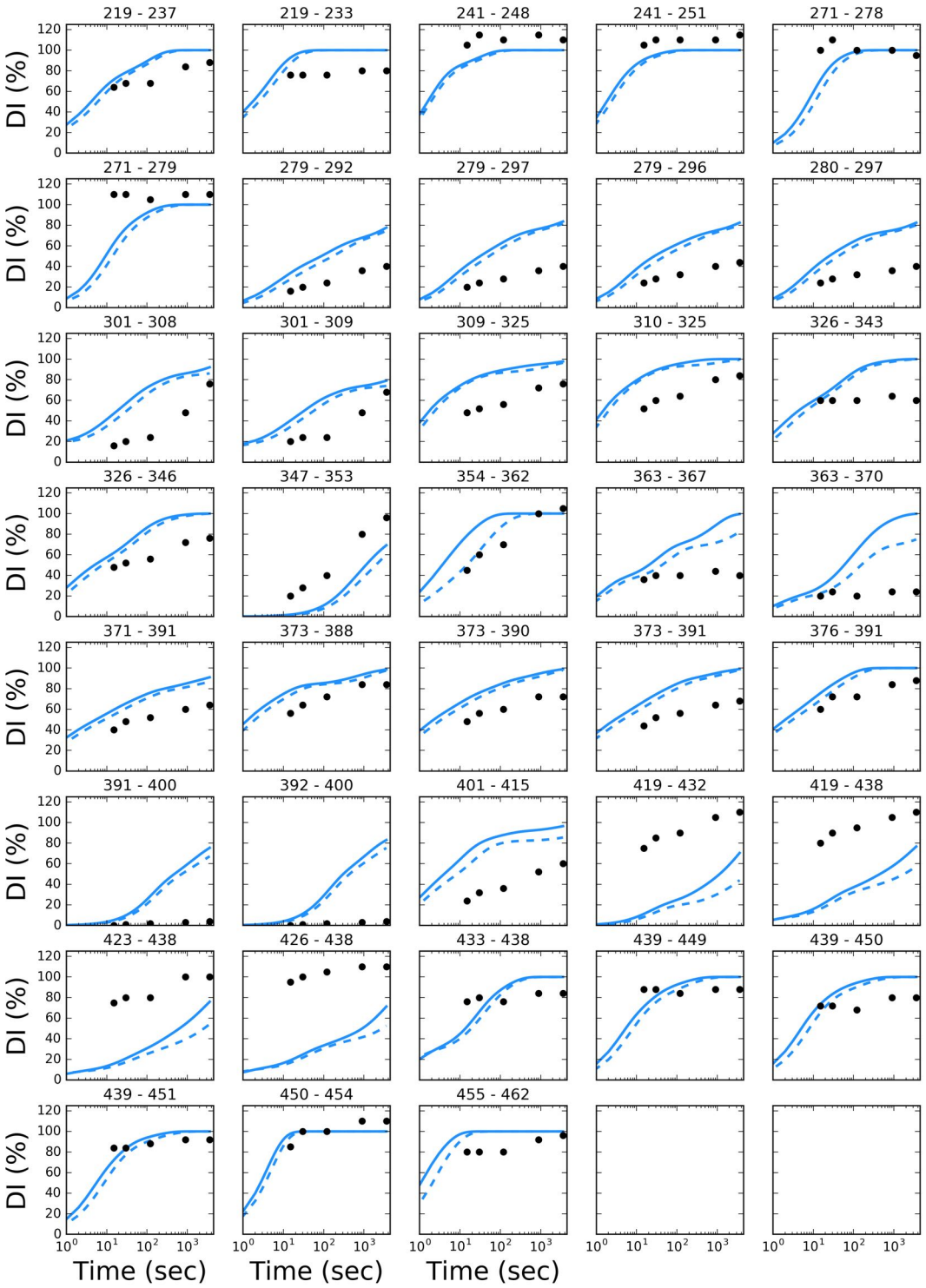


FIGURE 7.2 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme libre du LBD de RXRα (modèle simple). En bleu par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires entières aMD, en bleu par des traits pointillés ceux prédits à partir des trajectoires aMD où les 50 premières nanosecondes de chaque trajectoire aMD ont été supprimées et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

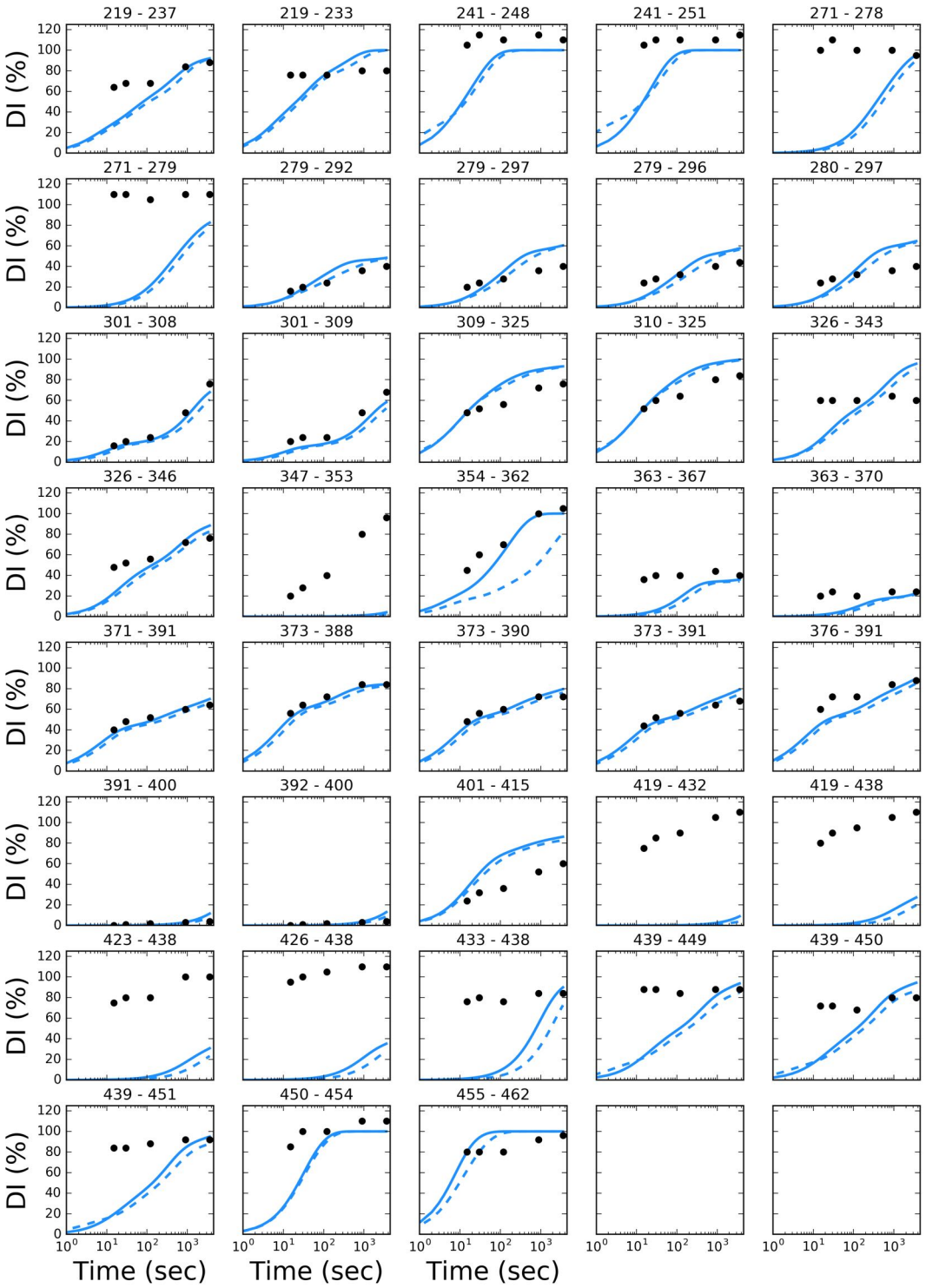


FIGURE 7.3 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme libre du LBD de RXRα (modèle concerté). En bleu par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires entières aMD, en bleu par des traits pointillés ceux prédits à partir des trajectoires aMD où les 50 premières nanosecondes de chaque trajectoire aMD ont été supprimées et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

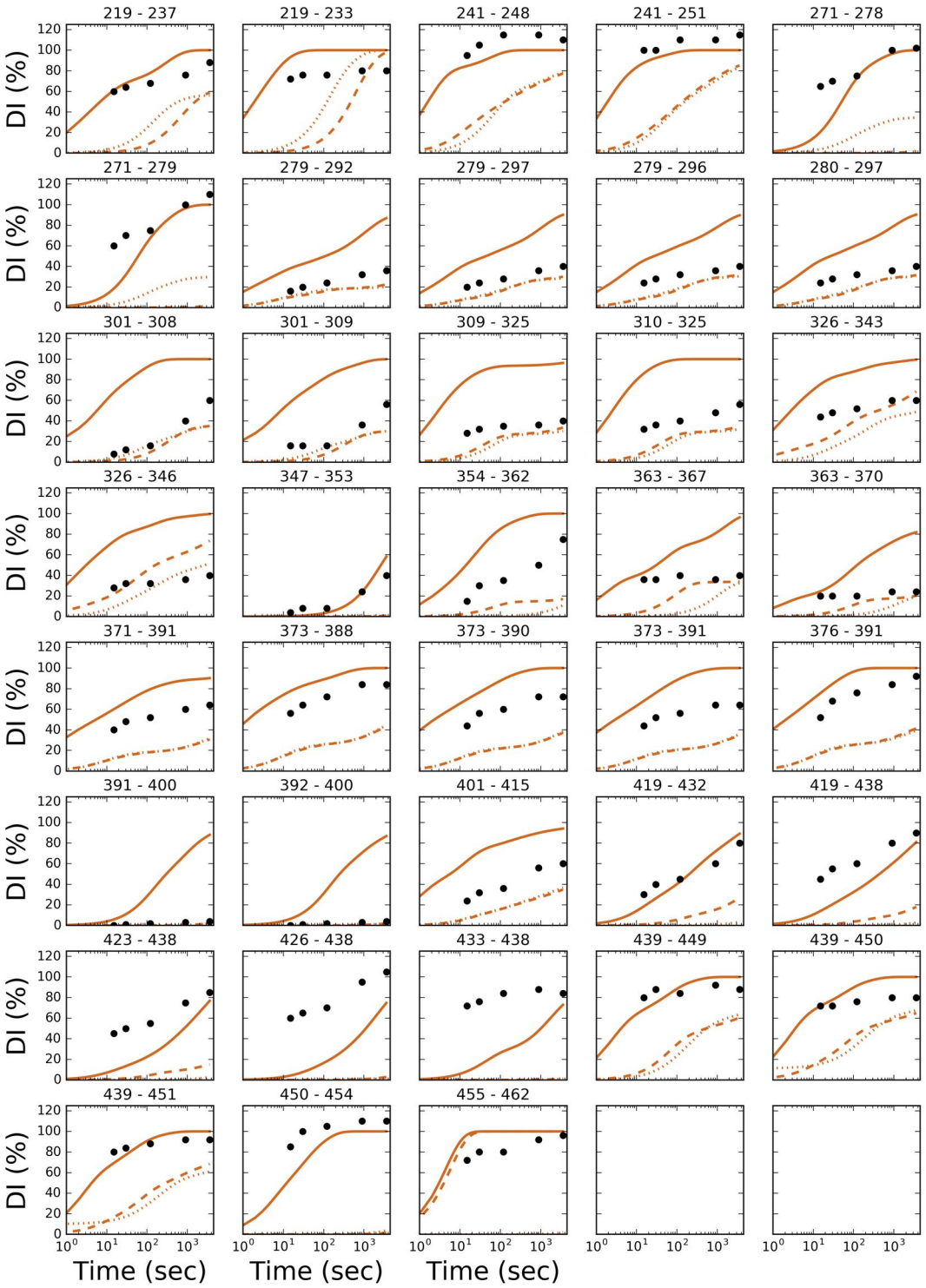


FIGURE 7.4 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme liée à 9cRA du LBD de RXRα (modèle simple). En orange par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires aMD, en orange par de larges pointillés ceux prédits à partir d'une trajectoire cMD de 10 ns de la forme liée à 9cRA et en orange par de petits pointillés ceux prédits à partir d'une trajectoire cMD de 10 ns de la forme liée dont 9cRA a été retiré et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

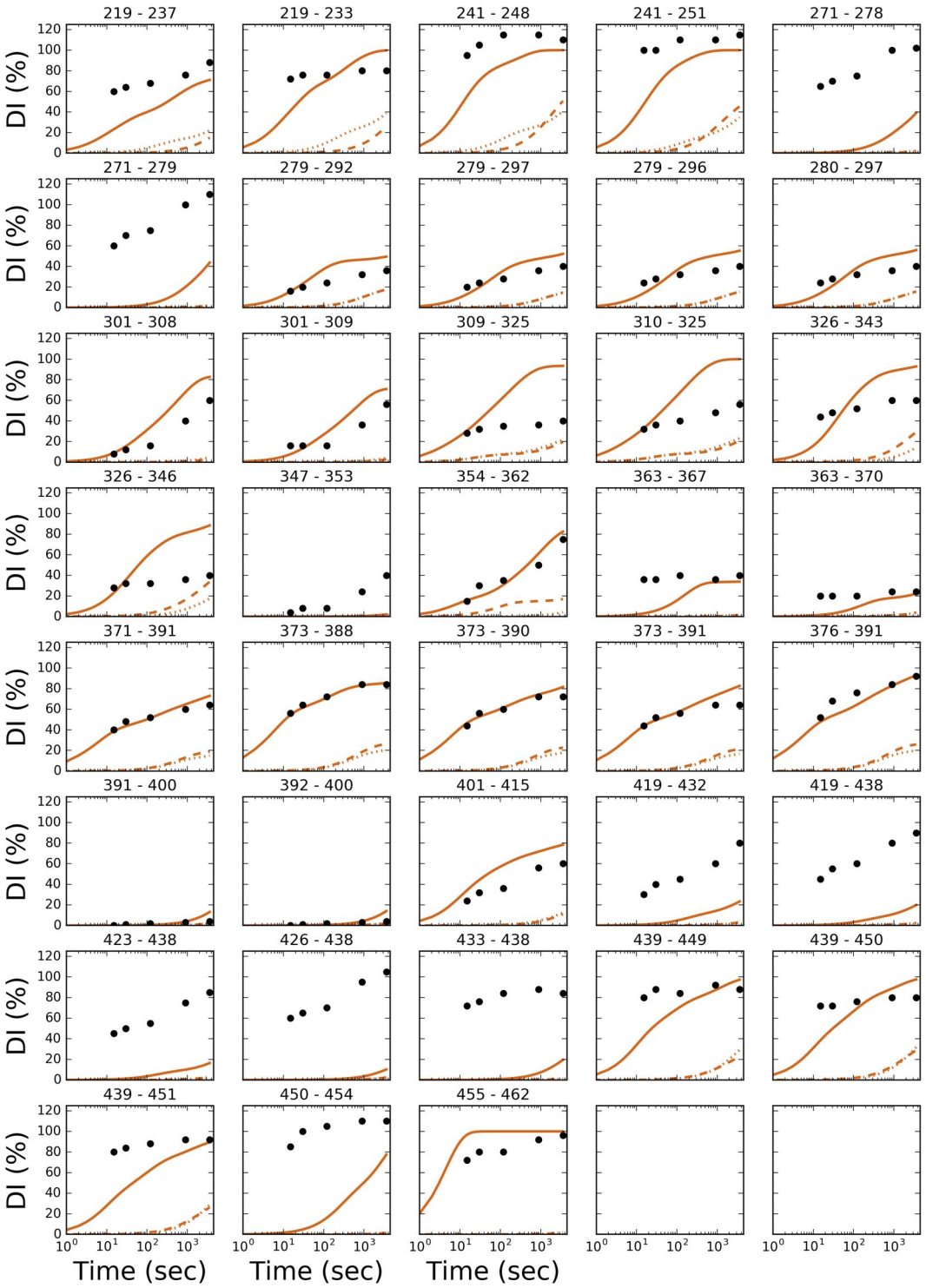


FIGURE 7.5 – Détails par peptide du taux d'incorporation en deutérium (exprimé en %) au cours du temps (exprimé en seconde sur une échelle logarithmique) pour la forme liée à 9cRA du LBD de RXRα (modèle concerté). En orange par un trait plein est représenté le taux d'incorporation en deutérium prédit à partir des trajectoires aMD, en orange par de larges pointillés ceux prédits à partir d'une trajectoire cMD de 10 ns de la forme liée à 9cRA et en orange par de petits pointillés ceux prédits à partir d'une trajectoire cMD de 10 ns de la forme liée dont 9cRA a été retiré et en noir le taux d'incorporation en deutérium obtenu expérimentalement à 15, 30, 90, 120 et 3600 secondes.

Publication

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Structural Insights into the Molecular Mechanism of Vitamin D Receptor Activation by Lithocholic Acid Involving a New Mode of Ligand Recognition

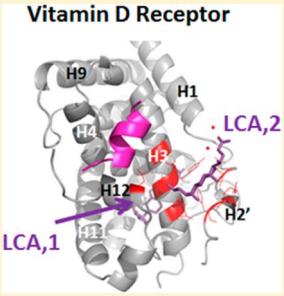
Anna Y. Belorusova,[†] Jérôme Eberhardt,[†] Noëlle Potier,[‡] Roland H. Stote,[†] Annick Dejaegere,[†] and Natacha Rochel*[,†]

[†]Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964, Centre National de Recherche Scientifique (CNRS) UMR 7104, Université de Strasbourg, 67404 Illkirch, France

[‡]Institut de Chimie LC3, CNRS, UMR 7177, 67008 Strasbourg, France

Supporting Information

ABSTRACT: The vitamin D receptor (VDR), an endocrine nuclear receptor for 1α,25-dihydroxyvitamin D3, acts also as a bile acid sensor by binding lithocholic acid (LCA). The crystal structure of the zebrafish VDR ligand binding domain in complex with LCA and the SRC-2 coactivator peptide reveals the binding of two LCA molecules by VDR. One LCA binds to the canonical ligand-binding pocket, and the second one, which is not fully buried, is anchored to a site located on the VDR surface. Despite the low affinity of the alternative site, the binding of the second molecule promotes stabilization of the active receptor conformation. Biological activity assays, structural analysis, and molecular dynamics simulations indicate that the recognition of two ligand molecules is crucial for VDR agonism by LCA. The unique binding mode of LCA provides clues for the development of new chemical compounds that target alternative binding sites for therapeutic applications.



INTRODUCTION

The vitamin D nuclear receptor (VDR) binds the active form of vitamin D, calcitriol (1α,25-dihydroxyvitamin D3, 1,25-$(OH)_2D_3$), and functions as a ligand-regulated transcription factor.[1] It stimulates intestinal calcium absorption, bone calcium resorption, and renal calcium reabsorption. In addition, it regulates growth and differentiation of many cell types and displays immunoregulatory and anti-inflammatory activities.[2] Besides its role in vitamin D physiology, VDR is a sensor for secondary bile acids including lithocholic acid (LCA) and 3-keto-LCA and serves as a means of protection from their harmful effects in the intestine.[3] Indeed, LCA is a considered promoter of colon carcinogenesis in mammals.[4] VDR is activated by LCA, and even more efficiently by 3-keto-LCA, but still with less potency than calcitriol.[3c] Via VDR, LCA stimulates its own degradation in a classic negative feedback regulation by inducing the expression of *CYP3A*, the gene encoding a catabolic enzyme that detoxifies bile acids in human intestinal cells.[5] LCA has been shown to act as a vitamin D substitute in rats although its action is weaker.[6] Interestingly, LCA activates VDR in target intestine and kidneys organs without causing hypercalcemia,[7] the major side effect of calcitriol.

Detailed information on the molecular mechanism of action of calcitriol and synthetic agonist ligands have been obtained by the elucidation of the crystal structure of its complexes with the VDR ligand binding domain (LBD).[8] Structural characterization of VDR−ligand interactions have been investigated using human, rat, or zebrafish VDR LBDs.[9] The crystal structure of the hVDR LBD complex[9a] has revealed a conformation similar to that of other nuclear receptors LBDs with a general fold that consists of a three-layered α-helical sandwich composed of 12 helices (H1−H12). Upon ligand binding, helix H11 is repositioned in the continuity of helix H10, and helix H12, also referred as the activating domain (AD) of the AF-2 function, swings to seal the binding cavity. This stabilizes ligand binding by contributing to the hydrophobic environment, making in some cases additional contacts with the ligand. This agonist conformation of H12 allows interactions with coactivator proteins through surface-exposed residues. Structure−function study analysis has predicted that LCA binds to VDR in a mode distinct from that of calcitriol.[7,10] Although the crystal structures of the rVDR in complex with LCA and its derivatives together with the Med1 coactivator peptide have been recently described,[11] these structures reveal a similar mode of binding through direct and water mediated hydrogen bonds with the residues forming the anchors of calcitriol in its ligand binding pocket (LBP) with no significant differences in interactions with coactivators. To get further insight into the mode of recognition of LCA by VDR, we decided to study the structure and function of its complex using the wild-type zVDR and hVDR. In the past, the zebrafish construct was successfully used in order to circumvent the

Received: February 16, 2014
Published: May 12, 2014

 © 2014 American Chemical Society

packing constraints of the crystal form of the human hVDR deletion mutant lacking the insertion region of 50 residues between helices H2 and H3.[12] We have previously shown that the binding of calcitriol is identical in both zVDR LBD and hVDR LBD.[9c] Furthermore, LCA and derivatives are known to activate zebrafish VDR and other nonmammalian VDRs.[13]

We now report the crystallographic structure of the zVDR LBD in complex with the secondary bile acid LCA at a resolution of 1.9 Å. Unexpectedly, this crystal structure shows a second LCA molecule bound to VDR in a novel binding surface together with the binding of an LCA molecule in the calcitriol LBP. This is the first structure of VDR showing two bound ligand molecules. Examination of the protein−ligand interactions reveals its mode of binding to VDR and highlights both similarities and differences with respect to the calcitriol-bound VDR structure. Biophysical characterization in solution using isothermal titration calorimetry (ITC) and electrospray mass spectrometry (ESI-MS) shows that VDR binds the two LCA molecules with different affinities in a two-step mechanism. Structural observations together with molecular dynamics (MD) simulations show that the second site promotes stabilization of the VDR agonist conformation. This multidisciplinary functional and structural analysis demonstrates the functional importance of this second LCA molecule bound to VDR.

■ RESULTS

Crystal Structure Reveals Two Molecules of LCA Interacting with VDR LBD. To unravel the molecular mechanism of action of LCA, we cocrystallized LCA with the zebrafish wild-type VDR LBD in the presence of coactivator peptide corresponding to the second LXXLL motif of SRC-2 and solved the structure to a resolution of 1.9 Å. The data collection and refinement statistics of the structure are summarized in Table 1. All protein residues were found in allowed regions of the Ramachandran plot, and the validation parameters exhibit the quality of final model. The overall structure is highly homologous to the VDR-calcitriol structures with the insertion region between helices H2 and H3 not visible in the electron density map indicating local disorder of this region (Figure 1A). Unexpectedly, two molecules of LCA were observed in the electron density maps (Figure 1B). One is located in the cognate LBP site recognized by calcitriol, and a second site is located on the LBD surface between loops 1−3, H2′, and H3. The position and conformation of the activation helix H12 in the zVDR-LCA complex is strictly maintained in its agonistic position.[9c,14] When compared to the zVDR LBD-calcitriol structure, the Cα atoms of the zVDR LBD-LCA complex have a root-mean-square deviation (rmsd) of 0.5 Å over 248 residues. However, significant shifts in secondary structural elements are observed with a rearrangement of the protein going from helix H6 through the β-sheet region, the last turn of helix H11, and to a lesser extent, the loop between helices H1 and H3. The rmsd values for the Cα atoms of the region of H6 (residues 316−330 of zVDR) is 1.3 Å (Figure 2A).

The h-, r-, and zVDR LBPs are lined with identical residues and have homologous structures.[9c] Note that the residue numbers used in the text correspond to the zVDR sequence. When the human VDR is referred to, the corresponding human sequence numbers are indicated as hAAxxx.

Consistent with the previous crystal structure of the rVDR-LCA complex,[11] LCA in the primary binding site adopts an orientation opposite that of calcitriol with the A-ring facing H12 and the side chain pointing toward the β-sheet (Figure 2A). LCA contains polar groups at each end that allow it to maintain most of the hydrogen bonds observed in the VDR-calcitriol structure. Only two direct H-bonds are formed between zTyr175 (H1) and zSer306 (H5) and the 24-carboxyl group. The other anchoring residues of calcitriol, zSer265 (H3) and zArg302 (H5), interact with the 24-carboxyl group of LCA through water molecules. Residues zHis333 (loop 6−7) and zHis423 (H11) interact with the hydroxyl group of the A-ring of LCA (Figure 2B). The LCA-receptor hydrophobic interactions involve the same residues as those for calcitriol, although for the former, the number of interactions is less (41

Table 1. Data Collection and Refinement Statistics

	zVDR LBD-LCA
Data Processing	
X-ray source	ID29 ESRF
detector	Pilatus 6M
λ	0.976 Å
temperatue	100 K
resolution (Å)	20.0−1.9 (1.97−1.9)
crystal space group	$P6_522$
cell parameters (Å)	$a = b = 66.213$; $c = 262.571$
unique reflections	28123
mean redundancy	10.8 (7.7)
R_{sym} (%)	5.5 (21.1)
completeness (%)	97.1 (76.1)
mean I/σ	37 (3.8)
Refinement	
rmsd bond length (Å)	0.008
rmsd bond angles (deg)	1.25
R_{cryst} (%)	21.3
R_{free} (%)	24.4
no. of non-H atoms	
protein	2033
ligands	54
water	87
average B factor	
protein	46.16
ligands	36.23
water	49.16
Ramachandran plots	
favored (%)	97.53
allowed (%)	2.47

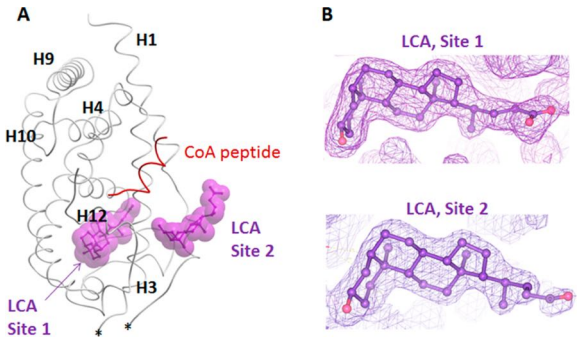


Figure 1. zVDR LBD crystal structure bound to two LCA molecules: (A) overall structure of the VDR LBD receptor indicating the region of the two ligand binding sites; (B) $2F_o - F_c$ electron density (violet) contoured to 1σ showing the two bound LCA molecules.

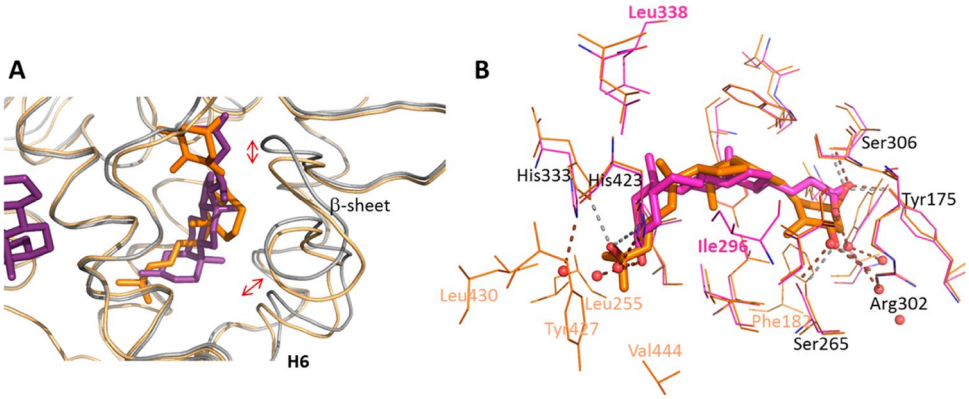


Figure 2. LCA binding to the canonical binding pocket. (A) Superposition of LCA (pink) and calcitriol (orange) molecules in the canonical VDR binding pocket showing the rearrangement of the secondary structure elements H6 and the β-sheet upon LCA binding. (B) Details of interactions in the first site. Residues that interact at a 4.0 Å cutoff are shown. The hydroxyl groups of LCA making H-bonds interactions are labeled in black. Specific interactions with calcitriol but absent in the LCA complex are labeled in orange, and specific interaction with LCA absent in the calcitriol complex are labeled in pink. Water molecules are indicated by red spheres.

interactions with LCA instead of 56 for calcitriol at a 4.0 Å distance cutoff). Even though the interactions with the receptor are globally very similar between LCA and calcitriol, there are some significant differences between the two ligands. Because of its shorter length, 14 Å for LCA versus 15.1 Å for calcitriol, and its position closer to the β-sheet, some interactions between the side chain of calcitriol and H11 and H12 C-terminal region of VDR are absent in the VDR-LCA complex, notably with zTyr427, zLeu430, or zVal444. On the A-ring site of calcitriol, the interaction with zPhe182 is also not formed by LCA. On the other hand, hydrophobic interactions mediated by LCA, and absent in the calcitriol complex, are observed with zIle296 (H5, distance 3.8 Å for LCA instead of 4.5 Å for calcitriol) and zLeu338 (H8, distance of 3.7 Å instead of 4.7 Å for calcitriol).

Analysis of the Second LCA Binding Site. In addition to the first cognate binding site, LCA associates with a second site localized on the surface of VDR close to the loop 1−3 and H3 (Figure 3A). Unlike the first, the second molecule is not fully buried. The conformation of the second LCA is, however, similar to that of the first with some difference in the positioning of the side chain. This second LCA ligand is primarily stabilized by nonpolar interactions with zAsp181, zArg184, zPhe185 (H2), zAsp260, zSer263, zTyr264, zGln267 (H3), and zLeu443 (H12) (Figure 3B). The hydroxyl group of the A-ring forms two direct H-bonds with zSer263 (H3) and zGln267 (H3). The side chain is more loosely positioned, and its 24-carboxyl group interacts through a water molecule with zAsp181 (H2′) and zLys268 (H3). Analysis of the crystal packing around the second LCA ligand indicates only one hydrophobic interaction of the A-ring of LCA with a symmetrical molecule (Supporting Information Figure 1), suggesting that the two LCA ligands are specifically bound to VDR.

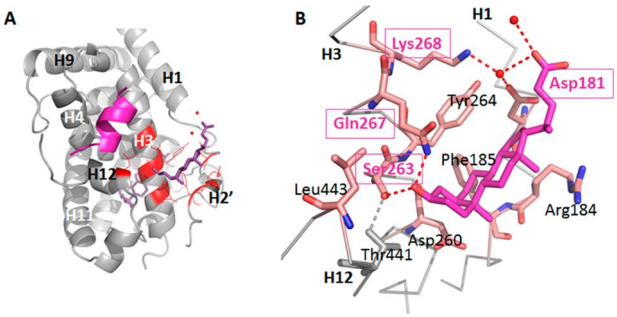


Figure 3. Second LCA binding site. (A) The secondary structural elements involved in the second LCA binding site are shown in red in the overall crystal structure. (B) Ligand and residue interactions occurring in the newly identified second site at a 4.0 Å cutoff. Residues involved in hydrogen bonds to the second LCA are labeled in pink. Red dashed lines represent hydrogen bonds, and red spheres indicate water molecules.

The observation of two LCA binding sites within zVDR LBD was rather unexpected because of the fact that only one molecule of ligand was seen in the previously reported crystal structure of the rVDR-LCA complex.[11] We cannot bypass the fact that VDR LBDs from different species have been used. However, the binding residues forming the second binding site are conserved between VDR family members, with the exception of zArg184 that is replaced in human and rat VDR LBD by amino acids hGln152 and rAsp152. Thus, the preferential LCA binding by zVDR is not likely. Furthermore, the two VDR-LCA complexes differ in both crystallization conditions and packing of formed crystals. Therefore, the absence of the second ligand bound to the rVDR LBD can be explained by lower ligand concentration used for crystallization and the limited access to the second site caused by the constraints of crystal packing.

A comparison of the existing crystal structures of VDR LBD complexes containing only one bound ligand shows that the surface corresponding to the second LCA binding site is always occupied by water molecules. Of particular interest is a highly conserved water that interacts with zSer263 and zGln267, indicating that the identified binding site is favored to be occupied by a polar ligand atom. Indeed, this conserved water molecule gets displaced by the hydroxyl group of the LCA A-ring upon ligand binding.

To confirm the hypothesis that VDR is able to bind two ligand molecules in solution, we performed a biophysical characterization of the binding of LCA to the VDR LBD by ITC. To determine binding parameters, we performed a direct titration experiment for hVDR LBD with LCA (Figure 4A).

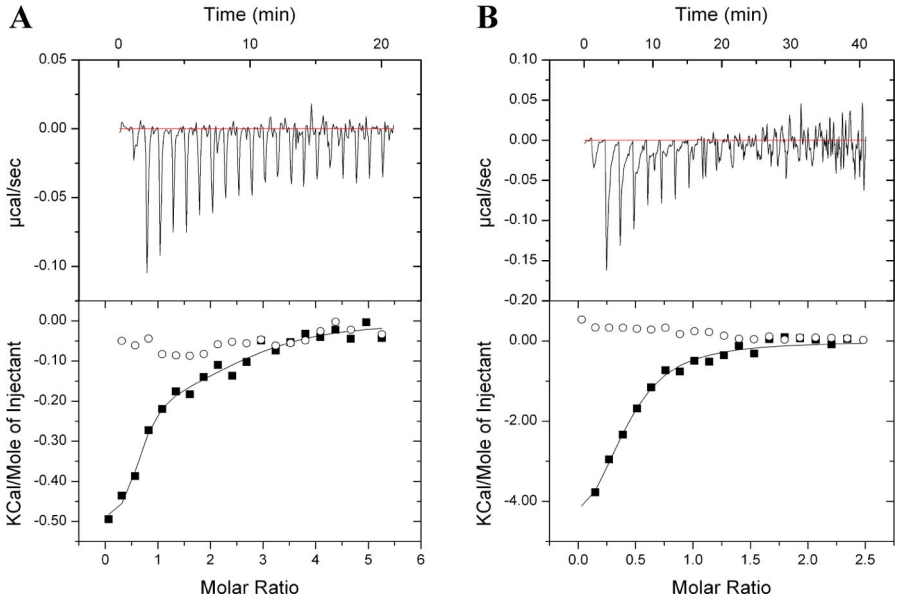


Figure 4. VDR binds two LCA molecules in solution. (A, B) Representative ITC isotherms for the binding of LCA to hVDR LBD. The top panels show the raw ITC data expressed as the change in thermal power with respect to time over the period of titration. For the lower panels, the change in molar heat is expressed as a function of molar ratio of corresponding LCA to VDR. Filled squares represent experimental data, and open circles correspond to control titration experiments. The solid lines in the lower panels represent the fit of data using the ORIGIN software. Shown are results from (A) direct titration and (B) reverse titration.

Table 2. Quantification of the Interaction between VDR and LCA by ITC

expt type	K_D (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	N_1	χ^2
direct titration	0.33(±0.26)	−8.84	−0.52(±0.05)	−8.31	0.57(±0.07)	325.1
	9.52(±6.82)	−6.86	−0.18(±0.06)	−6.68	2.11(±0.36)	
reverse titration	18.87(±5.52)	−6.45	−2.19(±0.09)	−4.26	2.50(±0.24)	1.67×10^4

The raw data are of rather low quality because of the poor LCA solubility and precipitation of protein during the experiment. However, binding isotherms were reproducible, and thermodynamic parameters were determined. The binding process appears to be complex and consists of several binding events with different affinities. A two-independent-site binding model best fits the data, giving affinities of 0.3 and 9.5 μM for the first and the second binding events, respectively. Both processes are mostly entropy-driven, with entropy contribution ($T\Delta S$) of 8.31 kcal mol^{-1} versus 6.68 kcal mol^{-1}. The enthalpy change is slightly favorable for both events (ΔH_1 = −0.52 kcal mol^{-1} and ΔH_2 = −0.18 kcal mol^{-1}). All thermodynamic parameters are summarized in Table 2. Unexpectedly, the stoichiometry of the second binding event was determined to be N = 2.11(±0.37), indicating that two LCA molecules bind to the low-affinity site. However, the inflection point of the binding isotherm can be observed near a 2:1 and not a 3:1 ratio. In this case, the chosen model with two ligand molecules that bind distinct sites simultaneously may not reflect the correct sequence of binding events.

To further verify the stoichiometry of the binding between hVDR LBD and LCA, we performed a reverse titration experiment (Figure 4B). The binding isotherm has an inflection point near molar ratio 0.4, giving N = 2.50(±0.24) when the fitting algorithm is adjusted to the reverse titration (ligand in the cell). Noteworthy in the case of reverse titration, heat peaks are wide with the time required to reach the baseline of ≥120 s. This indirectly confirms multiple binding events distinct in rate and time taking place during injection of protein into the ligand solution. Thus, our ITC experiments confirm the presence of two LCA binding sites within the hVDR LBD and it is likely that the first LCA molecule occupies the binding site within the canonical LBP and sequentially the second LCA binds to the low-affinity site. However, additional experiments and ITC data of higher quality are needed to determine and characterize the cooperativity between the two binding sites.

Binding of LCA to VDR was also monitored by ESI-MS under nondenaturating conditions using purified bacterially produced zVDR and hVDR LBD and calcitriol or LCA. In the presence of a large excess (10- to 20-fold molar excess of LCA), multiaddition of LCA to VDR could be observed (Supporting Information Figure 2).

For the biophysical assays, both zVDR and hVDR LBDs were used and similar results were obtained on LCA binding, confirming that the sequence difference zArg184/hGln152 in the second LCA binding site does not affect the binding of the second LCA.

Interactions at the Second Site Stabilize Active Conformation. Binding of LCA to the second binding site by direct interaction with H12 stabilizes the active agonist conformation of the complex and compensates for the weak stabilization of the C-terminal part of VDR by the binding of the first LCA in the canonical binding pocket. Analysis of the average temperature factors on the overall structure and its comparison with those observed for the crystal structures of the rVDR LBD in complex with calcitriol or LCA reveal additional stabilization of the loop 11−12, helix H12, and the coactivator peptide in the presence of the two LCA ligands (Figure 5).

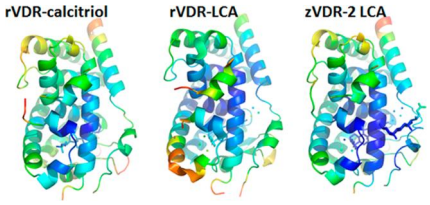


Figure 5. Influence of the binding of a second LCA molecule to VDR on the overall stabilization of the complex as shown by the *B*-factor plot on the overall structure of the complexes. The secondary structural elements are colored by *B*-factors on a relative scale: highest *B*-factors, red; lowest, blue.

To explore the function of the second binding site of VDR, we carried out site-directed mutagenesis. The roles of zSer263(hSer235), zGln267(hGln239), zAsp181(hAsp149), and zLys268(hLys240) on LCA binding and activation were investigated via site-directed mutagenesis of the corresponding hVDR residues to alanine. The mutations were expected to disrupt the binding of the second LCA by preventing the anchoring interactions without decreasing LCA binding to the first VDR binding site. To check that the mutations of hSer235Ala and hGln239Ala do not perturb the primary binding site, calcitriol binding to the VDR LBD double mutants was monitored by ESI-MS under nondenaturating conditions. In the presence of a 2-fold molar excess of calcitriol, the mutant VDR (hSer235Ala, hGln239Ala) is fully bound to the ligand with similar affinity as the wild-type VDR protein, thus indicating that these mutations do not affect the primary ligand binding site (Supporting Information Figure 3).

We next analyzed the effects of these mutations by transactivation functional assay in HEK293 cells using *CYP3A4* reporter gene (Figure 6). In agreement with published

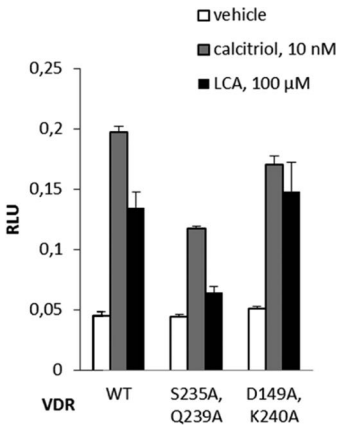


Figure 6. The second LCA is important for VDR function. Transient transfection assays with expression vectors encoding full-length VDR wild-type or mutant proteins were performed in HEK293 EBNA cells to evaluate their transcriptional activity on a *CYP3A1* promoter containing luciferase reporter gene in response to 10 nM calcitriol, 100 μM LCA, or vehicle. Bars represent the mean ± SD (n = 3).

data,[3a,c] LCA induces transcriptional response of hVDR wild-type at 100 μM ligand concentration. Mutations of the residues hAsp149 and hLys240 that interact with the second LCA through water molecules have little or no effect on the transcriptional response of VDR upon LCA or calcitriol stimulation. At the same time, the substitution of hSer235 and hGln239 by alanine drastically decreased the transcriptional activity of VDR to 48% of that of wild-type upon the LCA treatment. This dramatic effect correlates with the structural observations showing the importance of these residues for the binding at the second site, as they form direct H-bonds with the second LCA molecule. Interestingly, these mutations also affect the response of VDR upon calcitriol activation (60% of wild-type activity), suggesting that this novel VDR binding surface may also be important for full transcriptional response of other VDR ligands.

Structural Dynamics of zVDR-LCA Complex. To analyze the structural dynamics of the zVDR complex and the effect of the second LCA, MD simulations were run on zVDR LBD in complex with two LCA ligands and on zVDR LBD where the second LCA was removed. As a measure of the structural stability of zVDR, the overall rmsd of the backbone coordinates was calculated as a function of time for each simulation of the one and two LCA ligand/zVDR complexes. The time series (Supporting Information Figure 4) indicate that the structure is stable for both complexes given the relatively small rmsd values, with fluctuation of around 1.5 Å. The second binding site is not an enclosed pocket, as is the first binding site, but rather a cleft on the receptor surface that contains hydrophobic amino acids and is accessible to solvent in the absence of a second ligand. More precisely, in the simulations of zVDR with one bound LCA, the average solvent accessible surface of the amino acids that will find themselves within 5 Å of the second LCA ligand is 458 ± 52 Å^2. The magnitude of local structural fluctuations was analyzed from rms fluctuations around average structures (Figure 7 and Supporting Information Figure 5). The

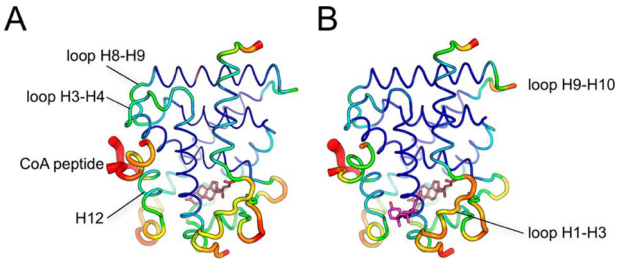


Figure 7. The second LCA stabilizes the zVDR agonist conformation. Shown are the rms fluctuations (rmsf) of the backbone atoms of zVDR LBDs in the presence of one (A) and two (B) LCA ligands mapped onto the crystal structure. Fluctuations are represented by a color gradient ranging from blue for rmsf less than 0.5 Å to red for rmsf greater than 1.5 Å. They were calculated from the final 30 ns of the molecular dynamics simulations (see text for details). The numerical data corresponding to the color code are presented in Supporting Information Figure 4.

fluctuations as a function of amino acid sequence were not significantly different for the two complexes (with one and two ligands) except in loop L9−10 and of the coactivator peptide, where indeed differences in the structural dynamics are observed. The side chains of the second binding pocket are generally mobile in the absence and in the presence of the second LCA molecule. The rmsf values for the coactivator peptide are smaller when an LCA is bound at the second site (Figure 7). This suggests that the recruitment of the second LCA provides additional stabilization of the coactivator peptide. Interestingly, the rmsf is higher for the loop L9−10 when the two ligands are bound (Figure 7); the biological consequence of this long-range effect remains to be assessed. Together, these

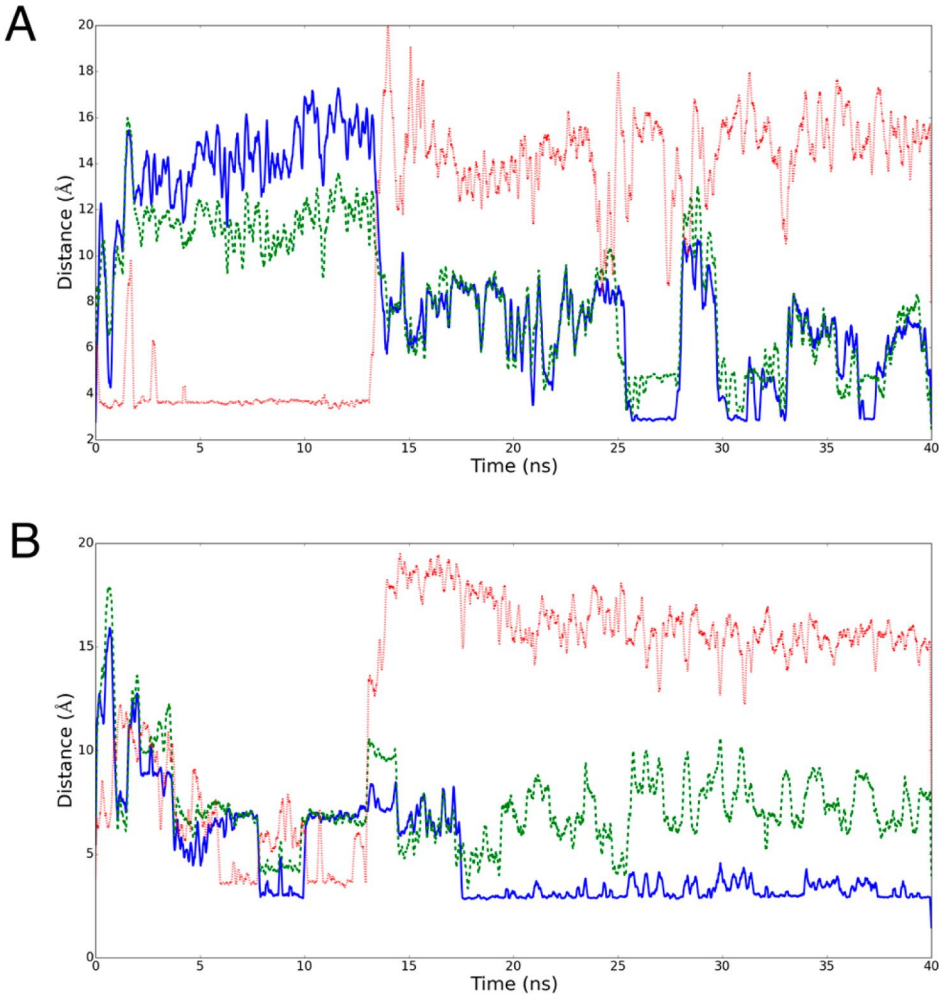


Figure 8. Analysis of the interactions between the second ligand and zVDR-SRC-2 coactivator peptide complex as a function of time. Shown are the times series for the distance between the second LCA and zVDR LBD. The distance (in Å) between LCA hydroxyl oxygen and zSer263 OG (blue full line), LCA hydroxyl oxygen and zGln267 NE2 (green dotted line), and LCA carboxyl carbon and C18−zLys268 NZ (red full line) is displayed as a function of time (ns) for the two molecular dynamics trajectories (A, B). The interactions monitored during the trajectories correspond to those presented in Figure 3b.

results indicate that the binding of LCA at the second site does not induce major changes in the global structural dynamics of zVDR, but the second ligand nevertheless affects the structural dynamics of the coactivator peptide, as well as that of the distant loop 9−10.

Next we characterized the interactions of the LCA ligand at the second site of zVDR by computing the van der Waals (VdW) and electrostatic interaction energies between the zVDR (with coactivator peptide) and the ligand (Supporting Information Figure 6). The results show that VdW interaction energy between the zVDR and the second site LCA is favorable (<0 kcal/mol) and stable during both simulations. The electrostatic interaction energies do show significant fluctuations during the simulations, but this is in line with the high mobility of the ligand carboxy terminal in the binding cleft observed during the simulations (Supporting Information Figure 6).

The relative distances between the LCA ligand at second site and zSer263, zGln267, and zLys268 were measured over each simulation. The minimum distance observed between the ligand hydroxyl group of the A ring and zSer263 is about 2.8 Å, i.e., the approximate distance of a hydrogen bond, in agreement with the crystallographic data (Figure 3B and Figure 8). Interestingly, in the simulations of the zVDR with one LCA ligand bound, a water molecule is found where the hydroxyl group of the A ring of the second ligand binds for at least 50% of the simulation time. Concerning zGln267, it also remains close to the ligand hydroxyl group of the A ring (Figure 8) while the distance between zLys268 and the ligand 24-carboxyl group is about 15 Å on average (Figure 8), which is too far for an interaction even through a water molecule, as was seen in the crystal structure of zVDR (Figure 3B). Thus, during the simulations, the carboxy-terminal part of the ligand is quite mobile while the hydroxyl-terminal part remains anchored close to zSer263. These results confirm the important role of the second LCA molecule in the agonist conformation of the complex and suggest that the residue zSer263 is more significant for the binding of the LCA to the second site of zVDR than the other residues identified in the crystal structure.

DISCUSSION AND CONCLUSION

VDR acts as an endocrine nuclear receptor for calcitriol as well as a bile acid sensor, binding the secondary bile acid LCA with low affinity and regulating its enzymatic degradation. LCA is highly toxic; it exerts carcinogenic effects in the colon[15] and causes hepatotoxicity in the liver when it overaccumulates.[16] Thus, VDR plays a critical role in the metabolism and detoxification of LCA. As with calcitriol, LCA induces nongenomic action that involves VDR in the plasma membrane of human hepatocyte cells.[17] LCA has been also shown to induce the recruitment of cofactors differently from calcitriol,[18] and interestingly, it does not induce the hypercalcemic side effects of calcitriol.[7] Structure−function analysis and mutational studies have previously suggested that LCA interacts with the VDR LBP differently from calcitriol, in particular with helices 3 and 4/5.[7,18a] Although the crystal structures of rVDR LBD bound to LCA and LCA derivatives have been reported,[11] structural observations demonstrated that most of the hydrophobic and hydrophilic interactions between the LCA and VDR LBP are similar to those of calcitriol; therefore, these structures do not provide an explanation for the selective VDR modulator action of LCA.

Using a combination of structural, biophysical, and in vitro transcriptional assay, we identified a novel unique binding mode of LCA to VDR. The crystal structure of zVDR LBD in complex with LCA reveals that LCA binds to two distinct sites on the VDR LBD. The first molecule binds to the canonical LBP with a reversed orientation compared to calcitriol, as previously described.[3c] This binding requires some rearrangement of the protein near helix 6. LCA interacts with the canonical VDR LBP mainly through hydrophobic interactions and weaker H-bonds with the anchoring residues through water molecules with zSer265, zArg302, zHis333, and zHis423 and direct contacts with zTyr175 and zSer306. No direct contacts are observed between this LCA molecule and the C-terminal part of VDR. Thus, weaker interactions explain the low affinity observed for the binding of LCA to VDR and its ability to activate VDR-mediated transcription at comparatively high micromolar concentration. The second LCA molecule is not fully buried, like the first, and is anchored to VDR through its hydroxyl A-ring moiety forming direct H-bonds with zSer263 (H3) and zGln267 (H3). The second binding site is of considerably lower affinity compared to the first site, but we show by structural analysis and MD simulations that the binding of a second molecule promotes stabilization of the active protein conformation. The simulations also identified zSer263 as the key residue for the anchoring of second LCA molecule, which is consistent with results of the mutational analysis carried out on the hVDR. Binding residues forming both the canonical LBP and the second binding site are conserved between VDR family members, suggesting a common mechanism of their activation by LCA.

LCA is also known as a weak agonist of other nuclear receptors such as the farnesoid X receptor (FXR) and the pregnane X receptor (PXR).[19] A sequence alignment reveals that the amino acids of the second binding site that are involved in H-bonds formation, indicatively hAsp232(zAsp181), hSer235(zSer263), hTyr236(zTyr264), and hGln239(zGln267) within H3 of VDR, are closely conserved between hVDR, hFXR, and hPXR (Supporting Information Figure 7), indicating that these nuclear receptors might share a similar mechanism of activation by LCA. The residues responsible for the binding of the second LCA molecule (zSer263 and zGln267, H3) are not identical in the hFXR (Thr306 and Gln310) and hPXR (Thr248 and Lys252); however, docking of LCA to the crystal structures of these receptors shows that corresponding amino acids can form direct H-bonds with the hydroxyl group of the A-ring of LCA (Supporting Information Figure 8). Interestingly, the crystal structure of rFXR LBD in complex with the bile acid 6α-ethylchenodeoycholic acid and coactivator peptide has revealed that two coactivator peptides can be simultaneously bound to the rFXR LBD surface.[20] Besides the canonical coactivator peptide binding cavity, the second peptide binding site is localized on the LBD surface and matches the novel LCA recognition site of zVDR. Binding of two coactivator peptides to the FXR was shown to provide additional receptor stabilization.[20]

It is possible that the mode of LCA binding and its role in stabilization of active conformation may be part of a larger mechanism of cofactor selectivity for this ligand. Being a weak agonist for VDR, LCA has been shown to affect the recruitment of classical co-regulators, such as SRC-1 and SMRT, or β-catenin in a manner distinct from that of calcitriol.[10,21] Furthermore, LCA stimulates interaction between VDR and another ligand-dependent co-repressor, Hairless (Hr), more efficiently than calcitriol.[18b,22] Interestingly, this co-regulator does not contact H12 of VDR but requires a part of H3 and H6 helices for interaction. The interaction site has been identified as being between residues 202 and 303 in the central portion of the LBD, indicating a possible participation of the second LCA-binding site in the co-regulator association.

Interestingly, the two-pocket model based on the computational and mutational analysis of VDR has been proposed earlier,[23] where the genomic and alternative binding pockets of VDR overlap. This model provided some insights into the specific actions of calcitriol and its analogues, including their genomic and nongenomic activities and allosteric communications between different conformers. However, the two-pocket model for VDR was not validated by structural data. The crystal structure presented here is the first one that reveals two ligands bound to VDR. While other nuclear receptors have been shown to be able to bind two ligands, in most of the reported structural data the second ligand binds on a nuclear receptor LBD surface near the coactivator binding site and prevents the binding of co-regulators. This is the case for 4-hydroxytamoxifen bound to ERβ estrogen receptor[24] or amino ketones that bind to the thyroid receptor TR.[25] New binding sites for compounds that allosterically inhibit interactions with coactivators were also described for the androgen receptor, AR.[26] Recent data report the crystal structure of a PPARα LBD complex bound to two molecules of the synthetic agonist ligand WY14643.[27] This structure revealed that this fibrate analogue induces activation through a two-step mechanism involving direct H12 stabilization and an alternative ligand binding mode that involves a second ligand in the pocket.[27] Although a similar mechanism requiring the binding of two molecules is observed for LCA, in VDR the second binding site is different from the one observed within the PPARα complex.

In conclusion, we have identified a novel molecular mechanism of VDR activation by LCA ligand that is different from that of calcitriol and requires two ligand molecules bound to distinct binding sites. The low affinity LCA binds to VDR in a two-step mechanism that leads to the stabilization of the receptor active conformation. This new mechanism may be important for other ligands that bind near the β-sheet and do

not interact directly with H12. This unique binding mode of LCA may also be implicated in cofactor selectivity for this ligand and explain its dissociate profile with no hypercalcemic effect. The present results open interesting new perspectives in the design of specifically targeted ligands acting as alternative site nuclear receptor modulators.

■ EXPERIMENTAL PROCEDURES

Ligands and Peptides. Crystalline calcitriol (Sigma) and lithocholic acid (Sigma) were dissolved in 100% ethanol and stored at −80 °C until use.

Vectors, Expression, and Purification. cDNAs encoding hVDR LBD (118−427 AA), zVDR LBD (154−456 AA), and hVDR LBD (118−427, hSer235Ala, hGln239Ala) were cloned into pET28b vector to generate N-terminal His-tag fusion proteins. Purification was carried out as previously described,[9c,28] including metal affinity chromatography and gel filtration. The reporter plasmid pLuc-MCS (Stratagene, La Jolla, CA) contained two copies of the distal VDRE from the human *CYP3A4* gene.

Transient Transfection and Luciferase Reporter Gene Assays. HEK293 EBNA cells were plated into 24-well plates at 10^5 cells per well and grown overnight in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% charcoal-treated fetal bovine serum (FCS) and 40 μg/mL gentamycin. Cells were transfected with 150 ng of the expression plasmid pSG5-hVDR wild-type or mutants (hSer235Ala, hGln239Ala, hAsp149Ala, hLys240Ala), 150 ng of the pSG5-hRXRα, 150 ng of the reporter plasmid pLuc-MCS (Stratagene, La Jolla, CA) containing two copies of the distal VDRE from the human *CYP3A4* gene, 3 ng of the pRLplasmid (Promega) containing the Renilla luciferase gene (transfection and cell viability control), and 497 ng of the carrier plasmid pBluescript (Stratagene). Transfection was performed with jetPEI (Polyplus transfection) according to the manufacturer's instructions. Six hours after transfection, tested compounds were added. Cells were harvested after 18 h of incubation with ligands. The amounts of reporter gene product (firefly luciferase) and constitutively expressed Renilla luciferase produced in the cells were measured using Dual-Luciferase reporter assay system (Promega) on a luminometer plate reader LB96P (Berthold Technologies). Luminescence of firefly luciferase values was normalized to the Renilla luciferase activity. Luciferase activities are expressed as relative units of light intensity. Data points represent the mean of assays performed in triplicate for at least three independent experiments. For every triplicate, the mean and the standard deviation of the mean were calculated.

Isothermal Titration Calorimetry Studies. ITC measurements were performed at 25 °C on a MicroCal ITC200 (MicroCal). Purified protein was dialyzed extensively against the buffer 25 mM HEPES−KOH, pH 7.0, 50 mM NaCl, 2% glycerol, and 1 mM TCEP. LCA was prepared as 20 mM stock solution in 100 mM NaOH (the compound is very water-insoluble) and diluted prior to titration with the dialysis buffer. In the direct titration experiment 2 μL aliquots of LCA at 1−1.5 mM were injected into a 40−75 μM hVDR LBD solution (200 μL sample cell). In the reverse titration the protein (500 μM) was injected into the cell containing the ligand (20 μM).The duration of each injection was 4 s with a delay between injections of 100−120 s. ITC titration curves were analyzed using the software Origin 7.0 (OriginLab). Standard free energies of binding and entropic contributions were obtained, respectively, as $\Delta G = -RT \ln(K_a)$ and $T\Delta S = \Delta H - \Delta G$, from the K_a and ΔH values derived from ITC curve fitting.

Electrospray Ionization Mass Spectrometry. Prior to ESI-MS analysis, samples were desalted on Zeba Spin desalting columns (Pierce) in 200 mM ammonium acetate (pH 8.0). ESI-MS measurements were performed on an electrospray time-of-flight mass spectrometer (MicrOTOF, BrukerDaltonic, Germany). Purity and homogeneity of the VDR LBD proteins were verified by mass spectrometry in denaturing conditions (samples were diluted at 2 pmol/μL in a 1:1 water−acetonitrile mixture (v/v) acidified with 1% formic acid). The mass measurements of the noncovalent complexes were performed in ammonium acetate (200 mM, pH 8.0). Samples were diluted to 8 pmol/μL in the previous buffer and continuously infused into the ESI ion source at a flow rate of 3 μL/min through a Harvard syringe pump (Harvard Apparatus model 11). A careful optimization of the interface parameters was performed to obtain the best sensitivity and spectrum quality without affecting the noncovalent complexes stability. For ligand-interaction analysis, ligands were added to the proteins in a 2- to 20-fold molar excess.

Crystallization and Structure Determination. The crystallization experiments were carried out by hanging drop vapor diffusion at 290 K by mixing equal volume (1 μL) of protein−LCA complex and reservoir solution (0.1 M BTP, pH 7.0, 3 M NaOAc). The crystals of the complex were transferred to artificial mother liquor containing 15% glycerol and flash-cooled in liquid nitrogen. Data from crystals of zVDR-LCA were collected on the ID29 beamline at the ESRF. The raw data were processed and scaled with the HKL2000 program suite.[29] The structure was solved and refined using CNS[30] and iterative model building using COOT.[31] Crystallographic refinement statistics are presented in Table 1. All structural figures were prepared using PyMOL (www.pymol.org/).

Molecular Dynamics. The X-ray crystal structure of the zVDR complexed to two LCA molecules and a SRC-2 coactivator peptide was used as a starting point for all calculations. All heavy atoms were present in the experimental coordinate files except the region between helix H2 and H3 (Arg190−Arg251). As this loop is too long (about 60 residues) to be fully reconstructed and not relevant in this study, we replaced the missing sequence by four Gly residues. This strategy is justified by the fact that an experimentally analogous variant was used in earlier studies of hVDR and rVDR mutants.[9,32] This construction was built using the Modeller program.[33] The protonation states of all His residues at physiological pH (7.4) were obtained by PROPKA, an empirical method for pK_a calculation.[34] All His residues were neutral. The final construction of proton positions was done using the HBUILD module[35] in the CHARMM program.[36] The topology and parameters of the LCA ligand were obtained from the CHARMM generalized force field (CGENFF)[37] and were used without any additional optimization. The CHARMM36 all-atom force field was used for the protein.[38]

To study the effects of LCA bound to the second site on the structural dynamics of the zVDR LBD, molecular dynamics simulations were run on zVDR LBD in complex with two LCA ligands and on the same zVDR LBD where the second binding site LCA was removed. The coactivator peptide was present in both complexes, and both systems were subjected to the same simulation protocol, which follows one used in an earlier study of a nuclear receptor LBD.[39] The NAMD 2.9 program[40] was used for all molecular dynamics simulations, and the CHARMM program[36] was used for the analysis. The simulations were conducted under NVE conditions at 298.15 K with a physiological ionic strength of approximately 0.15 M arrived at by adding Na^+/Cl^- counterions to the simulation box. A production run of 40 ns was simulated, and the first 10 ns of each simulation was excluded from the analysis. This protocol was repeated twice for both the zVDR with two LCA ligands and zVDR with only one LCA ligand yielding 80 ns total simulation time for each form. For each simulation, the backbone root-mean-square coordinate difference (rmsd) was calculated, as well as the backbone atomic root-mean-square fluctuations (rmsf) using standard procedures.[39] The computation of the rmsd was done relative to the initial protein structure, which was the same in all simulations, and the structures from the trajectories were reoriented on the backbone of the initial structure. The rmsf values around average structures were computed for 10 segments of 3 ns extracted from the last 30 ns of each trajectory. For each 3 ns trajectory segment, an average structure and rmsf around that structure were computed. The rmsf values were averaged by residue and compared to the experimental *B*-factors. For the complex of zVDR with two LCA ligands, the van der Waals and electrostatic interaction energies between the protein and the LCA at the second site were computed using the CHARMM program, as were the distances between atom pairs LCA hydroxyl oxygen and zSer263, LCA

hydroxyl oxygen and zGln267, and LCA carboxyl carbon and C18-zLys268.

■ ASSOCIATED CONTENT

Supporting Information

Eight figures showing crystal packing, ESI-MS results, rms results, interaction energies, amino acid sequence alignment, and docking results. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The atomic coordinates and structure factors have been deposited in the Protein Data Bank with entry code 4Q0A.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 33388655293. E-mail: rochel@igbmc.fr.

Author Contributions

A.Y.B. and N.R. designed the experiments. A.Y.B. purified proteins and carried out biophysical, functional, and structural experiments. N.P. carried out the ESI-MS experiments. J.E., R.H.S., and A.D. designed the molecular dynamics experiments. J.E. executed the molecular simulations and the analysis. J.E., R.H.S., and A.D. contributed to the manuscript. A.Y.B. and N.R. analyzed the data and wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dino Moras of the Institut de Génétique et de Biologie Moleculaire et Cellulaire (IGBMC) for support and Alastair McEwen (IGBMC) for his kind assistance during X-ray data collections and corrections of the manuscript. We are grateful to Peter Jurutka (Arizona State University) for the *CYP3A4* reporter gene. We thank the staff of the beamline at the ESRF for the experimental assistance during data collection, Pascal Eberling (IGBMC) for the peptide synthesis, and the staff of the cell culture facility from IGBMC for technical assistance.

The project was supported by the Fondation pour la Recherche Médicale (FRM) and the Agence Nationale de Recherche (Grant ANR-13-BSV8-0024-01). This work was also supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Université de Strasbourg (UDS), Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions, the French Infrastructure for Integrated Structural Biology (FRISBI) (Grant ANR-10-INSB-05-01). J.E. was supported by a grant from the French Ministry of Education and Research. Computing time was provided at the French national computing centers by GENCI (Grand Equipement National de Calcul Intensif) and the Meso-Center for High Performance Computing at the University of Strasbourg, France, and supported by the project EQUIP@MESO.

■ ABBREVIATIONS USED

calcitriol, 1α,25-dihydroxyvitamin D3; LCA, lithocholic acid; VDR, vitamin D receptor; LBD, ligand binding domain; LBP, ligand binding pocket; rmsd, root-mean-square deviation; rmsf, root-mean-square fluctuation; ITC, isothermal titration calorimetry; MS, mass-spectrometry

■ REFERENCES

(1) Haussler, M. R.; Jurutka, P. W.; Mizwicki, M.; Norman, A. W. Vitamin D receptor (VDR)-mediated actions of 1alpha,25-$(OH)_2$vitamin D3: genomic and non-genomic mechanisms. *Best Pract. Res., Clin. Endocrinol. Metab.* **2011**, *25*, 543−559.

(2) Campbell, M. J.; Adorini, L. The vitamin D receptor as a therapeutic target. *Expert Opin. Ther. Targets* **2006**, *10*, 735−748.

(3) (a) Makishima, M.; Lu, T. T.; Xie, W.; Whitfield, G. K.; Domoto, H.; Evans, R. M.; Haussler, M. R.; Mangelsdorf, D. J. Vitamin D receptor as an intestinal bile acid sensor. *Science* **2002**, *296*, 1313−1316. (b) Adachi, R.; Shulman, A. I.; Yamamoto, K.; Shimomura, I.; Yamada, S.; Mangelsdorf, D. J.; Makishima, M. Structural determinants for vitamin D receptor response to endocrine and xenobiotic signals. *Mol. Endocrinol.* **2004**, *18*, 43−52. (c) Adachi, R.; Honma, Y.; Masuno, H.; Kawana, K.; Shimomura, I.; Yamada, S.; Makishima, M. Selective activation of vitamin D receptor by lithocholic acid acetate, a bile acid derivative. *J. Lipid Res.* **2005**, *46*, 46−57.

(4) Hofmann, A. F.; Hagey, L. R. Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cell. Mol. Life Sci.* **2008**, *65*, 2461−2483.

(5) (a) Matsubara, T.; Yoshinari, K.; Aoyama, K.; Sugawara, M.; Sekiya, Y.; Nagata, K.; Yamazoe, Y. Role of vitamin D receptor in the lithocholic acid-mediated CYP3A induction in vitro and in vivo. *Drug Metab. Dispos.* **2008**, *36*, 2058−2063. (b) Pavek, P.; Pospechova, K.; Svecova, L.; Syrova, Z.; Stejskalova, L.; Blazkova, J.; Dvorak, Z.; Blahos, J. Intestinal cell-specific vitamin D receptor (VDR)-mediated transcriptional regulation of CYP3A4 gene. *Biochem. Pharmacol.* **2010**, *79*, 277−287.

(6) Nehring, J. A.; Zierold, C.; DeLuca, H. F. Lithocholic acid can carry out in vivo functions of vitamin D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10006−10009.

(7) Ishizawa, M.; Matsunawa, M.; Adachi, R.; Uno, S.; Ikeda, K.; Masuno, H.; Shimizu, M.; Iwasaki, K.; Yamada, S.; Makishima, M. Lithocholic acid derivatives act as selective vitamin D receptor modulators without inducing hypercalcemia. *J. Lipid Res.* **2008**, *49*, 763−772.

(8) (a) Rochel, N.; Moras, D. Ligand binding domain of vitamin D receptors. *Curr. Top Med. Chem.* **2006**, *6*, 1229−1241. (b) Carlberg, C.; Molnar, F.; Mourino, A. Vitamin D receptor ligands: the impact of crystal structures. *Expert Opin. Ther. Pat.* **2012**, *22*, 417−435.

(9) (a) Rochel, N.; Wurtz, J. M.; Mitschler, A.; Klaholz, B.; Moras, D. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol. Cell* **2000**, *5*, 173−179. (b) Vanhooke, J. L.; Benning, M. M.; Bauer, C. B.; Pike, J. W.; DeLuca, H. F. Molecular structure of the rat vitamin D receptor ligand binding domain complexed with 2-carbon-substituted vitamin D3 hormone analogues and a LXXLL-containing coactivator peptide. *Biochemistry* **2004**, *43*, 4101−4110. (c) Ciesielski, F.; Rochel, N.; Moras, D. Adaptability of the Vitamin D nuclear receptor to the synthetic ligand Gemini: remodelling the LBP with one side chain rotation. *J. Steroid Biochem. Mol. Biol.* **2007**, *103*, 235−242.

(10) Choi, M.; Yamamoto, K.; Itoh, T.; Makishima, M.; Mangelsdorf, D. J.; Moras, D.; DeLuca, H. F.; Yamada, S. Interaction between vitamin D receptor and vitamin D ligands: two-dimensional alanine scanning mutational analysis. *Chem. Biol.* **2003**, *10*, 261−270.

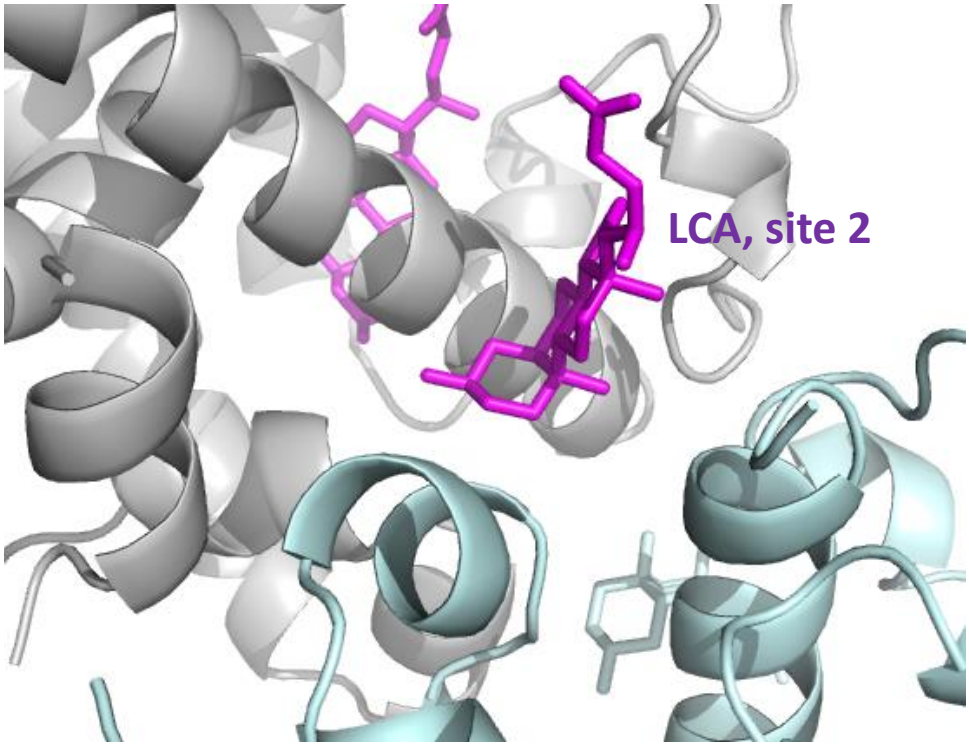
(11) Masuno, H.; Ikura, T.; Morizono, D.; Orita, I.; Yamada, S.; Shimizu, M.; Ito, N. Crystal structures of complexes of vitamin D receptor ligand-binding domain with lithocholic acid derivatives. *J. Lipid Res.* **2013**, *54*, 2206−2213.

(12) Ciesielski, F.; Rochel, N.; Mitschler, A.; Kouzmenko, A.; Moras, D. Structural investigation of the ligand binding domain of the zebrafish VDR in complexes with 1alpha,25(OH)2D3 and Gemini: purification, crystallization and preliminary X-ray diffraction analysis. *J. Steroid Biochem. Mol. Biol.* **2004**, *89−90*, 55−59.

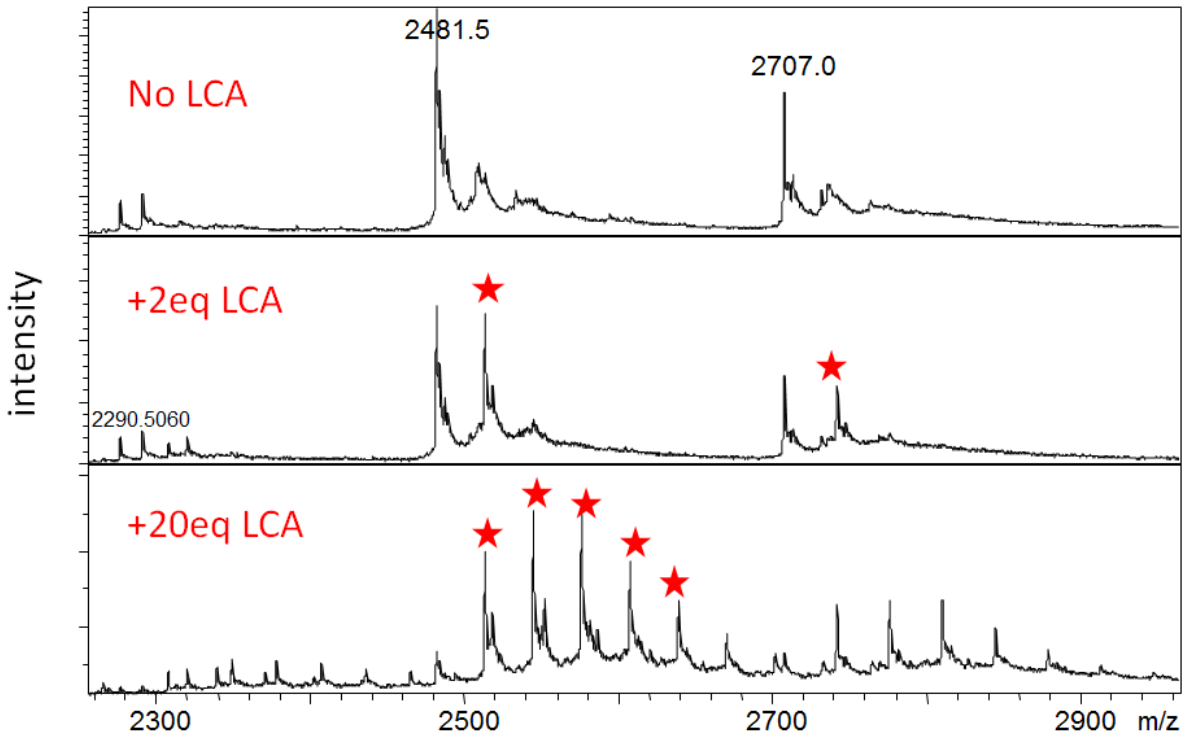
(13) Reschly, E. J.; Bainy, A. C.; Mattos, J. J.; Hagey, L. R.; Bahary, N.; Mada, S. R.; Ou, J.; Venkataramanan, R.; Krasowski, M. D. Functional evolution of the vitamin D and pregnane X receptors. *BMC Evol. Biol.* **2007**, *7*, 222.

(14) Antony, P.; Sigueiro, R.; Huet, T.; Sato, Y.; Ramalanjaona, N.; Rodrigues, L. C.; Mourino, A.; Moras, D.; Rochel, N. Structure−function relationships and crystal structures of the vitamin D receptor bound 2 alpha-methyl-(20S,23S)- and 2 alpha-methyl-(20S,23R)-epoxymethano-1 alpha,25-dihydroxyvitamin D3. *J. Med. Chem.* **2010**, *53*, 1159−1171.
(15) Kozoni, V.; Tsioulias, G.; Shiff, S.; Rigas, B. The effect of lithocholic acid on proliferation and apoptosis during the early stages of colon carcinogenesis: differential effect on apoptosis in the presence of a colon carcinogen. *Carcinogenesis* **2000**, *21*, 999−1005.
(16) Cheng, J.; Fang, Z. Z.; Kim, J. H.; Krausz, K. W.; Tanaka, N.; Chiang, J. Y.; Gonzalez, F. J. Intestinal CYP3A4 protects against lithocholic acid-induced hepatotoxicity in vitamin D receptor intestine-deficient mice. *J. Lipid Res.* **2013**, *55*, 455−465.
(17) Han, S.; Li, T.; Ellis, E.; Strom, S.; Chiang, J. Y. A novel bile acid-activated vitamin D receptor signaling in human hepatocytes. *Mol. Endocrinol.* **2010**, *24*, 1151−1164.
(18) (a) Choi, M.; Yamada, S.; Makishima, M. Dynamic and ligand-selective interactions of vitamin d receptor with retinoid x receptor and cofactors in living cells. *Mol. Pharmacol.* **2011**, *80*, 1147−1155. (b) Jurutka, P. W.; Thompson, P. D.; Whitfield, G. K.; Eichhorst, K. R.; Hall, N.; Dominguez, C. E.; Hsieh, J. C.; Haussler, C. A.; Haussler, M. R. Molecular and functional comparison of 1,25-dihydroxyvitamin D(3) and the novel vitamin D receptor ligand, lithocholic acid, in activating transcription of cytochrome P450 3A4. *J. Cell. Biochem* **2005**, *94*, 917−943.
(19) Li, T.; Chiang, J. Y. Nuclear receptors in bile acid metabolism. *Drug Metab. Rev.* **2013**, *45*, 145−155.
(20) Mi, L. Z.; Devarakonda, S.; Harp, J. M.; Han, Q.; Pellicciari, R.; Willson, T. M.; Khorasanizadeh, S.; Rastinejad, F. Structural basis for bile acid binding and activation of the nuclear receptor FXR. *Mol. Cell* **2003**, *11*, 1093−1100.
(21) Egan, J. B.; Thompson, P. A.; Vitanov, M. V.; Bartik, L.; Jacobs, E. T.; Haussler, M. R.; Gerner, E. W.; Jurutka, P. W. Vitamin D receptor ligands, adenomatous polyposis coli, and the vitamin D receptor FokI polymorphism collectively modulate beta-catenin activity in colon cancer cells. *Mol. Carcinog.* **2010**, *49*, 337−352.
(22) (a) Hsieh, J. C.; Sisk, J. M.; Jurutka, P. W.; Haussler, C. A.; Slater, S. A.; Haussler, M. R.; Thompson, C. C. Physical and functional interaction between the vitamin D receptor and hairless corepressor, two proteins required for hair cycling. *J. Biol. Chem.* **2003**, *278*, 38665−38674. (b) Chuma, M.; Endo-Umeda, K.; Shimba, S.; Yamada, S.; Makishima, M. Hairless modulates ligand-dependent activation of the vitamin D receptor-retinoid X receptor heterodimer. *Biol. Pharm. Bull.* **2012**, *35*, 582−587.
(23) Mizwicki, M. T.; Keidel, D.; Bula, C. M.; Bishop, J. E.; Zanello, L. P.; Wurtz, J. M.; Moras, D.; Norman, A. W. Identification of an alternative ligand-binding pocket in the nuclear vitamin D receptor and its functional importance in 1alpha,25(OH)2-vitamin D3 signaling. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12876−12881.
(24) Wang, Y.; Chirgadze, N. Y.; Briggs, S. L.; Khan, S.; Jensen, E. V.; Burris, T. P. A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor beta. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9908−9911.
(25) Hwang, J. Y.; Arnold, L. A.; Zhu, F.; Kosinski, A.; Mangano, T. J.; Setola, V.; Roth, B. L.; Guy, R. K. Improvement of pharmacological properties of irreversible thyroid receptor coactivator binding inhibitors. *J. Med. Chem.* **2009**, *52*, 3892−3901.
(26) Estebanez-Perpina, E.; Arnold, L. A.; Nguyen, P.; Rodrigues, E. D.; Mar, E.; Bateman, R.; Pallai, P.; Shokat, K. M.; Baxter, J. D.; Guy, R. K.; Webb, P.; Fletterick, R. J. A surface on the androgen receptor that allosterically regulates coactivator binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16074−16079.
(27) Bernardes, A.; Souza, P. C.; Muniz, J. R.; Ricci, C. G.; Ayers, S. D.; Parekh, N. M.; Godoy, A. S.; Trivella, D. B.; Reinach, P.; Webb, P.; Skaf, M. S.; Polikarpov, I. Molecular mechanism of peroxisome proliferator-activated receptor alpha activation by WY14643: a new mode of ligand recognition and receptor stabilization. *J. Mol. Biol.* **2013**, *425*, 2878−2893.
(28) Rochel, N.; Tocchini-Valentini, G.; Egea, P. F.; Juntunen, K.; Garnier, J. M.; Vihko, P.; Moras, D. Functional and structural characterization of the insertion region in the ligand binding domain of the vitamin D nuclear receptor. *Eur. J. Biochem.* **2001**, *268*, 971−979.
(29) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307−326.
(30) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect.D: Biol. Crystallogr.* **1998**, *54*, 905−921.
(31) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect.D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(32) (a) Perakyla, M.; Malinen, M.; Herzig, K. H.; Carlberg, C. Gene regulatory potential of nonsteroidal vitamin D receptor ligands. *Mol. Endocrinol.* **2005**, *19*, 2060−2073. (b) Perakyla, M. Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. *Eur. Biophys J.* **2009**, *38*, 185−198.
(33) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinf.* **2006**, No. Chapter 5, Unit 5.6.
(34) Rostkowski, M.; Olsson, M. H.; Sondergaard, C. R.; Jensen, J. H. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **2011**, *11*, 6.
(35) Brunger, A. T.; Karplus, M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins* **1988**, *4*, 148−156.
(36) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(37) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(38) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(39) Chebaro, Y.; Amal, I.; Rochel, N.; Rochette-Egly, C.; Stote, R. H.; Dejaegere, A. Phosphorylation of the retinoic acid receptor alpha induces a mechanical allosteric regulation and changes in internal dynamics. *PLoS Comput. Biol.* **2013**, *9*, e1003012.
(40) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

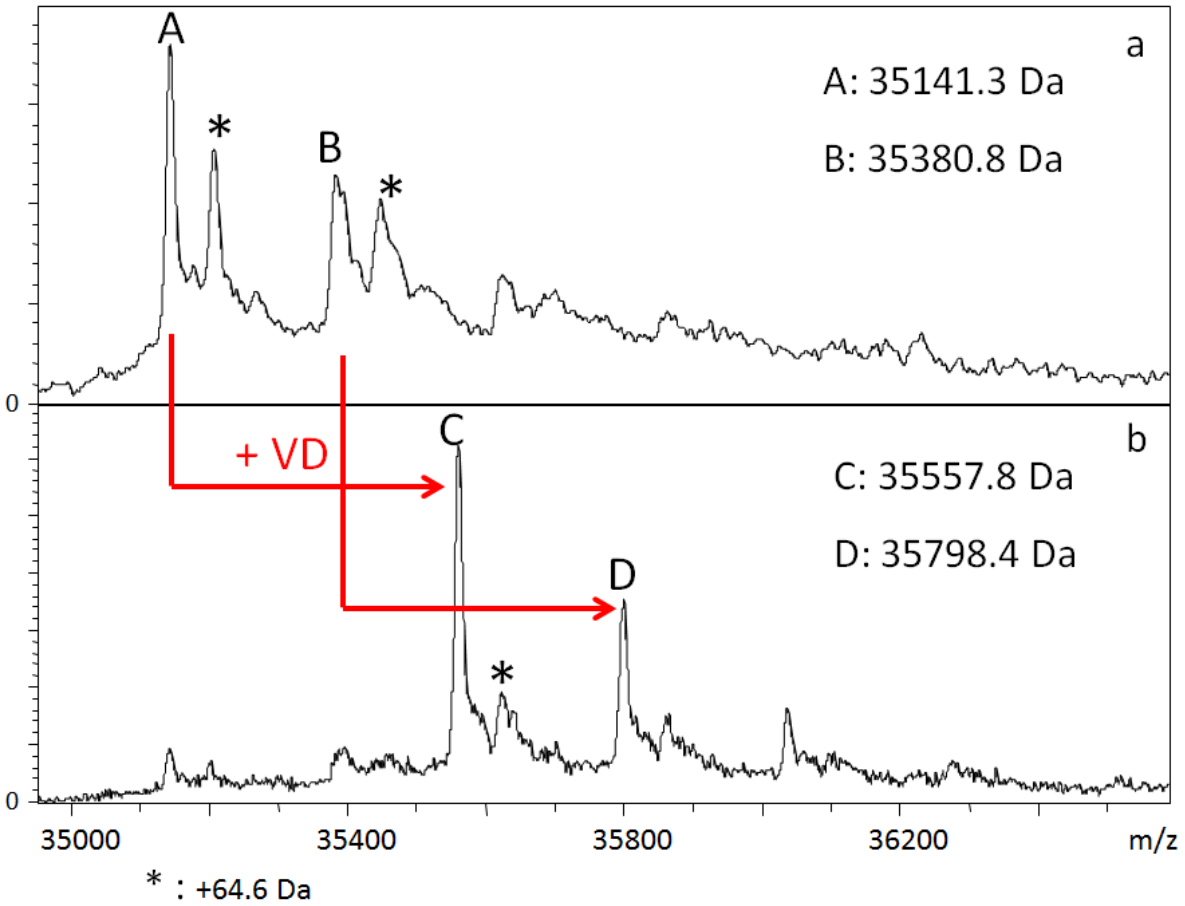
Supplementary Figure 1. Crystal packing of zVDR-LCA complex around the second LCA binding site. The asymmetric unit content is shown in grey and the symmetrical molecule in cyan.



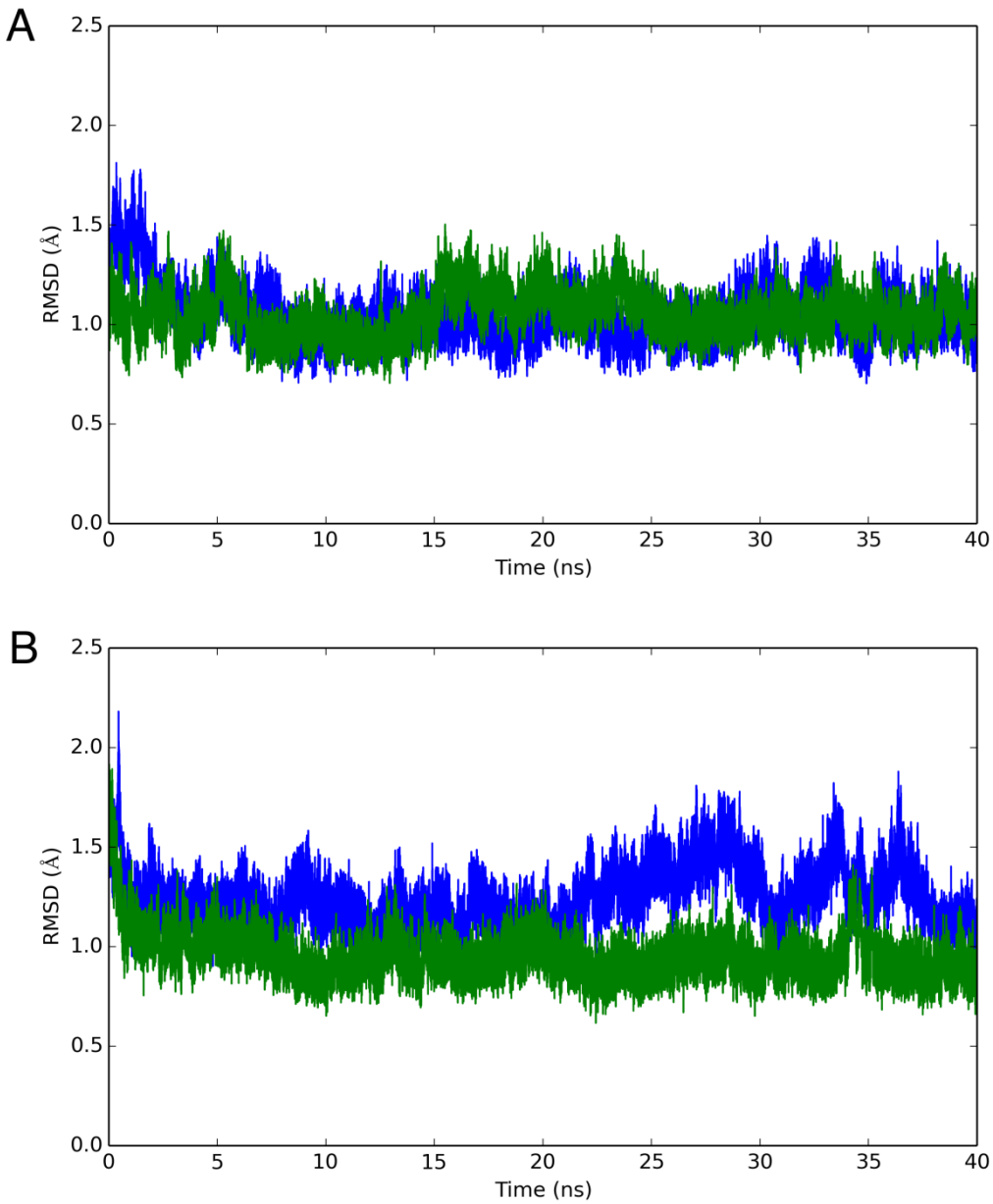
Supplementary Figure 2. Non-denaturing ESI-MS analysis of LCA binding to zVDR. Enlarged view of the 12+, 13+ and 14+ ions of the electrospray ionization mass spectrum of zVDR LBD in the absence of ligand and in the presence of two-fold or twenty-fold molar excess of LCA. Red stars indicate the VDR-LCA complex. In presence of twenty-fold excess of LCA, multi-addition of LCA is observed.



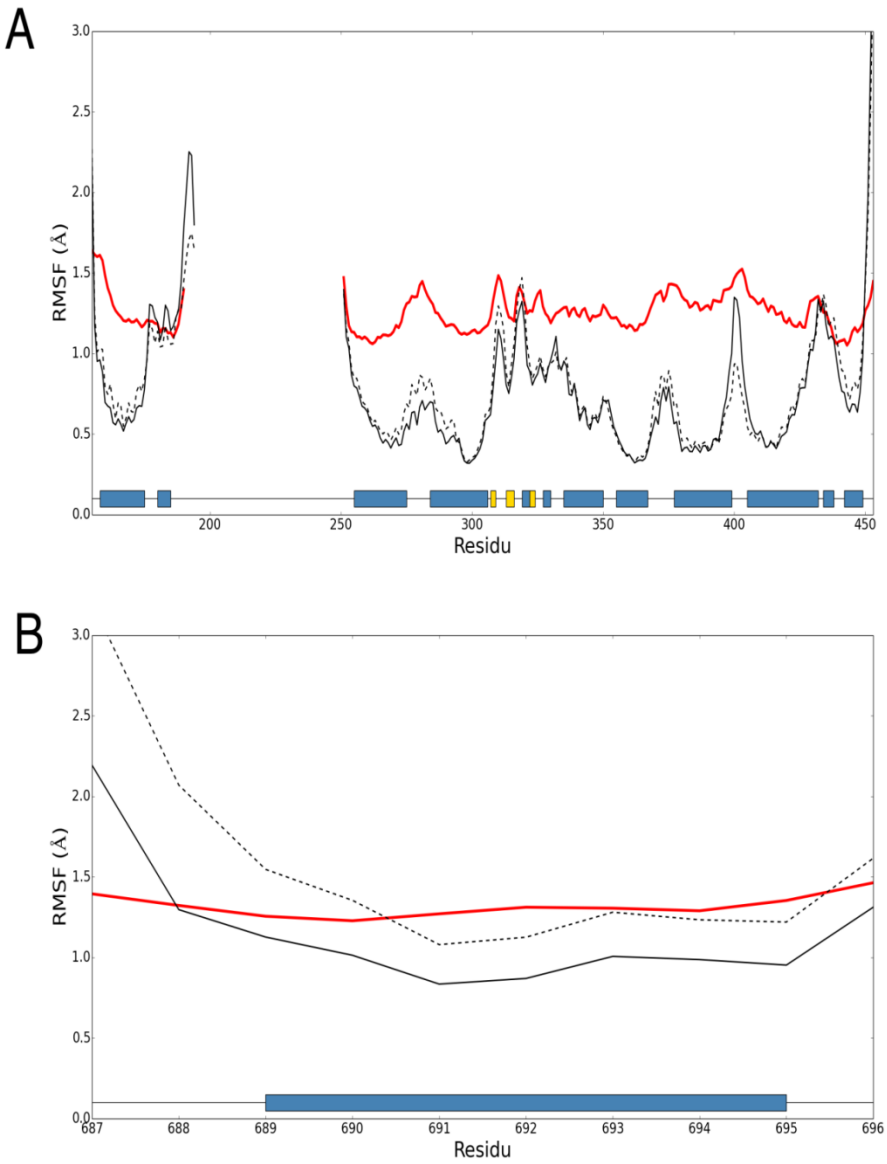
Supplementary Figure 3. Non-denaturing ESI-MS analysis of ligand binding to hVDR Ser235Ala, Gln239Ala mutant. (a-b) Deconvoluted mass spectrum of hVDR LBD mutant in the absence of ligand (a) and in presence of twofold molar excess of calcitriol(VD) (b) Arrows indicate the shift of the peak upon ligand binding. Peaks labeled with A, B or * correspond to different truncated species.



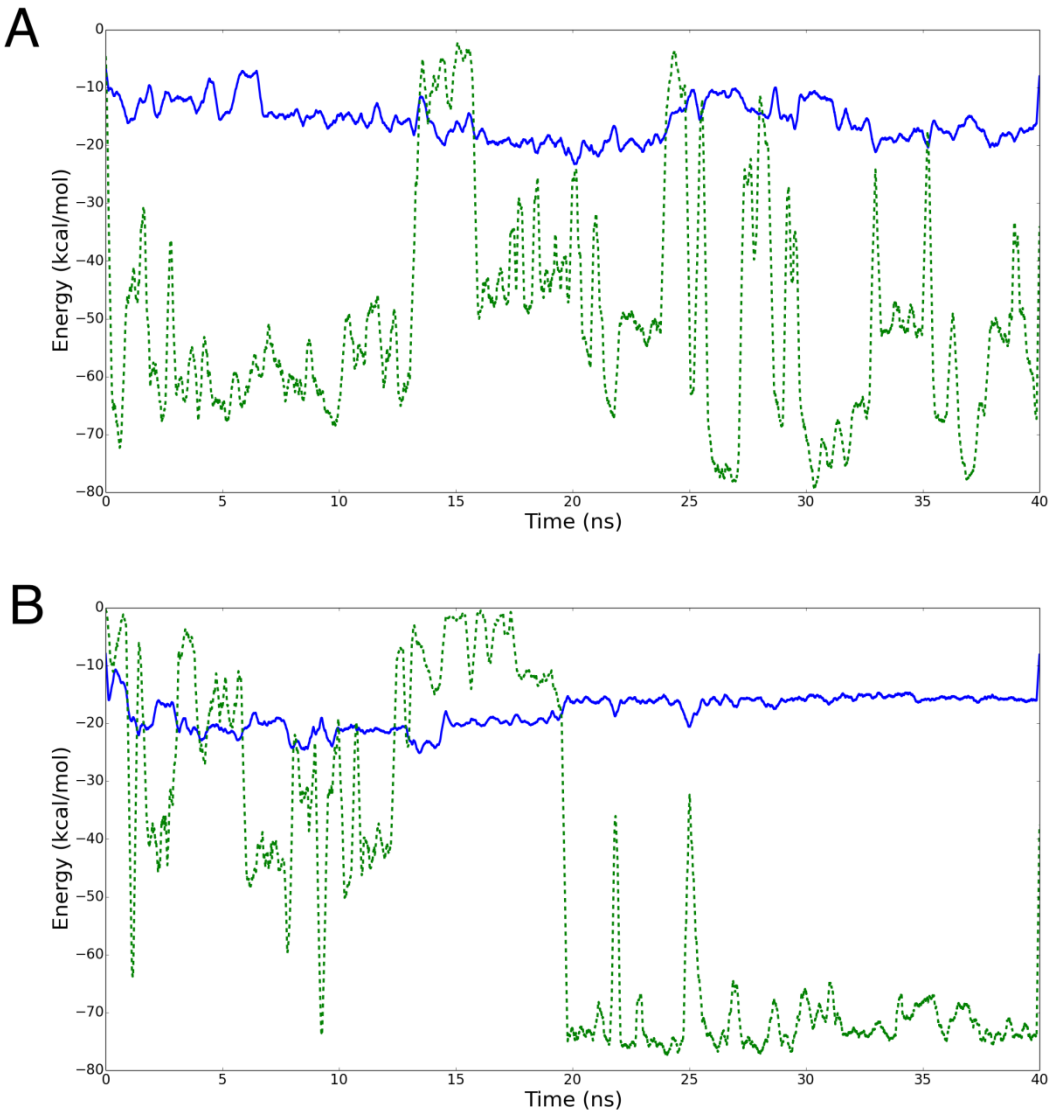
Supplementary Figure 4. RMS deviation (in angstrom) from the initial crystallographic structure of zVDR with two ligands, for the two simulations with one LCA ligand (A) and the two simulations with two ligands (B).



Supplementary Figure 5. **(A)** RMS atomic fluctuations of the backbone atoms of zVDR LBDs in the presence of one (black dotted line) and two (black full line) LCA ligands. Global motion was removed by reorienting trajectory frames onto the backbone of the initial LBD structure. Fluctuations were calculated from the final 30 ns of the molecular dynamics simulations and given in Angstrom (see Text for details). The fluctuations obtained from the experimental B-factor of the crystallographic structures of zVDR with two ligands are given in red. Residues number and positions of the secondary structure elements are schematized along the x-axis (α-helices in blue, β-sheet in yellow) and the L9-10 is situated around residue 400. (**B**) Fluctuations of the backbone atoms of the SRC-2 coactivator peptide bound to zVDR LBD in the presence of one (black dotted line) and two (black full line) LCA ligands. The helical segment of SRC-2 extends between AA 689 and 695.



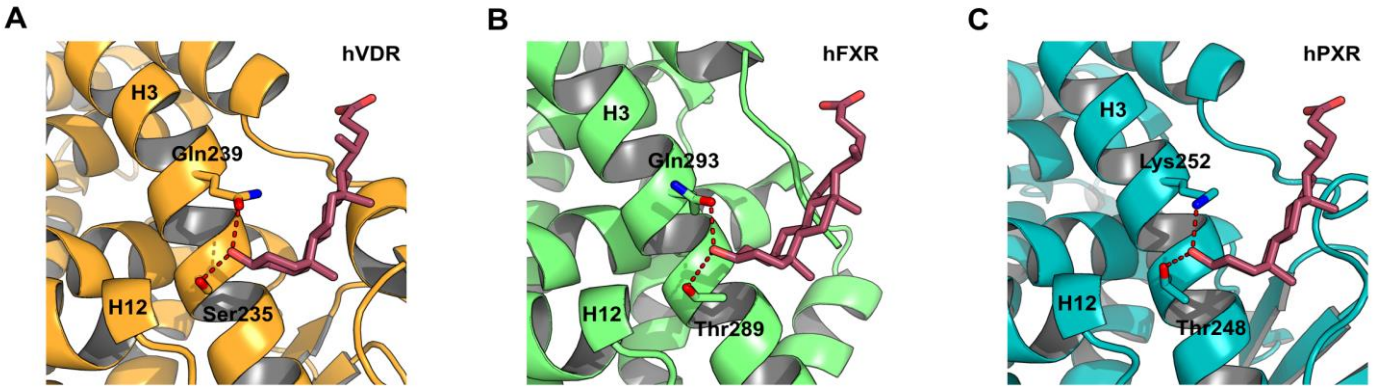
Supplementary Figure 6. Analysis of the interaction energies. Van der Waals (blue full line) and electrostatic (green dotted line) interaction energies between the second LCA ligand and zVDR-SRC-2 coactivator peptide complex. The interactions energies (kcal/mol) are displayed as a function of time (ns) for the two molecular dynamics trajectories (A, B). The interaction energies, specially the electrostatic component, show large fluctuations, but remain favorable throughout the simulations.



Supplementary Figure 7. Amino acid sequence alignment generated by Clustal Omega. The residues are colored according to their physico-chemical properties using Clustal color code. VDR residues forming direct H-bonds with LCA and corresponding amino acids of FXR and PXR are framed.

```
zVDR|260-267        DLVSYSIQ
hVDR|232-239        DLVSYSIQ
hFXR|303-310        EMATNHVQ
hPXR|245-252        DMSTYMFK
                    :: :  .:
```

Supplementary Figure 8. Docking of the second LCA molecule to hVDR (A), hFXR (B) and hPXR (C). Using the PyMOL program, the crystallographic structures of human VDR (4G2I), FXR (1OSH) and PXR (1ILG) were superposed onto the structure of zebrafish VDR with two LCA ligands. The coordinates of the second LCA ligand were then transferred from the zVDR to the hVDR, hFXR and hPXR. This simple comparative docking procedure has been shown to provide good quality starting structures for complexes (Martin L, Catherinot V, Labesse G. *kinDOCK: a tool for comparative docking of protein-kinase ligands*. *Nucleic Acid Res*. *2006;34:W325-W329)*. The resulting complex was then energy minimized in vacuum using the CHARMM software (300 steps of Steepest Descent followed by 900 steps of Conjugate Gradient). Harmonic constraints derived from the crystallographic B-factors were applied during the energy minimization to the carboxyl atoms (C18, O21 and O22) of the LCA ligand to retain their original orientation towards the solvent, which could not be correctly modeled due to the absence of explicit water molecules around the structures. The remainder of the structure (ligand and protein) was fully relaxed. The figure presents the resulting minimized structures for LCA (second site) bound to hVDR (A), hFXR (B) and hPXR (C). The hydrogen bonds between the LCA hydroxyl group of the A ring and the amino acids equivalent to zVDR Ser 263 and zVDR Gln 267 are highlighted in dotted red.



UNIVERSITÉ DE STRASBOURG

Jérôme EBERHARDT

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Étude de la dynamique structurale du domaine de liaison au ligand de RXRα et implication de la phosphorylation dans la transcription

De nombreuses études révèlent que le domaine de liaison au ligand de RXRα est très dynamique, même en présence d'un ligand agoniste. Nous avons utilisé les données expérimentales (HDX, RMN et Xray) disponibles sur ce domaine pour mettre en place un protocole, basé sur la dynamique moléculaire accélérée, permettant d'explorer efficacement la dynamique conformationnelle du domaine de liaison au ligand de RXRα et de valider les ensembles conformationnels obtenus. Ce protocole a été appliqué pour analyser l'influence de la phosphorylation pSer260, située à proximité de la surface d'interaction avec les protéines coactivatrice et impliquée dans le développement de carcinomes hépatocellulaires, sur la structure de ce domaine et sa dynamique. Parallèlement, une méthode de réduction de la dimensionnalité a été développé afin d'analyser de longues trajectoires de dynamique moléculaire. Ainsi grâce à cette méthode, nous avons pu identifier plusieurs nouvelles conformations alternative stables du domaine de liaison au ligand de RXRα.

Many studies reveal that the ligand binding domain of RXRα is very dynamic, still even in a presence of an agonist ligand. Therefore, the availability of experimental data (HDX, NMR and Xray) on the domain was used as a leverage in order to set up a protocol, based on accelerated molecular dynamics, to explore its conformational dynamic and to validate it. This protocol was applied to understand the influence of the pSer260 phosphorylation, closed to the binding surface of coactivator proteins and implied in the hepatocellular carcinoma growth, on its structure and its dynamic. At the same time, a dimensional reduction method was developed to analyse long molecular dynamic trajectories. Thus, with this approach, we identified a couple of new alternative and stable conformations of the ligand binding domain of RXRα.