



**HAL**  
open science

# Recherche par modélisation moléculaire de signatures RMN et DC caractéristiques pour les coudes $\beta$ et $\gamma$ dans les peptides bioactifs.

Mattia Migliore

## ► To cite this version:

Mattia Migliore. Recherche par modélisation moléculaire de signatures RMN et DC caractéristiques pour les coudes  $\beta$  et  $\gamma$  dans les peptides bioactifs.. Chimie analytique. Normandie Université, 2020. Français. NNT : 2020NORMR001 . tel-03112593

**HAL Id: tel-03112593**

**<https://theses.hal.science/tel-03112593v1>**

Submitted on 17 Jan 2021

**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.



Normandie Université

# THÈSE

Pour obtenir le diplôme de doctorat

Spécialité Chimie

Préparée au sein de l'Université de Rouen Normandie

**Recherche par modélisation moléculaire de signatures RMN et DC caractéristiques pour les coudes  $\beta$  et  $\gamma$  dans les peptides bioactifs**

Présentée et soutenue par  
**Mattia MIGLIORE**

Thèse soutenue publiquement le 16 janvier 2020  
devant le jury composé de

M. Alexandre DE BREVERN	Directeur de Recherche Université Paris Diderot	Rapporteur
M. Sylvain BROUSSY	Maître de Conférence Université Paris Descartes	Rapporteur
M. Philippe JUBAULT	Professeur des Universités, Université de Rouen Normandie	Examineur
M. Olivier LEQUIN	Professeur des Universités, Université Pierre et Marie Curie	Examineur
Mme Laure GUILHAUDIS	Maître de Conférence Université de Rouen Normandie	Codirecteur de thèse
Mme Isabelle SEGALAS-MILAZZO	Professeur des Universités, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par le Pr. Isabelle SEGALAS-MILAZZO et le Dr. Laure GUILHAUDIS, laboratoire COBRA, UMR 6014, CNRS





***There is a rough analogy between proteins structures and a text.***

*The amino acids correspond to letters, the secondary structures to words, super-secondary structures to phrases (or even to cliché), elements of tertiary structure to sentences – this is the level at which true individuality makes its appearance- domains to paragraphs, the structure of a fully polypeptide chain to a chapter, and the quaternary structure to the assembly of chapter into a complete book.*

*“Nothing in life is to be feared,  
it is only to be understood.  
Now is the time to understand more,  
so that we may fear loss.”  
Marie Curie*



*A mes deux grands-pères*  
*A mes deux grands-mères*  
*A mes parents Bruna et Franco*



## *Remerciements*

Tout d'abord, je tiens à adresser mes sincères remerciements au Dr. Alexandre DE BREVERN et au Dr. Sylvain BROUSSY pour avoir accepté d'être rapporteurs de ce mémoire. Je voudrais également remercier le Pr. Olivier LEQUIN d'avoir accepté d'examiner ces travaux de thèse. Je voudrais, en outre, exprimer mes remerciements au Pr Philippe JUBAULT d'avoir accepté d'être examinateur de mon jury de thèse.

Ce projet de thèse a été réalisé au sein de l'équipe « Analyse et modélisation » de l'UMR CNRS 6014 COBRA dirigée par le Pr. Hassan OULYADI. Je tiens à remercier Hassan de m'avoir accueilli au sein du laboratoire. Je vous remercie également pour le temps que vous nous avez consacré pendant les séminaires du groupe RMN, malgré votre emploi du temps très chargé, afin que l'on puisse partager les projets scientifiques, échanger et trouver les solutions aux problématiques rencontrées au cours des années et, enfin, d'aider la vie d'équipe. Je vous remercie pour avoir été toujours à l'écoute pendant ces trois années de thèse.

Un grand merci au Pr. Isabelle SEGALAS-MILAZZO et au Dr. Laure GUILHAUDIS pour avoir encadré ce travail de thèse. Merci à vous deux pour vos conseils, pour les nombreux échanges et pour le temps que vous m'avez dédié. Je vous remercie ensuite pour votre patience tout au long de ces trois années. Merci à vous Isabelle de m'avoir remotivé quand je perdais le moral quand quelque chose ne marchait pas. Je vous remercie énormément pour avoir toujours été disponibles. Merci à vous aussi Laure pour toutes les discussions que l'on a eu pendant ces trois années. Merci pour avoir été une guide dans ce travail de thèse. Merci beaucoup d'être resté jusqu'à tard au laboratoire afin de faire avancer ce projet de recherche.

Je souhaiterais également remercier le Dr. Muriel Sebban avec qui j'ai pu partager les enseignements pour les TP de RMN. Merci pour toutes vos explications, pour votre gentillesse et pour votre disponibilité chaque fois que j'en avais besoin.

Je tiens surtout à remercier le Pr. Laurent Joubert, le Dr. Vincent Tognetti ainsi qu'au Dr. Andrea Bonvicini pour la possibilité qui m'a été donnée de collaborer avec vous.

Je remercie particulièrement Andrea pour toutes les heures que l'on a passé dans la salle de modélisation à travailler ensemble sur Miri pour atteindre l'objectif d'une partie du travail de thèse et pour toutes les rigolades qu'on a fait même quand tout était compliqué ! Tu as été fondamental !!!!

Je tiens à remercier également tous les permanents de l'équipe Analyse et modélisation Gaël, Carlos, Isabelle, Hélène, Corinne, Marie et Albert.

Je voudrais maintenant à remercier tous les étudiants de l'équipe « Analyse et Modélisation » avec qui j'ai partagé beaucoup de moments. Asma, merci pour les années passées ensemble dans le bureau. Tu m'as toujours aidé. Merci pour ta joie, ta bonne humeur et pour toutes les rigolades. T'es une personne en or !!! Riham (Riri), ma marraine au labo, merci pour ton amitié, pour ta disponibilité, pour ton soutien, les supers soirées et pour ton franc parler ! T'as été une marraine parfaite ! Je ne pouvais pas rêver mieux ! Sébastien (Séb), le seul français dans le groupe pour presque deux années ! Merci beaucoup pour ton sourire et pour ton épaule sur laquelle on peut toujours s'appuyer dès que quelque chose ne va pas trop ! T'es une personne formidable ! Andrea (Andreeeee), le premier italien arrivé au RDC ! En fait, merci pour tout ! Merci pour avoir été ami, collègue, pote de soirées, pour ta confiance ! T'es une personne géniale ! Un grand merci va à Lina pour ton aide, pour ton sourire, ta bonne humeur au labo, pour les soirées et pour ton soutien chaque jour. Merci pour la remotivation que t'as su me donner quand j'étais au plus bas. T'es une vraie amie. Matthieu, merci pour tous nos discussions, tes blagues, nos pauses et ton soutien pendant ces trois années de thèse. Amélie, la « nouvelle » doctorante dans le bureau ! Merci pour ta joie, ton soutien, ta franchise. Je t'encourage à rester comme tu es : une vraie guerrière ! A tous les deux, merci pour toutes les relectures au manuscrit et pour nos délires ! Guillaume, merci beaucoup pour tous nos débats, nos discussions, nos soirées au « ghetto ». Ta bonne humeur, tes blagues, tes mots encourageants et ton amitié ont été toujours présents. Anaïs, ta gentillesse, ta bonne humeur, ta gaieté et ton amitié ont rempli les moments compliqués ! Valentina (Vale), la troisième italienne du RDC. Merci parce que chaque jour t'as été une personne sincère avec moi ! Merci pour tout le temps passé ensemble, pour les soirées italiennes, pour ta positivité et pour ton aide ! Johann et Julien, merci pour nos délires et nos fous rires. Merci à Sophie, tu m'as soutenu durant la dernière ligne droite. Merci aussi aux deux nouveaux Hend et Aël, les deux nouveaux du groupe, qui débutent leur thèse quand je termine la mienne.

Merci aussi à Daniele et Celeste, les deux biologistes au CURIB avec qui j'ai partagé des soirées italiennes. Vous m'avez toujours soutenu ! Merci

Je remercie également tous les massistes que je n'ai pas eu l'occasion de trop connaître : Valentin, Estelle, Clément, Olivier et Oscar.

Un grand merci à Silvia, Stefano, Aurora et Emiliano qui m'ont soutenu chaque jour de ces trois années de thèse de près ou de loin. Vous êtes des personnes formidables. J'ai une sœur, un frère et deux parents acquis !

Un grand merci à mes amies et mes amis italiens (Mony, Debby, Eccia, Vale Vale, Ila, Ally, Dondy, Cassi, Cicci, Giuly Sch, Mo' et Tiii) qui ont toujours été à mon côté grâce à leurs messages, à leurs appel-vidéos. Je vous aime tous.

Toutes ces années ne seraient pas été possibles si ma famille ne m'avait pas soutenu. Maman, Papà c'est seulement grâce à vous qu'aujourd'hui je suis là et je n'oublierai jamais tout ce que vous avez dû faire pour moi. Vous m'avez toujours appris à croire en moi-même, à montrer de quoi j'étais capable, même quand mon espoir n'était plus là. Ceci m'a permis d'avancer sans peur et avec fierté chaque jour dans ma vie. J'espère que vous serez fiers de votre « petit » Mattia. Merci au reste de ma famille, mes grands-mères, mes oncles, mes tantes et mes cousins pour vos encouragements.



## ***Table des matières***

<b><i>Liste des abréviations</i></b> .....	<b>17</b>
<b><i>Chapitre 1 : Introduction</i></b> .....	<b>21</b>
I.La structure des protéines et des peptides.....	22
I.1 La structure primaire .....	24
I.2 La structure secondaire .....	25
I.2.1 Les hélices .....	26
I.2.2 Les feuillets .....	30
I.3 La structure tertiaire .....	32
I.4 La structure quaternaire .....	32
II.Une structure particulière : les coudes .....	33
II.1 Les coudes $\gamma$ .....	33
II.2 Les coudes $\beta$ .....	35
II.2.1 Evolution de la classification des coudes $\beta$ .....	35
II.2.2 Caractéristiques des coudes $\beta$ .....	45
III.Méthodes d'analyse de la structure des protéines et des peptides.....	51
III.1 Méthodes pour l'étude de la structure secondaire .....	51
III.1.1 La Spectroscopie Infrarouge (IR) .....	51
III.1.2 Le dichroïsme circulaire .....	52
III.1.3 La RMN .....	53
III.2 Méthodes pour l'étude de la structure tertiaire.....	54
III.2.1 La cristallographie aux rayons X .....	54
III.2.2 La RMN en solution couplée à la modélisation moléculaire (MM) .....	54
III.2.3 Cryo-microscopie électronique (Cryo-EM) .....	55
III.3 Méthodes <i>in silico</i> .....	55
III.3.1 Les méthodes <i>ab initio</i> pour l'étude du repliement.....	56
III.3.2 Les méthodes basées sur la mécanique quantique .....	56
III.3.3 Modélisation par homologie .....	57
<b>Objectifs de la thèse</b> .....	<b>58</b>

## **Chapitre 2 : Techniques utilisées pour l'analyse structurale de peptides et protéines en solution... 63**

I. Le dichroïsme circulaire .....	63
I.1 Dichroïsme circulaire des protéines et peptides .....	63
II. La Résonance magnétique nucléaire (RMN) en solution des polypeptides .....	71
II.1 Le déplacement chimique secondaire .....	72
II.2 Le diagramme NOE .....	74
II.3 La constante de couplage $^3J_{\text{HN-H}\alpha}$ .....	78
III. Modélisation moléculaire sous contraintes .....	80
III.1 Le champ de forces .....	81
III.2 La fonction d'énergie $E_{\text{empirique}}$ .....	81
III.3 Les termes d'énergies de contraintes .....	85
III.4 Calculs des structures .....	86
III.5 Validations des structures .....	87
III.5.1 La valeur d'énergie et les violations .....	87
III.5.2 Le diagramme de Ramachandran .....	87
III.5.3 Les liaisons hydrogène .....	92
III.5.4 La convergence des structures .....	93

## **Chapitre 3 : Recherche expérimentale de paramètres RMN discriminatoires à travers des peptides modèles ..... 97**

I. Constitution de la banque de données et sélection des peptides modèles .....	98
II. Synthèse des peptides modèles .....	102
III. Étude expérimentale des peptides modèles .....	103
III.1 Analyse conformationnelle par DC des peptides modèles .....	103
III.1.1 Résultats et discussion .....	104
III.2 Etude par RMN des peptides 1IGD, 8ADH et 1MOL .....	110
III.2.1 Résultats et discussion .....	111
III.3 Conclusions .....	119
IV Matériels et méthodes .....	120
IV.1 Dichroïsme circulaire .....	120
IV.1.1 Condition d'acquisition et de traitement .....	120

IV.1.2 Préparation des échantillons .....	121
IV.2 Résonance Magnétique Nucléaire .....	121
IV.2.1 Conditions d'acquisition et de traitement .....	121
IV.2.2 Paramètres d'acquisition des spectres RMN des peptides modèles.....	122
<b>Chapitre 4 : Recherche de paramètres structuraux de RMN des coudes <math>\beta</math> et <math>\gamma</math> par modélisation moléculaire .....</b>	<b>125</b>
I.Méthodes mises en œuvre.....	125
I.1 Constitution de la base de données de peptides .....	125
I.2 Optimisation de la géométrie des peptides.....	133
I.2.1 Paramètres du recuit simulé.....	133
I.2.2 Contraintes de distance et d'angle utilisées pour les calculs de modélisation moléculaire ...	134
I.3 Structures finales.....	135
I.3.1 Constitution du jeu final de structures .....	135
I.3.2 Evaluation de la qualité des structures .....	135
I.4 Evaluation de la qualité de la base de données .....	136
I.4.1 Analyse des angles dièdres $\Phi$ et $\Psi$ des résidus centraux .....	136
I.4.2 Analyse des angles dièdres $\Phi$ et $\Psi$ du premier et du dernier résidu .....	139
II. Recherche de paramètres de RMN caractéristiques .....	141
II.1 Distances $^1\text{H}-^1\text{H}$ .....	142
II.2 Constantes de couplage $^3\text{J}_{\text{HN-H}\alpha}$ .....	143
III. Résultats .....	145
III.1 Les coudes $\beta$ .....	145
III.1.1 Cas général.....	145
III.1.1.1 Les distances .....	145
III.1.1.1.1 Distances entre les hydrogènes du squelette peptidique .....	145
III.1.1.1.2 Le diagramme NOE .....	148
III.1.1.1.3 Distances entre la chaîne latérale et le squelette peptidique .....	153
III.1.1.1.4 Influence du motif.....	157
III.1.1.1.4.1 Influence du premier résidu en conformation hélice $\alpha$ .....	162

III.1.1.1.4.2 Influence du premier résidu en conformation brin $\beta$ .....	167
III.1.1.2 La constante de couplage $^3J_{\text{HN-H}\alpha}$ .....	172
III.1.1.3 Combinaison distances inter hydrogènes et constante de couplage $^3J_{\text{HN-H}\alpha}$ .....	179
III.1.1.4 Comparaison avec les données de la littérature .....	181
III.1.2 Cas des coudes $\beta$ avec préférence de proline <i>trans</i> en position 2 .....	183
III.1.2.1 Analyse des distances et des diagrammes NOE.....	183
III.1.2.2 La constante de couplage $^3J_{\text{HN-H}\alpha}$ .....	191
III.1.2.3 Distinction avec les autres types de coude.....	191
III.2 Les coudes $\gamma$ .....	193
III.2.1 Le cas général .....	193
III.2.1.1 Les distances .....	193
III.2.1.1.1 Distances entre les hydrogènes du squelette peptidique .....	193
III.2.1.1.2 Le diagramme NOE .....	194
III.2.1.1.3 Distances entre les hydrogènes de la chaîne latérale et du squelette peptidique .....	195
III.2.1.2 Influence du motif.....	197
III.2.1.2.1 Influence du premier résidu en conformation hélice $\alpha$ .....	200
III.2.1.2.2 Influence du premier résidu en conformation brin $\beta$ .....	201
III.2.1.2 La constante de couplage .....	202
III.2.1.3 Combinaison distances inter protons et constante de couplage $^3J_{\text{HN-H}\alpha}$ .....	203
III.2.1.4 Conclusions.....	204
III.2.2 Le cas des coudes $\gamma$ inverse avec une proline en position 2.....	204
III.2.2.1 Analyse des distances et du diagramme NOE .....	204
III.2.2.2 Conclusion .....	206
IV.Conclusions .....	206
<b>Chapitre 5 : Recherche de paramètres de DC des coudes <math>\beta</math> de type I, I', II et II' par TDDFT .....</b>	<b>211</b>
I.Préambule .....	211
II.Théorie de la fonctionnelle de la densité ( <i>Density Functional Theory</i> , DFT).....	214
II.1. De la fonction d'onde à l'équation de Schrödinger .....	214

II.2 La densité électronique .....	215
II.3 Approche de Kohn-Sham.....	217
II.4 Traitement de l'échange corrélation et fonctions de base.....	217
II.5 Méthode pour l'étude des états excités .....	219
II.6 Théorie de la fonctionnelle de la densité dépendant du temps ( <i>Time-Dependent Density Functional Theory</i> , TDDFT).....	220
II.6.1 Théorie de la réponse linéaire .....	220
II.6.2 Propriétés spectroscopiques obtenues par la TDDFT .....	221
II.6.3 Transition électronique.....	222
II.6.4 Point de vue spectral .....	224
II.6.5 Les effets de l'environnement .....	224
II.6.6 Méthode de l'inclusion polarisable (Polarizable Embedding, PE).....	225
II.7 La dynamique moléculaire.....	226
II.7.1 Le principe .....	227
II.7.2 Le système vu comme un ensemble thermodynamique .....	228
II.7.3 Conditions périodiques aux limites .....	228
II.7.4 La dynamique moléculaire sous contraintes .....	229
III.Article .....	231
IV.Discussion des résultats et conclusions .....	257
<b>Conclusions et perspectives .....</b>	<b>263</b>
<b>Annexes .....</b>	<b>269</b>
<i>Annexes 1 : Matériels et Méthodes</i> .....	269
<i>Annexes 2 : Déplacements chimiques</i> .....	280
<b>Bibliographie .....</b>	<b>289</b>



## Liste des abréviations

1D, 2D :	Expérience RMN à une ou deux dimensions
3D :	Tridimensionnelle
$^3J_{\text{HN-H}\alpha}$ :	La constante de couplage entre HN et H $\alpha$
COSY :	CORrelation SpectroscopY
COSY-DQF	Double Quantum Filtered CORrelation SpectroscopY
CSD :	Chemical Shift Deviation
CNS :	Crystallography and NMR System
CHARMM :	Chemistry at HARvard Macromolecular Mechanics
Cryo-EM	Microscopie électronique cryogénique
DC :	Dichroïsme Circulaire
DCE :	Dichroïsme Circulaire Electronique
DFT	Density Functional Theory
DSS :	2,2-Diméthyl-2-Silapentane-5-Sulfonate
FT-IR :	Spectroscopie Infrarouge à transformée de Fourier
HF	Hartree-Fock
MeOH :	Méthanol
MM	Mécanique moléculaire
MQ	Mécanique quantique
NOE :	Nuclear Overhauser Effect
NOESY :	Nuclear Overhauser Effect SpectroscopY
PDB :	Protein Data Bank
PE :	Polarizable Embedding
PIV:	Acide pivalique
PSF:	Protein Structure File
RMN :	Résonance Magnétique Nucléaire
RCPG :	Récepteur Couplé à la Protéine G
RSA :	Relation Structure-Activité
RMSD :	Root Mean Square Deviation
TDDFT	Time-dependent Density Functional Theory
TFE :	TriFluoroEthanol
TOCSY :	TOTAL Correlation SpectroscopY
$\tau_c$ :	Temps de Corrélation
$\tau_m$ :	Temps de Mélange
UV :	Ultra-Violet

### ***Abréviations des acides aminés naturels :***

<b>Nom</b>	<b>Code à 3 lettres</b>	<b>Code à 1 lettre</b>
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Acide aspartique	Asp	D
Cystéine	Cys	C
Glutamine	Gln	Q
Acide glutamique	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Méthionine	Met	M
Phénylalanine	Phe	F
Proline	Pro	P
Sérine	Ser	S
Thréonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

### ***Abréviations des acides aminés non naturels :***

<b>Nom</b>	<b>Code à 3 lettres</b>
Acide 2-aminoisobutyrique	Aib
Cyclohexylalanine $\beta$	Cha
Ornithine	Orn

# **Chapitre 1**

## **Introduction**



## ***Chapitre 1 : Introduction***

### **Contexte général**

Les protéines et les peptides sont parmi les molécules biologiques les plus importantes du monde vivant. Ce sont en quelque sorte les abeilles ouvrières de tout organisme. Elles remplissent, entre autres, des fonctions enzymatiques, nerveuses ou motrices. Ces molécules biologiques sont constituées d'acides aminés, mais se différencient principalement par le nombre d'acides aminés les constituant. En effet, un peptide peut être constitué de moins de 50 à 60 résidus et une protéine de plus d'une cinquantaine d'acides aminés. Grâce à la longueur de chaîne polypeptidique, les protéines se replient sur elles-mêmes d'une manière plus compliquée que les peptides. Les propriétés fonctionnelles des protéines comme des peptides dépendent essentiellement de la structure tridimensionnelle (3D) qu'ils adoptent.<sup>1</sup>

La découverte du rôle physiologique d'un grand nombre de peptides a stimulé de nouveaux axes de recherche, tels que la peptidomimétique. Depuis plusieurs décennies, les peptides bioactifs apparaissent attractifs en raison du fait qu'ils possèdent certaines caractéristiques structurales ou fonctionnelles permettant la reconnaissance spécifique, la liaison ainsi que l'activation de leur cible biologique.<sup>2,3</sup> Puisque les peptides naturels peuvent rarement être utilisés comme agents thérapeutiques, en raison des problèmes associés à une faible absorption, un métabolisme rapide et une faible biodisponibilité orale, de nombreux efforts visent à modifier la séquence naturelle des acides aminés des peptides bioactifs. La conception de peptidomimétiques est basée sur cette approche. Les peptidomimétiques conservent donc les éléments clés requis pour l'activité mais remplacent les liaisons les plus labiles du peptide bioactif naturel par des liaisons plus stables pour stabiliser la molécule ou pour éviter sa dégradation.<sup>2-4</sup> Les modifications structurales ont l'avantage de fournir de nouvelles fonctionnalités qui peuvent contourner les processus naturels. Par exemple, les peptidomimétiques deviennent capables d'exercer des fonctions qui ne sont pas disponibles avec les peptides naturels, comme la liaison aux membranes cellulaires, leur pénétration ainsi que la résistance à la dégradation par les enzymes. L'une des méthodes de conception de ces peptidomimétiques, l'approche « *Structure-based drug design* », repose sur la connaissance de la structure 3D de la molécule. Cela est fondamental pour la détermination des éléments structuraux et fonctionnels nécessaires pour l'activité biologique du peptide natif. Cette approche nécessite de caractériser le plus précisément possible l'organisation spatiale de la molécule et plus particulièrement les motifs impliqués dans la reconnaissance et la liaison du

partenaire. Ces connaissances permettent de mettre en œuvre une conception rationnelle pour le développement de nouvelles molécules à usage thérapeutique.

Après les hélices  $\alpha^5$  et les feuillets  $\beta^6$ , dans les structures peptidiques, les coudes  $\beta$  constituent la troisième structure secondaire la plus rencontrée.<sup>7</sup> Leur rôle structural intervient notamment dans le repliement des protéines, dans les processus de reconnaissance moléculaire entre protéines et dans les interactions entre les substrats peptidiques et les récepteurs.<sup>8</sup> En particulier, il a été montré que les coudes jouent un rôle essentiel dans les phénomènes de reconnaissance entre les récepteurs couplés aux protéines G (RCPG) activés et leurs ligands peptidiques endogènes.<sup>9,10</sup> La caractérisation structurale des coudes est indispensable pour bien comprendre les interactions moléculaires existant entre les partenaires et faciliter la conception d'analogues pseudo-peptidiques. Par conséquent, une description conformationnelle fine, particulièrement en solution, peut donner de nouvelles perspectives sur les études de relation structure-activité (SAR).

L'introduction de ce mémoire présente un contexte général sur les peptides et les protéines et leurs différents niveaux de repliement où seront détaillés plus particulièrement les éléments de structure secondaire. Il contient également une description générale des méthodes, expérimentales et théoriques, utilisées pour l'étude de la structure des peptides et des protéines.

## I. La structure des protéines et des peptides

La première organisation structurale des protéines a été mise en évidence en 1957 par Kendrew qui proposa une représentation tridimensionnelle de la myoglobine.<sup>11</sup> Cependant, c'est en 1951 que seulement, Pauling et Corey ont proposé les premières conformations de structures en hélice.<sup>5</sup> En 1959, Linderstrom-Lang a établi la hiérarchie des structures des protéines en quatre niveaux (primaire, secondaire, tertiaire et quaternaire) que Wetlaufer a étayé en 1973 en ajoutant deux niveaux (super secondaire et domaines).<sup>12</sup>

Les polypeptides possèdent donc quatre niveaux d'organisation qui découlent les uns des autres (Figure 1) :

- **La structure primaire** est la séquence linéaire de la protéine et du peptide dont la nature des acides aminés qui le compose est en grande partie responsable de la structure finale.
- **La structure secondaire** (hélices, feuillets et coudes) que pourra adopter la protéine ou le peptide est dictée par sa longueur et la nature des résidus. L'environnement de la molécule au moment de sa structuration, joue un rôle majeur ; ainsi de nombreux

peptides possèdent des structures secondaires différentes en fonction de la nature du milieu. C'est le cas de peptides amphiphiles qui se structurent en hélice au niveau des membranes lipidiques et qui n'ont pas de structure ordonnée dans le cytoplasme ou dans le milieu extracellulaire ;

- **La structure tertiaire** est composée par l'arrangement tridimensionnel de tous les atomes de la chaîne polypeptidique ;
- **La structure quaternaire** est l'association de plusieurs structures tertiaires dont la cohésion est parfois assurée par une molécule non-peptidique ou un ion.

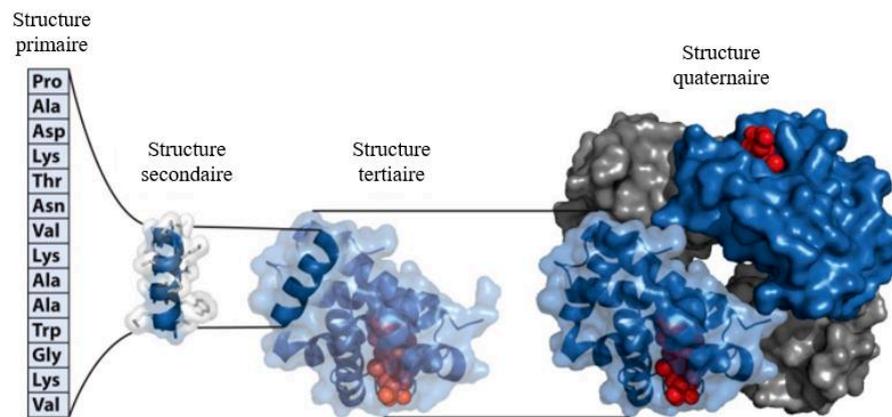


Figure 1 Vue schématique des niveaux de repliement d'une chaîne polypeptidique. L'ensemble des acides aminés constituant la structure primaire s'organise spatialement pour former les structures secondaires qui se replient ensuite pour donner la structure tertiaire. Plusieurs sous-unités peuvent s'unir pour former la conformation quaternaire.

En plus des quatre niveaux structuraux présentés, deux autres sont retrouvés dans la structure des protéines et des peptides :

- **La structure super secondaire** qui associe les divers éléments de structures secondaires (par exemple  $\beta$ -hairpins, hélices  $\alpha$  hairpins et motifs  $\beta$ - $\alpha$ - $\beta$ , Figure 2A) ;
- **Le domaine** qui est une partie d'un polypeptide capable d'adopter une structure de manière autonome ou partiellement autonome du reste de la molécule (par exemple coiled-coils, Figure 2B).

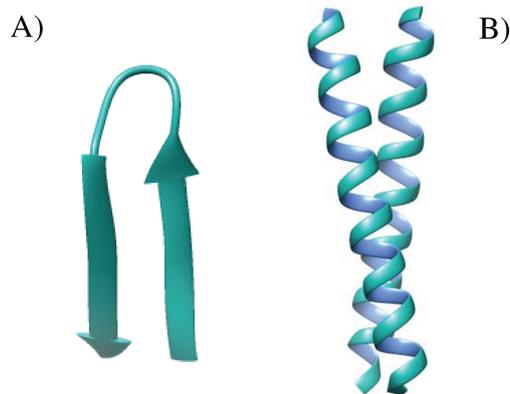


Figure 2 A). Représentation d'un exemple de structure super-secondaire de type  $\beta$  hairpin. B). Représentation d'un exemple d'un domaine protéique de type coiled coil.

### I.1 La structure primaire

Les protéines et les peptides sont des macromolécules constituées de blocs de base : les acides aminés. La grande majorité des protéines est synthétisée à l'aide de 21 acides aminés différents. Ces acides aminés contiennent à la fois un groupement amine basique et un groupement carboxyle acide (Figure ). Ils ont tous un squelette différent selon leur chaîne latérale (identifiée le plus souvent avec la lettre « R ») qui possède des caractéristiques physico-chimiques et structurales distinctes.

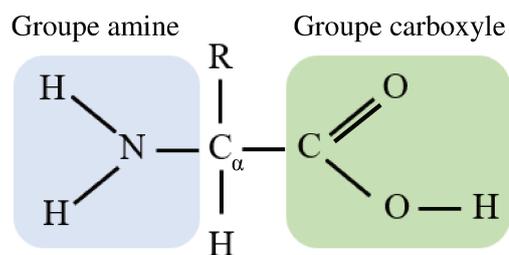


Figure 3 Structure d'un acide aminé. Le groupement amine est en bleu clair, le groupement carboxylique est en vert clair, le groupement R représente la chaîne latérale. C'est ce groupement qui diffère d'un acide aminé à un autre. Le carbone  $\alpha$  ( $C_\alpha$ ) est le centre stéréogène pour chaque acide aminé, à l'exception de la glycine.

De plus, les acides aminés possèdent une stéréoisomérisation D ou L due à la présence d'un carbone chiral, le carbone  $\alpha$ . Tous les acides aminés synthétisés par la voie ribosomiale sont des énantiomères L.<sup>13</sup>

Les acides aminés se lient entre eux pour former un peptide ou une protéine grâce à une réaction d'amidation entre les fonctions amine et carboxyle, qui aboutit à la formation d'une liaison

peptidique covalente, avec la perte d'une molécule d'eau (Figure 4). Cette liaison peptidique covalente peut présenter deux conformations différentes : *cis* et *trans*. Dans la majorité des cas, la liaison est en conformation *trans* afin d'éviter l'encombrement stérique des chaînes latérales. Cependant, la présence d'une proline dans une chaîne peptidique rend la conformation *cis* possible. Dans ce dernier cas, les deux conformations *cis* et *trans* peuvent être présentes.

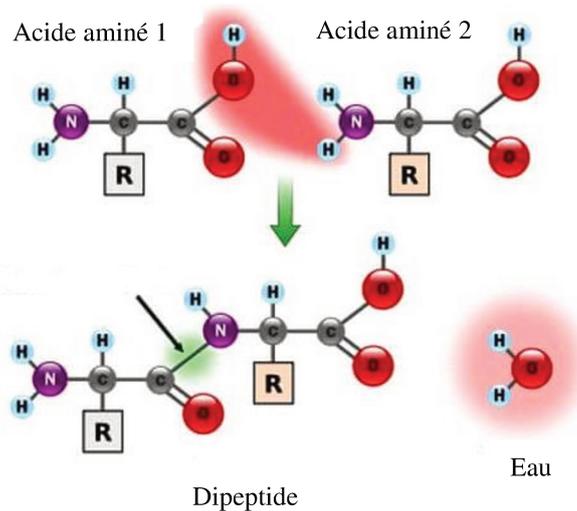


Figure 4 Formation d'une liaison peptidique entre deux acides aminés.

## I.2 La structure secondaire

La structure secondaire est définie par le repliement local de la chaîne principale d'un peptide (Figure 5A). Ce repliement est stabilisé par la présence de liaisons hydrogène entre les groupements amine (NH) et carbonyle (CO) du squelette peptidique. Il existe trois principales catégories de structure secondaire : les hélices  $\alpha$ , les feuillets  $\beta$  et les coudes. Pour définir les différents types de structures secondaires, la conformation des acides aminés a été étudiée en fonction de la combinaison de valeurs des angles dièdres  $\Phi$  et  $\Psi$  et il a été déduit que toutes les valeurs d'angles dièdres n'étaient pas possibles pour les acides aminés.<sup>14</sup> Ces angles sont mesurés entre deux plans autour d'une liaison (Figure 5B). L'angle  $\Phi$  est l'angle autour de la liaison  $C\alpha - N$ , et l'angle  $\Psi$  celui autour de la liaison  $C\alpha - C=O$  (Figure 5B).

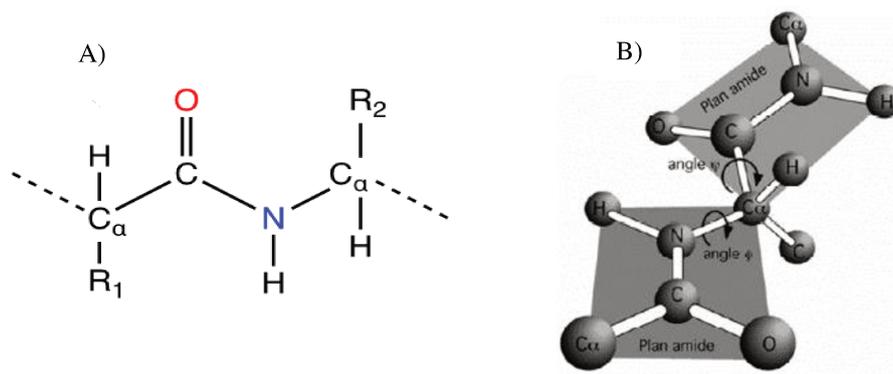


Figure 5 (A) Représentation de la chaîne principale. (B) Représentation des deux angles dièdres  $\phi$  et  $\psi$  avec le plan de liaisons. Les symboles  $R_1$  et  $R_2$  représentent la chaîne latérale des acides aminés.

Dans la suite de ce chapitre, nous présenterons les différentes structures secondaires qu'adoptent les peptides et les protéines et nous nous attarderons plus spécialement sur les structures en conformation coude.

### I.2.1 Les hélices

Les hélices constituent le type de structure secondaire le plus abondant dans les protéines et dans les peptides. Les hélices sont construites à partir d'un enchaînement de conformations élémentaires pour lesquels les valeurs respectives des angles dièdres  $\Phi$  et  $\Psi$  sont répétées. Pour décrire les hélices le long de l'axe de propagation, deux paramètres sont pris en compte, «  $n$  » le nombre de résidus par tour d'hélice et le pas «  $p$  » de résidus par tour.

Dans les protéines et les peptides, il existe plusieurs types d'hélices, dont la plus abondante est l'hélice  $\alpha$ . Nous rencontrons aussi de courts segments d'hélice  $3_{10}$ . En revanche, les hélices  $\pi$  sont très rares (5%). Les paramètres géométriques des différentes hélices sont donnés dans le Tableau 1.

Tableau 1 Paramètres caractéristiques des hélices dans les protéines et les peptides.  $n$  : nombre de résidus par tour et  $p$  : pas de résidus par tour.

Structure	$\Phi$	$\Psi$	$n$	$p$ (Å)	Liaisons hydrogène (CO...HN)
Hélice $\alpha$	$-57^\circ$	$-47^\circ$	3,6	5,4	$i, i+4$
Hélice $3_{10}$	$-49^\circ$	$-26^\circ$	3,0	6,0	$i, i+3$
Hélice $\pi$	$-57^\circ$	$-70^\circ$	4,4	5,0	$i, i+5$

## Les hélices $\alpha$

L'hélice  $\alpha$  fut décrite par Pauling en 1951 qui prédit que cette structure devait être stable et énergétiquement favorable dans les protéines.<sup>5</sup>

Elle possède 3,6 résidus par tour d'hélice avec une translation de 5,4 Å. Elle est stabilisée par des liaisons hydrogène entre le carbonyle en position  $i$  et le proton amide en position  $i+4$  (Figure 6). Tous les CO et NH sont impliqués dans des liaisons hydrogène à l'exception des trois premiers NH et des trois derniers CO de l'hélice. Comme toutes les liaisons hydrogène sont parallèles et orientées dans le même sens, l'hélice  $\alpha$  est très polarisée (Figure 6).<sup>13</sup> Son moment dipolaire correspond à une charge de 0,5-0,7 Debye à chaque extrémité. Le compactage des atomes est favorable aux interactions de Van Der Waals. Les chaînes latérales sont dirigées vers l'extérieur. Dans les protéines et les peptides, les valeurs d'angles dièdres idéaux théoriques, données dans le Tableau 1, sont  $-57^\circ$  pour l'angle dièdre  $\Phi$  et  $-47^\circ$  pour l'angle dièdre  $\Psi$ . Les possibles variations de ces valeurs sont fonction de la séquence de la chaîne peptidique, de la courbure de l'hélice et de la solvataion.

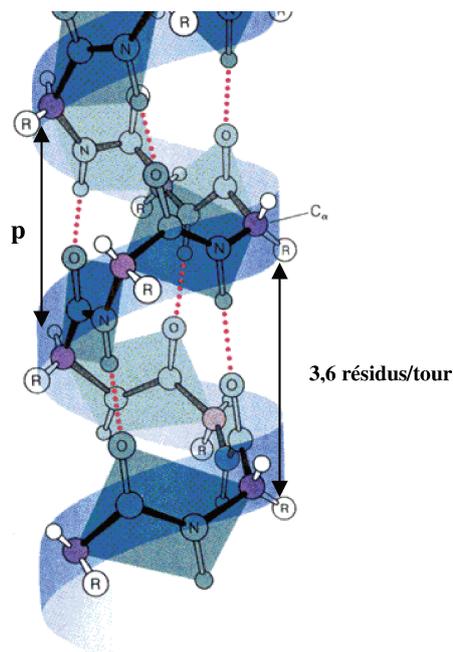


Figure 6 Représentation d'une hélice  $\alpha$  droite avec son élévation pour un tour complet ( $p$ ) et les liaisons hydrogène caractéristiques en pointillés (tiré de Lehninger et al.).<sup>13</sup>

## Les hélices $3_{10}$

L'hélice  $3_{10}$  n'est pas un élément structural fréquent dans les protéines. Au contraire, dans les peptides, la forme  $3_{10}$  est prédominante sur la forme  $\alpha$ .<sup>15,16</sup> Les hélices  $3_{10}$  sont souvent

rencontrées dans les protéines aux extrémités N- et C-terminales des hélices  $\alpha$  et comportent entre 4 et 6 résidus.<sup>17,18</sup> Elle fut proposée par Huggins en 1943.<sup>19</sup> Selon la base de données de Kabsch et Sander, seulement 3,4% des résidus sont dans cette conformation.<sup>20</sup>

Elle possède 3,0 résidus par tour d'hélice avec une translation de 6,0 Å. Elle est stabilisée par des liaisons hydrogène entre le carbonyle en position  $i$  et le proton amide en position  $i+3$ . Elle est donc plus longue que l'hélice  $\alpha$ , pour le même nombre de résidus. Les liaisons hydrogènes ont un décalage de 30° par rapport à l'axe de l'hélice (Figure 7).

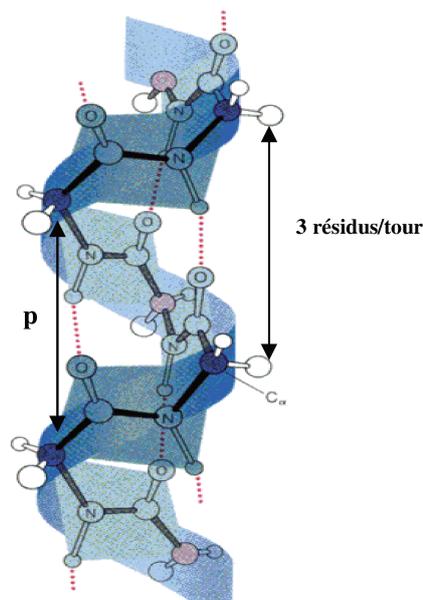


Figure 7 Représentation d'une hélice 3<sub>10</sub> avec son élévation pour un tour complet (p) et les liaisons hydrogène caractéristiques en pointillés (tiré de Lehninger et al.).<sup>13</sup>

### Les hélices $\pi$

L'hélice  $\pi$  ou hélice 4.4<sub>16</sub> est une structure secondaire extrêmement rare dans les protéines et les peptides. Elle a été prédite par Donohue en 1953 et découverte dans une catalase et dans quelques enzymes très récemment.<sup>21,22</sup> Elle possède 4,4 résidus par tour d'hélice avec une translation de 5,0 Å. Elle est stabilisée par des liaisons hydrogène entre le carbonyle en position  $i$  et le proton amide en position  $i+5$ . Elle est donc plus courte que l'hélice  $\alpha$ , pour le même nombre de résidus (Figure 8).

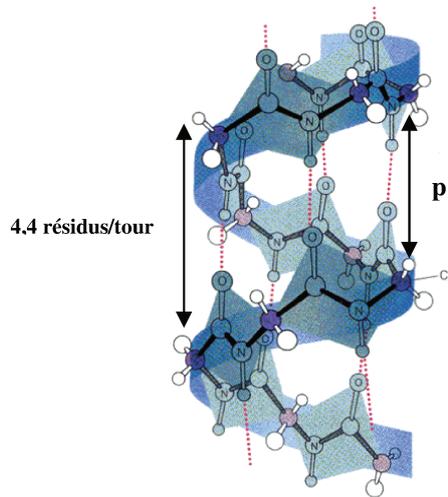


Figure 8 Représentation d'une hélice  $\pi$  avec son élévation pour un tour complet ( $p$ ) et les liaisons hydrogène caractéristiques en pointillés (tiré de Lehninger et al.).<sup>13</sup>

L'hélice  $\pi$  est très défavorable pour trois raisons : (i) la valeur des angles dièdres est énergétiquement défavorable par rapport à l'hélice  $\alpha$ , (ii) le trou central de 1 Å de diamètre est trop petit pour accepter une molécule d'eau et engendre une perte des interactions de Van Der Waals, et (iii) quatre résidus doivent être correctement alignés pour permettre la création de la liaison hydrogène  $i, i+5$ .<sup>23-25</sup>

De manière générale, la stabilité des hélices dépend de plusieurs paramètres :

- La nature des acides aminés est grandement impliquée ; ainsi Ala, Arg, Lys, Glu, Leu et Met sont très favorables à la formation d'hélice. En revanche, Gly est très défavorable (souvent rencontré en fin d'hélice) et Pro est un briseur d'hélice mais qui est souvent présent avant le premier tour de l'hélice du côté C-terminal.<sup>17</sup>
- La longueur de l'hélice  $\alpha$  est importante. La plupart des hélices présentes dans les protéines contiennent entre 10 et 15 résidus ; en dessous de cette longueur, c'est la forme  $3_{10}$  qui prédomine. Dans les solvants protiques comme le méthanol ou l'eau, 15 résidus sont nécessaires pour obtenir une structure en hélice  $\alpha$ .<sup>15</sup>
- La répartition des résidus dans l'hélice est primordiale. La nature hydrophile et hydrophobe des chaînes latérales est responsable de la structuration de l'hélice. Ainsi de nombreuses hélices ont un caractère amphiphile avec une face hydrophobe (dirigée vers l'intérieur) et une face hydrophile (dirigée vers l'extérieur).<sup>26</sup>
- L'environnement de l'hélice est souvent responsable de sa formation. Ainsi l'augmentation de la concentration de la molécule (*in-vitro*) favorise la forme structurée

au dépend de la pelote statistique. Les autres motifs peptidiques adjacents jouent le même rôle que la concentration et les interactions hydrophobes jouent probablement un rôle important dans la formation des structures secondaires.<sup>27</sup>

### I.2.2 Les feuillets

Les feuillets  $\beta$  sont les deuxièmes structures secondaires majeures rencontrées dans les protéines et dans les peptides.<sup>20</sup> L'unité de base du feuillet  $\beta$  est le brin qui possède des valeurs d'angles dièdres  $\Phi$  et  $\Psi$  de signes opposés. Le feuillet peut être considéré comme une hélice aplatie composée de deux résidus par tour. Les caractéristiques géométriques sont données dans le Tableau 2.

Les feuillets  $\beta$  sont présents sous deux formes, parallèle et antiparallèle, qui sont liées à la direction relative des deux brins. La longueur moyenne des feuillets est constituée de six résidus par brin et ils sont généralement formés de moins de six brins. Les chaînes latérales de deux résidus adjacents sont de part et d'autre du feuillet et comme les hélices  $\alpha$ , ils peuvent avoir un potentiel amphiphile (une face hydrophile et une face hydrophobe). La cohésion entre les brins est possible grâce à de nombreuses liaisons hydrogène. Les feuillets  $\beta$  sont formés à partir de la combinaison de différentes régions de la chaîne polypeptidique qui sont souvent éloignées les unes des autres.

Tableau 2 Paramètres caractéristiques des feuillets  $\beta$  parallèle et antiparallèle dans les protéines et les peptides.

Structure	$\Phi$	$\Psi$	Nombre de résidus par tour
Feuillet $\beta$ parallèle	-119°	+113°	2
Feuillet $\beta$ antiparallèle	-139°	+135°	2

Au niveau structural, il existe des différences entre les feuillets  $\beta$  parallèles et les feuillets  $\beta$  antiparallèles.

## Les feuillets antiparallèles

Les feuillets antiparallèles sont les plus fréquents car ils sont induits par la présence de boucles ou coudes à leurs extrémités. Les feuillets antiparallèles sont plus stables que les parallèles, du fait d'une orientation des liaisons hydrogène entre les chaînes plus favorables. Les liaisons hydrogène sont presque perpendiculaires à l'axe des chaînes étirées (Figure 9). L'oxygène carbonyle et l'hydrogène amide d'un résidu sont engagés dans une liaison hydrogène avec l'oxygène carbonyle et l'hydrogène amide d'un autre résidu de l'autre brin (Figure 9).

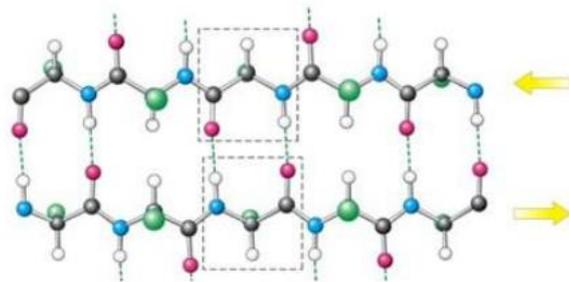


Figure 9 Représentation d'un feuillet  $\beta$  antiparallèle. Les liaisons hydrogène caractéristiques dans les feuillets  $\beta$  antiparallèles sont représentées en pointillés (tiré de Lehninger et al.).<sup>13</sup>

## Irrégularité $\beta$ (bulge $\beta$ )

L'irrégularité  $\beta$  est un petit morceau de structure non répétitive qui peut apparaître dans les régions aléatoires, mais se présente le plus souvent et se visualise plus facilement sous la forme d'une irrégularité dans la structure  $\beta$  antiparallèle.<sup>28</sup>

Un bulge  $\beta$  est défini comme une région située entre deux brins  $\beta$ , qui comprend deux résidus sur un brin et un seul résidu sur l'autre brin.<sup>28</sup> Ils affectent la direction de la chaîne polypeptidique.<sup>29</sup>

## Les feuillets parallèles

Les feuillets parallèles n'ont pas de liaison hydrogène perpendiculaire à l'axe du brin, cela a pour effet la formation d'une plus faible interaction dipolaire. Chaque résidu forme des liaisons hydrogène entre l'oxygène du carbonyle et l'hydrogène amide de deux différents résidus de deux brins adjacents (Figure 10).

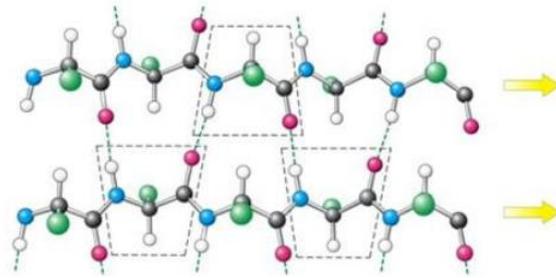


Figure 10 Représentation d'un feuillet  $\beta$  parallèle. Les liaisons hydrogène caractéristiques dans les feuillets  $\beta$  parallèles sont représentées en pointillés (tiré de Lehninger et al.).<sup>13</sup>

La troisième catégorie de structures secondaires rencontrée dans les protéines et les peptides est le coude. Ce motif permet d'inverser la direction de la chaîne polypeptidique. Il comporte différentes catégories et sous-types, dont le nombre a évolué au cours du temps. Etant le sujet central de ce travail de thèse, nous avons décidé de le traiter en détail dans la partie II de ce chapitre. Nous poursuivrons donc ce chapitre en expliquant la structure tertiaire et la structure quaternaire des protéines et des peptides.

### I.3 La structure tertiaire

L'arrangement tridimensionnel d'une protéine ou d'un peptide constitue sa structure tertiaire, et consiste au repliement et à l'agencement des structures secondaires afin d'obtenir la conformation de plus basse énergie. Cela va permettre à des acides aminés qui étaient séquentiellement éloignés de se rapprocher et de leur donner la possibilité d'effectuer des interactions diverses. Parmi ces interactions, on trouve (i) les liaisons hydrogène qui se forment entre un groupement donneur (oxygène le plus souvent) et un groupement accepteur (azote le plus souvent). Cette interaction peut se produire entre les atomes du squelette ou des chaînes latérales, (ii) les interactions électrostatiques, comme les liaisons ioniques, par exemple entre la chaîne latérale de la lysine et l'acide aspartique et (iii) les liaisons covalentes, comme la formation des ponts disulfures ou des ponts salins.

### I.4 La structure quaternaire

De nombreuses protéines sont constituées de multiples chaînes polypeptidiques, souvent appelées sous-unités protéiques. Ces sous-unités peuvent être identiques (comme dans un homodimère) ou différentes (comme dans un hétérodimère). La structure quaternaire correspond donc à l'association de ces différentes sous-unités protéiques. La forme finale du

complexe de protéines est à nouveau stabilisée par diverses interactions, y compris les liaisons hydrogène, les ponts disulfures et les interactions électrostatiques.

## II. Une structure particulière : les coudes

Les coudes sont le troisième type de structure secondaire rencontrée dans les structures des protéines et des peptides. Ils permettent de relier 2 structures secondaires de type hélices et/ou brins. La localisation des coudes à la surface des molécules a conduit à l'hypothèse que ces derniers jouent un rôle critique dans la reconnaissance dans les systèmes biologiques. En outre, plusieurs études ont suggéré qu'avec les hélices  $\alpha$ , les coudes sont les motifs de reconnaissance privilégiés des peptides bioactifs par leurs cibles biologiques.<sup>8,30</sup> Les coudes constituent un repliement particulier du squelette carboné. En effet, ils sont des motifs structuraux ordonnés, mais non répétitifs. Alors que les hélices et les brins  $\beta$  ont la propriété de répéter approximativement les mêmes angles dièdres  $\Phi$  et  $\Psi$  pour des résidus successifs, les éléments de structure non répétitive présentent une succession particulière de valeurs d'angles dièdres  $\Phi$  et  $\Psi$  différentes pour chaque résidu, de sorte que la notion de position du résidu dans la structure est plus influente que dans une structure répétitive. Plusieurs structurations de type coude sont possibles selon le nombre de résidus impliqués dans la structuration. Ainsi, cinq catégories de coudes ont été décrites dans la littérature : les coudes  $\delta$ , les coudes  $\gamma$ , les coudes  $\beta$ , les coudes  $\alpha$  et les coudes  $\pi$  qui sont formés par deux, trois, quatre, cinq et six résidus respectivement.

L'ensemble de ces coudes est caractérisé par une distance inférieure à 7 Å entre les carbones  $\alpha$  ( $C_\alpha$ ) du premier et celui du dernier résidu, par la formation d'une liaison hydrogène entre le groupement carbonyle du résidu  $i$  et le groupement amide du résidu  $i+n$  ( $n-1$  étant le nombre de résidus dans le coude) et, enfin, le dernier critère est basé sur le fait que la structure ne soit pas hélicoïdale ou pas trop étendu.

Chez les peptides, les catégories de coudes les plus rencontrées sont  $\gamma$  et  $\beta$ .<sup>31</sup> De ce fait, dans le reste de ce chapitre, seuls les coudes de nature  $\gamma$  et  $\beta$  seront détaillés à cause de leur importance structurale dans la reconnaissance et l'activité des peptides bioactifs vers leurs cibles biologiques (par exemple les récepteurs), en particulier les récepteurs couplés à une protéine G (RCPG).<sup>8,30,32</sup>

### II.1 Les coudes $\gamma$

Les coudes de nature  $\gamma$  sont des motifs dipeptidiques. Ils ont été décrits théoriquement pour la première fois par Némethy et Printz<sup>33</sup> et observés expérimentalement dans la structure

atomique de la thermolysine par Matthews.<sup>34</sup> Ces structures possèdent une liaison hydrogène très étroite, sur un cycle à sept atomes, situé entre le groupe carbonyle du premier résidu et le groupe amine du troisième (Figure ). Ils se retrouvent souvent dans les boucles et à la fin des motifs structuraux de type «  $\beta$ -hairpins ».<sup>35</sup>

Deux types de coude  $\gamma$ , classique et inverse, ont été caractérisés. Ils se différencient par les valeurs d'angles dièdres  $\Phi$  et  $\Psi$  que possède le résidu  $i+1$  (Tableau 3 et Figure 11).

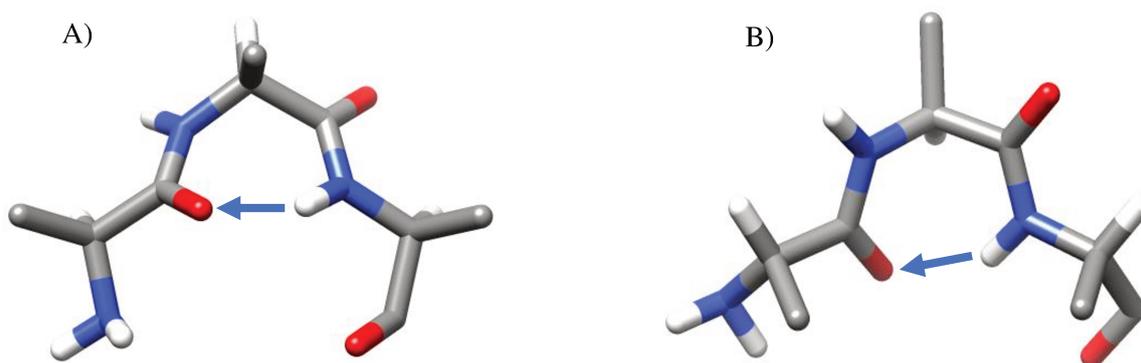


Figure 11 Représentation des deux types de coudes  $\gamma$ , (A) Coude  $\gamma$  de type classique, (B) Coude  $\gamma$  de type inverse. Les atomes sont représentés en rouge pour l'oxygène, bleu pour l'azote, gris pour le carbone. LA liaison hydrogène est représentée par une flèche en bleu clair.

Tableau 3 Valeurs des angles dièdres du résidu central impliqué dans les deux types de coudes  $\gamma$ .<sup>35,36</sup>

Structure	$\Phi_{i+1}$	$\Psi_{i+1}$
<b>Coude <math>\gamma</math> classique</b>	+75°	-64°
<b>Coude <math>\gamma</math> inverse</b>	-79°	+69°

Dans les coudes de nature  $\gamma$  classique, la glycine est souvent présente en position  $i+1$  (Figure 12).<sup>37</sup> Cet acide aminé est aussi le plus retrouvé en position  $i$  alors que l'histidine, la lysine et la sérine préfèrent la position  $i+2$  car ils peuvent former une liaison hydrogène supplémentaire par leur chaîne latérale avec le squelette (Figure 12).<sup>37</sup>

Dans les coudes de nature  $\gamma$  de type inverse, la proline a comme position préférentielle la position  $i+1$ , car la valeur de l'angle  $\Phi_{i+1}$  est favorable pour ce résidu cyclique.<sup>35,37</sup> L'autre acide aminé fréquemment retrouvé est l'asparagine (Figure 12). En position  $i+2$ , près d'un quart de la position est occupée par l'asparagine qui fait une liaison hydrogène supplémentaire par sa chaîne latérale avec l'oxygène du groupe carbonyle du résidu  $i$ .<sup>37</sup>

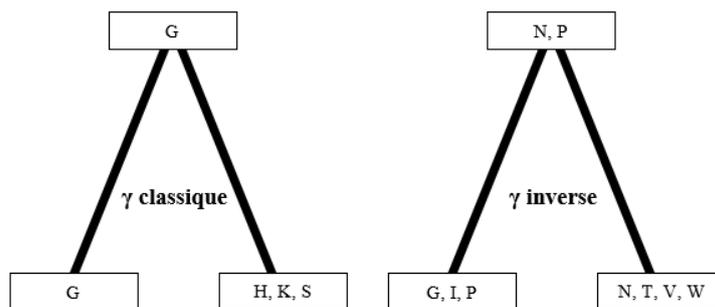


Figure 12 Représentation des préférences d'acides aminés statistiquement significatives à chaque position des coudes  $\gamma$  de type classique et inverse (tiré de Guruprasad et al, 2000).<sup>37</sup>

## II.2 Les coudes $\beta$

Les coudes de nature  $\beta$  sont les coudes les plus répandus. Ce sont des motifs composés de 4 résidus. Ils ont été proposés pour la première fois à partir d'une analyse de conformation théorique réalisée par Venkatachalam en 1968.<sup>38</sup> Il examina quelles conformations étaient disponibles pour un système de trois unités peptidiques liées (ou quatre résidus successifs) qui pourraient être stabilisées par une liaison hydrogène centrale entre l'oxygène du groupe carboxylique (CO) du résidu  $i$  et l'hydrogène du groupe amide (NH) du résidu  $i+3$ . Les coudes  $\beta$  sont classés grâce aux valeurs d'angles dièdres  $\Phi$  et  $\Psi$  des deux résidus centraux, notés  $i+1$  et  $i+2$ . Leur classification a évolué jusqu'à l'heure actuelle grâce à l'amélioration des méthodes de résolution de structures protéiques qui ont permis l'augmentation du nombre de structures disponibles. Dans la suite de cette partie, nous allons présenter l'évolution et la réorganisation de cette classification.

### II.2.1 Evolution de la classification des coudes $\beta$

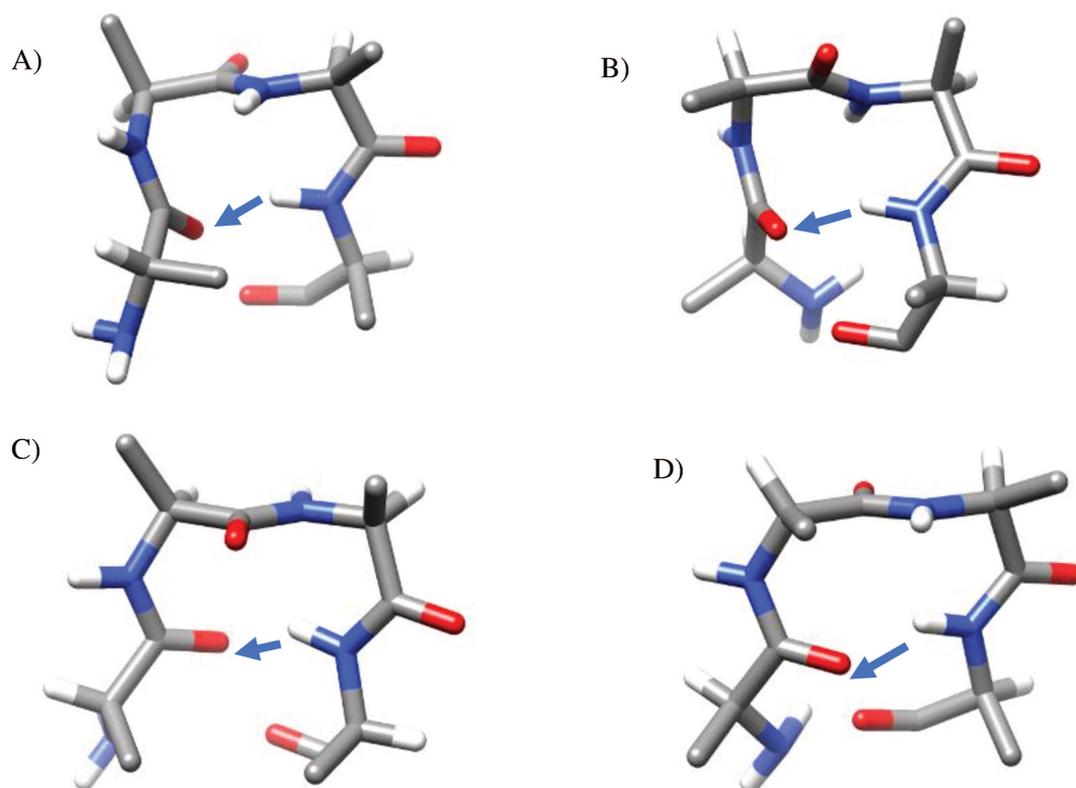
Venkatachalam et ses collaborateurs ont tout d'abord identifié six types généraux de coudes de nature  $\beta$ , qu'ils ont appelés type I, II et III avec leurs images miroir nommées I', II' et III' (Figure 13), en analysant les conformations que pouvait adopter un système modèle tripeptidique stabilisé par une liaison hydrogène entre le premier et le dernier résidu.<sup>38</sup> Ces types se distinguent les uns des autres par des valeurs d'angles dièdres  $\Phi$  et  $\Psi$  caractéristiques des résidus  $i+1$  et  $i+2$  qui sont résumées dans le Tableau 4.

Les types I et type II sont non hélicoïdaux et replient la chaîne sur elle-même autour d'un coin plutôt carré (Tableau 4 et Figure ), de sorte que les premier et quatrième carbones  $\alpha$  soient séparés d'une distance d'environ 5 Å. Au contraire des types I et II, les types III et III' ont une organisation hélicoïdale pour les deux résidus centraux (Tableau 4). Ces deux types sont liés

l'un à l'autre par un retournement de  $180^\circ$  de l'unité peptidique centrale. Les types I et III sont identiques pour le résidu  $i+1$  et ne diffèrent que de  $-30^\circ$  pour le résidu  $i+2$  (Tableau 4). Les coudes  $\beta$  de type I', II' et III' ont des valeurs d'angles dièdres  $\Phi$  et  $\Psi$  de signe opposé à leurs images miroirs (I, II et III) (Tableau 4)

Tableau 4 Valeurs des angles dièdres  $\Phi$  et  $\Psi$  pour les résidus  $(i+1)$  et  $(i+2)$  impliqués dans les coudes de nature  $\beta$  identifiés par Venkatachalam en 1968.<sup>38</sup>

Structure	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
Coude $\beta$ I	$-60^\circ$	$-30^\circ$	$-90^\circ$	$0^\circ$
Coude $\beta$ I'	$+60^\circ$	$+30^\circ$	$+90^\circ$	$0^\circ$
Coude $\beta$ II	$-60^\circ$	$+120^\circ$	$-80^\circ$	$0^\circ$
Coude $\beta$ II'	$+60^\circ$	$-120^\circ$	$+80^\circ$	$0^\circ$
Coude $\beta$ III	$-60^\circ$	$-30^\circ$	$-60^\circ$	$-30^\circ$
Coude $\beta$ III'	$+60^\circ$	$+30^\circ$	$+60^\circ$	$+30^\circ$



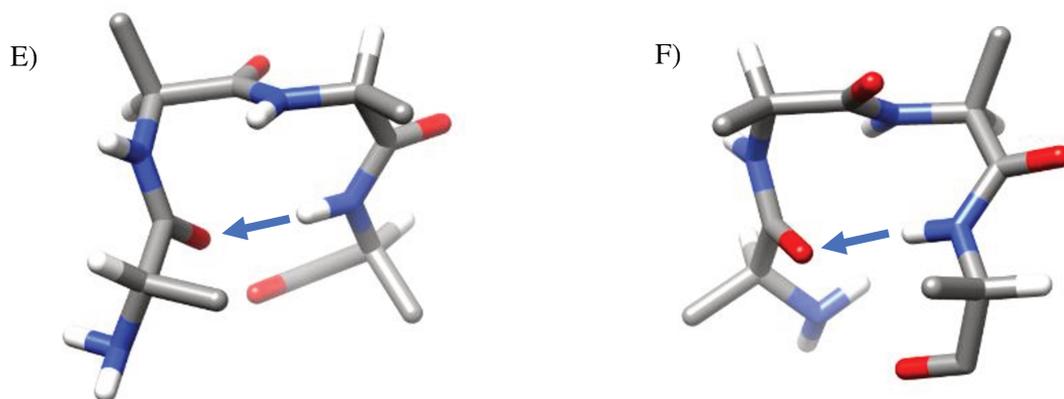


Figure 13 Représentation schématique des six types de coudes  $\beta$  identifiés par Venkatachalam en 1968.<sup>38</sup> (A) coude  $\beta$  de type I, (B) coude  $\beta$  de type I', (C) coude  $\beta$  de type II, (D) coude  $\beta$  de type II', (E) coude  $\beta$  de type III, (F) coude  $\beta$  de type III'. Les atomes sont représentés en bleu pour l'azote, en rouge pour l'oxygène, en gris pour le carbone et en blanc pour l'hydrogène. La liaison hydrogène est représentée par une flèche en bleu clair.

Suite au travail de Venkatachalam, Lewis et ses collaborateurs, au cours de l'année 1973, ont réalisé une étude sur les structures cristallographiques de huit protéines dans lesquelles 129 motifs structuraux de type coudes  $\beta$  ont été identifiés.<sup>39</sup> Ceux-ci ont été groupés en dix types différents (I, I', II, II', III, III', IV, V, VI et VII). Afin de déterminer leur stabilité conformationnelle, des calculs de minimisation d'énergie des peptides Ace-Ala-Ala-Ala-Ala-NHMe, Ace-Ser-Ala-Ser-Asp-NHMe, Ace-Asp-Ser-Gly-Gly-NHMe et Ace-Asp-Lys-Thr-Gly-NHMe dans les différentes conformations de coudes  $\beta$  identifiés, ont été réalisés. Parmi les dix types retrouvés, les six premiers (I, I', II, II', III et III') sont ceux décrits précédemment par Venkatachalam. Dans les nouveaux types, le type V a été déterminé à partir d'un coude de type II inhabituel. Les valeurs canoniques d'angles dièdres  $\Phi$  et  $\Psi$  pour les deux résidus centraux du type V sont présentés dans le Tableau 5.

Lewis et ses collaborateurs ont également proposé trois autres types de coudes  $\beta$  : les types VI, VII et IV. Le coude de type VI doit son existence à la présence d'une proline en conformation *cis* en position  $i+2$ . En effet, la construction du modèle indique qu'un résidu en conformation *cis* dans cette position spécifique est géométriquement adapté pour inverser la direction de la chaîne polypeptidique. Cependant, aucune valeur d'angle dièdre  $\Phi$  et/ou  $\Psi$  caractéristique n'a pu être identifiée pour ce type de coude.

Les coudes de type VII répondent uniquement au critère selon lequel la distance entre le  $C_\alpha$  du premier résidu et le  $C_\alpha$  du dernier résidu doit être inférieure à 7 Å. Ce type de coude  $\beta$  présente soit une valeur inhabituellement grande de  $\Psi_{i+1}$  couplée à une petite valeur de  $\Phi_{i+2}$ , soit une petite valeur de  $\Psi_{i+1}$  couplée à une valeur exceptionnellement grande de  $\Phi_{i+2}$ . Quand ces conditions sont réunies, ces coudes apparaissent comme des « nœuds » dans la structure de la

protéine plutôt que de véritables inversions de chaîne.<sup>39</sup> Les conditions qui caractérisent et permettent la formation du coude de type VII sont présentées dans le Tableau 5.

Enfin, nous retrouvons le type IV qui se démarque des autres types de coudes. En effet, ce type regroupe les structures de coude ayant deux ou plusieurs angles dièdres avec des valeurs qui diffèrent de  $\pm 40^\circ$  par rapport aux valeurs canoniques observées dans les autres types de coudes (Tableau 5).

Tableau 5 Valeurs des angles dièdres  $\Phi$  et  $\Psi$  pour les résidus (i+1) et (i+2) impliqués dans les coudes de nature  $\beta$  des types identifiés par Lewis et al en 1973.<sup>39</sup> En gras, les modifications apportées à la classification par Lewis et ses collaborateurs.

Structure	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
Coude $\beta$ I	-60°	-30°	-90°	0°
Coude $\beta$ I'	+60°	+30°	+90°	0°
Coude $\beta$ II	-60°	+120°	-80°	0°
Coude $\beta$ II'	+60°	-120°	+80°	0°
Coude $\beta$ III	-60°	-30°	-60°	-30°
Coude $\beta$ III'	+60°	+30°	+60°	+30°
<b>Coude <math>\beta</math> IV</b>	<b>Tous les coudes <math>\beta</math> qui possèdent deux ou plus angles dièdres (<math>\Phi_{i+1}</math>, <math>\Psi_{i+1}</math>, <math>\Phi_{i+2}</math> et/ou <math>\Psi_{i+2}</math>) différant au moins de <math>40^\circ</math> par rapport aux valeurs canoniques des autres types de coudes <math>\beta</math></b>			
<b>Coude <math>\beta</math> V</b>	<b>-80°</b>	<b>+80°</b>	<b>+80°</b>	<b>-80°</b>
<b>Coude <math>\beta</math> VI</b>	<b>Coude <math>\beta</math> caractérisé par la présence d'une proline en conformation cis en position i+2</b>			
<b>Coude <math>\beta</math> VII</b>		<b><math>\approx 180^\circ</math> ou <math>&lt;60^\circ</math></b>	<b><math>\approx 180^\circ</math> ou <math>&lt;60^\circ</math></b>	

A partir de cette classification, plusieurs modifications ont été ultérieurement apportées par le groupe de Richardson.<sup>40</sup> Certaines catégories ont ainsi été purement supprimées et d'autres ont été réintégrées à des catégories plus caractéristiques.

Tout d'abord, il a été suggéré que le type III est un cas particulier de coude  $\beta$  car il comporte en fait des valeurs d'angles  $\Phi$  et  $\Psi$  répétées de  $-60^\circ$  et  $-30^\circ$  pour les deux résidus centraux, ce qui en fait une structure identique à l'hélice  $3_{10}$  (Tableau 4). De même, son image miroir, le type III', possède des valeurs caractéristiques des angles dièdres qui sont proches du coude  $\beta$  de type I'. A cause de ces similarités structurales, ces types ont été écartés de la classification.<sup>40</sup> Ensuite, une analyse spécifique du type  $\beta$ -V a été réalisée sur 29 structures cristallographiques de protéines par Richardson et ses collaborateurs.<sup>40</sup> Les résultats ont montré une fréquence de 0,03%, pour ce type dans les structures analysées. Cette faible représentation ne justifiant pas

de garder ce type en tant que catégorie indépendante, cela a donc conduit à son élimination de la classification.<sup>40</sup>

En outre, comme les coudes de type VII ne sont définis que par deux angles dièdres et peuvent adopter deux valeurs différentes pour ceux-ci, leurs structures varient énormément. L'analyse de structures cristallographiques de protéines réalisée par le groupe de Richardson a permis de repérer neuf structures de coude  $\beta$  de type VII.<sup>40</sup> Parmi ces neuf exemples, au moins la moitié est douteuse. Par exemple, le coude retrouvé dans la nucléase staphylococcique (résidus 47-50) au bout d'une boucle partiellement désordonnée et celui retrouvé dans la rubredoxine (46-49) ne sont pas bien définis car les coordonnées des atomes des deux résidus centraux ne permettent pas de définir avec certitude les angles dièdres  $\Phi$  et  $\Psi$ . Par conséquent, le type VII a semblé injustifiable en tant que catégorie distincte et il a alors été inclus dans les autres catégories.<sup>40</sup>

Enfin, le type VI a été mieux caractérisé. En effet, une analyse sur les motifs qui présentaient la proline en conformation *cis* a été effectuée et les types VIa et le type VIb ont été différenciés (Figure 14).<sup>40</sup> Ils peuvent être distingués par le fait que, dans le type VIa, la proline a une orientation « concave », des valeurs d'angles dièdres de type hélicoïdal et la liaison hydrogène est typiquement présente. Au contraire, dans le type VIb, la proline a une orientation « convexe », des angles dièdres en conformation étendue et la liaison hydrogène n'est pas un requis obligatoire.<sup>40</sup>

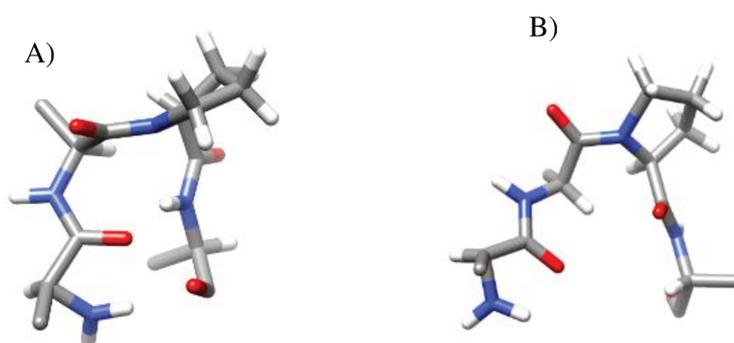


Figure 14 Représentation schématique des coudes  $\beta$  de type VIa (A) et VIb (B) identifiés par Richardson et al. en 1981.<sup>40</sup> La figure A montre l'orientation « concave » du coude de type VIa ; la figure B l'orientation « convexe » du coude de type VIb.

Dans la classification de Richardson, le type IV est essentiellement une catégorie hétérogène, qui inclut n'importe quel exemple avec deux des angles dièdres distants de plus de  $40^\circ$  des valeurs idéales pour l'un des autres types de coudes  $\beta$ .

A l'issue des travaux de Richardson et ses collaborateurs, la classification des coudes  $\beta$  ne comporte donc plus que 7 types (I, I', II, II', VIa, VIb et le type IV) (Tableau 6).

Tableau 6 Valeurs des angles dièdres pour les résidus ( $i+1$ ) et ( $i+2$ ) impliqués dans les coudes de nature  $\beta$  des types identifiés par Richardson et al. en 1981.<sup>40</sup> En gras, les modifications apportées à la classification par Richardson et ses collaborateurs.

Structure	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
Coude $\beta$ I	-60°	-30°	-90°	0°
Coude $\beta$ I'	+60°	+30°	+90°	0°
Coude $\beta$ II	-60°	+120°	-80°	0°
Coude $\beta$ II'	+60°	-120°	+80°	0°
Coude $\beta$ IV	Tous les coudes $\beta$ qui possèdent deux ou plus angles dièdres ( $\Phi_{i+1}$ , $\Psi_{i+1}$ , $\Phi_{i+2}$ et/ou $\Psi_{i+2}$ ) différant au moins de 40° par rapport aux valeurs canoniques des autres types de coudes $\beta$			
<b>Coude <math>\beta</math> VIa</b>	<b>cis-proline avec orientation « concave » en position <math>i+2</math></b>			
<b>Coude <math>\beta</math> VIb</b>	<b>cis-proline avec orientation « convexe » en position <math>i+2</math></b>			

La classification changea encore après une analyse statistique de la fréquence de tous les types de coudes  $\beta$  identifiés dans les structures protéiques menée par Wilmot et Thornton en 1988.<sup>41</sup> En effet, en plus d'avoir confirmé les résultats des travaux réalisés auparavant, ils ont observé qu'un autre motif formé de quatre acides aminés avec des valeurs d'angles dièdres  $\Phi$  et  $\Psi$  non conventionnelles. Ce motif est rencontré suffisamment souvent pour justifier l'introduction d'un autre type de coude spécifique, le type VIII, décrit aussi comme conformation  $\alpha_R \rightarrow \beta$  (Figure 15).<sup>41</sup> Cette définition est due au fait que les angles dièdres du résidu  $i+1$  sont ceux d'une conformation hélicoïdale, et les angles  $\Phi$  et  $\Psi$  du résidu  $i+2$  ont les valeurs d'une conformation en brin  $\beta$  (Tableau 7).<sup>41,42</sup>

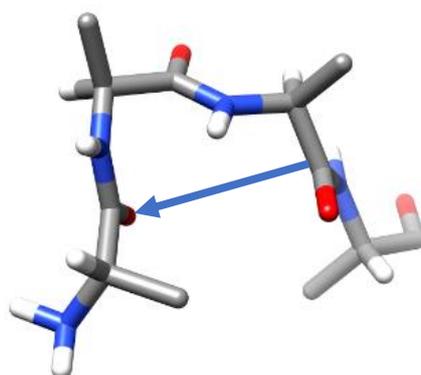


Figure 15 Représentation schématique du coude  $\beta$  de type VIII identifié par Wilmot et Thornton en 1988.<sup>41</sup> Les atomes sont représentés en bleu pour l'azote, en rouge pour l'oxygène, en gris pour le carbone et en blanc pour l'hydrogène. La liaison hydrogène est représentée par une flèche en bleu clair.

Dans cette étude, une analyse des valeurs des angles des structures en conformation coude  $\beta$  de type VIa et VIb a également été réalisée dans le but de préciser les possibles valeurs d'angles dièdres  $\Phi$  et  $\Psi$ . Ces valeurs sont reportées dans le Tableau 7.

Tableau 7 Valeurs des angles dièdres pour les résidus ( $i+1$ ) et ( $i+2$ ) impliqués dans les coudes de nature  $\beta$  des types identifiés par Wilmot and Thornton en 1988.<sup>41</sup> En gras, les modifications apportées à la classification par Wilmot et Thornton.

Structure	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
Coude $\beta$ I	-60°	-30°	-90°	0°
Coude $\beta$ I'	+60°	+30°	+90°	0°
Coude $\beta$ II	-60°	+120°	-80°	0°
Coude $\beta$ II'	+60°	-120°	+80°	0°
Coude $\beta$ IV	Tous les coudes $\beta$ qui possèdent deux ou plus angles dièdres ( $\Phi_{i+1}$ , $\Psi_{i+1}$ , $\Phi_{i+2}$ et/ou $\Psi_{i+2}$ ) différant au moins de 40° par rapport aux valeurs canoniques des autres types de coudes $\beta$			
<b>Coude <math>\beta</math> VIa</b>	<b>-60°</b>	<b>+120°</b>	<b>-80°</b>	<b>0°</b>
<b>Coude <math>\beta</math> VIb</b>	<b>-120°</b>	<b>+120°</b>	<b>-60°</b>	<b>0°</b>
<b>Coude <math>\beta</math> VIII</b>	<b>-60°</b>	<b>-30°</b>	<b>-120°</b>	<b>+120°</b>

De nouvelles modifications ont été effectuées par Hutchinson et Thornton au cours de l'année 1994.<sup>43</sup> En effet, afin d'éclaircir la classification des coudes  $\beta$  de type IV, une base de données de 1 666 structures a été construite. Ensuite, des calculs théoriques de minimisation d'énergie ont été réalisés. Les structures ont été regroupées selon les valeurs des angles dièdres caractéristiques d'un motif de type coude et un groupe majoritaire composé de 425 structures a été identifié.<sup>43</sup> Les angles  $\Phi$  et  $\Psi$  des résidus centraux de chaque structure ont été mesurés et ces mesures ont permis de définir les valeurs moyennes des angles dièdres pour ce type de coude  $\beta$  (Tableau 8). En outre, pour les coudes de type VIa, deux sous-groupes ont été identifiés : le sous-type VIa1 et le sous-type VIa2 (Figure 16, Tableau 8). La distinction se fait sur des valeurs de  $\Phi$  et de  $\Psi$  assez différentes adoptées par le résidu  $i+1$  (Tableau 8).<sup>40,43,44</sup>

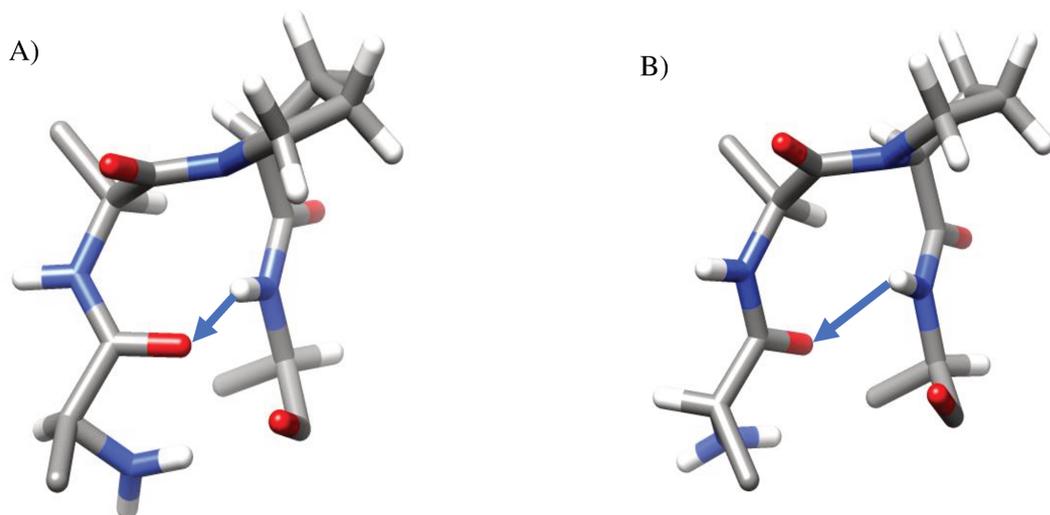


Figure 16 Représentation schématique des coudes  $\beta$  identifié par Hutchinson et Thornton en 1994.<sup>43</sup> (A) coude  $\beta$  de type VIa1, (B) coude  $\beta$  de type VIa2. Les atomes sont représentés en bleu pour l'azote, en rouge pour l'oxygène, en gris pour le carbone et en blanc pour l'hydrogène. La liaison hydrogène est représentée par une flèche en bleu clair.

Au contraire, pour le type VIb aucun sous-type n'a été identifié, mais des valeurs  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$ , et  $\Psi_{i+2}$  ont été attribuées pour ce type de coude (Figure 17, Tableau 8).<sup>43,44</sup>

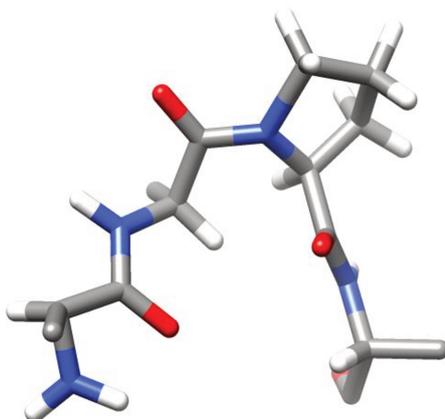


Figure 17 Représentation schématique du coude de nature  $\beta$  de type VIb identifié par Hutchinson et Thornton en 1994.<sup>43</sup> Les atomes sont représentés en bleu pour l'azote, en rouge pour l'oxygène, en gris pour le carbone et en blanc pour l'hydrogène. La liaison hydrogène est représentée par une flèche en bleu clair.

Tableau 8 Valeurs des angles dièdres pour les résidus ( $i+1$ ) et ( $i+2$ ) impliqués dans les coudes de nature  $\beta$  des types classifiés par Hutchinson et Thornton et al en 1994.<sup>43</sup> En gras, les modifications apportées à la classification par Hutchinson et Thornton.

Structure	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
Coude $\beta$ I	-60°	-30°	-90°	0°
Coude $\beta$ I'	+60°	+30°	+90°	0°
Coude $\beta$ II	-60°	+120°	-80°	0°
Coude $\beta$ II'	+60°	-120°	+80°	0°
<b>Coude <math>\beta</math> IV</b>	<b>-61°</b>	<b>+10°</b>	<b>-53°</b>	<b>+17°</b>
<b>Coude <math>\beta</math> VIa1</b>	<b>-60°</b>	<b>+120°</b>	<b>-90°</b>	<b>0°</b>
<b>Coude <math>\beta</math> VIa2</b>	<b>-120°</b>	<b>+120°</b>	<b>-60°</b>	<b>0°</b>
<b>Coude <math>\beta</math> VIb</b>	<b>-135°</b>	<b>+135°</b>	<b>-75°</b>	<b>+160°</b>
Coude $\beta$ VIII	-60°	-30°	-120°	+120°

Enfin, une analyse récente, effectuée en 2016 par le Dr. De Brevern, a proposé une nouvelle classification pour les coudes  $\beta$  de type IV.<sup>45</sup> En effet, ces structures représentent 1/3 des résidus de coudes  $\beta$ . Pour localiser de nouvelles conformations récurrentes dans ce type de coude  $\beta$ , une approche de regroupement automatique (clustering), sur la base des règles d'attribution utilisées pour définir un coude  $\beta$  a été réalisée. Les quatre groupes les plus fréquents ont été identifiés (IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub> et IV<sub>4</sub>) comme nouveaux sous-types et analysés (Figure 18). Pour chaque sous-type, des valeurs d'angles pour les résidus centraux ont été attribuées (Tableau 9).

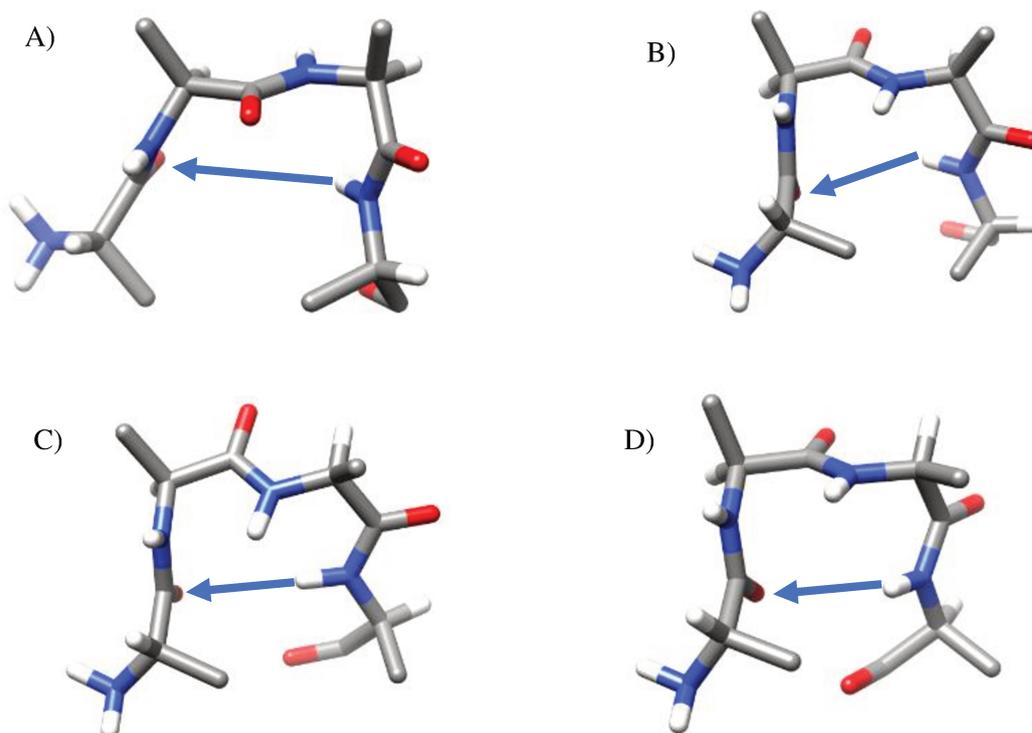


Figure 18 Représentation schématique de quatre types de coudes  $\beta$  identifiés par De Brevern en 2016.<sup>45</sup> (A) coude  $\beta$  de type  $IV_1$ , (B) coude  $\beta$  de type  $IV_2$ , (C) coude  $\beta$  de type  $IV_3$ , (D) coude  $\beta$  de type  $IV_4$ . Les atomes sont représentés en bleu pour l'azote, en rouge pour l'oxygène, en gris pour le carbone et en blanc pour l'hydrogène. La liaison hydrogène est représentée par une flèche en bleu clair.

Tableau 9 Valeurs des angles dièdres pour les résidus  $(i + 1)$  et  $(i + 2)$  impliqués dans différents types de coudes  $\beta$  en utilisant la classification de A. De Brevern proposée en 2016.<sup>45</sup> En gras, les modifications apportées à la classification par De Brevern..

Type de coude	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
I	-60°	-30°	-90°	0°
I'	+60°	+30°	+90°	0°
II	-60°	+120°	+80°	0°
II'	+60°	-120°	-80°	0°
<b>IV<sub>1</sub></b>	<b>-120°</b>	<b>+130°</b>	<b>+55°</b>	<b>+41°</b>
<b>IV<sub>2</sub></b>	<b>-85°</b>	<b>-15°</b>	<b>-125°</b>	<b>+55°</b>
<b>IV<sub>3</sub></b>	<b>-71°</b>	<b>-30°</b>	<b>-72°</b>	<b>-47°</b>
<b>IV<sub>4</sub></b>	<b>-97°</b>	<b>-2°</b>	<b>-117°</b>	<b>-11°</b>
VIa1	-60°	+120°	-90°	0°
VIa2	-120°	+120°	-60°	0°
VIb	-135°	+135°	-75°	160°
VIII	-60°	-30°	-120°	120°

En conclusion, la classification actuelle des coudes de type  $\beta$  regroupe 12 sous-types différents (I, I', II, II', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb, VIII et IV<sub>misc</sub>). Les valeurs des angles dièdres  $\Phi$  et  $\Psi$  des résidus centraux sont résumées dans le Tableau 9. Le travail de thèse présenté dans ce manuscrit s'appuiera sur cette classification.

## **II.2.2 Caractéristiques des coudes $\beta$**

### **Caractéristiques liées à leur localisation dans les protéines et les peptides**

L'analyse de la localisation des coudes  $\beta$  dans les protéines et les peptides a montré, comme déjà dit auparavant, que nous pouvons les retrouver à la surface de ces macromolécules et qu'ils relient d'autres types de structures secondaires de différentes manières.<sup>2</sup> Il a été mis en évidence que les coudes de nature  $\beta$  sont retrouvés dans une structure super-secondaire.<sup>46</sup> Plus précisément, les images miroirs (les types I' et II'), qui sont les moins favorables au niveau énergétique à cause de leur encombrement stérique, sont les coudes impliqués dans le motif structural «  $\beta$ -hairpin ». <sup>39,41,46</sup> Ceci est dû au fait que le type I' possède la torsion convenable qui correspond à la torsion relative qui est observée normalement dans deux brins  $\beta$  antiparallèles.<sup>47</sup> De même, le type II', qui est relativement planaire, est aussi l'une des conformations privilégiées pour relier un feuillet  $\beta$  antiparallèle.<sup>46</sup> Au contraire, les autres types de coudes  $\beta$  relient les structures secondaires (hélice  $\alpha$  et brin  $\beta$ ) avec plusieurs combinaisons possibles et ils peuvent être aussi rencontrés dans des parties en conformation aléatoire des protéines. En outre, des études sur une succession de coudes ont été aussi menées. Isogai et ses collaborateurs ont montré que l'enchaînement de coudes successifs forment soit des structures peptidiques approximativement hélicoïdales, soit des retournements de chaîne plus complexes.<sup>48</sup>

### **Analyse de la fréquence des types de coudes $\beta$**

En se basant sur une étude de prédiction de la fréquence des coudes  $\beta$  dans les structures des protéines réalisée par Fuchs et Alix<sup>49</sup> en 2005, l'analyse réalisée par le Dr. De Brevern<sup>45</sup> a mis en évidence que les coudes  $\beta$  de type I sont les plus répandus (38,2 %), suivis par les types II (11,8 %) et VIII (9,8 %). Ensuite, les types IV<sub>1</sub> (5,1 %), I' (4,1 %), IV<sub>2</sub> (3,9 %), IV<sub>3</sub> (3,5 %), IV<sub>4</sub> (2,7 %) et II' (2,5 %) sont ceux retrouvés avec des fréquences plus modérées. Les types VI sont les moins rencontrés avec une gamme de fréquence entre 0,2 et 0,9% (Tableau 10).<sup>45</sup>

Tableau 10 Fréquences des coudes  $\beta$  dans les structures cristallographiques de protéines reportées dans la PDB (tiré de De Brevern 2016).<sup>45</sup>

Type de coude $\beta$	(%)
I	38,2
I'	4,1
II	11,8
II'	2,5
IV <sub>1</sub>	5,1
IV <sub>2</sub>	3,9
IV <sub>3</sub>	3,5
IV <sub>4</sub>	2,7
VIa1	0,7
VIa2	0,2
VIb	0,9
VIII	9,8
IV <sub>misc</sub>	16,6

### Caractéristiques structurales

Les coudes  $\beta$  ont tous la même orientation standard du squelette peptidique. En effet, les quatre carbones  $\alpha$  ( $C_\alpha$ ) se retrouvent sur le même plan. Les carbones  $\beta$  ( $C_\beta$ ) sont toujours dans le plan ou au-dessus du plan pour tous les types de coudes  $\beta$ .

Bien que la présence de la liaison hydrogène ait conduit à la caractérisation initiale des coudes<sup>38</sup>, cette liaison a été supprimée comme condition obligatoire dès que des études sur des structures protéiques connues ont été effectuées.<sup>50</sup> En effet, de nombreux exemples de coudes ont été identifiés en dehors d'une distance plausible de liaison hydrogène mais avec une conformation par ailleurs très coudée. Les différentes conformations de coudes sont suffisamment énergétiquement favorisées pour ne pas nécessiter de stabilisation par la liaison hydrogène, comme dans le type VIb (Figure 17).<sup>44</sup> Cela n'est pas surprenant, car les différents types de coudes sont généralement retrouvés à la surface des protéines et des peptides. Cela permet à ces structures secondaires de se lier, à travers des liaisons hydrogène, au solvant qui peut les stabiliser.<sup>51,52</sup>

En outre, l'angle dièdre « virtuel » défini par les quatre atomes  $C_\alpha$  est proche de  $0^\circ$  pour le type II ou II', il est quelque peu positif (moyenne d'environ  $+45^\circ$ ) pour le type I et légèrement négatif ( $-45^\circ$ ) pour le type I'.<sup>40</sup> Dans les types I et II', l'oxygène du groupe carbonyle du deuxième

résidu pointe à environ  $90^\circ$  vers le bas par rapport au plan (Figure 13), tandis que dans les types I' et II, il pointe à environ  $90^\circ$  vers le haut par rapport au plan (Figure 13). Si l'on considère l'oxygène du groupe carbonyle du premier acide aminé, il est dirigé avec un angle d'environ  $90^\circ$  vers le bas par rapport au centre du plan formé par les  $C_\alpha$  dans le type I, vers le haut dans le type I', légèrement vers le bas dans le type II et légèrement vers le haut dans le type II' (Figure 13). La position du second oxygène carbonyle distingue alors les types I et II (ou I' et II'), tandis que la position du premier oxygène carbonyle distingue le type I contre le type II' (ou le type II vers le type I').<sup>40,42,43</sup>

En revanche, il est important de souligner que l'idée d'un coude approximativement planaire avec une liaison hydrogène linéaire n'est vérifiée que pour le type II. Ainsi, même un coude « idéal » de type I est non planaire car les groupes amide et carbonyle engagés dans la liaison hydrogène doivent être presque perpendiculaires l'un par rapport à l'autre (Figure 13). Cet oxygène se trouve dans le plan des trois derniers carbones  $\alpha$ , mais le premier  $C_\alpha$  et donc le premier résidu sont relevés hors du plan, produisant un angle dièdre virtuel de  $45^\circ$ . Pour les autres types de coudes de nature  $\beta$ , cette analyse n'a jamais été réalisée. Néanmoins, quelques informations supplémentaires ont été mises en évidence par de Brevern en 2016. En ce qui concerne les types de coudes  $IV_1$  et  $IV_2$ , de Brevern a mis en évidence une relation entre les coudes  $\beta$  de type  $IV_1$  et de type II ainsi qu'entre les types  $IV_2$  et VIII. En terme de valeurs d'angles dièdres, le type  $IV_1$  ressemble à une conformation légèrement décalée du type II (Tableau 9), alors que le type  $IV_2$  semble être un type VIII moins étendu (Tableau 9).<sup>45</sup> Les types  $IV_3$  et  $IV_4$  se distinguent des autres à cause des valeurs d'angles dièdres de type hélicoïdal (Tableau 9).<sup>45</sup>

### **Caractéristiques au niveau de la séquence en acides aminés**

Si nous nous intéressons maintenant aux séquences en acides aminés qui composent les structures secondaires de type coudes  $\beta$ , il a été montré qu'il existe des préférences pour certains résidus dans des positions spécifiques.<sup>41,42</sup> Ceci est dû aux caractéristiques structurales de chaque type de coude. Rose et Kuntz ont ainsi montré une forte tendance à l'hydrophilie des résidus impliqués au sein d'une structure de type coude  $\beta$ .<sup>36,53</sup> La préférence de résidus hydrophiles pourrait être le résultat de l'emplacement presque universel des coudes  $\beta$  à la surface des protéines où ils se trouvent pour relier des segments de structures secondaires différentes plus internes.<sup>36,53,54</sup> Les analyses de structures cristallographiques de protéines présentant des coudes  $\beta$  réalisées par Guruprasad et ses collaborateurs (2000)<sup>54</sup> et par de

Brevern (2016)<sup>45</sup> ont permis de déterminer la fréquence de chaque acide aminé dans les positions des différents types de coudes  $\beta$ . Ici, nous allons nous concentrer sur l'analyse des préférences en acides aminés pour les positions caractéristiques des coudes  $\beta$  ( $i+1$  et  $i+2$ ). Pour le coude  $\beta$  de type I, une grande variabilité est observée pour les résidus  $i+1$  et  $i+2$ . Cependant, des fortes propensions de l'acide aminé proline en position  $i+1$  ainsi que des résidus asparagine et aspartate en position  $i+2$  ont été remarquées (Figure 19).<sup>37,45</sup> Dans le type I', une propension pour les acides aminés asparagine, aspartate et histidine en position  $i+1$  peut être observée (Figure 19). La glycine est fortement préférée en position  $i+2$  pour ce type de coude. Dans le coude de type II, les acides aminés alanine, glutamate, lysine avec une importante présence de la proline ont été observés pour le résidu  $i+1$  alors que la glycine et l'asparagine sont les plus retrouvées pour le résidu  $i+2$  (Figure 19). Au contraire, dans l'image miroir, le type II', la glycine se trouve préférentiellement en position  $i+1$  et dans la position suivante, nous retrouvons les acides aminés serine, thréonine et asparagine (Figure 19).

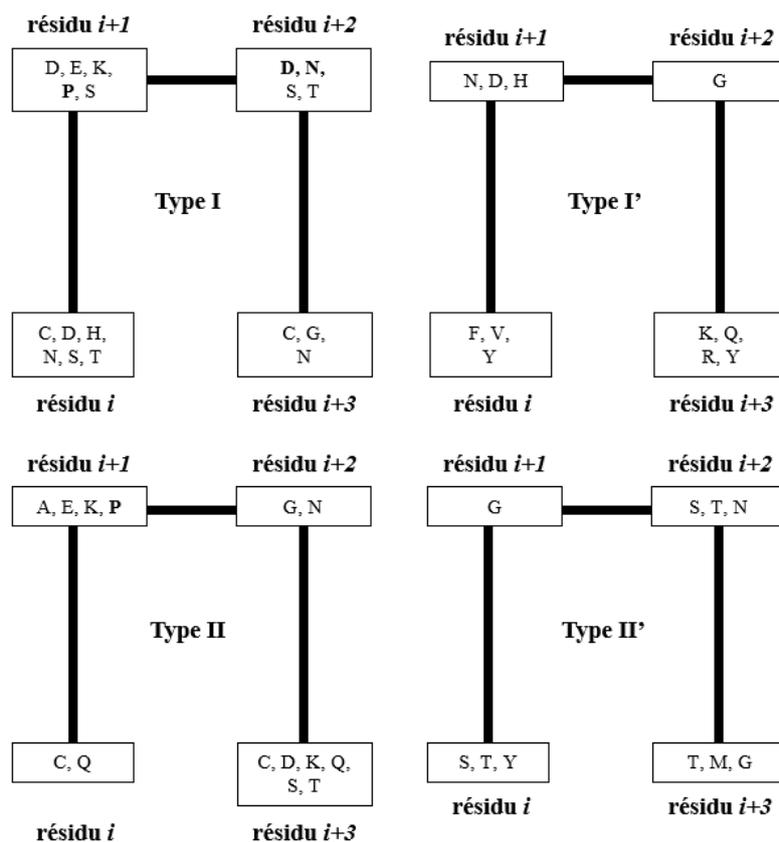


Figure 19 Représentation des préférences en acides aminés pour les coudes  $\beta$  de type I, I', II, II' (tiré de Guruprasad et al., 2000 et De Brevern, 2016)<sup>37, 45</sup>

En ce qui concerne les types IV, nous pouvons observer que tous les sous-types ont une grande variabilité d'acides aminés pour les deux résidus centraux, sauf le type IV<sub>3</sub> en position  $i+1$ , où seulement les acides aminés alanine et proline ont été observés (Figure 20).

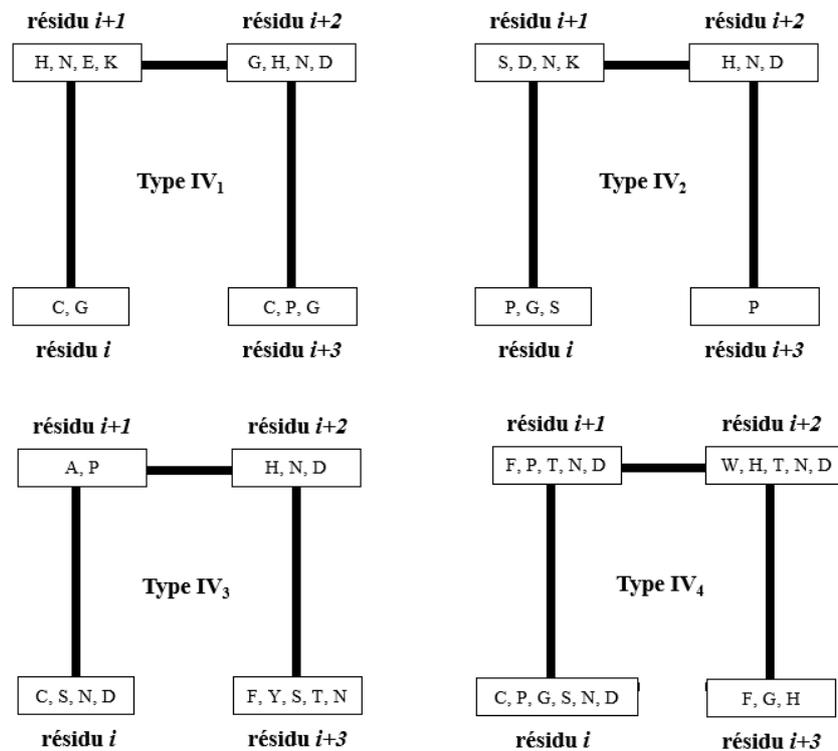


Figure 20 Représentation des préférences en acides aminés pour les coudes  $\beta$  de type IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub> et IV<sub>4</sub> (tiré de De Brevern, 2016)<sup>45</sup>

Les types VI sont caractérisés par la présence d'une proline, en conformation *cis*, qui est obligatoirement en position  $i+2$  (Figure 21). Pour ce qui concerne la position  $i+1$ , le type VIa1 est caractérisé par trois préférences d'acides aminés : la proline, la tyrosine et la phénylalanine (Figure 21).<sup>37</sup> Dans le type VIa2, on retrouve l'asparagine et phénylalanine (Figure 21) et dans le type VIb, la thréonine et la tyrosine (Figure 21).<sup>37,43,45</sup>

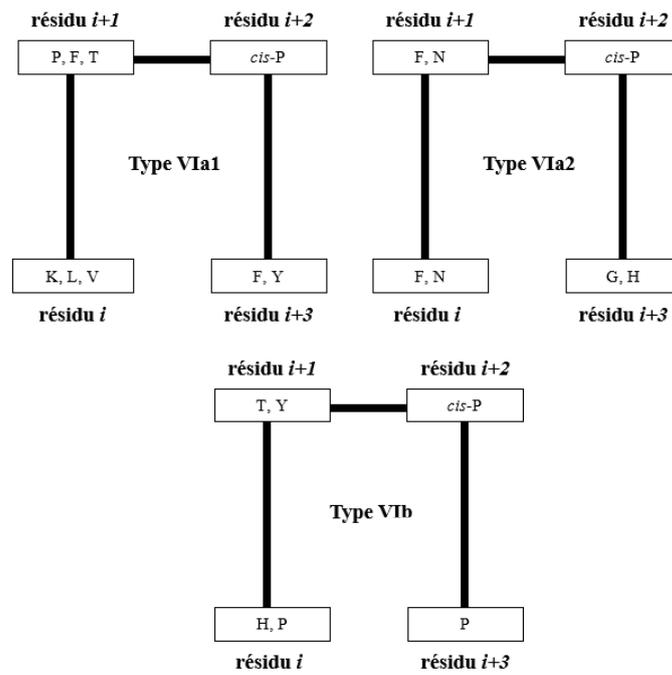


Figure 21 Représentation des préférences en acides aminés pour les coudes  $\beta$  de type VIa1, VIa2, VIb (tiré de Guruprasad et al., 2000 et De Brevern, 2016)<sup>37, 45</sup>

Enfin, le coude  $\beta$  de type VIII peut présenter la proline en conformation *trans* et l'aspartate en position  $i+1$  (Figure 22).<sup>37,43,45</sup> L'histidine, la tyrosine, l'isoleucine, la phénylalanine et l'asparagine sont caractéristiques de la position  $i+2$  (Figure 22).<sup>37,43,45</sup>

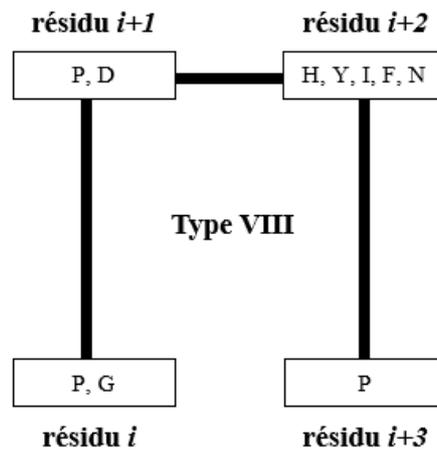


Figure 22 Représentation des préférences en acides aminés pour les coudes  $\beta$  de type VIII (tiré de Guruprasad et al., 2000 et De Brevern, 2016)<sup>37,45</sup>

### III. Méthodes d'analyse de la structure des protéines et des peptides

La détermination de la structure des protéines et des peptides est un exercice difficile mais des outils performants sont à la disposition des chercheurs. Il existe plusieurs méthodes expérimentales qui permettent d'étudier la structure de peptides et de protéines apportant des informations plus ou moins détaillées. Parmi ces techniques, certaines rendent possible l'identification des éléments structuraux de la structure secondaire (dichroïsme circulaire, spectroscopie Infra-Rouge et Résonance Magnétique Nucléaire), d'autres permettent d'étudier la structure à des niveaux plus élevés (structure super secondaire et tertiaire) au niveau de l'échelle atomique (cristallographie aux rayons X, cryo-microscopie électronique et Résonance Magnétique Nucléaire couplée à la modélisation moléculaire).

L'analyse en solution de ces molécules biologiques est réalisée en utilisant la spectroscopie Infra-Rouge (IR), le dichroïsme circulaire (DC), la Résonance Magnétique Nucléaire (RMN) du liquide. L'étude à l'état solide est réalisée grâce à la cristallographie aux rayons X, la RMN du solide et le cryo-microscopie électronique (Cryo-EM).

En parallèle aujourd'hui des méthodes expérimentales, les méthodes *in silico* pour les études structurales sont de plus en plus développées. Ces techniques incluent les méthodes *ab initio*, les méthodes basées sur la mécanique quantique et la modélisation par homologie.

#### III.1 Méthodes pour l'étude de la structure secondaire

##### III.1.1 La Spectroscopie Infrarouge (IR)

La spectroscopie IR est une méthode efficace pour étudier la structure secondaire des protéines et des peptides à travers l'analyse du spectre d'absorption dans l'infrarouge. Elle mesure l'excitation vibrationnelle des atomes autour de la liaison qui les unissent. L'empreinte observée dépend de la nature du groupe fonctionnel et de l'environnement proche. Les principales bandes d'absorption d'une chaîne polypeptidique en infrarouge sont dues aux modes vibrationnels des groupements amide de la chaîne peptidique, et sont appelées bande amide I ( $1700 - 1600 \text{ cm}^{-1}$ ), bande amide II ( $1600 - 1500 \text{ cm}^{-1}$ ) et bande amide III ( $1400 - 1200 \text{ cm}^{-1}$ ).<sup>55</sup> La bande amide I est la bande la plus utilisée pour l'analyse de la structure secondaire des peptides et des protéines.

Il est possible d'attribuer distinctement des composantes de la bande amide I à des éléments spécifiques de structures secondaires de ces molécules biologiques selon la position de leurs pics.<sup>56</sup>

Les hélices  $\alpha$  sont trouvées dans la région 1660 – 1648  $\text{cm}^{-1}$  (Tableau 11). La position de cette bande dépend de la longueur de l'hélice, de sa flexibilité et de son hydratation. Ainsi, les grands nombres d'onde correspondent à des hélices courtes et flexibles, tandis que les petites valeurs sont associées à des structures longues et rigides.<sup>57</sup> Les feuilletts  $\beta$  intramoléculaires possèdent deux bandes d'absorption de différentes intensités. Pour les structures antiparallèles, une bande basse fréquence est trouvée autour de 1633  $\text{cm}^{-1}$  et une bande haute fréquence autour de 1686  $\text{cm}^{-1}$  (Tableau 11). Pour les feuilletts  $\beta$  parallèles, la bande à basse fréquence est légèrement plus grande tandis que la bande à haute fréquence est plus faible que les feuilletts antiparallèles.<sup>56</sup> La bande amide I dans les structures en conformation aléatoire est due aux contributions des liaisons peptidiques dans différents environnements, menant par conséquent à une bande large centrée autour de 1654  $\text{cm}^{-1}$  (Tableau 11). Cependant, cette bande est superposée à celle correspondant aux hélices  $\alpha$ . Par échange H/D, un déplacement vers 1645  $\text{cm}^{-1}$  est observé permettant de discriminer cette bande par rapport à celle des hélices.<sup>55</sup> Deux bandes d'absorption dans entre 1678 et 1682  $\text{cm}^{-1}$  ont été attribuées aux structures de type coude  $\beta$ .<sup>56,58,59</sup> Cependant, il n'est pas encore possible d'identifier des bandes différentes pour chaque type de coude.

Tableau 11 Fréquences IR de la bande amide I selon les différentes structures secondaires.

Fréquence ( $\text{cm}^{-1}$ )	Nombre de bandes	Structure secondaire
1633-1686	2	Feuillet $\beta$
1654-1670	1	Conformation aléatoire
1648-1660	1	Hélice $\alpha$
1678-1682	2	Coude $\beta$

### III.1.2 Le dichroïsme circulaire

L'aptitude des composés asymétriques à interagir avec un rayonnement électromagnétique, leur confère une activité optique importante. Le dichroïsme circulaire (DC) est ainsi défini comme l'absorption différentielle ( $\Delta\epsilon$ ) des lumières polarisées circulairement à gauche ( $\epsilon_G$ ) et à droite ( $\epsilon_D$ ) à une longueur d'onde donnée. La spectroscopie de DC est un outil très sensible pour détecter les structures secondaires des peptides et des protéines en solution.<sup>60,61</sup> En effet, les macromolécules biologiques, comme les peptides ou les protéines, sont composés d'éléments optiquement actifs (liaisons peptidiques, cycles aromatiques, etc....) et produisent un spectre

DC qui leur est propre et est caractéristique de la structure secondaire qu'elles adoptent. L'étude du spectre DC d'un peptide ou d'une protéine va ainsi permettre de déterminer le principal type de structure secondaire en présence et d'en estimer sa proportion.<sup>62</sup> Le spectre de DC obtenu sera une combinaison linéaire de chaque spectre de référence, pondérée par la fraction de chaque type de structure secondaire. Les structures secondaires (hélice  $\alpha$  et brin  $\beta$ ) et la conformation aléatoire présentent des spectres de DC de référence, au contraire des motifs de type coude où seul un spectre de DC moyen est disponible.<sup>63</sup> L'absence des spectres de DC de référence pour tous les types de coudes  $\beta$  et  $\gamma$  limite l'analyse de molécules riches en coudes comme les peptides bioactifs.

Le principe général du DC et son fonctionnement sont décrits en Annexe 1. Dans le chapitre 2, nous détaillerons l'utilisation de cette technique pour caractériser la structure secondaire des polypeptides. Cette méthode peut présenter une limite pour l'analyse de molécules riches en coudes tandis que l'identification des structures hélicoïdales ou en brin  $\beta$  est réalisé sans difficulté.

### III.1.3 La RMN

La spectroscopie de résonance magnétique nucléaire (RMN) des protéines et des peptides étudie principalement le repliement le long de la chaîne polypeptidique. L'étude peut se faire par RMN homonucléaire ou hétéronucléaire, suivant que l'on observe les noyaux protons seuls ou avec les noyaux carbones et/ou azotes. Les expériences RMN permettent l'observation des couplages scalaires et dipolaires entre spins. L'observation du transfert d'aimantation, par couplage dipolaire, produit des effets Overhauser nucléaires (NOE) entre noyaux situés à moins de 5 Å l'un de l'autre, et permet l'estimation des distances entre les noyaux observés. Dans le cas d'une étude ( $^1\text{H}$ ) en solution, la comparaison des NOE expérimentaux avec ceux de référence, identifiés par Wagner et ses collaborateurs, permet de réaliser une interprétation qualitative de ce paramètre structural en fonction de la structure primaire pour l'identification des structures secondaires de type hélice et feuillet.<sup>64</sup> Cependant, la RMN homonucléaire présente certaines limitations pour l'étude des protéines et des peptides à cause du taux de superposition des pics sur les spectres et du mauvais transfert d'aimantation entre protons par couplage scalaire.

Ces problèmes, qui rendaient difficile l'utilisation de la RMN pour l'étude de protéines de taille supérieure à cent résidus, ont été réduits par l'utilisation des propriétés RMN de noyaux autres que le proton. En effet, des noyaux différents résonnent dans des gammes de fréquences

différentes, ce qui permet de résoudre les superpositions de déplacements chimiques. De plus, les constantes de couplage hétéronucléaires améliorent la sensibilité du couplage scalaire. L'interprétation qualitative des NOE en fonction de la structure primaire de la molécule permet l'identification des structures secondaires de type hélice ou feuillet. En revanche, celle des coudes est beaucoup plus difficile. En effet, ces derniers sont des motifs non répétitifs et impliquant peu d'acides aminés. De ce fait, le nombre de NOE disponibles pour cette structure secondaire est très limité. En outre, le diagramme NOE de référence qui a été établi est limité aux coudes  $\beta$  de type I, I', II et II'.<sup>64</sup>

## **III.2 Méthodes pour l'étude de la structure tertiaire**

### **III.2.1 La cristallographie aux rayons X**

La cristallographie a été l'une des premières méthodes permettant de déterminer la structure tridimensionnelle des protéines et reste l'outil le plus performant pour connaître la structure exacte d'une molécule à l'état solide car elle permet d'avoir une description directement à l'échelle atomique.<sup>65</sup> La condition principale pour pouvoir résoudre la structure tridimensionnelle d'une protéine par cristallographie aux rayons X est d'avoir un cristal bien ordonné qui diffracte correctement.<sup>66</sup> Le problème rencontré avec les protéines est la difficulté d'obtenir des cristaux réguliers. De même, son utilisation est limitée par la capacité de la molécule à cristalliser et il s'agit d'une difficulté pour les peptides dont la flexibilité entrave souvent la cristallisation.<sup>65,67</sup>

### **III.2.2 La RMN en solution couplée à la modélisation moléculaire (MM)**

La RMN en solution couplée à la modélisation permet également de déterminer la structure tridimensionnelle d'une protéine ou d'un peptide. Cela consiste à trouver un ensemble de conformations qui correspondent à des minima énergétiques et qui sont en accord avec les données expérimentales. Les calculs s'effectuent dans le cadre du formalisme de la mécanique classique empirique. Les données expérimentales de RMN sont utilisées en tant que contraintes dans les calculs. La relation entre les paramètres RMN et les paramètres géométriques de la protéine étant assez qualitative, les contraintes sont appliquées sous forme d'intervalles plus ou moins larges.

En 1985, la structure 3D de l'inhibiteur de trypsine de taureau<sup>68</sup> a été obtenue par la méthode de géométrie des distances développée par W. Braun dans l'équipe de K. Wüthrich.<sup>69</sup> Cette

méthode de géométrie des distances est de moins en moins utilisée, au bénéfice des méthodes de calcul de recuit simulé. Le principe du recuit simulé sera détaillé dans le chapitre 3.

### III.2.3 Cryo-microscopie électronique (Cryo-EM)

La cryo-microscopie électronique est une alternative aux autres méthodes biophysiques expérimentales lorsqu'on s'intéresse à des molécules d'une masse supérieure à 250 kDa.<sup>70</sup> Elle permet d'obtenir des informations sur des molécules isolées sur une large gamme de résolution. Le développement méthodologique et instrumental de la cryo-microscopie électronique a permis ces dernières années d'obtenir un nombre croissant de structures de larges macromolécules à des résolutions intermédiaires (6-9 Å).<sup>71,72</sup> À de telle résolution, il devient possible d'identifier certaines structures secondaires telles que les hélices  $\alpha$ . D'autre part, la structure d'un large complexe macromoléculaire peut être détaillée en déterminant la structure atomique par RMN ou cristallographie des éléments constitutifs du complexe. Ensuite, l'interaction entre ces éléments peut être modélisée en les plaçant correctement dans la densité électronique du complexe total déterminée par microscopie électronique. Par ailleurs, la microscopie électronique peut être appliquée à la compréhension détaillée de processus dynamiques. Ainsi, pour un système biologique dont le développement dans le temps est connu et maîtrisé, différents états d'un même système peuvent être isolés en congelant dans l'éthane liquide à différents temps.<sup>73,74</sup>

### III.3 Méthodes *in silico*

Le développement d'outils informatiques fiables couplé à la croissance de la puissance informatique a permis la mise en place de techniques de simulation numérique pour les molécules biologiques, comme les protéines et les peptides. Par analogie avec les expressions *in vivo* et *in vitro*, le terme « *in silico* » a été introduit pour qualifier les méthodes computationnelles mises en œuvre pour traiter de tels systèmes. Le domaine des méthodes *in silico* regroupe un très large ensemble de méthodes numériques fondées sur les lois de la physique et de la chimie qui, utilisant les approches des mathématiques, permettent de simuler ou de modéliser un système biologique et ses propriétés à l'aide de l'outil informatique. Ces méthodes de modélisation moléculaire permettent de déterminer la structure tridimensionnelle, l'énergie ainsi que d'autres propriétés chimique-physiques des molécules et des systèmes biologiques à partir de leur composition atomique.

Différentes approches de modélisation sont envisageables. Si celles concernant la mécanique moléculaire permettent de traiter des systèmes moléculaires de grande taille, les méthodes quantiques (*ab initio*, semi-empiriques, ou théorie de la fonctionnelle de la densité) sont quant à elles capables de calculer les propriétés électroniques des systèmes de petite taille (par exemple, 4 acides aminés).

### III.3.1 Les méthodes *ab initio* pour l'étude du repliement

Les méthodes *ab initio* n'utilisent pas d'informations directement issues de bases de données expérimentales en tentent de reconstruire la structure d'une protéine à partir de la seule connaissance de sa séquence.<sup>75</sup> Elle tentent d'denier, pour une protéine, la structure ayant l'énergie libre la plus basse, et ce en utilisant uniquement la séquence en acides aminés de cette dernière et les lois physique. Cette classe de prédiction de la structure des protéine peut *ab initio* être utilisée pour n'importe quelle séquence protéique avec des potentiels ayant un sens physique et une représentation atomique des modèles.<sup>76-78</sup> C'est de loin la catégorie la plus complexe, mais aussi la plus intéressante tant elle peut nous apprendre sur le repliement des protéines. Il s'avère qu'à l'heure actuelle, bien que ces méthodes donnent des résultats encourageants pour des protéines de petite taille (50 acides aminés), elles ne sont pas les plus performantes pour générer des modèles protéiques de taille plus importantes.<sup>75,79</sup>

### III.3.2 Les méthodes basées sur la mécanique quantique

Les méthodes fondées sur la mécanique quantique (MQ) s'appuient sur la résolution approchée de l'équation de Schrödinger.<sup>80</sup> Elles permettent la détermination de la configuration électronique de la molécule, sa structure tridimensionnelle, son énergie et ses propriétés physico-chimiques. Le plus souvent, ces propriétés sont déterminées de façon quantitative avec une précision qui avoisine celle de l'expérience. En revanche, les méthodes de MQ permettent seulement de traiter des systèmes dont la taille est modérée (4 acides aminés).<sup>81</sup> Pour des systèmes plus grands, les méthodes de MQ permettent néanmoins d'utiliser plusieurs niveaux de théorie pour étudier les différentes parties du système. De ce fait, une partie du système sera traité en utilisant la mécanique quantique alors que pour le reste la mécanique moléculaire sera employée.<sup>82,83</sup> Le choix du type de calcul dépendra donc du problème étudié (taille et précision souhaitée pour le calcul) et évidemment des ressources de calcul (puissance de calcul) mises à disposition de l'utilisateur.

### III.3.3 Modélisation par homologie

Les techniques de modélisation moléculaire par homologie constituent un autre moyen pour obtenir des informations sur la structure. La modélisation par homologie permet d'obtenir un modèle tridimensionnel théorique d'une protéine lorsque sa structure n'a pas été déterminée expérimentalement.<sup>84,85</sup> C'est un moyen de construire des modèles 3D d'une protéine sur la base d'une autre protéine « patron » de structure expérimentale connue.<sup>86,87</sup> Elle repose sur les idées suivantes :

- des séquences primaires ressemblantes présentent des structures 3D similaires,
- des structures 3D similaires engendrent des mécanismes d'action proches,

Les méthodes de *novo* visent à prédire les structures peptidiques à partir de séquences d'acides aminés. PEP-FOLD utilise cette approche qui repose sur l'alphabet structural lettres SA pour décrire les conformations de quatre résidus consécutifs.

## Objectifs de la thèse

Avec les hélices  $\alpha$  et les feuillets  $\beta$ , les coudes sont la troisième structure secondaire rencontrée au sein des peptides et des protéines. L'implication de ce motif structural dans le repliement de protéines et dans la reconnaissance des peptides bioactifs par leurs cible biologique fait de leur caractérisation un élément important dans la chimie médicinale afin de pouvoir concevoir des peptidomimétiques.

En dépit du fait que les coudes soient fréquemment rencontrés dans les peptides et dans les protéines, leur caractérisation structurale avec des méthodes expérimentales en solution est plus difficile que celle des autres motifs structuraux (hélice ou feuillet). En effet, la structure ordonnée mais non répétitive, la grande flexibilité et l'existence de plusieurs catégories, possédant des géométries différentes, rendent la caractérisation et la différenciation entre les différents types de coudes un enjeu de taille car peu d'outils expérimentaux sont disponibles. Ceci est en partie lié au fait que la majorité des études de biologie structurale porte sur des protéines, dans lesquelles sont principalement décrits les motifs structuraux de type hélice et feuillet. Ceci constitue une difficulté majeure lorsqu'on souhaite caractériser finement les coudes au sein de peptides bioactifs.

Dans ce contexte, l'objectif principal de ce travail de thèse est donc d'élargir la base de données de référence pour deux méthodes spécifiques, la RMN et le DC, afin de caractériser finement tous les types de coudes  $\beta$  et  $\gamma$ .

Afin d'identifier des paramètres de RMN discriminants, nous avons mené, dans un premier temps, une étude expérimentale sur une banque de peptides modèles en conformation  $\beta$ -hairpin (2 brins de feuillet  $\beta$  reliés par un coude). Pour cette étude, nous nous sommes limités aux coudes  $\beta$  de type I, I', II et II', et aux coudes  $\gamma$  de type classique et inverse. Ce choix a été effectué pour, d'une part, confirmer et éventuellement élargir le nombre de paramètres de RMN déjà disponibles pour certains de ces types de coudes (coudes  $\beta$  de type I, I', II et II') et d'autre part, tenter d'établir des paramètres de RMN discriminants pour des coudes pour lesquels aucune donnée expérimentale n'est présente dans la littérature. Pour tous ces types de coudes sélectionnés, un peptide modèle a été choisi. Les peptides modèles ont été étudiés par DC et RMN.

Dans un deuxième temps, nous avons orienté la thèse vers un côté théorique. Une étude de modélisation moléculaire en utilisant les méthodes de dynamique et de mécanique moléculaire

a été réalisée pour extraire des paramètres RMN discriminants pour les différents motifs de coudes. Pour cette partie de ce projet, nous avons élargi l'analyse à tous les types de coudes  $\beta$  (I, I', II, II', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb et VIII) et  $\gamma$  (classique et inverse).

Enfin, nous avons décidé de prédire, en couplant la dynamique moléculaire et la théorie de la fonctionnelle de la densité dépendante du temps (TDDFT), des spectres de référence de DC pour les coudes  $\beta$  de type I, I', II et II' pour lesquels des travaux contradictoires existent dans la littérature.

La suite de ce manuscrit de thèse se divisera en quatre parties. Dans le chapitre 2, je présenterai les techniques employées dans ces travaux de thèse, puis, dans le troisième chapitre, les résultats principaux de la partie expérimentale. Le quatrième chapitre sera dédié à la présentation des paramètres de RMN discriminants pour les coudes de nature  $\beta$  et  $\gamma$  identifiés par modélisation moléculaire ainsi qu'à leur discussion. Ensuite, dans le cinquième chapitre, sera exposée la simulation de spectres de DC pour les coudes  $\beta$  de type I, I', II et II'. Les résultats de cette étude seront présentés à travers l'article, sous forme d'épreuve, accepté au journal *Physical Chemistry Chemical Physics*. Enfin, les conclusions et les perspectives sur ce travail de thèse achèveront ce manuscrit.



## **Chapitre 2**

### **Techniques utilisées pour l'analyse structurale de peptides et de protéines en solution**



## ***Chapitre 2 : Techniques utilisées pour l'analyse structurale de peptides et protéines en solution***

Ce chapitre expose comment le DC et la RMN sont utilisés en biologie structurale pour la détermination de la structure tridimensionnelle de protéines et peptides en solution. Nous présenterons d'abord le dichroïsme circulaire et son application pour l'étude des protéines et peptides, en se focalisant en détail sur les problématiques de l'identification des coudes à travers ce type de spectroscopie. Dans un deuxième temps, la RMN des protéines et peptides en solution sera présentée. Comme pour le DC, nous expliquerons comment les coudes ont été caractérisés par RMN. Enfin, la troisième partie de ce chapitre portera sur la méthode de modélisation moléculaire basée sur la mécanique moléculaire. Nous présenterons le principe, le protocole de calcul ainsi que la procédure de validation des structures obtenues par les calculs.

### **I. Le dichroïsme circulaire**

La spectroscopie de dichroïsme circulaire est une technique extrêmement sensible pour analyser la structure secondaire des peptides ou des protéines en solution.<sup>88</sup> En raison de sa facilité de mise en œuvre, de sa rapidité et des faibles concentrations d'échantillons utilisées, elle est considérée comme un puissant outil comparatif permettant de détecter les variations de conformations entre plusieurs composés.

Le principe général du DC est décrit en Annexe I. Dans cette partie du manuscrit, nous allons détailler l'utilisation de cette technique pour caractériser la structure secondaire des polypeptides.

#### **I.1 Dichroïsme circulaire des protéines et peptides**

Les chaînes polypeptidiques sont des molécules biologiques qui sont formées par des éléments chimiques optiquement actifs (liaisons peptidiques et cycles aromatiques des acides aminés) et qui peuvent absorber de façon inégale la lumière polarisée circulairement à droite de la lumière polarisée circulairement à gauche. En effet, les liaisons peptidiques, qui sont des éléments intrinsèquement asymétriques, absorbent dans la région UV du spectre (240 nm et au-dessous). L'absorption électronique observée pour la liaison peptidique consiste en une transition faible (et large)  $n \rightarrow \pi^*$  centrée autour de 220 nm (paire d'électrons d'azote), et une transition intense  $\pi \rightarrow \pi^*$  à environ 190 nm (double liaison).<sup>89</sup> L'intensité et l'énergie de ces transitions dépendent des angles de liaison peptidique ( $\Phi$  et  $\Psi$ ), et donc de la structure secondaire de la chaîne peptidique.<sup>90</sup> Les transitions optiques sont affectées lorsque les chromophores amides du

squelette polypeptidique sont alignés dans des réseaux. Par conséquent, les différents éléments de structure secondaire produisent un spectre de DC caractéristique.<sup>61</sup>

Pour déterminer la structure secondaire d'un peptide ou d'une protéine en solution, une comparaison du spectre de DC obtenu avec des spectres de référence des structures en conformation hélicoïdale, en feuillet  $\beta$  ou en conformation aléatoire doit être réalisée.<sup>63,91</sup> Elle permet alors de caractériser la présence de la structure secondaire qui est majoritaire dans la structure du peptide ou d'une protéine en solution.

### **Spectres de DC des conformations aléatoire, hélice et feuillet**

Le spectre de DC de peptides ou d'une protéine non structurés en solution, ou dénaturés, présente un signal négatif et intense entre 190 et 200 nm (Figure 22).<sup>92</sup> Une faible valeur positive d'ellipticité est parfois observable aux alentours de 220 nm.<sup>93</sup> Des valeurs proches de zéro sont souvent observées sur le reste du spectre (220 – 260 nm).

Les polypeptides présentant une structuration en hélice en solution possèdent un spectre de DC composé de trois signaux caractéristiques : deux minima négatifs à 208 et 222 nm et un signal positif vers 190 nm (Figure 23).<sup>94,95</sup> Selon la nature des hélices (hélice  $3_{10}$ , hélice  $\alpha$  ou autres),<sup>96,97</sup> le nombre et le type de résidus impliqués dans cette structure hélicoïdale,<sup>95</sup> la présence d'autres chromophores (cycle aromatiques, ponts disulfure, par exemple) ainsi que l'existence d'un type supplémentaire de structure secondaire (comme les coudes), l'intensité de ces pics peut changer.<sup>89</sup>

Dans le cas d'un peptide ou d'une protéine adoptant une structuration en feuillet  $\beta$ , l'allure du spectre est caractérisée par la présence d'un seul minimum négatif entre 210 et 225 nm et d'un autre signal positif vers 190-200 nm (Figure 23).<sup>63,93</sup> Comme pour une structuration en conformation hélice, l'allure d'une structuration en feuillet  $\beta$  est dépendante du nombre de résidus par brin, de la présence de structures secondaires supplémentaires (notamment des coudes) et du solvant.<sup>33,39</sup>

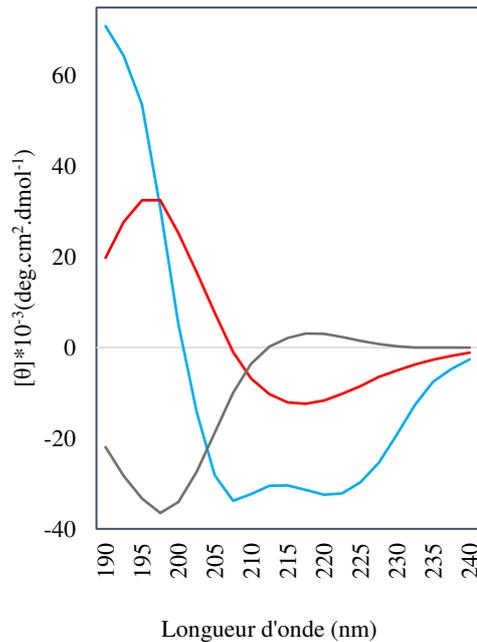


Figure 23 Spectres de référence de dichroïsme circulaire pour les hélices  $\alpha$  (bleu ciel), les feuillets  $\beta$  (rouge) et la conformation aléatoire (gris) (tiré de Reed et Reed, 1997).<sup>63</sup>

### Spectres de DC des coudes

Comme présenté au sein du Chapitre 1, les coudes sont reconnus comme étant des structures ordonnées mais non répétitives et comme la structure secondaire la plus flexible. En effet, chaque angle dièdre  $\Phi$  and  $\Psi$  des résidus  $i+1$  et  $i+2$  possède une gamme de degré de liberté égale à  $\pm 30^\circ$  par rapport aux valeurs des angles canoniques qui les caractérisent (Tableau 9, cf. Chapitre 1).<sup>45</sup> Par ailleurs, les angles des résidus  $i$  et  $i+3$  peuvent correspondre à plusieurs conformations structurales (hélice, feuillet et conformation aléatoire). Ceci donne, donc, aux liaisons peptidiques une orientation variable par rapport aux autres motifs.

Pour toutes ces raisons, la caractérisation à l'aide de la spectroscopie de DC d'une structure de type de coude est beaucoup plus difficile. Dans la littérature, il n'existe pas de spectres de DC de référence pour toutes les natures et tous les types de coudes. Ceci est principalement dû au fait qu'il est difficile de disposer d'un peptide adoptant un type de coude précis et stabilisé en solution.

Compte tenu de la difficulté d'obtenir des spectres expérimentaux de référence fiables pour les motifs de type coudes, l'utilisation de méthodes théoriques peut venir en aide afin de pouvoir les simuler.

D'un point de vue théorique, le premier travail de spectroscopie de DC sur les coudes de nature  $\beta$  fut réalisé par Woody en 1974, qui identifia *via* des calculs semi-empiriques le rôle

fondamental des transitions  $n \rightarrow \pi^*$  et  $\pi \rightarrow \pi^*$  pour les coudes  $\beta$  de type I, II et III.<sup>98</sup> Les calculs ont mis en évidence que tous les types analysés étaient généralement caractérisés par une bande négative  $n \rightarrow \pi^*$  à  $\sim 225$  nm, suivie par une bande positive et une négative  $\pi \rightarrow \pi^*$  à  $\sim 205$  et  $190$  nm, respectivement. Une telle allure de dichroïsme circulaire électronique, incluant les bandes positives et négatives, a été notée comme spectre de DC de classe B (Tableau 12).

Tableau 12 Schéma de classification des spectres DC calculés par Woody.<sup>98</sup> Les classes ont été établies en fonction du nombre et de la position des maxima et des minima.

Classe	Bande	Gamme de longueur d'onde [nm]
A	-	210-220
	+	195-200
	-	<190
B	-	>220
	+	200-210
	-	<190
C	-	200-210
	+	180-195
D	-	>225
	+	210-220
	-	190-200
	+	<190

Un second travail théorique a été réalisé par Sathyanarayana et Applequist en 1986.<sup>99</sup> Il présente les spectres de DC de peptides modèles Ac-X-X-NHMe (X = Ala, Gly) en conformation coude  $\beta$  de type I, II et III. Les spectres de DC obtenus montrent que les coudes  $\beta$  de type I et III présentent deux spectres similaires (Figure 24) avec une bande négative ( $\sim 205$  nm) et une bande positive ( $\sim 190$  nm). En revanche, un spectre de DC présentant un maximum à  $205$  nm et un minimum à  $190$  nm est caractéristique pour le coude  $\beta$  de type II (Figure 24).<sup>99</sup> Ces résultats sont en contradiction avec l'étude de référence réalisée par Woody auparavant.<sup>98</sup>

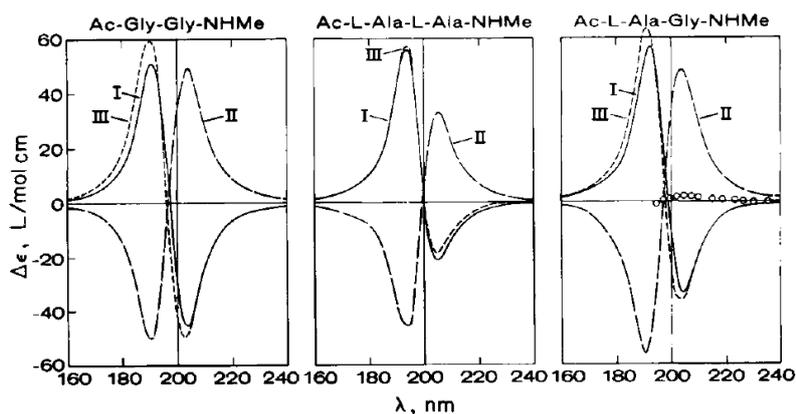


Figure 24 Spectres de DC théoriques des coudes de nature  $\beta$  de type I, II et III (tiré de Sathyanarayana et Applequist, 1986).<sup>99</sup>

Différentes approches ont été utilisées pour obtenir les spectres expérimentaux de référence de DC pour les coudes de nature  $\beta$ . La première consiste à soustraire les contributions pondérées des hélices, des feuillettes et de la conformation aléatoire des spectres de DC de protéines globulaires.<sup>100,101</sup> Cela conduit à un spectre de DC général pour les coudes, qui ne prend pas en compte les différents types possibles. Il reste donc complexe de corrélérer le spectre obtenu à une conformation précise de coude de nature  $\beta$ .

La deuxième approche est basée sur une étude comparative des spectres théoriques et expérimentaux de DC.<sup>102</sup> En utilisant cette méthode, le spectre moyen de référence pour les coudes  $\beta$  a été établi par Brahms et ses collaborateurs.<sup>102</sup> Ce spectre de DC a été obtenu à partir d'un peptide modèle en conformation coude et présente une allure similaire à celle obtenue par Woody.<sup>98</sup>

La troisième approche, la plus utilisée, repose sur des études structurales comparatives de peptides en conformation coude par cristallographie aux rayons X, DC et RMN.<sup>103-106</sup> Dans cette approche, des peptides cycliques ou avec des acides aminés modifiés ont été choisis pour limiter la flexibilité des motifs structuraux. Par exemple, des spectres de DC, composés d'une bande négative à 227-222 nm et d'un positif à 200-205 nm (Figure 25), ont été observés pour des hexapeptides cycliques qui adoptent une géométrie de coude de nature  $\beta$  de type II à l'état solide.<sup>104</sup>

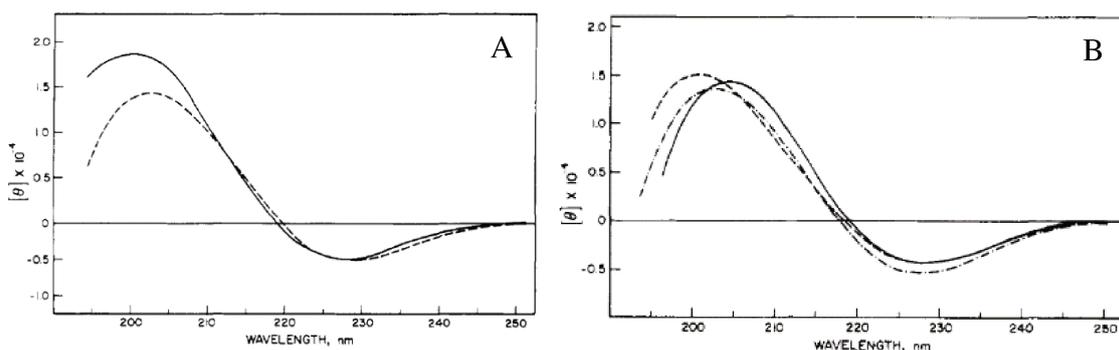


Figure 25 (A) Spectres de DC des peptides cycliques  $(L\text{-Orn-L-Pro-D-Phe})_2$  (---) et  $(L\text{-Orn-L-Pro-D-Cha})_2$  (\_\_\_) ; (B) Spectres de DC des peptides cycliques  $(Gly\text{-L-Pro-D-Phe})_2$  (\_\_\_),  $(Gly\text{-L-Pro-D-Cha})_2$  (----) et  $(L\text{-Ala-L-Pro-D-Phe})_2$  (-.-.-) dans du hexafluoro-2-propanol (tiré de Bush et al., 1978).<sup>104</sup>

De même, un spectre qui présente un minimum inférieur à 188 nm, un maximum à 202-204 nm et un autre minimum autour de 228-230 nm a été observé pour un autre peptide cyclique replié en coude  $\beta$  de type II (Figure 26).<sup>106</sup> Un spectre de DC différent, composé par un maximum proche de 190 nm et deux minima distincts à 206-208 et 218-222 nm, a été établi pour un peptide modèle en conformation coude  $\beta$  de type I (Figure 26).<sup>106</sup>

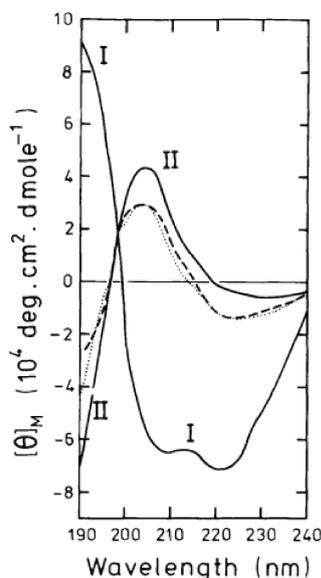


Figure 26 Spectres de DC du peptide cyclique  $L\text{-Ala-D-Ala-Aca}$  (coude  $\beta$  de type II) et du peptide  $L\text{-Ala-D-Ala-Aca}$  (coude  $\beta$  de type I) dans le méthanol (tiré de Bandekar et al., 1982).<sup>106</sup>

D'autres spectres expérimentaux ont été obtenus à partir de deux peptides modèles non cycliques en conformation de coudes  $\beta$  de types I et II (Figure 27).<sup>105</sup> Les peptides repliés en coude  $\beta$  de type I présentent une allure de spectre formée par deux maxima, un à 230 et l'autre

190 nm, et d'un minimum à ~200 nm (Figure 27A), alors que les spectres des peptides en coude  $\beta$  de type II montrent deux bandes positives à environ 230 nm et ~200 nm (Figure 27B).

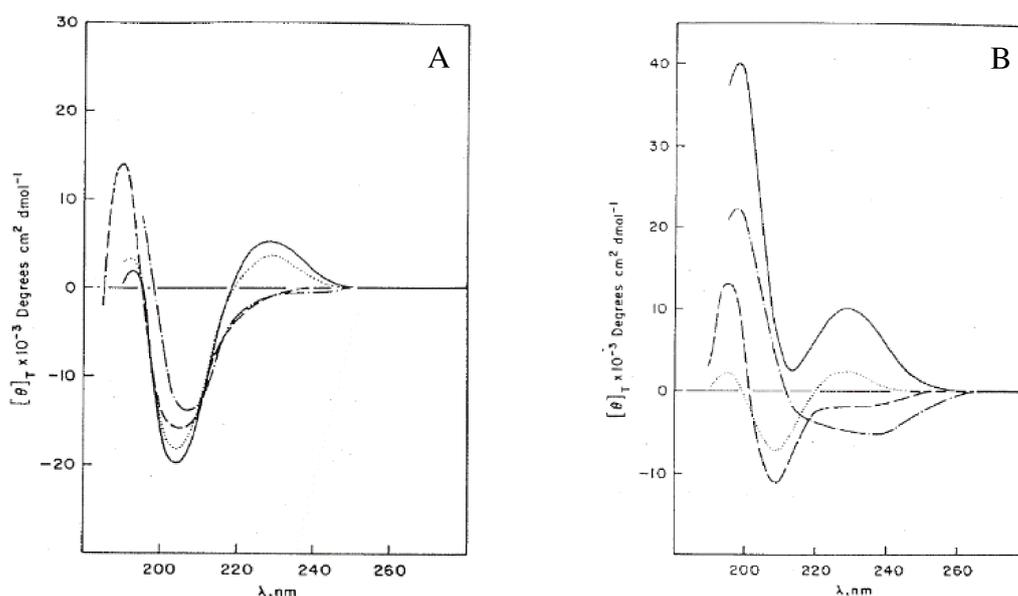


Figure 27 Spectres de DC du peptide Z-Aib-Pro-Aib-Pro-OMe (coude type I) dans le méthanol (—), dans l'eau (...) et dans le cyclohexane (-.-.-) ; (B) Spectres de DC du peptide Piv-Pro-Aib-NHMe (coude de type II) dans le méthanol (—), dans l'eau (...) et dans le cyclohexane (-.-.-) (tiré de Crisma et al., 1984).<sup>105</sup>

Il nécessaire de remarquer que, dans toutes ces études, le type de coude a été caractérisé à l'état solide à partir d'analyses par cristallographie aux rayons X et les auteurs ont supposé que le même repliement persiste également en solution.<sup>107,108</sup> En conséquence, les divergences, c'est-à-dire avoir trouvé trois spectres différents pour le coude de type I et deux spectres différents pour le coude de type II, observées dans ces études expérimentales pourraient être liées à la présence de conformations supplémentaires dans les solvants utilisés pour l'enregistrement des spectres de DC.

Il est intéressant de noter que les ambiguïtés n'ont pas été résolues par les spectres de DC rapportés sur des pseudohexapeptides qui adoptent un mélange conformationnel de coudes  $\beta$  de type I et de type II (déterminés par des expériences de RMN) dans différents solvants.<sup>103,108-</sup>

<sup>111</sup> Dans ces études, une approche différente a été adoptée puisque les peptides ont été caractérisés en solution et non plus à l'état solide par cristallographie aux rayons X.

Dans l'une de ces études, deux allures différentes ont été attribuées au coude  $\beta$  de type I : l'une présentant une bande négative entre 225 et 270 nm et une bande positive à 195 nm (Figure 28), et l'autre caractérisée par un minimum à environ 205 nm et un maximum à  $\sim 195$  nm (Figure 28). Pour le type II, un seul spectre a été identifié avec une bande positive à  $\sim 200$  nm (Figure 28).<sup>111</sup> Tous ces spectres n'ont pas été déduits directement, mais ont été obtenus à travers la décomposition des spectres expérimentaux des pseudohexapeptides analysés.

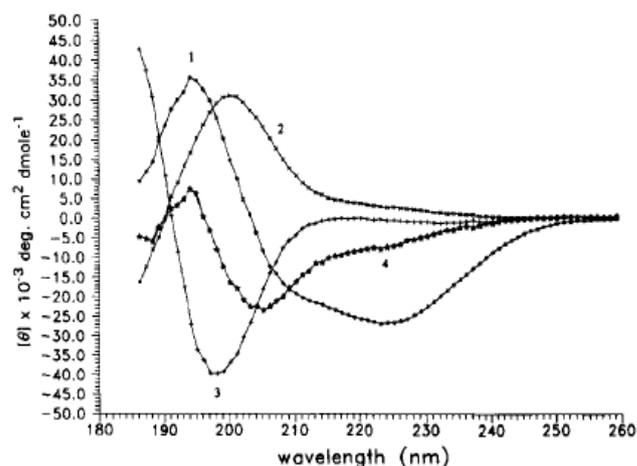


Figure 28 Spectres de DC mesurés pour des pseudohexapeptides différents possédant une structure  $\beta$  de type I (spectre 1 et 4), type II (spectre 2) et une structure désordonnée (spectre 3) (tiré de Perczel et al., 1992).<sup>111</sup>

Pour essayer d'éclaircir toutes ces incohérences sur le motif structural du coude, en 1997, Reed et Reed ont proposé un programme d'ajustement de courbe de DC (PEPFIT) en donnant des allures de référence pour les structures secondaires les plus communes.<sup>63</sup> Pour les coudes  $\beta$  de type I, II et III, les trois allures retenues sont représentées sur la Figure 29A.<sup>63</sup> En utilisant ces trois allures, une courbe générale pour les coudes  $\beta$  a été proposée (Figure 29B). Par ailleurs, pour la première fois, une courbe représentative pour les coudes  $\gamma$  a également été présentée (Figure 29C).<sup>63</sup> Pour établir ces spectres de DC, les auteurs ont choisi de prendre comme référence les spectres obtenus par d'autres travaux de la littérature<sup>101,102,105</sup> et compte tenu des divergences observées dans la littérature, ceci met en doute la fiabilité de ces allures pour les coudes  $\beta$  et  $\gamma$ .

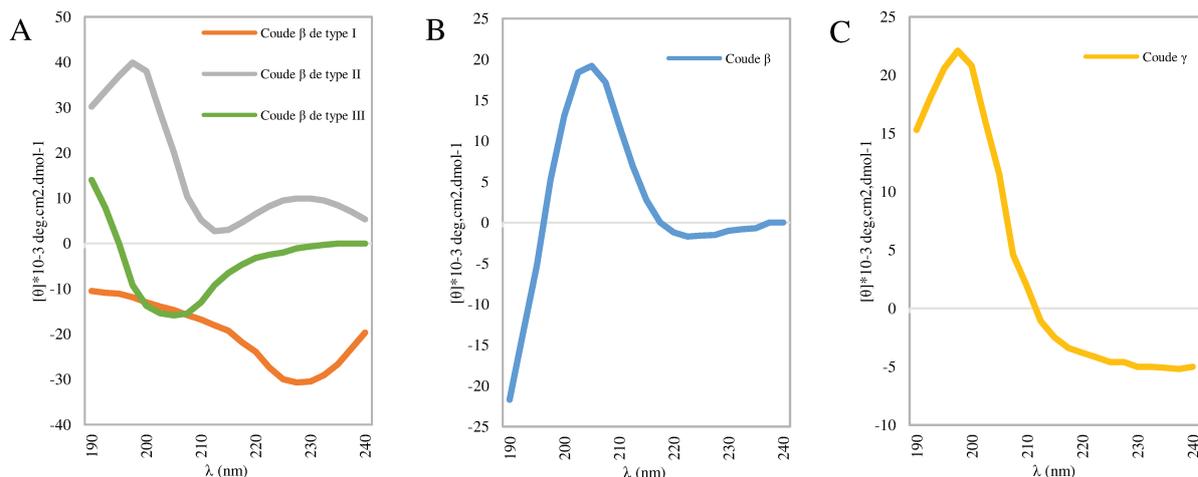


Figure 29 Spectres de DC de référence d'après les données proposés par Reed et Reed (A) Spectres pour trois types de coude  $\beta$  ; (B) Spectre général pour les coudes  $\beta$  ; (C) Spectre général pour les coudes  $\gamma$ .

Pour les structures de type coude, il n'existe donc pas de spectres de référence fiables. Par ailleurs, un peptide peut adopter un mélange de structures secondaires. La structure secondaire en hélice domine généralement sur les autres types de structures sur le spectre de DC. Ainsi, pour un peptide présentant en même temps une structure en hélice et une structure en feuillet, la structure hélicoïdale est la structure majoritaire qui apparaît sur le spectre de DC.<sup>112,113</sup> De la même manière, la présence d'une structure secondaire de type de coude avec une structure de type hélice ou feuillet, peut ne pas donner un spectre de DC comparable aux spectres de référence.

Bien qu'il existe des logiciels de déconvolution (DichroWeb par exemple) qui donnent le pourcentage de structures secondaires dans le spectre de DC, ces logiciels ne peuvent définir ni la localisation, ni la nature du coude et les estimations de pourcentages ne sont pas fiables. Le DC peut donc seulement être considéré comme une technique qualitative car il ne donne pas d'informations sur les régions impliquées dans chaque structure secondaire et ne fournit qu'une estimation globale sur tout le peptide.

## II. La Résonance magnétique nucléaire (RMN) en solution des polypeptides

La RMN est une technique d'analyse utilisée pour l'analyse de molécules, comme les peptides et les protéines. Cette technique permet d'étudier plusieurs noyaux présents au sein des chaînes polypeptidique, comme le  $^1\text{H}$ ,  $^{13}\text{C}$  et  $^{15}\text{N}$ . Cependant, l'abondance naturelle de certains noyaux est trop faible pour pouvoir les utiliser pour l'étude de ces molécules biologiques car cela impliquerait la préparation d'échantillons contenant une grande quantité de composé. Ainsi, le  $^1\text{H}$  est le noyau le plus utilisé car il est le plus abondant (99,9%) par

rapport au  $^{13}\text{C}$  (1,11%) et au  $^{15}\text{N}$  (0,36%). C'est pour cela que la stratégie généralement employée est d'effectuer des analyses de RMN, en étudiant l'hydrogène ( $^1\text{H}$ ), car il demande une faible quantité de produit.

Pour des protéines de petite taille, des expériences classiques de RMN 1D  $^1\text{H}$  et 2D  $^1\text{H}$ - $^1\text{H}$  (TOCSY, COSY et NOESY) sont utilisées pour caractériser leurs structures tridimensionnelles. La base théorique de cette technique ainsi que la méthode d'attribution des spectres sont reportées en Annexe 1. L'étude de cartes bidimensionnelles  $^1\text{H}$ - $^1\text{H}$  permet l'obtention d'un ensemble des paramètres spécifiques, nommés « paramètres structuraux » (déplacements chimiques secondaires, effets NOE, constantes de couplage) qui sont utilisés afin d'avoir une idée initiale sur la structure secondaire. Ces paramètres structuraux vont être présentés dans ce qui suit.

## II.1 Le déplacement chimique secondaire

Le premier paramètre qui donne des informations sur la structure secondaire est le déplacement chimique des protons. Du fait de sa grande dépendance vis-à-vis de l'environnement électronique, le déplacement chimique d'un noyau est lié à plusieurs facteurs structuraux : la géométrie locale des liaisons, l'implication dans une liaison hydrogène, la proximité de cycles aromatiques, mais également à des facteurs expérimentaux tels que le solvant ou le pH de l'échantillon.

Il n'est pas encore possible de calculer *ab initio* le déplacement chimique de molécules aussi grosses que des polypeptides avec précision, il a donc fallu essayer d'établir des corrélations entre déplacements chimiques et structure secondaire.

Par des études statistiques, il a été montré que le déplacement chimique des protons amides  $\text{H}_\text{N}$ , des protons  $\text{H}_\alpha$  ainsi que celui des carbones  $\text{C}_\alpha$  dépendent du type de résidu et de la structure secondaire dans laquelle ils sont engagés.<sup>114,115</sup> En 1991, pour la première fois, Wishart montre que la valeur moyenne du déplacement chimique des  $\text{H}_\alpha$  d'un résidu impliqué dans une structure secondaire s'écarte des valeurs moyennes observées en conformation aléatoire.<sup>114</sup> Selon la nature de la structuration, en hélice ou en feuillet, le déplacement chimique secondaire s'éloigne respectivement de  $\Delta\delta = -0,39$  ou  $+0,37$  ppm par rapport à la conformation aléatoire. Les auteurs ont mis ainsi au point une méthode qui permet d'estimer la nature et la localisation des éléments de la structure secondaire en fonction du déplacement chimique du proton  $\text{H}_\alpha$ , en définissant l'indice de déplacement chimique (CSD pour *chemical shift deviation*).

Celui-ci est calculé en faisant la différence entre le déplacement chimique expérimental ( $\delta_{\text{obs}}$ ) d'un noyau donné et une valeur de référence ( $\delta_{\text{ref}}$ ) pour le même noyau dans sa conformation aléatoire :

$$\text{CSD} = \delta_{\text{obs}} - \delta_{\text{ref}} \quad (2.1)$$

La définition des valeurs de référence applicables aux peptides en solution aqueuse a été effectuée par Wishart, en analysant les déplacements chimiques des différents acides aminés au sein d'un peptide composé de glycines et d'alanines (Ac-Gly-Gly-X-Ala-Gly-Gly-NH<sub>2</sub>, où le résidu noté X est le résidu d'intérêt).<sup>114,116</sup> De la même manière, Wishart et ses collaborateurs ont déterminé les déplacements chimiques de tous les résidus sur un peptide présentant une proline dans la séquence (Ac-Gly-Gly-X-Pro-Gly-Gly-NH<sub>2</sub>) afin de prendre en considération la présence d'une proline dans une séquence polypeptidique. L'enchaînement de glycines et d'alanines permet d'assurer que le peptide soit en conformation aléatoire.<sup>116</sup>

Les valeurs de CSD donnent donc une première information sur le type de structuration et permettent l'identification de trois conformations différentes :

- Hélice si la valeur de CSD d'au moins 4 résidus consécutifs est inférieure à  $-0,1$  ppm ;
- Feuillet si la valeur de CSD d'au moins 3 résidus consécutifs est supérieure à  $0,1$  ppm ;
- Conformation aléatoire si la valeur de CSD est entre  $-0,1$  et  $+0,1$  ppm.

Il est à noter que, pour une structure secondaire de type coude, il n'existe pas de valeurs de CSD caractéristiques comme pour les autres structures secondaires, car ces motifs sont considérés comme des conformations flexibles.

Des corrélations similaires ont été observées entre les déplacements chimiques des C $\alpha$ , C $\beta$  et CO et la structure secondaire.<sup>117</sup> Mais ces noyaux sont moins utilisables que le proton H $\alpha$ , pour des peptides non enrichis, puisque le <sup>13</sup>C existe en faible pourcentage d'abondance naturelle (1%). Des concentrations élevées de peptides ou un marquage isotopique sont alors nécessaires pour enregistrer des expériences 2D hétéronucléaires.

Ce paramètre est utilisé avec les autres paramètres structuraux (effet Overhauser nucléaire et constante de couplage) pour déterminer l'organisation structurale d'un peptide par RMN en solution.

## II.2 Le diagramme NOE

L'effet Overhauser nucléaire (NOE) est une corrélation qui repose sur l'interaction dipolaire entre deux protons proches dans l'espace dont la distance est inférieure ou égale à 5 Å. Cet effet NOE présente une importance considérable pour la reconstitution des structures secondaires des peptides et des protéines. En effet, chaque type de structure secondaire donne naissance à des effets NOE caractéristiques en nature et en intensité. Wagner et ses collaborateurs ont proposé une stratégie qui permet de regrouper l'enchaînement des corrélations NOE caractéristiques et de les représenter dans un diagramme NOE afin de visualiser et d'identifier rapidement la nature des structures secondaires présentes dans les peptides et les protéines (Figure 30).<sup>64,118</sup> Ces corrélations servent de référence pour l'identification des structures secondaires de type hélice et feuillet.

	$\beta, \beta_p$	$\alpha$ -Helix	$3_{10}$ -Helix	Turn I	Turn II	Turn I'	Turn II'	Half-Turn
$d_{\alpha N}(i, i+4)$		—————						
$d_{\alpha\beta}(i, i+3)$		—————	—————					
$d_{\alpha N}(i, i+3)$		—————	—————	—————		—————	—————	
$d_{NN}(i, i+2)$		—————	—————	—————	—————	—————	—————	
$d_{\alpha N}(i, i+2)$			—————	—————	—————	—————	—————	—————
$d_{NN}$	—————	—————	—————	—————	—————	—————	—————	—————
$d_{\alpha N}$	—————				—————	—————	—————	—————
$^3J_{HN-H\alpha}$ (Hz)	9 9 9 9 9 9 1 2 3 4 5 6	4 4 4 4 4 4 4 1 2 3 4 5 6 7	4 4 4 4 4 4 1 2 3 4 5 6	4 9 1 2 3 4	4 5 1 2 3 4	7 5 1 2 3 4	7 9 1 2 3 4	4 9 1 2 3 4

Figure 30 Diagramme des corrélations NOE et des constantes de couplages  $^3J_{HN-H\alpha}$  caractéristiques des différents types de structures secondaires généralement observées au sein des peptides et des protéines (tiré de Wagner et al., 1986).<sup>64</sup> Un trait épais correspond à une forte distance, un trait moyen correspond à une distance intermédiaire et un trait fin correspond à une faible distance.

### Identification des structures secondaires hélices et feuillets

Les structures hélicoïdales sont caractérisées par un enchaînement d'effets NOE de type  $d_{\alpha N}(i, i+3)$ ,  $d_{NN}(i, i+2)$  et  $d_{\alpha\beta}(i, i+3)$ . L'identification d'une hélice de type  $\alpha$  ou de type  $3_{10}$  se fait à l'aide de deux autres corrélations dipolaires NOE spécifiques : l'une est de type  $d_{\alpha N}(i, i+4)$ , observable pour les hélices  $\alpha$ , et l'autre de type  $d_{\alpha N}(i, i+2)$ , présente dans les hélices  $3_{10}$  (Figure 30).<sup>119</sup>

De la même manière, la présence d'effets NOE de type  $d\alpha N (i, i+1)$  de forte intensité, ainsi que de type  $dNN (i, i+1)$  de faible intensité (et même souvent inexistant), permet d'identifier une structuration en feuillet  $\beta$ . Cette structuration est caractérisée également par l'existence de corrélations NOE de longue distance de type  $H_N-H_N$ ,  $H_N-H_\alpha$  et  $H_\alpha-H_\alpha$  entre les résidus impliqués dans la formation du feuillet  $\beta$  (représentés en vert sur la Figure 31).<sup>119</sup>

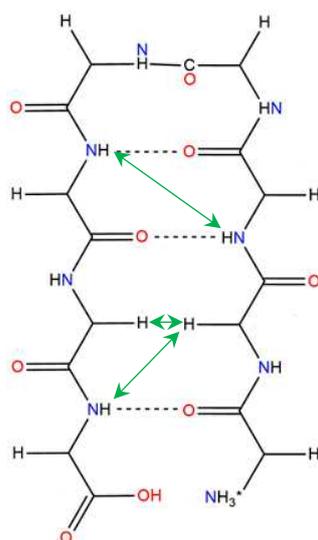


Figure 31 Représentation d'une structure en conformation  $\beta$ -hairpin (deux brins  $\beta$  reliés par un coude) qui montre les effets à longue distance de type  $H_N-H_N$ ,  $H_N-H_\alpha$  et  $H_\alpha-H_\alpha$  (en vert) entre les résidus impliqués dans la formation du feuillet  $\beta$ .

### Identification des structures secondaires de type coude

Si la comparaison des effets NOEs expérimentaux avec le diagramme de référence permet une bonne identification et caractérisation des hélices  $\alpha$  ou  $3_{10}$  et des feuillets  $\beta$ , celles des coudes est beaucoup plus difficile.<sup>64</sup> En effet, ces derniers sont des motifs non répétitifs et impliquant un nombre d'acides aminés limité (trois résidus pour les coudes  $\gamma$  et quatre résidus pour les coudes  $\beta$ ). De ce fait, la quantité de NOE disponibles pour ces structures secondaires est plus limitée. Par ailleurs, dans le diagramme de référence établi par Wagner et ses collaborateurs, il n'existe pas d'effets NOE caractéristiques pour tous les types de coudes. Le travail fondateur sur les coudes  $\beta$  porte sur une analyse théorique des distances  $^1H-^1H$  des types I et II. Cette analyse a été réalisée sur des peptides modèles de polyalanines par Wüthrich en 1984.<sup>119</sup> Cette étude montre que pour les coudes  $\beta$  de type I et de type II, de manière similaire aux structures en conformation hélicoïdale, un réseau dense de courtes distances inter-proton est retrouvé (Tableau 13).<sup>119</sup> Ainsi, des distances moyennes (3,0-3,8 Å) ont été retrouvées au sein de ces

structures secondaires.<sup>119</sup> En outre, cette analyse sur des chaînes peptidiques de polyalanines ne permet de prendre en compte ni tous les types de coude qui ont été identifiés jusqu'à l'heure actuelle (12 types pour les coudes  $\beta$  et 2 types pour les coudes  $\gamma$ ) ni l'influence de la nature des différentes chaînes latérales sur le type de coude rencontré.<sup>64,119,120</sup>

Tableau 13 Distances (en Å) inter-proton relevées dans des peptides modèles de polyalanines en conformation de coudes  $\beta$  de type I et II (tiré de Wüthrich et al., 1984).<sup>106</sup>

Distances	Coude $\beta$ de type I	Coude $\beta$ de type II
<b>dNN (2, 3)</b>	2,6	4,5
<b>dNN (3, 4)</b>	2,4	2,4
<b>dNN (2, 4)</b>	3,8	4,3
<b>d<math>\alpha</math>N (2, 3)</b>	3,4	2,2
<b>d<math>\alpha</math>N (3, 4)</b>	3,2	3,2
<b>d<math>\alpha</math>N (2, 4)</b>	3,6	3,3
<b>d<math>\alpha</math>N (1, 4)</b>	3,1-4,2	3,8-4,7
<b>d<math>\beta</math>N (2, 3)</b>	2,9-4,4 2,9-4,1	3,6-4,6 3,6-4,4
<b>d<math>\beta</math>N (3, 4)</b>	3,6-4,6 3,6-4,4	3,6-4,6 3,6-4,4

La première étude expérimentale par RMN pour l'établissement des distances  $^1\text{H}$ - $^1\text{H}$  caractéristiques pour les coudes  $\beta$  a été réalisée par Wagner et ses collaborateurs en 1986 sur la protéine metallotionine-2 du foie de lapin.<sup>64</sup> Dans cette structure, seuls quatre types de coude de nature  $\beta$  (I, I', II et II') ont été caractérisés. Ces derniers présentent des distances séquentielles d $\alpha$ N ( $i, i+1$ ), dNN ( $i, i+1$ ) courtes et souvent des effets NOE de type d $\alpha$ N ( $i, i+3$ ), d $\alpha$ N ( $i, i+2$ ) et dNN ( $i, i+2$ ), comme présenté à la Figure 30. En revanche, aucune distance d $\alpha\beta$  ( $i, i+3$ ) n'a été observée dans ces structures secondaires, qui, d'un autre côté, serait attendue pour les hélices  $\alpha$ . De la même manière, les effets NOE à longue distance, caractéristiques des feuillets  $\beta$ , pas observés dans les coudes  $\beta$ .<sup>119</sup> Par ailleurs, la distinction entre les quatre types de coude  $\beta$  (type I, I', II, et II') se joue souvent sur un ou deux effets. Il est donc difficile de déterminer la nature d'un coude seulement à partir des données de NOE.<sup>64</sup>

En outre, dans les diagrammes proposés tels que celui de la Figure 30, n'ont été gardées, en réalité, que les données discriminatoires entre les principaux motifs structuraux et non pas les différents types de coudes spécifiquement.

Les données de référence que nous avons à ce jour ne permettent que d'identifier 4 types de coudes  $\beta$  parmi les 12 présents dans la classification établie par de Brevern.<sup>45</sup> De plus, dans la pratique, la discrimination entre les différents coudes ne se fait donc que sur une ou deux distances et le plus souvent *a posteriori* lors de l'analyse de la structure tridimensionnelle. Le manque de paramètres expérimentaux caractéristiques pour toutes les natures et tous les types de coudes constitue donc une limite à la caractérisation de ces structures et conduit parfois à des conclusions erronées pour ce type de motifs.

### Etablissement du diagramme NOE expérimental

Le diagramme NOE peut se construire d'une façon qualitative ou quantitative. Pour l'analyse qualitative, les effets NOE caractéristiques sont reportés selon les intensités apparentes des taches de corrélation et classés en trois catégories : faibles, moyennes et fortes. L'analyse quantitative consiste à calculer la distance entre deux protons. Dans ce cas, l'intensité des pics hors diagonale est convertie en distance selon l'équation 2.1 :

$$\eta_{ij} = \frac{K}{d_{ij}^6} \quad (2.2)$$

où  $i$  et  $j$  sont les protons impliqués,  $K$  est une constante qui dépend du temps de corrélation  $\tau_c$  de la molécule. Une boîte est définie autour de chaque pic de façon à intégrer le minimum de bruit spectral. La calibration des effets NOE se fait en utilisant un pic de corrélation entre deux protons situés à une distance fixe. Les protons géminés  $\beta$  sont souvent utilisés pour une telle calibration ( $d = 1,8 \text{ \AA}$ ).

Les effets NOE sont classés en fonction de valeurs seuils. Les valeurs seuils utilisées sont variables mais des distances inférieures à  $2,8 \text{ \AA}$  sont généralement considérées comme fortes, des distances entre  $2,8$  et  $3,8 \text{ \AA}$  sont considérées comme moyennes, et des distances supérieures à  $3,8 \text{ \AA}$  sont considérées comme faibles.

Néanmoins, la proportionnalité entre la distance calculée et l'intensité intégrée n'est pas tout le temps juste. Le temps de mélange de l'expérience NOESY, c'est-à-dire la période pendant laquelle l'aimantation d'un spin, soumise à un champ de radiofréquence, est transférée à un spin relié à lui-même soit par un couplage scalaire, soit par un couplage dipolaire, a un impact sur l'intensité du pic de corrélation NOE. En effet, un temps de mélange trop bas ne permet pas de visualiser un nombre suffisant d'effets NOE. A l'inverse, un temps de mélange trop élevé aboutit à l'apparition de pics NOE supplémentaires qui sont dus à la présence de diffusion de

spins. Cette dernière est un effet indirect où l'aimantation entre deux protons passe à travers celle d'un troisième proton. Il est donc intéressant d'enregistrer différentes expériences NOESY à des temps de mélange croissants (*build-up*) pour pouvoir optimiser le temps de mélange.

De même, lorsque plusieurs attributions sont possibles pour un même pic, ce dernier est considéré comme « ambigu ». Dans ce cas, les pics ambigus sont différenciés sur le diagramme NOE.

### II.3 La constante de couplage $^3J_{\text{HN-H}\alpha}$

L'interaction entre les noyaux à travers les liaisons chimiques correspond au couplage scalaire. Ce couplage est mesuré à l'aide d'un paramètre nommé la constante de couplage. Comme pour le CSD, l'analyse des constantes de couplage  $^3J_{\text{HN-H}\alpha}$  sur plusieurs résidus apporte des informations sur la nature de la structure secondaire du peptide ou de la protéine et permet de confirmer la présence d'une structure hélicoïdale ou en brin  $\beta$ . Dans les chaînes peptidiques, l'une des constantes de couplage les plus intéressantes à regarder est celle à 3 liaisons entre le proton amide  $H_N$  et le proton  $H_\alpha$  du squelette peptidique, dénommée «  $^3J_{\text{HN-H}\alpha}$  ». Ce couplage est directement lié à la valeur de l'angle dièdre  $\Phi$ , par l'intermédiaire de l'angle  $\theta$  (qui correspond à la valeur absolue de l'angle  $\Phi$  auquel il faut soustraire  $60^\circ$ , selon l'équation de Karplus<sup>121</sup> (2.2) :

$$^3J_{\text{HN-H}\alpha} = A \cos^2\theta + B \cos \theta + C \quad (2.3)$$

Avec  $\theta = |\Phi - 60^\circ|$

Étant A, B, C trois coefficients

$$^3J_{\text{HN-H}\alpha} = 6,4 \cos^2 |\Phi - 60^\circ| - 1,4 \cos |\Phi - 60^\circ| + 1,9 \quad (2.4)$$

La résolution de l'équation de Karplus à travers une série de coefficients (dans l'équation 2.4, la série de coefficients proposée par Pardi et ses collaborateurs<sup>122</sup> est présentée) indique la présence de quatre possibilités d'angle  $\Phi$  pour les valeurs de constante de couplage inférieures à 6 Hz (Figure 32). Parmi ces quatre valeurs, la valeur d'angle  $\Phi = -57^\circ$  est caractéristique d'une structure hélicoïdale. De même, les valeurs de constante de couplage comprises entre 8 et 10 Hz donnent deux valeurs d'angles  $\Phi$  de l'ordre de  $-120^\circ$  qui sont caractéristiques d'un feuillet  $\beta$  (Figure 32). Il est généralement admis que plusieurs résidus ayant des valeurs  $^3J_{\text{HN-H}\alpha}$  inférieures à 6 Hz sont en conformation hélicoïdale alors qu'une succession de résidus ayant des valeurs de  $^3J_{\text{HN-H}\alpha}$  supérieures à 9 Hz sont plutôt engagés dans une structure en feuillet.

Des valeurs intermédiaires de  $^3J_{\text{HN-H}\alpha}$  ( $6 < ^3J_{\text{HN-H}\alpha} < 8$  Hz) ne permettent pas de définir de structuration particulière. En effet, ces valeurs peuvent être le signe d'une structuration étendue ou d'une structuration en hélice. De même, la caractérisation d'une structuration en coude en utilisant les valeurs des constantes de couplage est difficile à interpréter pour plusieurs raisons : (i) seules deux valeurs de constantes de couplage servent pour caractériser tous les types de coudes  $\beta$  et une seule pour le coude  $\gamma$ , (ii) la similarité des valeurs caractéristiques de certains types de coudes (coude  $\beta$  de type I et VIa1 par exemple ou coude  $\gamma$  classique et inverse) ne permet pas toujours de trouver le type de coude de façon univoque et (iii) la présence d'une proline dans une structuration en coude ne permet d'avoir accès qu'à une seule constante de couplage caractéristique du coude.

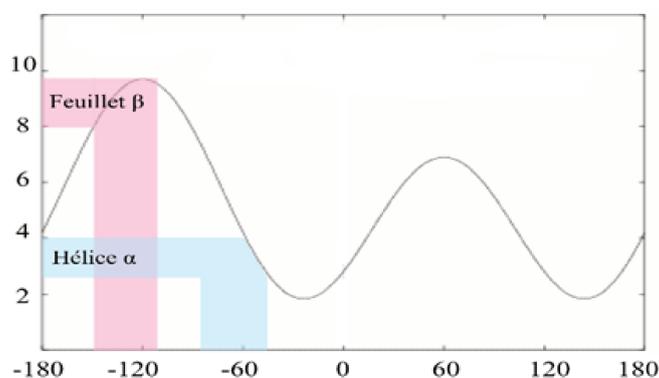


Figure 32 Représentation graphique de la résolution de l'équation de Karplus avec les coefficients de Pardi (adapté de Wüthrich et al., 1986).<sup>107</sup>

Différentes méthodes permettent de mesurer expérimentalement ces constantes de couplages. Dans le cas de peptides de faible poids moléculaire, les valeurs de  $^3J_{\text{HN-H}\alpha}$  peuvent souvent être mesurées directement à partir des spectres 1D. Malheureusement, la présence de superpositions des pics peut rendre la mesure sur le spectre 1D plus difficile. Pour résoudre ce problème, la mesure peut être effectuée sur des expériences à deux dimensions, soit sur un spectre COSY-DQF (*Double Quantum Filtered COSY*),<sup>123</sup> soit sur un spectre 2D TOCSY  $^1\text{H}$ - $^1\text{H}$  à très haute résolution.<sup>124</sup>

En conclusion, les paramètres structuraux permettent d'obtenir des informations sur la structure secondaire des peptides et protéines. Les déplacements chimiques secondaires suggèrent les structurations au sein de ces molécules ; les corrélations NOE localisent la structuration au sein de la séquence, confirment ou infirment les informations des CSD et permettent d'obtenir plus

d'informations sur la structuration. Ce paramètre structural est essentiel pour la détermination de la structuration 2D du peptide ou de la protéine. Les constantes de couplage  $^3J_{\text{HN-H}\alpha}$  viennent en soutien du diagramme NOE. Ce paramètre ne peut être utilisé seul pour la détermination de la structuration du fait de la similarité des valeurs caractéristiques des structuration ainsi que des erreurs de mesure.

Cependant, ces paramètres ne sont pas toujours suffisants pour définir avec précision la nature de la structure secondaire (type d'hélice ou de coude) et/ou délimiter exactement les différents éléments de structure secondaire. De ce fait, il est important de coupler l'ensemble des paramètres structuraux pour déterminer exactement les différentes structures secondaires présentes dans un peptide ou une protéine et réaliser ensuite la modélisation moléculaire qui permet, par l'obtention de la structure tridimensionnelle du peptide ou de la protéine, de confirmer et préciser la structuration secondaire.

### **III. Modélisation moléculaire sous contraintes**

La RMN est reconnue comme étant une technique indirecte pour la détermination des structures tridimensionnelles puisqu'elle ne permet pas d'avoir directement la représentation dans l'espace de la molécule. Elle permet cependant de recueillir un ensemble de paramètres structuraux qui peuvent ensuite être utilisés sous forme de contraintes pour calculer la structure tridimensionnelle. Ces calculs sont l'objet de la modélisation moléculaire. La modélisation moléculaire sous contraintes est ainsi une méthode qui permet de prendre en compte ces contraintes expérimentales mais aussi des contraintes théoriques afin de pouvoir calculer la structure tridimensionnelle de la molécule étudiée. Ces contraintes sont intégrées à l'ensemble des interactions atomiques qui sont prises en compte dans les calculs. Cet ensemble est appelé « champ de forces ». Cette dernière doit présenter un minimum d'énergie et répondre à certains critères de validation.

A ce jour, la méthode de calcul basée sur le protocole de recuit simulé est la méthode la plus adaptée pour explorer les conformations des peptides et des protéines. C'est une méthode utilisant la dynamique et la mécanique moléculaires, qui consiste à augmenter la température afin de franchir les barrières énergétiques. Après cette étape, une étape de refroidissement est appliquée pour stabiliser les structures.

La notion de champ de forces, le recueil des contraintes expérimentales nécessaires pour calculer les structures ainsi que les critères d'évaluation de la qualité des structures seront développés dans les parties qui suivent.

### III.1 Le champ de forces

La structure tridimensionnelle d'une molécule résulte des interactions intramoléculaires et des interactions avec le solvant. La modélisation moléculaire consiste donc à déterminer les minima de l'énergie globale d'interaction qui est en général réduite aux interactions intramoléculaires. Les simulations de mécanique et de dynamique moléculaire s'appuient sur une représentation simplifiée de la surface d'énergie appelée « champ de forces ». Selon ce principe, pour chaque état conformationnel du peptide ou de la protéine, une énergie potentielle empirique est définie grâce au champ de forces. Cette énergie potentielle prend en compte des interactions attractives et répulsives entre les atomes. La prise en compte de contraintes expérimentales dans les protocoles de modélisation moléculaire permet de restreindre l'espace conformationnel pour le peptide ou la protéine et de favoriser des configurations qui sont en accord avec les données expérimentales. Pour intégrer les données expérimentales, le champ de forces utilisé est constitué d'un terme d'énergie potentielle empirique  $E_{\text{empiriques}}$  représentant les contraintes physiques internes à la molécule entre atomes liés ou non et d'un terme énergétique supplémentaire  $E_{\text{contraintes}}$  qui englobe toutes les contraintes expérimentales :

$$E = E_{\text{empirique}} + E_{\text{contraintes}} \quad (2.5)$$

### III.2 La fonction d'énergie $E_{\text{empirique}}$

L'énergie potentielle empirique d'une molécule comprend l'ensemble des énergies définissant sa géométrie spatiale. Elle est alors la somme de deux termes, l'un représentant les interactions entre atomes liés par des liaisons covalentes  $E_{\text{liés}}$ , et l'autre, les interactions entre atomes non liés  $E_{\text{non liés}}$ , selon l'équation décrite ci-dessous :

$$E_{\text{empirique}} = E_{\text{liés}} + E_{\text{non liés}} \quad (2.6)$$

L'énergie entre les atomes liés ( $E_{\text{liés}}$ ) est définie comme :  $E_{\text{liaisons}} + E_{\text{angles}} + E_{\text{dièdres}} + E_{\text{impropres}}$

### Energie d'élongation des liaisons

Le potentiel de liaisons représente l'énergie mise en jeu lors des variations de la longueur  $r$  (Figure 33) des liaisons par rapport à une longueur idéale d'équilibre  $r_0$ . Ces liaisons sont considérées comme des oscillateurs harmoniques (« ressorts ») et l'énergie est décrite comme suit :

$$E_{liaisons} = \sum K_r (r - r_0)^2$$

où  $K_r$  est la constante de force de la liaison. (2.7)



Figure 33 Représentation schématique du terme d'énergie de liaison, à gauche. L'équation de cette énergie avec,  $K_r$  constante de force de la liaison,  $r$  la distance entre les deux atomes et  $r_0$  la distance idéale d'équilibre, à droite

### Energie de déformation des angles

L'énergie d'angles prend en compte les fluctuations de l'angle  $\theta$  formé par deux liaisons contiguës (Figure 34), en induisant une énergie de déformation angulaire qui est estimée par un potentiel harmonique :

$$E_{angles} = \sum K_\theta (\theta - \theta_0)^2$$
 (2.8)

où  $\theta_0$  est la valeur de l'angle à l'équilibre et  $K_\theta$  est la constante de force.

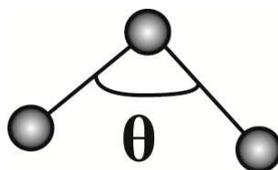


Figure 34 Représentation schématique du terme d'énergie d'angles, à gauche. L'équation de cette énergie avec  $K_\theta$  la constante de force de l'angle,  $\theta$  l'angle formé entre les deux atomes et  $\theta_0$  la valeur de l'angle à l'équilibre, à droite.

### Energie de torsion des angles dièdres

L'Énergie des angles dièdres reflète la déformation des angles dièdres  $\Phi$  formés par quatre atomes (Figure 35). La contribution énergétique lors de la rotation de la molécule autour d'une liaison se décrit selon l'équation suivante :

$$E_{dièdres} = \sum K_{\phi} (1 + \cos n\phi) \quad (2.9)$$

où  $K_{\phi}$  est une constante de force et  $n$  la périodicité de la rotation.

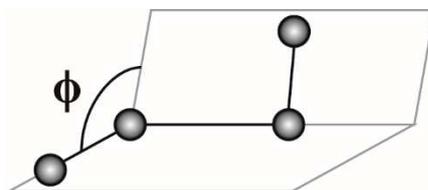


Figure 35 Représentation schématique du terme d'énergie des angles dièdres  $\phi$ , à gauche. L'équation de cette énergie avec,  $K_{\phi}$  la constante de force et  $n$  la périodicité de la rotation, à droite.

### Energie des angles impropres

L'énergie des angles impropres représente la fluctuation de l'angle impropre  $\omega$ . Cet angle sert à maintenir la chiralité et la planéité d'une liaison. L'angle  $\omega$  est constitué de trois atomes, liés à un quatrième atome entre deux plans (Figure 36).

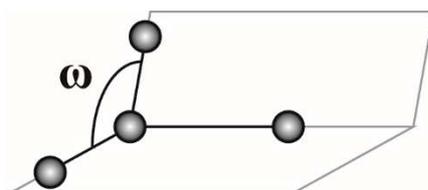


Figure 36 Représentation schématique du terme d'énergie des angles impropres  $\omega$ .

$$E_{impropres} = \sum K_{\omega} (\omega - \omega_0)^2 \quad (2.10)$$

L'énergie entre atomes non liés ( $E_{\text{non liés}}$ ) sont définies comme suit :  $E_{\text{Van der Waals}} + E_{\text{électrostatique}}$

### Potentiel de Van der Waals

Le potentiel de Van der Waals ( $E_{vdw}$ ) traduit l'attraction et la répulsion entre deux atomes séparés par une distance  $r_{ij}$  et constituant un dipôle (Figure 37). Il est décrit par un potentiel de Lennard-Jones avec une partie répulsive en  $1/r^{12}$  et une partie attractive en  $1/r^6$  :

$$E_{vdw} = \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (2.11)$$

où  $A_{ij}$  et  $B_{ij}$  sont des paramètres dépendant de la nature des atomes  $i$  et  $j$ .

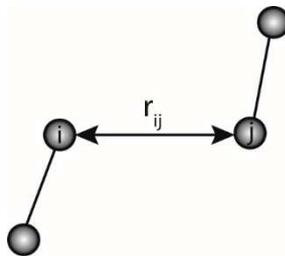


Figure 37 Représentation schématique de l'interaction de Van der Waals entre deux atomes non liés,  $i$  et  $j$ .

### Potentiel électrostatique

L'interaction électrostatique ( $E_{elec}$ ) entre deux charges partielles  $q_i$  et  $q_j$  distantes de  $r_{ij}$  est définie par un potentiel de type Coulombien :

$$E_{elec} = \sum_{i,j} \frac{1}{4\pi\epsilon_r} \frac{q_i q_j}{\epsilon_0 r_{ij}} \quad (2.12)$$

où  $q_i$  et  $q_j$  sont les charges partielles des atomes  $i$  et  $j$ ,  $r_{ij}$  est la distance entre les atomes,  $\epsilon_0$  représente la permittivité du vide et  $\epsilon_r$  celle du milieu.

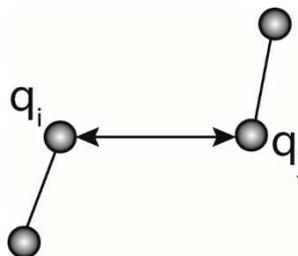


Figure 38 Représentation schématique de l'interaction électrostatique entre deux charges  $q_i$  et  $q_j$ .

Selon la description des termes de la fonction  $E_{\text{non liés}}$ , des interactions de type électrostatique et de Van der Waals peuvent en théorie être observées entre deux atomes très éloignés en distance. La prise en compte de l'ensemble des combinaisons possibles entre paires d'atomes d'une macromolécule impliquerait un temps de calcul considérable. Pour réduire la durée de ces calculs, il est possible de limiter la portée des interactions en définissant un seuil de coupure (en anglais, cut-off). Selon ce principe, seules les interactions entre atomes séparés par une distance inférieure à ce seuil seront prises en compte par le champ de forces. Aussi, pour éviter des discontinuités d'énergie induite par l'introduction de ces seuils, la fonction d'énergie  $E_{\text{non liés}}$  est multipliée par une fonction d'amortissement. Ce type de fonction permet de faire tendre progressivement les énergies électrostatiques et de Van der Waals vers zéro, et donc d'éviter des coupures brusques pouvant générer des erreurs de calcul.

### III.3 Les termes d'énergies de contraintes

Les contraintes expérimentales sont introduites au sein du protocole de recuit simulé et elles sont associées à des termes énergétiques supplémentaires dans le champ de forces. Ces termes traduisent le degré de prise en compte de ces données. Le non-respect d'une contrainte aboutit à une pénalité énergétique, appelée violation, qui informe de l'écart entre une valeur de distance ou d'angle dièdre dans une structure calculée et un intervalle de valeurs autorisées.

Dans le cadre de notre étude, nous avons utilisé deux types de contraintes, les contraintes de distances et les contraintes angulaires. Dès lors, le terme énergie de contraintes peut s'écrire sous la forme suivante :

$$E_{\text{cont}} = E_{\text{dist}} + E_{\text{CDIH}} \quad (2.13)$$

#### L'énergie des contraintes de distance

L'énergie des contraintes de distance ( $E_{\text{dist}}$ ) peut être décrite par différentes fonctions potentielles. Dans le cadre de l'étude de modélisation moléculaire sous contraintes, la fonction potentielle carrée a été utilisée :

$$E_{\text{dist}} = K_{\text{dist}} \cdot \delta^n \quad \begin{array}{l} \text{Avec } \delta = r_{\text{min}} - d \text{ pour } d < r_{\text{min}} \\ \delta = d - r_{\text{max}} \text{ pour } d > r_{\text{max}} \end{array} \quad (2.14)$$

où  $r_{\text{min}}$  et  $r_{\text{max}}$  représentent les limites de l'intervalle de distances permises pour une contrainte donnée,  $d$  est la distance correspondante dans le modèle,  $K_{\text{dist}}$  est la constante de force, et  $n$  est généralement égal à 2. Le terme  $E_{\text{dist}}$  est nul lorsque la valeur  $\delta$  de distance est encadrée par les valeurs  $r_{\text{min}}$  et  $r_{\text{max}}$ .

## L'énergie des contraintes angulaires

Le terme énergétique  $E_{CDIH}$  est lié aux contraintes angulaires  $\Phi$  (ou  $\Psi$ ) et il est représenté par une fonction potentielle carrée comparable à celle définie ci-dessus et de la forme suivante :

$$E_{CDIH} = K_{CDIH} (\Phi - \Phi_x)^2 \quad (2.15)$$

avec  $\Phi_x$  la limite minimum ou maximum de l'intervalle de valeurs autorisées pour une contrainte donnée,  $\Phi$  la valeur d'angle dans le modèle, et  $K_{CDIH}$  la constante de force associée.

### III.4 Calculs des structures

L'introduction des contraintes mentionnées en point III.3 de ce chapitre dans un protocole de recuit simulé permet de calculer la structure tridimensionnelle des peptides ou des protéines.

Le protocole de recuit simulé permet d'atteindre les structures les plus stables avec un minimum d'énergie. Le protocole de recuit est formé essentiellement de quatre étapes :

1. Une dynamique moléculaire à haute température : Pendant cette étape, le système est chauffé à haute température ( $T > 10\ 000\ \text{K}$ ) par rotation des angles de torsion. L'énergie cinétique est ainsi augmentée pour dépasser les barrières énergétiques. Les contraintes expérimentales sont intégrées sous forme de termes énergétiques supplémentaires au champ de forces. Dans celui-ci, le terme  $E_{elec}$  est supprimé et le terme  $E_{vdw}$  est restreint à la contribution répulsive. Par ailleurs, la constante de forces de ce terme est faible ( $E_{DIST}$  pour les contraintes de distance et  $E_{CDIH}$  pour les contraintes angulaires).
2. Une première étape de refroidissement : Le système est refroidi progressivement en diminuant la température par palier afin de descendre dans un puits de potentiel. Le potentiel de contraintes expérimentales augmente à cette étape ainsi que le potentiel de répulsion dans le but de restreindre la pénétration des atomes.
3. Une deuxième étape de refroidissement : Cette étape est facultative pour l'étude des peptides. Nous l'avons cependant conservée dans notre étude. Pendant cette étape, le potentiel de répulsion va augmenter afin de stabiliser la conformation du système alors que le potentiel de contraintes garde le même ordre de grandeur que l'étape précédente.
4. Une minimisation finale : Au sein du protocole de recuit simulé une minimisation est intégrée en utilisant l'algorithme de Powell. Cela consiste à incorporer tous les termes d'énergie ( $E_{liaisons}$ ,  $E_{angle\ de\ valence}$ ,  $E_{angle\ dièdres}$ ,  $E_{angle\ impropres}$ ,  $E_{vdw}$ ,  $E_{elec}$ ,  $E_{DIST}$ ,  $E_{angle\ \Phi}$ ) et permet aux structures calculées d'atteindre le minimum d'énergie.

Une fois les structures tridimensionnelles générées, elles seront par la suite analysées et validées selon plusieurs paramètres.

### **III.5 Validations des structures**

L'évaluation de la qualité des structures tridimensionnelles générées se fait selon les critères suivants : (i) la valeur d'énergie et les violations ; (ii) le diagramme de Ramachandran ; (iii) les liaisons hydrogène, (iv) la convergence des structures sur les atomes du squelette peptidique et des chaînes latérales.

#### **III.5.1 La valeur d'énergie et les violations**

Le premier critère à satisfaire après la finalisation des calculs est celui de l'énergie des structures calculées. Celle-ci est en relation avec les paramètres du champ de forces utilisé ainsi que les contraintes introduites lors des calculs. Lorsque les valeurs d'énergie sont faibles, les structures sont considérées comme plus stables. Ces valeurs basses d'énergie se traduisent également par un nombre minimum des violations. Seules les structures possédant des violations de distance inférieures à 0.2 Å et des violations d'angles  $\Phi$  inférieures à 10° sont conservées.

#### **III.5.2 Le diagramme de Ramachandran**

La distribution des angles dièdres  $\Phi$  et  $\Psi$  sur le diagramme de Ramachandran permet de visualiser dans quelle structuration secondaire chaque acide aminé est engagé (Figure 39).<sup>125</sup>

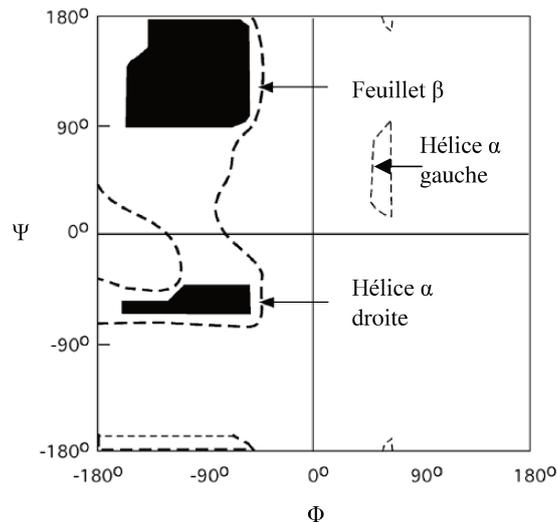


Figure 39 Diagramme de Ramachandran avec les différentes zones définissant les structures secondaires. Il est possible de distinguer trois régions principales : une pour les feuillets  $\beta$ , une pour les hélices  $\alpha$  droites et une pour les hélices  $\alpha$  gauches.<sup>125</sup> Les zones en noir sont les zones énergétiquement permises, les traits pointillés bornent les zones permises mais moins favorables au niveau énergétique.

En 1963, Ramachandran proposa ce diagramme et le construisit en se basant sur la variation de l'encombrement stérique d'un dipeptide de L-Ala en fonction des valeurs des angles dièdres  $\Phi$  et  $\Psi$ .<sup>125</sup> Il définit trois zones favorables énergiquement, spécifiques des trois structures secondaires feuillet  $\beta$ , hélice  $\alpha$  droite et hélice  $\alpha$  gauche. Lorsqu'un résidu converge vers un couple de valeurs d'angles dièdres ( $\Phi$  et  $\Psi$ ) situé dans l'une des zones favorables, il est alors considéré comme engagé dans une structure secondaire bien précise (Figure 39). Au contraire, un couple d'angles dièdres ( $\Phi$  et  $\Psi$ ) en dehors des zones favorables révèle que ce résidu présente une conformation plus flexible.

Les régions de structures secondaires de type brin  $\beta$  antiparallèle ou hélicoïdal ont été bien repérées dans ce diagramme. En revanche, ce dernier ne prend pas bien en considération les structures secondaires de type coude. Ceci est dû au fait que les coudes  $\beta$  n'ont été identifiés qu'en 1968 et les coudes  $\gamma$  qu'en 1972. Les valeurs des angles dièdres pour les coudes de nature  $\gamma$  et  $\beta$  ne sont pas forcément localisées dans les zones énergétiquement favorables. Ramachandran et ses collaborateurs ont donc construit ce diagramme sans pouvoir prendre en considération les liaisons hydrogène ( $i, i+2$ ) pour un coude  $\gamma$  et ( $i, i+3$ ) pour un coude  $\beta$ , qui sont l'une des interactions essentielles pour la stabilisation de ces structures secondaires. Il est donc moins adapté pour l'analyse de ce type de structuration secondaire.

Pour élargir les zones énergétiquement favorables, en 2003, Lovel et ses collaborateurs ont réalisé une étude sur 500 structures de protéines prenant en compte les énergies des liaisons hydrogène ainsi que les énergies de Van der Waals.<sup>49</sup> Cette étude a été menée dans le but de

définir un nouveau diagramme contenant toutes les structures secondaires disponibles. Les auteurs ont alors défini de nouvelles zones de structures correspondant à des régions permises mais énergétiquement moins favorables (Figure 40).

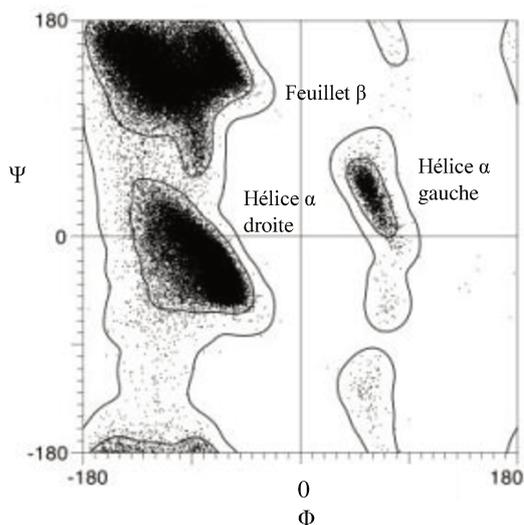
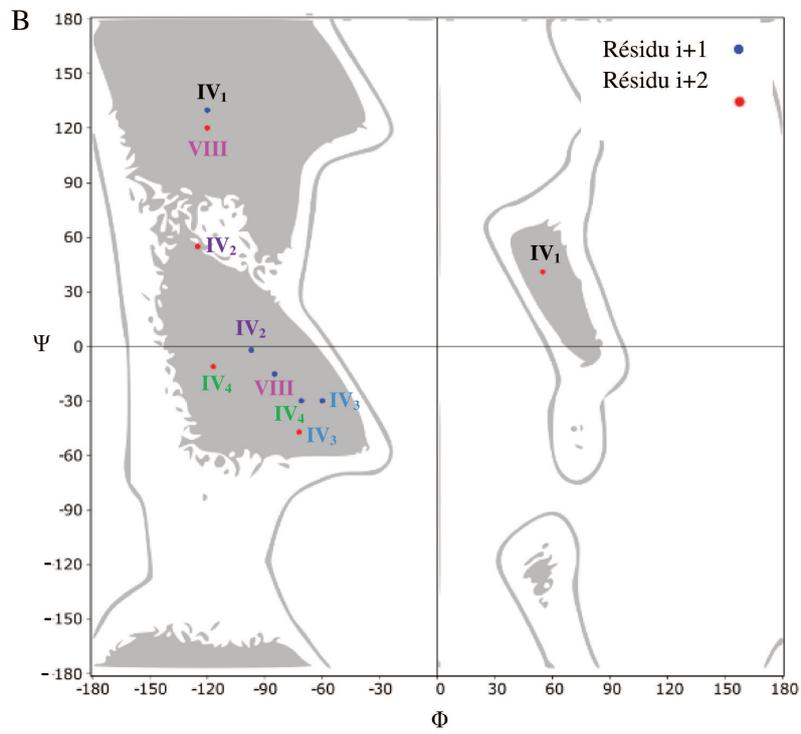
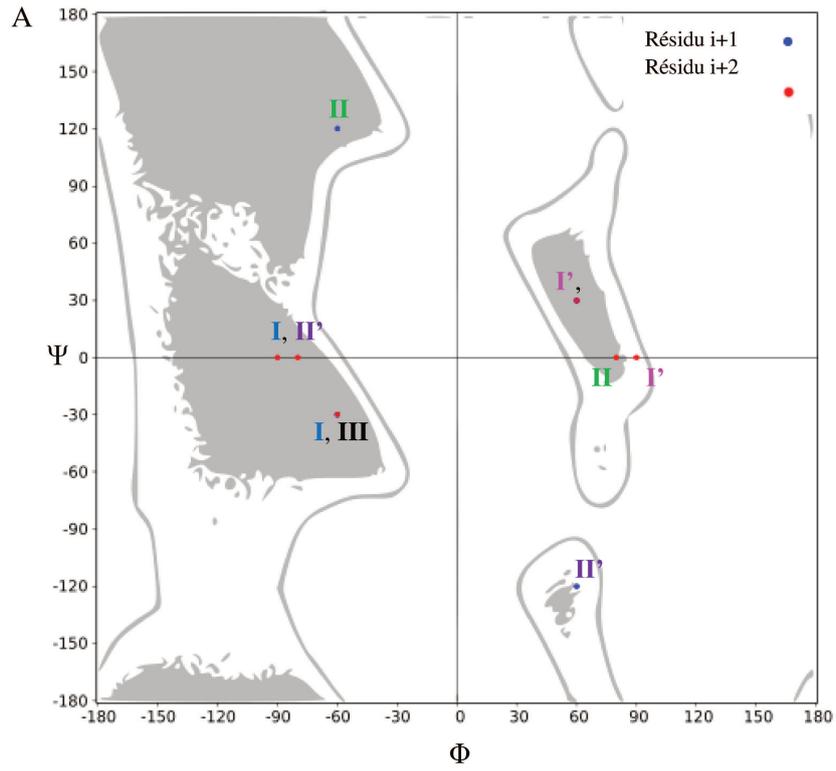


Figure 40 Diagramme de Lovel présentant les différentes zones de structures secondaires après analyse de 500 protéines (tiré de Lovel et al., 2003).<sup>49</sup> Il est possible de distinguer trois régions principales : une pour les feuillets  $\beta$ , une pour les hélices  $\alpha$  droites et une pour les hélices  $\alpha$  gauches. Les zones en noir sont les zones énergétiquement permises, les lignes bornent les zones permises mais moins favorables au niveau énergétique.

Sur la base du diagramme de Lovel et à partir de la classification des différents coudes  $\beta$  et  $\gamma$ , quatre diagrammes peuvent être utilisés pour visualiser la position des résidus centraux pour les coudes de nature  $\beta$  (Figures 41A, 41B et 41C) et celle du résidu central pour les coudes de nature  $\gamma$  (Figure 42).

Nous pouvons constater que, pour la plupart des coudes  $\beta$ , les résidus centraux tombent dans des zones énergétiquement favorables (zones de feuillet  $\beta$ , d'hélice  $\alpha$  droite et d'hélice  $\alpha$  gauche). Seuls quelques résidus (le  $i+1$  du coude  $\beta$  de type II', le  $i+2$  du coude  $\beta$  de type IV<sub>2</sub>) sont localisés dans des zones non permises par Ramachandran (Figures 41A, 41B et 41C).



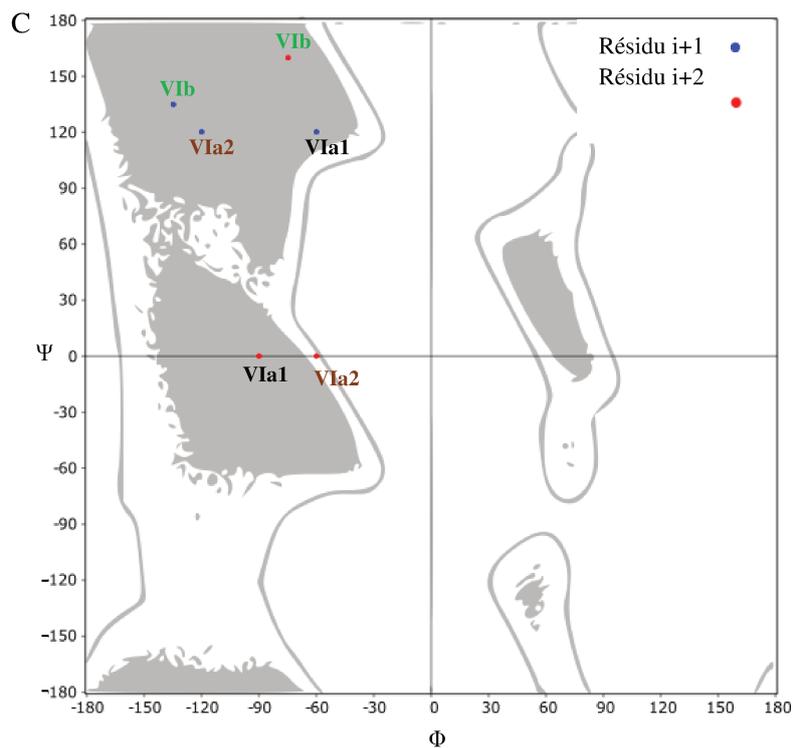


Figure 41 Diagramme de Lovel présentant les différentes zones de structures secondaires de type coude  $\beta$ . (A) Diagramme de Lovel pour les coudes  $\beta$  de type I, I', II, II', III et III' ; (B) Diagramme de Lovel pour les coudes  $\beta$  de type IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub> et VIII ; (C) Diagramme de Lovel pour les coudes  $\beta$  de type VIa1, VIa2 et VIb.

De la même manière, pour les coudes  $\gamma$  de type classique et inverse, le résidu central se trouve dans des zones énergétiquement non favorables (Figure 42).

Ces diagrammes peuvent être utilisés en complément pour analyser la convergence de résidus présentant une structuration de type coude  $\beta$  ou  $\gamma$ .

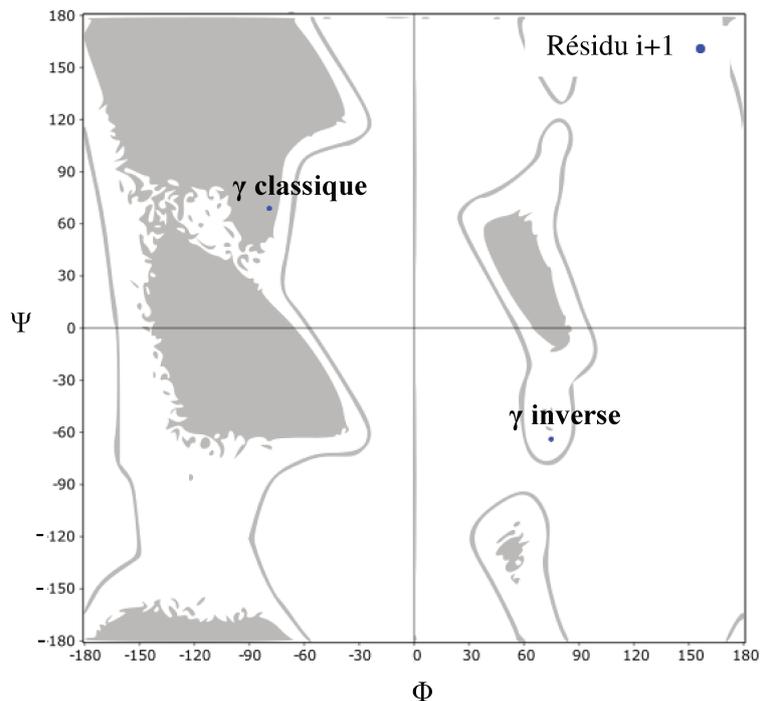


Figure 42 Diagramme de Lovel présentant les différentes zones de structures secondaires de type coude  $\gamma$

### III.5.3 Les liaisons hydrogène

Les liaisons hydrogène sont des forces intramoléculaires dans les structures des peptides et des protéines. Elles jouent un rôle primordial dans la stabilisation des structures secondaires au sein de ces molécules biologiques. Elles sont établies entre un groupement donneur, le proton amide  $H_N$ , et l'oxygène accepteur du groupement carbonyle (Figure 43). Dans l'analyse structurale d'un peptide ou d'une protéine, ces liaisons sont couramment utilisées pour caractériser finement et distinguer les hélices en hélice  $\alpha$  ou hélice  $3_{10}$ , les brins  $\beta$  parallèles ou antiparallèles et les coudes de nature  $\beta$  ou  $\gamma$  (cf. Chapitre 1).

Pour considérer une liaison hydrogène avérée, cette interaction devra assumer une valeur d'angle et de distance définies. La distance entre l'atome d'oxygène du carbonyle et le proton  $H_N$  du résidu donneur ( $d_{H_N \cdots O}$ ) devra être comprise entre 1,8 et 3,0 Å, alors que la valeur d'angle ( $\theta$ ) formé entre l'oxygène du carbonyle, le proton  $H_N$  et le N du résidu donneur doit être comprise entre 120° et 180°.



Figure 43 Représentation de la liaison hydrogène entre le groupement donneur, le proton amide  $H_N$ , et l'oxygène accepteur du groupement carbonyle (En rouge les oxygènes, en blanc les hydrogènes, en bleu les azotes et en gris les carbones). La liaison hydrogène est représentée par des traits pointillés. La liaison hydrogène doit avoir une longueur entre 1,8 et 3,0 Å et un angle  $\theta$  (formé entre l'oxygène du carbonyle, le proton  $H_N$  et le N du résidu donneur) entre 120° et 180°.

### III.5.4 La convergence des structures

Une structure est dite convergente si elle est caractérisée par une bonne superposition sur les atomes du squelette peptidique (N, C et  $C_\alpha$ ) ainsi que sur les atomes des chaînes latérales. Cette convergence est mesurée à travers l'écart quadratique moyen  $\sigma$  (ou RMSD, de l'anglais Root Mean Square Deviation) qui est calculé selon l'équation suivante :

$$\sigma = \sqrt{\sum \frac{|x - \bar{x}|^2}{n}} \quad (2.16)$$

où  $x$  est la position d'un atome dans une structure,  $\bar{x}$  la position d'un atome dans la structure moyenne et  $n$  le nombre de structures.

Les structures convergent lorsque la valeur de RMSD sur les atomes du squelette est inférieure à 1 Å. Pour les chaînes latérales, une valeur de RMSD inférieure à 2 Å indiquent qu'elles sont dans une conformation privilégiée.



## **Chapitre 3**

**Recherche expérimentale de paramètres  
RMN discriminatoires à travers des  
peptides modèles**



### ***Chapitre 3 : Recherche expérimentale de paramètres RMN discriminatoires à travers des peptides modèles***

Comme présenté au sein du chapitre 2 (partie II.2), dans la littérature le nombre de paramètres de référence pour l'identification des coudes par RMN dans les peptides et protéines de faible poids moléculaire est limité. Dans un premier temps, nous avons donc choisi d'utiliser une approche expérimentale afin de caractériser les coudes de nature  $\beta$  de type I, I', II et II' ainsi que les coudes  $\gamma$  de type classique et inverse. Ce choix a été effectué car les paramètres de RMN établis pour les coudes  $\beta$  ne sont pas suffisants pour les distinguer. En outre, aucun paramètre de RMN n'a été déterminé afin d'identifier les coudes  $\gamma$  au sein d'un chaîne polypeptidique. Il existe toutefois peu de peptides natifs qui présentent, dans leur structure, un coude dont la nature est connue de manière précise. Nous avons donc décidé de concevoir des peptides modèles, en se référant à la structure d'édifices macromoléculaires équivalents que sont les protéines. La majorité des structures de protéines disponibles a été résolue par cristallographie aux rayons X, qui permet d'identifier sans ambiguïté la nature et le type du coude par analyse géométrique des angles dièdres  $\Phi$  et  $\Psi$ . Dans la banque de données des protéines (PDB, Protein Data Bank), de nombreuses structures protéiques, déterminées par cristallographie de rayons X, présentent des motifs de type coude de nature ( $\beta$  versus  $\gamma$ ) et de types différents. Comme évoqué dans l'introduction (*cf.* partie II, Chapitre 1), les coudes permettent le plus souvent de relier des éléments de structures secondaires tels que les brins  $\beta$ . Ils sont majoritairement présents dans les brins  $\beta$  qui s'organisent de manière tête-bêche pour former un feuillet antiparallèle. Cette organisation est appelée «  $\beta$ -hairpin ». De manière intéressante, des études ont montré qu'un  $\beta$ -hairpin isolé peut être stabilisé en solution dans les milieux couramment utilisés pour des études structurales par RMN (eau/TFE, méthanol, ...).<sup>127-130</sup> Cet arrangement structural nous est donc apparu comme un support structural de choix pour concevoir le système modèle.

L'étude expérimentale a été divisée en 4 tâches principales :

- Nous avons identifié dans la PDB des structures polypeptidiques, résolues par cristallographie aux rayons X, présentant le motif  $\beta$ -hairpin et les types de coudes  $\beta$  I, I', II, II',  $\gamma$  classique et  $\gamma$  inverse afin de constituer une banque de peptides modèles ;
- Pour chaque type de coude  $\beta$  et  $\gamma$ , un peptide représentatif a été choisi et synthétisé ;
- Les analyses par dichroïsme circulaire, des peptides, dans différents milieux ont été réalisées ;
- Des expériences RMN ont été enregistrées et les spectres ont été analysés.

Dans ce chapitre, nous présenterons d'abord la constitution de la banque de données de peptides modèles. Ensuite, nous montrerons les analyses de DC qui vont permettre de contrôler leur solubilité et vérifier si les peptides choisis adoptent bien la structuration attendue en feuillet  $\beta$ . Dans un troisième temps, nous exposerons d'une façon détaillée et séparée les résultats de RMN. Enfin, les conclusions de la partie expérimentale seront présentées. Les matériels et méthodes utilisés se trouveront à la fin du chapitre.

## I. Constitution de la banque de données et sélection des peptides modèles

La première étape du travail a été la recherche de séquences peptidiques afin de constituer une banque de peptides pour les coudes  $\beta$  (types I, I', II et II') et les coudes  $\gamma$  (classique et inverse). Une analyse approfondie de la Protein Data Bank (PDB, <http://www.rcsb.org/>)<sup>131</sup> a été menée pour identifier, pour chaque nature et type de coudes, des motifs  $\beta$ -hairpin qui devaient satisfaire les critères suivants :

- Être situé de préférence à la surface de la protéine ;
- Présenter des séquences suffisamment variables en acides aminés au niveau du coude afin d'obtenir des valeurs consensuelles et caractéristiques des paramètres de RMN ;
- Avoir une distance entre les carbones  $\alpha$  des résidus  $i$  et  $i + n$  (ou  $n=2$  ou  $3$ ) inférieure à  $7\text{\AA}$  ;
- Présenter une liaison hydrogène entre l'oxygène du carbonyle du premier acide aminé ( $i$ ) et le proton amide du dernier acide aminé ( $i+n$ , ou  $n=2$  ou  $3$ ).

La Figure 44 présente un exemple de structure cristalline d'une protéine pour laquelle un motif  $\beta$ -hairpin (contenant un coude de type I) a été identifié.

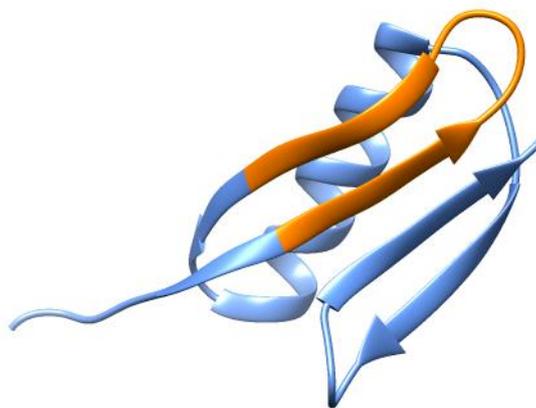


Figure 44 Structure cristalline du troisième domaine de la protéine G du streptocoque formé par deux motifs  $\beta$ -hairpin et une hélice  $\alpha$  (Code PDB : 1IGD, Derrick et al., 1994). En orange, le motif  $\beta$ -hairpin composé par deux brins  $\beta$  antiparallèles reliés par un coude  $\beta$  de type I.

La stabilité structurale du motif  $\beta$ -hairpin étant liée à l'existence de liaisons hydrogène entre les groupements -CO et -NH des brins, les peptides étudiés devaient contenir dans leur séquence des acides aminés appartenant aux brins  $\beta$ . Toutefois, afin de limiter les phénomènes d'agrégation, les superpositions sur les spectres RMN et de permettre une analyse détaillée des paramètres au niveau du coude, nous souhaitions travailler sur des peptides les plus courts possibles. Les peptides sélectionnés devaient donc contenir, en plus des acides aminés du coude, 2 à 5 résidus appartenant à chaque brin  $\beta$ .

Outre les peptides identifiés dans la PDB, une recherche bibliographique a également été réalisée afin de retenir, dans notre banque de données, des peptides déjà étudiés qui adoptent, en solution, une structuration en  $\beta$ -hairpin.<sup>127</sup>

Afin de construire cette banque de données, seules les structures résolues par cristallographie aux rayons X ont été prises en compte. Ce choix de méthode de détermination de structure a été fait car nous ne voulions pas de biais dus à la RMN (coude non ou mal défini). Les structures présentaient une résolution entre 1,00 et 2,50 Å et une valeur du facteur-R basse (inférieure à 0,26). Les structures au format .pdb ont été téléchargées et analysées en utilisant le logiciel Chimera.<sup>132</sup> Pour les différentes natures ( $\beta$  et  $\gamma$ ) et types de coudes (I, I', II, II', classique et inverse), les angles dièdres  $\Phi$  et  $\Psi$  des deux acides aminés centraux (pour les  $\beta$ ) et de l'acide aminé central (pour les  $\gamma$ ) ont été mesurés afin de ne retenir que des structures dans lesquelles leurs valeurs ne s'écartaient pas plus de  $\pm 30^\circ$  par rapport aux valeurs des angles dièdres canoniques établis dans la littérature (cf. Tableau 9, Chapitre 1).

Enfin, afin de constituer une banque de données représentative, une analyse des séquences en acides aminés a également été menée. Ces séquences ont pu être identifiées selon les préférences de position des acides aminés : ceux-ci se localisent en positions  $i+1$  et  $i+2$  dans les coudes  $\beta$  et en position  $i+1$  dans les coudes  $\gamma$  (cf. partie II.2.2, chapitre 1).

Dans le Tableau 14 sont présentés les différents peptides identifiés. Un total de 31 séquences, répondant aux critères définis précédemment, a été trouvé. Nous pouvons observer que les coudes  $\beta$  de type I sont les plus fréquents, suivis par le type I', II et II'. En revanche, les coudes de nature  $\gamma$  sont normalement évités dans les structures cristallines des protéines et leur fréquence est donc très restreinte.

Tableau 14 Résumé des séquences des peptides sélectionnées pour les coudes de nature  $\beta$  de type I, I', II et II' et de nature  $\gamma$  de type classique et inverse. Parmi ces peptides, ceux qui ont été déjà étudiés en solution sont signalés avec le symbole « \* ».

<b>Type <math>\beta</math>-I</b>							
	<b>Code PDB</b>	<b>Résolution (Å)</b>	<b>R-factor</b>	<b>Brin de feuillet <math>\beta</math></b>	<b>Coude</b>	<b>Brin de feuillet <math>\beta</math></b>	<b>Référence</b>
1	5II0	2,09	0,170	K <sub>1</sub> H <sub>2</sub>	P <sub>3</sub> E <sub>4</sub> N <sub>5</sub> N <sub>6</sub>	R <sub>7</sub> T <sub>8</sub>	Johansson and Reedtz-Runge, 2016
2	1IGD*	1,10	0,193	K <sub>1</sub> L <sub>2</sub> V <sub>3</sub> L <sub>4</sub> N <sub>5</sub>	G <sub>6</sub> K <sub>7</sub> T <sub>8</sub> L <sub>9</sub>	K <sub>10</sub> G <sub>11</sub> E <sub>12</sub> T <sub>13</sub> T <sub>14</sub>	Derrick <i>et al.</i> , 1994
3	1BBP	2,00	0,200	C <sub>1</sub> K <sub>2</sub> Y <sub>3</sub> D <sub>4</sub>	E <sub>5</sub> D <sub>6</sub> K <sub>7</sub> K <sub>8</sub>	G <sub>9</sub> H <sub>10</sub> Q <sub>11</sub> D <sub>12</sub>	Huber <i>et al.</i> , 1990
4	3SGB	2,00	0,210	A <sub>1</sub> L <sub>2</sub> Y <sub>3</sub>	S <sub>4</sub> S <sub>5</sub> T <sub>6</sub> G <sub>7</sub>	R <sub>8</sub> C <sub>9</sub> S <sub>10</sub>	Weber <i>et al.</i> , 2011
5	1Z2V	1,90	0,186	L <sub>1</sub> L <sub>2</sub> Y <sub>3</sub> C <sub>4</sub>	S <sub>5</sub> N <sub>6</sub> G <sub>7</sub> G <sub>8</sub>	H <sub>9</sub> F <sub>10</sub> L <sub>11</sub> R <sub>12</sub>	Lee <i>et al.</i> , 2005
6	1QCS	1,90	0,210	H <sub>1</sub> V <sub>2</sub> I <sub>3</sub> V <sub>4</sub> R <sub>5</sub> T <sub>6</sub>	S <sub>7</sub> P <sub>8</sub> N <sub>9</sub> H <sub>10</sub>	K <sub>11</sub> Y <sub>12</sub> I <sub>13</sub> F <sub>14</sub> T <sub>15</sub> L <sub>16</sub>	Yu <i>et al.</i> , 1999
7	1BPI	1,09	0,187	I <sub>1</sub> R <sub>2</sub> Y <sub>3</sub> F <sub>4</sub> Y <sub>5</sub> N <sub>6</sub>	A <sub>7</sub> K <sub>8</sub> A <sub>9</sub> G <sub>10</sub>	L <sub>11</sub> C <sub>12</sub> Q <sub>13</sub> T <sub>14</sub> F <sub>15</sub> V <sub>16</sub>	Parkin <i>et al.</i> , 1996
8	3KQN	2,05	0,190	N <sub>1</sub> L <sub>2</sub> R <sub>3</sub>	T <sub>4</sub> G <sub>5</sub> V <sub>6</sub> R <sub>7</sub>	T <sub>8</sub> I <sub>9</sub>	Gu and Rice, 2010
<b>Type <math>\beta</math>-I'</b>							
1	1PP2	2,05	-	T <sub>1</sub> Y <sub>2</sub> S <sub>3</sub> E <sub>4</sub>	E <sub>5</sub> N <sub>6</sub> G <sub>7</sub> E <sub>8</sub>	I <sub>9</sub> I <sub>10</sub> C <sub>11</sub> G <sub>12</sub>	Brunie <i>et al.</i> , 1996
2	1EG9	1,60	0,190	R <sub>1</sub> V <sub>2</sub> E <sub>3</sub> S <sub>4</sub>	F <sub>5</sub> H <sub>6</sub> G <sub>7</sub> F <sub>8</sub>	I <sub>9</sub> Y <sub>10</sub> G <sub>11</sub> C <sub>12</sub>	Carredano <i>et al.</i> , 2000
3	7FAB	2,00	0,169	Y <sub>1</sub> C <sub>2</sub> Q <sub>3</sub> T <sub>4</sub>	H <sub>5</sub> E <sub>6</sub> G <sub>7</sub> S <sub>8</sub>	T <sub>9</sub> V <sub>10</sub> E <sub>11</sub> K <sub>12</sub>	Saul and Poljak, 1992
4	3SSI	2,30	0,185	T <sub>1</sub> V <sub>2</sub> D <sub>3</sub> G <sub>4</sub> V <sub>5</sub>	W <sub>6</sub> Q <sub>7</sub> G <sub>8</sub> K <sub>9</sub>	R <sub>10</sub> V <sub>11</sub> S <sub>12</sub> Y <sub>13</sub>	Suzuki, 1996
5	3APP	1,80	0,126	F <sub>1</sub> S <sub>2</sub> V <sub>3</sub> S <sub>4</sub>	I <sub>5</sub> S <sub>6</sub> G <sub>7</sub> Y <sub>8</sub>	T <sub>9</sub> A <sub>10</sub> T <sub>11</sub> V <sub>12</sub>	James <i>et al.</i> , 1983
6	2ALP	1,70	0,131	E <sub>1</sub> Y <sub>2</sub> S <sub>3</sub>	I <sub>4</sub> N <sub>5</sub> N <sub>6</sub> A <sub>7</sub>	S <sub>8</sub> L <sub>9</sub> C <sub>10</sub> S <sub>11</sub>	Fujinaga <i>et al.</i> , 1985
7	8ADH	2,40	-	F <sub>1</sub> I <sub>2</sub>	C <sub>3</sub> R <sub>4</sub> G <sub>5</sub> K <sub>6</sub>	P <sub>7</sub> I <sub>8</sub>	Jones and Ehlund, 1989
<b>Type <math>\beta</math>-II</b>							
1	2CPL	1,63	0,180	H <sub>1</sub> R <sub>2</sub> I <sub>3</sub>	I <sub>4</sub> P <sub>5</sub> G <sub>6</sub> F <sub>7</sub>	M <sub>8</sub> C <sub>9</sub> Q <sub>10</sub> G <sub>11</sub>	Ke <i>et al.</i> , 1992
2	2POR	1,80	0,186	Q <sub>1</sub> V <sub>2</sub> L <sub>3</sub> D <sub>4</sub>	I <sub>5</sub> D <sub>6</sub> T <sub>7</sub> I <sub>8</sub>	D <sub>9</sub> D <sub>10</sub> V <sub>11</sub> T <sub>12</sub> Y <sub>13</sub>	Weiss and Schulz, 1992
3	2PIA	2,00	0,186	S <sub>1</sub> K <sub>2</sub> E <sub>3</sub> K <sub>4</sub> L <sub>5</sub>	A <sub>6</sub> R <sub>7</sub> D <sub>8</sub> I <sub>9</sub>	W <sub>10</sub> S <sub>11</sub> F <sub>12</sub> E <sub>13</sub> L <sub>14</sub>	Correll <i>et al.</i> , 1992
4	1PGS	1,80	0,168	H <sub>1</sub> A <sub>2</sub> K <sub>3</sub> P <sub>4</sub> Y <sub>5</sub>	D <sub>6</sub> A <sub>7</sub> G <sub>8</sub> S <sub>9</sub>	R <sub>10</sub> G <sub>11</sub> C <sub>12</sub> A <sub>13</sub>	Norris <i>et al.</i> , 1994
5	1A1X	2,00	0,211	W <sub>1</sub> Q <sub>2</sub> L <sub>3</sub>	Y <sub>4</sub> P <sub>5</sub> E <sub>6</sub> E <sub>7</sub>	R <sub>8</sub> Y <sub>9</sub> M <sub>10</sub> D <sub>11</sub>	Fu <i>et al.</i> , 1998
6	1A7T	1,85	0,151	S <sub>1</sub> L <sub>2</sub> A <sub>3</sub> E <sub>4</sub>	I <sub>5</sub> E <sub>6</sub> G <sub>7</sub> W <sub>8</sub>	G <sub>9</sub> W <sub>10</sub> G <sub>11</sub> M <sub>12</sub> V <sub>13</sub>	Fitzgerald <i>et al.</i> , 1998

<b>Type <math>\beta</math>-II'</b>							
1	2SGA	1,50	0,126	S <sub>1</sub> L <sub>2</sub> F <sub>3</sub>	A <sub>4</sub> G <sub>5</sub> S <sub>6</sub> T <sub>7</sub>	A <sub>8</sub> L <sub>9</sub> G <sub>10</sub>	Moult <i>et al.</i> , 1985
2	1MDC	1,75	-	S <sub>1</sub> M <sub>2</sub> Y <sub>3</sub> T <sub>4</sub> V <sub>5</sub>	D <sub>6</sub> G <sub>7</sub> N <sub>8</sub> V <sub>9</sub>	V <sub>10</sub> T <sub>11</sub> H <sub>12</sub> V <sub>13</sub> K <sub>14</sub>	Benning and Holden, 1992
3	3APP	1,80	0,126	D <sub>1</sub> S <sub>2</sub> Y <sub>3</sub> T <sub>4</sub> A <sub>5</sub>	A <sub>6</sub> G <sub>7</sub> S <sub>8</sub> Q <sub>9</sub>	S <sub>10</sub> G <sub>11</sub> D <sub>12</sub> G <sub>13</sub> F <sub>14</sub>	James <i>et al.</i> , 1983
4	1UXA	1,50	0,180	N <sub>1</sub> F <sub>2</sub> R <sub>3</sub>	S <sub>4</sub> G <sub>5</sub> N <sub>6</sub> S <sub>7</sub>	N <sub>8</sub>	Burmeister <i>et al.</i> , 2004
5	2KIN	2,00	0,194	P <sub>1</sub> K <sub>2</sub> F <sub>3</sub>	K <sub>4</sub> G <sub>5</sub> E <sub>6</sub> E <sub>7</sub>	T <sub>8</sub> V <sub>9</sub> V <sub>10</sub>	Sack <i>et al.</i> , 1997
<b>Type <math>\gamma</math> classique</b>							
1	1AWS	2,50	0,252	D <sub>1</sub> I <sub>2</sub> A <sub>3</sub> V <sub>4</sub>	D <sub>5</sub> G <sub>6</sub> E <sub>7</sub>	P <sub>8</sub> L <sub>9</sub> G <sub>10</sub> R <sub>11</sub>	Vodjos <i>et al.</i> , 1997
2	1GND	1,81	0,198	V <sub>1</sub> K <sub>2</sub> S <sub>3</sub>	E <sub>4</sub> G <sub>5</sub> E <sub>6</sub>	V <sub>7</sub> A <sub>8</sub>	Schalk <i>et al.</i> , 1996
3	1TLX	2,10	0,153	T <sub>1</sub> Y <sub>2</sub>	S <sub>3</sub> T <sub>4</sub> Y <sub>5</sub>	Y <sub>6</sub> Y <sub>7</sub>	English <i>et al.</i> , 1999
<b>Type <math>\gamma</math> inverse</b>							
1	1MOL	1,70	0,174	K <sub>1</sub> T <sub>2</sub> I <sub>3</sub> Y <sub>4</sub> E <sub>5</sub>	N <sub>6</sub> E <sub>7</sub> R <sub>8</sub>	E <sub>9</sub> I <sub>10</sub> K <sub>11</sub> G <sub>12</sub> Y <sub>13</sub>	Somoza <i>et al.</i> , 1993
2	1YTB	1,80	0,201	A <sub>1</sub> L <sub>2</sub> I <sub>3</sub> F <sub>4</sub>	A <sub>5</sub> S <sub>6</sub> G <sub>7</sub>	K <sub>8</sub> M <sub>9</sub> V <sub>10</sub> V <sub>11</sub>	Kim <i>et al.</i> , 1993

Une fois les séquences identifiées pour constituer la base de données, un tri a été effectué afin de sélectionner les peptides modèles utilisés pour les analyses expérimentales. Cette sélection a été guidée par plusieurs critères : (i) les peptides modèles ne devaient pas contenir plus de 14 résidus afin de limiter les superpositions dans les spectres RMN, (ii) les séquences en acides aminés devaient être différentes et ne pas inclure d'enchaînements de résidus avec des chaînes latérales trop encombrantes (cycles aromatiques, chaînes longues) car ils pouvaient induire une difficulté de synthèse, (iii) le nombre de résidus du  $\beta$ -hairpin devait être varié pour avoir des peptides courts (8 acides aminés) et des peptides de longueur moyenne (10-14 acides aminés) ; (iv) un peptide, qui en solution se replie dans la conformation souhaitée, déjà étudié dans la littérature, a été retenu afin d'essayer de retrouver en solution le motif  $\beta$ -hairpin (1IGD).

Afin de limiter le nombre de synthèses, seule une séquence représentative pour chaque type de coudes  $\beta$  de type I, I', II et II' et  $\gamma$  de type classique et inverse a été choisie. Six peptides ont donc été au final retenus pour la suite de l'étude (Tableau 15). Pour des facilités de lecture, les peptides seront dénommés selon le code PDB de la protéine dont ils sont issus. Parmi les peptides sélectionnés, certains étaient de longueur restreinte (8 acides aminés, 8ADH et 1GND) et d'autres de longueur plus élevées (10-14 résidus, 1IGD, 1A7T, 2SGA et 1MOL). Pour ce qui concerne le peptide issu de la littérature, le peptide 1IGD, nous avons raccourci la longueur de la molécule de quatre résidus, en espérant pouvoir conserver le motif  $\beta$ -hairpin en solution.

Tableau 15 Résumé des peptides modèles sélectionnés pour les expériences de DC et de RMN. En gras, les acides aminés centraux pour les coudes  $\beta$  et l'acide aminé central pour les coudes  $\gamma$ .

	<b>Code PDB</b>	<b>Brin de feuillet <math>\beta</math></b>	<b>Coude</b>	<b>Brin de feuillet <math>\beta</math></b>	<b>Type de coude</b>
1	1IGD	K <sub>1</sub> L <sub>2</sub> V <sub>3</sub> I <sub>4</sub> N <sub>5</sub>	G <sub>6</sub> <b>K<sub>7</sub></b> T <sub>8</sub> L <sub>9</sub>	K <sub>10</sub> G <sub>11</sub> E <sub>12</sub> T <sub>13</sub> T <sub>14</sub>	$\beta$ -I
2	8ADH	F <sub>1</sub> I <sub>2</sub>	C <sub>3</sub> <b>R<sub>4</sub></b> G <sub>5</sub> K <sub>6</sub>	P <sub>7</sub> I <sub>8</sub>	$\beta$ -I'
3	1A7T	S <sub>1</sub> L <sub>2</sub> A <sub>3</sub> E <sub>4</sub>	I <sub>5</sub> <b>E<sub>6</sub></b> G <sub>7</sub> W <sub>8</sub>	G <sub>9</sub> W <sub>10</sub> G <sub>11</sub> M <sub>12</sub> V <sub>12</sub>	$\beta$ -II
4	2SGA	S <sub>1</sub> L <sub>2</sub> F <sub>3</sub>	A <sub>4</sub> <b>G<sub>5</sub></b> S <sub>6</sub> T <sub>7</sub>	A <sub>8</sub> L <sub>8</sub> G <sub>10</sub>	$\beta$ -II'
5	1GND	V <sub>1</sub> K <sub>2</sub> S <sub>3</sub>	E <sub>4</sub> <b>G<sub>5</sub></b> E <sub>6</sub>	V <sub>7</sub> A <sub>8</sub>	$\gamma$ classique
6	1MOL	K <sub>1</sub> T <sub>2</sub> I <sub>3</sub> Y <sub>4</sub> E <sub>5</sub>	N <sub>6</sub> <b>E<sub>7</sub></b> R <sub>8</sub>	E <sub>9</sub> I <sub>10</sub> K <sub>11</sub> G <sub>12</sub> Y <sub>13</sub>	$\gamma$ inverse

## II. Synthèse des peptides modèles

Les peptides 1IGD, 8ADH, 2SGA, 1GND et 1MOL ont été synthétisés chimiquement au sein de laboratoire DC2N (Laboratoire U1239, plateforme PRIMACEN (Rouen, France)) sur phase solide (Solid Phase Peptide Synthesis -SPPS) à l'aide d'un synthétiseur de peptide automatisé avec micro-ondes. Les peptides ont été ensuite purifiés en phase inversée à l'aide d'une HPLC (high pressure liquid chromatography) dans un système H<sub>2</sub>O/acétonitrile comportant un faible

pourcentage de TFA (0,1%) puis lyophilisés. Une première caractérisation par spectrométrie de masse MALDI-TOF a été ensuite réalisée afin de vérifier la bonne réussite de la synthèse des peptides en termes de séquence et de pureté. Pour ce qui concerne le peptide 1A7T, la présence de deux acides aminés aromatiques (tryptophane) dans la séquence n'a pas permis de réaliser une synthèse correcte du peptide et ne sera donc pas présenté dans la suite de ce travail.

### **III. Étude expérimentale des peptides modèles**

#### **III.1 Analyse conformationnelle par DC des peptides modèles**

La première étape du travail expérimental a consisté à analyser les peptides synthétisés par dichroïsme circulaire dans différents milieux afin de contrôler leur solubilité et de vérifier qu'ils adoptent la structuration attendue en conformation «  $\beta$ -hairpin ». La spectroscopie de DC est une technique extrêmement sensible pour analyser la structure secondaire des peptides ou des protéines en solution (*cf.* partie I.2.1, chapitre 2).

Dans cette partie du projet, cette technique nous a permis d'effectuer le choix du milieu pour poursuivre les analyses. Ce choix est une étape importante de l'étude structurale par RMN et s'effectue en tenant compte des caractéristiques du peptide étudié. Il s'agit de trouver un compromis entre la solubilité, la stabilité, la structuration et la compatibilité avec la technique d'analyse utilisée ultérieurement, à savoir la RMN en solution.

Dans les études structurales des peptides et protéines, le tampon phosphate est le plus souvent choisi afin de mimer l'environnement physiologique. Cependant, la flexibilité d'un peptide de petite taille ne permet généralement pas de stabiliser la structuration dans ce milieu. Pour favoriser le repliement, la stratégie employée est donc d'utiliser un mélange tampon phosphate/solvant organique à des concentrations modérées. En pratique, des solutions tampon phosphate/solvant organique sont préparées en augmentant progressivement les proportions du solvant. Le mélange, offrant les meilleures conditions d'analyses, est choisi. En effet, ce mélange doit favoriser la formation d'un nombre suffisant d'interactions caractéristiques permettant la structuration du peptide ainsi que la détermination de structures secondaires spécifiques au sein de la molécule. De ce fait, les solvants organiques comme le méthanol (MeOH) ou le trifluoroéthanol (TFE, 2,2,2-trifluoroéthanol) sont souvent utilisés. Dans le cadre de cette partie de ce projet de thèse, nous avons donc décidé d'utiliser des mélanges tampon phosphate/solvant organique avec le but d'identifier les conditions expérimentales dans lesquelles, les peptides choisis étaient solubles et leur structuration le mieux stabilisée.

Pour un peptide, les spectres de DC résultent de la superposition des spectres des différents éléments de structure secondaire présents au sein de la molécule. Or, comme présenté au chapitre 2, la présence du coude, dans nos peptides, va changer l'apparence des spectres qui n'auront pas nécessairement l'allure des spectres caractéristiques mais celle d'un mélange de ces conformations. Ceci nous donnera des spectres de DC qui ne pourront pas être comparés avec les spectres de référence des autres structures secondaires (Figure 23). De plus, le manque de spectres caractéristiques pour chaque type de coude, ne nous permet pas non plus d'estimer l'impact que le coude aura sur les spectres globaux de DC. Pour ces raisons, les analyses par DC ont été conduites en cherchant une allure différente de celle d'une conformation aléatoire pour voir si les peptides adoptaient une structuration dans les milieux et les conditions testées. Même si dans les peptides modèles, la présence d'une structuration de type coude peut affecter l'allure des spectres de DC, dans cette première partie de l'étude expérimentale, nous nous attendons à observer une importante contribution des brins  $\beta$  présents au sein de la molécule avec une allure globale des spectres similaires à celle d'un feuillet  $\beta$  antiparallèle car la contribution des brins  $\beta$  sera plus importante de celle du coude.

### **III.1.1 Résultats et discussion**

Les spectres de DC ont été enregistrés dans du tampon phosphate à pH 6,0 pour le peptide modèle 8ADH, à pH 5,3 pour les peptides 1IGD, 1MOL 1GND et à pH 3,5 pour le peptide 2SGA (Tableau 16, cf Matériels et Méthode de ce chapitre). Dans ce milieu, les peptides 1IGD, 8ADH, 2SGA, 1GND présentent seulement un minimum autour de 190-195 nm (Figure 45), caractéristique d'une chaîne peptidique qui adopte une conformation aléatoire. Ceci n'est cependant pas étonnant puisque le milieu aqueux n'est pas un milieu favorable pour la stabilisation des structurations dans les peptides de petite taille. Différemment des autres, le peptide 1MOL, dans ces conditions, ne présente pas de minimum autour de 190 nm, suggérant ainsi qu'il adopte une structure secondaire dans le tampon phosphate (Figure 45). L'allure observée n'est pas semblable aux spectres de DC de référence des différentes structures secondaires (Figure 23). De plus, la faible intensité des bandes indique une faible structuration.

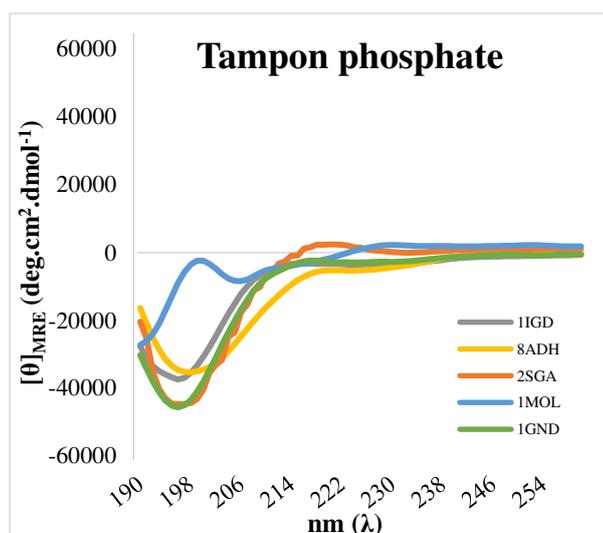


Figure 45 Spectres de DC de cinq peptides modèles dans du tampon phosphate. (A) En gris IIGD (20 mM, pH 6,0), en jaune 8ADH (20 mM, pH 6,0), en orange 2SGA (20 mM, pH 3,5), en bleu ciel IMOL (20mM, pH 5,3) et en vert 1GND (20 mM, pH 5,3).

Afin d'aider à la structuration, des co-solvants organiques, tels que le TFE et le MeOH à différents pourcentages (25%, 50%, 75%) ont été utilisés. Pour les deux peptides 2SGA et 1GND, les spectres de DC enregistrés dans ces milieux suggèrent qu'ils n'adoptent pas de structuration secondaire vue la présence du minimum à 195 nm caractéristique d'une chaîne polypeptidique en conformation aléatoire (Figures 46 et 47). Pour ces deux peptides, les conditions dans le milieu tampon phosphate/TFE et tampon phosphate/MeOH (75/25 v/v) n'ont pas été enregistrées (Figures 46 et 47) car nous avons pu observer que les peptides n'adoptaient pas de structuration secondaire en présence de co-solvants organiques à plus hauts pourcentages (50/50 v/v et 25/75 v/v). De plus, pour ces deux peptides, il est possible d'observer une diminution de l'intensité des signaux avec l'augmentation du pourcentage du co-solvant organique. Ceci indique une faible solubilité de ces deux molécules dans les milieux de travail choisis.

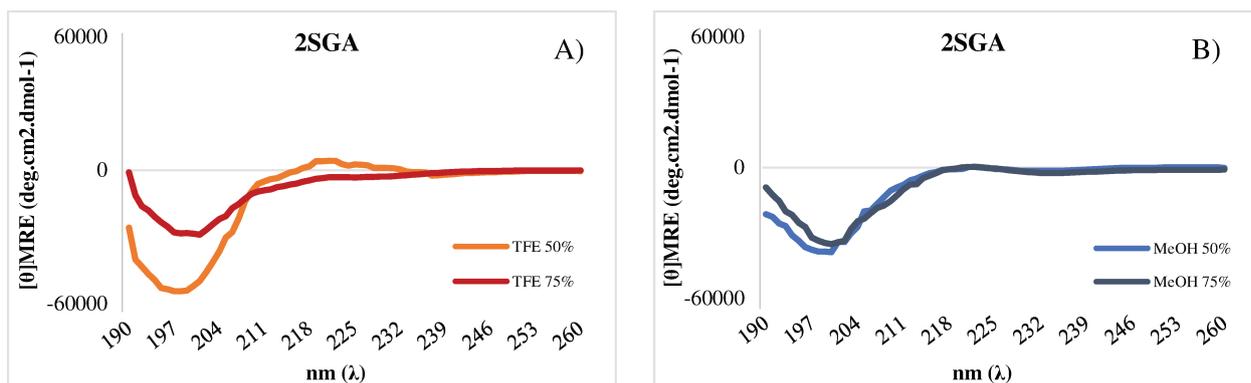


Figure 46 Spectres de DC du peptide 2SGA dans différents mélanges : (A) en milieu tampon phosphate/TFE (20 mM, pH 5,3) –en orange (50/50 v/v), en rouge (25/75 v/v) ; (B) en milieu tampon phosphate/MeOH (20 mM, pH 5,3) – en bleu (50/50 v/v), en bleu foncé (25/75 v/v).

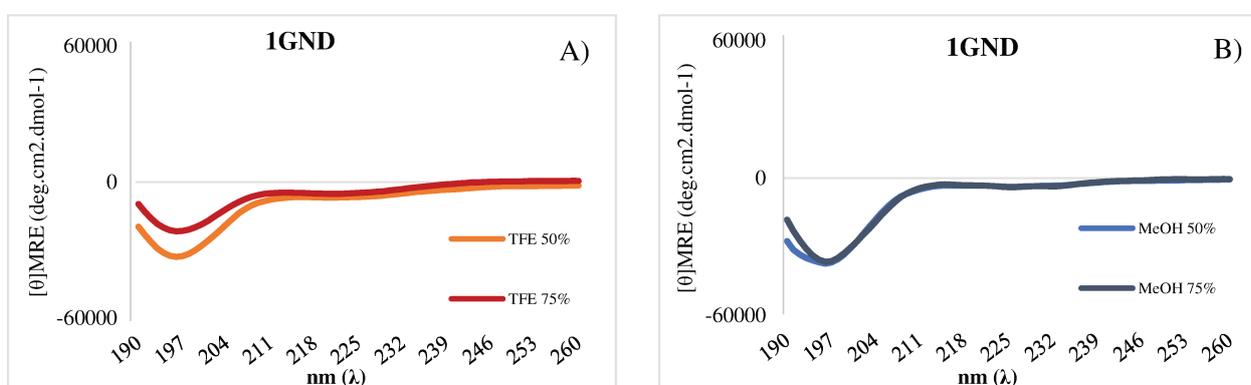


Figure 47 Spectres de DC du peptide 1GND dans différents mélanges : (A) en milieu tampon phosphate/TFE (20 mM, pH 5,3) –en orange (50/50 v/v), en rouge (25/75 v/v) ; (B) en milieu tampon phosphate/MeOH (20 mM, pH 5,3) – en bleu (50/50 v/v), en bleu foncé (25/75 v/v).

Les spectres de DC du peptide 1IGD en présence d'un mélange tampon phosphate/TFE aux pourcentages de 25%, 50% et 75% présentent un premier minimum vers 202 nm et un deuxième minimum vers 218 nm (Figure 48A). Ces derniers permettent de mettre en évidence que l'addition d'un co-solvant organique est à l'origine d'un changement de la conformation du peptide par rapport aux spectres enregistrés dans le tampon phosphate.

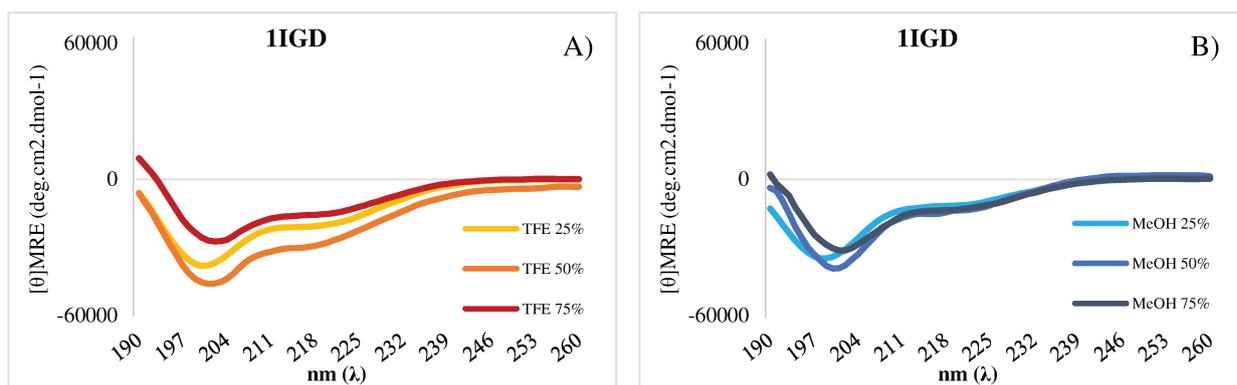


Figure 48 Spectres de DC du peptide IIGD dans différents mélanges : (A) en milieu tampon phosphate/TFE (20 mM, pH 3,5) – en jaune (75/25 v/v), en orange (50/50 v/v), en rouge (25/75 v/v) ; (B) en milieu tampon phosphate/MeOH (20 mM, pH 3,5) – en bleu ciel (75/25 v/v), en bleu (50/50 v/v), en bleu foncé (25/75 v/v).

Le spectre de DC dans ce milieu ne présente pas les deux minima et le maximum typiques des hélices  $\alpha$ , mais il pourrait être compatible avec un mélange de coudes et feuillet  $\beta$  (Figure 48A).<sup>90,133</sup> Ceci apparaît logique car, dans la structure cristallographique, dix résidus sont engagés dans une structuration en feuillet  $\beta$  antiparallèle et seuls quatre résidus sont impliqués dans une structuration de type coude  $\beta$ -I. Les résultats obtenus pour ce peptide sont en accord avec ceux de F. J. Blanco sur la même séquence, mais plus longue.<sup>133</sup> De notre côté, nous avons étudié une version plus courte du peptide afin de voir si la structuration en  $\beta$ -hairpin était conservée. Les analyses de DC suggèrent que la diminution du nombre de résidus des brins  $\beta$  ne modifie pas la conformation attendue.

L'analyse de ce peptide dans le mélange tampon phosphate/MeOH met en évidence des spectres similaires à ceux observés dans le mélange tampon phosphate/TFE indiquant que le MeOH aide aussi à la structuration de ce peptide en solution (Figure 48B).

De la même manière que pour les peptides 2SGA et 1GND présentés précédemment, nous observons, pour le peptide IIGD, une diminution de l'intensité des signaux avec l'augmentation du pourcentage du co-solvant organique. Ceci indique des problèmes de solubilité dans les deux milieux (Figure 48).

Le spectre de DC du peptide 8ADH, dans le mélange tampon phosphate/TFE (75/25 v/v) présente un minimum assez large entre 200 et 225 nm, tandis que dans le mélange tampon phosphate/TFE (50/50 v/v) il semble présenter deux minima, un à 212 nm et un autre en dessous de 200 nm (Figure 49A). Ce deuxième minimum n'est plus présent à un plus haut pourcentage de TFE puisque le spectre dans le mélange tampon phosphate/TFE (25/75 v/v) est caractérisé par un seul minimum à 210 nm (Figure 49A).

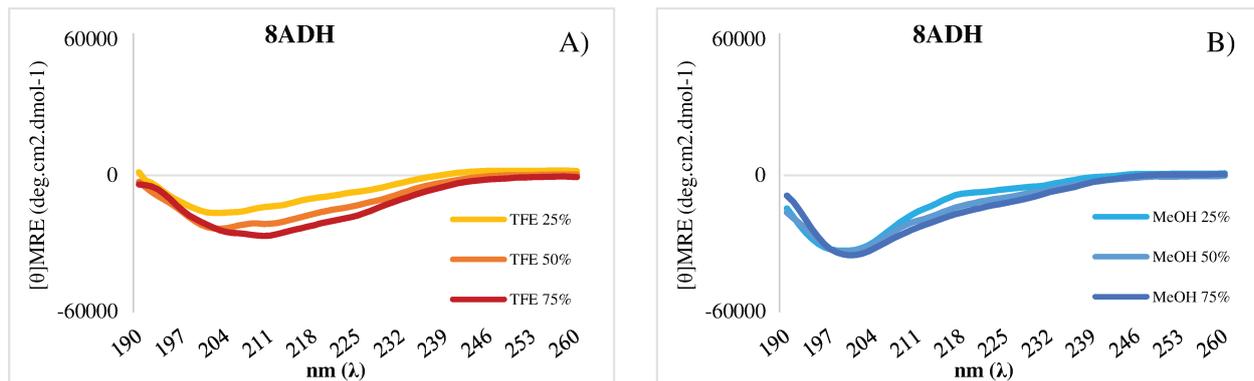


Figure 49 Spectres de DC du peptide 8ADH dans différents mélanges : (A) en milieu tampon phosphate/TFE (20 mM, pH 6,0) – en jaune (75/25 v/v), en orange (50/50 v/v), en rouge (25/75 v/v) ; (B) en milieu tampon phosphate/MeOH (20 mM, pH 6,0) – en bleu ciel (75/25 v/v), en bleu (50/50 v/v), en bleu foncé (25/75 v/v).

Ces derniers indiquent donc que le peptide adopte une structuration secondaire en solution dans ces conditions expérimentales. Cependant, la comparaison des spectres de DC avec ceux de référence ne permet pas d'identifier une structuration en feuillet  $\beta$  au sein de cette molécule. Ceci est peut-être dû au fait que le nombre d'acides aminés qui devrait être impliqué dans une conformation étendue est le même que celui des résidus engagés en conformation coude  $\beta$ . L'observation des spectres de DC enregistrés dans le mélange tampon phosphate/MeOH à tous les pourcentages laisse supposer que ce peptide n'adopte pas de structure secondaire dans ce milieu car son allure est caractérisée par un seul minimum à  $\sim 200$  nm (Figure 49B). Même si le signal d'une conformation aléatoire est normalement à 190 nm, la longueur de la molécule et la séquence en acides aminés pourraient expliquer ce décalage en nm. De manière intéressante, par rapport aux autres peptides présentés précédemment, 8ADH ne montre pas de problèmes de solubilité dans les deux milieux.

Sur les spectres de DC du peptide 1MOL enregistrés dans le mélange tampon phosphate/TFE (25%, 50% et 75%) à pH 5,3, nous pouvons observer un maximum à 195 nm et deux minima à  $\sim 206$  et  $\sim 222$  nm (Figure 50A).

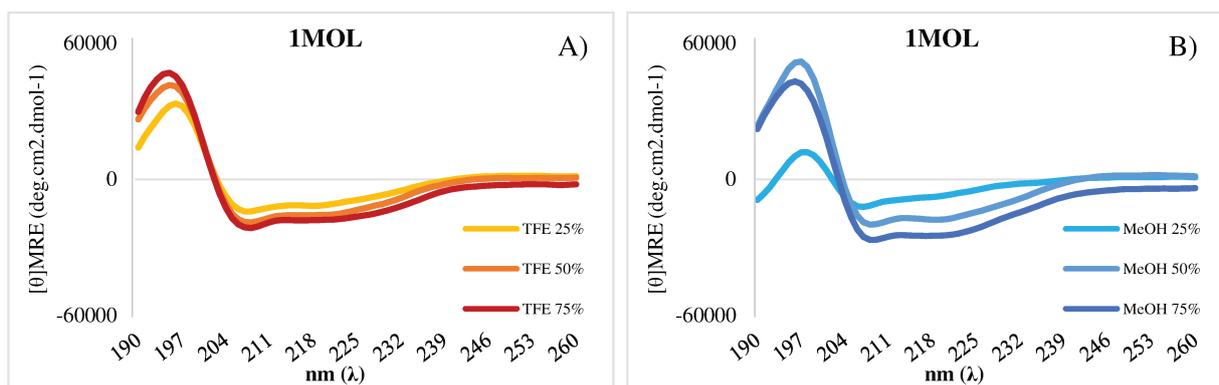


Figure 50 Spectres de DC du peptide 1MOL dans différents mélanges : (A) en milieu tampon phosphate/TFE (20 mM, pH 5,3) - en jaune (75/25 v/v), en orange (50/50 v/v), en rouge (25/75 v/v) ; (B) en milieu tampon phosphate/MeOH (20 mM, pH 3,5) – en bleu clair (75/25 v/v), en bleu (50/50 v/v), en bleu foncé (25/75 v/v).

Ces signaux suggèrent que le peptide en solution adopte une structure en conformation hélicoïdale (Figure 50A). Ceci peut être expliqué par le fait que le TFE est connu pour être un solvant organique qui induit et stabilise plus facilement une séquence peptidique en conformation hélice plutôt qu'en feuillet  $\beta$  antiparallèle.<sup>134-137</sup>

L'analyse du peptide dans le mélange tampon phosphate/MeOH semble indiquer que le peptide est mieux structuré dans ce solvant (Figure 50B). Les deux minima à 222 et 210 nm et le maximum à 197 nm sont plus nets que dans le mélange tampon phosphate/TFE. Le repliement adopté est toujours de type hélicoïdal en présence de 50% et 75% de méthanol. En revanche, dans le mélange tampon phosphate/MeOH (75/25 v/v), les signaux caractéristiques de l'hélice sont peu marqués : seuls un minimum à 210 nm et un maximum à 198 nm peuvent être remarqués. Ces derniers indiquent probablement une plus faible stabilisation du motif hélicoïdal à 25% de MeOH. Il est toutefois possible que la modification de l'allure du spectre soit liée à la présence d'un autre motif tel qu'un coude.

L'analyse qualitative effectuée par dichroïsme circulaire sur les peptides modèles choisis nous a permis, d'une part, d'écarter les peptides modèles qui ne se structurent pas en solution, et d'autre part, d'identifier les milieux les plus adaptés permettant une stabilisation de la structuration de ces derniers pour poursuivre l'analyse structurale par RMN. Ainsi, en raison des résultats de DC obtenus pour les peptides 1GND et 2SGA, qui, dans les conditions explorées, présentent une conformation aléatoire, l'étude n'a pas été poursuivie. Pour ce qui concerne les peptides 1IGD et 8ADH, nous avons pu observer qu'ils adoptent, en présence de TFE, une structuration secondaire. Au contraire, dans le mélange tampon phosphate/MeOH, parmi ces deux peptides, seul le peptide 1IGD se replie avec une structuration différente de la

conformation aléatoire. En revanche, nous ne pouvons pas *a priori* déterminer, en utilisant seulement la spectroscopie de DC, le type de structure adoptée par ces deux molécules car les signaux retrouvés ne correspondent pas aux spectres de références des feuillets ou des hélices. Enfin, l'analyse du peptide 1MOL nous suggère qu'en présence de TFE (dans toutes les conditions) ainsi qu'en présence de MeOH (50% et 75%) le peptide adopte une conformation en hélice  $\alpha$ . En revanche, dans le mélange tampon phosphate/MeOH (75/25 v/v) l'allure de DC observée n'est pas celle d'une structure hélicoïdale. Celle-ci pourrait suggérer un mélange de structuration de type hélice et coude.

Pour poursuivre l'analyse des peptides modèles par spectroscopie RMN, nous avons donc choisi d'analyser le peptide 1IGD dans le mélange tampon phosphate/TFE (75/25 v/v), le peptide 8ADH dans le mélange tampon phosphate/TFE (50/50 v/v) et le peptide 1MOL dans le mélange tampon phosphate/MeOH (75/25 v/v).

Le reste de ce chapitre est divisé en trois sous-parties présentant chacune les résultats des analyses de RMN obtenus pour les trois peptides modèles (1IGD, 8ADH et 1MOL). Chacune de ces sous-parties contient l'analyse des paramètres structuraux issus des spectres RMN 2D  $^1\text{H}$ - $^1\text{H}$ .

### **III.2 Etude par RMN des peptides 1IGD, 8ADH et 1MOL**

Dans cette partie nous allons présenter l'analyse structurale par RMN des peptides modèles 1IGD, 8ADH et 1MOL qui sont les trois peptides qui adoptaient une structuration secondaire en solution d'après les analyses conformationnelles réalisées par DC. Plus précisément, notre intérêt était d'aller vérifier que pour les peptides représentatifs des coudes de nature  $\beta$  de type I (1IGD) et I' (8ADH), le repliement observé correspondait à celui attendu, à savoir deux brins de feuillets  $\beta$  reliés par un coude. Pour 1MOL (censé représenter le coude  $\gamma$  de type inverse), nous avons décidé de poursuivre l'analyse structurale par RMN afin de confirmer la structuration du peptide en solution.

La démarche utilisée est celle proposée par K. Wüthrich et ses collaborateurs.<sup>118</sup> La première étape de cette stratégie consiste à identifier l'ensemble des protons de chaque acide aminé via l'observation des couplages scalaires dans les expériences RMN homonucléaires 2D de type COSY et TOCSY. La seconde étape permet de déterminer la position de chaque acide aminé dans la séquence du peptide via l'observation des couplages dipolaires entre les protons proches dans l'espace avec une distance inférieure à 5 Å dans les expériences 2D NOESY.

Une fois l'attribution des signaux  $^1\text{H}$  réalisée, l'analyse détaillée des distances inter-protons sur les spectres NOESY permet d'apporter des informations sur la structure secondaire. En particulier, les corrélations  $\text{H}_\alpha\text{-H}_\text{N}$ ,  $\text{H}_\text{N}\text{-H}_\alpha$ ,  $\text{H}_\text{N}\text{-H}_\text{N}$  et  $\text{H}_\alpha\text{-H}_\alpha$ , caractéristiques pour un feuillet  $\beta$  antiparallèle, ont été recherchées au sein de nos peptides modèles. La base théorique de cette technique ainsi que la méthode d'attribution des spectres sont reportées en Annexe 1.

### III.2.1 Résultats et discussion

#### III.2.1.1 Le peptide IIGD

Dans un premier temps, les expériences RMN du peptide IIGD ont été enregistrées à 298K (température ambiante) dans un mélange de tampon phosphate/TFE (20 mM, 75/25 v/v) à pH 5,3. Le pH choisi correspond à un pH légèrement acide et loin du pI. Dans ces conditions, aucun effet structurant n'a pu être observé, en dehors des effets séquentiels (Figure 51).

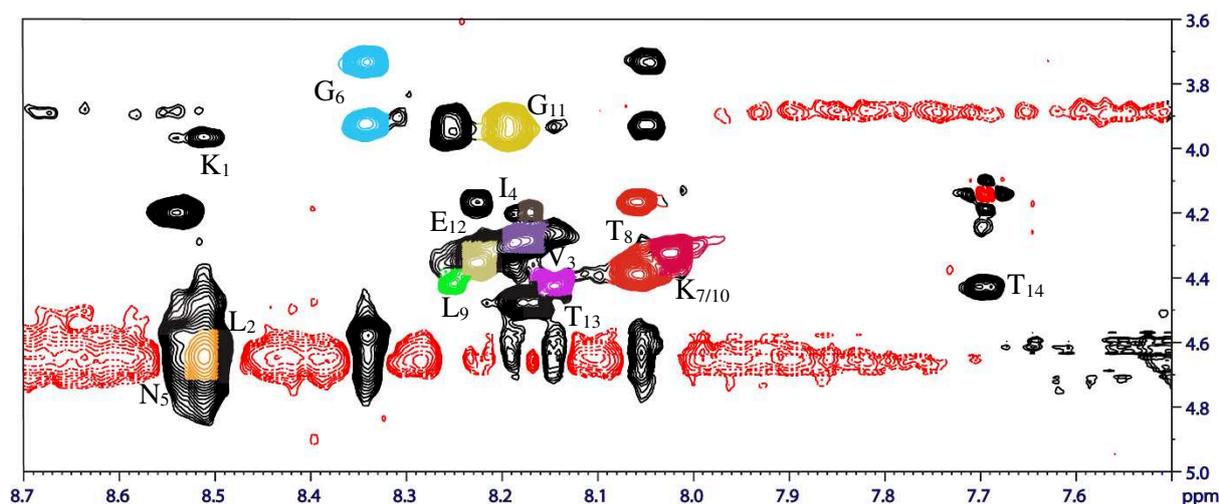


Figure 51 Zoom de la région  $\text{H}_\text{N}\text{-H}_\alpha$  du spectre NOESY du peptide IIGD à 298K dans un mélange de tampon phosphate/TFE (20 mM, 75/25 v/v) à pH 5,3 ; 600 MHz ;  $\tau = 120$  ms.

Nous avons donc décidé de diminuer la température d'analyse afin de stabiliser la structure de la molécule en solution. L'attribution des signaux de résonance  $^1\text{H}$  a donc été à nouveau réalisée à 278K. Le changement de ce paramètre nous a permis d'identifier des corrélations structurantes  $^1\text{H}\text{-}^1\text{H}$  supplémentaires, autres que celles séquentielles. Ceci indique donc que le peptide est mieux structuré à plus basse température (Figure 52).

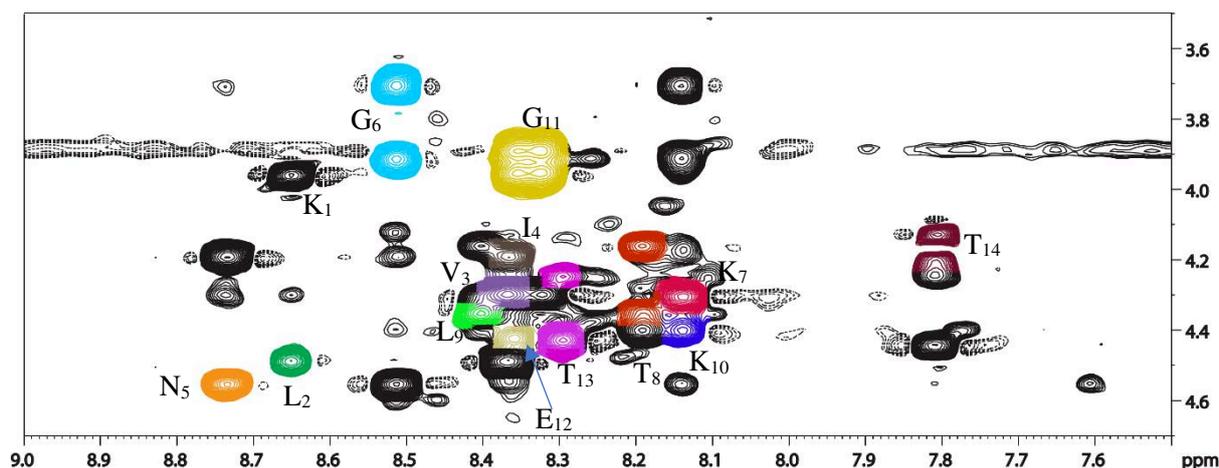


Figure 52 Zoom de la région  $H_N-H_\alpha$  du spectre NOESY du peptide IIGD à 278K dans un mélange de tampon phosphate/TFE (20 mM, 75/25 v/v) à pH 5,3 ; 600 MHz ;  $\tau = 120$  ms.

Nous avons ensuite souhaité réaliser les expériences RMN à un pH plus bas (3,5). Cette valeur est celle utilisée par Blanco et ses collaborateurs<sup>133</sup> pendant la caractérisation du même peptide plus long. L'analyse des cartes bidimensionnelles NOESY a permis d'observer des corrélations  $^1H-^1H$  similaires à celles déterminées précédemment. Nous avons donc choisi de continuer l'analyse de ce peptide à 278K et aux deux pH.

### Les déplacements chimiques secondaires

La Figure 53 présente les valeurs des déplacements chimiques secondaires (CSD) pour les protons  $\alpha$  du peptide IIGD aux deux pH (5,3 et 3,5) à 278K. L'absence d'une succession d'indices inférieurs à -0,1 ppm, caractéristiques d'une structuration en hélice, ou supérieurs à 0,1 ppm, caractéristiques d'une structuration en feuillet  $\beta$  ne permet pas de déterminer la nature de la structure secondaire de ce peptide. Nous pouvons toutefois noter que des valeurs supérieures à 0,1 ppm pour les acides aminés L2 et V3 ainsi que pour G12 et T13 sont compatibles avec l'éventuelle présence d'une structuration en brin  $\beta$ .

On peut également noter que le pH a une faible influence sur les déplacements chimiques  $H_\alpha$  suggérant que la structure secondaire du peptide n'est pas modifiée lors de la diminution du pH. Pour pouvoir obtenir des informations sur la structure secondaire adoptée par le peptide, nous avons analysé en détail les corrélations NOE.

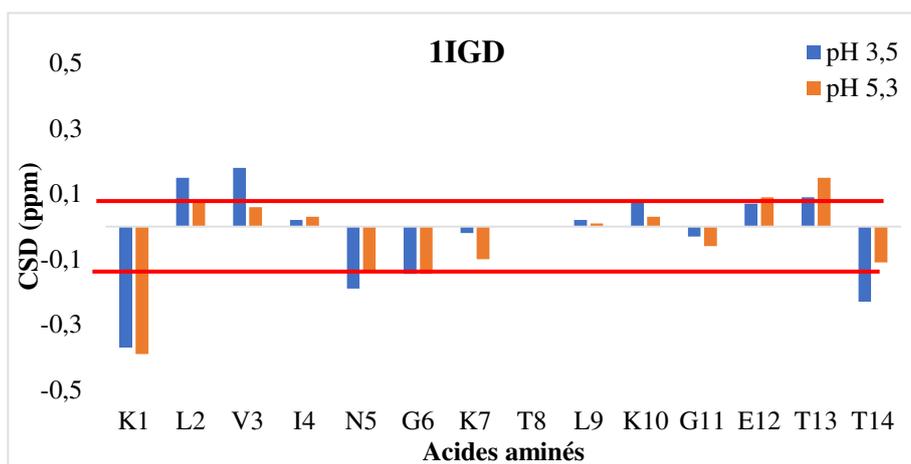


Figure 53 Déplacements chimiques secondaires (CSD) des protons  $H_{\alpha}$  du peptide 1IGD à pH 5,3 (bleu) et à pH 3,5 (orange) dans le mélange tampon phosphate/TFE (20 mM, 75/25 v/v) à 278K calculés à partir des valeurs de référence (Wishart et al., 1995).<sup>101</sup>

### Le diagramme NOE

L'analyse détaillée des spectres NOESY  $^1\text{H}$ - $^1\text{H}$  pour le peptide 1IGD dans le mélange tampon phosphate/TFE (75/25 v/v) à 278K et pH 5,3 a permis d'extraire des corrélations NOE. Grâce à ces corrélations NOE, il est possible de construire le diagramme NOE, qui permet d'avoir une idée initiale sur la structure secondaire du peptide (Figure 54). La présence d'effets séquentiels  $d_{\alpha\text{N}}$  forts nous suggère une probable structuration en deux brins de feuillet  $\beta$  entre les résidus K<sub>1</sub>-T<sub>4</sub> et L<sub>9</sub>-T<sub>14</sub> (Figure 54). En outre, les effets  $d_{\text{NN}}$ ,  $d_{\text{NN}}(i, i+2)$  et  $d_{\alpha\text{N}}(i, i+2)$  entre les acides aminés N<sub>5</sub>-K<sub>7</sub> et entre G<sub>6</sub>-T<sub>8</sub> suggèrent une structuration en conformation coude entre les deux brins de feuillet  $\beta$  (Figure 54). Nous ne retrouvons cependant pas les effets  $H_{\text{N}}\text{-}H_{\alpha}$  et  $H_{\alpha}\text{-}H_{\alpha}$  à longue distance, caractéristiques d'une conformation en feuillet  $\beta$ , ce qui montre que le motif  $\beta$ -hairpin n'est pas stabilisé dans ces conditions.

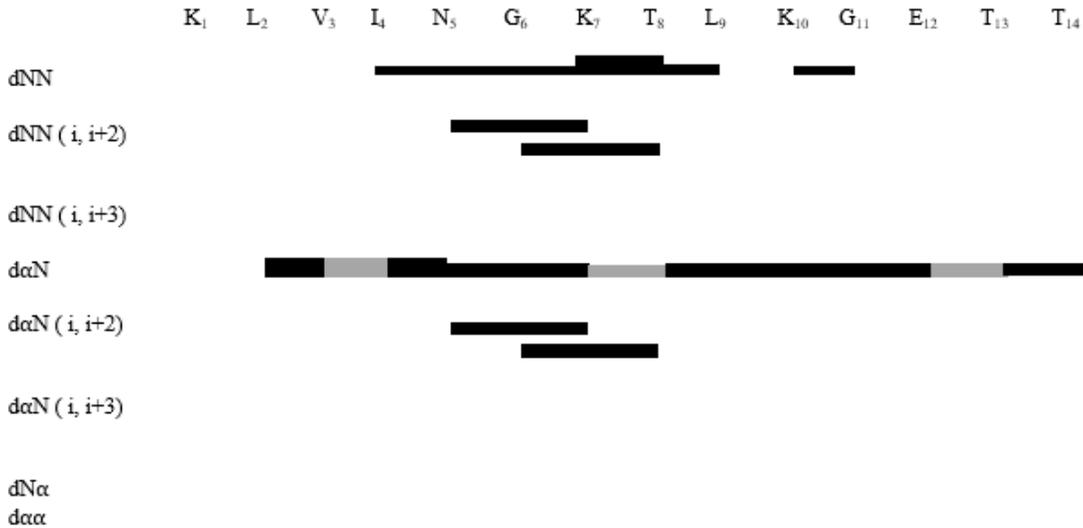


Figure 54 Diagramme des corrélations NOE séquentielles, à moyenne et longue distances obtenues pour le peptide IIGD dans le milieu tampon phosphate/TFE (20 mM, 75/25 v/v), pH 5,3 à 278K,  $\tau_m$  120 ms. Un trait épais correspond à corrélation NOE forte, un trait moyen correspond à une corrélation intermédiaire et un trait fin correspond à une corrélation NOE faible. Les NOE ambigus sont représentés en gris avec la même gamme d'intensité que les NOE certains.

L'analyse du peptide IIGD a également été réalisée à pH 3,5. Nous avons diminué le pH de la solution afin de ralentir la vitesse d'échange des protons amides avec l'eau et ainsi stabiliser la structure. Le changement de pH a permis de mieux stabiliser le motif  $\beta$ -hairpin puisque nous observons des effets  $d_{NN}(i, i+2)$  supplémentaires entre les acides aminés G<sub>6</sub>-T<sub>8</sub> et des effets  $d_{NN}(i, i+3)$  et  $d_{\alpha N}(i, i+3)$  entre N<sub>5</sub> et T<sub>8</sub> ont été observés dans cette condition, confirmant ainsi la possible présence d'un coude entre ces résidus (Figure 55).

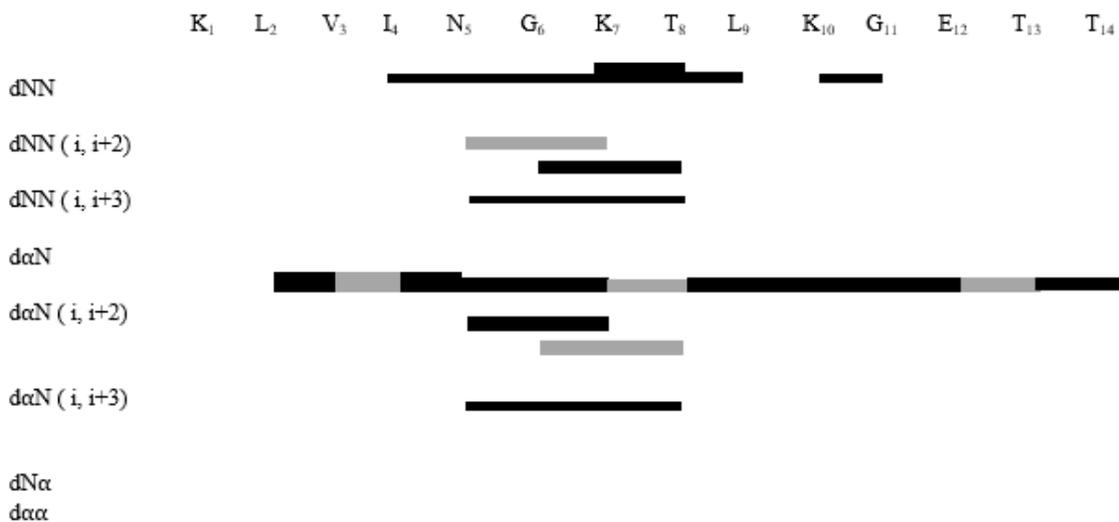


Figure 55 Diagramme des corrélations NOE séquentielles, à moyenne et longue distances obtenues pour le peptide IIGD dans le milieu tampon phosphate/TFE (20 mM, 75/25 v/v), pH 3,5 à 278K,  $\tau_m$  120 ms. Un trait épais correspond à corrélation NOE forte, un trait moyen correspond à une corrélation intermédiaire et un trait fin correspond à une corrélation NOE faible. Les NOE ambigus sont représentés en gris avec la même gamme d'intensité que les NOE certains.

Les paramètres structuraux obtenus sont en accord avec ceux reportés dans l'étude réalisée par Blanco et ses collaborateurs en 1994.<sup>133</sup> Lors de cette étude, le peptide plus long adopte une structuration «  $\beta$ -hairpin-like » comme dans la macromolécule à partir de laquelle il a été extrait. Cependant, dans notre cas, la suppression de quatre acides aminés des brins  $\beta$  a eu pour effet de déstabiliser ce motif structural en solution.

Nous avons alors essayé de modifier le milieu d'étude et d'analyser le peptide 1IGD en présence d'un mélange de tampon phosphate/MeOH (25/75 v/v) à 278K. Ce choix nous a été suggéré par les analyses de DC. Comme dans le mélange tampon phosphate/TFE, les mêmes corrélations à courte et moyenne distance ont été observées. Les NOE caractéristiques d'un motif  $\beta$ -hairpin ( $H_N-H_\alpha$  et  $H_\alpha-H_\alpha$  à longue distance) n'ont pas été repérés sur les cartes 2D NOESY. Ces résultats nous confirment que la structure du peptide en solution est trop flexible et que le motif  $\beta$ -hairpin n'est pas stabilisé.

Néanmoins, nous avons pu remarquer que, dans toutes les conditions, un coude était présent, ayant les acides aminés G<sub>6</sub>-K<sub>7</sub> comme centraux, alors que, dans la protéine native, les acides aminés K<sub>7</sub> et T<sub>8</sub> sont les centraux.

### **III.2.1.2 Le peptide 8ADH**

L'enregistrement des expériences RMN homonucléaires 2D <sup>1</sup>H-<sup>1</sup>H a été réalisé sur le peptide 8ADH dans un mélange tampon phosphate/TFE (50/50 v/v), pH 6,0 à 278K. Le choix de la condition expérimentale a été fait en se basant sur l'analyse par DC. L'analyse a été réalisée de manière similaire à celle du peptide 1IGD. L'attribution des cartes TOCSY, COSY et NOESY nous a permis d'extraire les paramètres structuraux (déplacements chimiques secondaires et effets NOE) afin de pouvoir déterminer le repliement en solution de ce peptide.

#### **Les déplacements chimiques secondaires**

La Figure 56 montre l'absence d'une succession d'indices inférieurs à -0,1 ppm, caractéristiques d'une structuration en hélice, ou supérieurs à 0,1 ppm, caractéristiques d'une structuration en feuillet  $\beta$  et donc ne permet pas de définir la nature de la structure secondaire.

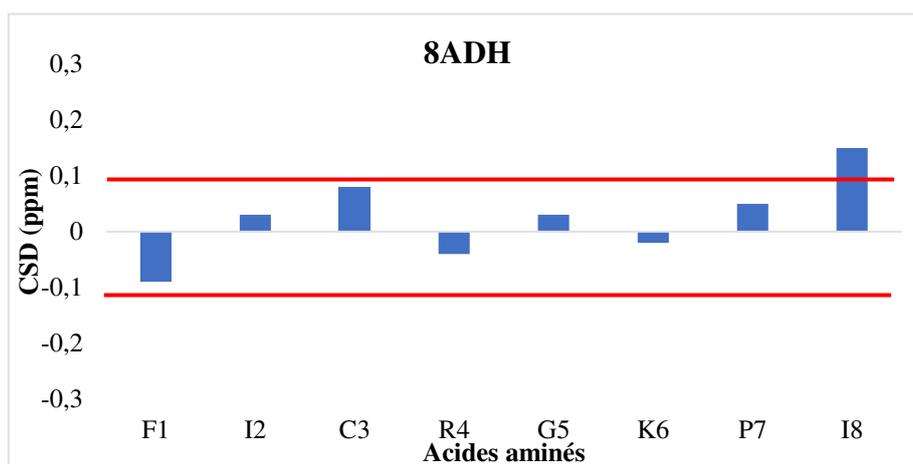


Figure 56 Déplacements chimiques secondaires des protons  $H_{\alpha}$  du peptide 8ADH à pH 6,0 dans le mélange tampon phosphate/TFE (20 mM, 50/50 v/v) à 278K à partir des valeurs de référence (Wishart et al., 1995).<sup>101</sup>

Ceci peut être dû au fait que le peptide est formé d'un nombre de résidus réduit et il n'est pas possible d'extraire des informations pertinentes à partir du diagramme de CSD lorsqu'on est en présence d'un peptide court.

### Le diagramme NOE

Le diagramme NOE a été construit à 278K dans un mélange tampon phosphate/TFE (50/50 v/v) à pH 6,0 (Figure 57). L'observation des seuls effets  $H_{\alpha}$ - $H_N$  séquentiels indique que cette molécule n'adopte pas de structuration secondaire précise en solution (Figure 57). Cependant, la présence d'effets NOE  $H_N$ - $H_N$  séquentiels entre R<sub>4</sub> et G<sub>5</sub> ainsi qu'entre G<sub>5</sub> et K<sub>6</sub> pourraient indiquer un début de structuration de type coude.

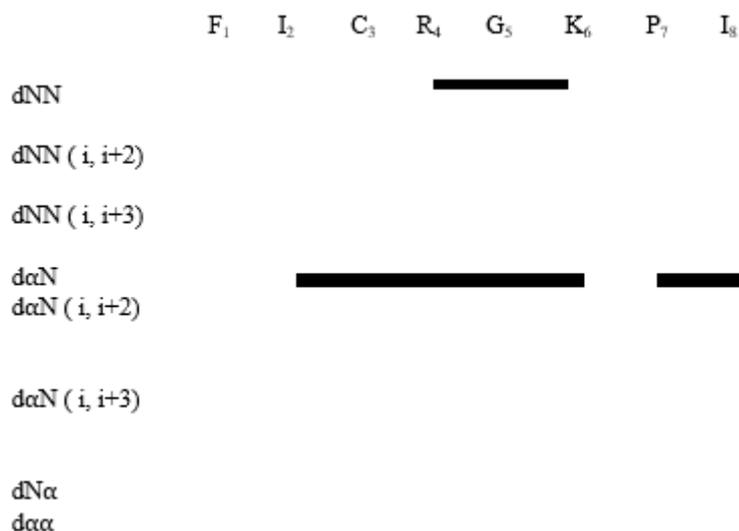


Figure 57 Diagramme des corrélations NOE séquentielles, à moyenne et longue distances obtenues pour le peptide 8ADH dans le milieu tampon phosphate/TFE (20 mM, 50/50 v/v), pH 6,0 à 278K,  $\tau_m$  120 ms. Un trait épais correspond à corrélation NOE forte, un trait moyen correspond à une corrélation intermédiaire et un trait fin correspond à une corrélation NOE faible.

### III.2.1.3 Le peptide 1MOL

1MOL est le dernier peptide dont nous avons déterminé les paramètres structuraux par RMN. Il s'agit d'un peptide de 11 acides aminés qui est apparemment structuré en conformation hélicoïdale selon les analyses de DC dans toutes les conditions, sauf peut-être dans le mélange tampon phosphate/MeOH (75/25 v/v).

Pour ce peptide, les expériences RMN 2D  $^1\text{H}$ - $^1\text{H}$  (TOCSY, COSY et NOESY) ont donc été enregistrées à 278K à pH 5,3 en présence de 25% de MeOH. Cette condition expérimentale a été choisie car les analyses de DC suggèrent que la molécule adopte dans ces conditions une plus faible hélicité, compatible avec la présence d'un mélange de structuration de type hélice et coude.

#### Les déplacements chimiques secondaires

Les valeurs des déplacements chimiques secondaires pour les protons  $\alpha$  du peptide 1MOL sont cohérentes avec les valeurs attendues pour une structuration en hélice (Figure 58). Au contraire des autres peptides, il est possible d'observer l'enchaînement de plusieurs valeurs de CSD inférieures à -0,1 ppm entre les acides aminés E5 et I10. Ceci semble confirmer que, même dans cette condition, le peptide présente une conformation hélicoïdale. En outre, les valeurs de CSD

indiquent que la structuration de type hélice pourrait également être présente dans les régions N- et C-terminales.

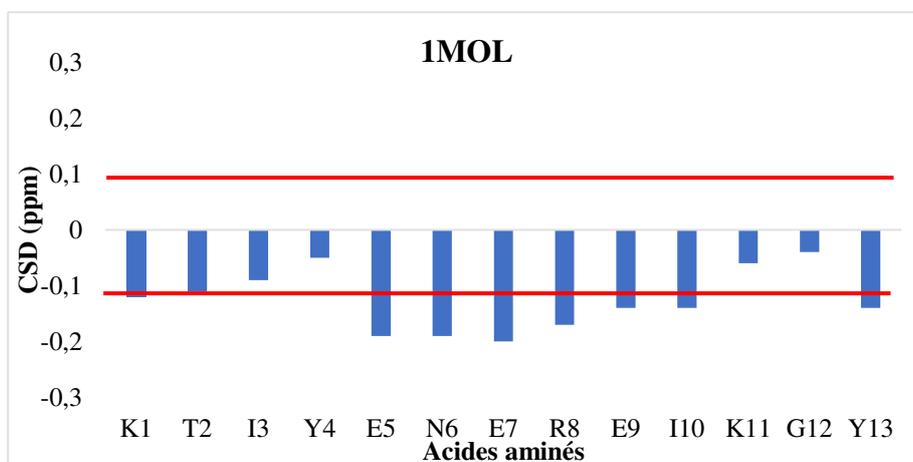


Figure 58 Déplacements chimiques secondaires des protons  $H_{\alpha}$  du peptide 1MOL aux deux pH 5,3 (bleu) et 3,5 (orange) dans le mélange tampon phosphate/ MeOH (20 mM, 75/25 v/v) à 278K à partir des valeurs de référence (Wishart et al., 1995).<sup>101</sup>

### Le diagramme NOE

Comme pour les deux premiers peptides modèles, le diagramme NOE a été tracé (Figure 59). L'analyse des spectres NOESY du peptide 1MOL en présence d'un mélange tampon phosphate/MeOH (20 mM) (75/25 v/v) à pH 5,3 et à 278K nous a permis de remarquer la présence d'effets NOE de type  $d_{\alpha N}(i, i+2)$ ,  $d_{\alpha N}(i, i+3)$ ,  $d_{NN}(i, i+2)$  et  $d_{\alpha\beta}(i, i+3)$  (Figure 59) caractéristiques d'une structure hélicoïdale. Plus dans le détail, sont présentes des effets  $d_{\alpha N}(i, i+2)$  entre les résidus T<sub>2</sub> et R<sub>8</sub>, un enchaînement d'effets  $d_{\alpha N}(i, i+3)$  entre les acides aminés T<sub>2</sub> et K<sub>11</sub> ainsi que des effets  $d_{NN}(i, i+2)$  entre les résidus Y<sub>4</sub> et G<sub>12</sub>. Enfin, des corrélations  $d_{\alpha\beta}(i, i+3)$  ont été relevées entre T<sub>2</sub> et I<sub>10</sub> (Figure 59).

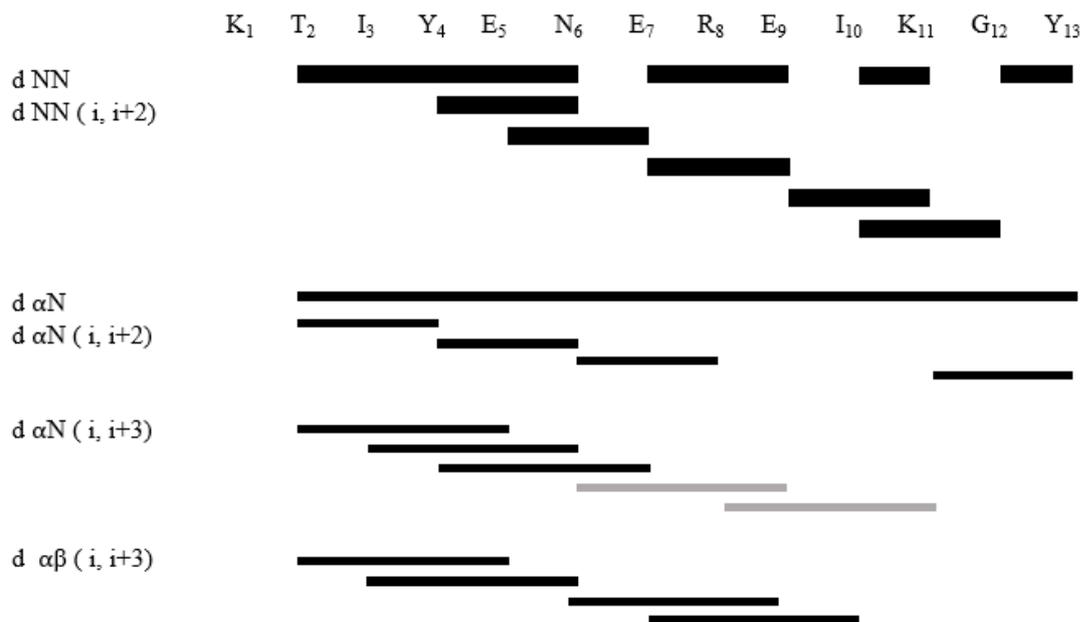


Figure 59 Diagramme des corrélations NOE séquentielles, à moyenne et longue distances obtenues pour le peptide 1MOL dans le milieu tampon phosphate/MeOH (20 mM, 75/25 v/v), pH 5,3 à 278K,  $\tau_m$  120 ms. Un trait épais correspond à corrélation NOE forte, un trait moyen correspond à une corrélation intermédiaire et un trait fin correspond à une corrélation NOE faible. Les NOE ambiguës sont représentés en gris avec la même gamme d'intensité que les NOE certains.

L'analyse qualitative du diagramme NOE du peptide 1MOL permet donc d'identifier une structuration de type hélice entre les acides aminés T<sub>2</sub> et K<sub>11</sub> dans ces conditions expérimentales. Ceci confirme que dans le mélange tampon phosphate/MeOH (75/25 v/v) ce peptide adopte en solution un repliement différent de celui attendu en  $\beta$ -hairpin.

### III.3 Conclusions

L'étude expérimentale indique que les peptides modèles choisis n'adoptent pas la structuration attendue «  $\beta$ -hairpin ». Les résultats obtenus en analyse structurale par RMN de ces peptides sont probablement dus au fait que les peptides choisis ne sont pas suffisamment longs en termes de résidus pour permettre le repliement de la molécule dans la structuration attendue. En effet, il faudrait rallonger les brins  $\beta$  pour aider les acides aminés à créer les liaisons hydrogène inter-brins qui normalement stabiliseraient le motif  $\beta$ -hairpin. Cependant, cette stratégie ne paraît pas indiquée dans le cadre de notre étude, car l'ajout d'acides aminés dans les séquences choisies pourrait avoir comme conséquence la présence de superpositions sur les spectres RMN, ne permettant pas l'identification des effets NOE caractéristiques du coude. L'un des autres problèmes de ces expériences pourrait être le milieu choisi lors de l'étude structurale par

DC et RMN. Des études montrent que l'utilisation de co-solvants organiques comme le TFE et le MeOH devraient aider à la formation de liaisons hydrogène qui permettraient de stabiliser le peptide de petite taille moléculaire.<sup>134,138</sup> Cependant, dans notre étude, nous avons pu observer que ces solvants n'apportent qu'une structuration partielle de nos peptides modèles : ils ne permettent pas de stabiliser la structure secondaire comme dans la protéine native. Une autre stratégie qui pourrait être envisagée serait la cyclisation de nos peptides modèles à travers l'ajout de cystéines surnumériques aux extrémités N- et C-terminales de la molécule. Ceci permettrait aux peptides d'adopter la structuration en feuillet  $\beta$ , mais les techniques de synthèse sophistiquées et le temps pour les réaliser sont des contraintes importantes. Par ailleurs, des études ont montré que les modifications au niveau des brins  $\beta$  dans des motifs structuraux de type  $\beta$ -hairpin peuvent impacter la nature et le type de coude.<sup>139-142</sup> En l'absence d'un repliement correct des peptides modèles, il n'était pas envisageable poursuivre cette étude en utilisant cette approche expérimentale et nous avons donc décidé de réorienter l'étude vers un côté théorique.

## IV Matériels et méthodes

### IV.1 Dichroïsme circulaire

#### IV.1.1 Condition d'acquisition et de traitement

Les expériences de dichroïsme circulaire pour les peptides modèles 1IGD, 8ADH, 2SGA, 1GND et 1MOL ont été enregistrées sur un appareil MOS-500 (Biologic) avec des cuves rectangulaires de trajet optique égal à 1 mm. Les acquisitions ont été effectuées entre les longueurs d'ondes ( $\lambda$ ) 190 et 260 nm par pas de 1 nm toutes les 2 secondes et à température ambiante. Chaque spectre est la moyenne de 3 scans.

Pour tous les spectres, la contribution du solvant a été corrigée par soustraction du spectre du solvant seul (blanc) pour chaque échantillon. Pour pouvoir comparer les valeurs indépendamment des biais engendrés par le trajet optique et par la concentration, les valeurs ont été exprimées en ellipticité molaire moyenne par résidu selon l'équation suivante :

$$\theta_{molaire} (deg.cm^2.dmol^{-1}) = \frac{100 \theta_{mesuré} (deg)}{lcn} \quad (3.1)$$

où  $\theta$  mesuré est la mesure expérimentale de l'ellipticité (en degrés),  $l$  est le trajet optique de la cuve (en cm),  $c$  la concentration du peptide (en mol/L) et  $n$  le nombre d'acides aminés du peptide.

#### IV.1.2 Préparation des échantillons

Les échantillons de peptides ont tous été préparés de la même manière pour les études structurales par DC. Une solution mère de tampon phosphate (mélange de  $K_2HPO_4$  et  $KH_2PO_4$ , 20 mM, pH 6) a été réalisée. Le pH du tampon phosphate a été adapté aux points isoélectriques (pI) de chaque peptide en ajoutant du HCl ou du NaOH 0,1M (Tableau 16).

*Tableau 16 Valeurs du poids moléculaire, du point isoélectrique et du pH utilisé pour les expériences de DC pour chacun des peptides modèles choisis.*

Code PDB	Poids moléculaire	Point isoélectrique (pI)	pH utilisé
<b>1IGD</b>	1501,79	9,70	5,3
<b>8ADH</b>	933,186	9,51	6,0
<b>2SGA</b>	923,034	5,24	3,5
<b>1GND</b>	817,89	4,53	5,3
<b>1MOL</b>	1642,83	6,23	5,3

Les analyses ont été faites dans trois solutions différentes : (i) un tampon phosphate (20 mM, pH 3,5 à 6,0), (ii) un mélange de tampon phosphate et TFE (75/25, 50/50, 25/75) et (iii) un mélange de tampon phosphate et MeOH (75/25, 50/50, 25/75).

Des solutions mères de peptides ont été réalisées en dissolvant 150  $\mu$ g de peptide dans 300  $\mu$ L de tampon phosphate afin d'avoir une concentration de 0,05 mg/mL. Les différents échantillons ont été préparés en prélevant un volume de 100  $\mu$ L de solution mère et en y ajoutant 75  $\mu$ L, 150  $\mu$ L et 225  $\mu$ L pour avoir les pourcentages souhaités de 25%, 50% ou 75% de co-solvant organique dans la solution complétée à 300  $\mu$ L avec du tampon phosphate.

### IV.2 Résonance Magnétique Nucléaire

#### IV.2.1 Conditions d'acquisition et de traitement

L'analyse structurale par RMN des peptides modèles a été effectuée dans une solution (550  $\mu$ L) de (i) mélange tampon phosphate (20 mM) /TFE<sub>d2</sub> et (ii) tampon phosphate (20 mM) /MeOH<sub>d3</sub>. La concentration de peptide de chaque échantillon était de 0,5-1 mM. 2  $\mu$ L de 2,2-

diméthyl-2-silapentane-5-sulfonate (DSS) ont été ajoutés comme référence interne sur les échantillons une fois la concentration finale obtenue (Tableau 17).

Tableau 17 Valeurs des concentrations et conditions expérimentales utilisées pour les expériences de RMN pour chacun des peptides modèles IIGD, 8ADH et 1MOL.

Peptide	Concentration (mM)	Milieu	pH	Température (K)
<b>IIGD</b>	0,45	Tampon phosphate/TFE <sub>d2</sub> (75/25 v/v)	5,3	298
<b>IIGD</b>	0,40	Tampon phosphate/TFE <sub>d2</sub> (75/25 v/v)	3,5	278
<b>8ADH</b>	0,47	Tampon phosphate/TFE <sub>d2</sub> (50/50 v/v)	6,0	278
<b>1MOL</b>	0,45	Tampon phosphate/MeOH <sub>d3</sub> (75/25 v/v)	5,3	278

#### IV.2.2 Paramètres d'acquisition des spectres RMN des peptides modèles

Toutes les expériences RMN ont été enregistrées sur un spectromètre BRUKER AVANCE III (600 MHz, Wissembourg, France) équipé d'une cryosonde (CPTXI (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N), gradient z). L'utilisation d'une cryosonde permet d'améliorer la qualité des spectres en augmentant le rapport signal/bruit.

Les expériences RMN 1D <sup>1</sup>H et homonucléaires 2D <sup>1</sup>H-<sup>1</sup>H (TOCSY, COSY et NOESY) ont été enregistrées à deux températures différentes : 298K et 278K pour effectuer l'attribution des peptides. Les expériences homonucléaires 2D <sup>1</sup>H-<sup>1</sup>H TOCSY et NOESY ont été enregistrées avec 'excitation sculpting water suppression', alors que pour les expériences COSY un 'low power pre-saturation' a été utilisé. Le temps de mélange ( $\tau_m = 80$  ms) a été utilisé dans les expériences 2D TOCSY. Les expériences NOESY ont été réalisées avec des temps de mélange différents de 120 et 150 ms. Les spectres ont été collectés avec 512 et 2048 points de données complexes dans les dimensions  $t_1$  et  $t_2$ . La fenêtre spectrale était de 11,0035 ppm x 11,0035 ppm. Pour les expériences homonucléaires 2D COSY (cosygpprpf) le nombre de scan (NS) était de 32, pour les expériences TOCSY (dipsi2esgpph) le NS utilisé était de 48 et un NS de 48 ou 64 pour les NOESY (noesyegpph). Les spectres ont été traités par la suite avec le logiciel TOPSPIN 3.2 (Bruker).

# **Chapitre 4**

## **Recherche de paramètres structuraux de RMN des coudes $\beta$ et $\gamma$ par modélisation moléculaire**



## ***Chapitre 4 : Recherche de paramètres structuraux de RMN des coudes $\beta$ et $\gamma$ par modélisation moléculaire***

Dans la première partie de ce travail, nous avons essayé de caractériser les coudes  $\beta$  de type I, I', II et II' ainsi que les coudes  $\gamma$  de type classique et inverse avec une étude expérimentale de peptides modèles en conformation  $\beta$ -hairpin issus de structures cristallographiques de protéines trouvées dans la PDB. Cependant, nous ne sommes pas arrivés à stabiliser le motif de choix en solution. Ceci ne nous a donc pas permis de caractériser les coudes  $\beta$  de type I, I', II et II' ni les coudes  $\gamma$  de type classique et inverse.

Pour ces raisons, nous avons alors choisi une approche théorique, en utilisant la modélisation moléculaire sous contraintes.

Afin de réaliser cette partie de l'étude, nous avons constitué une banque de données des différents types de coudes  $\beta$  et  $\gamma$  à partir de structures cristallines de la PDB. Une fois les structures identifiées à l'aide d'une analyse des paramètres géométriques, des peptides (4 acides aminés pour les coudes  $\beta$  et 3 acides aminés pour les coudes  $\gamma$ ) ont été extraits de la macromolécule native. Leur structure a ensuite été soumise à un protocole de recuit simulé sous contraintes de distance et angulaires pour optimiser leur géométrie. Les paramètres géométriques (angles dièdres  $\Phi$ ,  $\Psi$  et distances inter-hydrogènes) des structures obtenues ont alors été analysés. Les distances mesurées ont été utilisées pour générer des diagrammes NOE pour chaque type de coude. Enfin, à partir des angles dièdres  $\Phi$  et en utilisant les équations disponibles dans la littérature, nous avons calculé les valeurs de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  pour les résidus centraux des coudes  $\beta$  et  $\gamma$ .

Ce chapitre sera divisé en trois parties. D'abord, j'expliquerai les méthodes mises en œuvre pour la constitution de la banque de données de peptides, comment la géométrie des peptides sélectionnés a été optimisée et le choix du jeu de structures finales utilisé pour l'analyse. Je présenterai ensuite comment ont été obtenus les paramètres RMN pour les coudes de type  $\beta$  et  $\gamma$ . Enfin, je conclurai sur l'ensemble des résultats obtenus et les perspectives de cette étude.

### **I. Méthodes mises en œuvre**

#### **I.1 Constitution de la base de données de peptides**

Pour l'étude des paramètres structuraux de RMN, la base de données de coudes  $\beta$  et  $\gamma$  a été construite à l'aide de structures présentes dans la PDB (<http://www.rcsb.org/>)<sup>131</sup>. Une recherche de structures a été réalisée afin d'obtenir des structures de tous les types de coude  $\beta$  (I, I', II, II',

III, III', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb et VIII) et  $\gamma$  (classique et inverse). Les structures sélectionnées étaient composées que de 4 acides aminés pour les coudes  $\beta$  et de 3 acides aminés pour les coudes  $\gamma$ .

Pour la construction de cette banque de données, seules les structures résolues par cristallographie aux rayons X ont été prises en compte. Ce choix a été fait car peu de structures de type coude dont la nature est connue de manière précise ont été résolues par RMN en solution et leur utilisation pourrait constituer un biais dans notre base de données. Ces structures devaient présenter une résolution correcte (entre 1,0 et 2,8 Å) et une valeur du facteur-R basse (inférieur à 0,26). Les codes PDB ont été sélectionnés à partir de la littérature, dans des études statistiques sur les coudes<sup>37</sup> ou des études structurales mentionnant la présence de coudes dans la structure étudiée.<sup>127,129</sup> Une vérification des structures a été réalisée car la nature et le type de coude annoncés dans ces études n'étaient pas toujours ceux retrouvés dans les structures cristallographique des protéines déposées dans la PDB.

Ainsi, les motifs de coudes  $\beta$  et  $\gamma$  ont été sélectionnés selon les critères établis par Venkatachalam en 1968<sup>38</sup> pour définir un coude : (i) la distance entre les atomes C $_{\alpha}$  des résidus  $i$  et  $i+n$  devait être inférieure à 7Å (où  $n$  est égal à 2 pour les coudes  $\gamma$  et 3 pour les coudes  $\beta$ ) ; (ii) le résidu central ou les résidus centraux ne devaient pas être en conformation hélicoïdale ou trop étendue, donc les angles dièdres  $\Phi$  et  $\Psi$  ne devaient pas posséder respectivement des valeurs d'angles de  $-57^{\circ}$  et  $-47^{\circ}$  ; (iii) la liaison hydrogène entre le groupe carboxylique du premier résidu et le groupe amide du dernier résidu devait être présente.

La géométrie du premier et du dernier résidu peut avoir un impact sur les paramètres de RMN que nous souhaitons déterminer. Une analyse des angles dièdres  $\Phi$  et  $\Psi$  des résidus  $i$  et  $i+n$  (avec  $n = 3$  ou  $2$ ) de la structure en conformation coude a donc été menée afin de caractériser le motif ou les motifs l'entourant.

Nous avons identifié neuf motifs structuraux :

- Brin-coude-brin (S-T-S),
- Hélice-coude-hélice (H-T-H),
- Hélice-coude-brin (H-T-S),
- Brin-coude-hélice (S-T-H),
- Hélice-coude-conformation aléatoire (H-T-R),
- Conformation aléatoire-coude-hélice (R-T-H),
- Brin-coude-conformation aléatoire (S-T-R),
- Conformation aléatoire-coude-brin (R-T-S),

- Conformation aléatoire-coude-conformation aléatoire (R-T-R).

Pour les différentes natures de coude, les angles dièdres  $\Phi$  et  $\Psi$  des deux acides aminés centraux (pour les  $\beta$ ) et de l'acide aminé central (pour les  $\gamma$ ) ont été examinés afin de ne retenir que des structures dans lesquelles leurs valeurs ne s'écartaient pas plus de  $\pm 30^\circ$  par rapport aux valeurs des angles dièdres canoniques (Tableau 10, chapitre 1).<sup>45</sup>

Enfin, afin de constituer une banque de données représentative, une analyse des séquences en acides aminés a également été réalisée pour tenir en compte des préférences de position mais aussi, si possible, diversifier les séquences pour ne pas réduire les paramètres de RMN à analyser.<sup>37,45</sup> Notamment, des peptides contenant une proline ont été sélectionnés car en RMN l'identification des différents motifs structuraux est réalisée grâce aux distances impliquant les protons amides et cet acide aminé a la particularité de ne pas avoir de proton amide. Or, la proline est souvent rencontrée dans certains coudes. Ainsi, dans la littérature, il a été montré que la proline est l'acide aminé le plus fortement observé en position  $i+1$  dans les coudes  $\beta$  de type I et II, III et dans les coudes de nature  $\gamma$  de type inverse.<sup>37,41</sup> La rigidité de cet acide aminé cyclique ne permet à ce résidu qu'une valeur de  $-60^\circ$  pour l'angle  $\Phi$ . Dans les images miroirs (I', II' et III'), la proline n'est pas présente en position  $i+1$ , car la valeur de l'angle  $\Phi$  ( $+60^\circ$ ) ne peut pas être accessible à ce résidu.<sup>37,41</sup> De plus, pour tous les coudes  $\beta$  de types VIa1, VIa2 et VIb, une proline en position  $i+2$  est un requis nécessaire et elle doit être en conformation *cis*.<sup>44</sup> Au final, la proline peut donc se trouver préférentiellement en position  $i+1$  en conformation *trans* dans les coudes  $\beta$  de type I, II, III et de nature  $\gamma$  inverse et en position  $i+2$  en conformation *cis* dans les coudes  $\beta$  de type VI.

Le Tableau 18 résume les séquences des coudes  $\beta$  et  $\gamma$  retenues sur la base de ces critères. Nous avons essayé d'avoir un nombre de séquence en accord avec la fréquence des différents coudes dans les structures cristallographiques, tout en ayant une diversité de séquence pour chaque type de coude. Ce choix a eu pour effet de modifier la représentativité de certains types de coude car nous avons dû ajouter des séquences afin d'avoir une variabilité en acides aminés au sein de nos structures.

Un total de 104 peptides a été sélectionné. Les coudes  $\beta$  de type I, de type II et VIII sont les plus fréquents, suivis par les types I', II', III et III' qui ont une fréquence moyenne. Les coudes  $\beta$  de type VI sont représentés par un nombre limité de structures car ils sont un type très peu répandu. Pour les coudes  $\beta$  de type IV (IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>), nous avons essayé d'avoir un nombre similaire pour chaque catégorie. Enfin, les coudes  $\gamma$  sont des structures qui ont été étudiées de manière différente par les biochimistes structuraux. Ils sont retrouvés de façon

routinière dans les études RMN de peptides, peut-être influencées par le fait que les valeurs  $\Phi_{i+1}$  et  $\Psi_{i+1}$  du type classique ( $+75^\circ$ ,  $-64^\circ$ ) et de l'image-miroir  $\gamma$  inverse ( $-79^\circ$ ,  $+69^\circ$ ) sont les minimas globaux normalement obtenus pour les calculs d'énergie dans le vide des structures tridimensionnelles sous contraintes.<sup>143</sup> En revanche, ces structures sont évitées autant que possible, par les cristallographes de protéines, car les zones correspondantes aux angles  $\Phi_{i+1}$  et  $\Psi_{i+1}$  des coudes de nature  $\gamma$ , dans le diagramme de Ramachandran, sont classées comme interdites par PROCHECK (logiciel qui contrôle la qualité stéréochimique des structures protéiques résolues par cristallographie aux rayons X).<sup>144,145</sup> Nous avons sélectionné quatre structures pour les coudes  $\gamma$  classique et cinq pour les coudes  $\gamma$  inverse. Parmi ces dernières, deux séquences contenant une proline *trans* ont été incluses à notre base de données.

Dans sa classification, Venkatachalam avait inclus les coudes  $\beta$  de types III et III', mais ils ont ensuite été écartés parce qu'ils étaient considérés comme trop proches des hélices  $3_{10}$  et des coudes  $\beta$  de type I (et I'). Si les coudes  $\beta$  de type III et III' étaient encore reconnus, ils représenteraient la troisième catégorie la plus fréquemment rencontrée dans les chaînes polypeptidiques.<sup>45</sup> Si leur classification reste ambiguë en se basant sur les angles dièdres  $\Phi$  et  $\Psi$ , nous avons choisi de les inclure dans notre base de données car, en RMN, ils pourraient présenter des signatures caractéristiques qui les différencient des hélices  $3_{10}$  et des coudes  $\beta$  de type I (et I').<sup>64</sup>

Par ailleurs, dans le Tableau 18, on peut noter que tous les motifs identifiés ne sont pas présents de manière équilibrée. Ainsi, pour les coudes  $\beta$  de type I', seul le motif  $\beta$ -hairpin (ou S-T-S) est présent alors que pour les coudes  $\gamma$  inverse, on observe une plus grande diversité. Ceci est dû au fait que les caractéristiques géométriques et structurales des deux acides aminés centraux des différents types de coudes influencent le motif dans lequel ils se trouvent. Ceci implique que certains motifs soient plus représentés que d'autres pour certains types de coude (Tableau 18). Par exemple, les coudes  $\beta$  de type I', II' et III' sont retrouvés seulement dans le motif structural «  $\beta$ -hairpin » (Tableau 18). Ceci est dû au fait que leur structure est planaire et ils sont la conformation privilégiée pour relier deux brin  $\beta$  antiparallèles.

Tableau 18 Résumé des séquences des peptides sélectionnés pour les coudes de type  $\beta$  et  $\gamma$ . Les prolines ont été mises en gras et soulignées. Pour les structures 1ECA et 2CNA la valeur du facteur R n'a pas été retrouvé dans la PDB. Cette absence a été indiquée avec le symbole « - ».

#	Code PDB	Résolution (Å)	Facteur-R	Type de coude	Séquence	Motifs	Ref.
1	1BPI	1.09	0.187	I	A-K-A-G	S-T-S	Parkin <i>et al.</i> , 1996
2	1CYO	1.50	0.160	I	R-E-Q-A	H-T-H	Mathews <i>et al.</i> , 1996
3	1ECA	1.40	-	I	V-K-G-D	H-T-H	Steigemann <i>et al.</i> , 1979
4	1ECA	1.40	-	I	I-K-G-T	H-T-H	Steigemann <i>et al.</i> , 1979
5	1HLV	2.50	0.222	I	N- <u>P</u> -D-L	H-T-H	Tanaka <i>et al.</i> , 2010
6	1IGD	1.10	0.193	I	G-K-T-L	S-T-S	Derrick <i>et al.</i> , 1994
7	1JQZ	1.65	0.204	I	L- <u>P</u> -D-G	S-T-H	Brych <i>et al.</i> , 2001
8	1YT3	1.60	0.196	I	D- <u>P</u> -S-I	H-T-S	Zuo <i>et al.</i> , 2005
9	2CAB	2.00	0.193	I	K-A-D-G	H-T-S	Kannan <i>et al.</i> , 1984
10	2LHB	2.00	0.142	I	S-M-D-D	H-T-H	Honzatko <i>et al.</i> , 1985
1	1GVM	2.80	0.226	I'	I-A-D-K	S-T-S	Fernandez-Tonero <i>et al.</i> , 2002
2	1GVM	2.80	0.226	I'	Y-K-D-T	S-T-S	Fernandez-Tonero <i>et al.</i> , 2002
3	1GVM	2.80	0.226	I'	I-N-G-T	S-T-S	Fernandez-Tonero <i>et al.</i> , 2002
4	1TLX	2.10	0.153	I'	H-Y-G-V	S-T-S	English <i>et al.</i> , 1999
5	2ERF	1.45	0.186	I'	V-Q-G-K	S-T-S	Tan <i>et al.</i> , 2006
6	3APP	1.80	0.126	I'	V-G-G-V	S-T-S	James <i>et al.</i> , 1983
1	1A1X	2.00	0.211	II	Q-E-G-I	S-T-S	Fu <i>et al.</i> , 1998
2	1A1X	2.00	0.211	II	Y- <u>P</u> -E-E	S-T-S	Fu <i>et al.</i> , 1998
3	1A7T	1.85	0.151	II	I-E-G-W	S-T-S	Fitzgerald <i>et al.</i> , 1998
4	1JYK	1.50	0.218	II	L-A-N-S	H-T-S	Kwak <i>et al.</i> , 2002
5	1JYK	1.50	0.218	II	G-Y-L-K	S-T-H	Kwak <i>et al.</i> , 2002
6	1PEG	2.59	0.220	II	N-A-A-D	S-T-R	Zhang <i>et al.</i> , 2003
7	1PGS	1.80	0.168	II	R-A-G-W	R-T-R	Norris <i>et al.</i> , 1994
8	1PMI	1.70	0.184	II	F- <u>P</u> -N-D	H-T-H	Cleasby <i>et al.</i> , 1996
9	2CPL	1.63	0.18	II	I- <u>P</u> -G-F	S-T-S	Ke <i>et al.</i> , 1992
10	3N7S	2.10	0.209	II	A-A-G-T	S-T-S	Ter Haar <i>et al.</i> , 2010

11	8PAZ	1.60	0.178	II	N- <u>P</u> -G-D	S-T-S	Libeu <i>et al.</i> , 1997
1	1CBR	2.80	0.251	II'	D-G-D-Q	S-T-S	Kleywegt <i>et al.</i> , 1994
2	1EMA	1.90	0.175	II'	E-G-D-T	S-T-S	Ormo <i>et al.</i> , 1996
3	1UXA	1.50	0.180	II'	S-G-N-S	S-T-S	Burmeister <i>et al.</i> , 2004
4	2NUZ	1.85	0.176	II'	V-N-D-R	S-T-S	Argarwal <i>et al.</i> , 2006
5	3JZG	2.10	0.161	II'	G-S-N-R	S-T-S	Xu <i>et al.</i> , 2010
6	3KMT	1.78	0.196	II'	R-K-N-M	S-T-S	Wei <i>et al.</i> , 2010
7	3SEB	1.48	0.182	II'	E-N-E-N	S-T-S	Papageorgiou <i>et al.</i> , 1998
1	1CYO	1.50	0.160	III	S-K-T-F	H-T-S	Mathews <i>et al.</i> , 1996
2	1PCY	1.33	0.150	III	Q-G-A-G	H-T-R	Guss <i>et al.</i> , 1998
3	3KVU	2.00	0.220	III	Y- <u>P</u> -E-S	H-T-H	Dias <i>et al.</i> , 2010
4	2AZA (B)	1.80	0.157	III	K-E-A-D	R-T-H	Baker, 1998
5	4A7J	1.90	0.184	III	N- <u>P</u> -Q-S	S-T-S	Migliori <i>et al.</i> , 2012
6	5CH1	2.27	0.170	III	H-E-K-Q	H-T-S	Jiao <i>et al.</i> , 2015
7	5CH1	2.27	0.170	III	N-V-T-E	S-T-S	Jiao <i>et al.</i> , 2015
1	1A8D	1.57	0.187	III'	Y-N-N-N	S-T-S	Knapp <i>et al.</i> , 1998
2	1BP2	1.70	0.171	III'	S-N-N-E	S-T-S	Dijkstra <i>et al.</i> , 1981
3	1DFQ	2.60	0.200	III'	N-N-D-I	S-T-S	Emsley <i>et al.</i> , 2000
4	1IFB	1.96	0.188	III'	Y-E-G-V	S-T-S	Sacchettini <i>et al.</i> , 1989
5	1OBP	2.00	0.164	III'	R-D-G-K	S-T-S	Tegoni <i>et al.</i> , 1996
6	1TGN	1.65	0.223	III'	C-S-G-K	S-T-S	Kossiakoff <i>et al.</i> , 1996
7	1UBQ	1.80	0.176	III'	F-A-G-K	S-T-S	Vijay-Kumar <i>et al.</i> , 1996
8	3RN3	1.45	0.223	III'	W-D-G-K	S-T-S	Howlin <i>et al.</i> , 1989
1	1EMA	1.90	0.175	IV <sub>1</sub>	D-V-N-G	S-T-S	Ormo <i>et al.</i> , 1996
2	1JYK	1.50	0.218	IV <sub>1</sub>	C-T-N-E	R-T-S	Kwak <i>et al.</i> , 2002
3	1JYK	1.50	0.218	IV <sub>1</sub>	Q-V-N-Q	S-T-S	Kwak <i>et al.</i> , 2002
4	1KKO	1.33	0.162	IV <sub>1</sub>	R-I-D-G	S-T-S	Levy <i>et al.</i> , 2002
5	1KKO	1.33	0.162	IV <sub>1</sub>	G-A-G-G	R-T-R	Levy <i>et al.</i> , 2002
6	1QWG	1.60	0.170	IV <sub>1</sub>	G-K-G-K	R-T-R	Wise <i>et al.</i> , 2003

7	2ERF	1.45	0.186	IV <sub>1</sub>	M-K-K-T	S-T-S	Tan <i>et al.</i> , 2006
1	1A8D	1.57	0.187	IV <sub>2</sub>	G-I-N-G	S-T-S	Knapp <i>et al.</i> , 1998
2	1CBR	2.80	0.251	IV <sub>2</sub>	N-E-N-K	S-T-S	Kleywegt <i>et al.</i> , 1994
3	1DFQ	2.60	0.200	IV <sub>2</sub>	G-I-N-G	S-T-S	Emsley <i>et al.</i> , 2000
4	2AZA (A)	1.80	0.157	IV <sub>2</sub>	K-E-A-D	R-T-H	Baker, 1998
5	5L21	1.68	0.1931	IV <sub>2</sub>	G-S-K-V	S-T-R	Yao <i>et al.</i> , 2016
6	3GRS	1.54	0.186	IV <sub>2</sub>	T-L-S-G	S-T-S	Karplus <i>et al.</i> , 1987
1	1EMA	1.90	0.175	IV <sub>3</sub>	N-I-L-G	S-T-S	Ormo <i>et al.</i> , 1996
2	1EMA	1.90	0.175	IV <sub>3</sub>	V-T-T-F	S-T-S	Ormo <i>et al.</i> , 1996
3	1PA7	1.45	0.174	IV <sub>3</sub>	D- <u>P</u> -V-A	R-T-R	Hayashi <i>et al.</i> , 2003
4	1QJP	1.65	0.155	IV <sub>3</sub>	N-N-V-Y	S-T-S	Pautsch <i>et al.</i> , 2000
5	1SU8	1.10	0.146	IV <sub>3</sub>	A-A-E-A	S-T-R	Dobbek <i>et al.</i> , 2004
6	1URS	1.45	0.210	IV <sub>3</sub>	K-K-L-V	R-T-R	Schäfer <i>et al.</i> , 2004
7	2CNA	2.00	-	IV <sub>3</sub>	S-V-D-K	S-T-S	Reeke <i>et al.</i> , 1975
1	1CPS	2.25	0.174	IV <sub>4</sub>	G-N-R-G	S-T-S	Cappalonga <i>et al.</i> , 1992
2	1QWG	1.60	0.174	IV <sub>4</sub>	A-K-N-V	H-T-R	Wise <i>et al.</i> , 2003
3	1TLX	2.10	0.199	IV <sub>4</sub>	N-G-S-E	S-T-S	English <i>et al.</i> , 1999
4	3W95	1.85	0.2189	IV <sub>4</sub>	C-Q-H-G	S-T-S	Mu <i>et al.</i> , 2013
5	5CH1	2.27	0.170	IV <sub>4</sub>	D-D-Y-K	R-T-R	Jiao <i>et al.</i> , 2015
6	5VA6	2.40	0.249	IV <sub>4</sub>	S-R-E-G	R-T-S	Bergamin <i>et al.</i> , 2017
1	1ECA	1.40	-	VIa1	E-L- <u>P</u> -N	H-T-H	Steigemann <i>et al.</i> , 1979
2	1JYK	1.50	0.218	VIa1	L-R- <u>P</u> -L	R-T-R	Kwak <i>et al.</i> , 2002
3	2FBJ	1.95	0.194	VIa1	T-Y- <u>P</u> -L	S-T-S	Bhat <i>et al.</i> , 1988
4	5RSA	2.00	0.159	VIa1	K-Y- <u>P</u> -N	S-T-S	Wlodawer <i>et al.</i> , 1986
1	1EL1	1.90	0.159	VIa2	K-D- <u>P</u> -N	R-T-R	Koshiba <i>et al.</i> , 2000
2	1IQ6	1.50	0.203	VIa2	Q-L- <u>P</u> -G	H-T-R	Hisano <i>et al.</i> , 2003
1	1RNU	1.60	0.176	VIb	G-N- <u>P</u> -Y	S-T-S	Kim <i>et al.</i> , 1992
2	1VOK	2.10	0.198	VIb	R-E- <u>P</u> -K	S-T-S	Nikolov <i>et al.</i> , 1994
3	1VOK	2.10	0.198	VIb	K-V- <u>P</u> -K	S-T-S	Nikolov <i>et al.</i> , 1994

4	1YTB	1.80	0.201	VIb	R-E- <u>P</u> -K	S-T-S	Kim <i>et al.</i> , 1993
5	1YTB	1.80	0.201	VIb	V-K- <u>P</u> -K	S-T-S	Kim <i>et al.</i> , 1993
1	1LZ1	1.50	0.177	VIII	Q-D-N-I	H-T-H	Artymiuk <i>et al.</i> , 1981
2	1OBP	2.00	0.164	VIII	S-T-N-P	S-T-H	Tegoni <i>et al.</i> , 1996
3	1PGA	2.07	0.174	VIII	A-V-D-A	S-T-H	Gallagher <i>et al.</i> , 1994
4	1PIN	1.35	0.223	VIII	G-E-E-D	H-T-H	Ranganathan <i>et al.</i> , 1997
5	1URS	1.45	0.210	VIII	G-K-V-S	H-T-H	Schäfer <i>et al.</i> , 2004
6	1YT3	1.60	0.1961	VIII	I-T-D-W	H-T-H	Zuo <i>et al.</i> , 2005
7	2QSV	2.10	0.205	VIII	A-P-L-R	R-T-S	Binkowski <i>et al.</i> , 2007
8	3MHB	1.55	0.193	VIII	A-K-K-I	H-T-S	Roche <i>et al.</i> , 2012
9	3WIT	1.95	0.198	VIII	G-G-S-V	S-T-S	Uchida <i>et al.</i> , 201a
1	1TLX	2.10	0.153	$\gamma$ -classique	S-T-Y	S-T-S	English <i>et al.</i> , 1999
2	2PTN	1.55	0.193	$\gamma$ -classique	S-G-Y	S-T-S	Walter <i>et al.</i> , 1982
3	2QSV	2.10	0.205	$\gamma$ -classique	D-E-S	R-T-R	Binkowski <i>et al.</i> , 2007
4	4NZD	2.75	0.241	$\gamma$ -classique	S-G-Q	R-T-S	Hamming <i>et al.</i> , 2014
1	1H16	1.53	0.145	$\gamma$ -inverse	L- <u>P</u> -D	R-T-R	Becker <i>et al.</i> , 2002
2	1SU8	1.10	0.146	$\gamma$ -inverse	L- <u>P</u> -V	H-T-S	Dobbek <i>et al.</i> , 2004
3	1SU8	1.10	0.146	$\gamma$ -inverse	G-L-C	H-T-S	Dobbek <i>et al.</i> , 2004
4	1YTB	1.80	0.201	$\gamma$ -inverse	A-S-G	S-T-S	Kim <i>et al.</i> , 1993
5	5VA6	2.40	0.249	$\gamma$ -inverse	G-D-C	R-T-S	Bergamin <i>et al.</i> , 2017

## I.2 Optimisation de la géométrie des peptides

Les fichiers de coordonnées des structures de notre banque de données initiale ont été téléchargés à partir de la PDB. Les motifs de type coude composés par 4 résidus (coudes  $\beta$ ) et 3 résidus (coudes  $\gamma$ ) ont été isolés de la protéine native (Tableau 18). Les atomes d'hydrogène ont été ajoutés avec la fonction automatique de Chimera UCSF dans la structure peptidique. Pour éviter les charges, l'extrémité N-terminale a été acétylée et l'extrémité C-terminale a été amidée.

La géométrie de chaque structure a été optimisée en utilisant un protocole de recuit simulé sous contraintes. Ceci a été réalisé dans le but d'obtenir des structures avec des angles dièdres  $\Phi$  et  $\Psi$  des résidus  $i+1$  et  $i+2$  pour les coudes  $\beta$ , et  $i+1$  pour les coudes  $\gamma$ , proches des valeurs canoniques caractéristiques de chaque type de coude  $\beta$  ou  $\gamma$  (Tableau 10, chapitre 1).

### I.2.1 Paramètres du recuit simulé

Le logiciel de modélisation CNS (*Crystallography and NMR System*) a été utilisé pour notre étude.<sup>44</sup> Ce logiciel intègre les données expérimentales issues des analyses des spectres RMN ou de la cristallographie à la mécanique et la dynamique moléculaire pour déterminer les structures de plus basses énergies.

Le champ de force CHARMM (*Chemistry at HARvard Macromolecular Mechanics*) développé par l'équipe de Karplus en 1983 a été utilisé lors des calculs de modélisation moléculaire.<sup>147</sup> Afin de prendre en considération toutes les corrections de paramètres d'énergie potentielle, nous avons utilisé la version 22 du champ de force CHARMM adapté aux peptides et aux protéines. Ce champ de force est constitué essentiellement de deux fichiers :

- Topallh22x.pro : fichier de topologie contenant la représentation détaillée de tous les acides aminés avec leur type d'atome, les liaisons, les angles et les masses de chaque atome ainsi que des modifications éventuelles (N-ter, C-ter, etc....) ;
- Parallh22x.pro : fichier de paramètres contenant les valeurs d'équilibre des liaisons, des angles dièdres, impropres, de valence ainsi que les valeurs pour les interactions non-liées.

Il est important de signaler que le fichier Topallh22x.pro comporte toutes les informations étant en relation avec des acides aminés non modifiés.

En utilisant les deux fichiers Topallh22x.pro et Parallh22x.pro, ainsi que la séquence du peptide à modéliser, le logiciel CNS permet de créer un fichier .psf « *Protein Structure File* ». Ce dernier inclus toutes les informations du peptide.

Une fois les deux fichiers .psf et .pdb, créés ainsi que les contraintes angulaires et de distance préparées, (voir ci-dessous) les géométries ont été optimisées. Ceci a été réalisé en appliquant un protocole de recuit simulé sur chaque structure de peptide.

Les principaux paramètres utilisés dans le protocole de recuit simulé sont reportés dans le Tableau 19.

Tableau 19 Paramètres utilisés au cours du recuit simulé. La constante de force CDIH est exprimée en  $Kcal.mol^{-1}.rad^{-2}$  et les constantes de force  $K_{repel}$  et  $K_{DIST}$  sont exprimées en  $Kcal.mol^{-1}.Å^{-2}$ .

Étapes	Température (K)	Nombre de pas	$K_{repel}$	$K_{NOE}$	CDIH
<b>Chauffage</b>	20 000	1000	0.1	150	100
<b>1<sup>er</sup> refroidissement</b>	20 000 → 0	2000	0.1 → 1	150	200
<b>2<sup>ème</sup> refroidissement</b>	2000 → 0	3000	1 → 4	150	200
<b>Minimisation</b>	-	200	-	75	400

### I.2.2 Contraintes de distance et d'angle utilisées pour les calculs de modélisation moléculaire

La liaison hydrogène entre le résidu  $i$  et le résidu  $i+n$  (où  $n$  est égal à 2 pour les coudes de nature  $\gamma$  et 3 pour les coudes  $\beta$ ) a été utilisée en tant que contrainte de distance. Deux contraintes ont été incluses : une distance de  $2,0\pm 0,3$  Å entre l'hydrogène amide et l'atome oxygène accepteur du groupe carboxyle, et une distance de  $3,0\pm 0,3$  Å entre l'atome oxygène accepteur du groupe carboxyle et l'hydrogène amide.

Pour chaque nature et type de coude, les valeurs canoniques des angles dièdres  $\Phi$  et  $\Psi$  pour les résidus  $i+1$  (pour les coudes de nature  $\gamma$ , Tableau 3, chapitre 1) et pour les résidus  $i+1$  et  $i+2$  (pour les coudes de nature  $\beta$ , Tableau 10, chapitre 1) ont été introduites comme contraintes angulaires. Pour le premier et le dernier acide aminé de chaque peptide, des angles dièdres  $\Phi$  et  $\Psi$  ont également été ajoutés comme contraintes angulaires. Les valeurs ont été définies en fonction de la nature des motifs structuraux entourant le coude (hélice  $\alpha$  :  $-57^\circ$ ,  $-47^\circ$  ; brin  $\beta$  :  $-139^\circ$ ,  $+135^\circ$  ; conformation aléatoire : pas de contraintes). Les angles ont été limités dans un intervalle de  $\pm 10^\circ$  autour de la valeur définie. Ce seuil a été choisi pour éviter un basculement entre les différents types de motif.

### I.3 Structures finales

Une fois les 104 structures optimisées, nous avons décidé d'augmenter la taille de la banque de données en générant pour chaque séquence les huit autres motifs structuraux au sein duquel le coude pourrait se trouver. Ceci a été réalisé afin de pas introduire de biais dans les paramètres de RMN. En effet, les acides aminés en position  $i$  et  $i+3$  pour les coudes  $\beta$  ou  $i$  et  $i+2$  pour les coudes  $\gamma$  peuvent avoir un impact sur les distances entre les hydrogènes et un nombre faible de motifs représentatifs pour certains types de coudes (*cf.* Tableau 18) pourrait conduire à l'identification de faux paramètres de RMN discriminatoires.

#### I.3.1 Constitution du jeu final de structures

Pour la constitution du jeu final de structures, différentes contraintes angulaires, en utilisant les valeurs canoniques des angles dièdres  $\Phi$  et  $\Psi$  correspondant à une structuration hélicoïdale et/ou à une structuration en feuillet  $\beta$ , ont été imposées pour le premier et dernier résidu du coude. L'optimisation des géométries a été effectuée en suivant la procédure décrite dans la partie I.2 de ce chapitre. Cela a permis de générer, pour chaque séquence, les motifs structuraux suivants : brin-coude-brin (S-T-S), hélice-coude-hélice (H-T-H), hélice-coude-brin (H-T-S), brin-coude-hélice (S-T-H). Des calculs ont également été effectués sans restriction angulaires pour le premier et/ou le dernier résidu, ce qui a permis une meilleure exploration de la conformation des angles autorisés des résidus du coude, mais sans modifier le motif initial. Cette stratégie ne permettait donc pas une exploration des zones en conformation aléatoire, en particulier celles qui se trouvent dans la partie droite du graphe de Ramachandran. Pour surmonter ce problème, des calculs supplémentaires ont été effectués avec les couples de contraintes angulaires suivants :  $(+120^\circ, +120^\circ)$  ;  $(+90^\circ, +90^\circ)$ ,  $(+50^\circ, +50^\circ)$  ;  $(+50^\circ, -50^\circ)$ ,  $(+90^\circ, -90^\circ)$  ;  $(+120^\circ, -120^\circ)$  pour le premier ou le dernier résidu. Ces valeurs ont été choisies pour couvrir les régions autorisées du côté droit du graphe de Ramachandran. Pour ces calculs supplémentaires, une plage de  $\pm 20^\circ$  a été permise autour de la valeur définie. Pour chaque séquence sélectionnée, 21 structures ont été donc générées. Le jeu de structures finales était donc composé de 2184 structures optimisées de peptides.

#### I.3.2 Evaluation de la qualité des structures

Les 2184 structures ont été analysées en se basant sur plusieurs critères. Dans un premier temps, la valeur d'énergie a été regardée. Nous avons retenu les structures qui possédaient une énergie avec un écart de  $\pm 25$  kcal/mol par rapport à la structure de départ. Dans un second temps, les

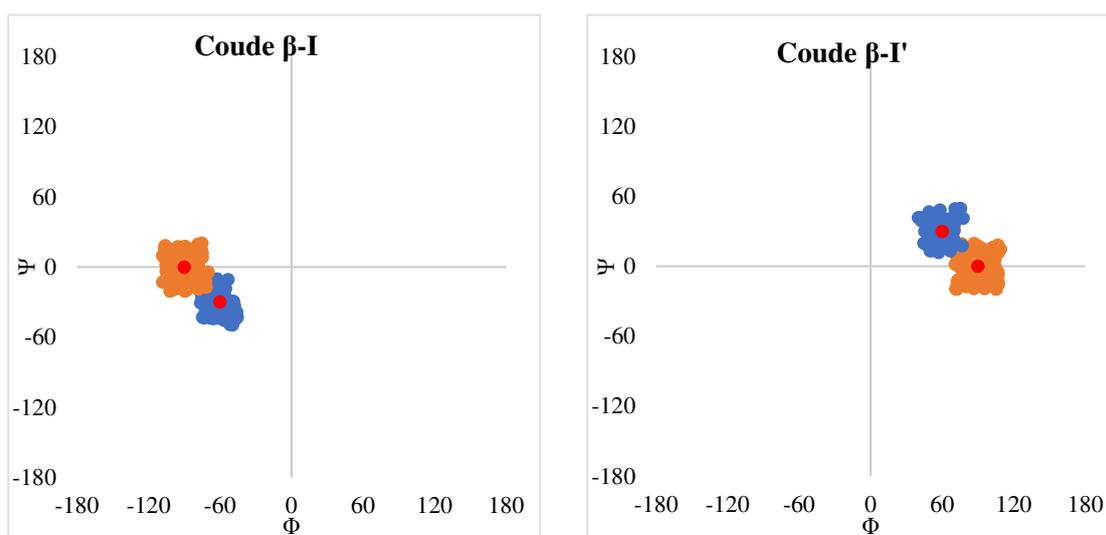
valeurs d'angles dièdres des acides aminés et la longueur de la liaison hydrogène ont été analysées et comparées aux valeurs canoniques. Les structures qui possédaient tous ces critères ont été retenues.

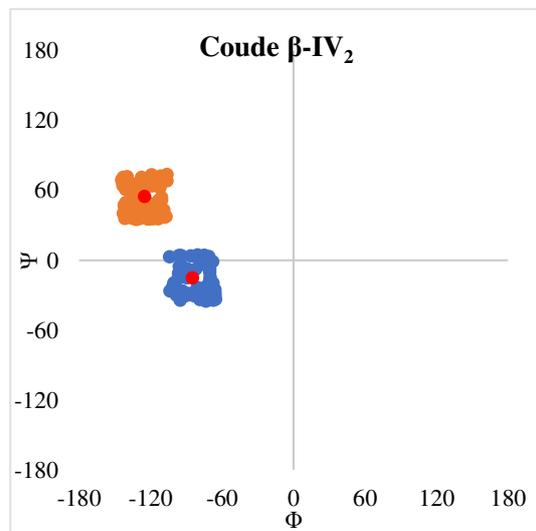
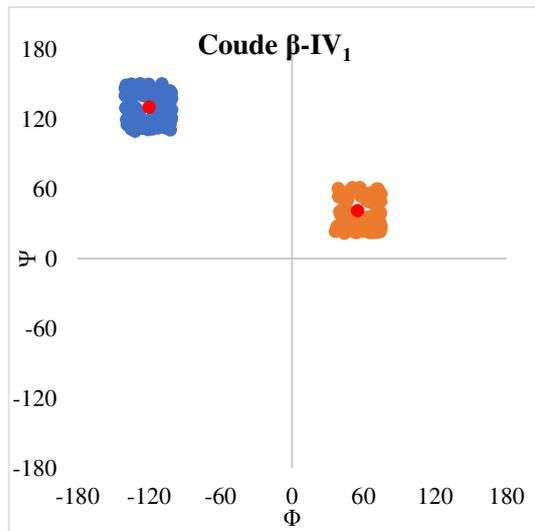
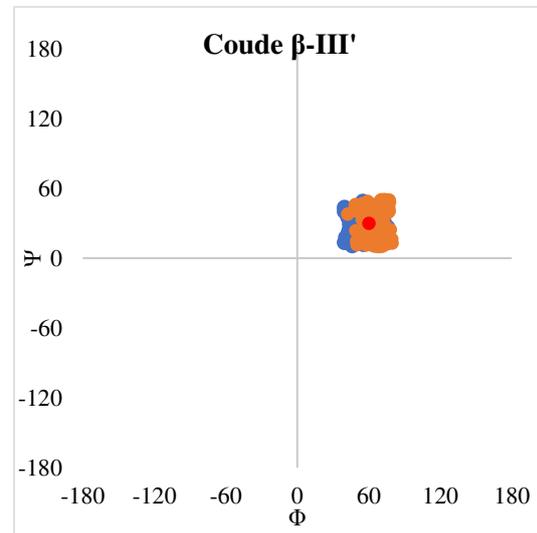
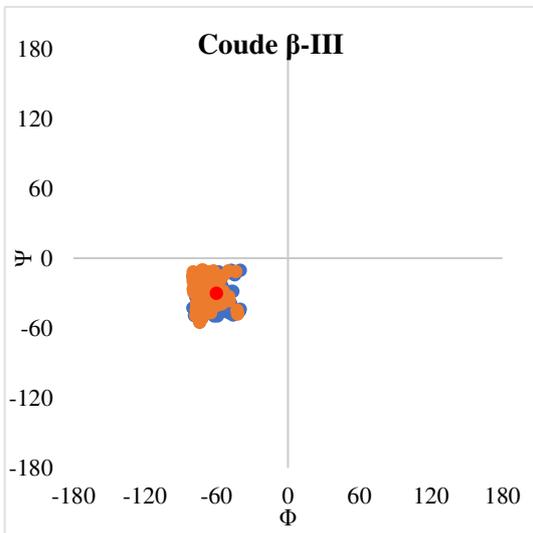
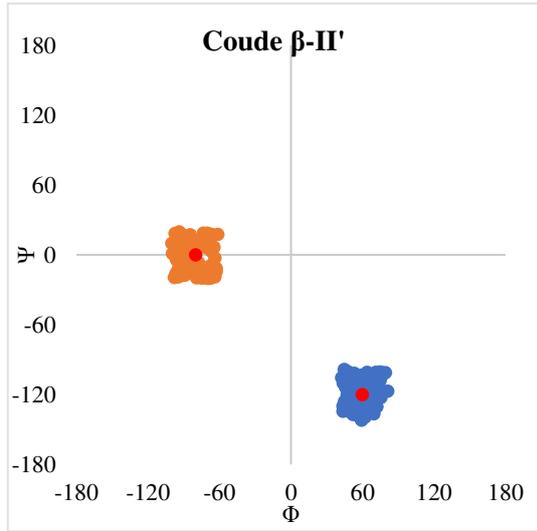
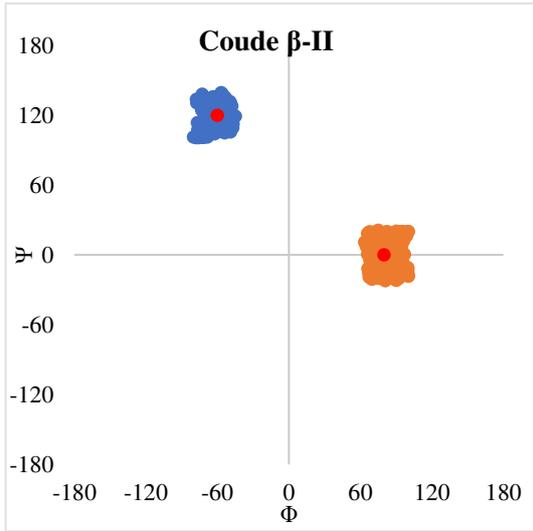
#### I.4 Evaluation de la qualité de la banque de données

Un fois l'ensemble des structures générées, une analyse a été menée afin de vérifier la bonne représentativité de la banque de données. Tout d'abord nous avons regardé l'exploration conformationnelle des angles  $\Phi$  et  $\Psi$  des deux résidus centraux, puis nous avons examiné celle du premier et du dernier résidu pour vérifier la bonne simulation des différents motifs.

##### I.4.1 Analyse des angles dièdres $\Phi$ et $\Psi$ des résidus centraux

L'analyse du diagramme de Ramachandran montre que la distribution des angles dièdres  $\Phi$  et  $\Psi$  du résidu central  $i+1$  (pour les coudes  $\gamma$ ) et des résidus centraux  $i+1$  et  $i+2$  (pour les coudes  $\beta$ ) est en accord avec celle reportée dans la littérature<sup>45</sup>, avec des angles qui s'écartent de moins de  $30^\circ$  par rapport aux valeurs des angles canoniques de chaque type de coude (Figures 60 et 61). Nous pouvons donc observer une bonne exploration conformationnelle dans les structures générées pour les résidus caractéristiques des coudes autour des valeurs canoniques. Tous les acides aminés centraux rentrent dans un jeu d'angles de  $\pm 20^\circ$  en comparaison avec les valeurs des angles canoniques (Figures 60 et 61).





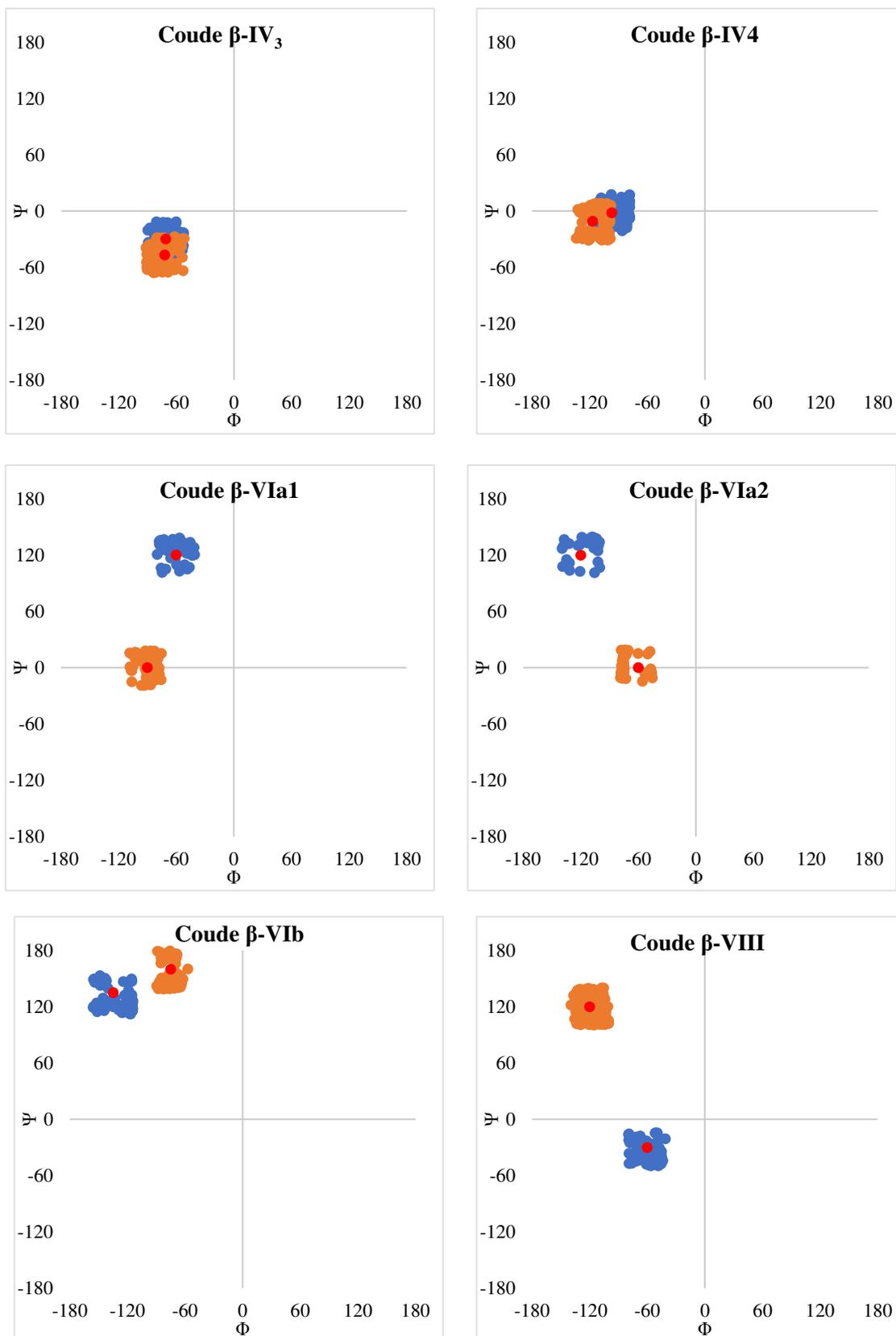


Figure 60 Distribution des angles dièdres  $\Phi$  et  $\Psi$  des résidus centraux pour tous les types de coudes  $\beta$  dans le graphe de Ramachandran. En bleu, le résidu  $i+1$  ; en orange, le résidu  $i+2$  ; en rouge, les angles canoniques de chaque type de coude.

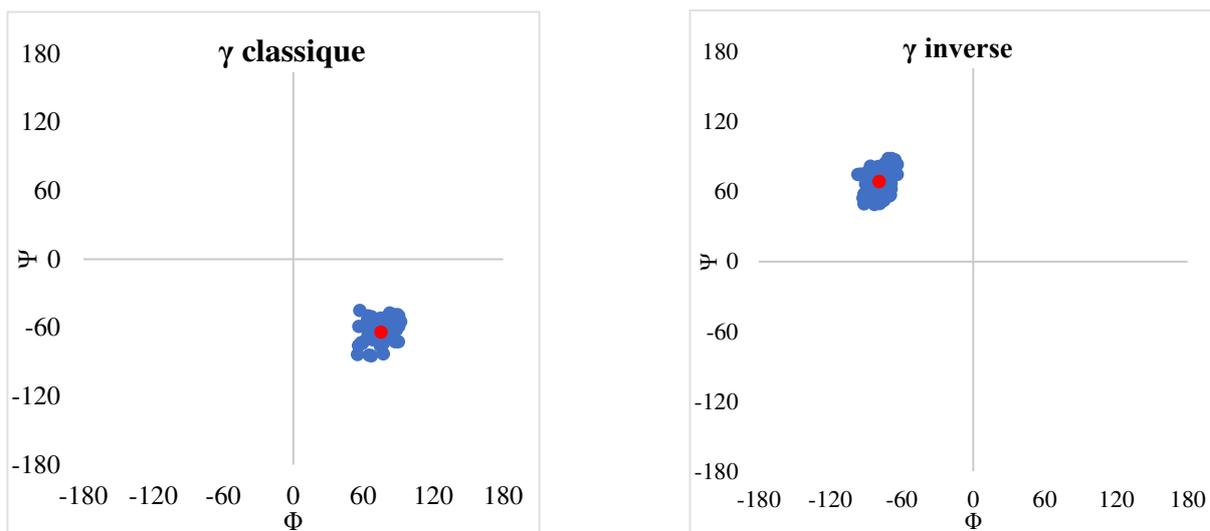
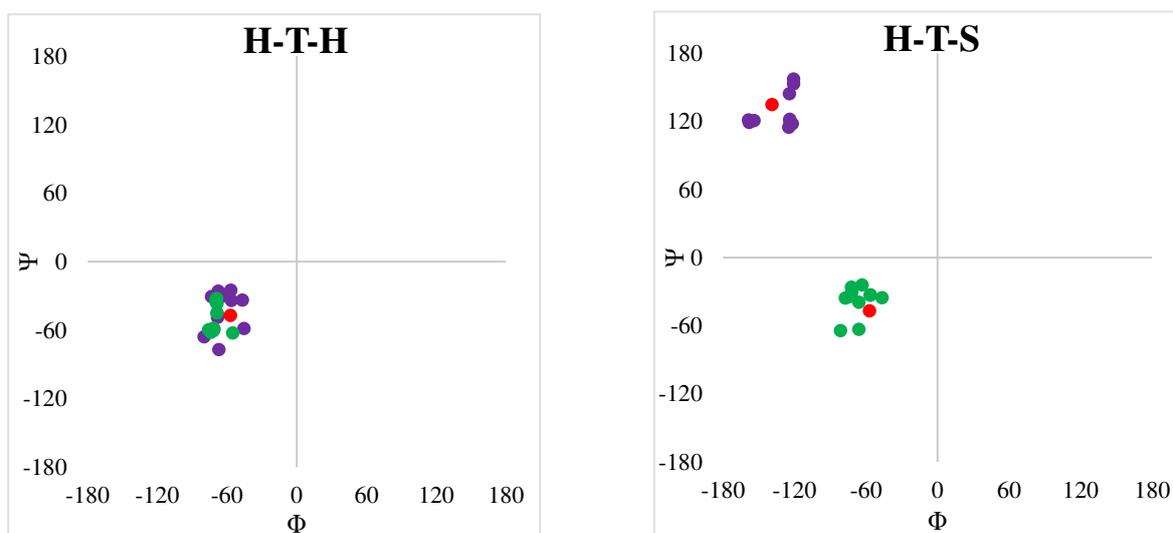


Figure 61 Distribution des angles dièdres  $\Phi$  et  $\Psi$  du résidu central pour les deux types de coudes  $\gamma$  dans le graphe de Ramachandran. En bleu, le résidu  $i+1$  ; en orange, le résidu  $i+2$  ; en rouge les angles canoniques de chaque type de coude.

#### I.4.2 Analyse des angles dièdres $\Phi$ et $\Psi$ du premier et du dernier résidu

De la même manière, une analyse des angles dièdres  $\Phi$  et  $\Psi$  a été effectuée pour le premier et le dernier résidu dans toutes les structures générées. Cette dernière a été réalisée afin de vérifier que, pour les motifs H-T-H, H-T-S, S-T-H et S-T-S, ces résidus sont localisés dans les zones énergétiquement favorables. Ainsi, les diagrammes indiquent clairement une bonne exploration conformationnelle des deux résidus autour des valeurs canoniques des structures en hélice  $\alpha$  ou en brin  $\beta$  comme montré dans la Figure 62. Nous avons décidé de présenter l'exemple du coude  $\beta$  de type I.



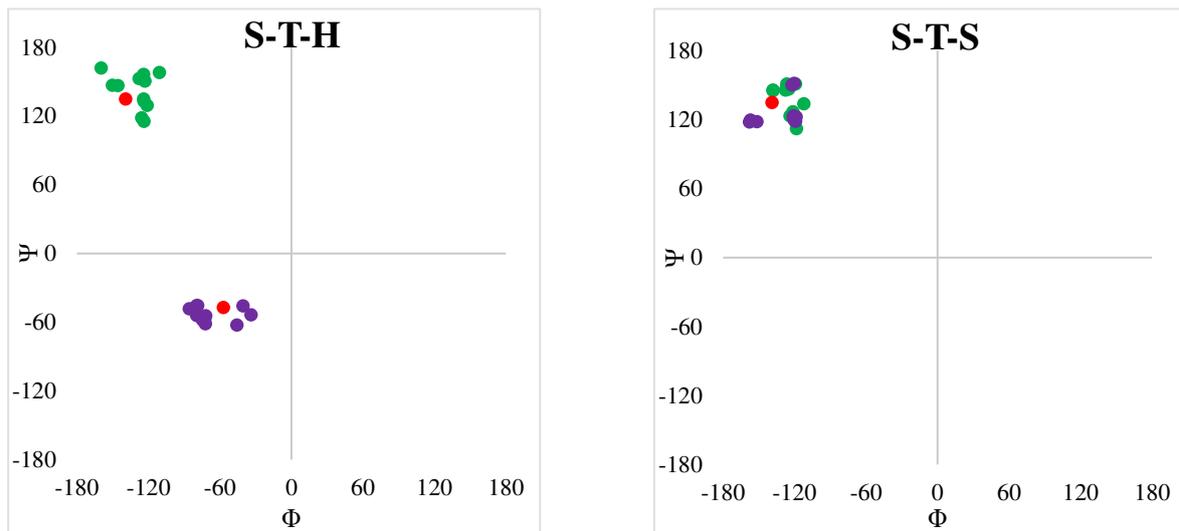
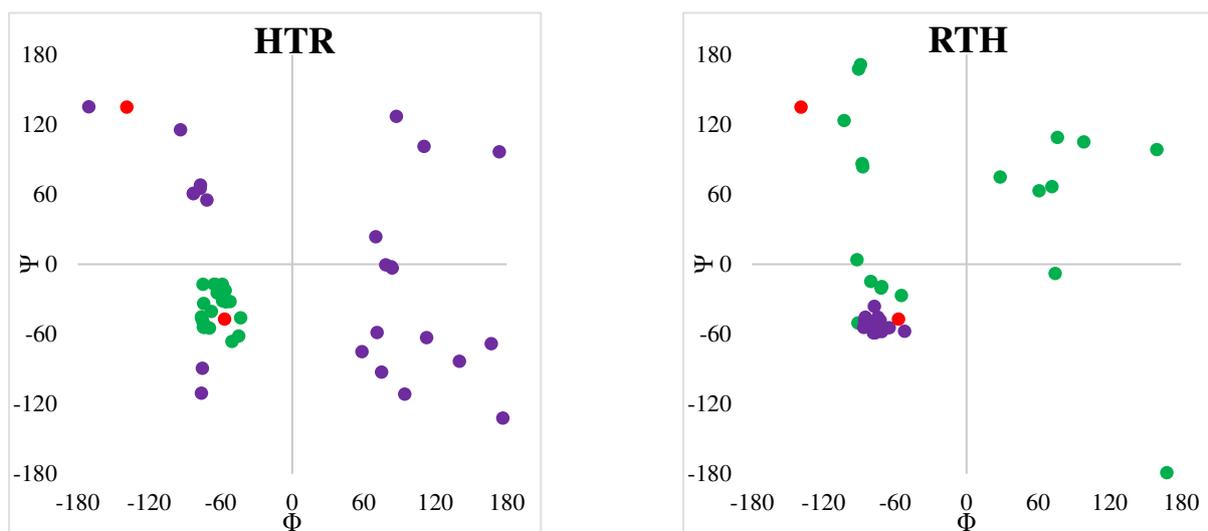


Figure 62 Distribution des angles dièdres  $\Phi$  et  $\Psi$  des résidus  $i$  et  $i+3$  pour les motifs  $H-T-H$ ,  $H-T-S$ ,  $S-T-H$  et  $S-T-S$  dans le graphe de Ramachandran pour le coude  $\beta$  de type I. En vert, le résidu  $i$ , en violet le résidu  $i+3$ , en rouge les angles canoniques d'une conformation en hélice  $\alpha$  (H) ou feuillet  $\beta$  (S).

Pour les autres motifs qui incluent en première ou dernière position une conformation aléatoire ( $H-T-R$ ,  $R-T-H$ ,  $S-T-R$ ,  $R-T-S$  et  $R-T-R$ ), nous pouvons observer une exploration conformationnelle plus restreinte. Ceci est lié à la difficulté d'explorer tous les domaines possibles de cette conformation aléatoire. Néanmoins, en imposant des contraintes d'angles ( $\Phi$  et  $\Psi$ ) dans toutes les zones du diagramme de Ramachandran, nous avons généré des structures suffisamment différentes pour obtenir un échantillonnage correct (exemple du coude  $\beta$  de type I, Figure 63). Ces observations sont générales pour tous les types de coudes  $\beta$  et  $\gamma$  étudiés.



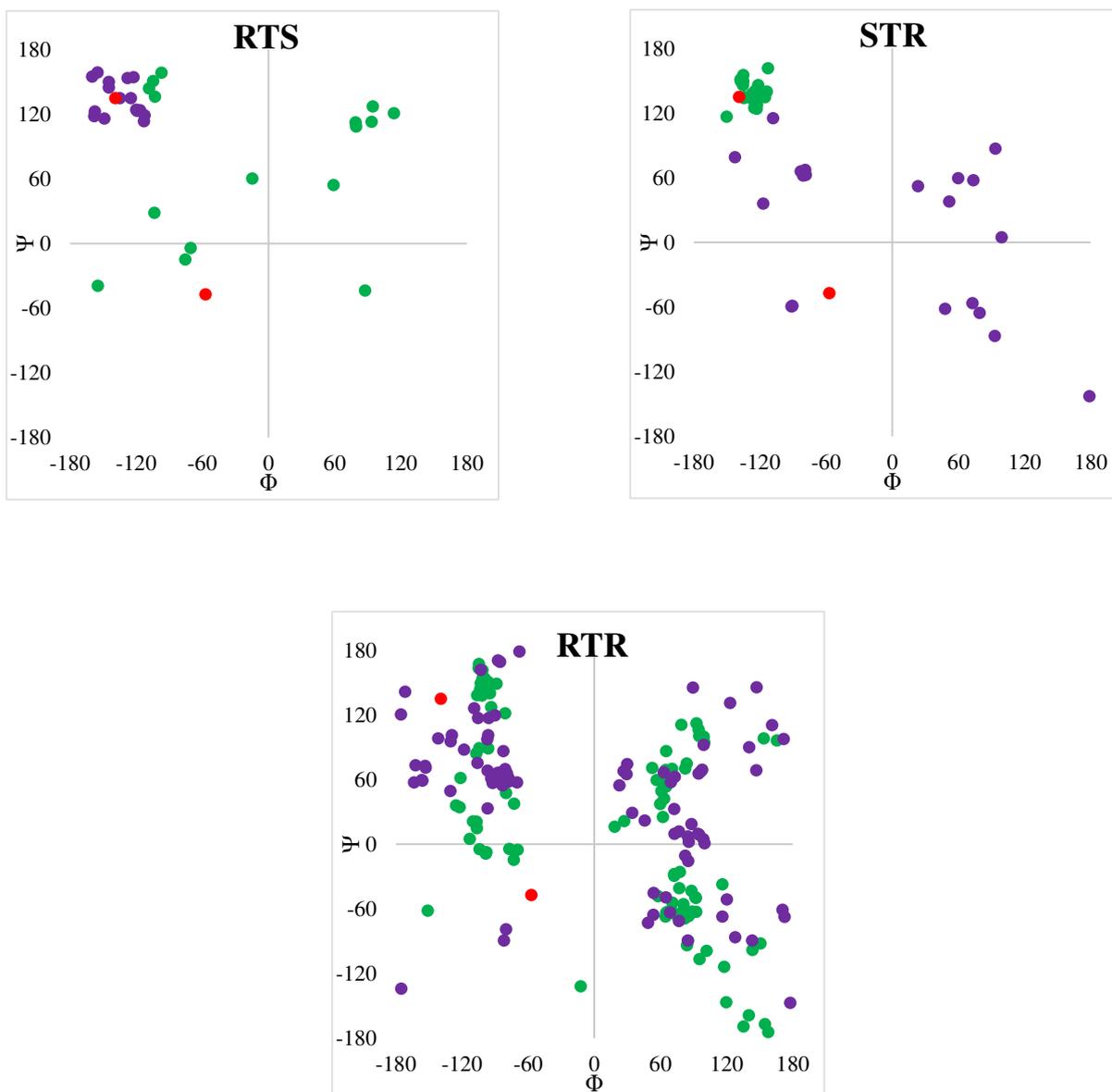


Figure 63 Distribution des angles dièdres  $\Phi$  et  $\Psi$  des résidus  $i$  et  $i+3$  pour les motifs H-T-R, R-T-H, R-T-S, S-T-R et R-T-R dans le graphe de Ramachandran pour le coude de type I. En vert, le résidu  $i$ , en violet le résidu  $i+3$ , en rouge les angles canoniques d'une conformation en hélice  $\alpha$  ou feuillet  $\beta$ .

## II. Recherche de paramètres de RMN caractéristiques

Comme vu au chapitre 2, les paramètres de RMN utilisés pour la caractérisation de la structure secondaire d'une chaîne polypeptidique sont les distances  $^1\text{H}-^1\text{H}$  et la constante de couplage  $^3\text{J}_{\text{HN-H}\alpha}$ . Nous sommes donc allés les rechercher pour chaque type de coude. Dans la suite du chapitre, je présenterai ces deux paramètres et comment ils ont été mesurés dans notre banque de données. Puis je présenterai les résultats obtenus. Enfin une conclusion générale achèvera ce chapitre.

## II.1 Distances $^1\text{H}$ - $^1\text{H}$

Les mesures des distances dans les structures générées ont été effectuées en prenant en compte uniquement les distances entre les atomes d'hydrogène. Nous sommes allés à la recherche de distances entre les hydrogènes du squelette peptidique ( $\text{H}_\text{N}$  et  $\text{H}_\alpha$ ), mais aussi entre les hydrogènes de la chaîne latérale (en se limitant aux  $\text{H}_\beta$ , ainsi qu'aux  $\text{H}_\delta$  en présence d'une proline) et les protons du squelette peptidique. Les acides aminés présents au sein de nos coudes n'ayant pas tous des  $\text{H}_\beta$  (glycine) ou pouvant présenter un nombre différent de ce type d'hydrogènes (un, deux ou trois), nous avons choisi de faire le traitement suivant. Pour les glycines, qui possèdent deux hydrogènes  $\text{H}_\alpha$ , l'hydrogène indiqué comme  $\text{H}_{\alpha 2}$  dans le fichier PDB a été traité comme un atome d'hydrogène  $\text{H}_{\beta 2}$  de la chaîne latérale. Pour les acides aminés ayant une chaîne latérale  $\text{H}_\beta$  unique (isoleucine, thréonine et valine), cet hydrogène a été traité comme un hydrogène  $\text{H}_{\beta 2}$ . Pour les alanines, qui présentent trois hydrogènes  $\text{H}_\beta$ , les trois distances entre les  $\text{H}_\beta$  et l'hydrogène amide d'intérêt ont été calculées. Etant conscient que pour ce résidu, seule une distance est observée dans une expérience bidimensionnelle NOESY, les trois distances relevées ont été moyennées en utilisant l'équation suivante :

$$\left( \frac{\frac{1}{d_1^6} + \frac{1}{d_2^6} + \frac{1}{d_3^6}}{3} \right)^{-\frac{1}{6}}$$

La valeur de distance obtenue a été utilisée en tant que  $\text{H}_{\beta 2}$ .

Afin d'avoir des données compatibles avec celles obtenues par la RMN, nous avons choisi d'adopter la présentation suivante. Les distances entre les atomes d'hydrogène de deux acides aminés consécutifs ont été présentées en tant que séquentielles ( $i, i+1$ ). Celles qui ont été présentées comme moyennes ( $i, i+j$  où  $j$  est égal à 2 ou 3) sont les distances inter-protons entre des acides aminés séparés par un nombre d'acides aminés égal à  $j-1$ .

Les distances inter-hydrogènes observées pour tous les peptides représentant les coudes  $\beta$  comprennent les  $\text{H}_\text{N}$ ,  $\text{H}_\alpha$ ,  $\text{H}_\beta$  séquentielles ( $i, i+1$ ), à moyenne portée ( $i, i+2$ ) ( $i, i+3$ ). Pour les peptides présentant une proline en position  $i+1$  ou  $i+2$ , les distances observées ont été les mêmes avec toutefois l'ajout des distances  $\text{H}_\delta$ - $\text{H}_\text{N}$ ,  $\text{H}_\text{N}$ - $\text{H}_\delta$ ,  $\text{H}_\delta$ - $\text{H}_\alpha$  et  $\text{H}_\alpha$ - $\text{H}_\delta$  séquentielles ( $i, i+1$ ). Les peptides en coudes  $\gamma$  présentent 3 acides aminés. Les distances analysées ont donc été :  $\text{H}_\text{N}$ ,  $\text{H}_\alpha$ ,  $\text{H}_\beta$  séquentielles ( $i, i+1$ ) et à moyenne portée ( $i, i+2$ ).

Afin de pouvoir comparer les informations obtenues aux données de référence présentes dans la littérature, les distances mesurées ont été représentées sous forme de diagramme NOE. Les distances relevées ont été divisées en quatre catégories : courtes ( $d \leq 2,8 \text{ \AA}$ ), moyennes ( $2,8 < d \leq 3,8 \text{ \AA}$ ), longues ( $3,8 < d \leq 5 \text{ \AA}$ ) et non visibles ( $d > 5 \text{ \AA}$ ). Les distances hydrogène-hydrogène ont été indiquées par des lignes reliant les résidus contenant les hydrogènes impliqués. Les barres épaisses, moyennes et fines correspondent aux distances courtes, moyennes et longues, respectivement. Les distances indétectables ( $> 5,0 \text{ \AA}$ ) ne sont pas représentées. La nature de l'hydrogène ( $H_N$ ,  $H_\alpha$ ,  $H_\beta$  ou  $H_\delta$ ) et le type de connectivité (séquentielle ou moyenne) sont indiqués à gauche du diagramme. Les chiffres du bas de ce diagramme représentent la position de l'acide aminé dans le coude.

Pour les distances couvrant deux catégories, nous avons décidé d'appliquer les traitements suivants : si la distance mesurée dépassait de  $0,1 \text{ \AA}$  dans une autre des catégories, la catégorie de référence principale a été conservée, sinon la distance a été classée en tant qu'intermédiaire entre deux catégories.

## II.2 Constantes de couplage $^3J_{HN-H\alpha}$

Le second paramètre de RMN utilisé dans les analyses structurales de polypeptides est la valeur de la constante de couplage  $^3J_{HN-H\alpha}$ . En effet, le couplage scalaire  $J$  entre les protons amide  $H_N$  et  $H_\alpha$  est lié à l'angle dièdre  $\Phi$ . Sa valeur peut donc être utilisée pour identifier des motifs de structure secondaire.

Pour les hélices et les brins  $\beta$ , des valeurs proches de 4 Hz et 9 Hz sont attendues, respectivement. Ces motifs, qui sont par nature répétitifs, sont plus faciles à caractériser que les coudes, car ces derniers sont des motifs structuraux formés d'un nombre de résidus limité (3 ou 4) et non répétitifs, pour lesquels les angles dièdres ne sont fixés que pour un (coudes  $\gamma$ ) ou deux résidus (coudes  $\beta$ ). Néanmoins, la valeur attendue de constante  $^3J_{HN-H\alpha}$  pour le ou les résidus centraux doit dépendre du type de coude.

Les valeurs des constantes de couplage canoniques  $^3J_{HN-H\alpha}$ , pour les résidus centraux ( $i+1$  et  $i+2$ ) dans les coudes  $\beta$  et pour le résidu central ( $i+1$ ) dans les coudes  $\gamma$ , ont été calculées à l'aide de l'équation de Karplus, en utilisant trois séries de coefficients disponibles dans la littérature :

$$^3J_{HN-H\alpha} = 6.4 \cos^2(\Phi-60) - 1.4 \cos(\Phi-60) + 1.9 \quad (1)$$

$$^3J_{HN-H\alpha} = 6.7 \cos^2(\Phi-60) - 1.3 \cos(\Phi-60) + 1.5 \quad (2)$$

$$^3J_{HN-H\alpha} = 6.51 \cos^2(\Phi-60) - 1.76 \cos(\Phi-60) + 1.6 \quad (3)$$

Le premier ensemble de coefficients correspond à celui proposé par Pardi<sup>122</sup> (1). La deuxième et la troisième série ont été proposées plus tard respectivement par Ludvigsen<sup>148</sup> (2) et Vuister<sup>149</sup> (3).

Afin de déterminer l'impact des coefficients, nous avons calculé, pour chaque type de coude, les constantes de couplage  $^3J_{\text{HN-H}\alpha}$  associées aux valeurs des angles canoniques, en appliquant ces trois équations, puis la moyenne a été effectuée (Tableau 20).

Tableau 20 Valeurs moyennes de constante de couplage  $^3J_{\text{HN-H}\alpha}$  pour les résidus centraux des différents types de coudes  $\beta$  et  $\gamma$ . Ces valeurs ont été calculées à partir de l'équation de Karplus en utilisant les trois séries de coefficients de Pardi, Ludvigsen et Vuister.<sup>132-134</sup> Les valeurs d'angle ou de constante de couplage  $^3J_{\text{HN-H}\alpha}$  qui ne sont pas disponibles ont été indiqués comme « nd ».

Type de coude	$\Phi_{i+1}$	$\Phi_{i+2}$	$^3J_{\text{HN-H}\alpha}$ (Hz) (i+1)				$^3J_{\text{HN-H}\alpha}$ (Hz) (i+2)			
			Pardi	Ludvigsen	Vuister	Moyenne	Pardi	Ludvigsen	Vuister	Moyenne
$\beta$ -I	-60°	-90°	4,2	3,8	4,0	4,0	7,9	7,6	7,8	7,8
$\beta$ -I'	+60°	+90°	6,9	6,9	6,5	6,7	5,5	5,4	5,1	5,2
$\beta$ -II	-60°	+80°	4,2	3,8	4,0	4,0	6,2	6,2	5,9	6,0
$\beta$ -II'	+60°	-80°	6,9	6,9	6,5	6,7	6,4	6,7	6,5	6,6
$\beta$ -III	-60°	-60°	4,2	3,8	4,0	4,0	4,2	3,8	4,0	4,0
$\beta$ -III'	+60°	+60°	6,9	6,9	6,5	6,7	6,9	6,9	6,5	6,7
$\beta$ -IV <sub>1</sub>	-120°	+55°	9,7	9,5	9,6	9,6	6,6	6,4	6,5	6,6
$\beta$ -IV <sub>2</sub>	-85°	-125°	7,3	7,1	7,2	7,2	9,6	9,4	9,5	9,6
$\beta$ -IV <sub>3</sub>	-71°	-72°	5,6	5,2	5,5	5,6	5,7	5,3	5,5	5,6
$\beta$ -IV <sub>4</sub>	-97°	-117°	8,7	8,5	8,5	8,6	9,6	9,4	9,5	9,6
$\beta$ -VIa1	-60°	-90°	4,2	3,8	4,0	4,0	nd	nd	nd	nd
$\beta$ -VIa2	-120°	-60°	9,7	9,5	9,6	9,6	nd	nd	nd	nd
$\beta$ -VIb	-135°	-75°	9,2	9	9,1	9,2	nd	nd	nd	nd
$\beta$ -VIII	-60°	-120°	4,2	3,8	4,0	4,0	9,7	9,5	9,6	9,6
$\gamma$ classique	+75°	nd	6,5	6,5	6,5	6,4	nd	nd	nd	nd
$\gamma$ inverse	-79°	nd	6,6	6,6	6,5	6,4	nd	nd	nd	nd

Nous avons pu remarquer de faibles variations, allant de 0,3 à 0,6 Hz, entre les valeurs obtenues avec les différentes séries de coefficients, les plus importantes étant rencontrées pour les angles positifs. Cela pourrait s'expliquer facilement par le petit nombre de valeurs expérimentales utilisées pour l'ajustement de l'équation pour des angles compris entre 0° et 150°. Compte tenu du faible impact des changements de coefficients, nous avons décidé d'utiliser la moyenne des trois valeurs calculées en tant que valeur de constante de couplage canonique  $^3J_{\text{HN-H}\alpha}$ . Cependant, nous avons conservé les trois séries de coefficients pour déterminer les valeurs de

constante de couplage  ${}^3J_{\text{HN-H}\alpha}$  associés aux angles  $\Phi_{i+1}$  et  $\Phi_{i+2}$  des structures qui compose notre jeu final afin de pouvoir déterminer l'intervalle le plus large pour chaque valeur.

### **III. Résultats**

La recherche de paramètres de RMN pour la caractérisation des coudes  $\beta$  et  $\gamma$  a été réalisée en analysant les distances et les angles présentés auparavant dans ce chapitre. Afin de simplifier la présentation, je présenterai dans un premier temps les résultats obtenus pour les coudes  $\beta$  puis dans un second temps ceux des coudes  $\gamma$ . Pour chaque nature de coude, je présenterai d'abord une analyse des distances obtenues de manière générale. Puis, une analyse détaillée de ces distances sous forme de diagramme NOE sera réalisée afin de pouvoir identifier les signatures caractéristiques pour distinguer les différents types de coudes  $\beta$  et  $\gamma$ . Après, la constante de couplage sera présentée.

#### **III.1 Les coudes $\beta$**

##### **III.1.1 Cas général**

Pour les coudes de nature  $\beta$ , dans un premier temps, nous présenterons les paramètres de RMN distances  ${}^1\text{H}$ - ${}^1\text{H}$  et constantes de couplage, puis nous aborderons le cas des coudes  $\beta$  de type I, II et III qui ont une forte préférence de la proline en position 2. Pour les distances, une analyse en deux temps sera réalisée : (i) distances entre les hydrogènes du squelette peptidique ( $\text{H}_\alpha$  et  $\text{H}_\text{N}$ ), (ii) distances entre la chaîne latérale et le squelette peptidique.

##### **III.1.1.1 Les distances**

###### **III.1.1.1.1 Distances entre les hydrogènes du squelette peptidique**

Toutes les distances entre les hydrogènes du squelette peptidique ( $\text{H}_\alpha$  et  $\text{H}_\text{N}$ ) ont été déterminées et elles sont présentées dans le Tableau 21.

Six distances  $\text{H}_\text{N}$ - $\text{H}_\text{N}$  peuvent être mesurées (entre les acides aminés  $1-2$  ;  $2-3$  ;  $3-4$  ;  $1-3$  ;  $2-4$  et  $1-4$ ). Nous pouvons remarquer que les intervalles obtenus pour les distances  $1-2$  et  $1-4$  sont similaires pour l'ensemble des coudes  $\beta$ . Pour la distance  $1-4$ , bien que l'intervalle soit plus restreint pour certains types ( $\text{VIa}_2$  et  $\text{IV}_1$ ), les valeurs observées, qui peuvent varier de plus de  $4,0 \text{ \AA}$ , ne peuvent pas être utilisées pour distinguer les différents types. Une conclusion similaire peut être tirée pour la distance  $1-2$ . En revanche, nous pouvons constater des différences significatives pour la distance  $2-3$  qui est faible ( $\sim 2-3 \text{ \AA}$ ) pour certains types (I, I',

III, III', IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub> et VIII) et élevées (>4 Å) pour d'autres (II, II' et IV<sub>1</sub>). La distance 3-4 est, quant à elle, relativement similaire pour tous les coudes β même si on peut toutefois noter qu'elle peut prendre des valeurs supérieures à 3,6 Å pour les types IV<sub>1</sub> et VIII. La distance 1-3 peut assumer des valeurs plutôt différents (entre 2,7 et 6,4 Å) hormis les types II, II' IV<sub>1</sub> pour lesquels une valeur supérieure à 4,4 Å est toujours retrouvée. De manière similaire, la distance 2-4 peut prendre des valeurs relativement différentes comprises entre 2,6 et 6,7 Å pour tous les types à l'exception du type VIII pour lequel la valeur mesurée est toujours élevée (>4,4 Å).

Une analyse analogue peut être menée pour les six distances H<sub>α</sub>-H<sub>N</sub> (entre les acides aminés 1-2 ; 2-3 ; 3-4 ; 1-3 ; 2-4 et 1-4). Comme pour les distances H<sub>N</sub>-H<sub>N</sub>, les intervalles obtenus pour les distances H<sub>α</sub>-H<sub>N</sub> 1-2 et 1-4 sont semblables pour l'ensemble des coudes β. Pour les distances 2-3 et 3-4, bien que les intervalles de distances se chevauchent, on peut noter que, pour certains coudes β, ces distances sont courtes alors que pour d'autres, elles prennent des valeurs intermédiaires. Ainsi la distance 2-3 est courte pour le type II (< 2,3 Å) et prend des valeurs moyennes (3-3,8 Å) dans les coudes de type I, III, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub> et VIII. De la même manière, la distance 3-4 est petite (< 2,5 Å) pour le type VIb et prend des valeurs moyennes (2,9-3,8 Å) dans les coudes de type I, I', II, III, IV<sub>3</sub>, IV<sub>4</sub>, VIa1 et VIa2. La distance H<sub>α</sub>-H<sub>N</sub> 2-4 peut varier fortement dans les coudes de type I, I', II, II', III, III', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub> et IV<sub>4</sub> mais elle prend des valeurs plutôt courtes dans les trois coudes de type VI (VIa1, VIa2 et VIb) alors qu'elle est longue pour le coude de type VIII. Enfin la distance H<sub>α</sub>-H<sub>N</sub> 1-3 est toujours élevée (> 4,6 Å) pour les coudes de type II, II' et IV<sub>1</sub>.

Tableau 21 Tableau des distances inter-hydrogènes (en Å) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) pour tous les types de coudes de nature  $\beta$ . ; « nd » indique les distances non disponibles.

Distances	Résidus	Type I	Type I'	Type II	Type II'	Type III	Type III'	Type IV <sub>1</sub>	Type IV <sub>2</sub>	Type IV <sub>3</sub>	Type IV <sub>4</sub>	proline en conformation <i>cis</i>			Type VIII
												Type VIa1	Type VIa2	Type VIb	
d $H_N H_N$	1-2	2,0-4,6	2,5-4,6	2,4-4,8	2,1-4,7	2,0-4,7	2,5-4,7	2,5-4,8	2,6-4,6	2,2-4,6	2,0-4,7	2,4-4,6	2,1-4,5	2,0-4,5	2,0-4,8
	2-3	2,4-3,5	2,4-3,6	<u>4,0-4,7</u>	<u>4,2-4,8</u>	2,0-3,5	2,4-3,6	<u>4,0-4,6</u>	2,1-3,2	2,0-3,4	1,7-3,3	nd	nd	nd	2,4-3,6
	3-4	2,0-3,3	2,1-3,1	2,3-3,5	2,1-3,6	2,2-3,2	2,3-3,3	2,4-4,8	2,1-3,5	2,1-3,2	2,0-2,9	nd	nd	nd	<u>2,9-4,1</u>
	1-3	2,9-6,1	2,7-6,0	4,4-7,2	4,6-7,3	3,9-6,5	2,7-6,0	4,6-8,1	3,6-6,4	3,2-6,4	2,5-6,1	nd	nd	nd	3,2-6,0
	2-4	3,5-4,7	3,4-4,9	2,9-5,0	3,7-5,1	3,1-5,0	3,5-5,0	2,7-4,8	2,6-5,3	3,3-5,5	2,8-6,7	3,9-5,0	3,0-4,5	3,5-5,2	4,4-5,2
	1-4	2,7-6,2	2,6-6,1	2,9-6,4	3,1-6,4	3,0-6,0	2,3-6,2	3,3-6,4	2,8-6,2	2,5-7,3	2,4-7,0	3,1-6,5	3,5-6,0	3,0-6,6	3,3-6,7
d $H_\alpha H_N$	1-2	2,2-3,8	2,1-3,6	2,2-3,7	2,0-3,6	2,1-3,7	2,1-3,6	2,2-3,7	2,1-3,7	2,2-3,7	2,2-3,7	2,1-3,6	2,3-3,7	2,2-3,7	2,1-3,7
	2-3	3,3-3,7	2,5-3,6	<u>2,0-2,3</u>	2,1-3,6	3,4-3,6	2,5-3,2	2,0-4,0	3,1-3,7	3,3-3,7	3,0-3,7	nd	nd	nd	3,4-3,8
	3-4	3,0-3,5	2,9-3,5	3,0-3,5	2,5-3,6	3,1-3,7	2,4-3,6	2,6-4,0	2,0-3,3	3,3-3,7	3,0-3,8	3,0-3,6	3,1-3,5	<u>2,1-2,5</u>	2,2-3,8
	1-3	3,0-5,3	3,1-5,7	4,6-6,4	4,9-6,4	3,1-5,7	3,0-5,2	4,9-7,6	3,3-5,7	3,0-5,7	2,1-5,3	nd	nd	nd	3,3-5,8
	2-4	3,2-4,5	3,5-4,8	2,7-4,2	3,1-4,9	3,2-5,1	3,7-5,4	2,1-4,0	3,5-5,0	3,4-6,1	3,2-6,1	2,1-3,6	2,0-3,3	2,0-3,7	3,9-5,2
	1-4	3,1-5,1	3,5-5,4	3,4-5,6	3,5-5,6	2,9-5,8	3,3-5,4	3,4-5,6	3,5-5,6	2,9-6,2	2,2-4,6	3,6-5,6	3,6-5,3	3,6-5,5	3,5-5,6

L'ensemble de ces observations suggère que, bien qu'il n'y ait pas de signatures nettement différentes pour chaque type de coude  $\beta$ , l'analyse simultanée de plusieurs distances  $H_N-H_N$  et  $H_\alpha-H_N$  peut permettre de discriminer les différents types.

### III.1.1.1.2 Le diagramme NOE

Comme vu précédemment au sein du chapitre 2, l'identification par RMN des structures secondaires dans une chaîne polypeptidique est réalisée (i) en générant à l'aide des distances expérimentales un diagramme NOE où ces distances sont classées en fonction de leur intensité en trois catégories : forte, moyenne et faible et (ii) en comparant le diagramme obtenu à un diagramme de référence pour chaque type de structure secondaire.

A partir du Tableau 21, nous avons donc généré un premier diagramme NOE général pour tous les types de coudes  $\beta$  (Figure 64).

Comme abordé au sein du chapitre 1, les structures de type coudes  $\beta$  présentent quatre résidus. Le premier résidu de ces structures secondaires est identifié comme  $i$ , le deuxième en tant que  $i+1$ , le troisième résidu est nommé  $i+2$  et le dernier  $i+3$ .

Dans ce diagramme, les distances entre hydrogènes sont représentées par des lignes d'épaisseur différente selon leur catégorie. Le type de distance (séquentielle ou à moyenne portée) est identifié sur le côté gauche de ce diagramme avec une nomenclature propre aux études de RMN. Par exemple, la distance entre deux  $H_N$  de deux résidus séquentiels dans la chaîne polypeptidique est identifiée comme  $d_{H_N-H_N}(i, i+1)$ .

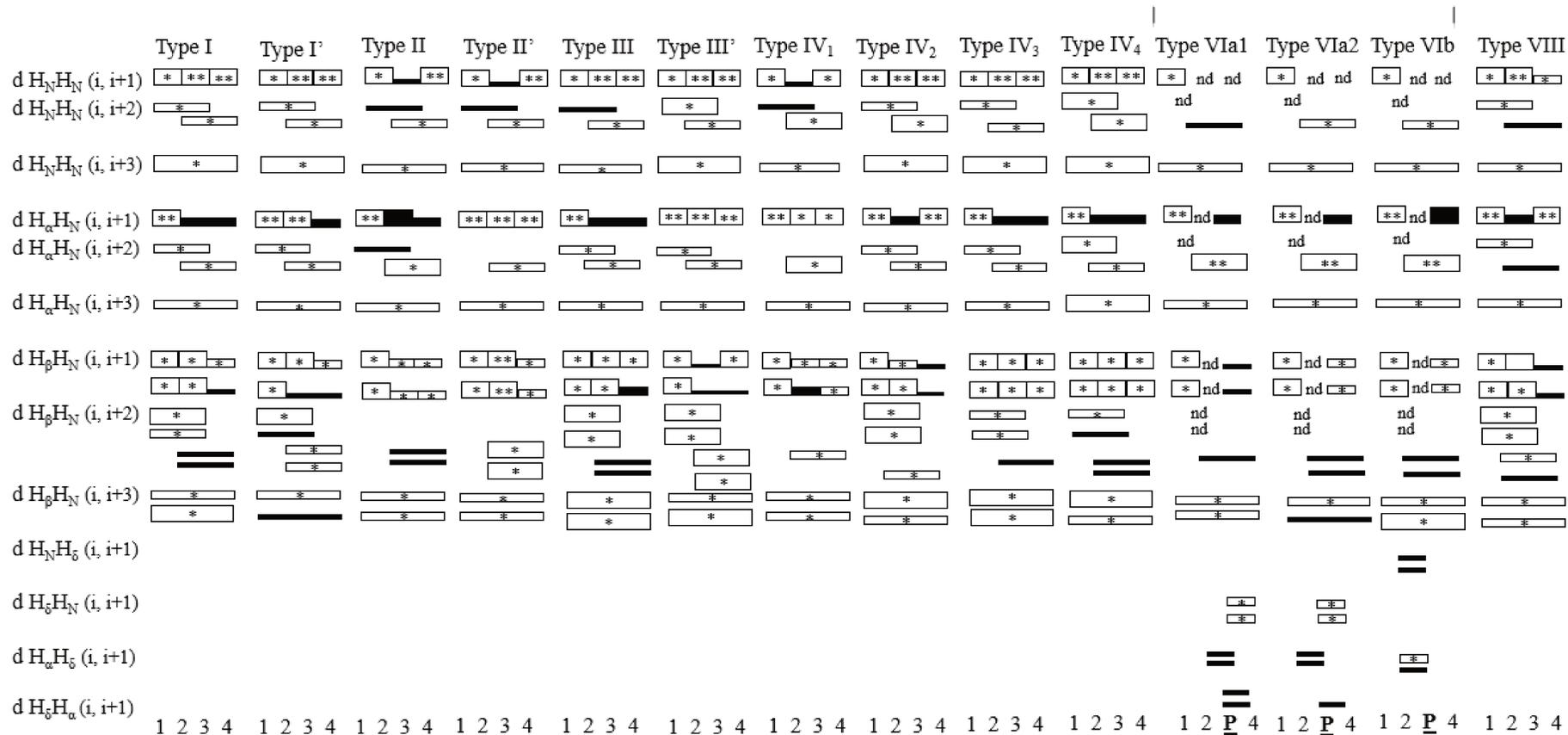


Figure 64 Représentation graphique en diagramme NOE des distances calculées entre les hydrogènes pour les tous les types coudes β. Les coudes VIa1, VIa2 et VIb sont caractérisés par la présence d'une proline en conformation cis en position 3. Un trait épais blanc avec deux étoiles ( **\*\*\*** ) correspond à une distance courte à moyenne (2,0 < d < 3,8 Å), un trait épais blanc avec une étoile ( **\*\*** ) correspond à une distance courte à longue (2,0 < d < 4,9 Å), un trait moyen blanc avec une étoile ( **\* \*** ) correspond à une distance moyenne à longue (2,9 < d < 4,9 Å), un trait épais noir ( **■** ) correspond à une distance courte (d < 2,8 Å) et un trait moyen noir ( **▬** ) indique une distance moyenne (2,9 < d < 3,8 Å) un trait fin noir ( **—** ) correspond à une distance longue (3,9 < d < 4,9 Å), « nd » indique les distances non disponibles.

Pour les coudes  $\beta$  sans proline, trois distances  $H_N-H_N$  séquentielles (1-2, 2-3 et 3-4) peuvent être observées. Parmi ces trois distances, la distance entre les résidus 1 et 2 n'est pas caractéristique selon le type de coude  $\beta$  (Figure 64). En revanche, l'analyse des deux autres distances  $H_N-H_N$  séquentielles montre que deux distances courtes à moyennes entre les acides aminés 2 et 3 ainsi qu'entre les résidus 3 et 4 sont caractéristiques des coudes de type I, I', III, III', IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub> (Figure 64). Les types II, II' et VIII sont caractérisés par une seule distance courte à moyenne. Pour le type VIII, cette signature est retrouvée entre les acides aminés 2 et 3, tandis que pour les types II et II' elle est entre les résidus 3 et 4 (Figure 64). Enfin, le type IV<sub>1</sub> est caractérisé par une distance courte à longue entre les acides aminés 3 et 4 (Figure 64). Nous pouvons donc proposer le premier organigramme d'aide à la décision suivant (Figure 65) :

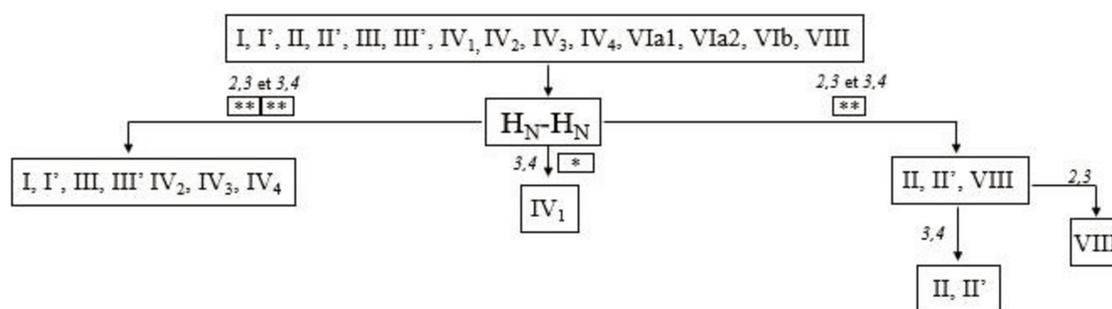


Figure 65 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  en fonction des distances  $H_N-H_N$  séquentielles. Un trait épais blanc avec deux étoiles (\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ) et un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ).

Les distances  $H_N-H_N$  (1-3, 2-4 et 1-4) ne montrent aucune signature caractéristique car des distance moyennes et des longues distances sont retrouvées dans tous les types de coudes  $\beta$  qui ne présentent pas, au sein de leur séquence, une proline.

Pour distinguer les coudes de type I, I', III, III', IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, les distances séquentielles entre les  $H_\alpha$  et les  $H_N$  doivent être analysées. Trois distances  $H_\alpha-H_N$  séquentielles peuvent être observées (1-2, 2-3 et 3-4). Les distances  $H_\alpha-H_N$  entre les résidus 1 et 2 ne sont pas caractéristiques selon le type de coude  $\beta$ . L'analyse des deux autres distances permet de mettre en évidence que les types I' et III' sont caractérisés par une distance courte à moyenne entre les acides aminés 2 et 3 (Figure 64). Pour le type IV<sub>2</sub>, nous pouvons observer une distance similaire, mais entre les acides aminés 3 et 4 (Figure 64). Dans les types I, III, IV<sub>3</sub> et IV<sub>4</sub> n'ont été observées que des distances moyennes entre les acides aminés centraux ainsi qu'entre le troisième et le quatrième acide aminé du coude (Figure 64).

La différenciation entre les coudes  $\beta$  de type II, II' est aussi possible en analysant les distances séquentielles  $H_\alpha-H_N$ . Pour le type II, nous retrouvons toujours une courte distance séquentielle entre le  $H_\alpha$  du deuxième résidu et le  $H_N$  du troisième acide aminé (Figure 64). En revanche, le type II' présente une gamme de distances courte-moyenne entre les acides aminés 2 et 3.

Enfin, le type IV<sub>1</sub> montre que la gamme de distances est encore plus étendue car elle va de courtes distances à des longues distances entre les deux résidus centraux de ce type de coude (2 et 3) comme montré dans le Tableau 21 et Figure 64.

De manière similaire aux distances entre les hydrogènes amides à moyenne portée, les distances  $H_\alpha-H_N$  (1-3, 2-4 et 1-4) n'apportent pas d'informations supplémentaires pour réaliser une distinction plus fine entre les différents types de coudes  $\beta$ . Ces observations permettent de continuer l'organigramme d'aide à la décision (Figure 66) :

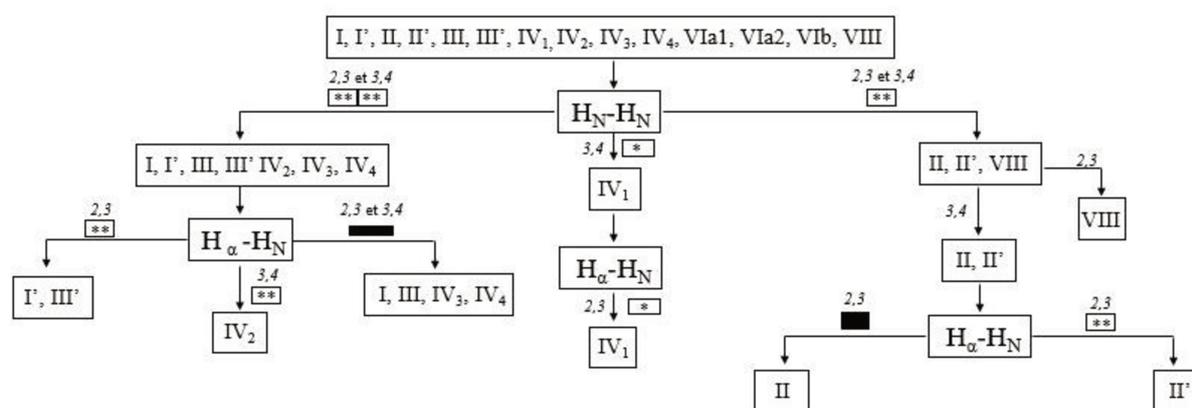


Figure 66 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  en fonction des distances  $H_\alpha-H_N$  séquentielles. Un trait épais blanc avec deux étoiles (\*\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait épais noir (■) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ) et un trait moyen noir (▬) indique une distance moyenne ( $2,8 < d < 3,8 \text{ \AA}$ )

Parmi les coudes  $\beta$ , les types VIa1, VIa2 et VIb sont des cas particuliers car ils possèdent une proline en conformation *cis* en position 3. La présence de la proline ne permet pas d'observer les distances séquentielles  $H_N-H_N$  entre les résidus 2 et 3, ainsi qu'entre les résidus 3 et 4 (Tableau 21 et Figure 64). De plus, la distance  $H_\alpha-H_N$  séquentielle entre les résidus 2 et 3 est absente (Tableau 21 et Figure 64). L'attention particulière que nous avons par rapport à ces distances est dû au fait qu'auparavant elles ont été nécessaires pour effectuer une différenciation entre les autres types de coudes  $\beta$ . Nous sommes tout de même aller rechercher des distances caractéristiques qui pourraient être caractéristiques des coudes  $\beta$  de type VIa1, VIa2 et VIb parmi les autres distances disponibles entre les hydrogènes du squelette peptidique. L'analyse des distances  $H_N-H_N$  (1-3 et 2-4) ne montre aucune signature particulière car des

distances longues sont retrouvées dans les trois types VIa1, VIa2 et VIb (Figure 64). En revanche, des informations intéressantes ont été observées en analysant les distances  $H_{\alpha}$ - $H_N$  séquentielles. Une courte distance  $H_{\alpha}$ - $H_N$  séquentielle entre les résidus 3 et 4 est caractéristique pour le type VIb (souligné dans le Tableau 21 et Figure 64) alors que pour les types VIa1 et VIa2 une distance moyenne a été observée (souligné dans le Tableau 21 et Figure 64). L'organigramme d'aide à la décision précédent peut, par conséquent, être complété comme ci-dessous (Figure 67).

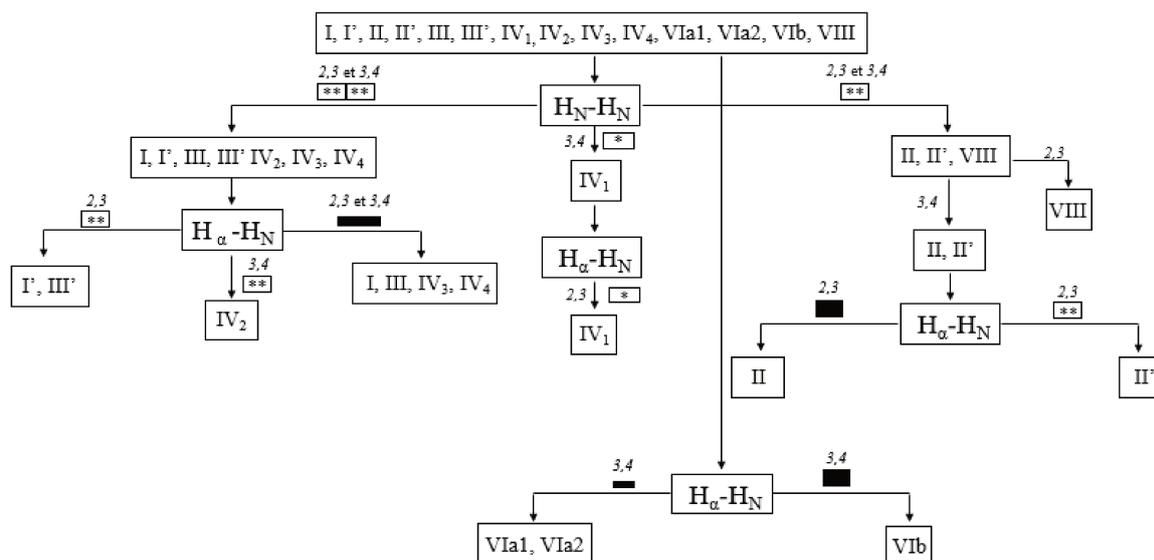


Figure 67 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  en fonction des distances  $H_{\alpha}$ - $H_N$  séquentielles. Un trait épais blanc avec deux étoiles ( **\*\*** ) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile ( **\*** ) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait épais noir ( **■** ) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ) et un trait moyen noir ( **▬** ) indique une distance moyenne ( $2,8 < d < 3,8 \text{ \AA}$ )

L'analyse des distances  $^1\text{H}$ - $^1\text{H}$  du squelette peptidique, nous a permis, dans un premier temps, de mettre en évidence que, pour les coudes  $\beta$ , une différenciation entre les différents types est possible. Cependant, l'ambiguïté entre certains types persiste. Ceci n'est pas aussi étonnant que ça. En effet, dans la classification présentée au sein du Chapitre 1, les types I'/III' et les types I/III ont été regroupés à cause de leur similarité au niveau géométrique. En outre, le fait de ne pas pouvoir effectuer une distinction entre les types I, IV<sub>3</sub> et IV<sub>4</sub> est dû au fait que les types IV<sub>3</sub> et IV<sub>4</sub> possèdent une forte similarité structurale et qu'ils sont dans la même région d'angles dièdres que le type I.<sup>45</sup> Enfin, l'ambiguïté entre les types VIa1 et VIa2 n'est pas surprenante car la différence entre ces deux types de coudes  $\beta$  se joue sur les valeurs des angles  $\Phi$  du résidu 2 qui sont respectivement  $-60^\circ$  et  $-120^\circ$  (Tableau 10).

### III.1.1.1.3 Distances entre la chaîne latérale et le squelette peptidique

Afin de réaliser une analyse complète, d'apporter des informations supplémentaires et de permettre de résoudre les ambiguïtés qui persistent entre certains types de coude, les distances entre les hydrogènes de la chaîne latérale (limités aux  $H_{\beta}$ ) et ceux du squelette peptidique ont été analysées. Six types de distances  $H_{\beta}$ - $H_N$  (entre les acides aminés 1-2 ; 2-3 ; 3-4 ; 1-3 ; 2-4 ; 1-4) ont pu être obtenues (Tableau 22). De plus, pour les coudes  $\beta$  de type VIa1, VIa2 et VIb, qui ont une *cis* proline en position 3, l'analyse entre les hydrogènes  $H_{\delta}$  et ceux du squelette peptidique ( $H_{\alpha}$  et  $H_N$ ) a été aussi effectuée. Toutes ces distances sont présentées dans le tableau 22. Pour chaque ligne, la première valeur correspond à une distance impliquant l'hydrogène  $H_{\beta 2}$  et la seconde à une distance impliquant l'hydrogène  $H_{\beta 3}$ . Nous pouvons constater que les intervalles sont similaires pour les deux types de  $H_{\beta}$  pour la plupart des distances mesurées. Nous ne ferons par conséquent pas de distinction lors de l'analyse du tableau 22 entre les deux protons  $H_{\beta}$  sauf lorsque des différences spécifiques sont observées.

Tableau 22 Tableau des distances inter-hydrogènes entre les atomes de la chaîne latérale ( $H_\beta$  et  $H_\delta$ ) et les hydrogènes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) pour tous les types de coudes  $\beta$ . La première ligne correspond à la gamme de distances pour les  $H_{\beta 2}$ ; la deuxième ligne correspond à la gamme de distances pour les  $H_{\beta 3}$ . Pour les coudes présentant une proline en position 3 (VIa1, VIa2 et VIb), pour les distances impliquant une  $H_\delta$ : la première ligne correspond à la gamme de distances pour les  $H_{\delta 2}$ ; la deuxième ligne correspond à la gamme de distances pour les  $H_{\delta 3}$ ; « nd » indique les distances non disponibles.

Distances	Résidus	Type I	Type I'	Type II	Type II'	Type III	Type III'	Type IV <sub>1</sub>	Type IV <sub>2</sub>	Type IV <sub>3</sub>	Type IV <sub>4</sub>	proline en conformation cis			Type VIII
												Type VIa1	Type VIa2	Type VIb	
d $H_\beta H_N$	1-2	1,9-4,6	1,8-4,6	1,9-4,7	2,0-4,8	1,7-4,7	1,8-4,8	2,0-4,9	1,8-4,7	1,9-4,7	1,9-5,4	2,0-4,7	1,9-4,5	1,8-4,8	1,9-4,6
		2,2-4,7	2,0-4,7	1,9-4,8	2,1-4,8	1,9-4,7 <sup>c</sup>	2,0-4,8	2,0-4,7	1,7-4,4	2,1-4,7	2,1-4,9	3,1-4,7	1,9-4,7	1,9-4,6 <sup>c</sup>	1,7-4,7
	2-3	2,3-4,5 2,5-4,5	2,6-4,6 4,0-4,6	3,5-4,5 3,3-4,5	1,6-3,8 1,8-3,5	2,2-4,4 2,4-4,1	3,9-4,7 3,9-4,6	3,0-4,6 2,9-3,7	3,4-4,5 2,2-4,6	2,4-4,2 2,5-4,7	2,8-4,7 2,5-4,8	nd	nd	nd	2,3-4,0 2,6-4,4
d $H_\beta H_N$	3-4	3,2-4,9 3,9-4,7	3,1-4,5 4,2-4,7	3,0-4,7 3,3-4,4	2,9-4,9 2,9-4,7	2,3-4,5 2,9-3,9	2,6-4,6 3,8-4,8	2,9-4,8 3,8-4,6	4,0-4,8 4,1-4,8	2,0-4,3 1,9-4,2	2,7-4,5 2,6-4,6	4,0-5,0 4,0-4,9	3,6-4,8 3,4-4,8	3,3-4,5 3,1-4,4	4,4-5,4 3,9-4,8
		1-3	2,7-6,3 2,9-6,0	2,8-6,1 4,3-6,3	5,0-7,7 5,0-7,5	4,9-7,9 4,5-7,7	2,5-6,1 2,6-6,5	2,8-6,4 2,7-6,3	4,8-8,9 5,0-8,1	2,5-5,8 2,3-6,9	2,9-6,2 2,7-6,2	2,9-6,3 3,5-6,5	nd	nd	nd
d $H_\beta H_N$	2-4	4,5-5,7 4,1-6,0	3,0-5,7 3,3-5,7	4,8-6,0 4,1-5,7	2,2-5,4 2,6-5,3	4,2-6,3 4,3-6,2	2,6-6,2 2,4-6,1	2,8-5,8 4,8-5,6	4,6-6,5 3,7-6,2	4,7-6,5 5,3-7,2	4,5-7,4 4,6-8,2	4,2-5,8 4,8-5,7	4,6-5,2 4,2-5,2	4,4-6,4 4,2-6,3	3,7-6,4 4,3-6,6
		1-4	3,0-6,1 2,9-6,0	2,9-6,7 4,2-6,6	3,0-6,6 3,0-6,1	3,2-6,7 2,9-6,7	2,0-6,1 2,5-6,9	3,5-6,8 2,4-6,5	2,9-6,6 2,9-6,5	2,6-5,9 3,1-6,2	2,3-7,6 2,0-6,7	2,5-9,3 3,4-6,4	3,5-6,2 3,4-6,7	3,7-5,9 4,0-5,7	3,4-6,4 2,2-6,4
d $H_N H_\delta$	2-3											4,9-5,9 4,8-6,1	4,7-5,9 4,8-6,0	4,3-5,6 4,3-6,0	
d $H_\delta H_N$	3-4											2,9-4,7 2,9-4,8	3,7-4,9 3,4-4,8	5,4-5,9 5,4-5,8	
d $H_\alpha H_\delta$	2-3											4,0-4,8 3,9-4,7	3,9-4,8 4,3-4,7	3,5-4,7 4,0-4,7	
d $H_\delta H_\alpha$	3-4											4,3-6,6 4,1-7,0	4,9-6,9 4,5-7,1	5,7-7,7 5,9-7,5	

Comme pour les distances du squelette peptidique, les distances  $H_{\beta}$ - $H_N$  1-2 et 1-4 ont des intervalles similaires pour l'ensemble des coudes  $\beta$  (Tableau 22). Pour les distances 2-3 et 3-4, bien que les gammes de distances soient étendues, nous pouvons remarquer que, pour quelques types de coudes, certains intervalles sont plus étroits avec des distances longues (3,9-4,9 Å) (Tableau 22). Par exemple, les distances 2-3 sont longues pour le type III' (3,9-4,7 Å) ainsi que la distance  $H_{\beta 3}$ - $H_N$  pour le type I'. Ces distances ont une plage de valeurs très large (2,0-4,8 Å) pour les coudes  $\beta$  de type I, III, IV<sub>3</sub>, IV<sub>4</sub> et VIII alors que des intervalles moins grands sont retrouvés pour le type IV<sub>1</sub>. Ces distances ne sont pas disponibles pour les trois types VI à cause de l'absence de l'hydrogène  $H_N$  pour le résidu 3 (Tableau 22). De la même manière, les distances 3-4 ont des intervalles de distances amples pour tous les types de coudes sauf pour les types IV<sub>2</sub>, VIa1 et VIII (Tableau 22).

Les distances  $H_{\beta}$ - $H_N$  1-3 peuvent varier fortement dans tous les types de coudes  $\beta$ . Nous pouvons toutefois noter qu'elles ne seront pas détectables par RMN dans les types II et IV<sub>1</sub> car leur valeur est supérieure à 5 Å (Tableau 22). Les distances 2-4 sont élevées (> 4,0 Å) dans les types I, II, III, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb et VIII tandis que, dans les types II' et III', elles peuvent prendre des valeurs plus faibles (jusqu'à 2,2 Å) mais arriver jusqu'à 5 Å (Tableau 22).

L'analyse des distances entre les  $H_{\delta}$  et le squelette peptidique permet de mettre en évidence des différences entre les types VIa1, VIa2 et VIb. Comme fait précédemment pour les  $H_{\beta}$ , lors de l'analyse des distances impliquant les hydrogènes  $H_{\delta}$ , nous ne ferons pas la distinction entre les deux hydrogènes  $H_{\delta}$ . De manière remarquable, la distance  $H_{\delta}$ - $H_N$  3-4 est élevée (>5,4 Å) dans le type VIb alors qu'elle peut prendre des valeurs plus faibles dans le type VIa1 (2,9 Å) et le type VIa2 (3,4 Å) (Tableau 22).

L'étude de la chaîne latérale nous a permis de mettre en évidence que des intervalles de distances plutôt larges peuvent être retrouvés. Cependant, nous pouvons observer que des gammes de valeurs plus restreintes peuvent apporter des informations supplémentaires afin d'affiner distinction entre certains types de coude  $\beta$ . Par exemple, la distance moyenne à faible  $H_{\beta 2}$ - $H_N$  (3-4) ainsi que la longue distance  $H_{\beta 3}$ - $H_N$  (3-4) sont une signature pour le type I alors que, dans les types IV<sub>3</sub> et IV<sub>4</sub>, les intervalles de distances sont plus étendus. Le type III est caractérisé par une distance moyenne  $H_{\beta 3}$ - $H_N$  (3-4) (Figure 64).

De la même manière, les longues distances  $H_{\beta}$ - $H_N$  2-3 sont une signature pour le type III' tandis que dans le type I' la distance  $H_{\beta 2}$ - $H_N$  (2-3) est forte à faible (Figure 64).

Pour les types II et II', des signatures ont été aussi identifiées. Les distances  $H_{\beta}$ - $H_N$  2-3 moyennes à faibles sont caractéristiques du type II tandis que les mêmes distances fortes à moyennes sont retrouvées pour le type II' (Figure 64). Celles-ci nous permettent de confirmer le type de coude  $\beta$ .

Enfin, les distances  $H_{\delta}$ - $H_N$  entre le troisième et quatrième résidu permettent de mettre en évidence que des distances supérieures à 5 Å sont retrouvées dans le type VIb alors que des distances moyenne à longue sont observables dans les types VIa1 et VIa2 (Figure 64).

Il est alors possible de compléter l'organigramme d'aide à la décision selon la figure 68.

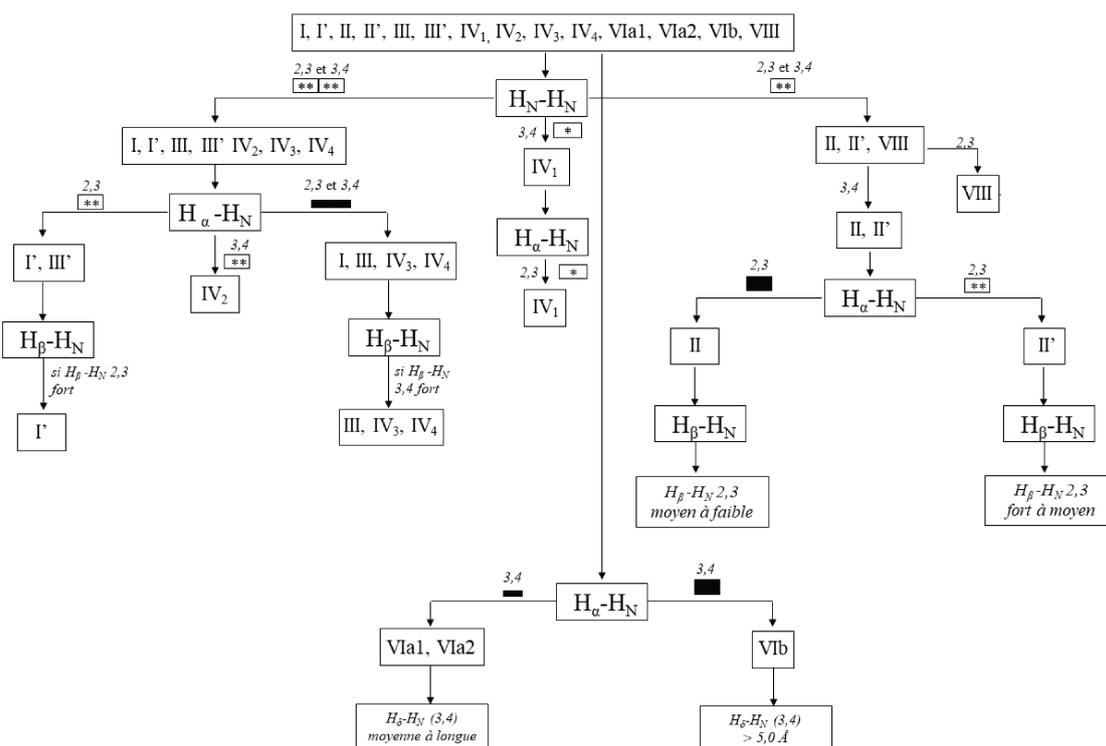


Figure 68 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  en fonction des distances entre les atomes d'hydrogènes de la chaînes latérale (limité aux  $H_{\beta}$ , ainsi qu'aux  $H_{\delta}$  en présence d'une proline) et les atomes d'hydrogènes du squelette peptidique ( $H_N$  et  $H_{\alpha}$ ). Un trait épais blanc avec deux étoiles (\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait épais noir (■) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ) et un trait moyen noir (▬) indique une distance moyenne ( $2,8 < d < 3,8 \text{ \AA}$ ).

#### **III.1.1.1.4 Influence du motif**

Afin d'apporter des informations supplémentaires, nous avons analysé les distances pour chaque motif présent dans notre banque de données. En effet, les conformations structurales (hélice  $\alpha$ , brin  $\beta$  et/ou boucle) dans lesquelles le premier résidu ou le dernier résidu sont engagés peuvent avoir un impact sur les distances mesurées.

Parmi l'ensemble des 9 motifs disponibles, l'analyse a montré que des différences n'étaient rencontrées que lorsque la conformation du premier résidu était différente. Nous avons donc regroupé, pour l'analyse, les motifs H-T-H, H-T-R et H-T-S dans le motif général H-T-X ; les motifs R-T-H, R-T-R et R-T-S dans le motif général R-T-X et, enfin, les motifs S-T-H, S-T-R et S-T-S dans le motif général S-T-X. Ceci nous a permis d'établir trois tableaux différents de distances (Tableaux 23, 26 et 27) et deux différents diagrammes NOE (Figures 69 et 71) selon la conformation dans laquelle le premier résidu est engagé.

Tableau 23 Tableau des distances inter-hydrogènes (en Å) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) pour tous les types de coudes de nature  $\beta$ . La première ligne indique quand le résidu est en conformation hélice, la deuxième ligne quand il est conformation aléatoire et la dernière ligne quand il est en conformation brin  $\beta$ . « nd » indique les distances non disponibles.

Distances	Residues	Type I	Type I'	Type II	Type II'	Type III	Type III'	Type IV <sub>1</sub>	Type IV <sub>2</sub>	Type IV <sub>3</sub>	Type IV <sub>4</sub>	proline en conformation <i>cis</i>			Type VIII
												Type VIa1	Type VIa2	Type VIb	
d $H_N H_N$	1-2	2,5-3,3	2,5-2,8	2,6-2,9	2,1-3,5	2,6-3,1	2,5-3,6	2,7-3,2	2,6-3,6	2,5-3,2	2,6-3,4	2,5-2,8	2,5-2,8	2,4-2,8	2,5-3,8
		2,0-4,6	2,6-4,6	2,4-4,8	2,4-4,7	2,0-4,7	2,7-4,7	2,5-4,8	2,7-4,6	2,2-4,6	2,0-4,7	2,4-4,6	2,1-4,4	2,0-4,4	1,9-4,8
		4,2-4,5	4,1-4,4	4,0-4,5	3,9-4,4	3,9-4,5	4,1-4,6	4,1-4,5	4,1-4,6	3,9-4,5	3,9-4,7	3,9-4,4	4,1-4,5	3,5-4,5	4,2-4,5
	2-3	2,5-3,5	2,8-3,2	4,2-4,6	4,2-4,7	2,7-3,0	2,8-3,6	4,1-4,5	2,1-3,2	2,6-3,3	1,8-2,9	nd	nd	nd	2,4-3,5
		2,4-3,5	2,4-3,6	4,0-4,7	4,3-4,8	2,2-3,5	2,4-3,6	4,0-4,6	2,1-3,2	2,0-3,4	1,7-3,3				2,6-3,6
		2,5-3,1	2,6-3,1	4,2-4,6	4,3-4,7	2,0-3,3	2,5-3,3	4,2-4,6	2,7-2,9	2,2-3,0	2,0-2,9				2,7-3,3
	3-4	2,0-3,1	2,5-3,0	2,4-3,2	2,2-2,9	2,3-2,8	2,3-3,0	2,5-4,6	2,1-3,3	2,3-3,1	1,9-2,5	nd	nd	nd	3,4-3,8
		2,2-3,3	2,1-3,1	2,3-3,5	2,1-3,6	2,2-3,2	2,4-3,2	2,4-4,8	2,1-3,5	2,1-3,2	1,9-2,8				2,9-3,8
		2,1-2,9	2,4-3,1	2,4-3,3	2,0-3,3	2,4-3,0	2,6-3,3	2,6-4,6	2,3-3,4	2,1-3,2	2,0-2,7				2,9-3,8
	1-3	3,6-5,8	4,5-5,5	6,4-7,1	6,3-7,3	4,2-4,6	4,7-5,4	6,4-7,2	3,6-5,9	3,2-5,5	3,3-4,5	nd	nd	nd	3,2-6,0
		2,8-6,1	2,7-6,0	4,4-7,2	4,6-7,3	3,7-6,5	2,7-6,0	4,6-8,1	3,6-6,4	3,4-6,3	2,5-6,1				3,4-6,3
		4,8-5,4	3,3-4,6	4,5-6,1	5,2-6,6	4,8-5,7	3,3-4,9	4,6-7,7	4,7-5,3	4,5-6,0	4,1-5,8				4,7-6,0
	2-4	3,8-4,7	4,0-4,6	4,1-4,8	3,8-4,7	4,0-5,0	3,9-4,8	2,9-4,7	2,7-4,7	3,9-4,5	2,9-4,7	3,9-4,8	3,4-4,4	4,7-5,0	4,5-5,3
		3,5-4,7	3,4-4,9	2,9-5,2	3,7-5,1	3,1-4,9	3,5-5,3	2,7-4,8	2,6-5,2	3,7-5,5	2,8-6,7	3,9-5,0	3,2-4,5	3,5-5,2	4,4-5,2
		3,8-4,7	4,1-4,7	4,2-5,0	4,3-4,9	3,9-4,6	4,1-5,0	3,4-4,8	4,2-5,3	3,3-5,2	3,2-4,6	3,9-4,6	3,0-4,5	3,9-5,0	4,7-5,2
	1-4	4,4-6,4	5,4-6,1	5,2-6,4	4,8-6,2	5,0-6,0	3,7-5,9	5,0-6,4	2,6-5,9	4,0-6,3	2,9-5,7	5,3-6,3	4,8-6,0	5,2-6,5	4,2-6,7
		2,7-6,2	2,6-5,9	2,8-6,2	3,1-6,4	3,0-6,5	2,3-6,2	3,3-6,3	2,8-6,2	2,5-7,3	2,4-7,0	3,1-6,5	3,5-5,8	3,0-6,6	3,3-6,7
		3,6-4,8	2,5-4,1	3,2-4,6	3,4-4,6	3,5-5,1	2,4-4,3	3,9-4,9	3,5-4,4	3,7-6,5	3,1-6,8	3,3-5,4	3,7-4,9	3,9-5,7	3,5-4,7
d $H_\alpha H_N$	1-2	3,4-3,6	3,5-3,6	3,5-3,7	2,4-3,6	3,5-3,6	3,3-3,6	2,6-3,7	2,5-3,7	3,5-3,7	2,6-3,7	3,2-3,6	3,6-3,7	3,4-3,7	2,1-3,7
		2,2-3,8	2,1-3,6	2,2-3,6	2,0-3,6	2,1-3,7	2,1-3,6	2,2-3,7	2,1-3,7	2,2-3,6	2,2-3,7	2,1-3,6	2,3-3,7	2,2-3,7	2,1-3,7
		2,1-2,4	2,1-2,3	2,2-2,6	2,1-3,2	2,1-2,4	2,2-2,4	2,2-3,5	2,3-3,4	2,1-2,4	2,2-3,6	2,1-2,5	2,4-2,6	2,2-3,4	2,1-2,4
	2-3	3,4-3,7	2,5-3,5	2,1-2,2	2,2-3,6	3,5-3,6	2,5-3,2	2,1-3,7	3,2-3,6	3,3-3,7	3,1-3,6	nd	nd	nd	3,4-3,7
		3,3-3,7	2,5-3,6	2,0-2,3	2,1-3,6	3,4-3,6	2,6-3,2	2,0-4,0	3,1-3,7	3,4-3,7	3,0-3,7				3,5-3,8
		3,4-3,7	2,6-3,6	2,1-2,3	2,2-3,5	3,4-3,6	2,5-3,1	2,0-4,0	3,2-3,7	3,4-3,6	3,0-3,6				3,5-3,8

	3-4	3,1-3,5 3,0-3,5 3,1-3,5	3,1-3,4 2,9-3,5 2,9-3,5	3,1-3,6 3,0-3,6 3,0-3,5	3,0-3,6 2,5-3,6 2,9-3,6	3,4-3,6 3,1-3,7 3,2-3,6	2,8-3,3 2,4-3,6 2,6-3,6	3,1-3,9 2,6-4,0 2,7-3,0	2,5-3,1 2,3-3,1 2,0-3,3	3,5-3,6 3,3-3,7 3,5-3,7	3,9-5,3 3,0-4,0 3,1-3,7	3,3-3,6 3,0-3,5 3,1-3,5	3,3-3,4 3,1-3,5 3,1-3,5	2,2-2,3 2,1-2,5 2,2-2,4	2,2-3,0 2,1-3,5 2,2-2,5
	1-3	3,3-5,6 3,0-5,3 3,4-4,6	4,1-5,3 3,1-5,7 3,4-4,2	5,1-6,2 4,6-6,4 5,3-6,3	4,9-6,4 5,1-6,6 5,6-6,4	3,9-4,4 3,2-5,7 3,1-4,9	3,9-5,2 3,0-4,9 3,1-4,2	5,3-7,1 4,9-7,6 5,3-7,4	4,2-5,7 3,3-5,6 3,3-4,6	3,6-5,3 3,0-5,7 3,4-4,5	3,4-5,5 2,1-5,3 2,5-4,4	nd	nd	nd	3,9-5,8 3,3-5,7 3,9-5,1
	2-4	3,2-4,5 3,4-4,4 3,4-4,3	3,9-4,5 3,5-4,8 3,7-4,8	2,7-3,9 3,0-4,2 3,2-4,0	3,4-4,6 3,1-4,9 3,3-4,6	3,6-5,1 3,2-4,9 3,9-4,9	3,9-4,8 3,7-5,4 4,2-5,2	2,3-4,0 2,1-4,0 2,4-3,7	3,6-5,0 3,5-4,8 3,7-4,8	3,5-5,1 3,4-6,1 3,4-5,3	3,9-4,1 3,2-6,1 3,4-5,0	2,1-3,4 2,1-3,6 2,1-3,5	2,4-3,1 2,0-3,3 2,4-2,7	2,5-3,2 2,1-3,6 2,0-3,7	3,9-4,9 3,9-5,2 4,0-5,0
	1-4	3,1-4,4 3,1-5,1 4,3-5,3	3,9-4,8 3,5-5,4 4,1-5,2	3,4-4,6 3,5-5,6 4,8-5,5	4,0-5,4 3,5-5,6 4,8-5,5	3,5-4,2 2,9-5,8 2,9-5,0	3,4-4,6 3,3-5,4 4,2-5,3	4,0-4,7 3,4-5,6 4,6-5,4	4,0-6,0 3,5-5,4 3,5-5,6	2,9-5,2 2,9-6,2 3,0-6,1	3,6-6,0 2,2-7,6 2,9-5,6	4,1-4,8 3,6-5,6 4,6-5,4	4,1-4,7 3,6-5,3 3,9-5,3	4,0-4,7 3,6-5,5 4,1-5,5	3,9-5,3 3,5-5,6 5,1-5,6

### Distances entre les hydrogènes du squelette peptidique ( $H_\alpha$ et $H_N$ )

Nous présenterons d'abord les coudes  $\beta$  sans proline et ensuite les coudes  $\beta$  de type VIa1, VIa2 et VIb qui présentent une proline en conformation *cis* en position 3. L'analyse du tableau 23 indique que la conformation structurale (hélice  $\alpha$ , brin  $\beta$  et/ou aléatoire) dans laquelle le premier résidu est engagé impacte principalement les distances  $H_N-H_N$  et  $H_\alpha-H_N$  séquentielles ( $I-2$ ) et impacte peu les autres distances entre les hydrogènes du squelette peptidique ( $H_\alpha$  et  $H_N$ ).

Si le premier résidu est en conformation aléatoire une gamme de distances qui peut être comprise entre une courte distance ( $>1,9 \text{ \AA}$ ) et une longue distance ( $<4,8 \text{ \AA}$ ) peut être observée pour les distances  $H_N-H_N$  séquentielles entre les résidus  $I$  et  $2$  de tous les type de coudes  $\beta$ . De manière similaire, pour les distances  $H_\alpha-H_N$  séquentielles, une gamme de distances courtes à moyennes est retrouvée dans tous les types de coudes  $\beta$  entre le résidu  $I$  et  $2$  (Tableau 23 et Tableau 24).

L'analyse de ces mêmes distances quand le résidu  $I$  est engagé dans une conformation hélicoïdale nous a permis de mettre en évidence des distances courtes à moyennes ( $2,0-3,8 \text{ \AA}$ ) pour les  $H_N-H_N$  ( $I,2$ ). Des distances moyennes ( $3,2-3,7 \text{ \AA}$ ) ont été retrouvées pour les distances  $H_\alpha-H_N$  séquentielles ( $I,2$ ) dans les coudes de types I, I', II, III, III' et IV<sub>3</sub>. Ces distances peuvent être plus courtes ( $>2,1 \text{ \AA}$ ) dans les types II', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>4</sub> et VIII (Tableau 23 et Tableau 24). Enfin, quand le premier résidu adopte une conformation en brin  $\beta$ , des distances  $H_N-H_N$  longues ( $3,9-4,9 \text{ \AA}$ ) et de distances  $H_\alpha-H_N$  courtes ( $2,0-2,8 \text{ \AA}$ ) sont observées pour tous les types de coudes  $\beta$  (Tableau 23 et Tableau 24).

Tableau 24 Tableau des distances inter-hydrogènes (en  $\text{\AA}$ ) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) des résidus  $I$  et  $2$  selon la conformation aléatoire, hélice ou brin  $\beta$  du résidu  $I$  pour les coudes  $\beta$  sans proline.

Conformation du résidu $I$	Distances séquentielles $H_N-H_N$ ( $I-2$ ) en $\text{\AA}$	Distances séquentielles $H_\alpha-H_N$ ( $I-2$ ) en $\text{\AA}$
<b>Aléatoire</b>	1,9-4,8	2,0-3,8
<b>Hélice</b>	2,0-3,8	2,1-3,8 (types II', IV <sub>1</sub> , IV <sub>2</sub> , IV <sub>4</sub> , VIII) 3,3-3,7 (types I, I', II, III, III', IV <sub>3</sub> )
<b>Brin <math>\beta</math></b>	3,9-4,9	2,0-2,8

Pour les coudes  $\beta$  de type VIa1, VIa2 et VIb, quand le premier résidu est engagé dans une conformation aléatoire, des intervalles de distances importants sont retrouvés pour les distances  $H_N-H_N$  et  $H_\alpha-H_N$  séquentielles (Tableau 23 et Tableau 25). Quand l'acide aminé  $I$  est en hélice, une faible distance  $H_N-H_N$  et une distance moyenne  $H_\alpha-H_N$  sont observées. Enfin, la

conformation en brin  $\beta$  a pour effet d'augmenter la distance  $H_N-H_N$  et de réduire la distance  $H_\alpha-H_N$  (Tableau 23 et Tableau 25).

Tableau 25 Tableau des distances inter-hydrogènes (en Å) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) des résidus 1 et 2 selon la conformation aléatoire, hélice ou brin  $\beta$  du résidu 1 pour les coudes  $\beta$  de type VIa1, VIa2, VIb.

Conformation du résidu $i$	Distances séquentielles $H_N-H_N$ (1-2) en Å	Distances séquentielles $H_\alpha-H_N$ (1-2) en Å
<b>Aléatoire</b>	2,0-4,6	2,0-3,8
<b>Hélice</b>	2,4-2,8	3,2-3,7
<b>Brin <math>\beta</math></b>	3,9-4,5	2,1-2,8

De manière générale, pour toutes les autres distances  $H_N-H_N$  et  $H_\alpha-H_N$  séquentielles et à moyenne portée, quand le premier résidu est en conformation aléatoire, des gammes de distances plus larges sont retrouvées par rapport au cas où l'acide aminé 1 est engagé dans une conformation hélice ou feuillet. Ces gammes correspondent aux distances et effets présentés précédemment (*cf.* partie III.1.1.1.1).

La conformation en hélice ou en brin  $\beta$  pour le premier acide aminé permet en général de restreindre l'intervalle de distance par rapport à une conformation désordonnée pour ce résidu (Tableau 23). Nous pouvons observer un impact du motif sur certaines distances, notamment les distances 1-3 et 1-4, pour certains types de coude  $\beta$  comme par exemple dans le type I' une distance moyenne 1-4 est observable quand le premier résidu est en conformation brin  $\beta$  et la même distance n'est plus mesurable quand le résidu est en conformation hélicoïdale.

Ces effets restent toutefois marginaux et ne donnent pas d'informations discriminantes supplémentaires.

## **Distances entre les hydrogènes de la chaîne latérale et les hydrogènes du squelette peptidique**

Afin de compléter notre analyse, nous nous sommes également intéressés à l'effet du motif sur les distances impliquant des hydrogènes de la chaîne latérale. Comme pour le squelette peptidique, nous avons constaté que c'est la conformation adoptée par le premier acide aminé du coude qui a un effet sur les distances mesurées (Tableaux 26 et 27).

Lorsque le premier acide aminé est engagé dans une structuration secondaire désordonnée, les intervalles observés sont importants et correspondent à ceux présentés au paragraphe III.1.1.1.3 (Tableau 22). Nous avons donc choisi de reporter uniquement les valeurs obtenues lorsque le premier acide aminé est engagé dans une conformation en hélice  $\alpha$  (HTX) (Tableau 26) ou en brin  $\beta$  (STX) (Tableau 27). Ceci nous a permis de relever des intervalles de distances discriminants supplémentaires en fonction du motif HTX ou STX. Ces intervalles sont en gras et soulignés dans les tableaux.

Pour la suite de l'analyse, nous avons choisi de générer des diagrammes NOE pour l'ensemble des types de coudes par motif HTX ou STX et de présenter, à l'aide de ces diagrammes, l'impact de l'influence de la conformation du premier résidu sur les organigrammes d'aide à la décision.

### **Analyse du diagramme NOE**

Contrairement au cas général, l'analyse par motif fait apparaître des différences entre les protons  $H_{\beta}$ . Ces derniers seront donc distingués dans la suite de l'analyse. Sur chaque diagramme NOE et dans chaque tableau, la première ligne correspond à l'intervalle du  $H_{\beta 2}$  alors que la deuxième ligne correspondra au  $H_{\beta 3}$ .

#### **III.1.1.1.4.1 Influence du premier résidu en conformation hélice $\alpha$**

La figure 69 représente le diagramme NOE quand le premier résidu est en conformation hélicoïdale. Sur ce diagramme, nous pouvons retrouver des signatures qui nous permettent de différencier le type I' du type III', le type VIa1 du VIa2 ainsi que les types I, III, IV<sub>3</sub> et IV<sub>4</sub> entre eux.

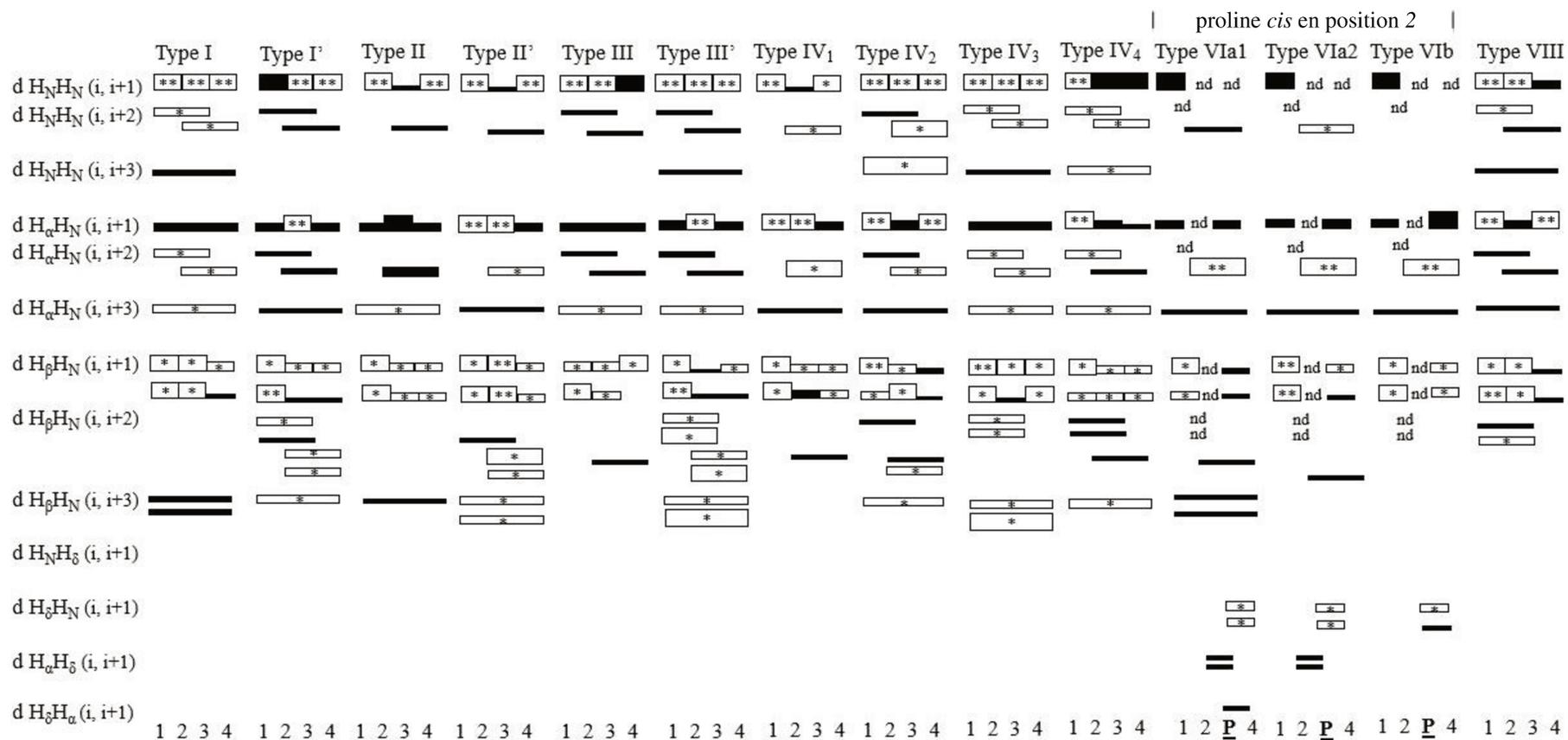


Figure 69 Représentation graphique en diagramme NOE des distances calculées entre les hydrogènes pour les tous les types coudes β quand le premier résidu est en conformation hélice. Les coudes VIa1, VIa2 et VIb sont caractérisés par la présence d'une proline en conformation *cis* en position 3. Un trait épais blanc avec deux étoiles (\*\*) correspond à une distance courte à moyenne (2,0 < d < 3,8 Å), un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue (2,0 < d < 4,9 Å), un trait moyen blanc avec une étoile (\*) correspond à une distance moyenne à longue (2,9 < d < 4,9 Å), un trait épais noir (■) correspond à une distance courte (d < 2,8 Å), un trait moyen noir (▬) indique une distance moyenne (2,9 < d < 3,8 Å) et un trait fin noir (—) correspond à une distance longue (3,9 < d < 4,9 Å). « nd » indique une distance non disponible.

La discrimination des types I' et III' peut être réalisée à travers la combinaison de plusieurs distances. Le type III' est caractérisé par la présence de deux distances longues  $H_{\beta 2}-H_N$  et  $H_{\beta 3}-H_N$  entre les résidus 2 et 3 (Figure 69). De plus, deux distances courtes à longues  $H_{\beta 3}-H_N$  1-3 et 1-4 sont des informations supplémentaires pour caractériser le type III'. Au contraire, pour le type I' les signatures retrouvées sont une distance longue  $H_{\beta 3}-H_N$  entre les résidus 2 et 3, une distance longue entre le  $H_{\beta 3}$  du résidu 1 et le  $H_N$  du résidu 3 et l'absence de distance détectable  $H_{\beta 3}-H_N$  1-4 (Figure 69).

La distinction entre les types VIa1 et VIa2 est possible en analysant les distances  $H_{\beta}-H_N$  entre le premier et le dernier résidu du coude (Figure 69). La présence de longues distances est caractéristique du type VIa1 alors qu'elles ont une valeur supérieure à 5 Å dans le type VIa2 (Tableau 26) et ne sont donc pas détectables (Figure 69).

La discrimination entre les types I, III, IV<sub>3</sub> et IV<sub>4</sub> peut être faite grâce aux distances  $H_{\beta}-H_N$  entre les résidus 1 et 3. Ces distances sont détectables (< 5,0 Å) pour les types IV<sub>3</sub> et IV<sub>4</sub> tandis qu'elles ne peuvent pas être observées (> 5,0 Å) dans les types I et III (Figure 69). De plus, pour le type I, la présence d'une distance longue  $H_{\beta 3}-H_N$  entre le troisième et le quatrième résidu ainsi que de deux distances longues  $H_{\beta 2}-H_N$  et  $H_{\beta 3}-H_N$  entre les résidus 1 et 4 permet de le différencier du type III (en gras et souligné dans le Tableau 26 et Figure 69). Pour ce dernier, la distance  $H_{\beta 3}-H_N$  1-4 est en effet trop élevée pour être détectable par RMN. Enfin, l'ambiguïté entre les types IV<sub>3</sub> et IV<sub>4</sub> peut être résolue car le type IV<sub>3</sub> est caractérisé par la présence d'une distance moyenne à longue  $H_{\beta 2}-H_N$  et d'une distance courte à longue  $H_{\beta 3}-H_N$  entre le premier résidu et le dernier résidu alors que pour le type IV<sub>4</sub> seule une distance moyenne à longue  $H_{\beta 2}-H_N$  est observable (Figure 69).

Les distances caractéristiques entre la chaîne latérale et le squelette peptidique pour les coudes  $\beta$  de type II, II', IV<sub>1</sub>, IV<sub>2</sub>, VIb et VIII ont été aussi identifiées. Elles sont mises en évidence dans le Tableau 26 (en rouge et souligné) et représentées dans le diagramme NOE à la Figure 69. Elles permettent de confirmer le type de coude.

A l'aide de ces nouvelles signatures, l'organigramme d'aide à la décision peut être continué comme représenté sur la Figure 70.

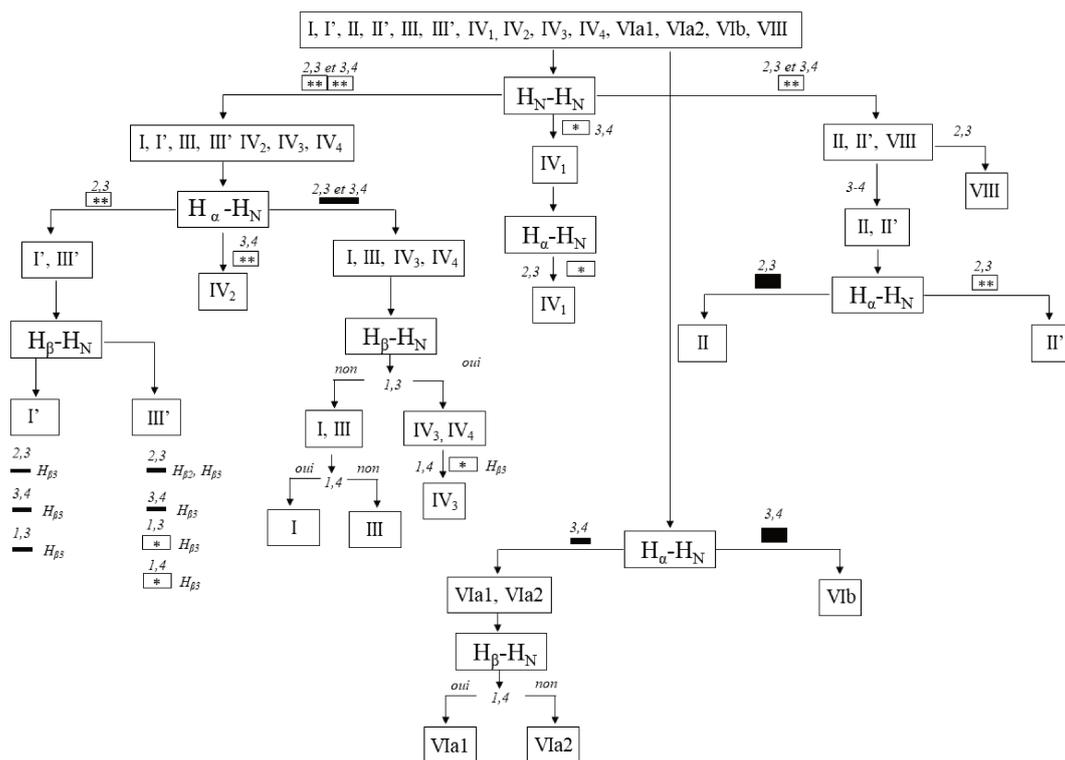


Figure 70 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  lorsque le premier résidu est en hélice  $\alpha$ . Un trait épais blanc avec deux étoiles (\*\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait moyen blanc avec une étoile (\*) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir (■) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ), un trait moyen noir (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) et un trait fin noir (—) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ).

Tableau 26 Tableau des distances inter-hydrogènes (en Å) entre les atomes de la chaîne latérale ( $H_\beta$  et  $H_\delta$ ) et les hydrogènes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\beta$  quand le résidu 1 est en conformation hélice  $\alpha$ . La première ligne correspond à la gamme de distances pour les  $H_{\beta 2}$ ; la deuxième ligne correspond à la gamme de distances pour les  $H_{\beta 3}$ . Pour les coudes présentant une proline en position 3 (VIa1, VIa2 et VIb), pour les distances impliquant une  $H_\delta$ : la première ligne correspond à la gamme de distances pour les  $H_{\delta 2}$ ; la deuxième ligne correspond à la gamme de distances pour les  $H_{\delta 3}$ ; « nd » indique les distances non disponibles. Les distances caractéristiques des coudes II, II', IV<sub>1</sub>, IV<sub>2</sub> et VIII sont en rouge.

Distances	Résidus	Type I	Type I'	Type II	Type II'	Type III	Type III'	Type IV <sub>1</sub>	Type IV <sub>2</sub>	Type IV <sub>3</sub>	Type IV <sub>4</sub>	proline en conformation cis			Type VIII
												Type VIa1	Type VIa2	Type VIb	
d $H_\beta H_N$	1-2	2,6-4,1	2,2-4,1	2,8-4,2	2,6-4,7	3,5-4,0	1,9-4,1	2,7-4,9	2,3-3,7	2,4-3,6	2,3-4,2	2,4-4,3	2,3-3,5	2,3-4,5	2,6-4,2
		2,4-4,1	2,0-3,7	2,2-4,1	2,7-4,5	2,6-4,3	2,0-3,8	2,3-4,6	3,6-4,0	2,1-4,0	3,0-4,4	3,2-4,3	2,4-3,7	2,5-4,4	1,7-3,2
	2-3	2,4-4,5 2,5-4,4	2,9-4,5 <b>4,2-4,5</b>	3,7-4,5 3,4-4,5	1,9-3,6 1,9-3,5	2,8-3,3 3,7-4,1	<b>3,9-4,6</b> <b>3,9-4,6</b>	3,0-4,6 2,9-3,7	3,5-4,5 2,2-4,5	2,4-4,2 4,0-4,5	2,9-4,6 2,9-4,6	n n	n n	n n	2,4-4,0 2,7-4,2
3-4	3,2-4,9 <b>4,3-4,5</b>	3,3-4,1 4,3-4,6	3,1-4,7 3,6-4,3	3,2-4,9 3,0-4,7	2,3-3,7 n	3,5-4,6 3,9-4,6	3,1-4,7 3,9-4,6	<b>4,0-4,7</b> <b>4,3-4,8</b>	2,7-4,1 2,4-4,2	2,8-4,4 2,8-4,5	4,1-4,7 4,2-4,9	3,7-4,8 3,9-4,7	3,5-4,4 3,5-4,3	4,4-5,1 <b>4,1-4,7</b>	
	d $H_\beta H_N$	1-3	5,0-5,9 5,0-5,9	2,8-5,5 <b>4,3-5,3</b>	5,5-7,5 5,8-7,2	5,3-7,3 <b>4,5-7,3</b>	5,4-5,8 5,0-6,0	<b>2,9-5,1</b> <b>2,7-5,9</b>	5,9-8,8 5,8-8,1	4,1-5,8 6,1-6,9	<b>3,3-5,9</b> <b>3,0-6,1</b>	<b>4,0-5,0</b> <b>4,7-5,0</b>	n n	n n	n n
2-4		4,8-5,7 4,1-5,8	3,5-5,0 3,3-4,2	5,1-5,8 4,8-5,4	2,3-5,2 3,4-4,8	4,4-5,8 5,3-6,2	3,0-5,3 2,4-5,3	<b>4,1-5,5</b> 4,8-5,3	4,6-6,5 3,7-6,0	4,8-5,9 5,5-6,5	4,7-6,4 5,0-6,5	4,3-5,6 4,8-5,5	4,8-5,1 4,6-5,1	4,9-6,1 5,4-6,1	5,2-6,2 5,4-6,6
d $H_\beta H_N$	1-4	<b>4,5-5,6</b> <b>4,4-5,7</b>	2,8-6,7 <b>5,9-6,6</b>	<b>4,1-6,6</b> 5,4-6,0	3,2-5,4 2,9-5,8	5,0-6,0 5,5-6,6	<b>3,7-5,3</b> <b>2,4-6,0</b>	4,8-6,6 4,8-6,5	3,4-5,9 5,8-6,2	<b>3,1-5,9</b> <b>2,0-5,0</b>	3,8-6,5 5,1-6,4	<b>3,9-5,0</b> <b>3,9-5,0</b>	5,3-5,6 5,3-5,6	5,2-6,3 4,9-6,1	4,0-6,2 5,5-6,1
		d $H_N H_\delta$	2-3										4,9-5,8 4,8-6,0	5,2-5,7 4,8-5,9	5,0-5,5 4,9-5,7
d $H_\delta H_N$	3-4											3,0-4,4 2,9-4,5	3,7-4,5 3,5-4,7	5,5-5,9 5,5-5,7	
d $H_\alpha H_\delta$	2-3											4,1-4,6 3,9-4,7	4,4-4,7 4,4-4,7	3,7-4,7 4,3-4,7	
d $H_\delta H_\alpha$	3-4											5,1-6,6 <b>4,1-6,7</b>	4,9-6,6 4,8-7,0	5,7-7,5 5,9-7,1	

#### **III.1.1.1.4.2 Influence du premier résidu en conformation brin $\beta$**

Le diagramme NOE pour tous les types de coudes  $\beta$  quand le premier résidu est en conformation brin  $\beta$  est représenté dans la Figure 71. Comme précédemment, nous pouvons observer de nouvelles signatures qui permettent de différencier le type I' du type III', le type VIa1 du type VIa2 ainsi que les types I, III, IV<sub>3</sub> et IV<sub>4</sub>.

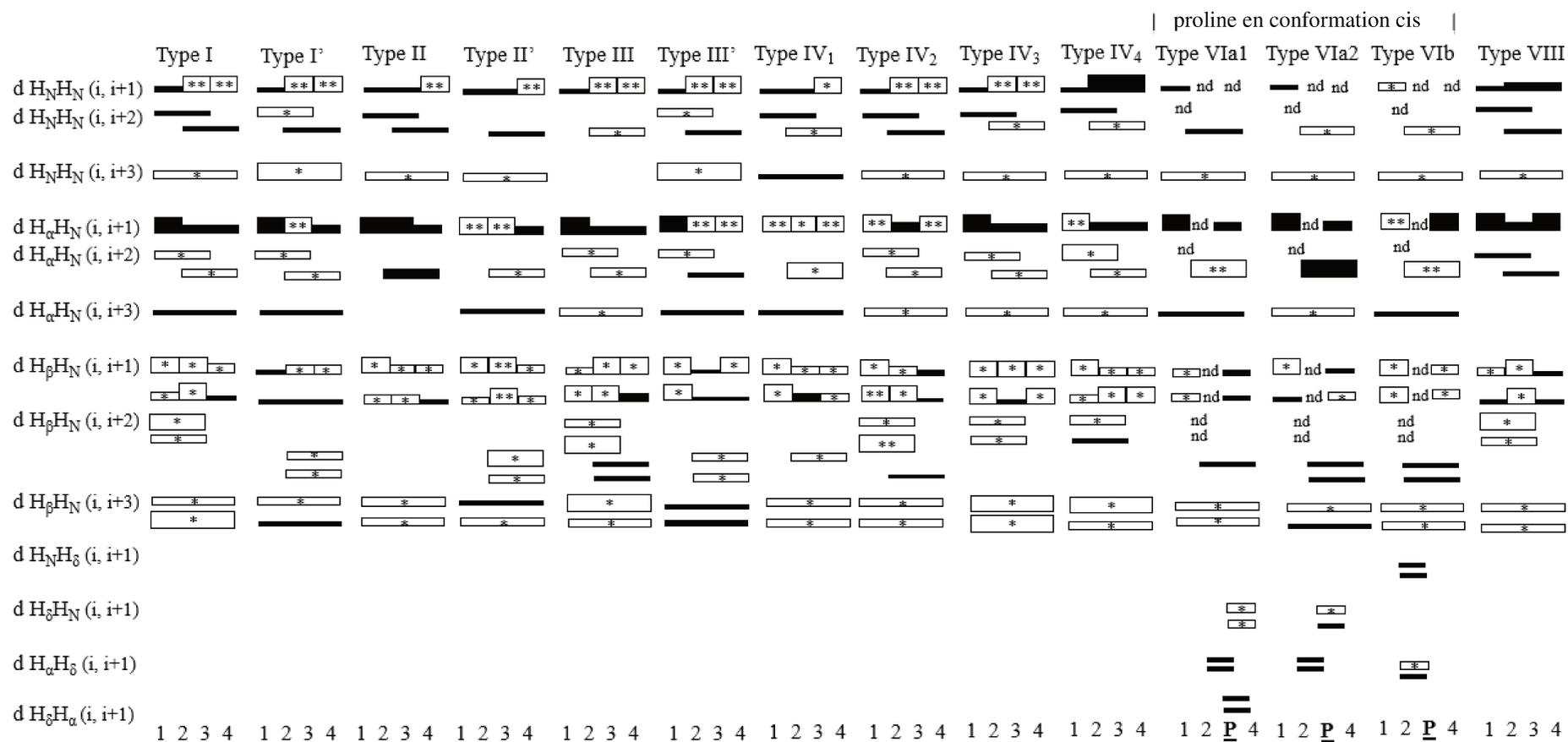


Figure 71 Représentation graphique en diagramme NOE des distances calculées entre les hydrogènes pour tous les types coudes β quand le premier résidu est en conformation brin β. Les coudes VIa1, VIa2 et VIb sont caractérisés par la présence d'une proline en conformation cis en position 3. Un trait épais blanc avec deux étoiles ( **\*\*\*** ) correspond à une distance courte à moyenne (2,0 < d < 3,8 Å), un trait épais blanc avec une étoile ( **\* \* \*** ) correspond à une distance courte à longue (2,0 < d < 4,9 Å), un trait moyen blanc avec une étoile ( **\* \* \*** ) correspond à une distance moyenne à longue (2,9 < d < 4,9 Å), un trait épais noir ( **■** ) correspond à une distance courte (d < 2,8 Å) et un trait moyen noir ( **■** ) indique une distance moyenne (2,9 < d < 3,8 Å), un trait fin noir ( **—** ) correspond à une distance longue (3,9 < d < 4,9 Å). « nd » indique une distance non disponible.

La distinction entre le type I' et III' peut être réalisée à travers l'analyse de plusieurs distances. Le type I' possède deux distances longues  $H_{\beta 2}-H_N$  et  $H_{\beta 3}-H_N$  entre les résidus 1 et 2 ainsi qu'une distance moyenne à longue entre le  $H_{\beta 2}$  du premier résidu et le  $H_N$  du dernier résidu du coude (Figure 71). Pour le type III', les distances  $H_{\beta 2}-H_N$  et  $H_{\beta 3}-H_N$  1-2 peuvent être courtes à longues et deux distances longues  $H_{\beta 2}-H_N$  et  $H_{\beta 3}-H_N$  entre les résidus 1 et 4 peuvent être observées (Figure 71).

La présence de deux distances longues entre le  $H_{\delta 1}$  et  $H_{\delta 2}$  de l'acide aminé 2 et le  $H_{\alpha}$  de l'acide aminé 3 pour le type VIa1 permet de le distinguer du type VIa2 (Figure 71 et Tableau 27).

Enfin, pour différencier les types I, III, IV<sub>3</sub> et IV<sub>4</sub>, les distances longues  $H_{\beta 2}-H_N$  et  $H_{\beta 3}-H_N$  entre les résidus 2 et 4 doivent être analysées (Figure 71 et Tableau 27). En effet, elles sont détectables dans le type III et ne le sont pas dans les autres (Figure 71 et Tableau 27). Le type III a comme signature supplémentaire la présence d'une distance moyenne entre le  $H_{\beta 3}$  du résidu 3 et le  $H_N$  du résidu 4 (Figure 71 et Tableau 27). La différenciation entre les types I, IV<sub>3</sub> et IV<sub>4</sub> peut être réalisée à l'aide de la distance  $H_{\beta 3}-H_N$  longue (3,4) pour le type I, de la distance longue  $H_{\beta 3}-H_N$  (2,3) pour le type IV<sub>3</sub> et de la distance longue  $H_{\beta 3}-H_N$  (1,3) pour le type IV<sub>4</sub> (Figure 71).

Pour les autres types de coudes (II, II', IV<sub>1</sub>, IV<sub>2</sub>, VIb et VIII), les distances caractéristiques entre la chaîne latérale et le squelette peptidique ont été aussi identifiées. Elles sont mises en évidence (en rouge et souligné) dans le Tableau 27 et représenté dans le diagramme NOE à la figure 71. Comme dans le cas où le premier résidu du coude est en conformation hélicoïdale, elles permettent de confirmer le type de coude.

L'organigramme d'aide à la décision pour les coudes  $\beta$  quand le premier résidu est en conformation brin  $\beta$  est présenté à la figure 72.

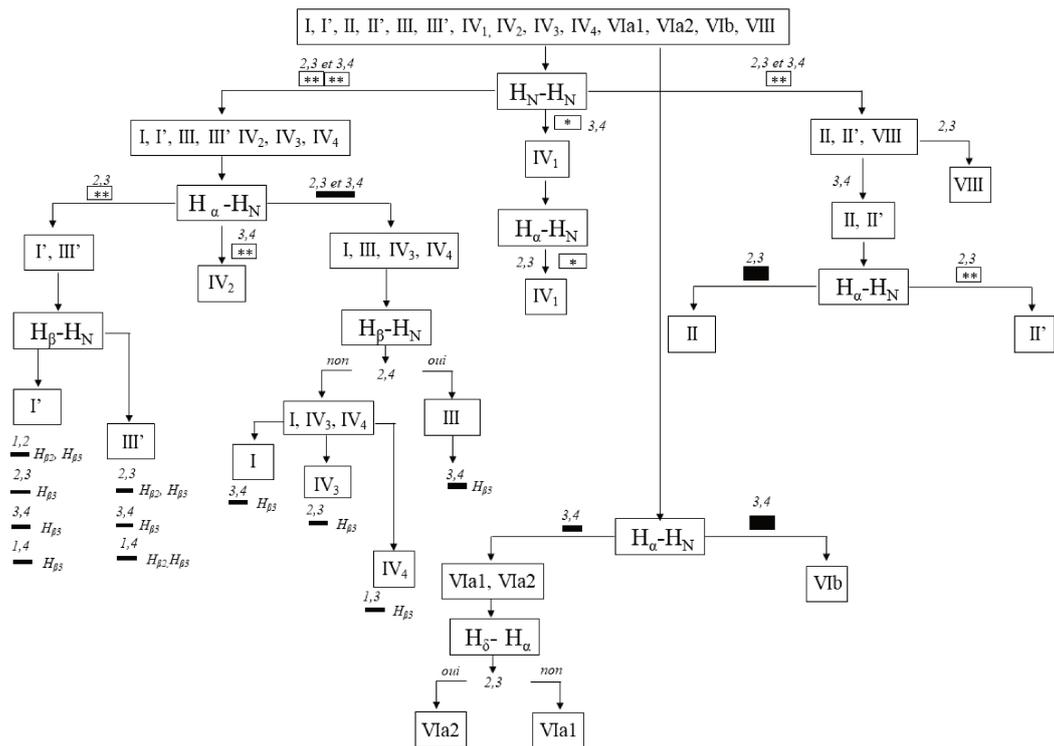


Figure 72 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  lorsque le premier résidu est en brin  $\beta$ . Un trait épais blanc avec deux étoiles (\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait moyen blanc avec une étoile (◻) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir (■) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ) et un trait moyen noir (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) un trait fin noir (—) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ).

Tableau 27 Tableau des distances inter-hydrogènes (en Å) entre les atomes de la chaîne latérale ( $H_\beta$  et  $H_\delta$ ) et les hydrogènes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\beta$  quand le résidu 1 est en conformation brin  $\beta$ . La première ligne correspond à la gamme de distances pour les  $H_{\beta 2}$ ; la deuxième ligne correspond à la gamme de distances pour les  $H_{\beta 3}$ . Pour les coudes présentant une proline en position 3 (VIa1, VIa2 et VIb), pour les distances impliquant une  $H_\delta$ : la première ligne correspond à la gamme de distances pour les  $H_{\delta 2}$ ; la deuxième ligne correspond à la gamme de distances pour les  $H_{\delta 3}$ ; « nd » indique les distances non disponibles. Les distances caractéristiques des coudes I', II, III', IV<sub>1</sub>, IV<sub>2</sub>, VIa1, VIa2 et VIII sont en rouge.

Distances	Résidus	Type I	Type I'	Type II	Type II'	Type III	Type III'	Type IV <sub>1</sub>	Type IV <sub>2</sub>	Type IV <sub>3</sub>	Type IV <sub>4</sub>	proline en conformation cis			
												Type VIa1	Type VIa2	Type VIb	Type VIII
d $H_\beta H_N$	1-2	2,5-4,3	<b><u>4,0-4,5</u></b>	2,4-4,5	2,3-4,6	3,6-4,5	2,6-4,6	2,4-4,2	2,2-4,1	2,7-4,5	2,3-5,2	3,4-4,5	2,4-4,3	1,9-4,6	3,0-4,3
		3,1-4,1	<b><u>4,0-4,6</u></b>	3,6-4,5	2,9-4,7	2,6-4,7	2,8-4,7	2,0-4,4	2,5-3,0	2,6-4,4	3,5-4,7	3,2-4,6	<b>3,9-4,6</b>	2,2-4,4	<b>4,1-4,5</b>
	2-3	2,4-4,2	2,9-4,6	3,8-4,3	2,0-3,7	2,3-4,1	<b>3,9-4,6</b>	3,4-4,4	3,8-4,5	2,4-4,1	3,0-4,5	n	n	n	2,3-3,9
		2,8-4,5	<b>4,3-4,5</b>	3,3-4,4	1,9-3,5	2,4-4,0	<b>4,2-4,6</b>	<b>3,4-3,7</b>	2,5-4,6	<b><u>3,8-4,5</u></b>	2,7-4,6	n	n	n	2,7-4,3
	3-4	3,3-4,7	3,1-4,5	3,1-4,6	3,3-4,8	2,6-4,4	2,8-4,6	3,1-4,7	<b>4,2-4,8</b>	2,4-4,3	3,1-4,4	4,2-5,0	<b>4,3-4,8</b>	3,5-4,4	<b>4,5-4,8</b>
<b><u>3,9-4,7</u></b>		<b>4,3-4,7</b>	<b>4,0-4,4</b>	3,2-4,7	<b><u>3,3-3,6</u></b>	<b>4,0-4,8</b>	3,8-4,6	<b>4,2-4,7</b>	2,7-4,2	2,7-4,4	<b>4,4-4,9</b>	3,5-4,8	3,6-4,4	<b>4,1-4,8</b>	
d $H_\beta H_N$	1-3	2,7-4,4	4,9-6,1	5,2-7,6	4,9-7,5	3,2-5,7	4,9-6,3	5,4-7,3	2,9-4,6	2,9-5,9	2,9-6,3	n	n	n	2,4-4,7
		2,9-4,0	5,4-5,9	5,0-6,9	5,4-7,3	2,8-5,3	5,1-6,0	5,1-6,2	2,6-3,8	2,9-4,8	<b><u>4,4-5,5</u></b>	n	n	n	3,5-4,8
2-4	4,9-5,6	3,2-5,6	5,0-5,8	2,7-5,0	<b><u>4,5-5,9</u></b>	3,0-5,8	2,9-5,6	5,3-6,5	4,8-6,3	4,9-6,5	4,3-5,3	4,6-5,2	4,4-5,9	5,2-6,1	
	4,9-5,9	3,4-5,7	4,9-5,7	3,0-4,8	<b><u>4,6-5,9</u></b>	3,6-6,0	5,2-5,5	4,7-6,0	5,7-6,6	5,1-6,7	4,8-5,3	4,4-4,9	4,2-6,2	5,6-6,6	
d $H_\beta H_N$	1-4	3,1-4,4	<b><u>3,9-6,4</u></b>	3,3-6,1	4,1-6,2	2,0-5,6	<b><u>4,0-6,5</u></b>	3,6-5,7	2,9-5,3	2,5-5,2	2,5-6,9	3,5-5,9	3,8-5,5	3,6-5,5	3,4-4,9
		2,8-4,2	4,5-5,6	3,0-5,4	3,8-6,3	2,9-4,2	<b><u>4,3-6,4</u></b>	3,3-5,0	3,6-4,6	2,1-5,2	3,6-6,1	3,4-5,9	4,9-5,2	3,4-5,6	3,7-5,5
d $H_N H_\delta$	2-3											5,3-5,7	5,1-5,9	4,3-5,5	
													5,2-5,9	4,8-5,4	4,3-6,0
d $H_\delta H_N$	3-4											3,1-4,7	3,7-4,9	5,5-5,8	
													3,1-4,8	4,6-4,7	5,6-5,7
d $H_\alpha H_\delta$	2-3											4,3-4,8	4,0-4,8	3,5-4,7	
													4,1-4,7	4,5-4,6	4,4-4,7
d $H_\delta H_\alpha$	3-4											<b><u>4,0-5,2</u></b>	5,1-6,3	5,9-7,2	
													<b><u>4,0-5,0</u></b>	5,0-7,0	5,9-7,0

### III.1.1.2 La constante de couplage $^3J_{\text{HN-H}\alpha}$

La constante de couplage  $^3J_{\text{HN-H}\alpha}$  est le deuxième paramètre utilisé en RMN pour caractériser la structure secondaire d'une protéine ou d'un peptide. Dans les coudes  $\beta$ , les valeurs de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  intéressantes à analyser sont celles correspondant aux angles dièdres  $\Phi$  des résidus centraux (2 ou  $i+1$  et 3 ou  $i+2$ ), caractéristiques de ce type de coude. Il est à noter que les coudes  $\beta$  qui présentent une proline dans leur séquence (VIa1, VIa2 et VIb) n'auront qu'une seule valeur qui pourra être mesurée. Pour ceux qui ne présentent pas de proline, le calcul des valeurs canoniques de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  pour les résidus centraux  $i+1$  et  $i+2$  a montré que des couples de valeurs de constantes de couplage différentes peuvent être identifiés pour la plupart des sous-types de coudes  $\beta$  (Tableau 28). En revanche, pour les types I', II', III' et IV<sub>3</sub> les couples de valeurs se chevauchent entre eux et elles ne sont donc pas discriminantes (Tableau 28). Pour ceux-ci, les valeurs calculées sont très similaires (Tableau 28). Pour les coudes  $\beta$  de type VI, la valeur J canonique du résidu 2 (ou  $i+1$ ), qui est la seule que nous pouvons mesurer, n'est caractéristique que pour le type VIa1 car pour les types VIa2 et VIb les valeurs sont très similaires (Tableau 28). Ceci pourrait suggérer que les couples de valeurs canoniques de constante de couplage  $^3J_{\text{HN-H}\alpha}$  pourraient être discriminantes pour les types I, II, III, IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>4</sub>, VIa1 et VIII.

Tableau 28 Valeurs moyennes de constante de couplage  $^3J_{\text{HN-H}\alpha}$  pour les résidus centraux des différents types de coudes  $\beta$ . Ces valeurs moyennes ont été calculées à partir de l'équation de Karplus en utilisant les trois séries de coefficients de Pardi, Ludvigsen et Vuister.<sup>133-135</sup> Les valeurs de constante de couplage  $^3J_{\text{HN-H}\alpha}$  qui ne sont pas disponibles ont été indiquées comme « nd ». En gras sont mises en évidence les valeurs qui sont similaires.

Type de coude	$\Phi_{i+1}$	$\Phi_{i+2}$	$^3J_{\text{HN-H}\alpha}$ (i+1) (en Hz)	$^3J_{\text{HN-H}\alpha}$ (i+2) (en Hz)
$\beta$ -I	-60°	-90°	4,0	7,8
$\beta$ -I'	+60°	+90°	<b>6,7</b>	<b>5,2</b>
$\beta$ -II	-60°	+80°	4,0	6,0
$\beta$ -II'	+60°	-80°	<b>6,7</b>	<b>6,6</b>
$\beta$ -III	-60°	-60°	4,0	4,0
$\beta$ -III'	+60°	+60°	<b>6,7</b>	<b>6,7</b>
$\beta$ -IV <sub>1</sub>	-120°	+55°	9,6	6,6
$\beta$ -IV <sub>2</sub>	-85°	-125°	7,2	9,6
$\beta$ -IV <sub>3</sub>	-71°	-72°	<b>5,6</b>	<b>5,6</b>
$\beta$ -IV <sub>4</sub>	-97°	-117°	8,6	9,6
$\beta$ -VIa1	-60°	-90°	4,0	nd
$\beta$ -VIa2	-120°	-60°	9,6	nd
$\beta$ -VIb	-135°	-75°	9,2	nd
$\beta$ -VIII	-60°	-120°	4,0	9,6

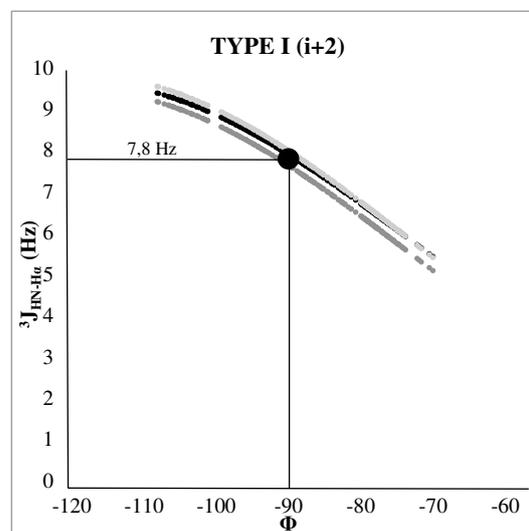
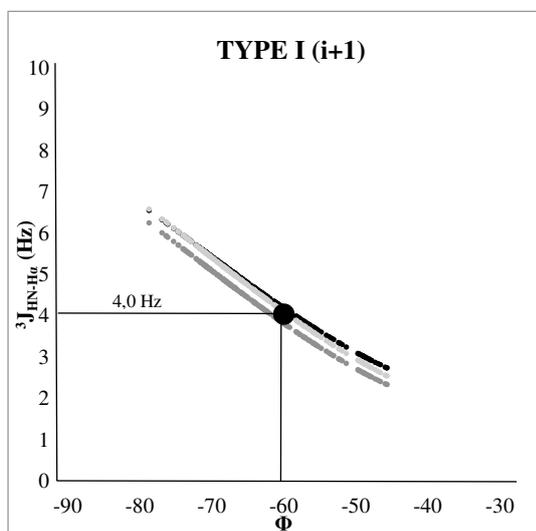
Cependant, les valeurs des angles dièdres  $\Phi$  des résidus centraux observés dans les différents types de coude peuvent s'éloigner des valeurs canoniques. Ainsi, des écarts allant jusqu'à  $30^\circ$  peuvent être relevés dans la littérature.<sup>45</sup> Nous avons donc décidé d'évaluer l'impact de la variabilité géométrique en calculant, pour l'ensemble des angles  $\Phi_{i+1}$  et  $\Phi_{i+2}$  des coudes  $\beta$  de notre base de données, les valeurs  ${}^3J_{\text{HN-H}\alpha}$  correspondantes. Le calcul a été effectué en utilisant les trois séries de coefficients disponibles (Pardi, Ludvigsen et Vuister) dans l'équation de Karplus. Pour chaque type de coude, l'évolution de la constante de couplage en fonction de la valeur de l'angle  $\Phi$  est représentée sur la Figure 73.

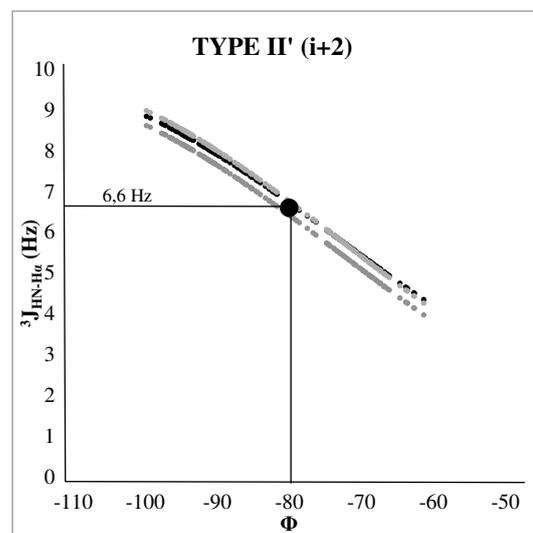
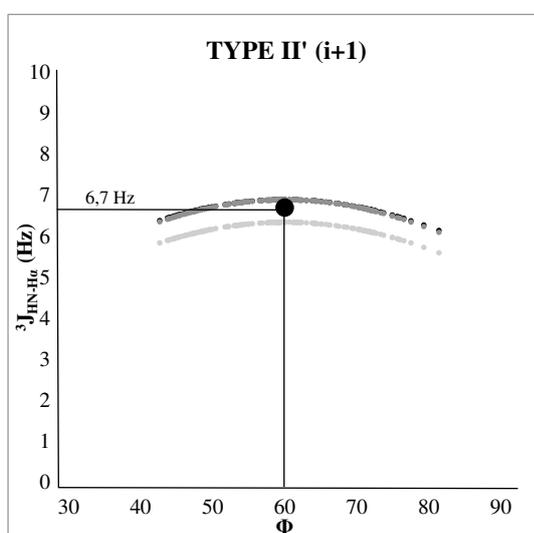
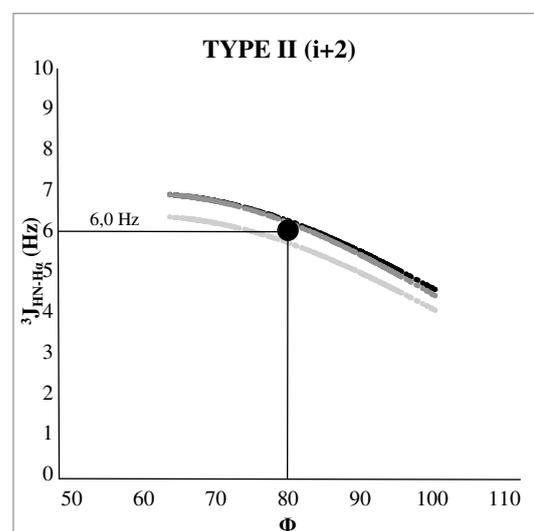
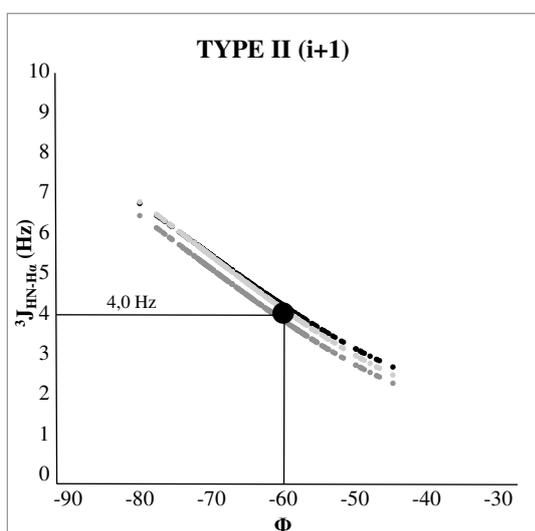
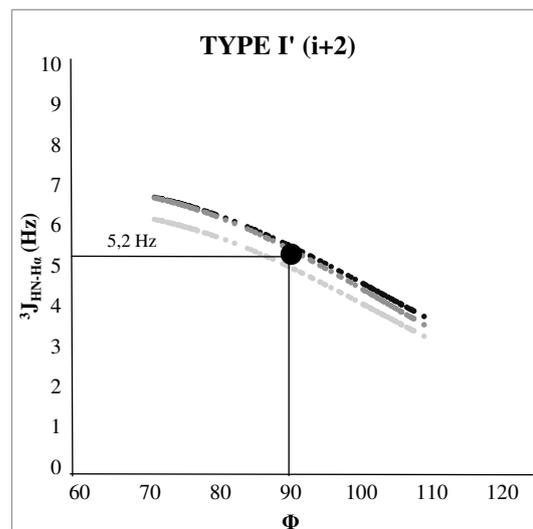
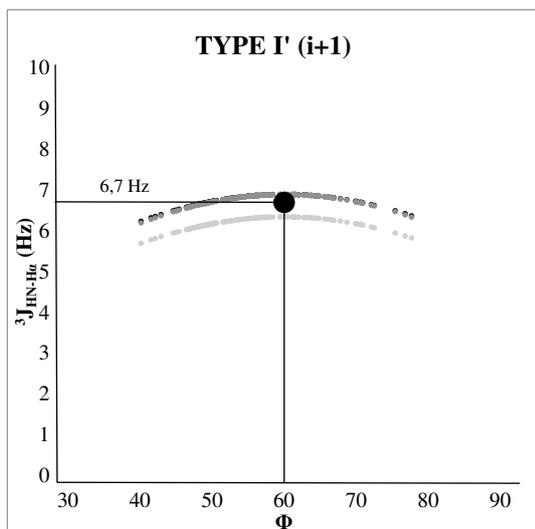
L'observation des graphes obtenus montre que la variation d'angle a un impact différent sur la valeur de la constante de couplage  ${}^3J_{\text{HN-H}\alpha}$  selon la valeur de l'angle et l'allure de la courbe de Karplus. Le relevé des valeurs extrêmes permet de définir un intervalle pour chaque constante de couplage (Tableau 29).

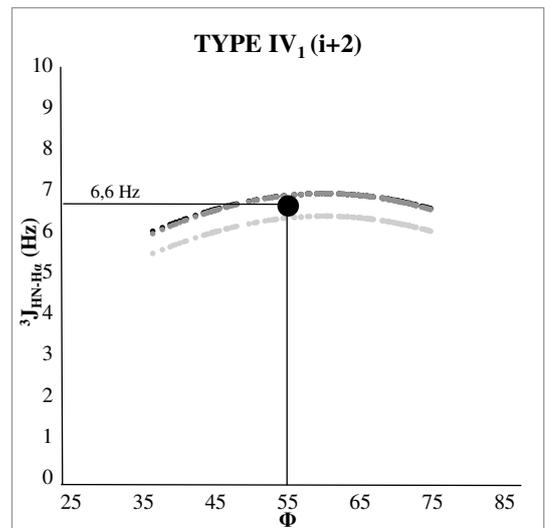
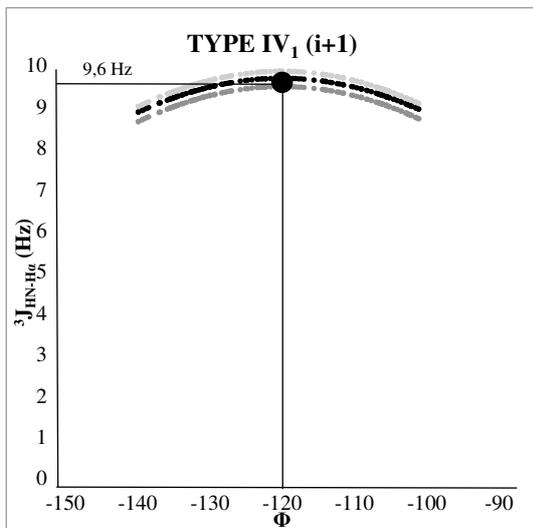
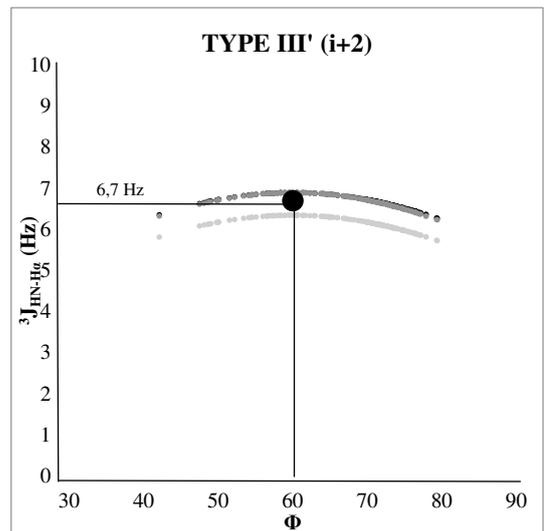
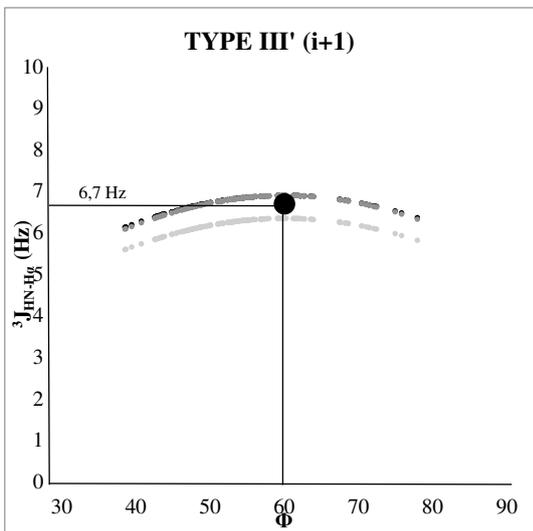
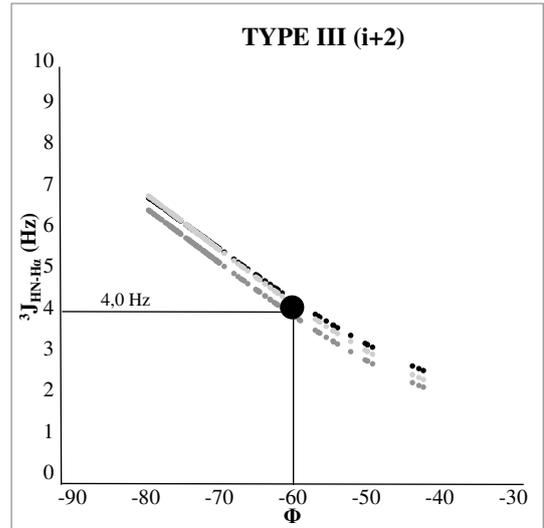
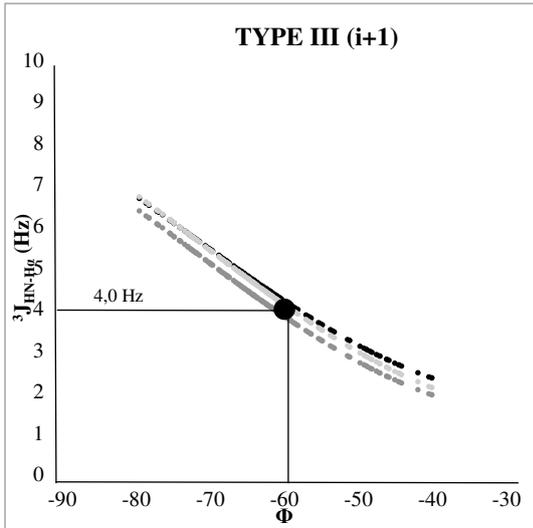
Des gammes de valeurs plutôt étendues sont retrouvées pour le résidu  $i+1$  des coudes de type I, II, IV<sub>2</sub>, VIa1 et VIII et pour le résidu  $i+2$  des coudes  $\beta$  de type I, I', II' (Figure 73 et Tableau 29). Ces gammes de constantes de couplage  ${}^3J_{\text{HN-H}\alpha}$  ne permettent pas de distinguer un coude d'un autre car leur valeur est comprise entre 2,0 et 7,0 Hz pour le résidu  $i+1$  ainsi qu'entre 4,0 et 9,0 Hz pour le résidu  $i+2$  (Figure 73 et Tableau 29). De manière intéressante, les types I', II', IV<sub>4</sub> et VIb pour le résidu  $i+1$  et les types II', IV<sub>1</sub> pour le résidu  $i+2$  sont caractérisés par des gammes de valeurs de J plus restreintes (avec un écart de 3 Hz) comme montré sur la Figure 73. Les coudes  $\beta$  qui présentent des angles  $\Phi$  qui se répètent (types III et III') ou avec des angles  $\Phi$  proches (type IV<sub>3</sub>) pour les deux acides aminés centraux possèdent les mêmes gammes de  ${}^3J_{\text{HN-H}\alpha}$  (Figure 73). En revanche, certains types de coudes  $\beta$  (les types IV<sub>1</sub> et VIa2 pour le résidu  $i+1$  et les types IV<sub>2</sub>, IV<sub>4</sub> et VIII pour le résidu  $i+2$ ) présentent des gammes des valeurs de J étroites et élevées entre 8,5 et 10 Hz qui sont donc des signatures caractéristiques (Figure 73).

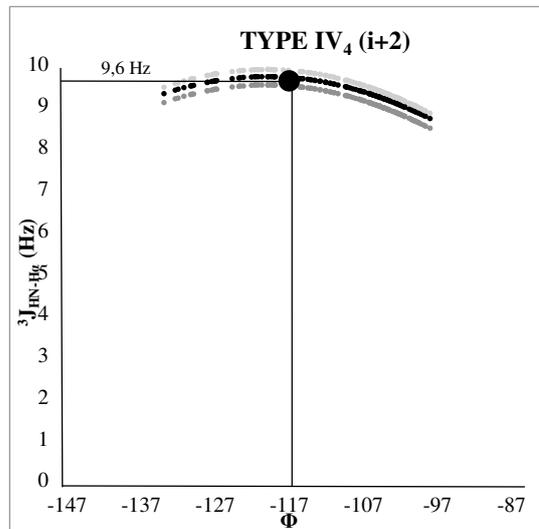
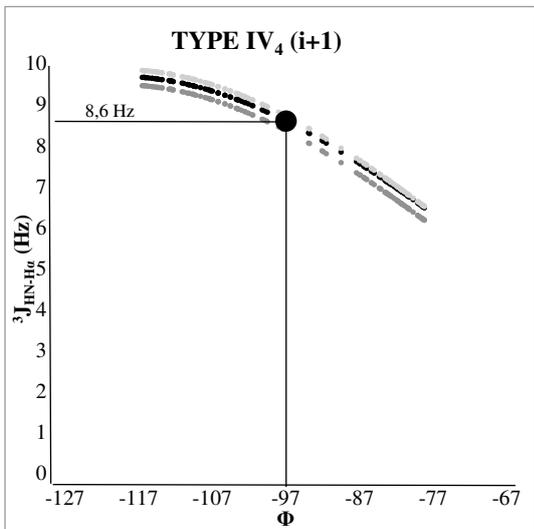
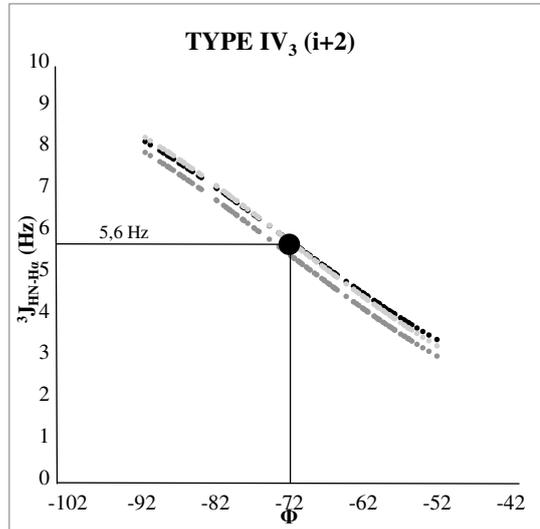
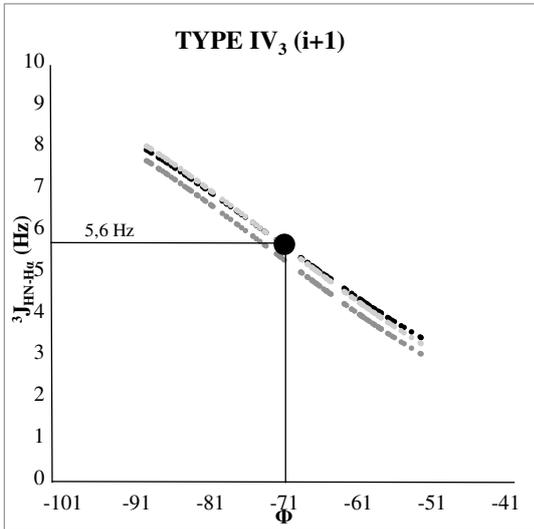
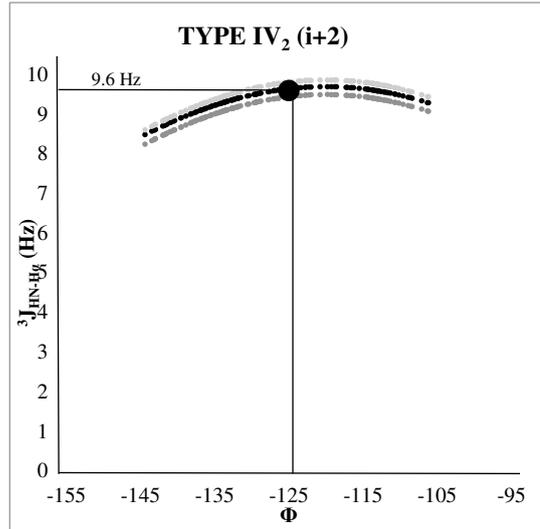
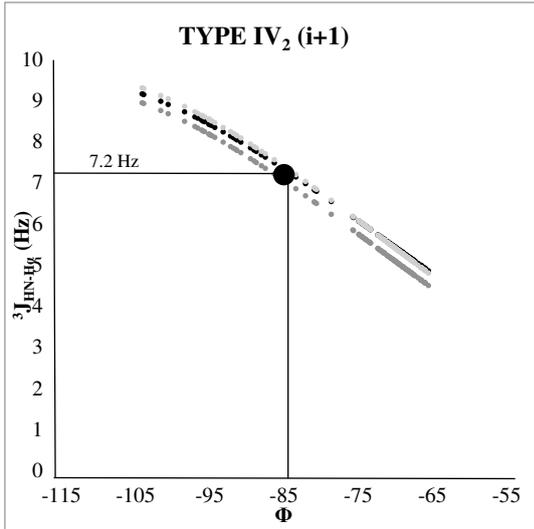
Tableau 29 Gammes de valeurs de la constante de couplage  $^3J_{\text{HN-H}\alpha}$  pour les résidus centraux des différents types de coudes  $\beta$ . Ces intervalles ont été définis à l'aide de la valeur la plus faible et de la valeur la plus élevée identifiées parmi les valeurs calculées sur toutes les structures de peptides générées à partir de l'équation de Karplus en utilisant les trois séries de coefficients de Pardi, Ludvigsen et Vuister.<sup>133-135</sup> Les valeurs de constante de couplage  $^3J_{\text{HN-H}\alpha}$  qui ne sont pas disponibles ont été indiquées comme « nd ». En gras sont mises en évidence les gammes de valeurs étroites.

Type de coude	$\Phi_{i+1}$	$\Phi_{i+2}$	$^3J_{\text{HN-H}\alpha}$ (i+1) (en Hz)	$^3J_{\text{HN-H}\alpha}$ (i+2) (en Hz)
<b><math>\beta</math>-I</b>	-60°	-90°	2,5-7,0	5,5-9,5
<b><math>\beta</math>-I'</b>	+60°	+90°	5,5-7,0	3,5-7,0
<b><math>\beta</math>-II</b>	-60°	+80°	2,5-7,0	4,0-7,0
<b><math>\beta</math>-II'</b>	+60°	-80°	5,5-7,0	4,0-9,0
<b><math>\beta</math>-III</b>	-60°	-60°	2,0-7,0	2,0-7,0
<b><math>\beta</math>-III'</b>	+60°	+60°	5,0-7,0	5,0-7,0
<b><math>\beta</math>-IV<sub>1</sub></b>	-120°	+55°	<b>8,5-9,5</b>	5,5-7,0
<b><math>\beta</math>-IV<sub>2</sub></b>	-85°	-125°	4,0-9,0	<b>8,0-9,5</b>
<b><math>\beta</math>-IV<sub>3</sub></b>	-71°	-72°	3,0-8,0	3,0-8,0
<b><math>\beta</math>-IV<sub>4</sub></b>	-97°	-117°	6,0-9,5	<b>9,0-9,6</b>
<b><math>\beta</math>-VIa1</b>	-60°	-90°	2,0-7,0	nd
<b><math>\beta</math>-VIa2</b>	-120°	-60°	<b>8,5-9,5</b>	nd
<b><math>\beta</math>-VIb</b>	-135°	-75°	7,0-10	nd
<b><math>\beta</math>-VIII</b>	-60°	-120°	2,0-7,0	<b>9,0-10</b>









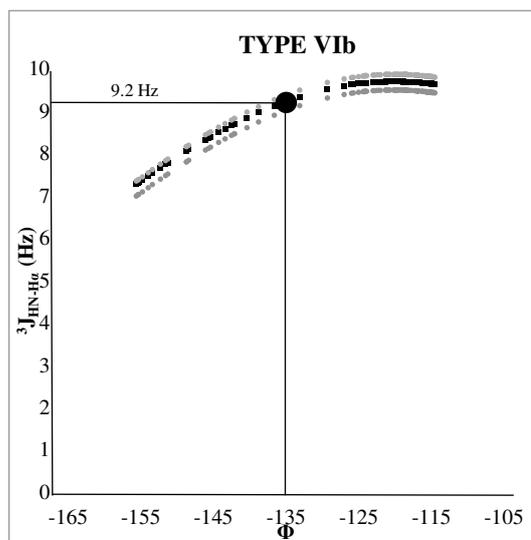
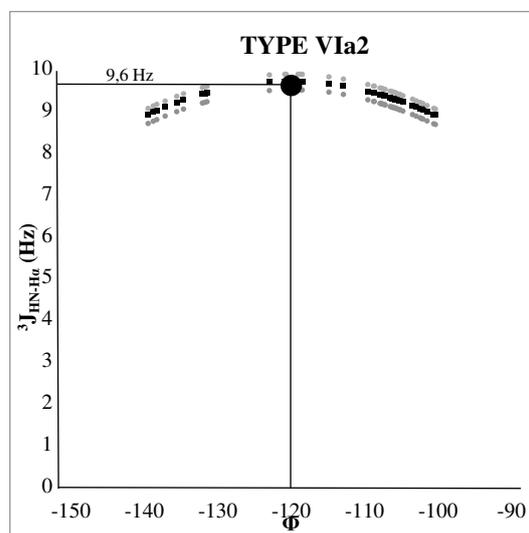
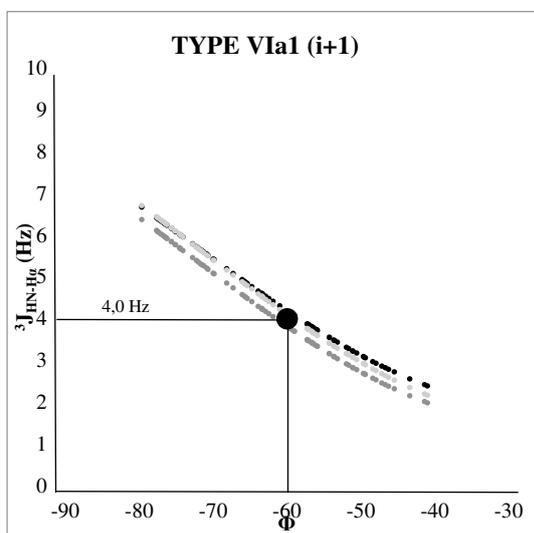
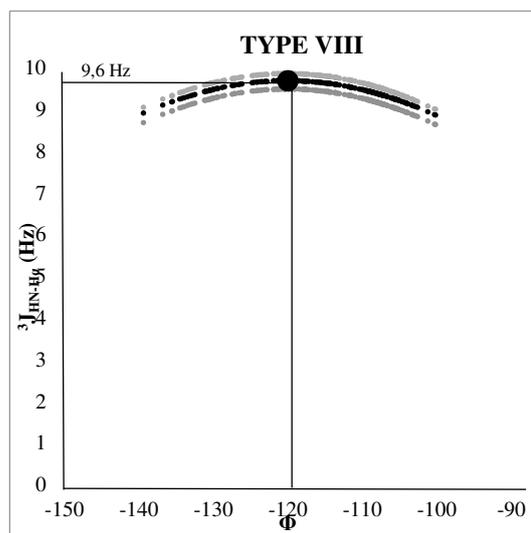
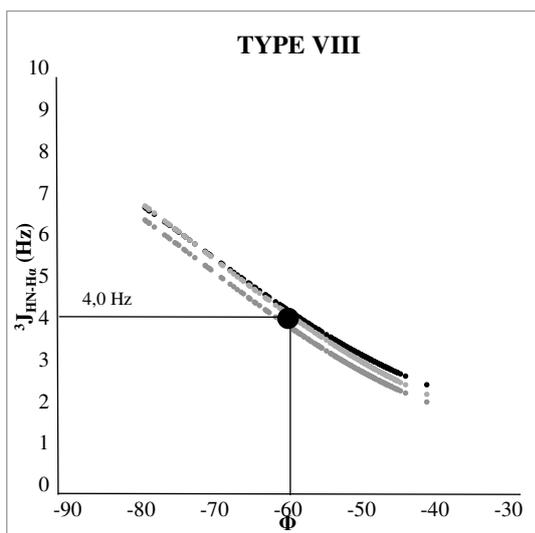


Figure 73 Graphes des gammes de valeurs de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  en fonction des valeurs d'angles  $\Phi$  des peptides en conformation coudée  $\beta$ . Chaque valeur  $^3J_{\text{HN-H}\alpha}$  a été calculée en utilisant les coefficients proposés par Pardi (1984)<sup>133</sup> en noir ; Ludvigsen (1991)<sup>134</sup> en gris et Vuister (1993)<sup>135</sup> en gris clair dans l'équation de Karplus. La valeur moyenne de  $^3J_{\text{HN-H}\alpha}$  pour chaque angle (Tableau 28) est représentée sur le graphe.

Dans la pratique, la valeur de constante de couplage ne pourra pas toujours être utilisée comme paramètre de RMN discriminant. En effet, la variabilité que certains angles dièdres des acides aminés centraux des coudes  $\beta$  peuvent adopter ne permet pas d'avoir des valeurs tranchantes de constante de couplage  $^3J_{\text{HN-H}\alpha}$ . Cependant, la valeur haute (8-10 Hz) de constante de couplage  $^3J_{\text{HN-H}\alpha}$  que nous retrouvons pour le résidu 2 des coudes  $\beta$  de type IV<sub>1</sub> et VIa2 ou pour le résidu 3 des coudes  $\beta$  de type IV<sub>2</sub>, IV<sub>4</sub> et VIII est caractéristique de ces types de coude  $\beta$  (Figure 73). On peut également noter que des faibles valeurs de constantes de couplage (< 3,0 Hz) ne pourront être observées que pour le résidu 2 des coudes de type I, II, III, VIa1 et VIII. Ce paramètre pourra donc être utilisé, dans certains cas (VIa2), pour exclure certains types de coude.

### III.1.1.3 Combinaison distances inter hydrogènes et constante de couplage $^3J_{\text{HN-H}\alpha}$

Afin de finaliser notre analyse, nous avons décidé de combiner les paramètres de distances inter hydrogènes et de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  pour voir s'il est possible de distinguer les différents types de coudes  $\beta$ , en particulier dans le cas où le premier résidu n'adopte ni une conformation hélice ni une conformation brin  $\beta$ . En effet, comme vu au paragraphe III.1.1.1, l'analyse des diagrammes NOE ne permet pas de différencier tous les types de coudes  $\beta$ . Si la distinction entre certains types est possible, elle n'est parfois basée que sur une seule distance. Les constantes de couplage  $^3J_{\text{HN-H}\alpha}$  adoptent, dans les coudes  $\beta$ , des gammes des valeurs plus ou moins larges (cf. III.1.1.2) et ne peuvent pas être utilisées seules pour discriminer les différents types. Certaines valeurs peuvent cependant être caractéristiques de certains types. La combinaison de ces deux paramètres pourrait donc permettre de lever des ambiguïtés sur le type de coude ou de confirmer son identification.

L'analyse du diagramme NOE général (Figure 64) permet d'identifier les coudes de type IV<sub>1</sub>, IV<sub>2</sub>, VIb et VIII. Les valeurs trouvées pour la constante de couplage pourront permettre de confirmer le type de coude. En effet, une valeur élevée ( $7 < J < 10$  Hz) est attendue pour la constante de couplage  $^3J_{\text{HN-H}\alpha}$  de l'acide aminé en position 2 (ou  $i+1$ ) dans les types IV<sub>1</sub>, VIa2 et VIb ou pour l'acide aminé en position 3 (ou  $i+2$ ) dans les types IV<sub>2</sub> et VIII.

Pour tous les autres types, certaines ambiguïtés peuvent persister à l'issue de l'étude du diagramme NOE. Ainsi, il peut être difficile de discriminer les types I' et III', II et II', I, III, IV<sub>3</sub> et IV<sub>4</sub> ainsi que les types VIa1 et VIa2. La valeur de  $^3J_{\text{HN-H}\alpha}$  peut aider à trancher entre les types I' et III'. En effet, si pour le type I' la valeur de constante de couplage attendue pour le résidu  $i+2$  peut être entre 3,5 et 7,0 Hz, dans le type III', pour les deux résidus, les valeurs de

$^3J_{\text{HN-H}\alpha}$  seront forcément entre 5,5 et 7,0 Hz. Une valeur faible ( $< 5,5$  Hz) permettra donc d'identifier le type I'. Pour les types II et II', l'observation d'une faible valeur ( $< 5,0$  Hz) de constante de couplage pour le résidu 2 (ou  $i+1$ ) permettra de confirmer que le coude est de type II. De manière similaire, une valeur élevée ( $> 7,0$  Hz) pour le résidu 3 (ou  $i+2$ ) sera caractéristique d'un coude de type II'.

Pour les coudes de type I, III, IV<sub>3</sub> et IV<sub>4</sub>, la discrimination sera plus difficile. Toutefois, des valeurs faibles de constantes de couplage ( $< 5,0$  Hz) pour les deux résidus centraux ne pourront pas être observées pour les coudes de type III et IV<sub>3</sub>. Des valeurs élevées pour les deux constantes de couplage ( $> 7,0$  Hz) ne pourront être relevées que pour les coudes IV<sub>3</sub> et IV<sub>4</sub>. Ces valeurs pourront être supérieures à 8,0 Hz pour le type IV<sub>4</sub> mais pas pour le type IV<sub>3</sub>. Une valeur faible de constante de couplage  $^3J_{\text{HN-H}\alpha}$  pour le résidu 2 ( $< 5,0$  Hz) et élevée pour le résidu 3 ( $< 8,0$  Hz) sera caractéristique du type I.

Enfin, la constante de couplage permettra également de trancher entre le type VIa1 et le type VIa2 car une forte valeur de constante de couplage (8,5 – 9,5 Hz) est attendue pour le type VIa2 tandis que pour le type VIa1 la valeur sera comprise entre 2,0 et 7,0 Hz.

Sur la Figure 74 est montré l'organigramme d'aide à la décision que nous avons proposé auparavant (Figure 68) en ayant analysé les distances pour les coudes  $\beta$  avec l'ajout des informations que l'analyse des constantes de couplage peut apporter pour lever les ambiguïtés.

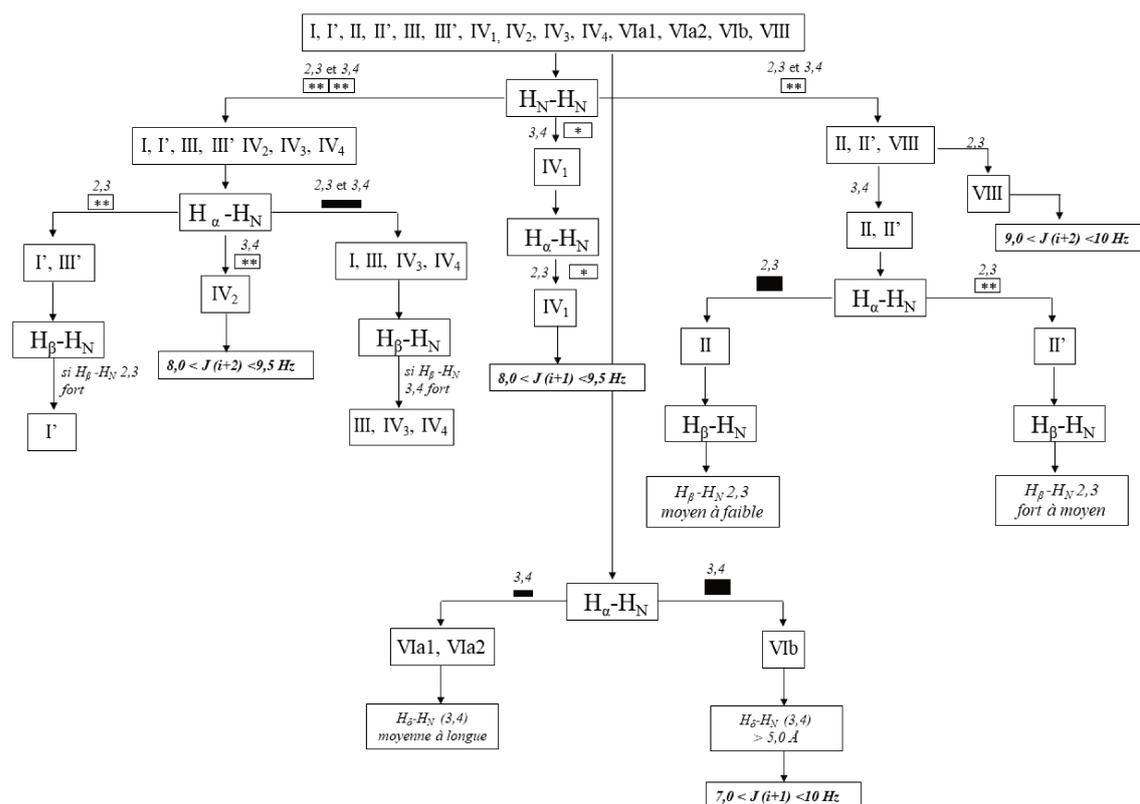


Figure 74 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  en fonction des distances inter-hydrogènes et des constantes de couplage  $^3J_{\text{HN-H}\alpha}$ . Un trait épais blanc avec deux étoiles (\*\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile (\*) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait épais noir (■) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ) et un trait moyen (▬) indique une distance moyenne ( $2,8 < d < 3,8 \text{ \AA}$ )

### III.1.1.4 Comparaison avec les données de la littérature

Comme présenté au sein du chapitre 2, les coudes  $\beta$  de type I, I', II et II' sont les seuls coudes  $\beta$  pour lesquels il existe des paramètres de RMN de référence (diagramme NOE et constantes de couplage  $^3J_{\text{HN-H}\alpha}$ ) dans la littérature.<sup>64,119</sup> La figure 75 donne une représentation des distances inter-protons utilisées dans ces travaux pour caractériser les coudes  $\beta$  de type I et II.

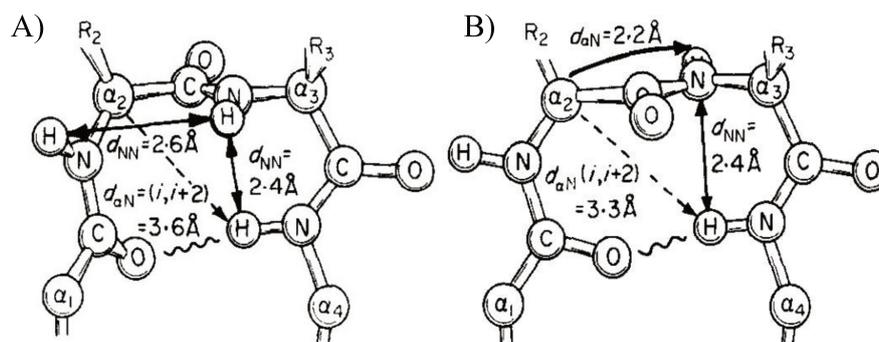


Figure 75 Représentation des coudes  $\beta$  de type I (A) et de type II (B) avec la géométrie canonique. Les lignes ondulées indiquent l'emplacement des liaisons hydrogène. Les flèches indiquent les distances  $^1\text{H-}^1\text{H}$  à courte et à moyenne portée nécessaire pour l'identification des coudes  $\beta$  par RMN (tiré de Wüthrich et al., 1984).<sup>105</sup>

Comme indiqué dans cette figure, tirée du travail de Wüthrich et al. réalisé en 1984,<sup>105</sup> seuls les types I et II sont présentés. Cependant, les distances inter-hydrogènes retrouvées pour le type I sont similaires à celles du type I' ; de même que les distances identifiées pour le type II sont similaires à celles du type II'. Ceci a été mis en évidence par Wagner et ses collaborateurs en 1986.<sup>64</sup>

Les valeurs de constante de couplage de référence ont été aussi déterminées expérimentalement et sont reportées dans le tableau 30.

Tableau 30 Valeurs de la constante de couplage  $^3J_{\text{HN-H}\alpha}$  pour les résidus centraux des coudes  $\beta$  de type I, I', II et II' (tiré de Wagner et al., 1986).<sup>64</sup>

Type de coude	$\Phi_{i+1}$	$\Phi_{i+2}$	$^3J_{\text{HN-H}\alpha} (i+1)$ (Hz)	$^3J_{\text{HN-H}\alpha} (i+2)$ (Hz)
$\beta$ -I	-60°	-90°	4	9
$\beta$ -I'	+60°	+90°	4	5
$\beta$ -II	-60°	+80°	7	5
$\beta$ -II'	+60°	-80°	7	9

Dans ce travail, nous avons retrouvé certaines distances qui avaient déjà été identifiées comme caractéristiques comme, par exemple, la distance  $\text{H}_\alpha$ - $\text{H}_\text{N}$  séquentielle courte entre les résidus 2 et 3 pour le coude de type II a été identifiée dans notre travail (Tableau 31).

Tableau 31 Comparaison des distances inter-hydrogènes relevées dans des peptides modèles de polyalanines en conformation de coudes  $\beta$  de type I et II et dans le présent travail.

Distances	Coude $\beta$ de type I		Coude $\beta$ de type II	
	Wüthrich (1984)	Ce travail	Wüthrich (1984)	Ce travail
dNN (2, 3)	<b>2,6</b>	<b>2,4-3,5</b>	<b>4,5</b>	<b>4,0-4,7</b>
dNN (3, 4)	<b>2,4</b>	<b>2,0-3,3</b>	<b>2,4</b>	<b>2,3-3,5</b>
dNN (2, 4)	3,8	3,5-4,7	<b>4,3</b>	<b>2,9-5,0</b>
d $\alpha$ N (2, 3)	3,4	3,3-3,7	2,2	2,0-2,3
d $\alpha$ N (3, 4)	3,2	3,0-3,5	3,2	3,0-3,5
d $\alpha$ N (2, 4)	3,6	3,2-4,5	3,3	2,7-4,2
d $\alpha$ N (1, 4)	<b><u>3,1-4,2</u></b>	<b><u>3,1-5,4</u></b>	<b><u>3,8-4,7</u></b>	<b><u>3,4-5,6</u></b>
d $\beta$ N (2, 3)	2,9-4,4	2,3-4,5	3,6-4,6	3,5-4,5
	2,9-4,1	2,5-4,5	3,6-4,4	3,5-4,5
d $\beta$ N (3, 4)	3,6-4,6	3,2-4,9	3,6-4,6	2,9-4,9
	3,6-4,4	3,9-4,7	3,6-4,4	2,9-4,7

Cependant, certaines distances ont été corrigées comme la distance  $\text{H}_\text{N}$ - $\text{H}_\text{N}$  entre les résidus 2 et 3 ou la distance  $\text{H}_\alpha$ - $\text{H}_\text{N}$  entre les résidus 1 et 4. Ces corrections sont indiquées en gras dans Tableau 31. Ainsi, nous avons montré que la distance  $\text{H}_\alpha$ - $\text{H}_\text{N}$  (1-4) pouvait être similaire dans

les deux types de coude  $\beta$  et ne constituait donc pas une distance discriminante (souligné dans le Tableau 31).

L'analyse des distances impliquant les hydrogènes de la chaîne latérale nous a par ailleurs permis d'identifier des distances discriminantes supplémentaires entre les  $H_\beta$  du résidu 1 et le  $H_N$  du résidu 3 qui peuvent être courtes dans le type I mais qui ne sont pas détectables pour le type II.

Nous avons également montré que la constante de couplage du résidu 2 ne pouvait pas être utilisée pour différencier le type I du type I' ou le type II du type II'. En effet, la valeur obtenue pour le résidu 2 n'était pas discriminante entre le type I et II ou I' et II'. De plus, l'écart qui peut exister entre les valeurs réelles d'angles dièdres et les valeurs canoniques pour ces types de coude peut modifier fortement la valeur de ces constantes de couplage. Les intervalles obtenus se recoupant (cf. III.1.1.2, Tableau 29), les valeurs mesurées pourront dans certains cas être similaires.

### III.1.2 Cas des coudes $\beta$ avec préférence de proline *trans* en position 2

Comme présenté au sein du chapitre 1, les coudes  $\beta$  de type I, II et III présentent une forte préférence pour l'acide aminé proline en position 2 (ou  $i+1$ ). Cet acide aminé a la particularité de ne pas posséder de proton amide. La présence de cette proline va donc réduire le nombre d'informations disponibles (distances  $H_N-H_N$ ,  $H_\alpha-H_N$  et  $H_\beta-H_N$ ) pour l'identification du type de coude.

En particulier, il ne sera plus possible de distinguer le type II du type I et III à l'aide de la distance courte à moyenne entre les protons amides  $H_N$  du résidu 2 et  $H_N$  du résidu 3. Par ailleurs, la présence d'un cycle dans la chaîne latérale de la proline pourrait avoir pour effet de réduire les intervalles des distances impliquant les protons  $H_\beta$  du résidu 2.

Afin de pouvoir évaluer l'impact de la présence de ce résidu, nous avons inclus, dans notre base de données, des séquences avec proline en position 2 (cf. Tableau 18) pour les coudes  $\beta$  de type I, II et III. Les distances séquentielles et les distances impliquant les hydrogènes de la chaîne latérale ( $H_\beta$  et  $H_\delta$  pour la proline) ont été mesurées et utilisées pour générer de nouveaux diagrammes NOE (Figures 76 et 78).

#### III.1.2.1 Analyse des distances et des diagrammes NOE

Les distances mesurées sont présentées dans les tableaux 32 et 33. L'analyse des distances mesurées entre les hydrogènes du squelette peptidique (Tableau 32 montre que deux distances

peuvent être utilisées pour distinguer le type II des types I et III. Le type II présente en effet une distance courte entre le  $H_\alpha$  du deuxième résidu et le  $H_N$  du troisième résidu alors qu'elle est moyenne pour les types I et III. Ces derniers possèdent, pour leur part, une distance  $H_N-H_N$  1-3 alors qu'elle sera toujours longue pour le coude  $\beta$  de type II. La représentation de ces distances sous forme de diagramme NOE (Figure 76) montre que ces deux distances peuvent être utilisées pour identifier le type II.

Des signatures supplémentaires caractéristiques du type II peuvent être identifiées lors de l'analyse des distances impliquant les hydrogènes de la chaîne latérale en particulier les  $H_\delta$  (Tableau 33 et Figure 76). Ainsi, des distances courtes à moyennes  $H_\delta-H_N$  (2-3) peuvent être observées pour les types I et III alors que dans le type II elles sont supérieures à 5 Å. De plus, une distance moyenne à longue entre le  $H_N$  du premier résidu et le  $H_{\delta 2}$  du deuxième résidu est observée dans le type III alors que pour les types I et II cette distance est forte à longue.

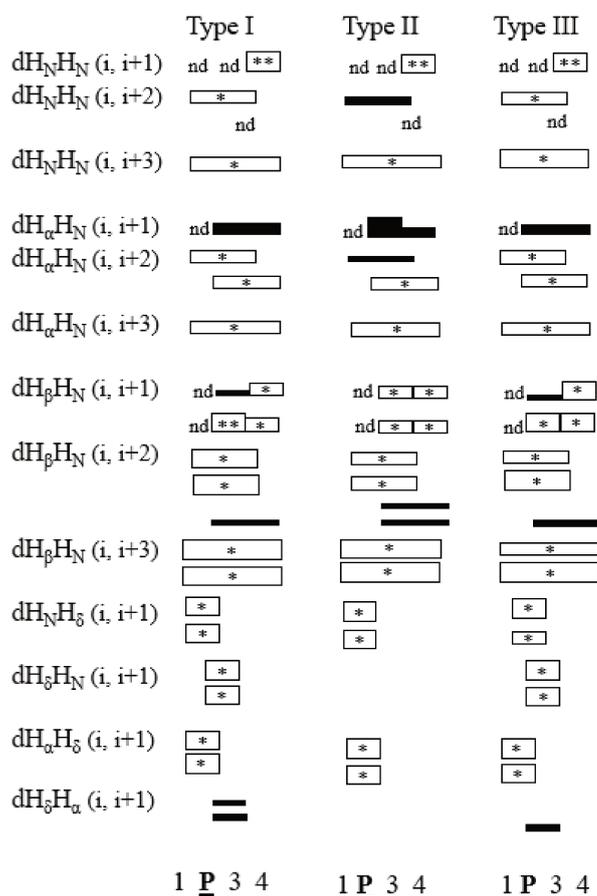


Figure 76 Représentation graphique en diagramme NOE des distances inter-hydrogènes calculées pour les coudes  $\beta$  de type I, II et III avec proline en position 2 (ou  $i+1$ ). Un trait épais blanc avec deux étoiles ( **\*\*** ) correspond à une distance courte à moyenne ( $2,0 < d < 3,8$  Å), un trait épais blanc avec une étoile ( **\*** ) correspond à une distance courte à longue ( $2,0 < d < 4,9$  Å), un trait moyen blanc avec une étoile ( **\*** ) correspond à une distance moyenne à longue ( $2,9 < d < 4,9$  Å), un trait épais noir ( **■** ) correspond à une distance courte ( $d < 2,8$  Å), et un trait moyen ( **■** ) indique une distance moyenne ( $2,9 < d < 3,8$  Å) un trait fin noir ( **■** ) correspond à une distance longue ( $3,9 < d < 4,9$  Å). « nd » indique que la distance n'est pas disponible.

En revanche, la distinction entre les types I et III est plus difficile. Les intervalles de distances entre les deux types sont, en effet, très similaires. Nous pouvons, toutefois, noter qu'une distance moyenne à longue est observée entre le  $H_N$  du premier résidu et le  $H_{\delta 2}$  du second résidu dans le type III alors que cette distance est plus courte dans le type I.

L'ensemble de ces observations permet de proposer l'organigramme de décision suivant (Figure 77) :

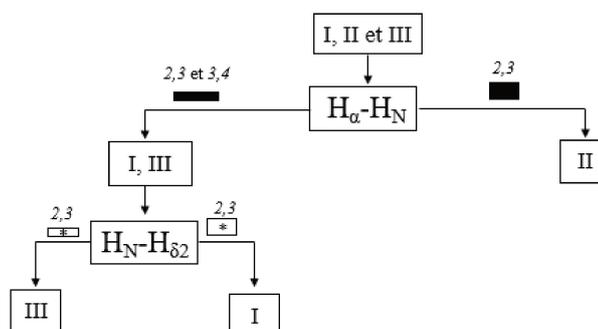


Figure 77 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\beta$  en fonction des distances  $H_\alpha-H_N$  et  $H_N-H_{\delta 2}$  pour les coudes  $\beta$  de type I, II et III comportant une proline en position 2. Un trait épais blanc avec deux étoiles ( **\*\*** ) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile ( **\*** ) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait moyen blanc avec une étoile ( **\*\_** ) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir ( **■** ) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ) et un trait moyen ( **▬** ) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) un trait fin noir ( **—** ) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ).

Tableau 32 Tableau des distances inter-hydrogènes (en  $\text{\AA}$ ) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\beta$  de type I, II et III présentant une proline en conformation *trans* en position 2. En gras la distance caractéristique.

proline en conformation <i>trans</i> en position 2				
Distances	Résidus	Type I	Type II	Type III
$dH_N H_N$	1-2	nd	nd	nd
	2-3	nd	nd	nd
	3-4	2,1-3,1	2,1-3,2	2,4-3,2 <sup>c</sup>
$dH_N H_N$	1-3	3,0-5,9	4,2-7,2	2,8-5,9
	2-4	nd	nd	nd
$dH_N H_N$	1-4	3,0-6,0	3,0-6,3	2,8-5,7
$dH_\alpha H_N$	1-2	nd	nd	nd
	2-3	3,4-3,7	<b>2,1-2,4</b>	3,4-3,6
	3-4	3,1-3,6	2,9-3,6	3,3-3,6
$dH_\alpha H_N$	1-3	3,1-5,1	4,4-6,4	2,9-5,1
	2-4	3,3-4,5	3,0-4,0	3,2-4,6
$dH_\alpha H_N$	1-4	3,2-5,5	3,3-5,6	3,1-5,2

Tableau 33 Tableau des distances inter-hydrogènes (en Å) entre les atomes de la chaîne latérale ( $H_\beta$  et  $H_\delta$ ) et les hydrogènes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\beta$  de type I, II et III présentant une proline en conformation *trans* en position 2. La première ligne correspond à la gamme de distances mesurées pour le  $H_{\beta 2}$ ; la deuxième ligne correspond à la gamme de distances mesurées pour le  $H_{\beta 3}$ .

proline en conformation <i>trans</i> en position 2					proline en conformation <i>trans</i> en position 2				
Distances	Résidus	Type I	Type II	Type III	Distances	Résidus	Type I	Type II	Type III
d $H_\beta H_N$	1-2	nd	nd	nd	d $H_\beta H_N$	1-4	2,5-5,9 2,2-5,4	2,8-6,1 2,6-6,4	2,9-5,9 2,1-5,4
	2-3	4,1-4,6 2,7-3,9	3,0-4,5 3,2-4,4	3,9-4,6 2,5-3,9	d $H_N H_\delta$	1-2	2,4-5,6 2,5-5,3	2,1-5,5 2,5-5,4	2,6-5,5 3,1-5,4
	3-4	3,1-4,6 3,1-4,7	3,1-7,7 3,4-4,4	2,5-4,4 2,7-4,5	d $H_\delta H_N$	2-3	2,4-4,6 2,5-4,5	5,0-5,9 4,8-5,8	2,7-5,6 2,4-4,7
d $H_\beta H_N$	1-3	2,8-5,8 2,6-5,5	3,5-7,0 3,7-7,6	3,7-6,2 2,2-5,4	d $H_\alpha H_\delta$	1-2	2,1-4,8 2,0-4,5	2,1-4,7 2,0-4,9	2,1-4,6 2,0-4,4
	2-4	5,2-6,1 4,7-5,4	4,7-6,0 4,7-5,9	5,2-6,2 4,7-5,4	d $H_\delta H_\alpha$	2-3	4,8-7,0 4,8-6,9	6,5-7,6 6,2-7,7	5,2-7,0 4,6-7,2

## Influence du motif

Afin de voir si la prise en compte de la conformation adoptée par le premier résidu du coude permet une identification plus facile des coudes de type I et III, nous avons réalisé une analyse de l'influence du motif structural sur les distances mesurées. Les distances relevées sont données dans les tableaux 34 et 35. Seules les valeurs correspondant aux motifs HTX et STX seront présentées. En effet, les distances pour la conformation aléatoire sont très proches de celles du cas général (Tableau 32 et 33) présenté au sein de la partie III.1.2.1 de ce chapitre.

Tableau 34 Tableau des distances inter-hydrogènes (en Å) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\beta$  de type I, II et III présentant une proline en conformation *trans* en position 2. La première ligne correspond à la gamme de distances mesurées quand le premier résidu est en conformation hélice ; la deuxième ligne correspond à la gamme de distances mesurées quand le premier résidu est en conformation brin  $\beta$ . « nd » indique les distances non disponibles.

proline en conformation <i>trans</i> en position 2				
Distances	Résidus	Type I	Type II	Type III
$dH_NH_N$	1-2	nd	nd	nd
	2-3	nd	nd	nd
	3-4	2,1-3,1 2,3-2,9	2,4-3,1 2,3-3,2	2,7-3,0 2,4-2,9
$dH_NH_N$	1-3	3,0-5,9 4,7-5,7	5,8-7,2 4,8-6,1	3,5-5,7 4,7-5,5
	2-4	nd	nd	nd
$dH_NH_N$	1-4	3,3-5,9 3,5-4,1	4,8-6,0 3,5-4,4	4,3-6,0 3,5-4,3
$dH_\alpha H_N$	1-2	nd	nd	nd
	2-3	3,4-3,7 3,5-3,6	2,1-2,3 2,1-2,4	3,4-3,6 3,5-3,6
	3-4	3,4-3,5 3,3-3,5	3,0-3,6 2,9-3,6	3,5-3,6 3,4-3,6
$dH_\alpha H_N$	1-3	4,4-5,2 4,0-4,7	5,2-6,1 5,4-6,4	4,3-4,5 3,7-4,9
	2-4	3,4-4,5 3,5-4,1	3,2-3,9 3,0-4,0	3,6-3,9 3,6-4,6
$dH_\alpha H_N$	1-4	3,8-4,4 4,4-5,2	3,7-5,0 4,8-5,5	3,5-5,0 4,4-5,3

Tableau 35 Tableau des distances inter-hydrogènes (en Å) entre les atomes de la chaîne latérale ( $H_\beta$  et  $H_\delta$ ) et les hydrogènes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\beta$  de type I, II et III présentant une proline en conformation *trans* en position 2. Pour chaque distance, la première ligne correspond à la gamme de distances entre le  $H_{\beta 2}$  et le  $H_N$  quand le premier résidu est en conformation hélice ; la deuxième ligne correspond à la gamme de distances entre le  $H_{\beta 2}$  et le  $H_N$  quand le premier résidu est en conformation brin  $\beta$  ; la troisième ligne correspond à la gamme de distances entre le  $H_{\beta 3}$  et le  $H_N$  quand le premier résidu est en conformation hélice et la quatrième ligne correspond à la gamme de distances entre le  $H_{\beta 3}$  et le  $H_N$  quand le premier résidu est en conformation brin  $\beta$ . « nd » indique les distances non-disponibles.

proline en conformation <i>trans</i> en position 2					proline en conformation <i>trans</i> en position 2					
Distances	Résidus	Type I	Type II	Type III	Distances	Résidus	Type I	Type II	Type III	
d $H_\beta H_N$	1-2	n	n	n	d $H_\beta H_N$	1-4	4,5-5,7	3,6-6,1	5,6-5,7	
		n	n	n			2,7-5,0	3,3-4,6	3,3-5,5	
		n	n	n			2,9-5,5	2,8-6,4	4,0-5,4	
		n	n	n			2,7-4,7	3,8-6,1	2,4-4,5	
	2-3	4,1-4,6	3,3-4,2	4,1-4,7		d $H_N H_\delta$	1-2	3,5-4,4	2,1-4,3	3,4-3,8
		4,2-4,6	3,2-4,4	4,1-4,4				4,6-5,5	4,5-5,5	4,2-5,5
		2,8-3,8	3,3-4,2	2,8-3,7				2,8-5,1	2,6-3,4	4,0-5,0
		3,2-3,7	3,4-4,4	2,7-3,2				4,8-5,3	4,7-5,4	4,7-5,5
	3-4	4,0-4,3	3,2-4,7	3,5-3,9		d $H_\delta H_N$	2-3	2,4-3,8	5,6-5,9	2,9-3,8
		3,1-4,4	3,1-4,7	2,5-4,4				2,5-4,6	5,2-5,8	2,7-4,2
3,2-3,7		3,6-4,3	2,8-4,2	2,5-3,9	4,9-5,7			2,4-4,1		
3,1-4,5		3,4-4,4	2,7-4,2	3,1-4,1	5,3-5,7			2,8-4,2		
d $H_\beta H_N$	1-3	3,5-5,7	4,2-6,6	5,6-6,2	d $H_\alpha H_\delta$	1-2	3,5-4,8	2,5-4,2	3,0-4,6	
		3,5-4,5	5,2-6,2	4,4-5,9			2,1-3,5	2,1-3,3	2,1-3,5	
		4,3-5,7	3,9-7,5	5,0-5,2			3,8-4,8	2,4-4,7	4,0-4,1	
		2,9-3,9	5,1-7,5	2,7-4,5			2,0-2,8	2,0-3,2	2,1-3,6	
	2-4	5,2-6,1	5,1-5,7	5,4-5,9	d $H_\delta H_\alpha$	2-3	4,8-6,4	7,1-7,6	5,3-7,0	
		5,4-5,8	4,7-6,0	5,4-6,2			5,2-7,0	6,7-7,6	5,2-6,8	
		4,8-5,4	5,2-5,8	4,9-5,4			4,9-6,6	6,3-7,7	4,6-6,7	
		4,9-5,3	4,9-5,8	4,8-5,4			5,6-6,7	7,2-7,7	5,4-6,8	

Comme observé lors de l'analyse de l'ensemble des coudes  $\beta$ , la prise en compte de la conformation du résidu 1 a par effet de restreindre certains intervalles de distances. L'élaboration des diagrammes NOE (Figure 78) permet de mettre en évidence l'existence de différences supplémentaires entre le type I et le type III.

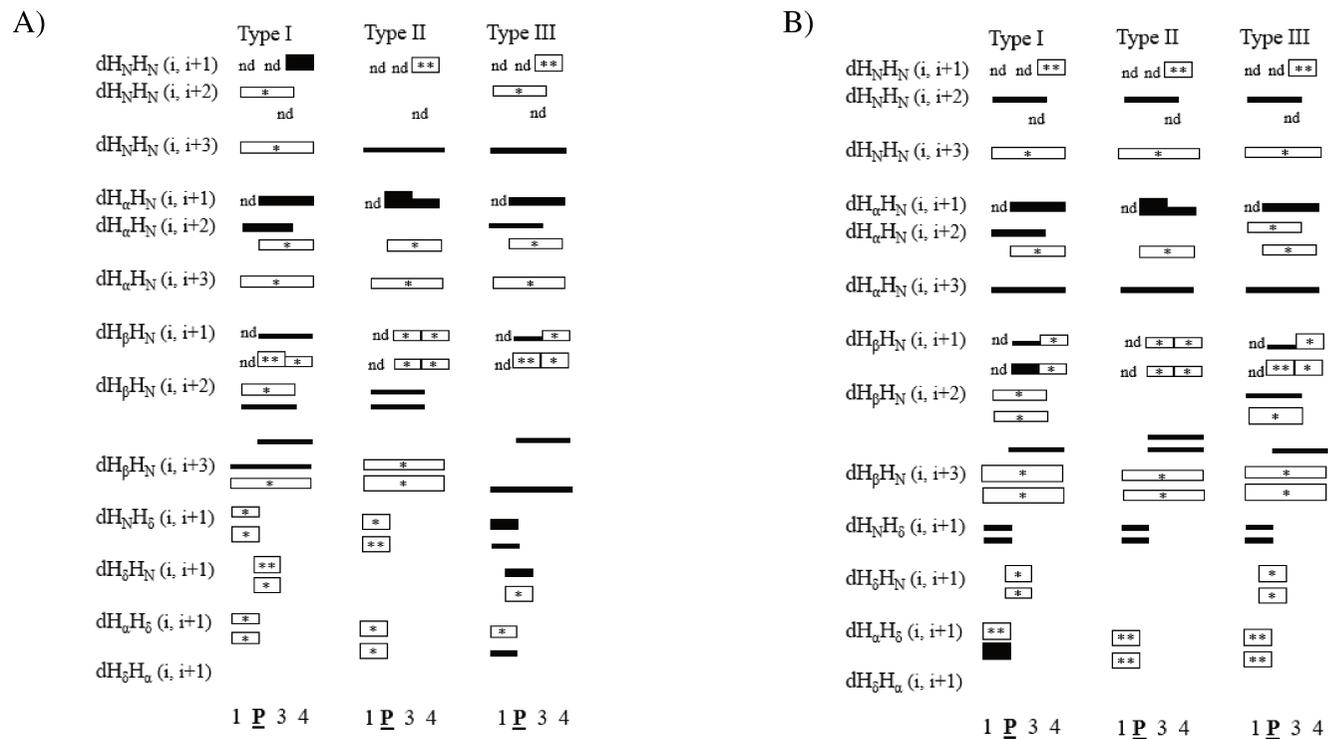


Figure 78 Représentation graphique en diagramme NOE des distances inter-hydrogènes calculées pour les coudes  $\beta$  de type I, II et III avec une proline en position 2. Un trait épais blanc avec deux étoiles ( **\*\*** ) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile ( **\*** ) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait moyen blanc avec une étoile ( **\*** ) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir ( **■** ) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ), un trait moyen ( **▬** ) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) et un trait fin noir ( **—** ) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ). (A) Diagramme NOE quand le premier acide aminé en conformation hélice ; (B) Diagramme NOE quand le premier acide aminé en conformation brin  $\beta$ .

Ainsi, lorsque l'acide aminé *I* adopte une conformation en hélice  $\alpha$ , une distance  $H_N-H_N$  *I-4* moyenne à longue est observée pour le type I alors que cette distance est toujours longue dans le type III. Une autre distance caractéristique du type I est la présence d'une distance  $H_{\beta 2}-H_N$  *I-3* moyenne à longue alors que cette distance est indétectable dans le type III. En outre, une distance  $H_{\beta 3}-H_N$  *I-4* qui peut prendre des valeurs courtes est remarquée dans le type I.

Des différences supplémentaires sont également présentes lorsque le résidu 1 est en conformation brin  $\beta$ . La distance  $H_\alpha-H_N$  *I-3* est toujours longue dans le type III mais elle peut être moyenne dans le type I. Pour ce qui concerne la chaîne latérale, la distance  $H_{\beta 3}-H_N$  *I-3* prend des valeurs moyennes à faibles pour le type III et des valeurs longues pour le type I. Enfin, la distance  $H_\alpha-H_{\delta 2}$  *I-2* prend des valeurs moyennes à faibles dans le type I tandis qu'une distance courte est toujours observée dans le type III.

L'ensemble de ces distances indique que la prise en compte de la conformation adoptée par le premier acide aminé du coude permettra une meilleure distinction entre les types I et III.

### III.1.2.2 La constante de couplage $^3J_{HN-H\alpha}$

Nous avons vu précédemment que les valeurs trouvées pour les constantes de couplage  $^3J_{HN-H\alpha}$  des acides aminés 2 et 3 peuvent permettre d'identifier certains types de coude.

Pour les coudes  $\beta$  de type I, II et III avec une proline en position 2, seule la valeur de constante de couplage du résidu 3 sera mesurable. Les intervalles de valeurs possibles pour cette constante seront identiques à ceux déterminés et présentés auparavant.

Pour le coude de type I, une gamme de valeurs entre 5,0 et 9,5 Hz sera attendue. Cette gamme sera entre 4,5 et 7,0 Hz pour le type II et entre 2,0 et 7,0 Hz pour le type III. Bien que ces intervalles se recoupent, la présence d'une valeur faible (< 5,0 Hz) ou d'une valeur élevée (> 7,0 Hz) pourra être un élément supplémentaire pour différencier les types I et III.

### III.1.2.3 Distinction avec les autres types de coude

La présence d'une proline en position 2 au sein des coudes  $\beta$  de type I, II et III peut également rendre plus difficile leur identification par rapport aux autres types de coudes  $\beta$ . Il faut toutefois souligner que la proline est un acide aminé qui ne peut pas être retrouvé à cette position dans tous les coudes  $\beta$ . Seules les coudes  $\beta$  de type IV<sub>3</sub>, IV<sub>4</sub> et VIII pourraient également présenter une proline en position 2. Pour cette raison, les types I', II', III', IV<sub>1</sub>, IV<sub>2</sub> peuvent déjà être

écartés. L'organigramme de décision que nous avons présenté dans le cas général (Figure 67) peut, de ce fait, être simplifié pour n'envisager que les types I, II, III, IV<sub>3</sub>, IV<sub>4</sub> et VIII.

La réduction des types envisagés permet malgré la suppression de l'effet séquentiel H<sub>N</sub>-H<sub>N</sub> 2-3 d'identifier le type VIII parmi les six types possibles car il présente une distance H<sub>N</sub>-H<sub>N</sub> 3-4 moyenne à longue et une distance H<sub>α</sub>-H<sub>N</sub> 3-4 courte à moyenne. Ces distances sont courtes à moyennes et moyennes respectivement dans les autres types. L'analyse des autres distances H<sub>α</sub>-H<sub>N</sub> permet de distinguer le type II des types I, III, IV<sub>3</sub> et IV<sub>4</sub>. Le type II présente en effet une courte distance entre le H<sub>α</sub> du deuxième résidu et le H<sub>N</sub> du troisième résidu.

L'ensemble de ces observations indique que la présence de la proline ne complexifie pas l'identification des différents types de coude puisque, comme dans le cas général, les types II et VIII peuvent être identifiés facilement alors que pour les types I, III, IV<sub>3</sub> et IV<sub>4</sub>, il faut recourir aux distances impliquant les H<sub>β</sub> ou à l'analyse prenant en compte la conformation du premier résidu.

Nous avons vu précédemment que les valeurs de constante de couplage <sup>3</sup>J<sub>HN-Hα</sub> peut permettre de confirmer le type du coude, voire de l'identifier. Cependant, dans le cas des coudes β présentant une proline en position 2, l'absence de la constante de couplage correspondante rend l'utilisation de ce paramètre RMN plus difficile en raison du recouvrement de gammes observées pour les coudes β I, II, III et IV<sub>3</sub> (Tableau 36). En revanche, les coudes β de type IV<sub>4</sub> et VIII présentent un intervalle de valeurs de constante de couplage restreinte, avec des valeurs élevées caractéristiques (en gras dans le tableau 36) qui permet de les identifier.

Tableau 36 Gammes des valeurs de la constante de couplage <sup>3</sup>J<sub>HN-Hα</sub> pour les résidus 3 (ou i+2) des coudes β de type I, II, III, IV<sub>3</sub>, IV<sub>4</sub> et VIII. Ces intervalles ont été définis à l'aide de la valeur la plus faible et de la valeur la plus élevées identifiées parmi les valeurs calculées sur toutes les structures de peptides générées à partir de l'équation de Karplus en utilisant les trois séries de coefficients de Pardi, Ludvigsen et Vuister.<sup>133-135</sup> En gras sont mises en évidence les gammes de valeurs étroites.

Type de coude	Φ <sub>i+2</sub>	<sup>3</sup> J <sub>HN-Hα</sub> (Hz) (i+2)
<b>β-I</b>	-90°	5,5-9,5
<b>β-II</b>	+80°	4,0-7,0
<b>β-III</b>	-60°	2,0-7,0
<b>β-IV<sub>3</sub></b>	-72°	3,0-8,0
<b>β-IV<sub>4</sub></b>	-117°	<b>9,0-9,6</b>
<b>β-VIII</b>	-120°	<b>9,0-10</b>

## III.2 Les coudes $\gamma$

A l'heure actuelle, dans la littérature, les informations structurales pour discriminer, *via* les paramètres de RMN, les coudes  $\gamma$  de type classique et de type inverse sont inexistantes. Nous avons donc souhaité exploiter les distances inter-hydrogènes et la constante de couplage  ${}^3J_{\text{HN-H}\alpha}$  (du résidu 2 ou  $i+1$ ) pour ces types de coudes afin de pouvoir les distinguer. Nous présenterons, dans un premier temps, les paramètres de RMN de distances entre les hydrogènes et de constantes de couplage des deux coudes de type  $\gamma$ , puis nous aborderons le cas particulier des coudes  $\gamma$  inverse qui ont une forte préférence de la proline en position 2. Pour les distances, une analyse en deux temps sera réalisée : (i) distances entre les hydrogènes du squelette peptidique ( $\text{H}_\alpha$  et  $\text{H}_\text{N}$ ), (ii) distances entre les hydrogènes de la chaîne latérale et ceux du squelette peptidique.

### III.2.1 Le cas général

#### III.2.1.1 Les distances

##### III.2.1.1.1 Distances entre les hydrogènes du squelette peptidique

Pour les coudes  $\gamma$ , trois distances  $\text{H}_\text{N}-\text{H}_\text{N}$  peuvent être mesurées (entre les acides aminés 1-2 ; 2-3 et 1-3). Nous pouvons remarquer que les intervalles obtenus pour ces trois distances sont similaires pour les deux types de coude  $\gamma$  même si la distance 1-2 peut être plus courte dans le type inverse que dans le type classique (Tableau 37).

Une analyse similaire peut être menée pour les trois distances  $\text{H}_\alpha-\text{H}_\text{N}$  (1-2 ; 2-3 et 1-3). La distance 1-2 est courte à moyenne (2,1-3,7 Å) pour les deux types de coude  $\gamma$ . En revanche, la distance 2-3 est courte pour le type inverse (<2,7 Å) et moyenne (3,5-3,8 Å) pour le type classique. Enfin la distance 1-3 est semblable pour les deux types avec des valeurs supérieures à 3,9 Å (Tableau 37).

Tableau 37 Tableau de distances inter-hydrogènes (en Å) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) dans les coudes  $\gamma$  classique et inverse.

Distances	Résidus	Classique	Inverse
$dH_NH_N$	1-2	2,5-4,7	1,7-4,5
	2-3	3,3-4,4	3,4-4,0
$dH_NH_N$	1-3	3,0-6,0	3,3-5,9
$dH_\alpha H_N$	1-2	2,1-3,6	2,3-3,7
	2-3	3,5-3,8	2,3-2,7
$dH_\alpha H_N$	1-3	3,9-5,1	3,9-5,0

### III.2.1.1.2 Le diagramme NOE

Comme présenté au sein de la partie III.1.1.1.2 pour les coudes  $\beta$ , les distances inter-hydrogènes ont été représentées sous forme graphique à travers un diagramme NOE où les distances ont été classées en fonction de leur intensité en trois catégories : forte, moyenne et faible. Pour les coudes  $\gamma$ , la notation des résidus est la suivante : l'acide aminé 1 est le  $i$ , l'acide aminé 2 est le  $i+1$  et le dernier acide aminé est le  $i+2$ .

A partir des tableaux 34 et 35, nous avons établi un premier diagramme NOE général pour les deux types de coude  $\gamma$  (Figure 79).

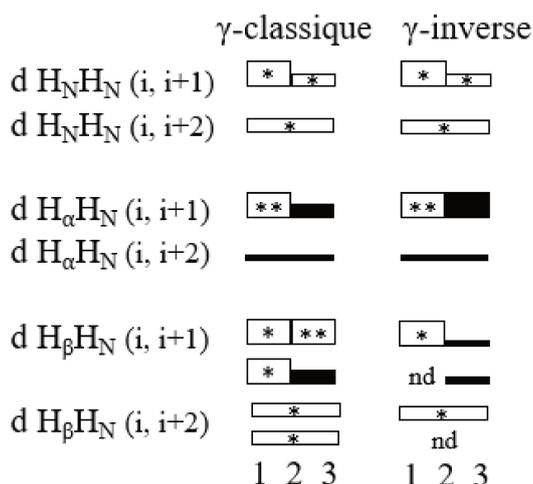


Figure 79 Représentation graphique en diagramme NOE des distances inter-hydrogènes calculées pour les coudes  $\gamma$  de type classique et inverse. Un trait épais blanc avec deux étoiles ( **\*\*** ) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait épais blanc avec une étoile ( **\*** ) correspond à une distance courte à longue ( $2,0 < d < 4,9 \text{ \AA}$ ), un trait moyen blanc avec une étoile ( **\***  ) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir ( **■** ) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ), un trait moyen noir ( **▬** ) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) et un trait fin noir ( **—** ) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ).

Deux distances  $H_N-H_N$  séquentielles (1-2, 2-3) peuvent être observées. Celles-ci ne sont pas caractéristiques selon le type de coude  $\gamma$ . En effet, la première distance (1-2) est courte à longue pour les deux types (classique et inverse) (Figure 79). De manière similaire, les distances  $H_N-H_N$  1-3 ne montrent aucune signature caractéristique. Pour distinguer le type  $\gamma$  classique du type  $\gamma$  inverse, les distances séquentielles  $H_\alpha-H_N$  doivent donc être analysées. Comme pour les coudes  $\beta$ , les distances  $H_\alpha-H_N$  1-2 et 1-3 ne sont pas caractéristiques selon le type de coude  $\gamma$ . En revanche, l'analyse de la même distance entre les acides aminés 2 et 3 montre qu'une courte distance est retrouvée pour le type inverse alors qu'une distance moyenne est observée pour le type classique (Figure 79).

Ces observations nous ont permis de proposer pour les coudes  $\gamma$  un premier diagramme d'aide à la décision qui est présenté en figure 80.

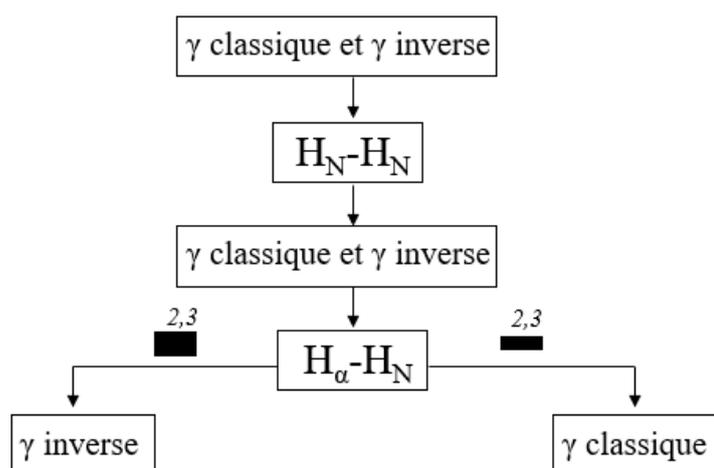


Figure 80 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\gamma$  en fonction des distances  $H_\alpha-H_N$ . Un trait épais noir (■) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ) et un trait moyen (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ).

### III.2.1.1.3 Distances entre les hydrogènes de la chaîne latérale et du squelette peptidique

Nous avons également analysé les distances entre les hydrogènes de la chaîne latérale ( $H_{\beta 2}$  et  $H_{\beta 3}$ ) et les hydrogènes amides du squelette peptidique. Cette analyse permet de mettre en évidence que, pour les distances 1-2 et 1-3, des gammes de valeurs amples peuvent être retrouvées (Tableau 38). Les distances 1-2 sont comprises entre 1,7 et 4,7  $\text{\AA}$  alors que les distances 1-3 sont comprises entre 3,0 et 6,6  $\text{\AA}$  (Tableau 38). En revanche, les gammes de

distances 2-3 sont plus différenciées. Ainsi, les distances 2-3 pour le type classique prennent des valeurs courtes à moyennes (2,1-3,4 Å) pour le  $H_{\beta 2}$  et des valeurs moyennes (3,1-3,8 Å) pour le  $H_{\beta 3}$  alors qu'un intervalle étroit de distances longues (4,1-4,7 Å) est retrouvé pour les deux distances dans le type inverse (Tableau 38).

Tableau 38 Tableau de distances inter-hydrogènes (en Å) entre les atomes de la chaîne latérale ( $H_{\beta}$ ) et du squelette peptidique ( $H_N$  et  $H_{\alpha}$ ) pour les coudes  $\gamma$  classique et inverse. « nd » correspond à une distance non disponible. La première ligne correspond à la valeur du  $H_{\beta 2}$  et la deuxième ligne correspond à la valeur du  $H_{\beta 3}$ .

Distances	Résidu	Classique	Inverse
d $H_{\beta}H_N$	1-2	1,7-4,7	2,2-4,6
		2,3-4,7	nd
	2-3	2,1-3,4	4,3-4,7
		3,1-3,8	4,1-4,5
d $H_{\beta}H_N$	1-3	3,2-6,6	3,5-5,7
		3,0-5,6	nd

Ceci permet de mettre en évidence des signatures caractéristiques supplémentaires pour ces types de coudes  $\gamma$  sur le diagramme NOE (Figure 79). En effet, pour le type classique, une distance courte à moyenne  $H_{\beta 2}$ - $H_N$  (2-3) et une distance moyenne  $H_{\beta 3}$ - $H_N$  (2-3) sont observées alors que deux distances longues 2-3 sont identifiées pour le type inverse (Figure 79). Cela permet de compléter l'organigramme d'aide à la décision selon la Figure 81 :

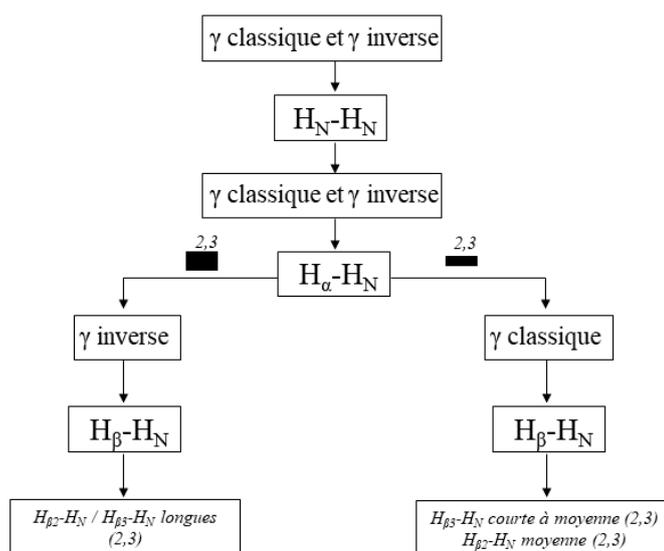


Figure 81 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\gamma$  en fonction des distances  $H_{\alpha}$ - $H_N$  et  $H_{\beta}$ - $H_N$ . Un trait épais noir (■) correspond à une distance courte ( $d < 2,8$  Å) et un trait moyen (▬) indique une distance moyenne ( $2,9 < d < 3,8$  Å).

### III.2.1.2 Influence du motif

Comme pour les coudes  $\beta$ , afin d'apporter des informations supplémentaires, nous avons voulu analyser l'impact de la conformation adoptée par le premier ou le dernier acide aminé sur les distances mesurées dans les coudes  $\gamma$ . L'analyse des neuf motifs structuraux montre que la conformation du premier résidu est la seule qui impacte les distances entre les hydrogènes. Nous retrouvons donc, pour les coudes  $\gamma$ , trois motifs principaux (H-T-X, R-T-X et S-T-X) comme auparavant. Ceci nous a permis d'établir trois nouveaux tableaux (Tableaux 36, 37A et 37B) de distances et deux nouveaux diagrammes NOE (Figure 82 et 84) selon la conformation dans laquelle le premier résidu du coude est engagé. Comme précédemment, nous analyserons les distances séquentielles puis celles impliquant les hydrogènes de la chaîne latérale. Nous proposerons, ensuite, des organigrammes d'aide à la décision par motif à l'aide des diagrammes NOE.

#### Distances séquentielles

Lorsque le premier résidu est en conformation aléatoire, les intervalles de distances relevées sont proches de ceux présentés dans le cas général (Tableau 39). Nous concentrerons donc notre analyse sur les motifs H-T-X et S-T-X. Nous présenterons les différences principales existantes entre les deux types de coude  $\gamma$ . La conformation en hélice ou brin  $\beta$  du premier acide aminé permet en général de restreindre l'intervalle de distances par rapport à la conformation désordonnée.

L'analyse de la distance  $H_N-H_N$  1-2 quand le résidu 1 est en conformation hélicoïdale montre la présence de distances courtes à moyennes (2,5-3,3 Å) pour le type classique et de distances courtes (2,4-2,8 Å) pour le type inverse (Tableau 39). La même distance est moyenne à faible pour les deux types de coude  $\gamma$  lorsque le premier résidu est en conformation brin  $\beta$ .

Des distances moyennes (3,4-3,6 Å) sont retrouvées pour les distances  $H_\alpha-H_N$  1-2 dans le type classique alors qu'une gamme de distances plus étendue (2,7-3,7 Å) est présente pour le type inverse (Tableau 39). Au contraire, quand le premier résidu adopte une conformation en brin  $\beta$ , de courtes distances  $H_\alpha-H_N$  (2,1-2,4 Å) sont présentes pour le type classique tandis que pour le type inverse un intervalle plus large, entre 2,3 et 3,6 Å, est observé (Tableau 39).

Tableau 39 Tableau de distances inter-hydrogènes (en Å) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ) pour les coudes de nature  $\gamma$  de type classique et inverse. La première ligne correspond à la gamme de distances mesurées quand le premier résidu est en conformation hélice ; la deuxième ligne correspond à la gamme de distances mesurées quand le premier résidu est en conformation aléatoire et la troisième ligne correspond à la gamme de distances mesurées quand le premier résidu est en conformation brin  $\beta$ .

Distances	Résidus	Classique	Inverse
$dH_NH_N$	1-2	<u>2,5-3,3</u>	<u>2,4-2,8</u>
		2,4-4,7	1,7-4,2
		<u>3,9-4,3</u>	<u>3,9-4,5</u>
$dH_NH_N$	2-3	3,4-3,6	3,5-3,8
		3,3-4,4	3,4-4,1
		3,4-3,8	3,5-3,8
$dH_NH_N$	1-3	4,9-5,7	5,1-5,8
		3,9-6,0	3,3-5,9
		3,0-4,4	4,0-5,4
$dH_\alpha H_N$	1-2	<u>3,4-3,6</u>	<u>2,7-3,7</u>
		2,1-3,6	2,3-3,6
		<u>2,1-2,4</u>	<u>2,3-3,6</u>
$dH_\alpha H_N$	2-3	3,5-3,6	2,3-2,6
		3,5-3,8	2,6-2,7
		3,5-3,6	2,3-2,6
$dH_\alpha H_N$	1-3	4,1-4,4	4,0-5,0
		3,9-5,1	3,9-4,8
		4,0-4,5	3,9-4,7

### Distances séquentielles entre les hydrogènes de la chaîne latérale et les hydrogènes du squelette

Afin de compléter notre analyse, nous nous sommes aussi intéressés à l'effet que le motif pouvait avoir sur les distances impliquant les hydrogènes de la chaîne latérale. De manière similaire au squelette peptidique, nous avons constaté que c'est la conformation adoptée par le premier résidu du coude qui a un impact sur les distances mesurées. Comme observé pour les coudes  $\beta$ , lorsque le premier acide aminé est engagé dans une conformation aléatoire, les intervalles observés sont larges et correspondent à ceux présentés dans le cas général (partie III.2.1.1.3). Nous présenterons donc seulement les valeurs mesurées lorsque le premier résidu est en conformation hélice (Tableau 40A) ou en brin  $\beta$  (Tableau 40B).

Tableau 40 Tableaux des distances inter-hydrogènes (en Å) entre les atomes de la chaîne latérale (limité aux  $H_{\beta}$ ) et du squelette peptidique ( $H_N$ ) pour les coudes de nature  $\gamma$  de type classique et inverse. (A) Quand le premier résidu est conformation hélice (B) Quand le premier résidu est conformation brin  $\beta$ . La première ligne correspond à la gamme de distances entre le  $H_{\beta 2}$  et le  $H_N$  et la deuxième ligne correspond à la gamme de distances entre le  $H_{\beta 3}$  et le  $H_N$ . « nd » indique une gamme de distances non disponible. Les principales différences sont soulignées et en gras.

A)				B)			
Distances	Résidu	Classique	Inverse	Distances	Résidu	Classique	Inverse
d $H_{\beta}H_N$	1-2	<b><u>3,9-4,2</u></b> 2,9-4,1	<b><u>2,8-3,7</u></b> nd	d $H_{\beta}H_N$	1-2	<b><u>3,0-4,5</u></b> 3,3-4,4	<b><u>2,2-3,7</u></b> nd
	2-3	2,5-2,8 3,7-3,8	4,3-4,7 4,2-4,5		2-3	2,4-2,8 3,7	4,3-4,7 4,3-4,5
d $H_{\beta}H_N$	1-3	3,2-4,9 3,0-4,7	<b><u>4,0-5,7</u></b> nd	d $H_{\beta}H_N$	1-3	5,0-6,0 4,3-5,3	<b><u>3,9-4,8</u></b> nd

Par rapport au cas général, l'analyse par motif fait apparaître de nouvelles signatures (en gras et soulignées dans les Tableaux 37A et 37B) entre les  $H_{\beta}$ , qui permettent de mieux caractériser les coudes  $\gamma$  de type classique et inverse. Des différences existantes entre les hydrogènes  $H_{\beta 2}$  et  $H_{\beta 3}$ , ces derniers seront donc distingués dans la suite de l'analyse. Cette dernière sera réalisée directement sur les diagrammes NOE. Comme précédemment, sur chaque diagramme, la première ligne correspond à l'intervalle  $H_{\beta 2}$ - $H_N$  alors que la seconde ligne correspond à l'intervalle  $H_{\beta 3}$ - $H_N$ .

### Analyse du diagramme NOE

Nous présenterons, dans un premier temps, l'influence de la conformation hélice pour le premier résidu. Dans un deuxième temps, nous décrirons l'influence de la conformation brin  $\beta$ .

### III.2.1.2.1 Influence du premier résidu en conformation hélice $\alpha$

La figure 82 représente le diagramme NOE quand le premier résidu est en conformation hélicoïdale. Sur ce diagramme, nous pouvons retrouver des signatures supplémentaires qui nous permettent de différencier le coude  $\gamma$  classique du coude  $\gamma$  inverse. Ainsi, une distance courte à moyenne  $H_{\beta 2}-H_N$  entre les résidus 1 et 2 est retrouvée pour le type inverse alors que dans le type classique la même distance est moyenne à faible.

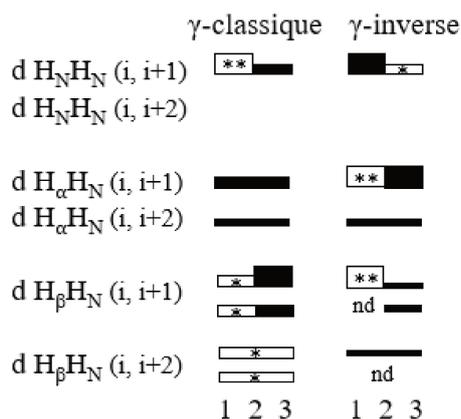


Figure 82 Représentation graphique en diagramme NOE des distances inter-hydrogènes calculées pour les coudes  $\gamma$  de type classique et inverse quand le premier résidu est en conformation hélice  $\alpha$ . Un trait épais blanc avec deux étoiles () correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait moyen blanc avec une étoile () correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir () correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ), un trait moyen noir () indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) et un trait fin noir () correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ). « nd » indique une mesure non disponible.

L'organigramme d'aide à la décision pour différencier les coudes  $\gamma$  de type classique et de type inverse quand le premier résidu est en conformation hélice est présenté sur la figure 83.

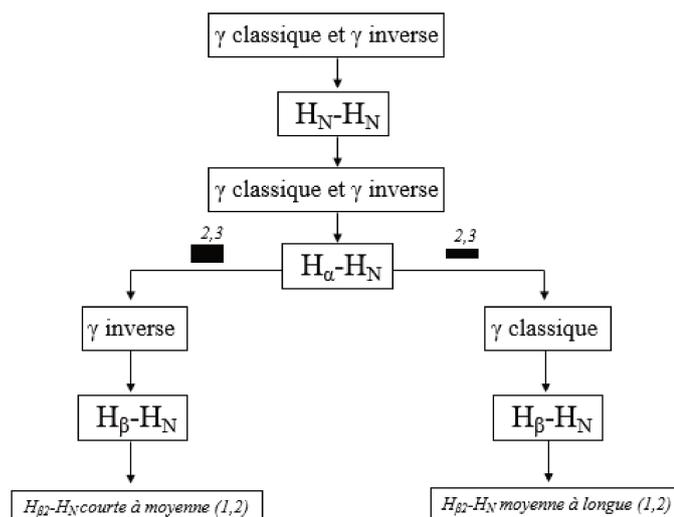


Figure 83 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\gamma$  lorsque le premier résidu est en conformation hélice  $\alpha$ . Un trait épais noir (■) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ) et un trait moyen (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ).

### III.2.1.2.2 Influence du premier résidu en conformation brin $\beta$

Le diagramme NOE pour les coudes  $\gamma$  quand le premier résidu est en conformation brin  $\beta$  est représenté sur la figure 84. Comme précédemment, nous pouvons retrouver des signatures supplémentaires qui nous permettent de différencier le coude  $\gamma$  classique du coude  $\gamma$  inverse. Une distance moyenne à longue  $H_{\beta 3}-H_N$  entre les résidus 2 et 3 est retrouvée pour le type classique alors que dans le type inverse la même distance est toujours longue.

	$\gamma$ -classique	$\gamma$ -inverse
$d H_N H_N (i, i+1)$	▬	▬
$d H_N H_N (i, i+2)$	▬*	▬
$d H_\alpha H_N (i, i+1)$	■	**■
$d H_\alpha H_N (i, i+2)$	▬	▬
$d H_\beta H_N (i, i+1)$	▬* ■ ▬* ▬*	** ▬ nd ▬
$d H_\beta H_N (i, i+2)$	▬	nd nd
	1 2 3	1 2 3

Figure 84 Représentation graphique en diagramme NOE des distances calculées pour les coudes  $\gamma$  de type classique et inverse quand le premier résidu est en conformation brin  $\beta$ . Un trait épais blanc avec deux étoiles (\*\*) correspond à une distance courte à moyenne ( $2,0 < d < 3,8 \text{ \AA}$ ), un trait moyen blanc avec une étoile (\*) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir (■) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ), un trait moyen noir (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) et un trait fin noir (▬) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ). « nd » indique une mesure non disponible.

L'organigramme d'aide à la décision pour différencier les coudes  $\gamma$  de type classique et de type inverse quand le premier résidu est en conformation brin  $\beta$  est présenté sur la figure 85.

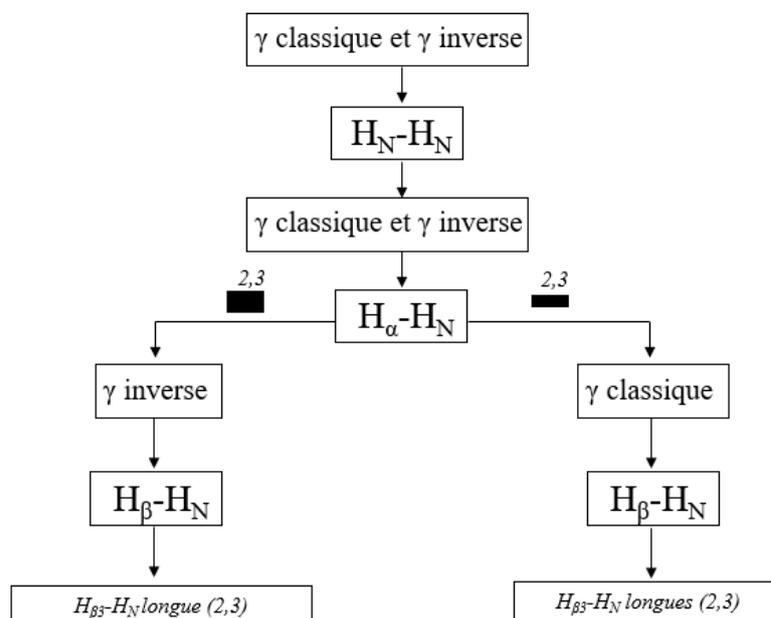


Figure 85 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\gamma$  lorsque le premier résidu est en conformation brin  $\beta$ . Un trait épais noir (■) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ) et un trait moyen (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ).

En conclusion, l'analyse de la conformation (hélice ou brin) dans laquelle le premier acide aminé est engagé fait apparaître des distances supplémentaires discriminatoires entre les coudes  $\gamma$  de type classique et inverse. La prise en compte de la conformation du premier résidu pourra donc permettre une identification plus facile du type de coude.

### III.2.1.2 La constante de couplage

Comme pour les coudes  $\beta$ , nous avons calculé les valeurs de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  du résidu 2 (ou  $i+1$ ) des coudes  $\gamma$  classique et inverse. Pour ces deux types de coude  $\gamma$ , nous ne pouvons malheureusement faire aucune distinction. En effet, les calculs en se basant sur les angles dièdres  $\Phi$  canoniques du résidu  $i+1$  ( $+75^\circ$ , pour le type classique et  $-79^\circ$ , pour le type inverse) donnent des valeurs de constantes de couplage similaires entre elles (6,4 Hz). Nous avons alors également évalué l'impact de la variabilité géométrique au sein des coudes  $\gamma$  sur ces valeurs. L'observation des graphes présentant l'évolution de la constante de couplage en fonction de  $\Phi$  montre un comportement différent pour les deux types de coudes. Nous pouvons voir que si, pour le type classique, une valeur entre 5 et 7 Hz est attendue (Figure 86), pour le type inverse, un intervalle de valeurs de constante de couplage entre 5 et 9 Hz peut être atteinte

(Figure 86). Cela indique qu'une différenciation à l'aide la constante de couplage  $^3J_{HN-H\alpha}$  pourrait être possible. Pour le type inverse, une valeur élevée de constante de couplage ( $> 7,0$  Hz) pourrait confirmer l'identification de ce type de coude.

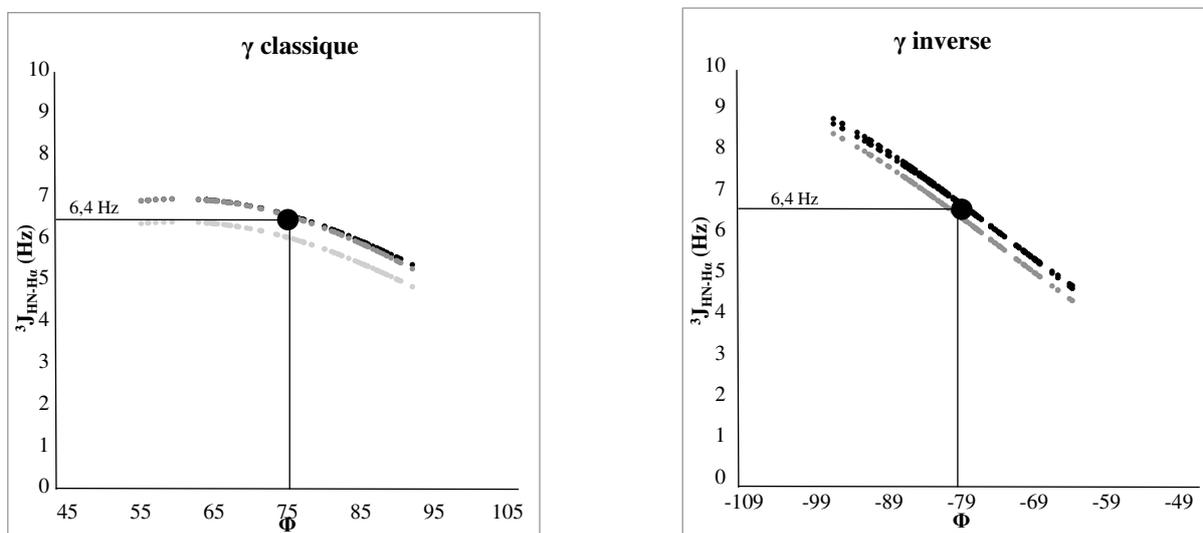


Figure 86 Graphes des gammes de valeurs de constantes de couplage  $^3J_{HN-H\alpha}$  en fonction des valeurs d'angles  $\Phi$  des peptides en conformation coude  $\gamma$ . Chaque valeur  $^3J_{HN-H\alpha}$  a été calculée en utilisant les coefficients proposés par Pardi (1984)<sup>133</sup> en noir ; Ludvigsen (1991)<sup>134</sup> en gris et Vuister (1993)<sup>135</sup> en gris clair dans l'équation de Karplus. La valeur moyenne de  $^3J_{HN-H\alpha}$  pour chaque angle (Tableau 28) est représentée sur le graphe.

### III.2.1.3 Combinaison distances inter protons et constante de couplage $^3J_{HN-H\alpha}$

Comme pour les coudes  $\beta$  nous avons finalisé notre analyse en combinant les deux paramètres RMN pour les coudes de nature  $\gamma$  (Figure 87).

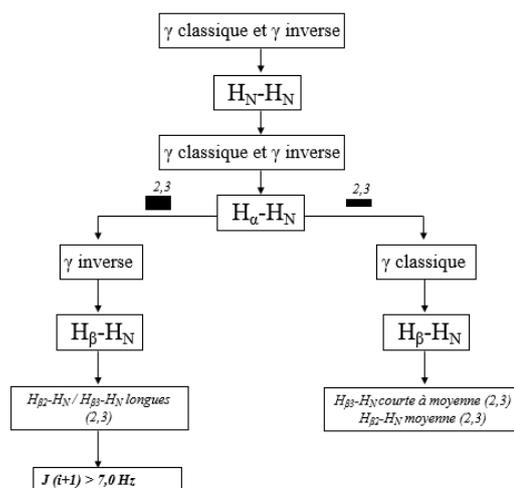


Figure 87 Organigramme d'aide à la décision pour différencier les sous-types de coudes  $\gamma$ . Un trait épais noir (■) correspond à une courte distance ( $d < 2,8 \text{ \AA}$ ) et un trait moyen (▬) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ).

### III.2.1.4 Conclusions

Pour la première fois nous avons établi les diagrammes NOE et les gammes de valeurs de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  pour les coudes  $\gamma$  de type classique et inverse. Ceci a permis de mettre en évidence des paramètres RMN caractéristiques pour ces types de coude.

### III.2.2 Le cas des coudes $\gamma$ inverse avec une proline en position 2

Comme présenté au sein du chapitre 1, le coude  $\gamma$  inverse présente une forte préférence pour la proline en position 2 (ou  $i+1$ ). Ceci est dû à la valeur d'angle  $\Phi$  de ce type de coude qui est  $-79^\circ$ , au contraire du type classique pour lequel l'angle  $\Phi$  de l'acide aminé central a une valeur positive qui est inaccessible pour une proline. Comme pour les coudes  $\beta$ , la présence de cet acide aminé va réduire le nombre d'informations disponibles (distances et constante de couplage  $^3J_{\text{HN-H}\alpha}$ ) pour l'identification de coudes de type  $\gamma$ . Ceci ne devrait cependant pas empêcher la distinction entre les deux types de coudes puisqu'elle est basée principalement sur les distances  $\text{H}_\alpha\text{-H}_\text{N } 2\text{-}3$  et  $\text{H}_\beta\text{-H}_\text{N } 2\text{-}3$  (Figures 88). Afin de vérifier cette absence d'impact, nous avons mesuré les distances pour des coudes  $\gamma$  inverse présentant une proline en position 2 (Tableau 18).

Les distances séquentielles et les distances impliquant les hydrogènes de la chaîne latérale ( $\text{H}_\beta$  et  $\text{H}_\delta$  pour la proline) ont été mesurées. Elles ont été utilisées pour générer un nouveau diagramme NOE pour les coudes  $\gamma$  inverse.

#### III.2.2.1 Analyse des distances et du diagramme NOE

Les distances mesurées sont présentées dans le tableau 41. L'analyse des distances montre qu'il y a peu d'impact de la proline sur les intervalles des distances  $\text{H}_\alpha\text{-H}_\text{N } 2\text{-}3$  et  $\text{H}_\beta\text{-H}_\text{N } 2\text{-}3$ . Le coude  $\gamma$  inverse présente toujours une distance courte  $\text{H}_\alpha\text{-H}_\text{N } 2\text{-}3$  et deux distances longues  $\text{H}_\beta\text{-H}_\text{N } 2\text{-}3$ .

Tableau 41 Tableaux des distances inter-hydrogènes (en Å) (A) entre les atomes du squelette peptidique ( $H_N$  et  $H_\alpha$ ), (B) ainsi qu'entre les hydrogènes de la chaîne latérale ( $H_\beta$  et  $H_\delta$ ) et le squelette peptidique ( $H_N$  et  $H_\alpha$ ) pour le coude  $\gamma$  de type inverse avec une proline en position 2.

A)

Distances	Résidus	Inverse
d $H_N H_N$	1-2	nd
	2-3	nd
d $H_N H_N$	1-3	2.6-5.5
d $H_\alpha H_N$	1-2	nd
	2-3	2.2-2.8
d $H_\alpha H_N$	1-3	3.8-5.1

B)

Distances	Résidus	$\gamma$ inverse
d $H_\beta H_N$	1-2	nd
	2-3	<b>4,1-4,8</b> <b>4,2-4,6</b>
d $H_\beta H_N$	1-3	2.8-5.5 2.3-5.7
		d $H_N H_\delta$
d $H_\delta H_N$	2-3	3.9-4.9 2.2-4.3
		d $H_\alpha H_\delta$
d $H_\delta H_\alpha$	2-3	5.0-6.0 5.0-7.2

La présence de la proline permet par ailleurs d'observer de nouvelles signatures caractéristiques pour le coude  $\gamma$  inverse avec la présence de distances  $H_\delta$ - $H_N$  2-3,  $H_N$ -  $H_\delta$  1-2 et  $H_\alpha$ -  $H_\delta$  1-2 (Figure 88) car dans le type classique une proline ne pourra pas être retrouvée.

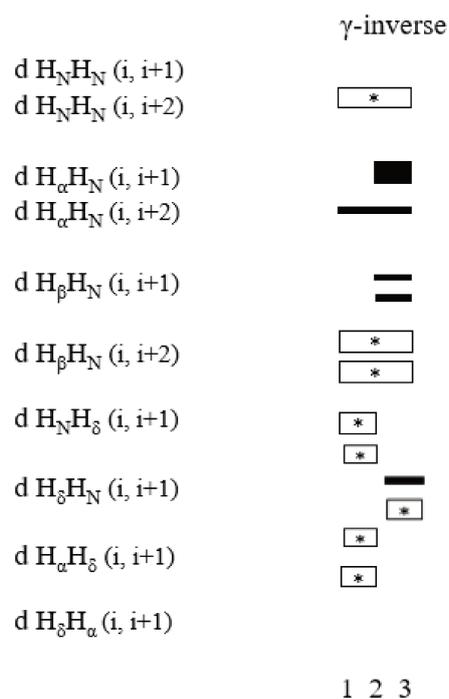


Figure 88 Représentation graphique en diagramme NOE des distances inter-hydrogènes calculées pour le coude  $\gamma$  de type inverse avec une proline en position 2. Un trait moyen blanc avec une étoile ( □\* ) correspond à une distance moyenne à longue ( $2,9 < d < 4,9 \text{ \AA}$ ), un trait épais noir ( ■ ) correspond à une distance courte ( $d < 2,8 \text{ \AA}$ ) et un trait moyen noir ( ▬ ) indique une distance moyenne ( $2,9 < d < 3,8 \text{ \AA}$ ) et un trait fin noir ( — ) correspond à une distance longue ( $3,9 < d < 4,9 \text{ \AA}$ ).

### III.2.2.2 Conclusion

En conclusion, la distinction entre les deux types de coude  $\gamma$  ne sera pas affectée par la présence de la proline. Cette dernière rendra même plus facile l'identification des coudes  $\gamma$  de type inverse car cet acide aminé ne pourra pas être rencontré dans les coudes  $\gamma$  de type classique.

## IV. Conclusions

L'ensemble des analyses de distances inter-hydrogènes menées sur la base de données de peptides représentatifs de coudes  $\beta$  et  $\gamma$  a permis de construire (i) des diagrammes NOE pour la totalité des types de coudes  $\beta$  identifiés (types I, I', II, II', III, III', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb et VIII) et (ii) pour la première fois, des diagrammes NOE pour les coudes  $\gamma$  (types classique et inverse). Ces analyses ont également conduit à corriger les diagrammes de référence déjà existants pour les coudes  $\beta$  de type I, I', II et II'.

L'étude des distances inter-hydrogènes a été complétée par une analyse de la constante de couplage  $^3J_{HN-H\alpha}$  du (ou des) résidu(s) centraux des coudes, calculée à partir de leur angle dièdre  $\Phi$ . Nous avons ainsi mis en évidence une variation importante de la valeur de cette

constante de couplage pour certains des coudes, en raison des variations autorisées pour l'angle  $\Phi$ . Ceci indique que la valeur canonique de la constante de couplage  ${}^3J_{\text{HN-H}\alpha}$  du (ou des) résidu(s) centraux des coudes ne peut pas être utilisée seule, comme paramètre discriminant d'un type de coude.

La combinaison des diagrammes NOE et des constantes de couplage  ${}^3J_{\text{HN-H}\alpha}$  a permis de proposer un organigramme d'aide à la décision pour différencier les types au sein de chaque nature de coude  $\beta$  ou  $\gamma$ . Les coudes  $\gamma$  classique et inverse peuvent toujours être différenciés. Les 14 types de coudes  $\beta$  sont, quant à eux, plus difficiles à discriminer. Ainsi, nous avons montré que, s'il est possible d'identifier les types IV<sub>1</sub>, IV<sub>2</sub>, VIb et VIII, il est plus difficile de distinguer le type I des types III, IV<sub>3</sub> et IV<sub>4</sub>, le type I' du III', le type II du II', ainsi que le type VIa1 du VIa2. La prise en compte de la conformation adoptée par le premier résidu du coude a laissé apparaître des distinctions supplémentaires, en particulier lorsque les distances entre les hydrogènes de la chaîne latérale et du squelette peptidique sont considérées. Ainsi, lorsque le premier résidu est en conformation hélice  $\alpha$  ou brin  $\beta$ , il est possible de distinguer le type I' du type III', le type II du type II', le type VIa1 du VIa2, ainsi que certains types au sein du groupe I, III, IV<sub>3</sub> et IV<sub>4</sub>.

Enfin, nous avons également étudié l'impact sur les organigrammes d'aide à la décision, de la forte préférence de la proline en position 2 pour les coudes  $\beta$  de type I, II et III et pour le coude  $\gamma$  de type inverse. La présence de la proline n'a pas d'influence sur la discrimination des coudes  $\gamma$ . Pour les trois types de coudes  $\beta$  étudiés, nous avons montré que le type II peut être différencié des types I et III. La distinction entre les deux derniers types reste possible lorsque le premier résidu du coude adopte une conformation en hélice  $\alpha$  ou en brin  $\beta$ .



# **Chapitre 5**

## **Recherche de paramètres de DC pour les coudes $\beta$ de type I, I', II et II' par TDDFT**



## ***Chapitre 5 : Recherche de paramètres de DC des coudes $\beta$ de type I, I', II et II' par TDDFT***

### **I. Préambule**

Au cours du précédent chapitre, nous avons présenté l'établissement de paramètres structuraux de RMN pour l'ensemble des coudes  $\beta$  et  $\gamma$ . Nous avons pu remarquer que plusieurs signatures caractéristiques nous permettent de différencier les différents types de coudes  $\beta$  et  $\gamma$ .

Comme évoqué au cours du chapitre 2, la RMN n'est pas la seule technique utilisée en biologie structurale pour caractériser les structures secondaires au sein des peptides et des protéines. En effet, le dichroïsme circulaire électronique (DCE) est une méthode spectroscopique qui est largement utilisée pour l'analyse des structures secondaires des polypeptides en solution.<sup>61</sup> De manière similaire à la RMN, des spectres expérimentaux de DCE de référence pour les hélices  $\alpha$  et les feuillets  $\beta$  existent déjà dans la littérature et sont employés pour comparaison dans les études structurales expérimentales de protéines et de peptides.<sup>150-155</sup> En revanche, des spectres de DCE expérimentaux fiables pour le motif coude  $\beta$  et/ou pour chaque type de coudes de nature  $\beta$  n'existent pas. Deux études théoriques<sup>99,156</sup> et plusieurs études expérimentales<sup>63,101-103,105,106</sup> ont essayé d'établir des spectres de référence pour les coudes  $\beta$ . Cependant, plusieurs allures de spectres DC, différentes, les unes des autres ont été observées (*cf.* partie I.1.2, chapitre 2). En effet, pour les motifs de type coude, plusieurs natures et types existent et, bien que certains types soient négligeables, ils peuvent être retrouvés au sein des structures des peptides ou protéines. De plus, la véritable difficulté pour ces structures secondaires est d'avoir un modèle de coude (quatre acides aminés pour les  $\beta$  et trois acides aminés pour les  $\gamma$ ) stable en solution. Dans ce contexte, nous avons voulu utiliser une approche de chimie théorique, la théorie de la fonctionnelle dépendante du temps (TDDFT), afin de simuler les spectres de DCE. Dans un premier temps, nous nous sommes focalisés sur les coudes  $\beta$  de type I, I', II et II'. Ce choix a été fait car ce sont ceux de la classification actuelle pour lesquels il existe des études théoriques contradictoires.<sup>98,99</sup> Par ailleurs, ce sont aussi les types les mieux caractérisés par d'autres techniques, comme la RMN. En outre, il est intéressant de souligner qu'en absence de spectres de DCE expérimentaux de référence, il est important de disposer au moins de spectres « théoriques ».

Au cours du chapitre 2, nous avons abordé les principaux travaux sur la caractérisation par DCE des coudes. D'un point de vue théorique, le premier travail sur la spectroscopie de DCE sur les coudes de nature  $\beta$  a été réalisé par Woody en 1974 qui identifia *via* des calculs semi-empiriques le rôle fondamental des transitions  $n \rightarrow \pi^*$  et  $\pi_{nb} \rightarrow \pi^*$  pour les coudes  $\beta$  de type I, II et III.<sup>156</sup> L'approche utilisée dans ce travail était de type statique. D'ailleurs, seuls ont été

considérés les changements de  $-10^\circ$ ,  $0^\circ$  et  $+10^\circ$  pour chaque angle caractéristique ( $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$ ,  $\Psi_{i+2}$ ) par rapport aux valeurs canoniques. Ceci a permis d'obtenir un total de  $3^4$  conformations possibles pour chaque type de coude  $\beta$ . Dans ce travail, Woody montre que les coudes  $\beta$  de type I, II et III présentent un seul spectre de DCE caractéristique. Celui-ci a été décrit comme présentant une allure de « classe B » (*cf.* Tableau 12, partie I.2.1, chapitre 2). Ceci ne rend pas possible la distinction entre plusieurs types de coudes  $\beta$  par la spectroscopie de DCE.

Le travail de Woody a été suivi par une autre étude réalisée par Sathyanarayana et Applequist en 1986 qui présente des spectres différents de ceux de Woody.<sup>98</sup> En effet, les auteurs ont montré que les coudes  $\beta$  de type I et II ont des spectres de DCE opposés alors que des allures similaires sont observées pour les coudes de type I et de type III.

Les contradictions entre ces deux études théoriques laissent ouvertes les questions suivantes :

- Le motif structural coude  $\beta$  est-il caractérisé par un seul spectre de DCE de référence ?
- La spectroscopie de DCE peut-elle permettre de distinguer tous les différents types de coudes  $\beta$  ?
- Le signal global des coudes  $\beta$  dans un spectre de DCE d'une protéine qui contient d'autres structures secondaires (hélices et/ou feuillettes) peut-il être distingué ?

Comme décrit en introduction du manuscrit, les coudes  $\beta$  sont des structures secondaires composée par 4 acides aminés. Ce système moléculaire est donc approprié pour effectuer des calculs de type quantique. Pour ces raisons, nous avons décidé de répondre aux questions restées en suspens en simulant les spectres de DC électronique des coudes  $\beta$  de types I, I', II et II'. Un modèle minimal (Ace-A-A-NHMe) (Figure 89) de coude  $\beta$  a été utilisé pour les simulations de dynamique moléculaire et les calculs de DCE dans un environnement aqueux. Ceci a été utilisé afin de focaliser l'attention seulement sur les chromophores structuraux (les trois liaisons peptidiques au sein de cette structure peptidique) qui sont communs pour tous les acides aminés.

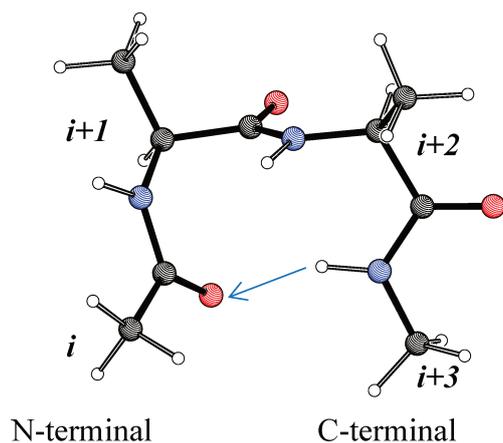


Figure 89 Structure chimique du peptide Ace-A-A-NHMe utilisé dans ce travail comme modèle minimal (H en blanc, C en gris, N en bleu, O en rouge). La liaison hydrogène est représentée avec une flèche bleu clair.

L'approche que nous avons employée pour ces calculs est de type *ab-initio*. Notamment, nous avons utilisé la méthode TDDFT à réponse linéaire couplée avec le schéma multi-échelle à inclusion polarisable (Polarizable-Embedding QM/MM method) et en y associant la dynamique classique. La présentation de ces méthodes théoriques constituera la deuxième partie de ce chapitre. Ensuite, le chapitre se poursuivra avec un article, sous forme d'épreuve, accepté au journal *Physical Chemistry Chemical Physics*. Enfin, le chapitre s'achèvera avec la discussion des résultats et les conclusions.

Cette partie de l'étude a été réalisée en collaboration avec le Dr. Andrea Bonvicini, le Dr. Vincent Tognetti et le Pr. Laurent Joubert qui font partie du groupe de chimie théorique du laboratoire COBRA.

## II. Théorie de la fonctionnelle de la densité (*Density Functional Theory, DFT*)

Au début du XX<sup>ème</sup> siècle, les physiciens découvrent que les lois de la mécanique classique ne permettent pas de décrire le comportement de petites particules telles que les électrons, les noyaux ou les molécules. Ceux-ci sont en effet régis par les lois de la mécanique quantique qui va permettre de calculer et de prédire les propriétés physiques et chimiques de systèmes atomiques et moléculaires. Ces propriétés trouvent leur origine dans le comportement des électrons présents au sein de tels systèmes et peuvent être évaluées à l'aide de calculs de dynamique moléculaire, des calculs de mécanique statistique et des calculs de structures électroniques. Ces derniers utilisent divers formalismes mathématiques afin de résoudre les équations fondamentales de la mécanique quantique.

L'étude détaillée des propriétés électroniques d'un système moléculaire nécessite la prise en compte des effets de corrélation électronique. En particulier, les méthodes Post Hartree-Fock (HF) permettent d'intégrer ces effets mais sont souvent lourdes et limitantes quant à la taille de systèmes étudiés. C'est pourquoi, au cours de ces trente dernières années, la Théorie de la Fonctionnelle de la Densité ou DFT<sup>157</sup> de l'anglais *Density Functionnal Theory* a été considérablement développée pour l'étude des systèmes chimiques et s'est imposée comme une alternative performante aux méthodes Post HF.

Initialement conçue et appliquée aux problèmes de l'état solide, plusieurs raisons ont contribué à sa popularité quant aux applications chimiques :

- Cette théorie inclut dans son formalisme une grande part de corrélation électronique.
- La méthode peut être appliquée à tout type de système : covalent, ionique ou métallique.
- Les ressources informatiques requises sont moins importantes que pour les calculs de type post-HF, rendant ainsi accessibles les études des systèmes moléculaires de plus grandes tailles.

### II.1. De la fonction d'onde à l'équation de Schrödinger

Inspiré par les travaux de Louis de Broglie mettant en évidence la dualité onde-corpuscule<sup>158</sup>, c'est en 1925 qu'Erwin Schrödinger postula l'équation portant encore aujourd'hui son nom.<sup>159</sup>

L'idée nouvelle de ses recherches résidait dans la construction d'une équation d'onde non relativiste capable de régir le comportement d'une particule placée dans un champ potentiel  $V$  ( $\vec{r}$ ). La particule était alors décrite par une fonction d'onde  $\psi$  ( $\vec{r}$ ,  $t$ ) dont la propagation dans le temps suivait l'équation :

$$i\hbar \frac{\partial}{\partial t} \psi(\vec{r}, t) = -\frac{\hbar^2}{2m} \nabla^2 \psi(\vec{r}, t) + V(\vec{r}, t) \psi(\vec{r}, t) \quad (5.1)$$

avec  $i$  l'unité imaginaire telle que  $i^2 = -1$ ,  $\hbar$  la constante de Planck réduite,  $\Psi(\vec{r}, t)$  la fonction d'onde décrivant l'état de la particule,  $m$  la masse de la particule considérée,  $\vec{r}$  sa position,  $\nabla^2$  le laplacien scalaire et  $t$  le temps.

Lorsque l'on s'intéresse ensuite à une molécule, qui est un système chimique multiélectronique, il est nécessaire de restreindre la résolution de cette équation aux états stationnaires (indépendants du temps) Ces états sont alors solutions de l'équation aux valeurs propres suivantes:

$$\hat{H}\psi_n = E_n\psi_n \quad (5.2)$$

où  $\hat{H}$  est l'opérateur Hamiltonien total du système noyaux-électrons correspondant à l'énergie totale  $E_n$  et  $\psi_n$  la fonction d'onde décrivant l'état  $n$  du système. Cette équation fait directement apparaître la quantification de l'énergie.

Pour la résoudre, plusieurs méthodes existent, se basant la plupart du temps sur l'approximation de Born-Oppenheimer<sup>160</sup>, où le problème moléculaire est séparé en deux problèmes électroniques et nucléaires.

En effet, la masse des électrons étant beaucoup plus faible que celle des noyaux, les électrons se déplacent beaucoup plus vite. On peut alors estimer que le mouvement des électrons peut être découplé de celui des noyaux, et ainsi, que la distribution électronique au sein d'une molécule dépendra principalement de la position des noyaux et non pas de leur vitesse.

Il est alors possible d'exprimer l'équation de Schrödinger non-relativiste et indépendante du temps décrivant les états stationnaires propres au système électronique total, et qui est l'équation fondamentale que nous cherchons à résoudre :

$$\hat{H}^{(el)}\Psi_n^{(el)} = E_n^{(el)}\psi_n^{(el)} \quad (5.3)$$

## II.2 La densité électronique

Un électron ne peut être localisé en tant que particule individuelle, en revanche sa probabilité de présence dans un élément de volume peut être estimée et est liée à la densité électronique ( $\rho$ ). Les électrons doivent donc être considérés dans leur aspect collectif (nuage électronique)

et la densité électronique permet de connaître les régions de l'espace où les électrons séjournent le plus souvent.

La densité électronique  $\rho(\vec{r})$  est une fonction positive dépendant uniquement des 3 coordonnées (x, y et z) de l'espace. Cette quantité s'annule à l'infini et vaut N-nombre total d'électrons lorsqu'elle est intégrée sur tout l'espace.

$$\begin{cases} \rho(r \mapsto \infty) = 0 \\ \int \rho(\vec{r}) d^3r = N \end{cases} \quad (5.4)$$

$\rho(\vec{r})$  représente donc, par définition, de nombre moyen d'électrons dans un volume unitaire  $d^3$  défini par  $r$ . Ainsi, la densité électronique, à la différence de la fonction d'onde, est une observable par diffraction des rayons X. Finalement, on peut noter que  $\rho(\vec{r})$  semble contenir assez d'informations pour décrire le système. On peut aussi remarquer que  $\Psi_n^{(el)}$  dispose de beaucoup plus d'informations mais certaines ne sont pas nécessaires pour la description du système chimique.

L'ensemble de ces arguments semble indiquer que la densité électronique suffit à la détermination complète des propriétés d'un système atomique et c'est pour cette raison que plusieurs tentatives de mise en place d'un formalisme quantique basé sur cette quantité ont été proposées. Mais c'est à Hohenberg et Kohn que nous devons la proposition d'un formalisme exact (exempt de toute approximation) énoncé sous la forme de deux théorèmes.

Le premier théorème de Hohenberg et Kohn indique ainsi que l'énergie électronique totale du système à l'état fondamental  $E_0^{(el)}$  dépend uniquement de la densité électronique, ce qui signifie que  $E_0^{(el)}$  peut s'écrire sous la forme d'une fonctionnelle de la densité électronique. Le deuxième théorème de Hohenberg et Kohn prouve qu'en minimisant l'énergie du système par rapport à la densité électronique, on peut obtenir l'énergie de l'état fondamental électronique.

$$E_0 = E_0[\rho] = F_{HK}[\rho] + \int_0 \rho(\vec{r}) v(\vec{r}) d^3r \quad (5.5)$$

où  $F_{HK}[\rho]$  est la fonctionnelle universelle de la densité,  $\rho(\vec{r})$  représente le nombre moyen d'électrons et  $v$  est un potentiel externe dans un volume unitaire  $d^3r$ .

Ainsi et *via* ces deux théorèmes, Hohenberg et Kohn montrent qu'il existe une fonctionnelle universelle reliant l'énergie totale à la densité électronique  $E_0 = F_{HK}[\rho]$ , mais son expression reste encore inconnue à ce jour de façon analytique.

En résumé, la DFT peut être décrite comme une méthode de calcul quantique permettant l'étude de la structure électronique, en principe de manière exacte, en se basant uniquement sur le calcul de l'énergie à partir de la densité électronique, et ce sans avoir *a priori* à déterminer la fonction d'onde exacte du système réel.

### II.3 Approche de Kohn-Sham

L'approche ultérieure de Kohn-Sham est la suite directe des théorèmes établis par Hohenberg et Kohn et est utilisée pour la DFT moderne. Dans les méthodes précédentes (méthodes de type Thomas-Fermi) ces théorèmes, les orbitales moléculaires n'étaient pas prises en compte. De ce fait, elles offraient généralement une description assez imprécise de l'énergie cinétique du système électronique. Ce problème a été contourné par Kohn et Sham en 1965, en réintroduisant les orbitales, ce qui permet un calcul plus précis de l'énergie cinétique à partir d'une fonction d'onde.<sup>161</sup> Ainsi, la méthode de Kohn-Sham propose d'abord le problème par le calcul de l'énergie cinétique d'un système fictif à N électrons indépendants (sans interactions) et de même densité à l'état fondamental que celle du système réel considéré. L'énergie du système est alors décomposée en quatre contributions énergétiques : l'énergie cinétique du système fictif (en général proche de celle du système réel), l'énergie d'interaction entre les électrons et les noyaux, l'énergie classique coulombienne de répulsion entre électrons et l'énergie d'échange-corrélation.

$$E_0 = T_s + \int \rho(\vec{r})v(\vec{r}) d^3r + \frac{1}{2} \iint \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3r d^3r + E_{xc} [\rho] \quad (5.6)$$

où  $T_s$  représente l'énergie cinétique et  $E_{xc}$  l'énergie d'échange-corrélation.

### II.4 Traitement de l'échange-corrélation et fonctions de base

La théorie de la fonctionnelle de la densité appliquée dans le cadre de l'approche orbitale de Kohn et Sham<sup>161</sup> demeure exacte dans son formalisme. Progressivement, la partie inconnue dans la fonctionnelle  $E[\rho]$  a été réduite à une fonctionnelle universelle  $F_{HK}[\rho]$  et finalement à une fonctionnelle de l'énergie d'échange et de corrélation  $E_{xc}[\rho]$ . A ce stade, il est nécessaire d'approcher l'expression de cette fonctionnelle d'échange et de corrélation, de sorte qu'elle offre une description aussi précise que possible du système.

Comme vu précédemment, c'est la qualité de la description de la densité électronique qui permet de déterminer de manière exacte les propriétés du système en DFT. En pratique, la

qualité de la description du système va dépendre des différents paramètres appliqués sur la méthode (principalement choix de la fonctionnelle d'échange-corrélation et de la base).

Il existe ainsi plusieurs niveaux de théorie dont les plus faibles ne tiennent compte que de la densité locale des électrons (LDA, *local density approximation*) et des niveaux supérieurs qui prennent également en compte la dépendance des orbitales non occupées. Ces différentes classes de fonctionnelles d'échange-corrélation sont généralement classées selon leur niveau de théorie sur une « échelle de Jacob de la DFT » de Perdew (Figure 90).<sup>162</sup>

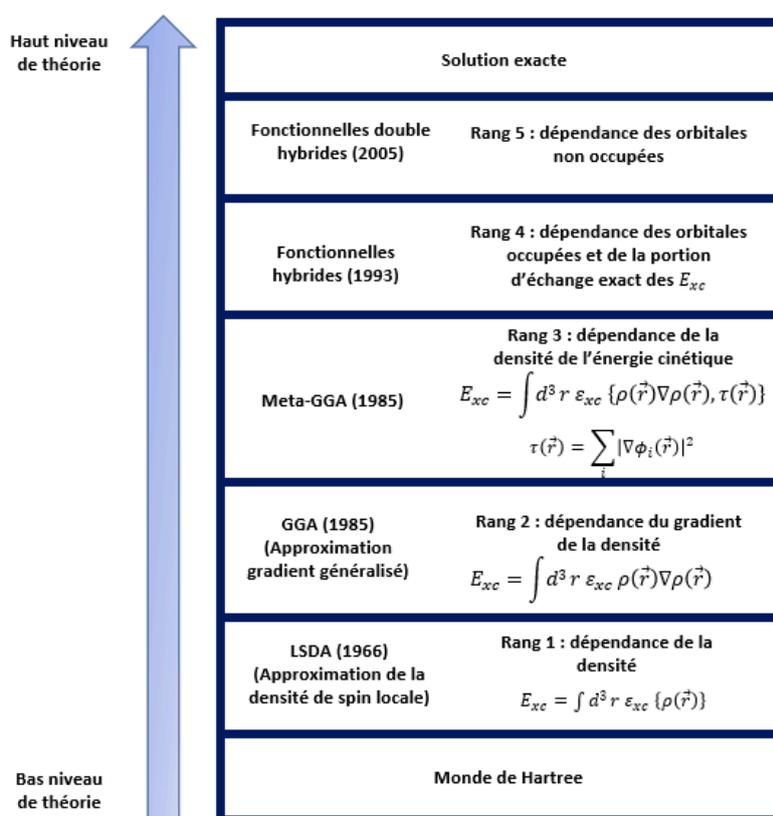


Figure 90 Echelle de Jacob" de la DFT de Perdew.<sup>146</sup>

Les fonctionnelles hybrides ont été introduites pour la première fois par Becke en 1993.<sup>163</sup> Elles ont pour caractéristique d'avoir l'échange basé sur la méthode de Hartree-Fock (HF) dépendante des orbitales occupées. La méthode HF vise à corriger l'erreur issue de l'interaction de l'électron avec lui-même (self-interaction). Deux exemples de fonctionnelles hybrides sont la LDA (Approximation de la densité de spin locale) et GGA (Approximation gradient généralisé).

En chimie quantique, une base est un ensemble de fonctions utilisées afin de modéliser des orbitales moléculaires, qui sont développées comme combinaisons linéaires de telles fonctions

avec des poids ou coefficients à déterminer. Ces fonctions sont habituellement des orbitales atomiques, car centrées sur les atomes, mais des fonctions centrées sur les liaisons ou les fonctions centrées des doublets non liants ont été utilisées comme l'ont été des paires de fonctions centrées sur les deux lobes d'une orbitale p. De plus, des bases composées d'ondes planes jusqu'à une longueur d'onde de coupure sont également utilisées, particulièrement dans les calculs sur des systèmes avec des conditions périodiques aux limites. De même que pour les fonctionnelles, il existe de très nombreuses bases permettant de développer les orbitales moléculaires, avec des descriptions électroniques plus ou moins détaillées. Il est également possible d'ajouter à ces bases des fonctions de polarisation, notées selon Pople par un astérisque \* ou deux astérisques \*\* pour les atomes légers (hydrogène et hélium).<sup>164</sup>

L'ajout de fonctions diffuses (par un signe + ou deux signes ++ pour les atomes légers) est également possible afin de donner de la flexibilité à la partie des orbitales atomiques éloignées du noyau, mais aussi pour décrire les anions ou les moments dipolaires. Dans tous les cas et quelle que soit la fonctionnelle ou la base utilisée, il est nécessaire, selon le niveau de théorie voulu, d'utiliser de manière adéquate les ressources disponibles, et de trouver un compromis entre temps de calcul et précision de la description du système étudié, une optimisation de structure selon un haut niveau de théorie consommant de pair d'importantes ressources numériques.

## II.5 Méthode pour l'étude des états excités

L'interaction entre la lumière (c'est-à-dire un champ électrique oscillant dans le temps et dans l'espace) et la matière doit être traitée en utilisant des méthodes qui prennent en compte les perturbations temporelles. Dans ce cas, l'hamiltonien est divisé en deux parties

$$\hat{H} = \hat{H}^0 + \hat{H}^1(t) \quad (5.7)$$

où  $\hat{H}^0$  est la partie indépendante du temps lié au système sans aucune perturbations et  $\hat{H}^1(t)$  est celle dépendante du temps qui décrit les interactions entre la lumière et la molécule.

Pour pouvoir décrire ce phénomène, il est nécessaire d'utiliser une méthode qui puisse prendre en compte tous les termes présents dans l'équation 5.7.

## II.6 Théorie de la fonctionnelle de la densité dépendant du temps (*Time-Dependent Density Functional Theory, TDDFT*)

La théorie de la fonctionnelle de la densité (DFT) est limitée à l'étude des processus indépendants du temps et permet seulement le calcul des propriétés de l'état fondamental. Cette théorie peut toutefois être généralisée aux processus dépendants du temps, et prend alors le nom de théorie de la fonctionnelle de la densité dépendante du temps (TDDFT, pour l'acronyme anglais). La TDDFT donne accès aux propriétés des états excités et aux polarisabilités (et hyperpolarisabilités) par l'intermédiaire de la théorie de la réponse linéaire. En effet, le champ potentiel étant considéré comme indépendant du temps, des solutions ont été cherchées sous la forme d'ondes stationnaires, dont la partie dépendant du temps n'était pas un objet d'étude. A partir de maintenant, l'hamiltonien de Schrödinger intègre une dépendance par rapport au temps ; l'équation sera donc traitée en conséquence.

Comme l'avait fait Hohenberg et Kohn<sup>165</sup> pour la DFT, Runge et Gross ont rendu légitime en 1984 l'utilisation de la densité électronique dépendant du temps.<sup>80</sup> Ceci donna naissance à la théorie de la fonctionnelle de la densité dépendant du temps.

### II.6.1 Théorie de la réponse linéaire

L'approche dépendant du temps permet, d'une part, d'étudier la perturbation du système à un temps  $t_0$ , puis, d'autre part, de propager cette perturbation pour un temps  $t > t_0$ . L'étude de l'évolution de la propagation de cette perturbation conduit à l'obtention du spectre d'absorption. L'un des moyens efficaces pour étudier cette propagation est la théorie de la réponse linéaire.<sup>81</sup> Cette dernière travaille dans l'espace des fréquences  $\omega$  plutôt que dans l'espace temporel. L'opération s'effectue par l'intermédiaire de la transformée de Fourier.

Soit  $f$  une fonction continue de la variable temporelle, la transformée de Fourier  $F$  de  $f$  donne :

$$F(f): \omega \mapsto \hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \quad (5.8)$$

Suivant cette approche, la densité électronique dépendante du temps s'écrit comme la somme de la densité électronique calculée à l'état fondamental  $\rho_{GS}$  additionnée d'une perturbation  $\delta_{\hat{\rho}}(\vec{r}, \omega)$  :

$$\hat{\rho}(\vec{r}, \omega) = \rho_{GS}(\vec{r}) + \delta\hat{\rho}(\vec{r}, \omega) \quad (5.9)$$

Dans le système des électrons en interaction, le potentiel extérieur s'écrit également comme la somme du potentiel extérieur calculé à l'état fondamental et additionné d'un potentiel perturbateur :

$$\hat{v}_{ext}(\vec{r}, t) = v_{ext}(\vec{r}) + \delta\hat{v}_{ext}(\vec{r}, \omega) \quad (5.10)$$

## II.6.2 Propriétés spectroscopiques obtenues par la TDDFT

A travers la TDDFT, il est possible de calculer les excitations verticales (c'est-à-dire des excitations de l'état fondamental à celui excité très rapides) entre l'état fondamental et les différents états excités. Ce calcul donne accès aux énergies d'excitation et aux forces d'oscillateur qui leur sont associées.

Dans la spectroscopie de DC, l'une des principales quantités d'intérêt est la force rotationnelle (« rotatory strength »)  $R$ , qui détermine le signe et l'intensité d'une bande de DC électronique. Expérimentalement, elle est obtenue en intégrant la zone située sous la bande correspondante et elle est représentée par l'équation suivante :

$$R_{\text{exp}} = 2,297 \times 10^{-39} \int_{\lambda_1}^{\lambda_2} \frac{\Delta\epsilon(\lambda)}{\lambda} d\lambda \quad (5.11)$$

où  $\lambda$  est la longueur d'onde et  $\Delta\epsilon$  est la différence entre les coefficients d'excitation de la lumière polarisée circulairement gauche et droite.

Au niveau théorique,  $R_{oi}$  (pour la transition entre l'état fondamental électronique et l'état excité  $i$ ) peut être obtenue selon Rosenfield par la partie imaginaire du produit scalaire entre les moments de transition électrique et magnétique :

$$R_{oi} = |\vec{\mu}_{oi}| |\vec{m}_{oi}| \cdot \cos(\vec{\mu}_{oi}, \vec{m}_{oi}) \quad (5.12)$$

où  $\vec{\mu}_{oi}$  et  $\vec{m}_{oi}$  sont respectivement les moments de transition électrique et magnétique. Ainsi, le signe et la valeur absolue de  $R$  dépendent des valeurs absolues des deux moments de transition et de leur orientation relative les uns par rapport aux autres.

Souvent, les spectres théoriques sont directement comparés aux spectres expérimentaux. Pour cela, il ne suffit pas de comparer les positions des bandes expérimentales avec les énergies

d'excitation et les intensités de transitions théoriques. Il est possible que plusieurs transitions se trouvent proches les uns des autres. Par conséquent, il est fortement recommandé de simuler réellement un spectre de DC. Cela se fait généralement en superposant des fonctions gaussiennes pour chaque transition :

$$\Delta \epsilon(E) = \frac{1}{2.297 \times 10^{-39}} \frac{1}{\sqrt{2\pi}\sigma} \sum_i^A \Delta E_i R_i e^{[-(E-\Delta E_i/2\sigma)]^2} \quad (5.13)$$

où  $\sigma$  est la largeur d'une bande d'absorption à une hauteur de  $\frac{1}{e}$ ,  $A$  est le nombre d'états excités, et  $E_i$  et  $R_i$  sont l'énergie d'excitation et la force de rotation de la transition  $i$ . Habituellement, une valeur de  $\sigma = 0,2$  eV est choisie pour la largeur de la bande.

### II.6.3 Transition électronique

Une transition purement électronique correspond à une excitation ou désexcitation verticale entre l'état fondamental, état singulet  $S_0$ , et un état singulet excité, noté comme  $S_n$ , où  $n$  est un indice strictement positif, représentant le numéro de l'état considéré. La TDDFT permet de calculer ce type de transitions. Dans le cas du processus d'absorption, l'excitation verticale part de la position d'équilibre de l'état fondamental  $S_0$ , pour rejoindre un point de la courbe potentielle de l'état excité considéré  $S_n$  (Figure 91). Dans le cas du processus d'émission, la désexcitation verticale part de la position d'équilibre de l'état singulet excité  $S_n$  pour rejoindre un point de la courbe potentielle de l'état fondamental  $S_0$  (Figure 91). Le calcul des transitions verticales est la première étape permettant l'obtention des spectres d'absorption ou d'émission. Cette première approche se révèle souvent suffisante pour décrire les propriétés spectroscopiques d'un système chimique, cependant certains systèmes nécessitent tout de même de coupler les transitions vibrationnelles aux transitions électroniques.

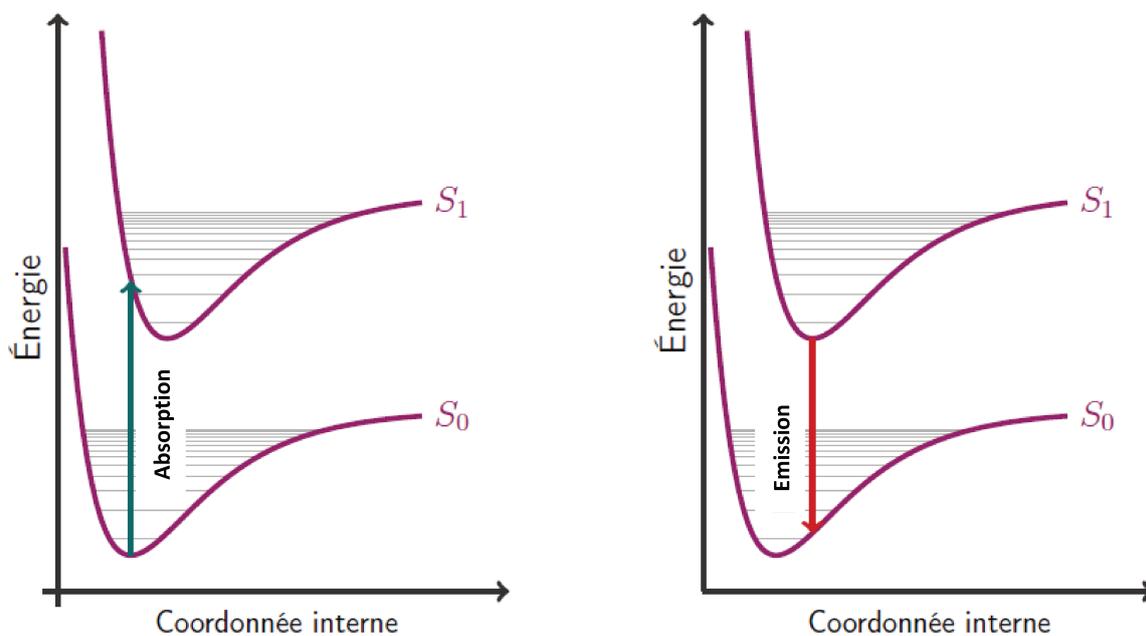


Figure 91 Transition verticale entre l'état fondamental  $S_0$ , et le premier état singulet excité  $S_1$ .  
Processus d'absorption (gauche) et d'émission (droite).

En réalité, une excitation électronique est toujours couplée avec des transitions vibrationnelles. Par différence avec une transition purement électronique, une transition vibrationnelle est une excitation ou désexcitation verticale entre un niveau vibrationnel de l'état fondamental et un niveau vibrationnel d'un état excité  $S_n$  (Figure 92). Ces transitions vibrationnelles, en absorption ou en émission sont une seconde étape à la modélisation spectroscopique.

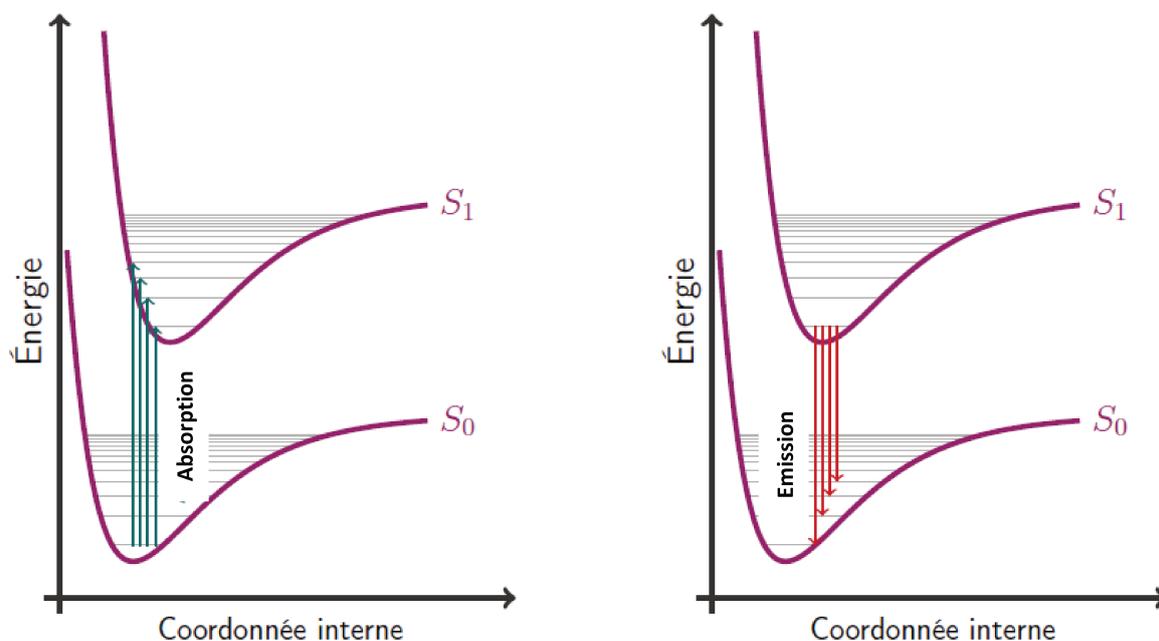


Figure 92 Transitions vibrationnelles entre l'état fondamental  $S_0$ , et le premier état singulet excité  $S_1$ . Processus d'absorption (gauche) et d'émission (droite).

#### II.6.4 Point de vue spectral

Dans le cas où le spectre expérimental d'un système chimique en absorbance révèle une bande d'absorption « de forme simple », une transition électronique est capable de décrire correctement cette même bande. Un simple profil gaussien centré sur la transition verticale en question permet alors de reproduire la structure de cette bande d'absorption. En revanche, dans le cas où le spectre expérimental d'un système chimique en absorbance révèle une bande d'absorption complexe, la seule prise en compte de la transition électronique ne suffit plus à reproduire la structure de bande.

#### II.6.5 Les effets de l'environnement

A l'heure actuelle, lors d'un calcul des propriétés spectroscopiques, il est important d'introduire les effets de l'environnement. En fait, l'environnement peut ajuster les propriétés optiques d'une molécule donnée (c'est-à-dire que le changement solvatochromique peut donner un aperçu précieux de la nature d'un état excité).

Parmi les modèles utilisés en chimie théorique pour décrire les effets du solvant, dans le cadre de ma thèse, nous avons utilisé l'une des méthodes multi-échelles QM/MM : la méthode de l'inclusion polarisable (Polarizable Embedding, PE). Dans cette partie du chapitre, nous allons présenter cette méthode de manière générale.

## II.6.6 Méthode de l'inclusion polarisable (Polarizable Embedding, PE)

Les méthodes utilisées pour montrer les effets de l'environnement (le solvant dans notre cas), comme le modèle de continuum polarisable (Polarizable Continuum Model, PCM)<sup>166</sup>, ne permettent pas de décrire les interactions intermoléculaires locales, comme les liaisons hydrogène qui peuvent se former entre la molécule et le solvant pendant la simulation. Afin de surmonter ce problème, les molécules du solvant (au moins celles proche du soluté) doivent être décrites de manière explicite.

En revanche, l'ajout des molécules du solvant à la région de mécanique quantique est très exigeant du point de vue informatique en raison du coût des calculs de chimie quantique (même au niveau des calculs DFT). Ainsi, les travaux fondateurs de Warshel et Levitt<sup>167</sup> ont proposé de diviser le système en deux sous-unités : une région QM (nommé core), traitée selon une méthode de chimie quantique donnée, et la région MM, décrite avec la mécanique classique. La partition du système dans les régions QM et MM et leur interaction se reflètent sous la forme de l'hamiltonien QM/MM.

Fondamentalement, nous pouvons avoir trois types d'hamiltoniens QM/MM : l'inclusion mécanique, l'inclusion électrostatique et l'inclusion polarisable. Parmi ces méthodes, celle de l'inclusion polarisable du solvant est le type de calcul le plus complexe. Dans cette méthode, il existe une polarisation mutuelle entre la densité électronique de la région QM, les charges fixes distribuées et les dipôles induits des molécules de l'environnement (par exemple ONIOM-PE<sup>168</sup>, QM/MMPol<sup>169</sup> et le PE<sup>83</sup>).

Parmi ces méthodes, dans ce travail de thèse nous avons utilisé une méthode multi-échelle d'inclusion polarisable développée par le groupe du Pr. Jacob Kongsted à l'Université du Danemark du Sud (University of Southern Denmark, SDU).<sup>82,83</sup>

La flexibilité du schéma de l'inclusion polarisable (PE) permet d'utiliser différentes méthodes chimiques quantiques comme la DFT. La méthode PE est formulée dans le cadre général de la théorie de la réponse dépendante du temps de la mécanique quantique et peut être utilisée dans le calcul des propriétés d'état excité au sein de la TDDFT. Différentes spectroscopies peuvent être simulées à l'aide de la théorie de la réponse linéaire, parmi lesquelles figure le DCE.

Dans le schéma de la PE, la région MM polarisable est représentée par l'expansion multipolaire (charges, dipôles, quadripôles, etc ...) des atomes MM et de polarisabilités afin de permettre la polarisation à partir de la densité électronique du système QM.

L'énergie totale du PE est donnée par l'équation :

$$E_{TOT} = E_{QM}^{PE} + E_{es}^{PE} + E_{ind}^{PE} \quad (5.14)$$

où  $E_{QM}^{PE}$  est l'énergie de la région « core » isolée (évaluée à un niveau donné de théorie),  $E_{es}$  est l'énergie d'interaction électrostatique et  $E_{ind}$  est l'énergie d'induction due à la polarisation des fragments de l'environnement.

De plus, dans la DFT, l'effet de l'environnement est pris en compte en utilisant un opérateur Kohn-Sham :  $\hat{f} = \hat{f}^{vac} + \hat{v}^{PE}$  (5.15)

où  $\hat{f}^{vac}$  est l'opérateur Kohn-Sham dans le vide et  $\hat{v}^{PE}$  est le potentiel d'inclusion qui décrit le potentiel exercé sur la région quantique par l'environnement.

## II.7 La dynamique moléculaire

Lorsqu'on décide d'étudier et de simuler les propriétés optiques en utilisant une méthode comme celle de l'inclusion polarisable, une étude préalable de l'espace conformationnel doit être réalisée. L'ensemble des géométries est éventuellement exploré et obtenu à travers des simulations de dynamique moléculaire (DM) classique.

L'image réelle d'un système moléculaire est loin d'être l'image statique usuellement suggérée. Les atomes et les groupes d'atomes sont en perpétuel mouvement. Ces mouvements peuvent être concertés et essentiels à la fonction des molécules biologiques. La dynamique moléculaire produit une succession de conformations qui décrit l'évolution au cours du temps de la position des atomes (leur trajectoire). Les variations des positions atomiques résultent des diverses forces auxquelles les atomes sont soumis au cours du temps. La mécanique et la dynamique moléculaires partagent la même fonction d'énergie potentielle  $U_p$  pour décrire l'ensemble des interactions interatomiques du système d'étude. Tout est décrit dans le champ de forces qui est une équation modélisant l'énergie potentielle d'un ensemble d'atomes comportant un certain nombre de paramètres. Pour des simulations de peptides et de protéines, les champs de forces les plus communs comme par exemple AMBER ff14SB, CHARMM, GROMOS, OPLS-AA.<sup>170-173</sup> contiennent les paramètres spécifiques des acides aminés. Pour certains types de molécules, contenant des acides aminés modifiés, qui ne sont pas décrits dans les champs de forces précédemment mentionnés, ont été développés des champs de forces qui contiennent les paramètres pour les décrire.<sup>174</sup>

## II.7.1 Le principe

La dynamique moléculaire consiste à étudier la trajectoire d'une molécule en lui appliquant les lois de la mécanique classique newtonienne. Une molécule est considérée comme un assemblage d'atomes dont les positions spatiales varient au cours du temps.

Chaque atome est considéré comme une masse ponctuelle dont le mouvement est déterminé par l'ensemble des forces qui s'exercent sur elle. Il en résulte un mouvement atomique correspondant à des oscillations autour d'un minimum d'énergie ou au passage d'un de ces minima à un autre. La dynamique moléculaire possède ainsi la faculté d'extraire la molécule d'un minimum local.

Cette méthode repose sur l'intégration des lois du mouvement de Newton pour chaque atome  $i$ . La seconde loi de Newton est exprimée par l'équation 5.16 où  $\vec{F}_i$  est la force exercée sur la particule  $i$  de masse  $m_i$ , de position  $\vec{r}_i(x, y, z)$  avec une accélération de  $\vec{a}_i$  :

$$\vec{F}_i(t) = m_i \vec{a}_i(t) \quad (5.16)$$

L'accélération d'un atome  $i$  étant la dérivée seconde des coordonnées  $\vec{r}_i(x, y, z)$  à un temps  $t$ , nous en déduisons que :

$$\vec{F}_i(t) = m_i \cdot \frac{d\vec{v}_i(t)}{dt} = \frac{m_i \cdot d^2\vec{r}_i(t)}{dt^2} \quad (5.17)$$

Si  $\vec{r}_1, \dots, \vec{r}_n$  sont les coordonnées des différents atomes à un instant  $t$ , on peut déterminer la force  $\vec{F}_i$  qui s'exerce à cet instant sur un atome  $i$  par dérivation par rapport à ses coordonnées spatiales de la fonction énergie potentielle  $E(\vec{r}_1, \dots, \vec{r}_n)$  :

$$\vec{F}_i(t) = - \frac{d}{d\vec{r}_i} E(r_1, \dots, r_n) \quad (5.18)$$

Pour obtenir la trajectoire des atomes en fonction du temps, on va réaliser une intégration numérique des équations de mouvements sur un court intervalle de temps. Plusieurs algorithmes de résolution numériques peuvent être utilisés, avec plus ou moins de succès suivant le système étudié. L'un des plus performants est l'algorithme de Verlet. Il repose sur le principe d'un développement limité pas à pas. L'algorithme de Verlet réduit le taux d'erreurs en calculant la position au pas de temps suivant à partir des positions courante et précédente, sans faire appel à la vitesse.

Ces équations de mouvements sont déterministes. Il faut, pour réaliser une simulation de dynamique moléculaire, partir d'une structure raisonnable du système (expérimentale ou modèle) que l'on veut étudier et on déterminera les coordonnées et vitesses initiales en procédant ainsi :

- Une minimisation d'énergie du système fournira le jeu de coordonnées de départ de la dynamique. Cette étape de minimisation nécessaire permet d'éviter qu'une énergie mécanique trop forte ne détériore l'assemblage moléculaire dès le début de la production de dynamique moléculaire.
- Les vitesses de départ des atomes sont attribuées aléatoirement à partir d'une distribution gaussienne de Maxwell-Boltzmann correspondant à la température  $T$ .

### **II.7.2 Le système vu comme un ensemble thermodynamique**

L'ensemble microcanonique NVE (nombre de particules, volume et énergie totale constants) est l'ensemble naturel de la dynamique moléculaire d'un système conservatif, car l'énergie est conservée au cours du temps. Cependant, cet ensemble se trouve relativement éloigné des conditions expérimentales d'un système réel. Les simulations ont donc été par la suite réalisées dans différents ensembles statistiques (canonique, isobare-isotherme, etc....) selon les conditions dans lesquelles on veut la réaliser et afin de se rapprocher au mieux des observations expérimentales.

### **II.7.3 Conditions périodiques aux limites**

En dynamique moléculaire, la protéine est le plus souvent placée au centre d'une boîte qui va contenir d'autres molécules (eau, ions, lipides) visant à reproduire son environnement biophysique. Toutefois, si cette boîte de simulation était dans le vide, on observerait rapidement la dispersion des éléments qui la composent. Pour maintenir les propriétés biophysiques du système, il est nécessaire d'appliquer des conditions périodiques aux limites (« Periodic Boundary Conditions » ou PBC). Cette méthode consiste à copier le système étudié dans les trois directions de l'espace (Figure 93). Ainsi, une particule qui sortirait par un côté de la boîte arriverait directement par le côté opposé. En utilisant ce biais, le système simulé paraît avoir des bornes infinies et garde ainsi une densité constante. Le calcul des interactions non liées tient bien évidemment compte des conditions périodiques et fait intervenir la version la plus proche de chacune des particules impliquées.

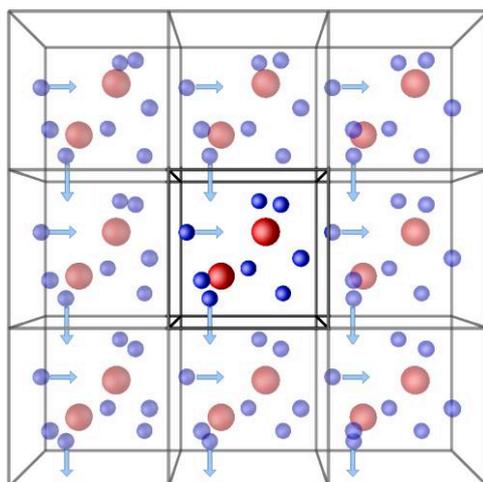


Figure 93 Conditions périodiques utilisées pendant les simulations de dynamique moléculaire classiques. La molécule initiale (en rouge) est immergée dans un solvant (en bleu) dans une boîte cubique (en noir). Pour effectuer le calcul des interactions entre atomes non liés, cette boîte est dupliquée à l'infini dans toutes les directions (les duplicata sont représentés en gris).

#### II.7.4 La dynamique moléculaire sous contraintes

Il est possible dans le cadre de certaines études de biaiser la simulation de dynamique moléculaire par l'ajout dans le champ de force d'une énergie de contrainte. Ces potentiels peuvent être de différentes formes selon le type de contraintes que l'on veut appliquer à la molécule pendant la simulation.

Cette opération consiste à rajouter comme énergie de contrainte au champ de force une énergie constituée par deux potentiels semiharmoniques.

Entre les valeurs  $r_2$  et  $r_3$  aucune contrainte n'est appliquée. En deçà et au-delà des valeurs  $r_1$  et  $r_4$ , des potentiels de contraintes linéaires sont appliqués. Le potentiel de contrainte harmonique n'est appliqué qu'entre les valeurs  $[r_1, r_2]$  et  $[r_3, r_4]$  (Figure 94).

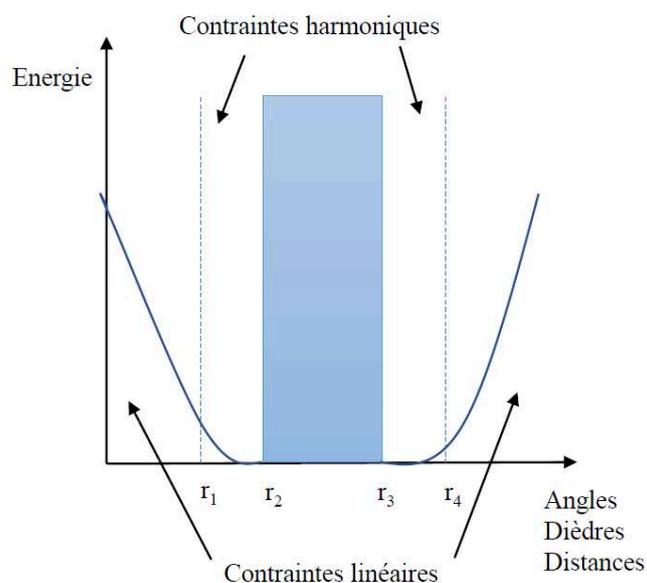


Figure 94 Représentation en fonction de la coordonnée de réaction choisie de la forme de l'énergie de contrainte imposée au système.

Au temps initial, nous avons une série de positions et de vitesses pour chaque particule dans le système. Dans les ensembles NVT (particules, volume total et température) et NPT (particules, pression et température), les vitesses sont attribuées en fonction d'une température fixe avec un bain de chaleur ; en fait, la température d'un corps est une mesure de l'énergie cinétique moyenne de ses particules et différents thermostats ont été développés. La pression peut également être contrôlée à l'aide d'un pressostat.<sup>175-181</sup>

Dans cette partie du travail de thèse, nous avons souhaité calculer les spectres de DCE pour les coudes  $\beta$  de type I, I', II et II'. Ces calculs ont été envisagés car, pour les structures secondaires de type coude, il n'est pas évident de stabiliser un peptide en solution dans ce type de conformation. De plus, plusieurs conformations pourraient être présentes en solution lors d'une étude expérimentale. Pour cette raison, nous avons décidé d'utiliser un peptide modèle minimal en conformation coude  $\beta$  de type I, I', II ou II'. Avant les calculs des spectres de DCE, l'exploration de l'espace conformationnel pour ce peptide modèle minimal a été réalisée à travers des simulations de dynamique moléculaire sous contraintes. L'approche *ab initio* utilisée est la méthode de la TDDFT à réponse linéaire qui permet de simuler la spectroscopie de DCE. Nous y avons couplé un schéma multi-échelle à inclusion polarisable afin de pouvoir prendre en compte l'effet du solvant.

### III. Article

Dans cette partie du chapitre, nous présenterons l'article, en format d'épreuve, accepté au journal *Physical Chemistry Chemical Physics*.



1 PCCP

5 PAPER

## Characterization of $\beta$ -turns by electronic circular dichroism spectroscopy: a coupled molecular dynamics and time-dependent density functional theory computational study<sup>†</sup>

10 Cite this: DOI: 10.1039/c9cp05776e

15 Mattia Migliore,<sup>‡a</sup> Andrea Bonvicini,<sup>‡a</sup> Vincent Tognetti,<sup>§b</sup> Laure Guilhaudis,<sup>\*a</sup> Marc Baaden,<sup>b</sup> Hassan Oulyadi,<sup>a</sup> Laurent Joubert<sup>a</sup> and Isabelle Segalas-Milazzo<sup>a</sup>

20 Electronic circular dichroism is one of the most used spectroscopic techniques for peptide and protein structural characterization. However, while valuable experimental spectra exist for  $\alpha$ -helix,  $\beta$ -sheet and random coil secondary structures, previous studies showed important discrepancies for  $\beta$ -turns, limiting their use as a reference for structural studies. In this paper, we simulated circular dichroism spectra for the best-characterized  $\beta$ -turns in peptides, namely types I, II, I' and II'. In particular, by combining classical molecular dynamics simulations and state-of-the-art quantum time-dependent density functional theory (with the polarizable embedding multiscale model) computations, two common electronic circular dichroism patterns were found for couples of  $\beta$ -turn types (namely, type I/type II' and type II/type I'), at first for a minimal di-peptide model (Ace-Ala-Ala-NHMe), but also for all sequences tested with non-aromatic residues in the central positions. On the other hand, as expected, aromatic substitution causes important perturbations to the previously found ECD patterns. Finally, by applying suitable approximations, these patterns were subsequently rationalized based on the exciton chirality rule. All these results provide useful predictions and pave the way for a possible experimental characterization of  $\beta$ -turns based on circular dichroism spectroscopy.

25 Received 23rd October 2019,  
Accepted 5th December 2019

30 DOI: 10.1039/c9cp05776e

35 rsc.li/pccp

### 1. Introduction

35  $\beta$ -turns (also known as tight-turns,  $\beta$ -bends or reverse turns), recognized in 1968 by Venkatachalam,<sup>1</sup> are motifs that allow reversal of the polypeptide chain direction in proteins and peptides. They constitute, after  $\alpha$ -helices and  $\beta$ -sheets, the third most common secondary structural elements present in these macromolecules.<sup>2</sup> Their structural role notably occurs in globular protein folding,<sup>3–7</sup> in molecular recognition processes between proteins, and in interactions between peptide substrates and receptors.<sup>8</sup> In particular,  $\beta$ -turns have been reported to play an essential role in recognition phenomena between peptide-activated G protein-coupled receptors (GPCRs) and their endogenous peptide ligands.<sup>9,10</sup> Moreover, the structural

35 characterization of  $\beta$ -turns is mandatory for a full understanding of the molecular interactions taking place between partners and to aid the design of pseudo-peptidic analogs.<sup>11,12</sup>

40 From a structural point of view,  $\beta$ -turns involve four sequential amino-acids (denoted as  $i$ ,  $i + 1$ ,  $i + 2$ , and  $i + 3$ , following the ordering from the N- to the C-terminal residues), and they can be classified into different types based on the  $\Phi$  and  $\Psi$  dihedral angles of their central residues ( $i + 1$  and  $i + 2$ ). They are thus composed by three peptide bonds and are stabilized by an intramolecular hydrogen bond between the  $H_N$  atom in residue  $i + 3$  and the CO group of residue  $i$ , as depicted in Fig. 1a.

45 Originally, six types of  $\beta$ -turns were defined (I, II, III and their mirror images I', II' and III').<sup>1</sup> Over the last 50 years, with the increasing number of solved protein structures, additional types were identified (IV, V, VI, VII and VIII), some subtypes were introduced (IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, and VIb), while some types were eliminated (III, III', V and VII) and incorporated into other categories (for example, III into I, and III' into I').<sup>1,2,6,13–19</sup> Although a new nomenclature has recently been proposed,<sup>19</sup> we decided, in order to allow direct comparisons with the dense literature on this topic, to keep the classification proposed by De Brevern in 2016.<sup>18</sup> As shown in Table 1, it comprises 12 types of  $\beta$ -turns.

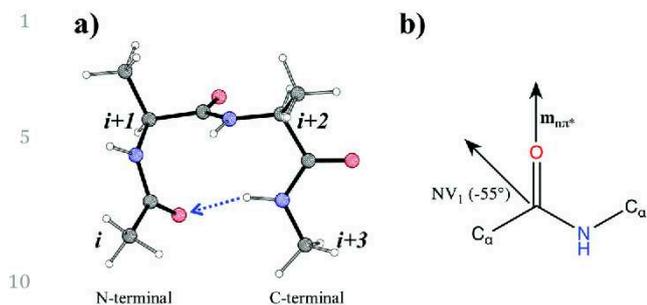
50 <sup>a</sup> Normandy Univ., COBRA UMR 6014 & FR 3038, Université de Rouen, INSA Rouen, CNRS, 1 rue Tesnière, 76821 Mont St Aignan, Cedex, France.

E-mail: vincent.tognetti@univ-rouen.fr, laure.guilhaudis@univ-rouen.fr

<sup>b</sup> Laboratoire de Biochimie Théorique, CNRS, UPR9080, Univ. Paris Diderot, Sorbonne Paris Cité, PSL Research University, 13 rue Pierre et Marie Curie, 75005 Paris, France

55 <sup>†</sup> Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cp05776e

<sup>‡</sup> A. B. and M. M. contributed equally to this work.



**Fig. 1** (a) Chemical structure of the Ace<sub>7</sub>-A<sub>i+1</sub>-A<sub>i+2</sub>-NHMe<sub>i+3</sub> peptide used in this work as a minimal model (H in white, C in grey, N in blue, and O in red). (b) Direction of the  $n \rightarrow \pi^*$  magnetic dipole transition moment and NV1 ( $\pi_{nb} \rightarrow \pi^*$ ) electric dipole transition moment with respect to the peptide bond.

**Table 1**  $\beta$ -turn classification based on the  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$  and  $\Psi_{i+2}$  dihedral angle canonical values according to De Brevern (ref. 18). Types in bold are those considered in this work

Type	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$
<b>I</b>	$-60^\circ$	$-30^\circ$	$-90^\circ$	$0^\circ$
<b>I'</b>	$+60^\circ$	$+30^\circ$	$+90^\circ$	$0^\circ$
<b>II</b>	$-60^\circ$	$+120^\circ$	$+80^\circ$	$0^\circ$
<b>II'</b>	$+60^\circ$	$-120^\circ$	$-80^\circ$	$0^\circ$
IV <sub>1</sub>	$-120^\circ$	$+130^\circ$	$+55^\circ$	$+41^\circ$
IV <sub>2</sub>	$-85^\circ$	$-15^\circ$	$-125^\circ$	$+55^\circ$
IV <sub>3</sub>	$-71^\circ$	$-30^\circ$	$-72^\circ$	$-47^\circ$
IV <sub>4</sub>	$-97^\circ$	$-2^\circ$	$-117^\circ$	$-11^\circ$
Via1	$-60^\circ$	$+120^\circ$	$-90^\circ$	$0^\circ$
Via2	$-120^\circ$	$+120^\circ$	$-60^\circ$	$0^\circ$
Vib	$-135^\circ$	$+135^\circ$	$-75^\circ$	$160^\circ$
VIII	$-60^\circ$	$-30^\circ$	$-120^\circ$	$120^\circ$

It should be emphasized that, as  $\beta$ -turns are flexible secondary structures, each  $\Phi$  and  $\Psi$  dihedral angle of the  $i + 1$  and  $i + 2$  residues is allowed to deviate up to  $\pm 30^\circ$  compared to its canonical value.<sup>18</sup> Last but not least, while  $\alpha$ -helices and  $\beta$ -sheets are ordered and repetitive secondary structures,  $\beta$ -turns are ordered but non-repetitive motifs. For all these reasons, the establishment of reference electronic circular dichroism (ECD) spectra for  $\beta$ -turns presents more difficulties than those encountered for the two other major secondary structures ( $\alpha$ -helix and  $\beta$ -sheet) and for the random coil conformation.<sup>20–23</sup> Accordingly, it still remains an open challenge in the field of structural biology, which may be tackled from either experimental or theoretical perspectives.

It should be mentioned that other experimental techniques, such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, may be used to determine the secondary and tertiary structures of peptides and proteins. However, although the first method allows identifying all secondary structural motifs present in the molecule in the solid state, its use is intrinsically limited by the ability of the molecule to crystallize. This is a particular issue for peptides whose flexibility often hinders crystallization.<sup>24,25</sup>

On the other hand, NMR gives access to the solution structure of peptides or proteins, for which helices and sheets

**Table 2** Classification scheme for calculated ECD spectra proposed by Woody (ref. 32). The classes were established as a function of the number and the position of maxima and minima

Class	Sign	Wavelength range [nm]
A	–	210–220
	+	195–200
	–	<190
B	–	>220
	+	200–210
	–	<190
C	–	200–210
	+	180–195
D	–	>225
	+	210–220
	–	190–200
	+	<190

are easily distinguished, whereas turns are more difficult to identify.<sup>26</sup> Indeed, this technique only allows characterization of four types of  $\beta$ -turns (*i.e.* I, I', II and II') among the twelve described in Table 1, and the distinction between these four types is often not absolute.<sup>26</sup> Moreover, this interpretation is made more complicated in the case of small size peptides, which frequently exhibit more than one conformation in solution.<sup>27–29</sup>

Besides, it should be pointed out that both methods are time and sample consuming compared to electronic circular dichroism, which requires a small amount of sample and is easy to implement from an experimental viewpoint. Circular dichroism spectroscopy is thus the preferred technique used to screen the secondary structure of polypeptides. Nevertheless, the lack of reference spectra for  $\beta$ -turns makes the estimations of polypeptide secondary structural content rather unreliable when the molecule contains a large amount of this kind of motif.<sup>22,30,31</sup>

From a theoretical point of view, the first work on ECD spectroscopy of  $\beta$ -turns was actually performed by Woody in 1974,<sup>32</sup> who identified *via* semi-empirical calculations (an improved version of the Kirkwood–Moffitt–Tinoco coupled oscillator theory based on the electrostatic interaction between chromophores,<sup>33</sup> also known as the matrix method,<sup>34</sup> which includes the description of charge transfer transitions between amide groups<sup>35</sup>) the fundamental role of the  $n \rightarrow \pi^*$  and  $\pi_{nb} \rightarrow \pi^*$  transitions for types I, II and III  $\beta$ -turns. More specifically, these calculations indicated that all analysed types were generally characterized by one negative  $n \rightarrow \pi^*$  band at  $\sim 225$  nm, followed by one positive and one negative  $\pi_{nb} \rightarrow \pi^*$  bands at  $\sim 205$  and  $190$  nm, respectively. Such an ECD pattern has been denoted as class B spectra (see Table 2 for all the ECD patterns found in this work).

Such findings were then challenged by the theoretical work by Sathyanarayana and Applequist<sup>36</sup> (still at the semi-empirical level based on DeVoe theory, a classical approach of dispersive oscillators with complex frequency-dependent polarizabilities), focusing only on the  $\pi_{nb} \rightarrow \pi^*$  (NV1) transition, thus omitting the  $n \rightarrow \pi^*$  transitions.<sup>33</sup> Indeed, their computed ECD spectra of Ace-X-X-NHMe peptides (with X = Ala, and Gly) showed, on the one hand, that types I and III share similar spectra with a

negative band ( $\sim 205$  nm) and a positive band ( $\sim 190$  nm), but that, on the contrary, a mirror spectrum (with a maximum at 205 nm and a minimum at 190 nm) was found for type II.

The available experimental ECD reference spectra for  $\beta$ -turns were derived using various approaches. The first one consists of subtracting the factored  $\alpha$ -helix,  $\beta$ -sheet and random coil contributions from the globular protein circular dichroism spectra.<sup>37,38</sup> This procedure leads to a global ECD  $\beta$ -turn spectrum, which does not take into account the various possible types. Thus, it remains challenging to correlate this obtained ECD spectrum with a precise  $\beta$ -turn conformation.

A second approach, based on a comparative study of theoretical and experimental peptide ECD spectra, was performed by Brahms and collaborators in 1977.<sup>39</sup> Using a poly(Ala<sub>2</sub>-Gly<sub>2</sub>) peptide, they obtained a single ECD curve similar to the ones calculated by Woody<sup>25</sup> and thus considered it as representative of both types I (or III) and II. However, no physical evidence has been shown to indicate which type of  $\beta$ -turn(s) may be present in this peptide, although the authors stated that a type II turn should be expected.<sup>40</sup>

The third and most used approach relies on the comparative structural studies of peptides by ECD, NMR and X-ray crystallography techniques.<sup>28,29,41–44</sup> In this approach, cyclic or modified amino-acid peptides were used to restrain the flexibility of the molecules. For instance, ECD spectra, composed of a negative band at 227–222 nm and a positive one at 200–205 nm, were observed for cyclic hexapeptides (cyclo(Gly/Orn-L-Pro-D-Phe/Cha)<sub>2</sub>) adopting a type II  $\beta$ -turn geometry.<sup>41</sup> Although ECD experiments were not recorded below 190 nm, the obtained pattern resembles class B of Woody's classification (Table 2).

Similarly, Bandekar *et al.* observed<sup>44</sup> a class B ECD spectrum, with a minimum below 188 nm, a maximum at 202–204 nm and another minimum around 228–230 nm, for the cyclo(L-Ala-D-Ala-Aca) peptide forming a type II  $\beta$ -turn. In the same study, a class C pattern (a maximum near 190 nm and two distinct minima at 206–208 and 218–222 nm) was observed for the cyclo(L-Ala-L-Ala-Aca) peptide that exhibits a type I (or III)  $\beta$ -turn.

A few years later, experimental ECD spectra of non-cyclic model peptides (Z-Aib-L-Pro-Aib-L-Pro-OMe (1), Piv-L-Pro-Aib-NHMe (2), Piv-L-Pro-D-Ala-NHMe (3) and Piv-L-Pro-L-Val-NHMe (4)) with various  $\beta$ -turn types (types I, II or III) were reported by Crisma *et al.*<sup>35</sup> They observed two different ECD patterns for type I: a class C for peptide 1 and a class B for peptide 4. These discrepancies led them to assign the class C pattern to a type III structure, rather than type I. Type II (peptides 2 and 3) was characterized by an ECD spectrum having two positive bands at  $\sim 230$  nm and  $\sim 200$  nm, a feature not predicted by Woody's calculations.

It should be emphasized that, in all these studies, the type of turn was characterized in solid-state from X-ray analyses, and that the authors assumed that the same folding also persists in solution. This assumption was based on NMR analyses, performed in solvents different from those used for ECD spectroscopy, and it thus might be arguable.<sup>28,45</sup> As a consequence, the

discrepancies observed in these experimental studies might be linked to the presence of additional conformations in the solvents used for recording ECD spectra.

Interestingly, such ambiguities were not removed by subsequent ECD spectra reported for pseudo-hexapeptides that adopt a conformational mixture of type I and type II  $\beta$ -turns (determined by NMR experiments) in different solvents.<sup>28,29,43,46,47</sup> In one of these studies,<sup>40</sup> two different patterns have been attributed to type I: one presenting a negative band between 225–230 nm and a positive band at  $\sim 195$  nm, and a second characterized by a minimum at  $\sim 205$  nm and a maximum at  $\sim 195$  nm, while one spectrum has been identified for type II featuring a positive band at  $\sim 200$  nm.

It can thus be concluded from this fast literature survey that, because of all of the inconsistencies found in theoretical and experimental works and owing to the fact that ECD is one of the most used experimental methods to study the secondary structures of peptides and proteins, the establishment of reliable reference ECD spectra in solution remains highly desirable.

For such reasons, we investigated in this paper if  $\beta$ -turns have characteristic ECD patterns and if ECD spectroscopy can distinguish between the four principal  $\beta$ -turn types described in solution by NMR spectroscopy,<sup>26</sup> *i.e.* types I, I', II and II'. In particular, we used state-of-the-art advanced quantum chemical computations, based on the time-dependent density functional theory (TD-DFT) formalism, coupled with classical molecular dynamics (MD) simulations on "minimal  $\beta$ -turn models" in aqueous solutions. In order to take into account solvent effects in the computed ECD spectra (both the works of Woody<sup>32</sup> and Sathyanarayana and Applequist<sup>36</sup> were actually performed in a vacuum), we used the polarizable embedding (PE) multiscale scheme,<sup>48,49</sup> as described in the following section.

## 2. Computational details

### 2.1. A minimal peptide model for $\beta$ -turns

Our "minimal  $\beta$ -turn model" can be described as a di-amino acid peptide with N-terminal acetyl and C-terminal N-Me amide capping groups (Ace<sub>T</sub>-X<sub>*i*+1</sub>-X<sub>*i*+2</sub>-NHMe<sub>*i*+3</sub>), as represented in Fig. 1a. For each conformation, the initial "canonical"  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$ , and  $\Psi_{i+2}$  dihedral angles (see Table 1) were set using the UCSF Chimera molecular modelling software.<sup>50</sup>

### 2.2. Classical molecular dynamics simulations

Each  $\beta$ -turn conformation was explored employing restrained classical MD simulations using the GROMACS 5.4.1 software<sup>51</sup> and the OPLS-AA force field.<sup>52</sup> In particular, all MD processes (energy minimization, equilibration, and production) were performed with energy penalty functions consisting of additional flat-bottom quadratic potentials placed on the  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$ , and  $\Psi_{i+2}$  dihedral angles with a force constant set to 10 000 kJ mol<sup>-1</sup> rad<sup>-2</sup> (except for the energy minimization step where it was reduced to 1000 kJ mol<sup>-1</sup> rad<sup>-2</sup>, in order to allow the molecule to reach the conformational minimum more easily, but always constrained to be close to a specific  $\beta$ -turn

conformation) applied when the dihedral angle deviations with respect to the canonical values (Table 1) were larger than  $\pm 10^\circ$ ,  $\pm 20^\circ$  or  $\pm 30^\circ$ .

For each  $\beta$ -turn type, the minimal peptide model was solvated in a water box ( $\sim 3 \times 3 \times 3 \text{ nm}^3$ ) using the TIP3P solvent water model<sup>53</sup> ( $\sim 700$  water molecules) using the tools provided by the GROMACS 5.1.4 simulation package. The solvation box was neutralized with a single counter-ion in the case of charged amino-acids (note that we considered the protonation state of amino acid side chains at physiological pH 7.4), not including any background salt. More precisely, we used  $\text{Cl}^-$  and  $\text{Na}^+$  ions for, respectively, positively and negatively charged amino-acids, described by the `opls_401` and `opls_407` associated atom types.

Once the equilibrium geometry was obtained from the energy minimization step, the temperature of the system was equilibrated in the *NVT* ensemble at 300 K for 1 ns using the velocity rescale thermostat proposed by Bussi and collaborators<sup>54</sup> (with a  $\tau_t = 0.1$  ps time-constant). After this *NVT* equilibration, a pressure equilibration was carried out in the *NPT* ensemble (with  $P = 1$  bar) still using the same modified thermostat<sup>54</sup> and with the Parrinello–Rahman barostat (with a  $\tau_p = 0.1$  ps time-constant).<sup>55</sup> Then, we performed the 10 ns production run, still in the *NPT* ensemble. This relatively short time was, however, found to be adequate due to the restrained degrees of freedom imposed on the dihedral angles  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$ , and  $\Psi_{i+2}$ . From each MD production, the minimal peptide model was centred in the box, and a total of 100 snapshots were extracted every 100 ps for the simulation of absorption and ECD spectra.

### 2.3. Quantum chemistry calculations

Absorption and ECD spectra were simulated by linear-response TD-DFT with the DALTON 2018 quantum chemistry software.<sup>56</sup> The CAM-B3LYP range-separated hybrid exchange–correlation functional<sup>57</sup> was used in conjunction with Pople’s 6-31+G\* and 6-311++G\*\* basis-sets.<sup>58</sup>

Environment effects were taken into account by using the polarizable embedding (PE) multiscale model of Kongsted and co-workers,<sup>48</sup> which was already used in the determination of ECD spectra of stereoisomers of the terrein molecule.<sup>49</sup>

The PE model divides the total system into two sub-systems: a “QM” region (usually the molecule or a part of the molecule containing the chromophore), which is described by a quantum mechanical method (Kohn–Sham (KS)-DFT in our case), and an “MM” region (usually the rest of the molecule and/or the solvent molecules), which is described classically using point charges, polarizabilities (and/or multipoles) that can be polarized by the electron density of the QM part. The effective Hamiltonian operator for the PE model will thus describe the interaction between the two sub-systems and can be written according to:

$$\hat{H}^{\text{eff}} = \hat{v}^{\text{KS}} + \hat{v}^{\text{emb}} = \hat{v}^{\text{KS}} + \hat{v}^{\text{es}} + \hat{v}^{\text{ind}} \quad (1)$$

where  $\hat{v}^{\text{KS}}$  is the conventional vacuum KS operator (defined by a specific DFT exchange–correlation functional), while  $\hat{v}^{\text{emb}}$  is the embedding operator, which is further divided into two

contributions:  $\hat{v}^{\text{es}}$  that describes the electrostatic interaction between fixed permanent charges and the QM region, and  $\hat{v}^{\text{ind}}$  that describes the polarization of the environment (induced dipoles) due to the QM part (which directly depends on the electron density of the former). The quantum-mechanical and classical polarization equations are solved by a self-consistent scheme in which the induced dipoles at each SCF iteration are updated until complete relaxation of both sub-systems.<sup>59</sup>

The PE model is extended to the calculation of excited state properties within the quantum-mechanical response theory<sup>60</sup> and can be used in the simulation of linear (*e.g.* electronic absorption<sup>48</sup> and ECD<sup>49</sup> spectroscopies) and even non-linear (*e.g.* two-<sup>61,62</sup> and three-photon<sup>63</sup> absorption spectroscopies) response properties. Further details about the mathematical form of  $\hat{v}^{\text{es}}$ ,  $\hat{v}^{\text{ind}}$  and their incorporation within the response theory can be found in ref. 48, 59 and 60.

The minimal peptide model (see Fig. 1) was thus treated as the QM region (also called the “core”). However, in order to check the effects of hydrogen bonds between the minimal model and solvent water molecules on the ECD spectra, water molecules within 0 (no water molecules added), 5 and 6 Å (in order to be sure to include all possible H-bonded water molecules) of the centre of mass of the minimal peptide model were also added to the QM sub-system. The remaining water molecules (and ions when present) were added to the polarizable MM region using the standard embedding parameters (charges and isotropic polarizabilities) provided by the “solvent embedding potential” (SEP).<sup>64</sup>

Please note that the impacts of higher-order multipoles and anisotropic dipole–dipole polarizabilities were not taken into account in this study.

For each frame, the potential files were constructed using the PyFraME library.<sup>59,65</sup> Then, for each snapshot, the total simulated absorption spectra ( $\epsilon_{\text{total}}$ , expressed in  $\text{M}^{-1} \text{cm}^{-1}$ ) were created by summation of the normalized Gaussian function ( $\epsilon_{0i}$ ) centred around each individual electronic transition, using the following equation (eqn (5) in ref. 66):

$$\begin{aligned} \epsilon_{\text{total}} &= \sum_{i=1}^n \epsilon_{0i} \\ &= \sum_{i=1}^n \frac{\sqrt{\pi} e^2 N_A}{1000 \ln(10) c^2 m_e A} f_{0i} \exp \left[ - \left( \frac{E - \Delta E_{0i}}{A} \right)^2 \right], \quad (2) \end{aligned}$$

where  $n$  is the total number of computed excited states (here fixed at 20) and  $\epsilon_{0i}$  is the Gaussian broadening function applied to the  $i$ -th transition,  $e$  is the electron charge in esu units,  $N_A$  is the Avogadro number,  $c$  stands for the speed of light in a vacuum (in  $\text{cm s}^{-1}$ ),  $m_e$  is the electron mass (in kg),  $A$  is half the bandwidth at  $1/e$  of peak height expressed in  $\text{cm}^{-1}$  (set to 0.3 eV but expressed in  $\text{cm}^{-1}$ ),  $\Delta E_{0i}$  is the transition energy from the ground state to the  $i$ -th excited state, and  $f_{0i}$  is the corresponding oscillator strength for the  $i$ -th transition.

Similar to the absorption spectra, the total simulated ECD spectrum (molar ellipticity,  $\Delta \epsilon_{\text{total}}$ , expressed in  $\text{M}^{-1} \text{cm}^{-1}$  units) for each snapshot was created by summation of the normalized Gaussian functions ( $\Delta \epsilon_{0i}$ ) placed around each

1 individual electronic transition, using the following equation  
(taken from ref. 67, eqn (8)):

$$\Delta\epsilon_{\text{total}} = \sum_{i=1}^n \Delta\epsilon_{0i} = \sum_{i=1}^n \frac{\Delta E_{0i}}{2.296 \times 10^{-39} \sqrt{\pi} \Delta} \times R_{0i} \exp \left[ - \left( \frac{E - \Delta E_{0i}}{\Delta} \right)^2 \right], \quad (3)$$

10 where, as before,  $n$  is the total number of computed excited states (20) and  $\Delta\epsilon_{0i}$  is the Gaussian broadening function applied to the  $i$ -th ECD transition,  $\Delta E_{0i}$  is the transition energy from the ground state to the  $i$ -th excited state,  $\Delta$  is half of the bandwidth at  $1/e$  peak height expressed in energy units (set to 0.3 eV if not otherwise stated), and  $R_{0i}$  is the corresponding rotatory strength (in  $10^{-40}$  esu<sup>2</sup> cm<sup>2</sup> units) for the  $i$ -th transition (expressed in the velocity form, in order to ensure its origin-independence).

20 Both absorption and ECD spectra were subsequently averaged for the 100 snapshots extracted from the restrained MD simulation.

### 3. Results and discussion

#### 3.1. Protocol validation

25 First, we will discuss the validation of our computational protocol on the minimal model  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  in the type I  $\beta$ -turn conformation. In such cases of aliphatic side-chains, the main chromophoric units are the peptide bonds (see Fig. 1b). There are three bonds of that type in the model peptide, and their interactions between their excited states are expected to be sensitive to different  $\beta$ -turn conformations.

30 However, such sequences,  $\text{Ace}_i\text{-X}_{i+1}\text{-X}_{i+2}\text{-NHMe}_{i+3}$ , are too short peptides to lead to strong structuration in a specific  $\beta$ -turn or even in a bend conformation without suitable modifications<sup>68</sup> as for example, cyclization,<sup>69</sup> chemical design of residues<sup>70,71</sup> and/or capping units in the sequence.<sup>72</sup> Without the use of any restraint, such peptides would be mainly unfolded during the MD simulation, and the  $\beta$ -turn types we would like to investigate would only be scarcely populated. This scenario would imply that only very few snapshots would be informative, and that, in consequence, the simulated averaged ECD spectra would not be statistically relevant. This issue was then circumvented by including dihedral restraints on each  $\Phi$  and  $\Psi$  dihedral angle for the two central residues ( $i+1$  and  $i+2$ ).

45 Although this approach precludes a straightforward thermodynamical study of these minimal  $\beta$ -turn models, it conversely constitutes a prolific source of conformations for each type of  $\beta$ -turn (generated within windows of  $\pm 10^\circ$ ,  $\pm 20^\circ$  or  $\pm 30^\circ$  allowed dihedral angle deviations with respect to the canonical values gathered in Table 1), which were subsequently used in the ECD spectral calculations.

50 Fig. S1–S4 in the ESI<sup>†</sup> report the  $\Phi$  and  $\Psi$  distributions produced by the classical MD simulations for the two central residues ( $i+1$  and  $i+2$ ). They show that the range of allowed angles ( $\pm 10^\circ$  in this case) around the canonical values was

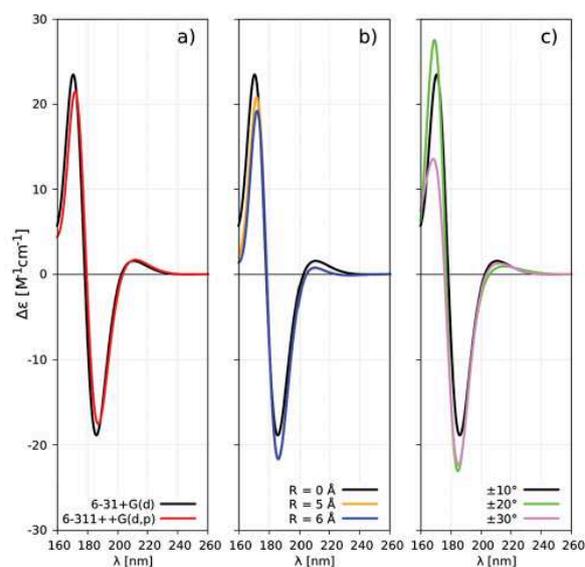


Fig. 2 (a) Averaged ECD spectra of type I  $\beta$ -turn using different basis-sets. (b) Effect of explicit quantum chemical water molecules on the averaged ECD spectra of type I using the 6-31+G\* basis-set. (c) Averaged ECD spectra for type I using various deviations for the  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$  and  $\Psi_{i+2}$  dihedral angles (with respect to the canonical values collected in Table 1). CAM-B3LYP level of theory.

almost uniformly explored, excluding the presence of conformational subgroups within each type of  $\beta$ -turn.

30 In the case of peptides with a charged residue (*i.e.*  $\text{Ace}_i\text{-K}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$ ,  $\text{Ace}_i\text{-A}_{i+1}\text{-K}_{i+2}\text{-NHMe}_{i+3}$ ,  $\text{Ace}_i\text{-D}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  and  $\text{Ace}_i\text{-A}_{i+1}\text{-D}_{i+2}\text{-NHMe}_{i+3}$ ), the system was neutralized by adding a single counter-ion in the simulation box ( $\text{Na}^+$  or  $\text{Cl}^-$  for negative or positive amino acids, respectively). The effect of this simplification was validated in the positively charged  $\text{Ace}_i\text{-K}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  case (type I) by comparing the previous simulations with those neutralized with 150 mM NaCl (a total of 3  $\text{Na}^+$  and 4  $\text{Cl}^-$  ions added) that represents a commonly used salt concentration for molecular dynamics simulations of proteins.<sup>73–75</sup> No statistically significant differences were found between the simulations for the  $\Phi$  and  $\Psi$  dihedral angle distributions as shown in Fig. S5 (ESI<sup>†</sup>), validating the chosen classical MD sampling protocol.

45 Shifting now to the quantum chemistry part, we assessed the basis-set size effect by using two different Pople's basis-sets: 6-31+G\* and the larger 6-311++G\*\* one. As shown in Fig. 2a for the  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  peptide in type I conformation, the less time-consuming 6-31+G\* basis-set can be safely considered as suitable for the semi-quantitative analysis we are aiming at.

50 The coupling between the QM and MM parts also deserves to be scrutinized. In particular, the impact of hydrogen bonds between the peptide model and the solvent molecules on average ECD spectra was checked including the closest water molecules in the QM part of the QM/MM system, since this radius parameter was shown in some cases to be critical.<sup>76</sup> The selection of these water molecules was based on the distance to

1 the peptide centre of mass: from 0 (no solvent molecules in the  
 QM-part), up to 5 and 6 Å. Due to timing reasons (the inclusion  
 of several water molecules in the QM region considerably  
 increases the computing time), this test was carried out only  
 5 for the type I  $\beta$ -turn conformation. It can be deduced from  
 Fig. 2b that the insertion of explicit solvent molecules in the  
 QM part has no significant impact on the average ECD spectra.  
 Consequently, all water molecules were then treated as part of  
 the MM sub-system.

10 Fig. 2c confirms that the three allowed deviations ( $\pm 10^\circ$ ,  
 $\pm 20^\circ$ , and  $\pm 30^\circ$ ) with respect to canonical dihedral angles  
 afford comparable patterns for the average ECD spectra. In  
 the following, only the  $\pm 10^\circ$  case will thus be further discussed  
 in detail, consistent with the previous study on  $\beta$ -turns by  
 15 Woody.<sup>32</sup>

The impact of the force field choice on such results also  
 deserves to be considered. The choice of OPLS-AA was driven by  
 two main reasons. First, it includes a balanced representation  
 for a variety of liquids as well as for peptides. The OPLS-AA  
 20 ability to treat or parametrize different solvents is also notably  
 in line with a future project to extend this first study to more  
 realistic systems involving mixed solvent environments as often  
 used in experimental ECD studies. For example, in order to  
 induce a secondary structure, a water:2,2,2-trifluoroethanol  
 25 (TFE) mixture is usually considered, and OPLS-AA will be  
 adequate for this. Secondly, it generally presents less secondary  
 structure biases than other force fields, as for instance recently  
 shown by Cruz and coworkers.<sup>77</sup>

In order to assess the influence of this FF choice, we  
 30 computed averaged ECD spectra using the CHARMM-27 FF<sup>78</sup>  
 for the  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  minimal model. Fig. S9 in the  
 ESI† shows that there are no significant differences between the  
 two profiles, letting us be confident about the transferability of  
 our conclusions to other popular force fields.

35 The last computational parameter we will discuss is related  
 to a more refined description of local field effects due to the  
 presence of the environment. To this aim, we used the effective  
 external field (EEF) extension<sup>62</sup> available in the RESPONSE  
 module in DALTON. Fig. S10 in the ESI† shows that the EEF  
 40 extension only affected the intensities of the average ECD  
 spectra (this screening effect due to the solvent molecules is  
 also expected to be similar for other secondary structures)  
 without altering the energy positions. More importantly, the  
 final pattern was unchanged, and even the ratio between the  
 45 minimal and maximal  $\Delta\epsilon$  values was conserved.

From our point of view, all previous comparisons contribute  
 to validating the choice of our computational protocol.

### 3.2. Electronic circular dichroism spectra

50  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  in types I, I', II or II' will be once more  
 our working case for the analysis of spectroscopic properties.  
 The corresponding average ECD spectra are plotted in Fig. 3. In  
 the four cases, there are two ECD bands with similar intensities  
 and opposite signs approximately around 180 nm. In particular,  
 55 types I and II' possess a common pattern: the first band at

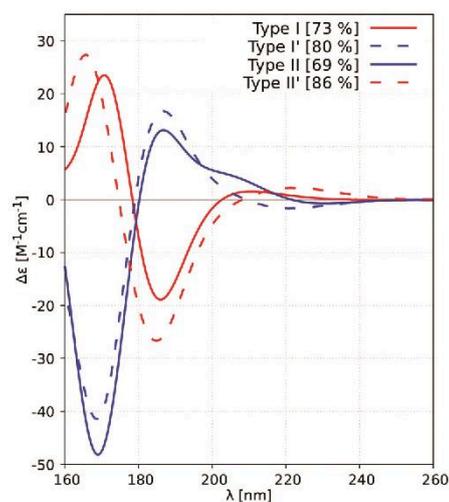


Fig. 3  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  averaged ECD spectra for the four selected  $\beta$ -turn types (types I, I', II and II'). CAM-B3LYP/6-31+G\* level of theory.

minor energies (*i.e.* high wavelengths) is negative in sign followed by a positive one at higher energies.

On the other hand, types I' and II reveal the opposite pattern: the first band at lower energies is positive in sign followed by a negative one at higher energies. It should be noticed that such a discrimination is not possible from the simple absorption spectra (see Fig. S6 in ESI†), since all types present a common pattern: a very weak band ( $\epsilon \sim 500 \text{ M}^{-1} \text{ cm}^{-1}$ ) at 220 nm, corresponding to an electric dipole-forbidden  $n \rightarrow \pi^*$  transition, followed by an intense band ( $\epsilon > 20\,000 \text{ M}^{-1} \text{ cm}^{-1}$ ) at  $\sim 175$  nm, ascribable to a  $\pi_{\text{nb}} \rightarrow \pi^*$  (NV1) electric dipole-allowed transition. This makes absorption spectroscopy fully inadequate for  $\beta$ -turn characterization.

Then, to deepen this shape analysis, we calculated the ratio ( $\kappa$ , unitless) between the negative and positive ECD band intensities, as well as the distance ( $\Delta\lambda$ , in nm) between the corresponding extrema, as defined by eqn (4) and (5):

$$\Delta\lambda = \arg \min_{\lambda}(\Delta\epsilon(\lambda)) - \arg \max_{\lambda}(\Delta\epsilon(\lambda)) \quad (4)$$

$$\kappa = \min(\Delta\epsilon(\lambda))/\max(\Delta\epsilon(\lambda)) \quad (5)$$

For example, for  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  in type I conformation, the ratio between the first negative and the second positive ECD bands is distributed around 1.0 (see Fig. S7 in ESI†), and  $\Delta\lambda$  between the positive and negative ECD band maxima shows a bi-modal distribution at  $\sim 15$  nm (majority) and another one at  $\sim -15$  nm (see Fig. S8 in ESI†). However, we noticed that the first mode (corresponding to positive values) is prevalently populated. Note that the percentage of ECD spectra that respect the “first negative”, types I/II', or “first positive”, types I'/II, is indicated in Fig. 3. Similar conclusions were obtained for the other  $\beta$ -turn conformations (*i.e.* types I', II and II').

We then inspected the impact of substitution in these “first negative/positive” patterns by modifying the amino-acids at the

1 **Table 3** Amino-acid classification based on the physical–chemical properties of their side-chains. Amino-acids in bold indicate the representative member chosen for the present study

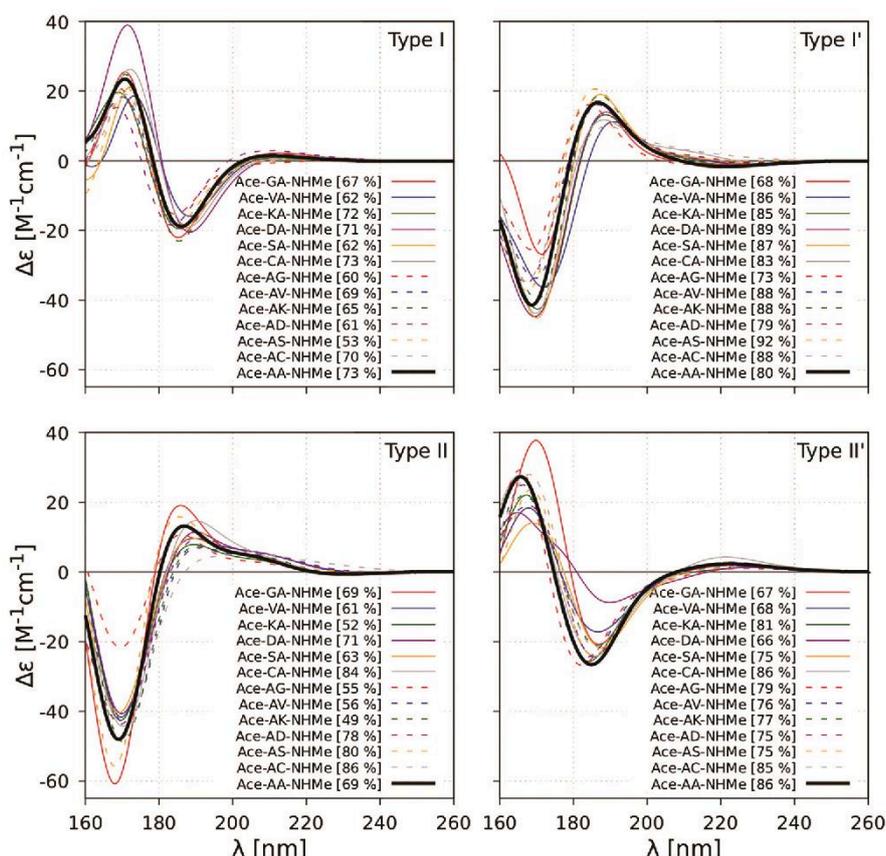
Amino-acid family	Components
5 Hydrophobic	<b>Ala, Val</b> , Ile, Leu, and Met
Polar uncharged	<b>Ser</b> , Thr, Asn, and Gln
Positively charged	<b>Lys</b> , Arg, and His
Negatively charged	<b>Asp</b> , and Glu
Special cases	<b>Cys, Gly</b> , and Pro
10 Aromatic	<b>Phe</b> , Tyr, and Trp

$i + 1$  or  $i + 2$  positions, leading to the following generic sequence:  $\text{Ace}_i\text{-X}_{i+1}\text{-X}_{i+2}\text{-NHMe}_{i+3}$ . Amino-acids are usually classified into six classes:<sup>79</sup> hydrophobic, polar uncharged, positively charged, aromatic, and special cases. We thus chose one or two representative amino-acids for each of these subclasses, as detailed in Table 3. For each of them, the influence of the substitution position was investigated. For instance, we considered both  $\text{Ace}_i\text{-G}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  and  $\text{Ace}_i\text{-A}_{i+1}\text{-G}_{i+2}\text{-NHMe}_{i+3}$  (where G denotes glycine). In total, 15 additional peptides were simulated for each of the four retained  $\beta$ -turn structures, resulting in 60 computed spectra.

In a first instance, we considered non-aromatic side chains. From Fig. 4, it can be inferred that non-aromatic amino-acids have an effect on the energetic position and on the intensity of the two opposite-in-sign ECD bands but that, globally, the patterns we found previously with the  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  minimal model for each kind of  $\beta$ -turn were unchanged.

Fig. 4 also indicates, for each sequence, the percentage of frames whose ECD spectra fulfil either the “first negative” (types I and II’) or the “first positive” (types I’ and II) patterns previously identified. The mean value of respected averaged ECD patterns calculated for the 13 sequences is 66%, 83%, 67% and 78% for types I, I’, II and II’, respectively. For type II, only 9 sequences respect the mean pattern with percentages greater than 60%. However, the fact that the average ECD spectra for a  $\beta$ -turn conformation calculated using different sequences converge to the same pattern with highly superimposable profiles strongly supports the two ECD patterns found with the minimal peptide model  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  (“first negative” and “first positive”).

Consequently, non-aromatic side chains do not seem to have a relevant impact on the averaged ECD spectra. This conclusion was actually already stated by Woody,<sup>80</sup> who underlined that among



55 **Fig. 4** Effect of residue substitutions in position  $i + 1$  (straight lines) or  $i + 2$  (dotted lines) on averaged ECD spectra compared to the  $\text{Ace}_i\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  peptide (black lines) for each amino-acid family (see Table 2). CAM-B3LYP/6-31+G\* level of theory.

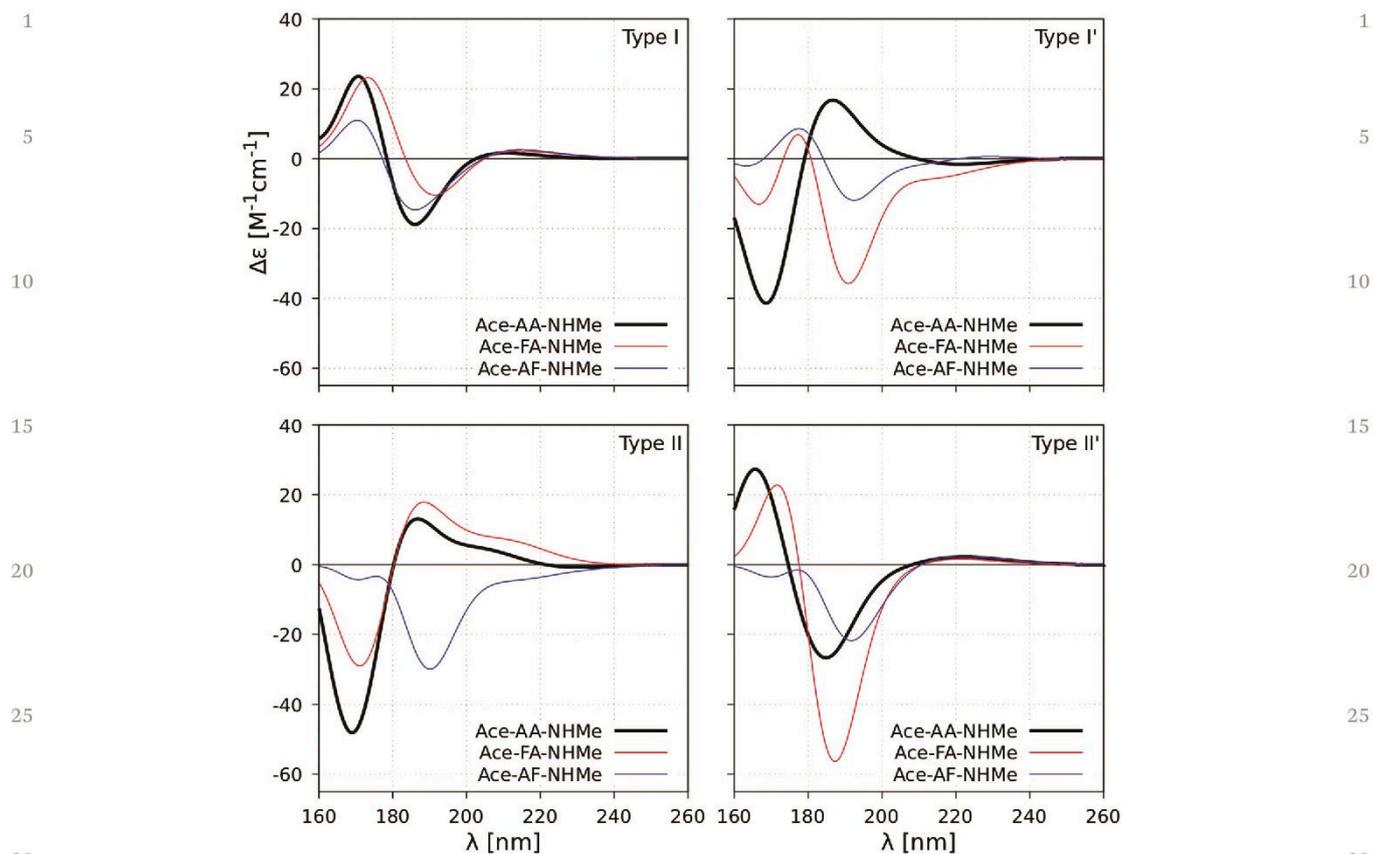


Fig. 5 Effect of an aromatic residue (Phe) substitution in position  $i + 1$  (red lines) or  $i + 2$  (blue lines) on averaged ECD spectra compared to  $\text{Ace}_7\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  ( $\text{Ace}_7\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  (black lines)). CAM-B3LYP/6-31+G\* level of theory.

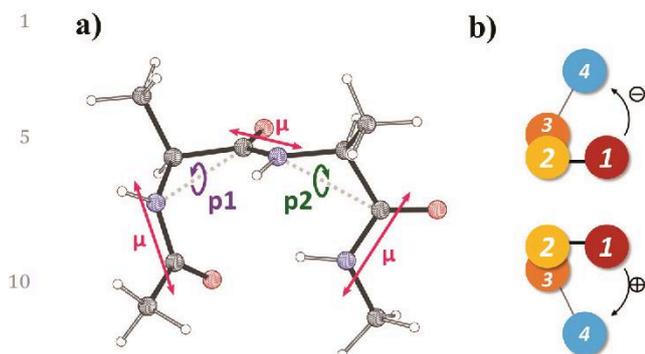
the amino acids with chromophoric side-chains (*i.e.* phenylalanine, tyrosine, tryptophan, disulfides, histidine, arginine, glutamine, asparagine, glutamic acid, and aspartic acid), only the aromatic and disulfide ones have electric or magnetic dipole transition moments large enough to make appreciable spectral contributions.<sup>80</sup>

In contrast, as stated before, amino-acids with aromatic side-chains (represented here by phenylalanine, F) greatly perturb the patterns found with the  $\text{Ace}_7\text{-A}_{i+1}\text{-A}_{i+2}\text{-NHMe}_{i+3}$  minimal peptide model, as illustrated by Fig. 5. This perturbation is in fact caused by the presence of other chromophores in the system with strong electric transition moments (phenyl in the case of phenylalanine). Similar perturbations have already been studied and characterized in several experimental and theoretical works.<sup>41,81–84</sup> Indeed, such a perturbing chromophore is spatially quite close to the three peptide bonds and, at the same time, its excited states ( ${}^1\text{L}_a$  and the degenerate states  ${}^1\text{B}_b$  and  ${}^1\text{B}_a$ ) are energetically close to the  $\pi_{\text{nb}} \rightarrow \pi^*$  transitions of the amide groups responsible for the exciton interactions.<sup>85</sup> As a consequence, these perturbations do not allow the identification of the previously found “first positive/negative” patterns.

This important shortcoming limits the validity of “first positive/negative” patterns as a unique reference for type I, I',

II and II'  $\beta$ -turns when at least one aromatic amino acid is included in the sequence of a short peptide. In such cases, ECD spectroscopy should certainly be complemented by other spectroscopic methods to univocally determine the conformation of the involved  $\beta$ -turn structures. However, it should be stressed that aromatic amino-acids (Phe, Tyr and Trp) are very rarely found (with a probability less than 1%) in position  $i + 1$  and/or  $i + 2$  for these types of  $\beta$ -turns,<sup>86</sup> so that this limitation will be fully ineffective in most cases.

Finally, even if we are not dealing with chirality determinations (all natural amino-acids are indeed  $\text{L}$ -amino-acids), we qualitatively tried to explain the origin of the “first negative” (characteristic for types I and II') and “first positive” (characteristic for type I' and II) ECD patterns by means of the non-empirical exciton chirality rule (ECR).<sup>87</sup> For this purpose, we made three strong approximations: (i) the peptide bond is a suitable chromophore for ECR (usually, chromophores with strong  $\pi_{\text{nb}} \rightarrow \pi^*$  absorption, as benzoate derivatives, are recommended for ECR<sup>88</sup>); (ii) the direction of the electric transition dipole moment (TDM) for the  $\pi_{\text{nb}} \rightarrow \pi^*$  ( $\text{NV}_1$ ) transition for each of the three peptide bonds is assumed to be parallel to the peptide bond and co-planar to the peptide group<sup>89</sup> (see Fig. 6a); (iii) only interactions and relative



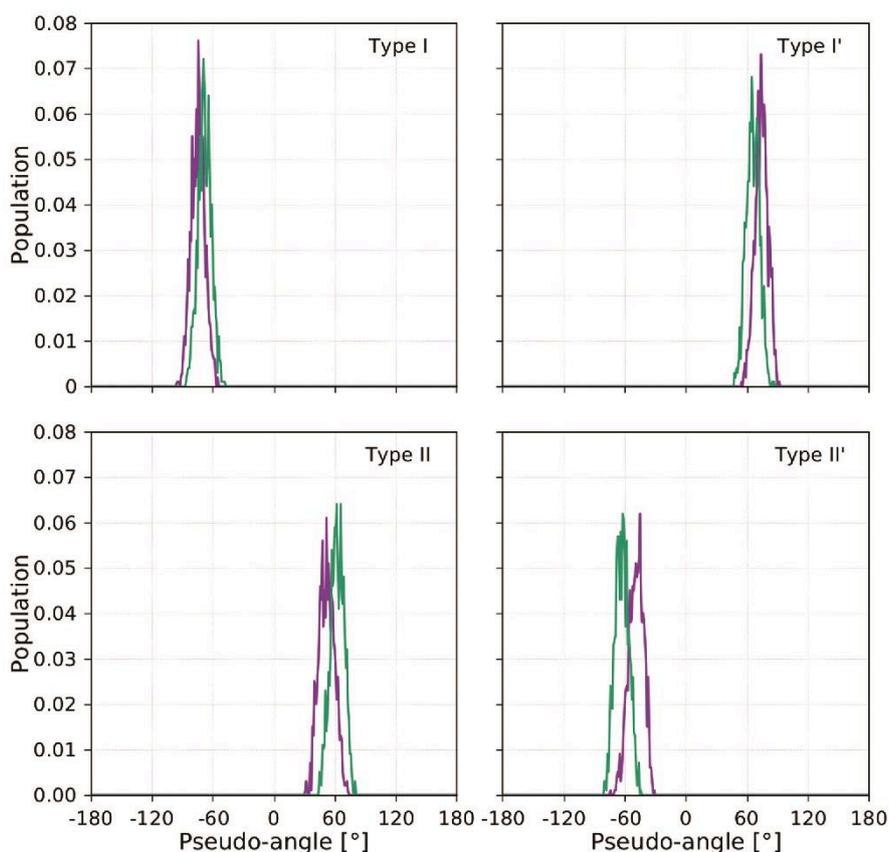
**Fig. 6** (a) Minimal  $\beta$ -turn model with the electric transition dipole moment (TDM) for the  $\pi_{nb} \rightarrow \pi^*$  ( $NV_1$ ) transition, approximated to lie along the C–N peptide bond (H in white, C in grey, N in blue, and O in red).  $p1$  ( $C_i-N_{i+1}-C_{i+1}-N_{i+2}$ ) and  $p2$  ( $C_{i+1}-N_{i+2}-C_{i+2}-N_{i+3}$ ) are the pseudo-dihedral angles between consecutive peptide bonds. (b) Definitions of negative (counterclockwise) and positive (clockwise twist) dihedral angles between atoms 1, 2, 3 and 4.

orientations between the electric TDMs between two consecutive chromophores are considered.

This last point implies that we have to analyse the relative orientations between the first and the second residues, and between the second and the third amide chromophores, through the definition of pseudo-dihedral angles “ $p1$ ” and “ $p2$ ”, represented in Fig. 6a. It also follows from this assumption that the interaction between the first and third peptide bonds is not considered. This can be crudely justified by observing that these bonds are far away, and by recalling that the exciton splitting intensity  $A$  is inversely proportional to  $R_{ij}^2$ , where  $R_{ij}$  denotes the interchromophoric vector between chromophores  $i$  and  $j$ . More specifically, the amplitude of the exciton ECD couplet  $A$  (defined as the difference between the first ECD band at lower energies and the second ECD band at higher energies, *i.e.*  $\Delta\epsilon_1 - \Delta\epsilon_2$ ) reads:<sup>87,90</sup>

$$A \propto R_{ij}^{-2} (\boldsymbol{\mu}_{i0a} \times \boldsymbol{\mu}_{j0a}) \cdot \mathbf{V}_{ij} = D_{i0a} D_{j0a} R_{ij}^{-2} \mathbf{e}_{ij} \cdot (\mathbf{e}_i \times \mathbf{e}_j) \{ \mathbf{e}_i \cdot \mathbf{e}_j - 3(\mathbf{e}_i \cdot \mathbf{e}_{ij})(\mathbf{e}_j \cdot \mathbf{e}_{ij}) \}, \quad (6)$$

where  $\boldsymbol{\mu}_{i0a}$  and  $\boldsymbol{\mu}_{j0a}$  are the electric transition dipole moments,  $D_{i0a}$  and  $D_{j0a}$  denote the electric dipole strength (squares of  $\boldsymbol{\mu}_{i0a}$  and  $\boldsymbol{\mu}_{j0a}$ ) of chromophores  $i$  and  $j$  for the transition  $0 \rightarrow a$  (the  $\pi_{nb} \rightarrow \pi^*$   $NV_1$  transition in this case), and  $\mathbf{e}_{ij}$ ,  $\mathbf{e}_i$  and  $\mathbf{e}_j$  are the unit vectors of  $R_{ij}$ ,  $\boldsymbol{\mu}_{i0a}$  and  $\boldsymbol{\mu}_{j0a}$ , respectively.



**Fig. 7** Distribution of pseudo-angles “ $p1$ ” (green) and “ $p2$ ” (violet) over the molecular dynamics production for  $\text{Ace-}A_{i+1}\text{-}A_{i+2}\text{-NHMe}_{i+3}$ . OPLS-AA and TIP3P force fields used.

1 Considering type I and type II' (belonging to the "first  
negative" pattern), we found (see Fig. 7) that the values for  
the pseudo-dihedral angles between consecutive amide bonds  
5  $p1$  and  $p2$  are both negative (see Fig. 6b for the adopted sign  
convention) along the whole restrained MD simulation. If the  
electric TDMs (for the  $\pi_{nb} \rightarrow \pi^*$  transitions) are assumed to lie  
along the peptide bonds, a clear correspondence between the  
sign of these dihedral angles and the sign of the first ECD band  
10 at lower energies in the region of the  $\pi_{nb} \rightarrow \pi^*$  transition is  
observed. On the other hand, if we consider type I' and type II  
("first positive" pattern), the  $p1$  and  $p2$  values are both positive  
(Fig. 7), a fact that is reflected in the sign of the first (positive)  
ECD band at lower energies.

15 These observations can be qualitatively rationalized by the  
ECM<sup>87</sup> that states that the sign of the first band at lower energy  
is determined by the dihedral angles between the electric TDMs  
between two interacting chromophores. We have thus  
"negative-exciton chirality" when the dihedral angle between  
20 two electric TDMs is negative, resulting in a negative ECD band  
at lower energies followed by a positive band at higher energies  
of similar intensity (pattern for types I and II'), while there is  
"positive-exciton chirality" when the dihedral angle between  
two electric TDMs is positive, resulting in a positive ECD band  
25 at longer wavelengths followed by a negative band at higher  
energies of similar intensity (patterns for types I' and II).

## 4. Conclusions

30 In this paper, state-of-the-art time-dependent density func-  
tional theory calculations coupled with the polarizable embed-  
ding model and with molecular dynamics simulations showed that  
two different electric circular dichroism patterns can be  
identified within four main  $\beta$ -turn types (namely I, I', II and II')  
35 in solution when no aromatic amino-acids are present in the  
peptide sequence. In particular, a "first negative" pattern can  
be associated with types I and II', while a "first positive"  
pattern is associated with types I' and II. This observation is  
in contrast with the seminal work by Woody<sup>32</sup> (in which types I  
40 and II presented the same pattern, coined class "B"), but in  
agreement with the work by Sathyanarayana and Applequist,<sup>36</sup>  
which are both semi-empirical calculations. Besides, these  
patterns were qualitatively rationalized by the exciton chirality  
method. These results thus pave the way towards a better  
45 theoretical and experimental discrimination of  $\beta$ -turns in  
peptides.

## Conflicts of interest



## Acknowledgements

55 We gratefully acknowledge LABEX SynOrg (ANR-11-LABX-0029),  
the Région Normandie and the European Union for funding. The  
"Centre Régional Informatique et d'Applications Numériques de

Normandie – CRIANN" (projects no. 2018019 and 2018012) and  
"Institut du Développement et des Ressources en Informatique  
Scientifique - IDRIS" (DARI project no. 100825) are thanked for  
computational support. A. B. warmly thanks Prof. Jacob Kongsted  
5 and Peter Reinholdt (University of Southern Denmark, SDU, Den-  
mark) for initiating him to the polarizable embedding model. M.  
B. acknowledges the support of "Initiative d'Excellence" program  
from the French State (Grants "DYNAMO", ANR-11-LABX-0011,  
and "CACISICE", ANR-11-EQPX-0008).

## References

- 1 C. M. Venkatachalam, Stereochemical criteria for polypep-  
tides and proteins. V. Conformation of a system of three  
15 linked peptide units, *Biopolymers*, 1968, **6**, 1425–1436.
- 2 J. S. Richardson, *Advances in Protein Chemistry*, Academic  
Press, 1981, vol. 34, pp. 167–339.
- 3 P. N. Lewis, F. A. Momany and H. A. Scheraga, Folding of  
Polypeptide Chains in Proteins: A Proposed Mechanism for  
20 Folding, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **68**, 2293–2297.
- 4 I. D. Kuntz, Protein folding, *J. Am. Chem. Soc.*, 1972, **94**, 4009–4012.
- 5 J. L. Crawford, W. N. Lipscomb and C. G. Schellman, The  
Reverse Turn as a Polypeptide Conformation in Globular  
Proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 1973, **70**, 538–542. 25
- 6 P. Y. Chou and G. D. Fasman,  $\beta$ -turns in proteins, *J. Mol.*  
*Biol.*, 1977, **115**, 135–175.
- 7 G. D. Rose, L. M. Gierasch and J. A. Smith, *Adv. Protein*  
*Chem.*, 1985, **37**, 1–109.
- 8 L. L. Videau, W. B. Arendall and J. S. Richardson, The *cis*-Pro  
30 touch-turn: A rare motif preferred at functional sites, *Pro-*  
*teins: Struct., Funct., Genet.*, 2004, **56**, 298–309.
- 9 J. D. A. Tyndall, B. Pfeiffer, G. Abbenante and D. P. Fairlie,  
Over One Hundred Peptide-Activated G Protein-Coupled  
Receptors Recognize Ligands with Turn Structure, *Chem.*  
35 *Rev.*, 2005, **105**, 793–826.
- 10 G. Ruiz-Gómez, J. D. A. Tyndall, B. Pfeiffer, G. Abbenante  
and D. P. Fairlie, Update 1 of: Over One Hundred Peptide-  
Activated G Protein-Coupled Receptors Recognize Ligands  
40 with Turn Structure, *Chem. Rev.*, 2010, **110**, PR1–PR41.
- 11 M. E. Johnson and M. Kahn, in *Biotechnology and Pharmacy*,  
ed. J. M. Pezzuto, M. E. Johnson and H. R. Manasse,  
Springer, Netherlands, Dordrecht, 1993, pp. 366–378.
- 12 P. Grieco, A. Carotenuto, R. Patacchini, C. A. Maggi,  
E. Novellino and P. Rovero, Design, Synthesis, Conforma-  
45 tional Analysis, and Biological Studies of Urotensin-II Lac-  
tam Analogues, *Bioorg. Med. Chem.*, 2002, **10**, 3731–3739.
- 13 P. N. Lewis, F. A. Momany and H. A. Scheraga, Chain  
reversals in proteins, *Biochim. Biophys. Acta, Protein Struct.*,  
1973, **303**, 211–229. 50
- 14 R. Huber and W. Steigemann, Two *cis*-prolines in the Bence-  
Jones protein Rei and the *cis*-pro-bend, *FEBS Lett.*, 1974, **48**,  
235–237.
- 15 C. M. Wilmot and J. M. Thornton, Analysis and prediction of  
the different types of  $\beta$ -turn in proteins, *J. Mol. Biol.*, 1988,  
55 **203**, 221–232.

- 1 16 E. G. Hutchinson and J. M. Thornton, A revised set of potentials for  $\beta$ -turn formation in proteins, *Protein Sci.*, 1994, **3**, 2207–2216.
- 17 K.-C. Chou, Prediction of Tight Turns and Their Types in Proteins, *Anal. Biochem.*, 2000, **286**, 1–16.
- 5 18 A. G. De Brevern, Extension of the classical classification of  $\beta$ -turns, *Sci. Rep.*, 2016, **6**, 33191.
- 19 M. Shapovalov, S. Vucetic and R. L. Dunbrack, A new clustering and nomenclature for beta turns derived from high-resolution protein structures, *PLoS Comput. Biol.*, 2019, **15**, e1006844.
- 10 20 G. Holzwarth and P. Doty, The Ultraviolet Circular Dichroism of Polypeptides 1, *J. Am. Chem. Soc.*, 1965, **87**, 218–228.
- 21 R. Townend, T. F. Kumosinski, S. N. Timasheff, G. D. Fasman and B. Davidson, The circular dichroism of the  $\beta$  structure of poly-L-lysine, *Biochem. Biophys. Res. Commun.*, 1966, **23**, 163–169.
- 15 22 J. Reed and T. A. Reed, A Set of Constructed Type Spectra for the Practical Estimation of Peptide Secondary Structure from Circular Dichroism, *Anal. Biochem.*, 1997, **254**, 36–40.
- 20 23 P. K. Sarkar and P. Doty, The optical rotatory properties of the beta-configuration in polypeptides and proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 1966, **55**, 981–989.
- 24 A. McPherson, *Crystallization of Biological Macromolecules*, Cold Spring Harbor Laboratory Press, 1999.
- 25 25 M. W. Parker, Protein Structure from X-Ray Diffraction, *J. Biol. Phys.*, 2003, **29**, 341–362.
- 26 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, 1986.
- 27 K. D. Kopple, T. J. Schamper and A. Go, Conformation of cyclic peptides. VIII. Cyclic hexapeptides containing the L-pro-D-phe sequence, *J. Am. Chem. Soc.*, 1974, **96**, 2597–2605.
- 30 28 A. Perczel, M. Hollósi, B. M. Foxman, G. D. Fasman and G. D. Fasman, Conformational Analysis of Pseudocyclic Hexapeptides Based on Quantitative Circular Dichroism (CD), NOE, and X-ray Data. The Pure CD Spectra of Type I and Type II  $\beta$ -Turns, *J. Am. Chem. Soc.*, 1991, **113**, 9772–9784.
- 35 29 L. M. Gierasch, C. M. Deber, V. Madison, C.-H. Niu and E. R. Blout, Conformations of (X-L-Pro-Y)<sub>2</sub> cyclic hexapeptides. Preferred beta-turn conformers and implications for beta-turns in proteins, *Biochemistry*, 1981, **20**, 4730–4738.
- 40 30 L. Whitmore and B. A. Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases, *Biopolymers*, 2008, **89**, 392–400.
- 45 31 N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc.*, 2006, **1**, 2876–2890.
- 32 R. W. Woody, in *Peptides, Polypeptides, and Proteins*, ed. E. R. Blout, F. A. Bovey, M. Goodman and N. Lotan, John Wiley & Sons, New York, 1974, pp. 338–350.
- 50 33 R. W. Woody, in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G. D. Fasman, 1996, pp. 25–67.
- 34 P. M. Bayley, E. B. Nielsen and J. A. Schellman, Rotatory properties of molecules containing two peptide groups: theory, *J. Phys. Chem.*, 1969, **73**, 228–243.
- 55 35 R. W. Woody, Improved Calculation of the  $n\pi^*$  Rotational Strength in Polypeptides, *J. Chem. Phys.*, 1968, **49**, 4797–4806.
- 36 B. K. Sathyanarayana and J. Applequist, Theoretical  $\pi-\pi^*$  absorption and circular dichroic spectra of  $\beta$ -turn model peptides, *Int. J. Pept. Protein Res.*, 1986, **27**, 86–94.
- 5 37 C. T. Chang, C. S. C. Wu and J. T. Yang, Circular dichroic analysis of protein conformation: Inclusion of the  $\beta$ -turns, *Anal. Biochem.*, 1978, **91**, 13–31.
- 10 38 L. A. Compton and W. C. Johnson, Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication, *Anal. Biochem.*, 1986, **155**, 155–167.
- 15 39 S. Brahms, J. Brahms, G. Spach and A. Brack, Identification of  $\beta$ -turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism, *Proc. Natl. Acad. Sci. U. S. A.*, 1977, **74**, 3208–3212.
- 20 40 S. Brahms and J. Brahms, Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism, *J. Mol. Biol.*, 1980, **138**, 149–178.
- 41 C. A. Bush, S. K. Sarkar and K. D. Kopple, Circular Dichroism of  $\beta$  Turns in Peptides and Proteins, *Biochemistry*, 1978, **17**, 4951–4954.
- 25 42 M. Crisma, G. D. Fasman, H. Balaram and P. Balaram, Peptide models for  $\beta$ -turns. A circular dichroism study, *Int. J. Pept. Protein Res.*, 1984, **23**, 411–419.
- 43 M. Hollósi, K. E. Kövér, S. Holly, L. Radics and G. D. Fasman,  $\beta$ -Turns in bridged proline-containing cyclic peptide models, *Biopolymers*, 1987, **26**, 1555–1572.
- 30 44 J. Bandekar, D. J. Evans, S. Krimm, S. J. Leach, S. Lee, J. R. McQuie, E. Minasian, G. Nemethy, M. S. Pottle, H. A. Scheraga, E. R. Stimson and R. W. Woody, Conformations of cyclo(L-alanyl-L-alanyl- $\epsilon$ -aminocaproyl) and of cyclo(L-alanyl-D-alanyl- $\epsilon$ -aminocaproyl); cyclized dipeptide models for specific types of  $\beta$ -bends, *Int. J. Pept. Protein Res.*, 2009, **19**, 187–205.
- 35 45 B. Drouillard, C. Peggion, B. Biondi, K. Wright, F. Couty, M. Crisma, F. Formaggio and C. Toniolo, Heterochiral Ala/ ( $\alpha$ Me)Aze sequential oligopeptides: Synthesis and conformational study, *J. Pept. Sci.*, 2019, **25**, e3165.
- 40 46 A. Perczel, M. Hollósi, G. Tusnády and G. D. Fasman, Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins, *Protein Eng.*, 1991, **4**, 669–679.
- 45 47 A. Perczel and G. D. Fasman, Quantitative analysis of cyclic  $\beta$ -turn models, *Protein Sci.*, 1992, **1**, 378–395.
- 48 J. M. Olsen, K. Aidas and J. Kongsted, Excited states in solution through polarizable embedding, *J. Chem. Theory Comput.*, 2010, **6**, 3721–3734.
- 50 49 M. S. Nørby, J. M. H. Olsen, C. Steinmann and J. Kongsted, Modeling Electronic Circular Dichroism within the Polarizable Embedding Approach, *J. Chem. Theory Comput.*, 2017, **13**, 4442–4451.
- 55 50 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

- 1 51 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
- 5 52 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.
- 10 53 P. Mark and L. Nilsson, Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.
- 54 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**, 014101.
- 15 55 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 56 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenaes, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjaergaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnaes, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawłowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Salek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Snegov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski and H. Ågren, The Dalton quantum chemistry program system, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 269–284.
- 40 57 T. Yanai, D. P. Tew and N. C. Handy, A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP), *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- 45 58 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions, *J. Chem. Phys.*, 1980, **72**, 650–654.
- 59 C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted and J. M. H. Olsen, Response properties of embedded molecules through the polarizable embedding model, *Int. J. Quantum Chem.*, 2018, **119**, e25717.
- 50 60 J. M. H. Olsen and J. Kongsted, *Adv. Quantum Chem.*, 2011, 107–143.
- 61 L. J. Nåbo, N. H. List, C. Steinmann and J. Kongsted, Computational Approach to Evaluation of Optical Properties of Membrane Probes, *J. Chem. Theory Comput.*, 2017, **13**, 719–726.
- 62 N. H. List, H. J. A. Jensen and J. Kongsted, Local electric fields and molecular properties in heterogeneous environments through polarizable embedding, *Phys. Chem. Chem. Phys.*, 2016, **18**, 10070–10080.
- 5 63 A. Bonvicini, P. Reinholdt, V. Tognetti, L. Joubert, D. Wüster and J. Kongsted, Rational design of novel fluorescent analogues of cholesterol: a “step-by-step” computational study, *Phys. Chem. Chem. Phys.*, 2019, **21**, 15487–15503.
- 64 M. T. P. Beerepoot, A. H. Steindal, N. H. List, J. Kongsted and J. M. H. Olsen, Averaged Solvent Embedding Potential Parameters for Multiscale Modeling of Molecular Properties, *J. Chem. Theory Comput.*, 2016, **12**, 1684–1695.
- 10 65 J. M. H. Olsen, *PyFraME: Python tools for Fragment-based Multiscale Embedding calculations (version 0.1.0)*, DOI: 10.5281/ZENODO.293765.
- 15 66 Gaussian Inc, Creating UV/VisiblePlots from the Results of Excited States Calculations.
- 67 P. J. Stephens and N. Harada, ECD cotton effect approximated by the Gaussian curve and other methods, *Chirality*, 2009, **22**, 229–233.
- 20 68 M. S. Searle, D. H. Williams and L. C. Packman, A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable non-native  $\beta$ -hairpin, *Nat. Struct. Biol.*, 1995, **2**, 999–1006.
- 25 69 Y. Mirassou, C. M. Santiveri, M. J. Pérez de Vega, R. González-Muñiz and M. A. Jiménez, Disulfide Bonds versus Trp · · Trp Pairs in Irregular  $\beta$ -Hairpins: NMR Structure of Vammin Loop 3-Derived Peptides as a Case Study, *ChemBioChem*, 2009, **10**, 902–910.
- 30 70 T. Tancredi, A. Pastore, S. Salvadori, V. Esposito and P. A. Temussi, Interaction of sweet proteins with their receptor, *Eur. J. Biochem.*, 2004, **271**, 2231–2240.
- 71 M. S. Searle, Peptide models of protein  $\beta$ -sheets: design, folding and insights into stabilising weak interactions, *J. Chem. Soc., Perkin Trans. 2*, 2001, 1011–1020.
- 35 72 B. L. Kier, I. Shu, L. A. Eidenschink and N. H. Andersen, Stabilizing capping motif for  $\beta$ -hairpins and sheets, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10466–10471.
- 73 G. T. Ibragimova and R. C. Wade, Importance of Explicit Salt Ions for Protein Stability in Molecular Dynamics Simulation, *Biophys. J.*, 1998, **74**, 2906–2911.
- 40 74 K. El Hage, F. Hédin, P. K. Gupta, M. Meuwly and M. Karplus, Valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size, *eLife*, DOI: 10.7554/eLife.35560. Q7
- 45 75 Y. Li, R. Sun, H. Liu and H. Gong, Molecular dynamics study of ion transport through an open model of voltage-gated sodium channel, *Biochim. Biophys. Acta, Biomembr.*, 1859, **2017**, 879–887.
- 50 76 J. Sjöqvist, M. Linares, K. V. Mikkelsen and P. Norman, QM/MM-MD Simulations of Conjugated Polyelectrolytes: A Study of Luminescent Conjugated Oligothiophenes for Use as Biophysical Probes, *J. Phys. Chem. A*, 2014, **118**, 3419–3428.
- 55 77 M. D. Smith, J. S. Rao, E. Segelken and L. Cruz, Force-Field Induced Bias in the Structure of  $A\beta_{21-30}$ : A Comparison of

- 1 OPLS, AMBER, CHARMM, and GROMOS Force Fields, *J. Chem. Inf. Model.*, 2015, **55**, 2587–2595.
- 78 A. D. MacKerell Jr, N. Banavali and N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, *Biopolymers*, 2001, **56**, 257–265.
- 5 79 D. L. Nelson and M. M. Cox, *Lehninger - Principles of Biochemistry*, Berlin, Heidelberg, 2001.
- 80 N. Sreerama and R. W. Woody, *Methods in Enzymology, Numerical Methods part D*, 2004, pp. 318–351.
- 10 81 C. Krittanai and W. C. Johnson, Correcting the Circular Dichroism Spectra of Peptides for Contributions of Absorbing Side Chains, *Anal. Biochem.*, 1997, **253**, 57–64.
- 82 A. Chakrabartty, T. Kortemme, S. Padmanabhan and R. L. Baldwin, Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities, *Biochemistry*, 1993, **32**, 5560–5565.
- 15 83 R. W. Woody, Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins, *Eur. Biophys. J.*, 1994, **23**, 253–262.
- 20 84 M. C. Manning and R. W. Woody, Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor, *Biochemistry*, 1989, **28**, 8609–8613.
- 85 N. Berova, P. L. Polavarapu, K. Nakanishi and R. W. Woody, *Comprehensive Chiroptical Spectroscopy*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012, vol. 1.
- 5 86 K. Guruprasad and S. Rajkumar,  $\beta$ - and  $\gamma$ -turns in proteins revisited: A new set of amino acid turn-type dependent positional preferences and potentials, *J. Biosci.*, 2000, **25**, 143–156.
- 87 N. Harada, K. Nakanishi and N. Berova, *Comprehensive Chiroptical Spectroscopy*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012, pp. 115–166.
- 10 88 N. Harada and K. Nakanishi, *Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry*, University Science Books, 1983.
- 15 89 N. Sreerama and R. W. Woody, *Circular dichroism: principles and applications*, Wiley-VCH, 2000, p. 877.
- 90 N. Harada, S.-M. L. Chen and K. Nakanishi, Quantitative definition of exciton chirality and the distant effect in the exciton chirality method, *J. Am. Chem. Soc.*, 1975, **97**, 5345–5352.
- 20
- 25
- 30
- 35
- 40
- 45
- 50
- 55



**Characterization of  $\beta$ -turns by electronic circular dichroism spectroscopy:  
A coupled molecular dynamics and time-dependent density functional  
theory computational study**

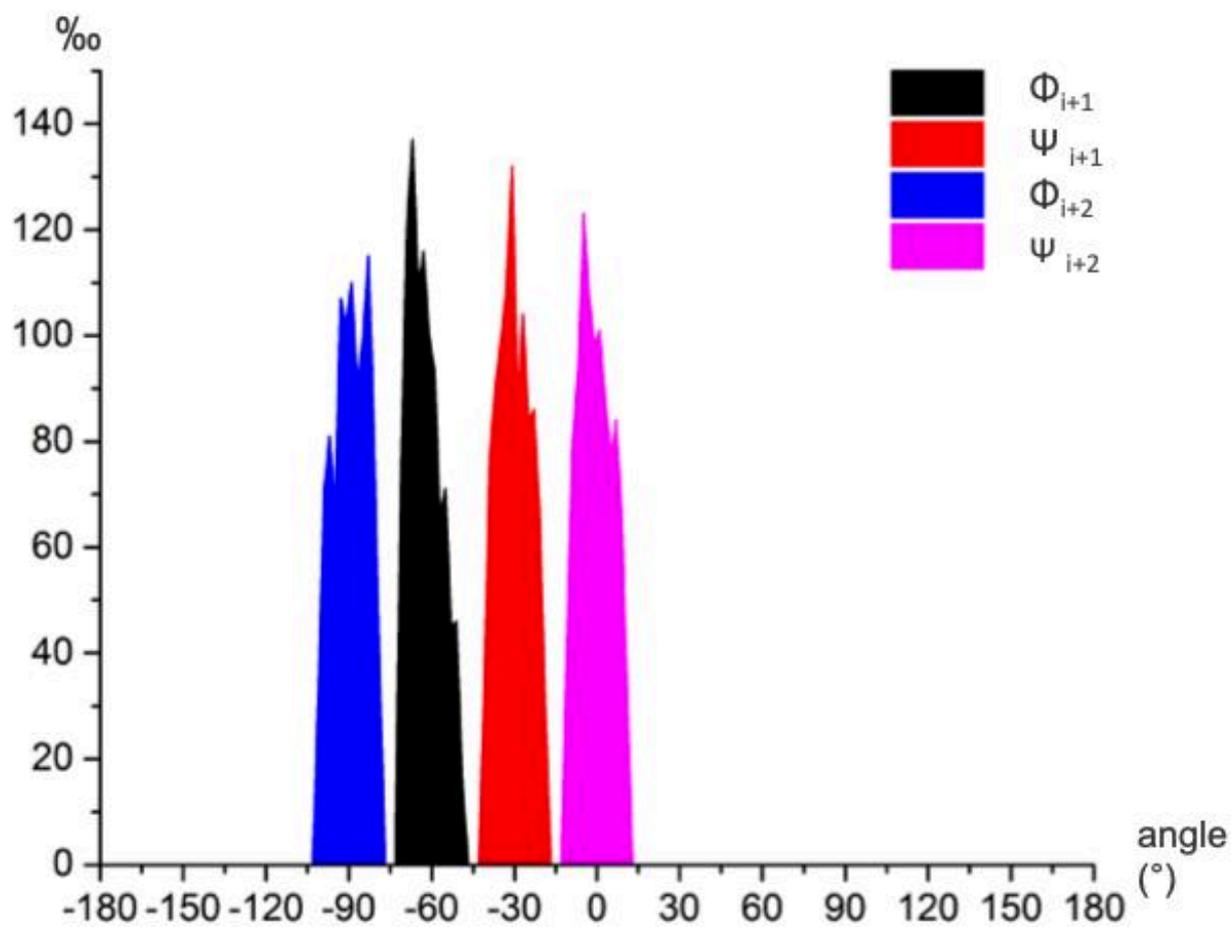
Mattia MIGLIORE, Andrea BONVICINI, Vincent TOGNETTI, Laure GUILHAUDIS, Marc  
BAADEN, Hassan OULYADI, Laurent JOUBERT, Isabelle SEGALAS-MILAZZO

***Supplementary information file***

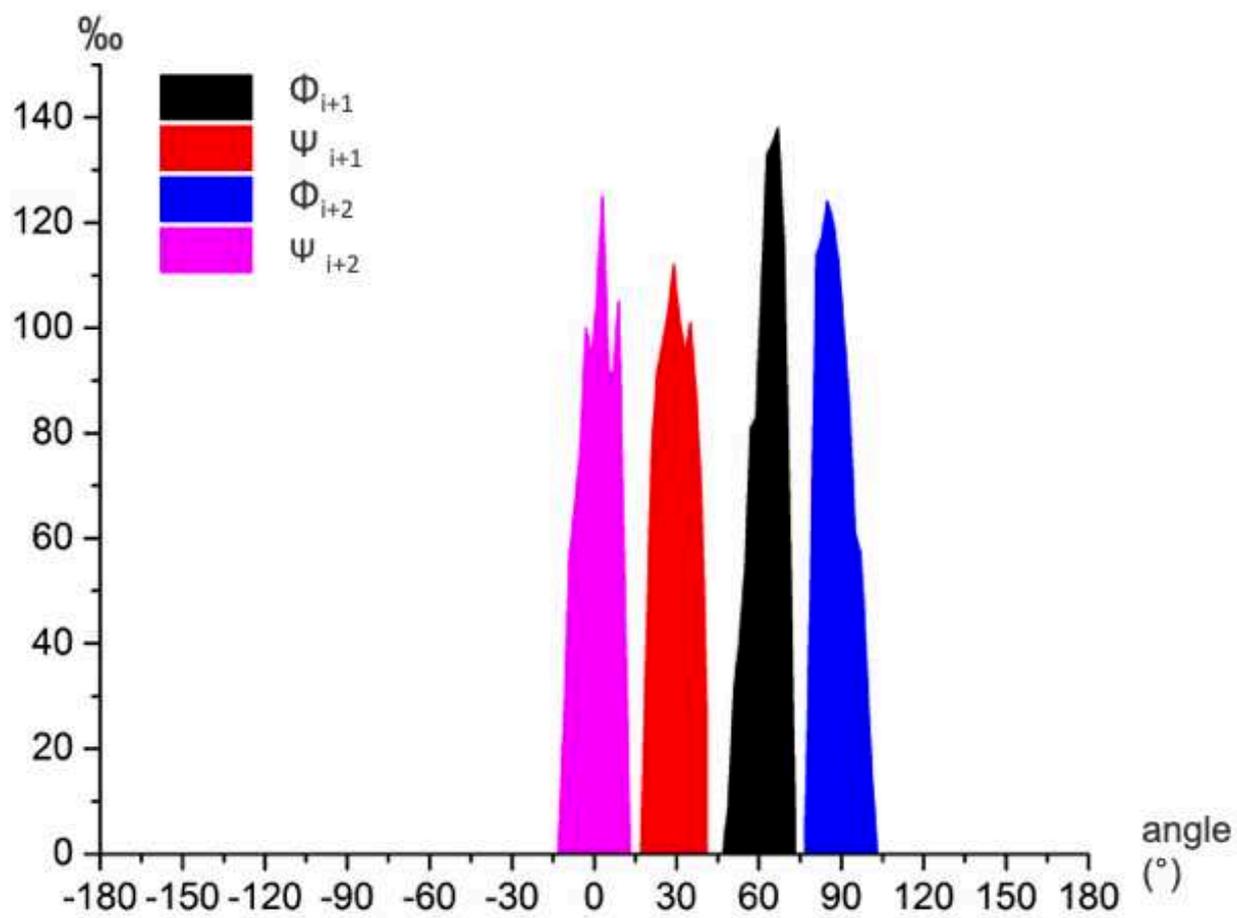
## Contents

<b>Figure S1.</b> Dihedral angles distributions for Ace-A-A-NHMe: type I $\beta$ -turn .....	S3
<b>Figure S2.</b> Dihedral angles distributions for Ace-A-A-NHMe: type I' $\beta$ -turn .....	S4
<b>Figure S3.</b> Dihedral angles distributions for Ace-A-A-NHMe: type II $\beta$ -turn .....	S5
<b>Figure S4.</b> Dihedral angles distributions for Ace-A-A-NHMe: type II' $\beta$ -turn .....	S6
<b>Figure S5.</b> Dihedral angles distributions for type I Ace-K-A-NHMe with only one Cl <sup>-</sup> and in a 150 mM NaCl solution.....	S7
<b>Figure S6.</b> Averaged absorption spectra for Ace-A-A-NHMe for the four $\beta$ -turns types..	S8
<b>Figure S7</b> $\kappa$ distributions for the four considered $\beta$ -turn types.....	S9
<b>Figure S8.</b> $\Delta\lambda$ distributions for the four considered $\beta$ -turn types.....	S10
<b>Figure S9.</b> Comparison between averaged ECD spectra computed from MD simulations using different force fields (OPLS-AA and CHARMM27).....	S11
<b>Figure S10.</b> Averaged ECD spectra for $\beta$ -turn type I computed with and without the PE-EEF extension .....	S12

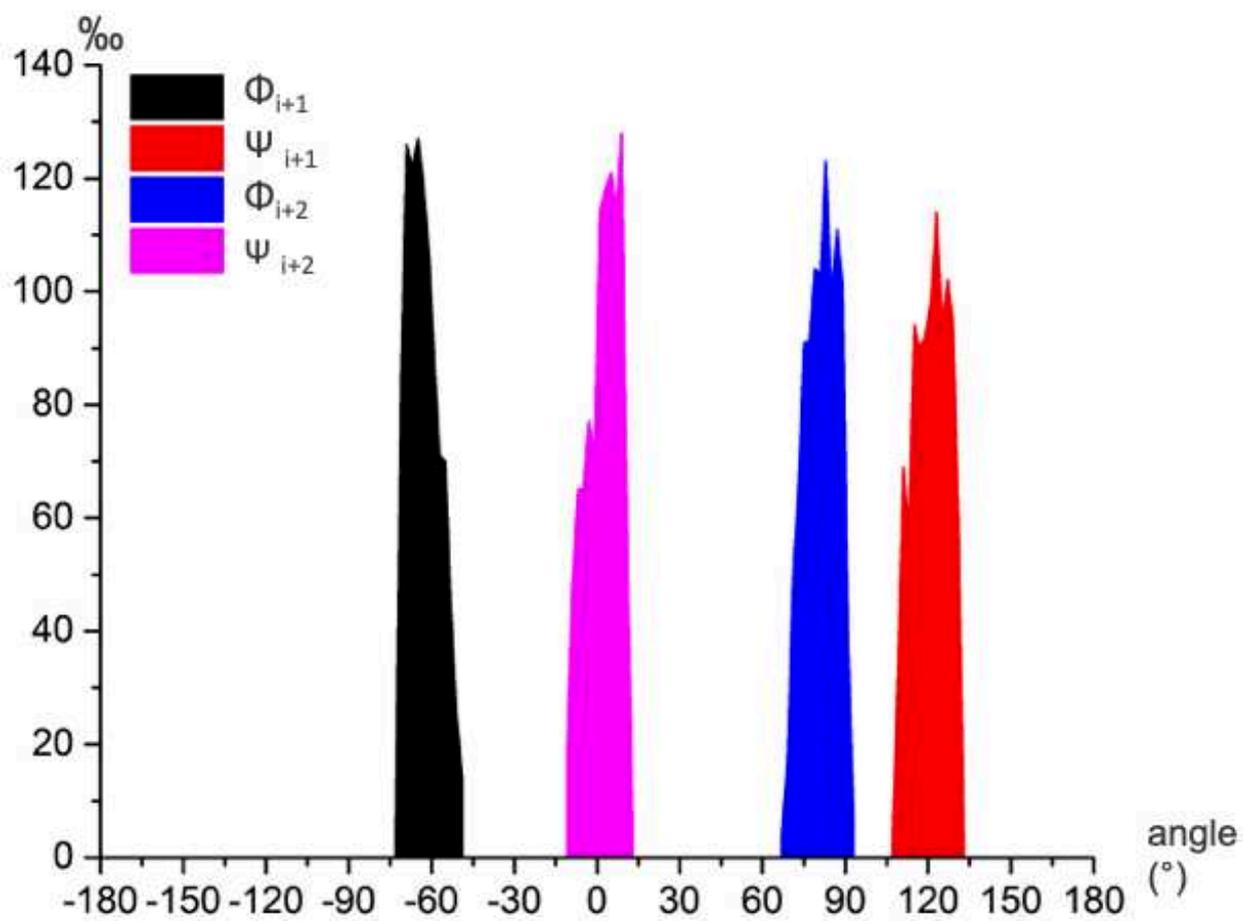
**Figure S1.** Dihedral angles (in °) distributions (in %) for Ace-A-A-NHMe in type I  $\beta$ -turn conformation.



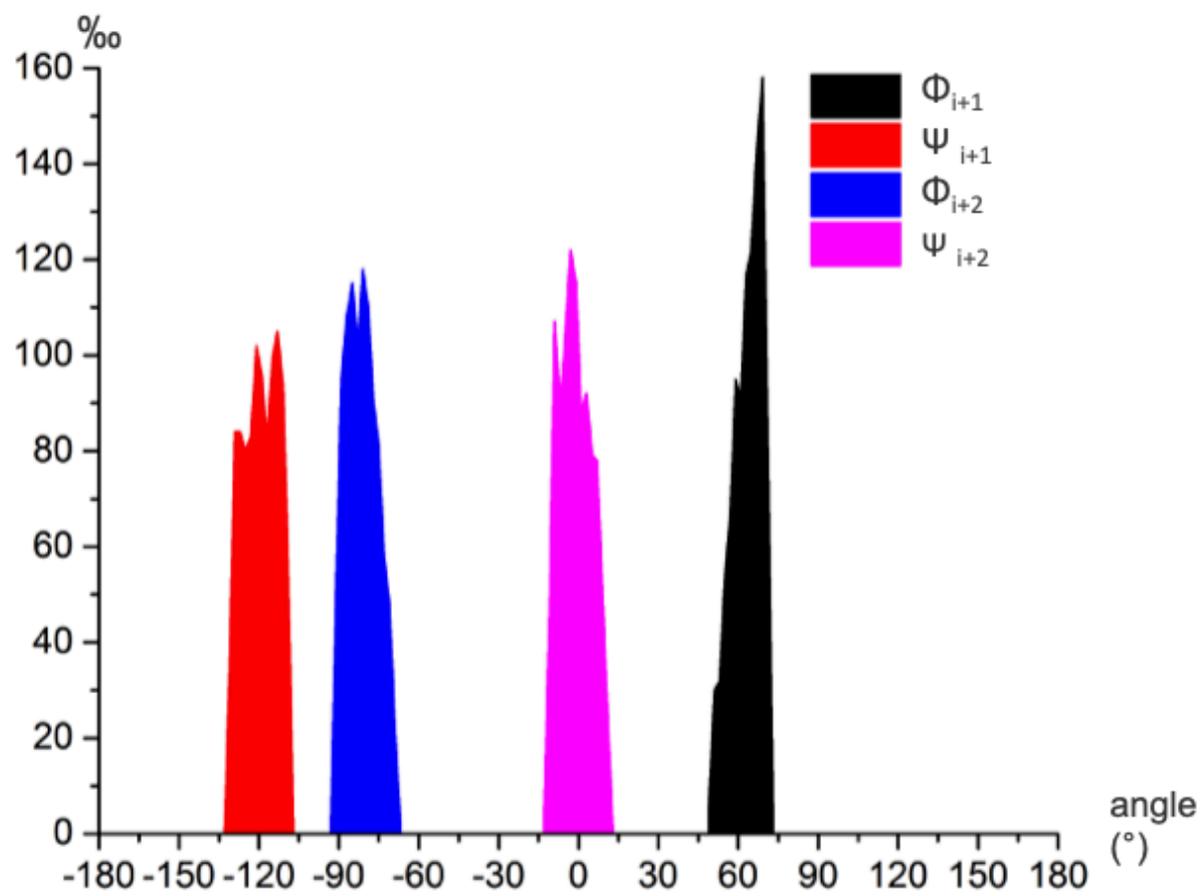
**Figure S2.** Dihedral angles (in °) distributions (in %) for Ace-A-A-NHMe in type I'  $\beta$ -turn conformation.



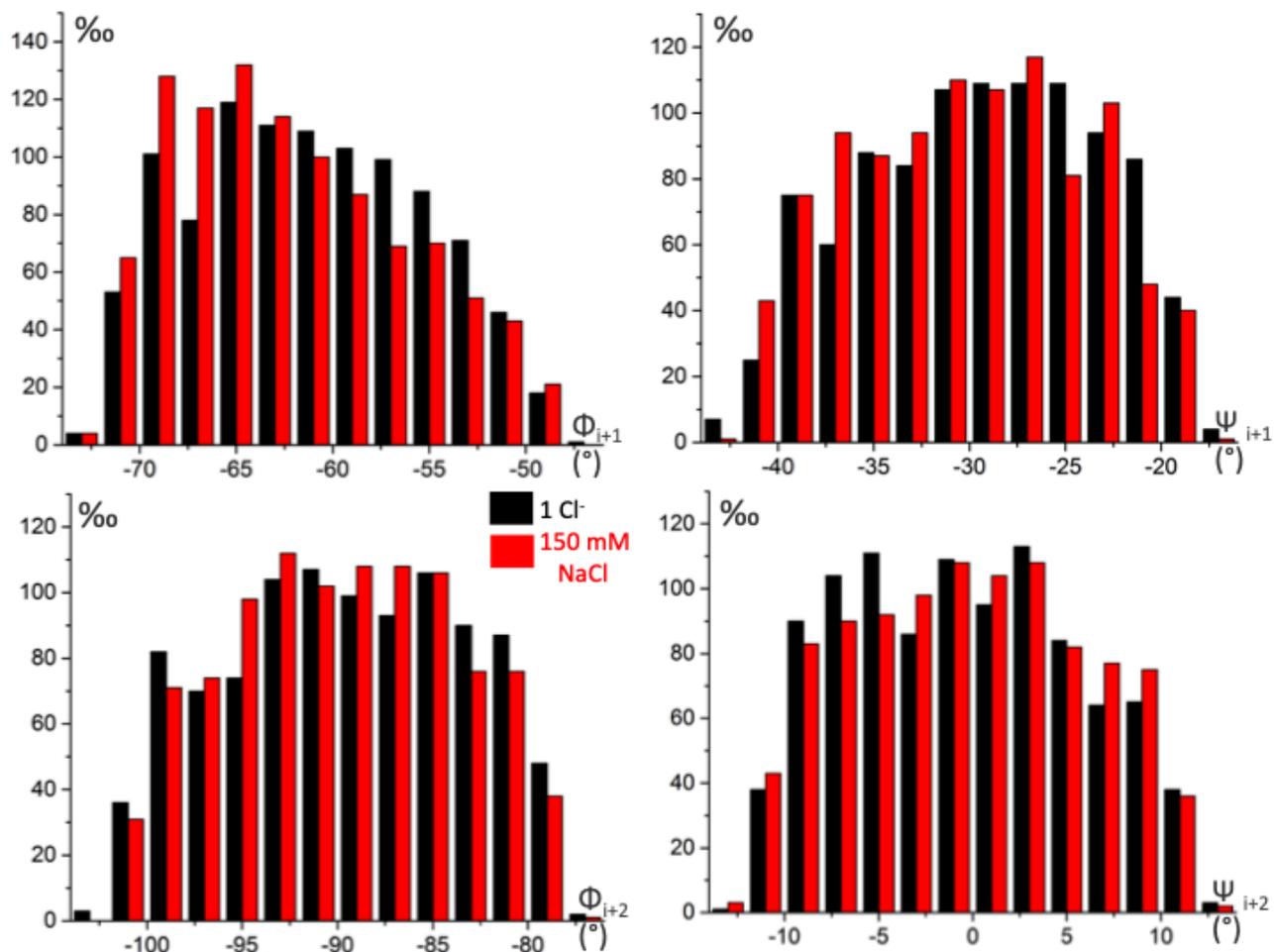
**Figure S3.** Dihedral angles (in °) distributions (in %) for Ace-A-A-NHMe in type II  $\beta$ -turn conformation.



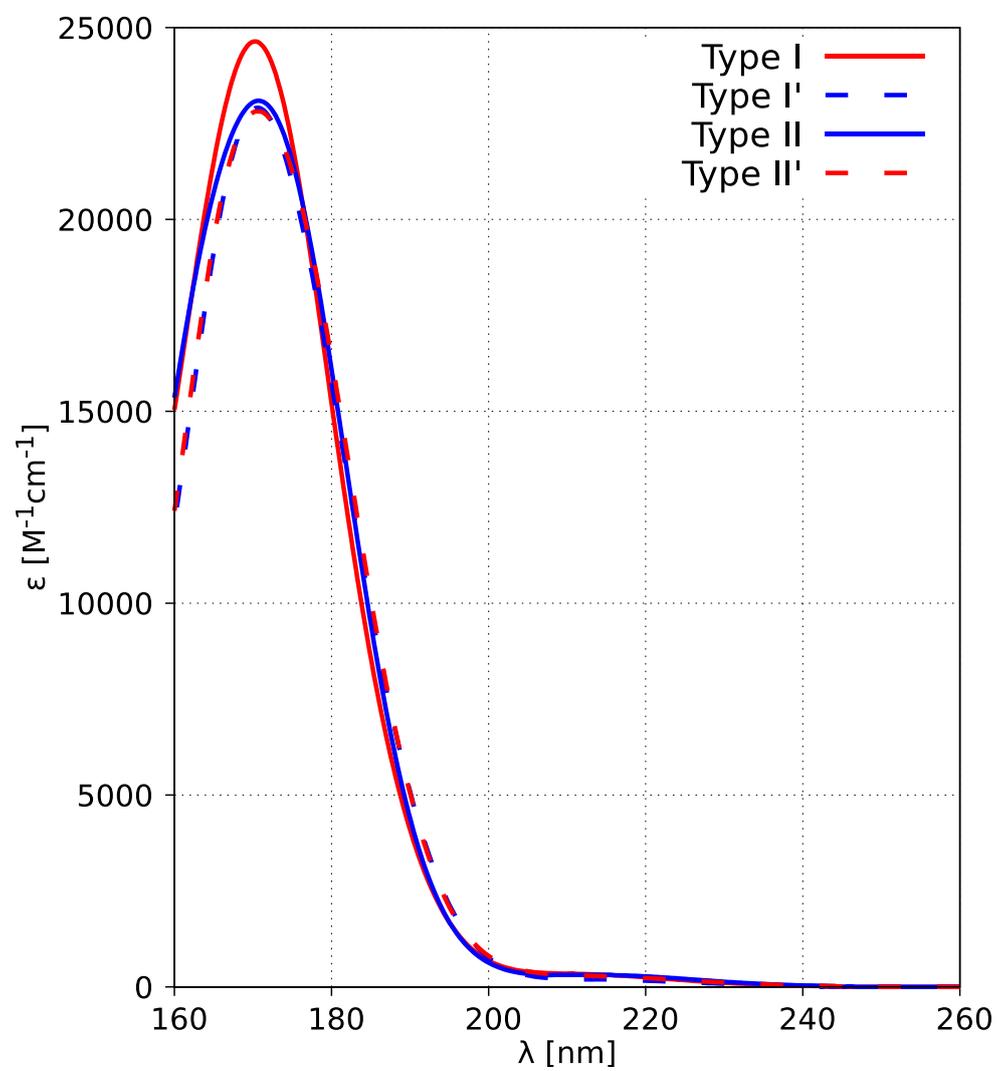
**Figure S4.** Dihedral angles (in  $^{\circ}$ ) distributions (in %) for Ace-A-A-NHMe in type II'  $\beta$ -turn conformation.



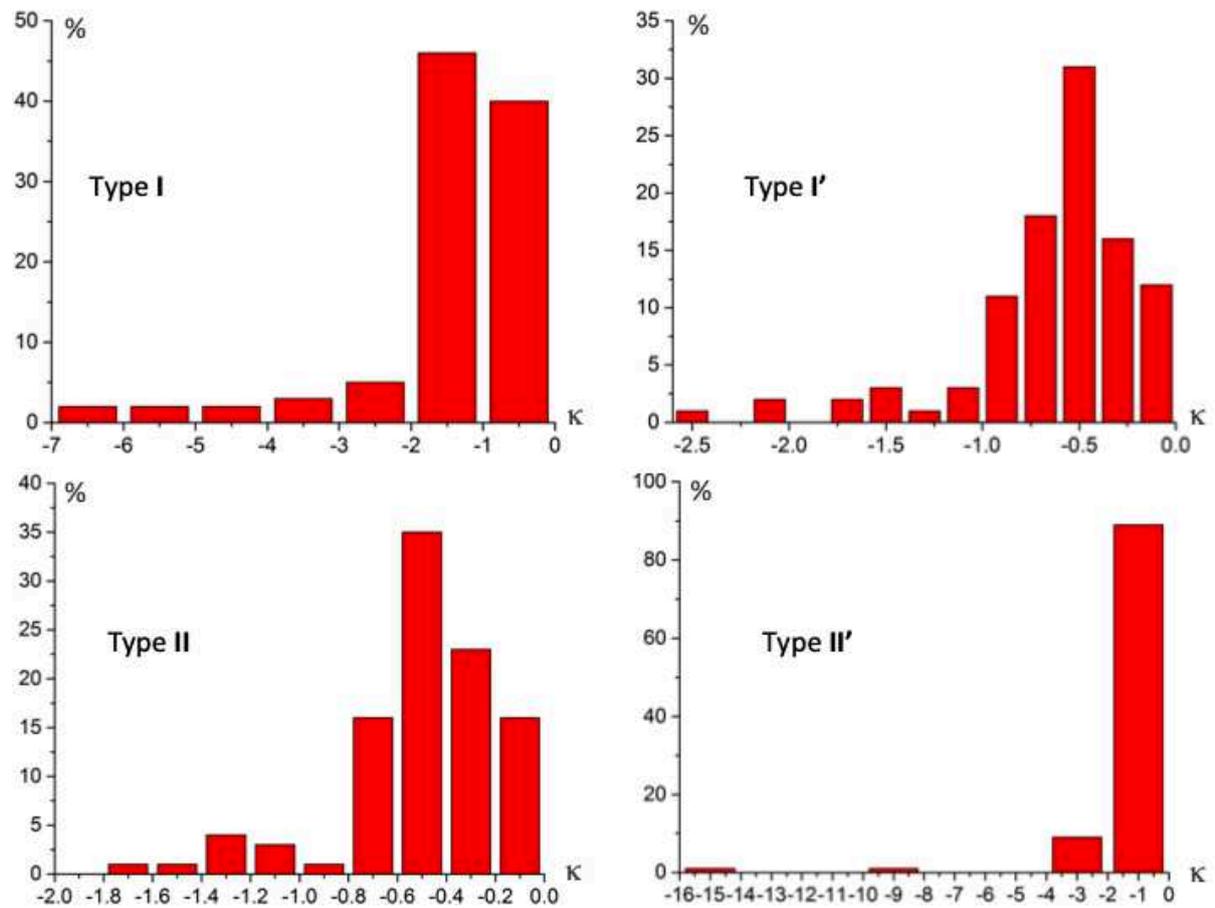
**Figure S5.** Dihedral angles (in  $^{\circ}$ ) distributions (in %) for Ace-K-A-NHMe in type I  $\beta$ -turn conformation with only one  $\text{Cl}^-$  as counteranion (in black) and in a 150 mM NaCl solution (in red).



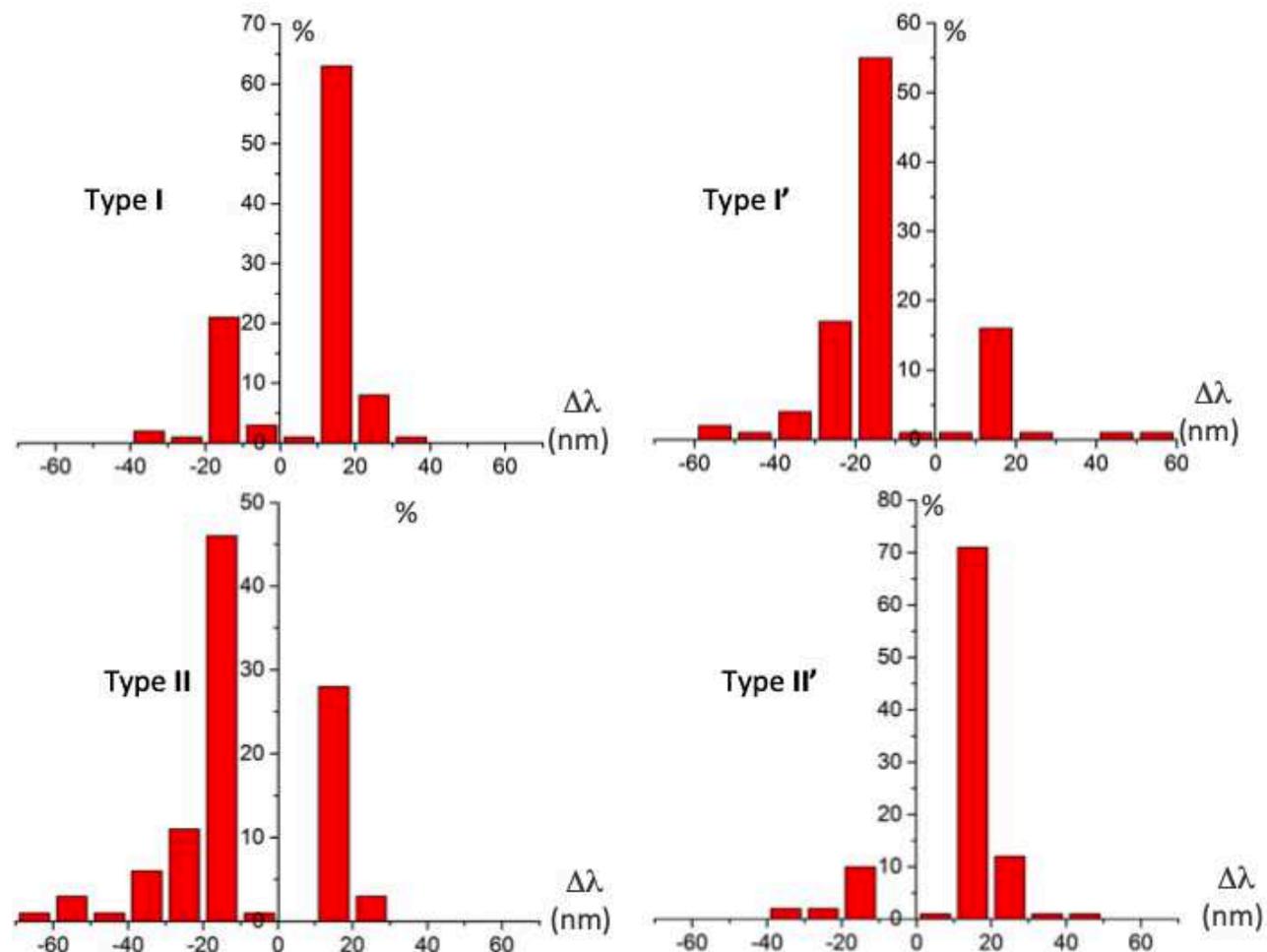
**Figure S6.** Averaged absorption spectra for Ace-A-A-NHMe for the four  $\beta$ -turns types considered here.



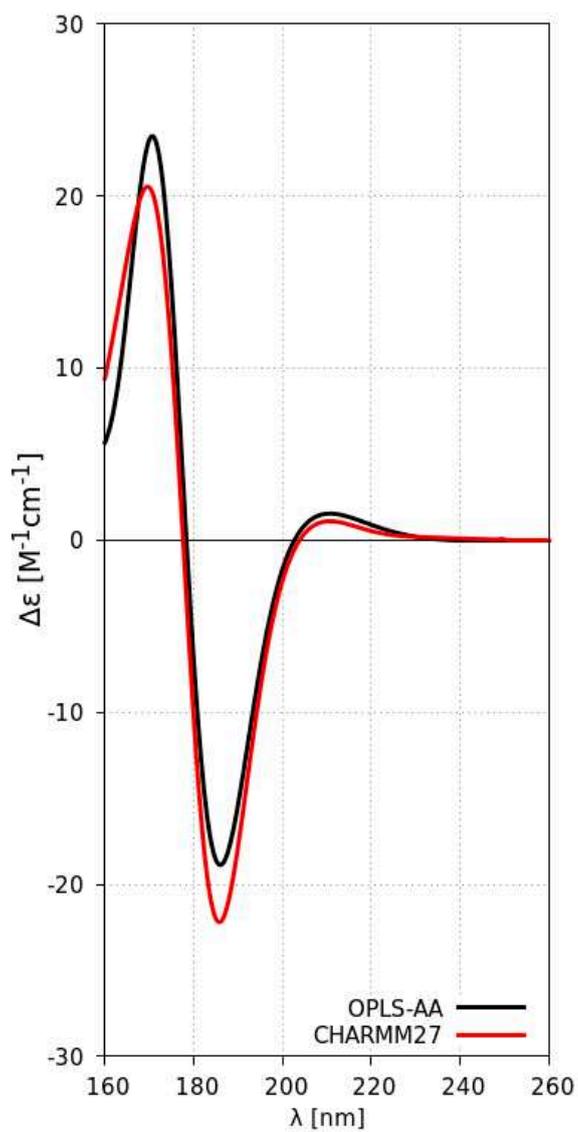
**Figure S7.** Distributions (in %) for the intensity ratio between the negative and positive ECD bands,  $\kappa$  (see eq. 5 in the paper) for the four considered  $\beta$ -turn types.



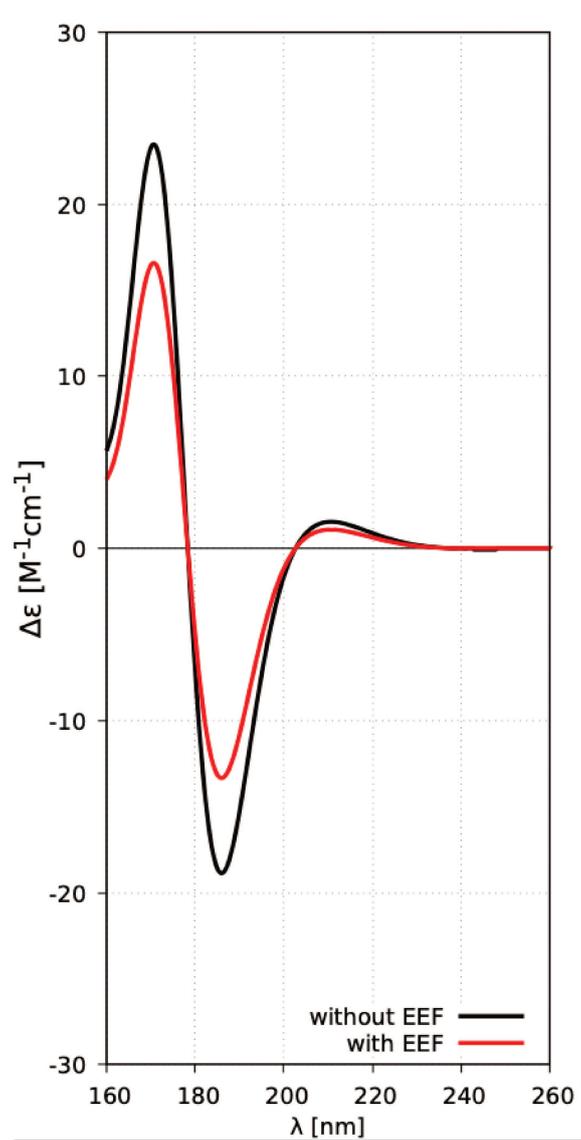
**Figure S8.** Distributions (in %) for the wavelength difference between the negative and positive ECD bands,  $\Delta\lambda$  (in nm, see eq. 4 in the paper) for the four considered  $\beta$ -turn types.



**Figure S9.** Comparison of averaged ECD spectra for the minimal model Ace-A-A-NHMe for the  $\beta$ -I turn conformation obtained from a restrained MD simulation (with allowed deviations of  $\pm 10^\circ$  around the canonical  $\Phi_{i+1}$ ,  $\Psi_{i+1}$ ,  $\Phi_{i+2}$  and  $\Psi_{i+2}$  dihedral angles) using either the OPLS-AA or the CHARMM27 force fields.



**Figure S10.** Averaged ECD spectra for  $\beta$ -turn type I computed with (using the RESPONSE module) and without (using the PROPERTIES module) the PE-EEF extension.



#### IV. Discussion des résultats et conclusions

Dans ce chapitre, nous avons présenté les travaux réalisés pour simuler les spectres de DCE des quatre types de coudes  $\beta$  (I, I', II et II'), qui sont les conformations caractérisées par RMN. Ces spectres ont été obtenus à travers les méthodes de calcul de la théorie de la densité de la fonctionnelle dépendante du temps, associés à un modèle d'inclusion polarisable (Polarizable Embedding) et en y couplant des simulations de dynamique moléculaire.

Les calculs nous ont permis de montrer que plusieurs spectres de DCE différents peuvent être identifiés au sein de ces quatre types de coudes  $\beta$  lorsqu'un acide aminé aromatique n'est pas présent dans la séquence peptidique. En particulier, nous avons pu remarquer que le type I est caractérisé par une allure composée d'une première bande négative ( $\sim 190$  nm) et une deuxième bande positive ( $\sim 170$  nm) (Figure 95). En revanche, un spectre miroir avec une première bande positive ( $\sim 190$  nm) et une deuxième bande négative ( $\sim 170$  nm) a été identifié pour le coude  $\beta$  de type I' (Figure 95). De manière similaire aux types I et I', les types II et II' possèdent deux spectres de DCE miroirs (Figure 95). En effet, le type II est caractérisé par une première bande positive ( $\sim 190$  nm) et une deuxième bande négative ( $\sim 170$  nm), alors que le spectre du type II' montre une première bande négative ( $\sim 190$  nm) et une deuxième bande positive ( $\sim 170$  nm).

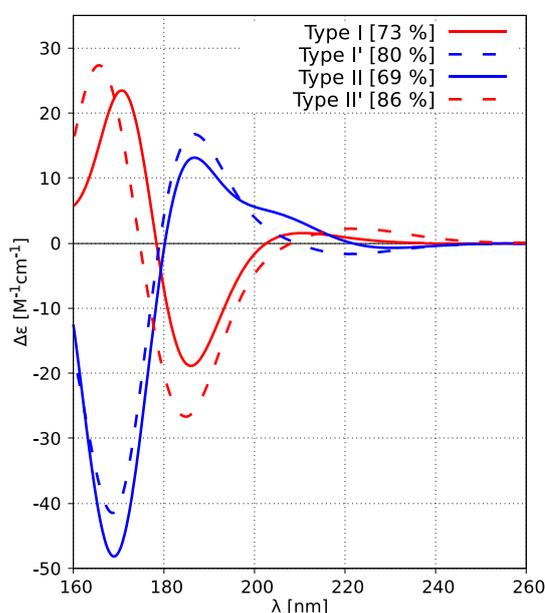


Figure 95 Spectres moyens de dichroïsme circulaire pour les coudes  $\beta$  de type I, I', II et II' calculés avec le niveau de théorie CAM-B3LYP/6-31+G\* pour le peptide modèle Ace-A-A-NHMe.

Notre étude contraste avec le travail fondateur de Woody<sup>98</sup>, dans lequel les types I et II présentaient la même allure, nommée « classe B », mais elle est en accord avec les travaux de Sathynarayana et Applequist<sup>99</sup> qui ont présenté deux spectres miroirs pour les coudes  $\beta$  de type I et II.

Les résultats de cette partie de l'étude nous ont permis de répondre aux questions posées au début de ce chapitre. En effet, le motif coude est caractérisé par un spectre de DCE différent par rapport aux autres structures secondaires, mais il est nécessaire souligner qu'un spectre général n'a pas été identifié car pour les types I/II' et I'/II nous retrouvons deux allures différentes. Ceci nous permet ensuite de répondre à la deuxième question sur la capacité de la spectroscopie de DCE de distinguer entre les différents types de coudes  $\beta$ . Effectivement, la spectroscopie de DCE peut aider à différencier les coudes  $\beta$  de type I, I', II et II' en deux groupes différents (I/II' et I'/II) selon les allures observées.

Enfin, la possibilité d'observer une signature de DCE caractéristique pour les coudes dans une protéine avec d'autres structures secondaires, telles que des hélices et/ou des feuillets, sera très difficile. Ceci en raison du nombre fixe et relativement court de résidus contenus dans un motif de type coude (3 ou 4 acides aminés) par rapport au nombre de résidus qui forment une hélice et un feuillet. Cependant, il devrait en principe être possible d'obtenir expérimentalement une allure de DC d'un coude  $\beta$  en utilisant un protocole expérimental très minutieux et après une déconvolution du signal de DC total à partir duquel on peut obtenir le pourcentage de structure secondaire présent dans une protéine (conformations hélicoïdale, étendue ou désordonnée) et traiter le signal résiduel comme dû au motif de type coude.

En revanche, pour de petites molécules telles que les peptides bioactifs, qui sont riches en coudes, la signature DCE de ce motif devrait pouvoir être distinguée de celles des autres éléments de structure secondaire. En effet, dans ces composés, la signature DCE du coude devrait représenter une proportion importante (~15-80%) du spectre expérimental total. Une telle analyse pourrait donc amener des informations structurales fines sur ce type de molécules.

En perspective de ce travail, il faut maintenant affiner le protocole de calcul quantique que nous avons utilisé en confrontant les spectres théoriques obtenus à une étude expérimentale de peptides modèles. Ceci permettra de valider ou corriger la position des différentes bandes observées. Une fois le protocole optimisé, les allures établies pourront servir de références générales pour les quatre types de coudes  $\beta$ . Ceci pourra être appliqué en particulier à l'analyse des spectres de DC de protéines ou des peptides riches en coudes.

Par ailleurs, nous avons montré que la présence d'une chaîne latérale de type aromatique (dans notre cas la phénylalanine) modifie l'allure des spectres de DCE des quatre coudes  $\beta$  de type I, II, I' et II'. Cependant, comme présenté dans le chapitre 1, les acides aminés aromatiques sont peu représentés dans les séquences de protéines contenant ces types de coudes. Il devrait donc être possible de négliger leur contribution pour ces chaînes peptidiques. En revanche, les

peptides bioactifs, qui, eux, présentent une forte proportion d'acides aminés aromatiques, la perturbation de ces derniers sur le spectre de DC total devra être prise en compte. Il est toutefois possible d'envisager la réalisation d'une librairie de spectres de DC de séquences qui contiennent des acides aminés aromatiques dans les positions  $i+1$  et  $i+2$  des coudes  $\beta$  dans les peptides bioactifs.

De manière intéressante, le protocole utilisé dans ce travail pourrait être employé afin de simuler les allures de spectres de DCE théoriques pour les autres types de coudes  $\beta$  (IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb et VIII) ainsi que pour les coudes de nature  $\gamma$  (classique et inverse). Ceci pourra être réalisé car le protocole est flexible, rapide et il suffit de construire les différents modèles de coudes pour obtenir les spectres de DC correspondants. De même, la rationalisation qualitative des résultats à travers l'« *Exciton Chirality Rule* » (ECR), utilisée dans notre travail pour les coudes  $\beta$  de type I, I', II et II', pourra être appliquée aux autres types de coudes. En particulier, pour les coudes  $\gamma$ , qui sont constitués seulement de deux chromophores, l'application de l'ECR serait plus adaptée car cette méthode avait été pensée et construite sur des molécules qui avait seulement deux chromophores séparés par un atome.<sup>182</sup>

Néanmoins, il est nécessaire de souligner que l'application de ce travail présentera quelques limites. En effet, pour quatre types de coudes  $\beta$ , seulement deux allures peuvent être déterminées. Ainsi, la distinction entre le type I et le type II' n'est pas possible car ces derniers présentent une allure similaire (une première bande négative et une deuxième positive). La même conclusion doit être faite pour les types I' et II car une allure miroir a été obtenue. Il est donc possible réaliser une répartition en groupes de coudes plutôt qu'une distinction des coudes de manière individuelle. Cette observation pourrait être retrouvée pour les autres types de coudes  $\beta$ .

En conclusion, le protocole de calcul utilisé a permis de mettre en évidence de nouvelles signatures DC caractéristiques pour les coudes  $\beta$  de type I, I', II et II'. En outre, à travers cette étude, pour la première fois, des spectres de DC pour les coudes  $\beta$  de type I' et II' ont été établis. L'application de ce protocole permettrait d'obtenir un spectre de DCE de référence pour chaque type de coude  $\beta$  et  $\gamma$  conduisant à la constitution de plusieurs catégories d'allures. L'utilisation de ces catégories pour l'analyse d'un spectre de DC permettrait une meilleure caractérisation des coudes  $\beta$  et  $\gamma$  présents au sein d'un peptide ou d'une protéine riche en coudes.



## **Conclusions et perspectives**



## Conclusions et perspectives

Depuis plusieurs décennies, l'utilisation des peptides en tant qu'agents thérapeutiques apparaît comme un axe de recherche de plus en plus attractif.<sup>1</sup> La connaissance de la structure tridimensionnelle de ces molécules est donc une information essentielle. En effet, elle permet d'obtenir une meilleure compréhension des relations structure-activité et contribue à identifier les éléments impliqués dans la reconnaissance des peptides avec leur cible.

Plusieurs études ont suggéré que les coudes, avec les hélices, sont les motifs de reconnaissance privilégiés des peptides bioactifs par leur cible biologique.<sup>7</sup> Plus précisément, la localisation des coudes à la surface des molécules a conduit à l'hypothèse que ces derniers jouent un rôle critique dans les phénomènes de reconnaissance au sein des systèmes biologiques.<sup>183</sup> Leur caractérisation fine est donc fondamentale pour comprendre les mécanismes d'activation de la cible et pour le développement de molécules à visée thérapeutique, comme les peptidomimétiques.<sup>184</sup>

Les coudes les plus fréquemment rencontrés dans les peptides bioactifs sont les coudes de type  $\gamma$  et  $\beta$  dont le schéma général est présenté à la Figure 96. Pour chacun, il a été proposé différents sous-types selon leurs caractéristiques géométriques. La classification actuelle répertorie 2 types de coudes  $\gamma$  (classique et inverse) et 12 types de coudes  $\beta$  (I, I', II, II', IV<sub>1</sub>, IV<sub>2</sub>, IV<sub>3</sub>, IV<sub>4</sub>, VIa1, VIa2, VIb, VIII et IV<sub>misc</sub>).<sup>45</sup>

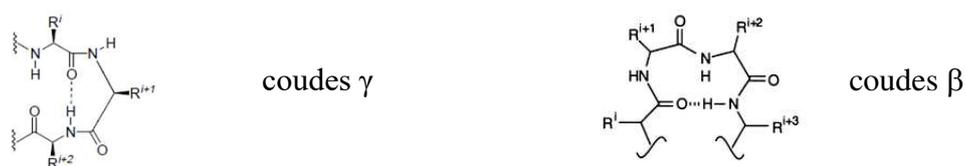


Figure 96 Schémas généraux de coudes  $\gamma$  et  $\beta$ . Les acides aminés constitutifs du coude sont numérotés à partir du résidu  $i$ .

La caractérisation structurale de structures secondaires (hélices  $\alpha$ , feuillets  $\beta$  et coudes) dans les peptides en solution est le plus souvent réalisée grâce aux techniques de dichroïsme circulaire (DC) et de résonance magnétique nucléaire (RMN) couplée à la modélisation moléculaire. A l'heure actuelle, les données disponibles permettent de déterminer sans ambiguïté les structures hélices et feuillets. En revanche, l'absence de données de référence rend la caractérisation structurale des coudes plus difficile. Cela est dû notamment à leur nombre important, à leur petite taille (3 ou 4 acides aminés), à leur flexibilité et à leur caractère non répétitif.

A cet égard, l'un des axes de recherche de l'équipe « Analyse et Modélisation » de l'UMR 6014 CNRS COBRA, où j'ai réalisé ces travaux de thèse, est la caractérisation de la structure tridimensionnelle de peptides bioactifs, par DC et RMN couplée la modélisation moléculaire, en vue d'une meilleure compréhension des relations structure-activité. Pour déterminer les structures de peptides bioactifs riches en coudes étudiés au sein de notre équipe, il est indispensable d'élargir les paramètres de RMN et de DC de référence pour les différents types de coudes  $\beta$  et  $\gamma$ . Ceci a donc constitué l'objectif principal de mes travaux de thèse.

Pour la RMN, deux approches ont été utilisées : l'une, expérimentale, et l'autre, théorique, toutes deux basées sur l'étude de peptides modèles. L'approche expérimentale n'a malheureusement pas permis de déterminer des données de référence, en raison d'un manque de structuration des molécules étudiées. En revanche, ces données de référence ont pu être obtenues par l'approche théorique, basée sur des méthodes de modélisation moléculaire. Ainsi, la mesure de distances inter-hydrogènes sur un ensemble de peptides représentatifs de coudes  $\beta$  et  $\gamma$  a permis de construire des diagrammes NOE pour la totalité des types de coudes  $\beta$  et les deux types de coudes  $\gamma$ . L'analyse de ces diagrammes conduit à une identification aisée des coudes de type IV<sub>1</sub>, IV<sub>2</sub>, VIb et VIII au sein des coudes  $\beta$ , ainsi que des deux types de coudes  $\gamma$ . Pour les autres types de coudes  $\beta$ , il est plus difficile de distinguer le type I des types III, IV<sub>3</sub> et IV<sub>4</sub>, le type I' du III', le type II du II', ainsi que le type VIa1 du VIa2.

L'étude des distances inter-hydrogènes a été complétée par un calcul de la constante de couplage  $^3J_{\text{HN-H}\alpha}$  du (ou des) résidu(s) centraux des coudes à partir de leur angle dièdre  $\Phi$ . Ceci a permis de mettre en évidence un impact important des variations autorisées pour l'angle  $\Phi$  sur la valeur de cette constante pour certains types de coudes. Ceci indique que la valeur canonique de la constante de couplage  $^3J_{\text{HN-H}\alpha}$  du (ou des) résidu(s) centraux des coudes ne peut pas être utilisée seule, comme référence d'un type de coude. En revanche, elle peut servir à confirmer son identification à l'issue de l'analyse du diagramme NOE. Pour certains types de coudes  $\beta$ , les valeurs observées pourront lever des ambiguïtés.

L'analyse combinée des diagrammes NOE et des constantes de couplage  $^3J_{\text{HN-H}\alpha}$  a alors conduit à proposer un organigramme d'aide à la décision pour différencier les types au sein de chaque nature de coude  $\beta$  ou  $\gamma$ . Cette analyse peut amener à l'identification non ambiguë d'un type de coude ou, à défaut, à une sélection d'un groupe de types de coudes.

Pour le DC, nous avons utilisé une approche théorique, couplant dynamique moléculaire et TDDFT, pour simuler les spectres de DC de référence de peptides modèles adoptant des conformations de coudes  $\beta$  de type I, I', II et II'. Ce choix de types de coudes a été guidé par des conclusions contradictoires au sein de travaux antérieurs. Ces calculs ont abouti à deux spectres différents pour les quatre types de coudes  $\beta$  étudiés, suggérant que le DC pourrait permettre de différencier ces coudes en deux groupes : I/II' et I'/II. Cette discrimination n'est cependant pas possible lorsqu'un acide aminé aromatique est présent dans la séquence peptidique.

L'ensemble de ces résultats indique que les coudes ne présentent pas nécessairement les mêmes signatures pour les deux techniques. Ainsi, les spectres DC des coudes  $\beta$  de type I et II' sont similaires, alors qu'ils possèdent des diagrammes NOE différents. On peut faire la même observation pour les coudes  $\beta$  de type I' et II. La combinaison des signatures discriminantes de RMN et de DC pourrait donc permettre une meilleure identification des différents types au sein des coudes  $\beta$ , au moins pour les types I, I', II et II'. Pour les autres types de coudes  $\beta$ , il sera nécessaire de déterminer leur spectre de DC de référence avant de tirer une conclusion analogue.

L'identification du type de coude pourrait être réalisée en deux temps. Dans un premier temps, l'analyse par DC permettrait d'identifier un groupe de types de coudes. Dans un second temps, l'analyse par RMN et la comparaison des données expérimentales (diagramme NOE et constantes de couplage  $^3J_{\text{HN-H}\alpha}$ ) avec les données générées dans ce travail servirait à identifier le type de coude.

A l'issue de ce travail, plusieurs perspectives peuvent être envisagées :

- Pour la RMN, les données de référence issues des calculs doivent maintenant être rapprochées de données expérimentales. En effet, les organigrammes de décision élaborés au cours de ce travail seront intéressants à appliquer dans les analyses structurales de peptides bioactifs (neuropeptides, par exemple) qui sont souvent riches en coudes. Ainsi, les distances inter-hydrogènes, couplées aux valeurs de constantes de couplage  $^3J_{\text{HN-H}\alpha}$  permettront d'identifier une ou plusieurs structure(s) de type coude au sein de la séquence peptidique. De plus, leur nature et leur type devraient pouvoir être caractérisés. Ces informations pourraient guider les études de relation structure-activité

dans une optique de conception rationnelle d'analogues de peptides bioactifs à visée thérapeutique.

- Pour le DC, les spectres théoriques obtenus précédemment pour les coudes  $\beta$  de type I, I', II et II' devront être confrontés à des études expérimentales. Pour atteindre cet objectif, il sera nécessaire de disposer du spectre DC expérimental pour au moins l'un des types de coudes étudiés afin de pouvoir obtenir la meilleure adéquation entre spectre calculé et spectre expérimental. Ceci permettra de valider ou de corriger la position des différentes bandes observées sur les spectres théoriques. Une fois le modèle de calcul validé, les spectres de DC théoriques des autres types de coudes  $\beta$  et des coudes  $\gamma$  pourront être calculés.
- Enfin, l'utilisation combinée des deux techniques, DC et RMN, pourra être testée dans l'analyse structurale de peptides bioactifs. Ceci permettra de vérifier si leur couplage permet une caractérisation structurale fine des coudes et leur identification de façon non ambiguë.

## **Annexes**



# ***Annexes***

## ***Annexes 1 : Matériels et Méthodes***

### **I. Le dichroïsme circulaire**

L'aptitude des composés asymétriques à interagir avec un rayonnement électromagnétique, leur confère une activité optique importante. Le dichroïsme circulaire (DC) est ainsi défini comme l'absorption différentielle ( $\Delta\epsilon$ ) des lumières polarisées circulairement à gauche ( $\epsilon_G$ ) et à droite ( $\epsilon_D$ ) à une longueur d'onde donnée.

Les macromolécules biologiques, comme les peptides ou les protéines, sont composés d'éléments optiquement actifs (liaisons peptidiques, cycles aromatiques, etc...) et produisent un spectre DC qui leur est propre et est caractéristique de la structure tridimensionnelle qu'elles adoptent. L'étude du dichroïsme circulaire d'un peptide ou d'une protéine va ainsi permettre de déterminer le principal type de structure secondaire en présence et d'en estimer sa proportion. Le spectre d'un peptide ou d'une protéine sera une combinaison linéaire de chaque spectre de référence, pondérée par la fraction de chaque type de structure secondaire. Ainsi, il sera possible d'extraire mathématiquement les informations concernant la structure secondaire de tout peptide ou protéine inconnus à partir de son spectre de DC.

Les bandes de dichroïsme circulaire sont localisées dans deux régions spectrales et donnent diverses informations sur la structure tridimensionnelle des polypeptides :

- La région d'ultra-violet (UV) lointain ou région des amides (170-250 nm) contient des informations sur la structure secondaire via le repliement des liaisons peptidiques. Ces bandes permettent donc d'apprécier quantitativement la proportion de structure secondaire régulière (hélice, feuillet, coude, conformation aléatoire) ;
- Les signaux dans la région d'UV proche (250-330 nm) correspondent aux contributions des acides aminés aromatiques, ponts disulfure et fournissent des informations sur le repliement tertiaire.

#### **I.1 Principe général du dichroïsme circulaire**

La lumière polarisée plane peut être considérée comme étant la somme de deux composantes lumineuses circulaires de même amplitude (Figure 97). Ces deux composantes tournent dans deux sens différents, l'une dans le sens inverse des aiguilles d'une montre (gauche, G) et l'autre dans le sens des aiguilles d'une montre (droite, D). Un signal de DC est basé sur la capacité des molécules chirales ou chromophores (groupements optiquement actifs) d'absorber

différemment ces deux composantes (Figure 97).<sup>185,186</sup> Par ailleurs, aucun signal n'est observable si les composantes G et D ont absorbé la lumière de manière égale (Figure 97). L'appareil de DC mesure et enregistre donc cette variation d'absorbance entre les deux composantes pour une longueur d'onde ( $\lambda$ ) donnée (Figure 97). Cette variation est définie par l'équation suivante :

$$\Delta A_{(\lambda)} = (A_G - A_D)_{(\lambda)}$$

Ainsi, en ajoutant un échantillon dans un faisceau dont la polarisation circulaire est alternativement à gauche et à droite, le DC mesure la différence de l'absorbance  $\Delta A$  entre ces deux lumières. On obtient donc des spectres de DC en fonction de la longueur d'onde (Figure 97).

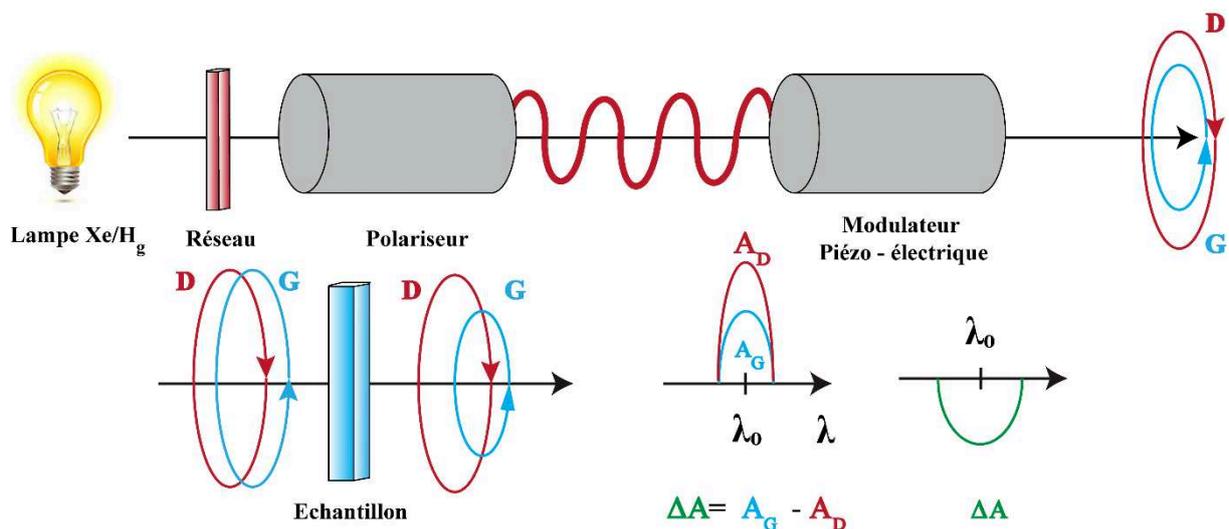


Figure 97 Schéma de principe général du dichroïsme circulaire

Le vecteur champ électrique  $\vec{E}$  d'une onde lumineuse polarisée peut être décomposé en deux vecteurs  $\vec{E}_D$  et  $\vec{E}_G$  (Figure 98). En l'absence d'échantillon, le faisceau polarisé est rectiligne et il est formé de deux composantes polarisées circulairement  $\vec{E}_D$  et  $\vec{E}_G$  de même amplitude  $E_0$  (Figure 96). En présence de l'échantillon possédant des molécules chirales, les deux composantes  $\vec{E}_D$  et  $\vec{E}_G$  ne seront plus égales. De ce fait, la polarisation du faisceau n'est plus rectiligne mais elliptique (Figure 98).

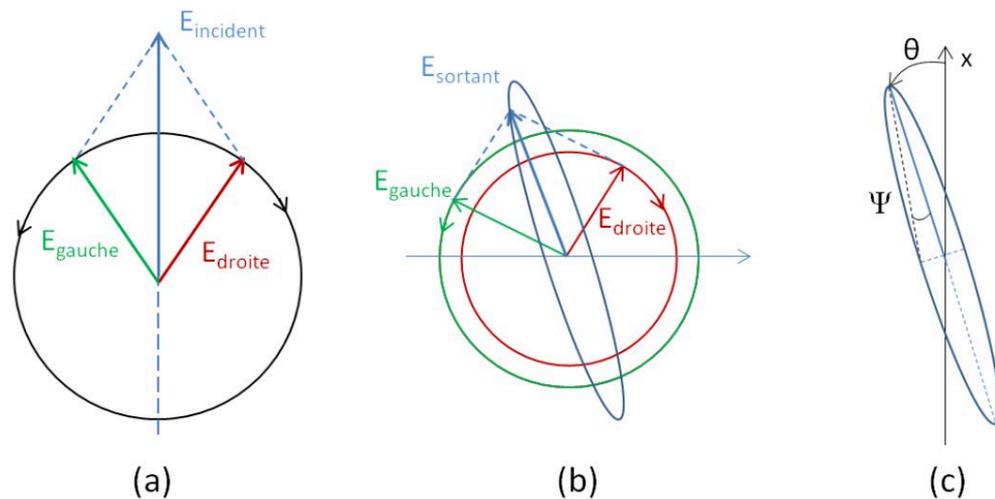


Figure 98 Schémas représentant les modifications du champ électrique à la traversée d'un milieu optiquement actif. (a) Champ électrique incident polarisé linéairement décomposé en ondes circulaires gauche et droite. (b) Les ondes gauche et droite sont absorbées et déphasées de façon différente à la traversée du milieu ; elles se recomposent à la sortie en une onde elliptique inclinée. (c) La direction du grand axe de la polarisation elliptique est incliné d'un angle  $\theta$  par rapport à la direction de polarisation incidente : c'est l'angle de rotation optique. L'ellipticité de la polarisation sortante est donnée par l'angle  $\Psi$ .

L'ellipticité est alors définie par la relation suivante :

$$\tan \theta_{(\lambda)} = \left( \frac{E_D - E_G}{E_D + E_G} \right)_{(\lambda)}$$

La relation entre  $\Delta A_{(\lambda)}$  et  $\theta_{(\lambda)}$ , exprimée en degrés, est donnée par l'équation ci-dessous :

$$\theta_{(\lambda)} = \frac{180}{4\pi} \ln(10) \times \Delta_{(\lambda)}$$

$$\text{Soit } \theta_{(\lambda)} = 32,98 \Delta A_{(\lambda)}$$

Par conséquent, le résultat de DC est affiché directement dans la plupart des instruments en ellipticité et non pas en  $\Delta A$ .

Le DC est donc une technique efficace qui permet de donner une idée initiale, en un minimum de temps, sur la structure secondaire des peptides ou des protéines. La déconvolution des spectres de DC rend possible l'estimation quantitative des différentes structures secondaires. Cependant, le DC n'apporte pas d'informations détaillées sur la localisation des structures

secondaires dans la séquence. A cet égard, il est important de coupler le DC à d'autres techniques afin de caractériser la structure tridimensionnelle d'un peptide ou d'une protéine .

## **II. Résonance magnétique nucléaire (RMN) en solution**

### **II.1 Principe général**

La RMN est la seule technique biophysique puissante qui permet de caractériser la structure des molécules en solution. Elle est également utilisée pour d'autres types d'études comme les études de dynamique des molécules et les études d'interactions intermoléculaires.<sup>187</sup> Le phénomène de base de la RMN a été mis en évidence en 1945 par Felix Bloch et Edward Purcell qui ont eu en 1952 le prix Nobel de physique pour cette découverte. Cependant, l'utilisation de cette technique en biologie structurale est beaucoup plus récente. La première structure d'un inhibiteur de protéase a été résolue en 1985 par Wüthrich et ses collaborateurs.<sup>118</sup> Après cette détermination, de nombreux développements technologiques et méthodologiques ont été apportés afin d'augmenter la qualité des spectres et par la suite l'information structurale obtenue par RMN. Malgré tous ces développements, la détermination de la structure des protéines de très grande taille moléculaire reste l'une des limitations de cette technique. En effet, la caractérisation des structures par la spectroscopie RMN est appliquée en général sur les protéines jusqu'à une taille moléculaire de 50 kDa.<sup>188</sup> Pour des macromolécule de petite taille moléculaire, l'analyse du noyau hydrogènes est suffisante alors que avec des molécule de taille plus importante il est nécessaire un marquage isotopique.

Grâce à la diversité de paramètres mesurables, la RMN haute résolution constitue une technique puissante pour la détermination de la structure secondaire tridimensionnelle d'un peptide ou d'une protéine en solution.

### **II.2 Expériences 1D**

L'analyse de composés biologiques par RMN se fait généralement dans des solutions aqueuses. L'un des problèmes majeurs de la RMN du proton est la présence du signal de l'eau. Ce dernier est de  $10^4$  à  $10^6$  plus intense que les signaux des protons du peptide étudié. Cependant, l'utilisation d'une solution complètement deutérée ( $D_2O$ ) aboutit à la disparition des informations structurelles essentielles.<sup>118</sup> Les protons amides des acides aminés sont en effet

très labiles et s'échangent rapidement avec le deutéron de D<sub>2</sub>O. De ce fait, plusieurs techniques de suppression de l'eau ont été mises en place afin de s'affranchir de ce problème :

- La présaturation du signal du solvant consiste à saturer le signal du solvant et permet ainsi de le diminuer d'un facteur de 10<sup>3</sup>.<sup>189</sup>
- Les séquences *WATERGATE* ou *Excitation sculpting* permettent de détruire le signal de l'eau en utilisant des impulsions gradients de champ magnétique.<sup>190,191</sup>

Après suppression du signal de l'eau, le spectre du peptide ou de la protéine apparaît plus nettement. Chaque signal observé correspond à un proton du peptide (Figure 99). Différentes régions sur le spectre 1D sont repérées selon le type et l'environnement de chaque proton (H<sub>N</sub>, CH, aromatique, etc...) comme montré dans la figure 99.<sup>188</sup>

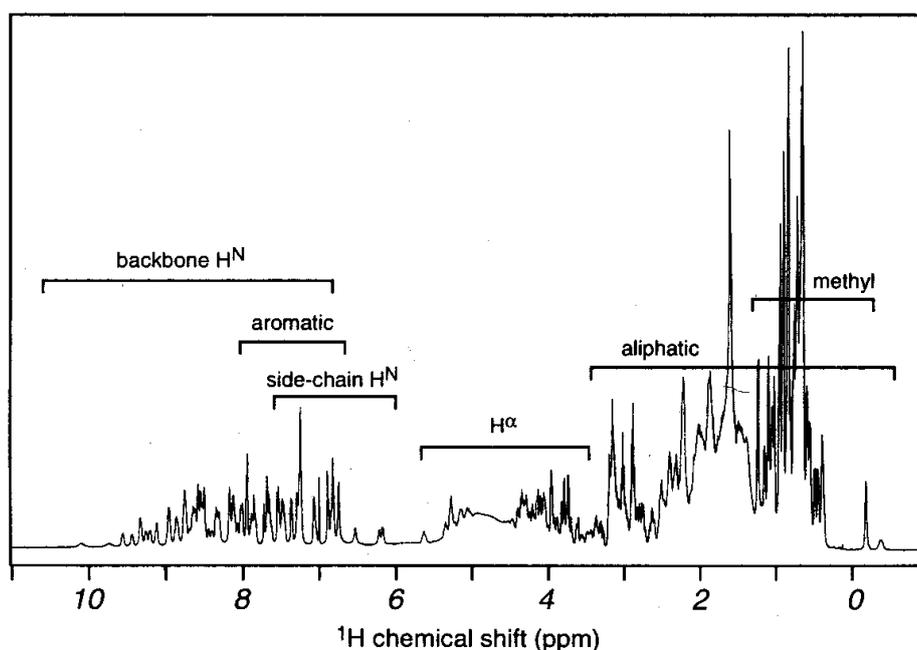


Figure 99 Positions des déplacements chimiques des <sup>1</sup>H des groupes chimiques présents dans les acides aminés, constitutifs des peptides et des protéines.

Les expériences 1D ne donnent pas d'information précise sur le peptide ou la protéine étudiés à cause de la superposition des signaux. Il est donc nécessaire d'utiliser des expériences bidimensionnelles homonucléaires <sup>1</sup>H – <sup>1</sup>H pour différencier les protons et extraire par la suite l'ensemble des informations essentielles dans l'analyse structurale.

## II.3 Expériences 2D

Une expérience à deux dimensions (2D) RMN est utilisée pour déterminer les corrélations entre deux noyaux  $^1\text{H}$ – $^1\text{H}$ , par exemple. Ces séquences de RMN comportent quatre périodes distinctes qui sont :

- La préparation pendant laquelle le système retourne vers son état d'équilibre. Cette période peut être relativement longue ( $>5 \cdot T_1$ ),  $T_1$  étant le temps de relaxation longitudinale des noyaux de la molécule ;
- L'évolution pendant le temps ( $t_1$ ) où le système de spins évolue sous l'influence de l'hamiltonien d'évolution. L'incrément de la valeur  $t_1$  permet de créer la seconde dimension au spectre RMN ;
- Le mélange qui permet de trier l'information RMN recherchée (couplage scalaire, effet nOe, etc...) en transférant les informations entre les noyaux par l'intermédiaire d'impulsions et de délais ;
- L'acquisition pendant laquelle le système de spins évolue sous l'influence de l'hamiltonien d'évolution et le signal est enregistré en fonction de  $t_2$ .

Un spectre RMN 2D est constitué d'un ensemble de FID (*Free Induction Decay*) dépendant de  $t_1$  et de  $t_2$  :  $S(t_1, t_2)$ . En appliquant deux transformées de Fourier dans les deux dimensions, on obtient alors un spectre 2D dans le domaine des fréquences.

Parmi les expériences homonucléaires  $^1\text{H}$ – $^1\text{H}$ , la *CORrelation Spectroscopy* (COSY), la *TOTAL Correlated Spectroscopy* (TOCSY) et la *Nuclear Overhauser Enhancement Spectroscopy* (NOESY) sont les expériences indispensables pour réaliser l'attribution des signaux  $^1\text{H}$  des différents acides aminés d'un peptide ou d'une protéine.

### II.3.1 Expérience CORrelation Spectroscopy (COSY)

L'expérience COSY permet d'identifier les corrélations homonucléaires  $^1\text{H}$ – $^1\text{H}$  dues à des couplages scalaires (par liaison). Elle permet donc l'observation des corrélations entre le proton amide et le proton  $\alpha$ , puis entre le proton  $\alpha$  et les protons  $\beta$  et ainsi de suite tout au long de la chaîne latérale sur un même acide aminé (Figure 100). L'enchaînement des protons de la chaîne latérale de chaque acide aminé peut ainsi être lu par une information dite « en escalier ».

L'efficacité du transfert de cohérence diminue fortement avec l'augmentation de la largeur de raie (qui est liée au poids moléculaire ou au solvant utilisé). L'expérience COSY est donc

presque exclusivement utilisée pour les peptides. Les positions des pics croisés dans le COSY sont caractéristiques des acides aminés (Figure 100).

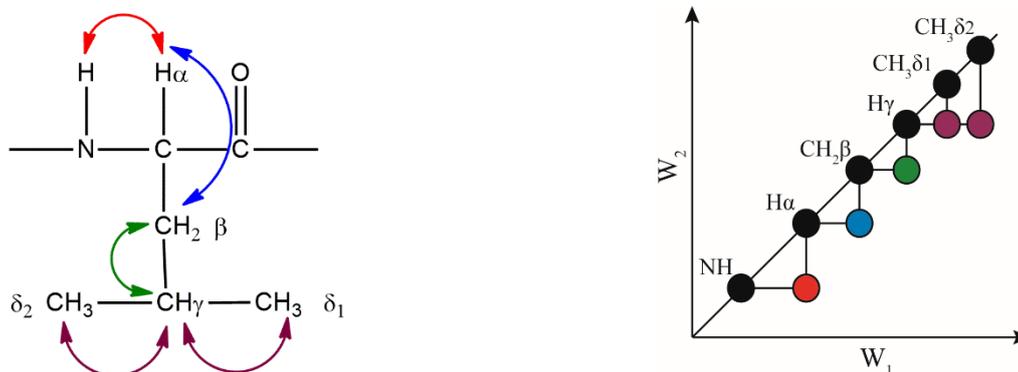


Figure 100 Couplages scalaires  $^3J$  observés pour un résidu leucine dans une expérience COSY.

### II.3.2 Expérience Total Correlation Spectroscopy (TOCSY)

Une expérience TOCSY contient tous les pics croisés dus aux protons du même système de spin (Figure 101). Cette expérience est donc utilisée afin de compléter les informations obtenues par l'expérience COSY.

Les protons de différents acides aminés appartiennent toujours à différents systèmes de spin, car il n'y a pas de couplage scalaire à travers la liaison amide (Figure 101). Certains acides aminés consistent en un système de spin unique (Ile), certains en contiennent deux (par exemple la phénylalanine Phe) ou trois (par exemple le tryptophane Trp).

L'analyse des systèmes de spin permet de décider à quel type d'acides aminés le système de spin fait partie (Figure 101). Les critères possibles sont les suivants : présence ou absence de groupes méthyle, longueur du système de spin, positions des déplacements chimiques dans le système de spin. Le temps de mélange ( $\tau_m$ ) (un paramètre qui peut être modifié lors de la configuration de l'expérience) détermine si des corrélations voisines (par exemple, pour un TOCSY de 12 ms, qui affiche des informations similaires à COSY) ou des corrélations à longue portée (par exemple, pour un TOCSY de 80 ms) sont détectées. Une telle analyse est effectuée dans une région du spectre qui affiche des moindres chevauchements. Cette partie est généralement la région contenant les protons amides. Les chevauchements possibles peuvent être supprimés en enregistrant un deuxième ensemble de spectres à une température légèrement différente.

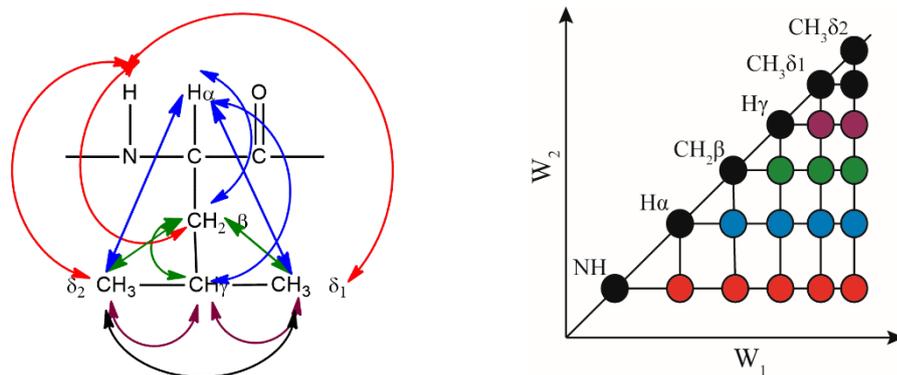


Figure 101 Couplages scalaires observés pour un résidu leucine dans une expériences TOCSY.

### II.3.3 Expérience Nuclear Overhauser Enhancement Spectroscopy (NOESY)

Les pics croisés dans le NOESY sont dus à des couplages dipolaires résultant d'interactions des noyaux dans l'espace ( $<5\text{\AA}$ ) et ne dépendent donc que de la distance, et non du nombre de liens intermédiaires. L'expression du couplage dipolaire est possible par un transfert de l'aimantation pendant le temps de mélange  $\tau_m$ . Pour un acide aminé donné, trois types de NOE sont observables sur un spectre NOESY : (i) les NOE intra-résidu, qui s'expriment entre les protons d'un même acide aminé, ces pics NOE sont identiques à ceux d'un spectre TOCSY, (ii) les NOE séquentiels de type  $d_{\alpha N}(i, i+1)$ ,  $d_{NN}(i, i+1)$  et  $d_{\beta N}(i, i+1)$  qui permettent de reconstituer l'enchaînement de la séquence peptidique, (iii) les NOE qui sont entre des protons de moyennes et longues distances (Figure 102).

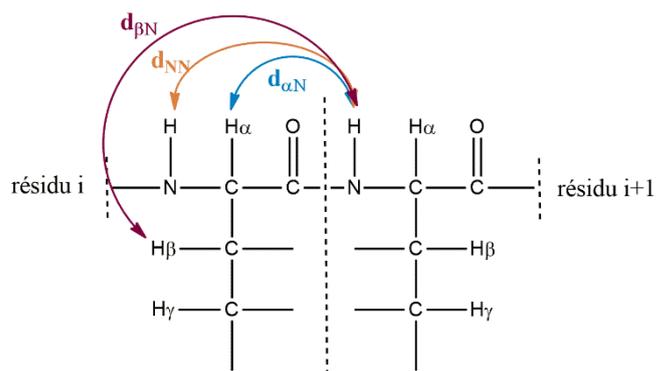


Figure 102 Deux résidus consécutifs avec les effets NOE séquentiels principaux.

L'intensité d'un pic NOE  $\eta$  varie selon plusieurs paramètres caractéristiques de la molécule étudiée. Elle dépend de la distance séparant les deux protons ( $r^{-6}$ ) ainsi que du temps de corrélation  $\tau_c$  de la molécule en fonction du champ magnétique  $\omega$  du spectromètre utilisé.

$$\eta = f(r^{-6}; \omega \tau_c)$$

La forte dépendance de l'intensité du pic croisé sur la distance de séparation explique pourquoi ce paramètre est le plus utile pour la détermination de la structure.

L'ensemble des expériences COSY, TOCSY et NOESY permettent donc d'attribuer les signaux protons d'un peptide ( $M < 10$  kDa). En 1986, Wüthrich a développé une méthode d'attribution pour l'étude structurale d'un polypeptide par RMN.<sup>118</sup>

## II.4 Méthode d'attribution des signaux de protons

La première étape de la méthode d'attribution des spectres RMN d'un peptide consiste en l'identification des systèmes de spins de chaque acide amine *via* les couplages scalaires à l'aide des expériences COSY et TOCSY. La seconde étape permet de déterminer la position de chaque système de spins dans la séquence *via* l'observation des couplages dipolaires à l'aide des expériences NOESY.

### II.4.1 L'identification des systèmes de spins

Les systèmes de spin sont tout d'abord classés. Ceci est fait en sélectionnant la région TOCSY des protons amides (environ 6,5-12 ppm) en abscisse et la région aliphatique (0-5,0 ppm) dans en ordonnée. Les systèmes de spin s'aligneront verticalement dans cette région, et le nombre et la position des pics donneront des informations précieuses.

La longueur des systèmes de spins (c'est-à-dire, le nombre de pics) est la première information. En effet, il existe des systèmes de spins courts (par exemple Ser, Figure 103) et des systèmes des spins longs (par exemple Lys, Figure 103). En outre, la répartition des pics est caractéristique pour la plupart des acides aminés (Figure 103). Certains acides aminés possèdent ainsi une topologie unique et sont donc facilement identifiables (Ala, Gly, Pro, Thr, Ser, Val, Arg) comme présenté dans la Figure 103. D'autres acides aminés sont plus difficiles

à identifier sans ambiguïté car ils possèdent des motifs similaires (Asp et Asn, Glu et Gln, Tyr, Phe...) comme présenté dans la Figure 103.

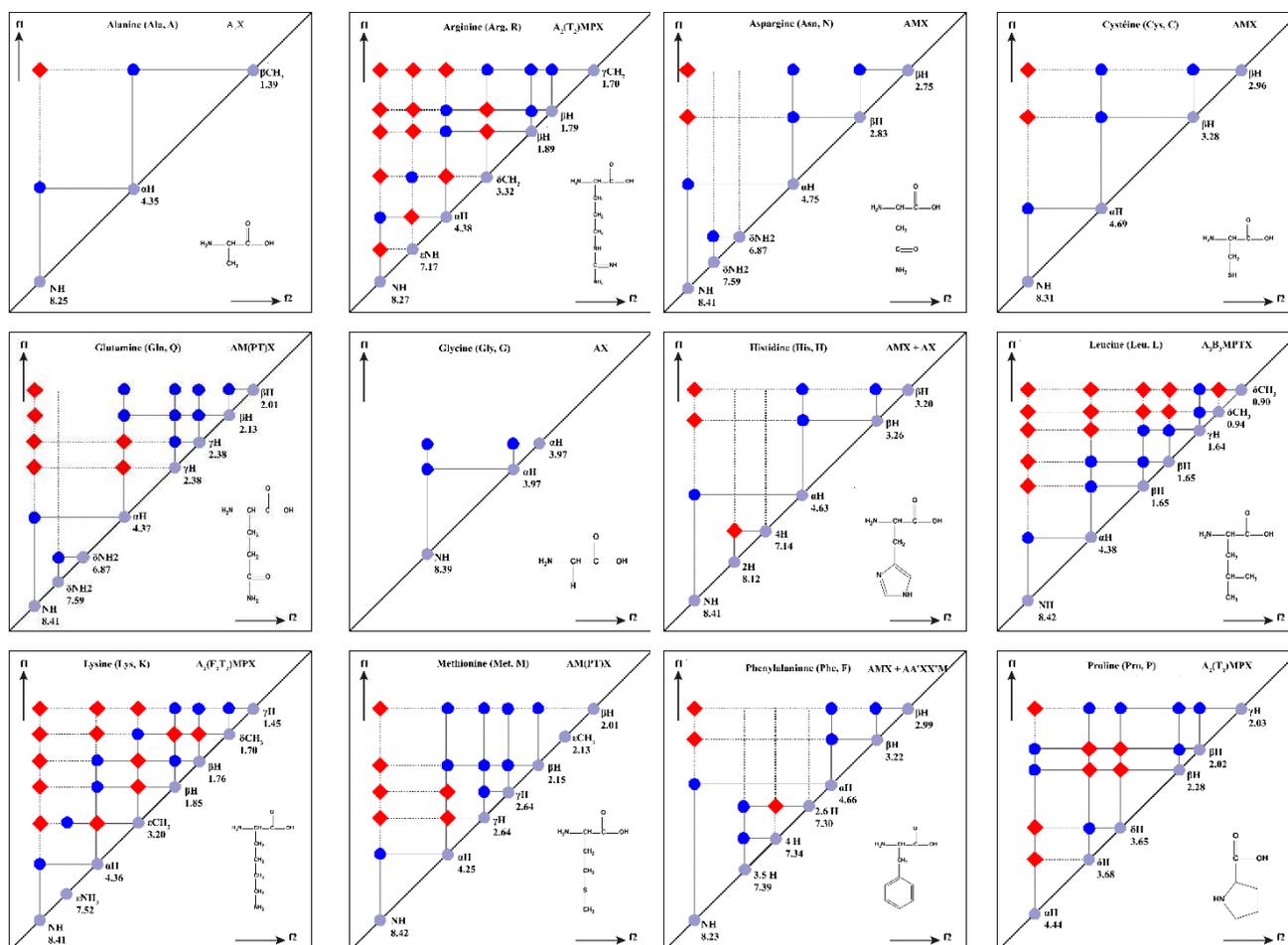


Figure 103 Représentation des couplages scalaires entre protons de quelques acides aminés.

## II.4.2 Attribution séquentielle

L'attribution séquentielle des signaux de résonance est la deuxième étape d'identification de la structure d'un peptide. Elle consiste en la connexion des systèmes de spin. Considérant le fait que des couplages scalaires ne se produiront jamais entre des protons de différents acides aminés, il est clair que les NOE doivent être utilisés à cette fin. Des NOEs peuvent être trouvés entre tous les protons proches dans l'espace pour une distance inférieure à 5,0 Å. À ce stade, il subsiste généralement des superpositions ou des ambiguïtés d'attributions. Il est alors habituel de faire l'acquisition de spectres COSY, TOCSY et NOESY à différentes températures afin d'évaluer la meilleure dispersion spectrale possible. Enfin, pour une température donnée, le temps de mélange  $\tau_m$  des expériences TOCSY et NOESY peut également être sensiblement

modifié. Une fois l'attribution séquentielle achevée, l'ultime étape consiste à distinguer les protons  $\beta_1/\beta_2$ ,  $\gamma_1/\gamma_2$ ,  $\delta_1/\delta_2$ , à l'aide des couplages scalaires des spectres COSY et les protons des chaînes aromatiques à l'aide des couplages dipolaires des spectres NOESY. Dans le cas particulier de la proline, des corrélations dipolaires spécifiques  $\alpha\alpha$  ( $X_i$ ,  $Pro_{i+1}$  dans la Figure 104) (isomère cis) ou  $\alpha\delta$  ( $X_i$ ,  $Pro_{i+1}$  dans la Figure 104) (isomère trans) permettent de distinguer la forme prédominante dans la séquence (cis ou trans).

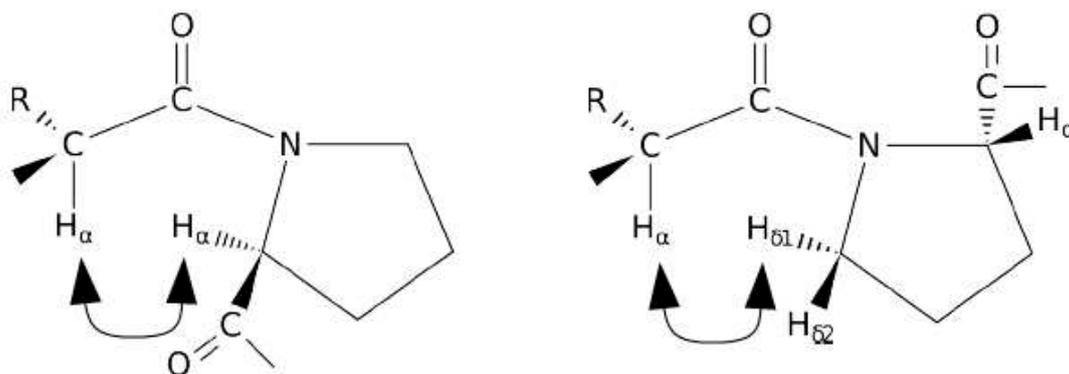


Figure 104 Isoméris cis et trans de la proline et interactions dipolaires.

*Annexes 2 : Déplacements chimiques*

**I. Tableaux des déplacements chimiques du peptide modèle IIGD**

**I.1 Spectres enregistrés à 298 K tampon phosphate/TFE (75/25 v/v, 20 mM, pH 5,3)**

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>K<sup>1</sup></b>	-	3,90	1,83 1,73	1,59 1,39	δ CH <sub>2</sub> : 1,62 – 1,50 ε CH <sub>2</sub> : 2,92 – 2,75 ζ NH <sub>3</sub> <sup>+</sup> : 7,56
<b>L<sup>2</sup></b>	8,50	4,45	1,55 1,46	0,80	δ CH <sub>3</sub> : 1,10 - 0,90
<b>V<sup>3</sup></b>	8,17	4,3	1,9	CH <sub>3</sub> : 0,83 CH <sub>3</sub> : 0,97	
<b>I<sup>4</sup></b>	8,18	4,2	1,73	CH <sub>3</sub> : 1,39 CH <sub>2</sub> : 1,05	δ1 CH <sub>3</sub> : 0,80
<b>N<sup>5</sup></b>	8,51	4,55	2,90 2,72		δ NH <sub>2</sub> : 6,7 -7,4
<b>G<sup>6</sup></b>	8,33	3,95 3,75			
<b>K<sup>7</sup></b>	8,02	4,3	1,82 1,70	1,42 1,10	δ CH <sub>2</sub> : 1,25 – 1,10 ε CH <sub>2</sub> : 2,70 – 2,60 ζ NH <sub>3</sub> <sup>+</sup> : 7,50
<b>T<sup>8</sup></b>	8,08	4,15	2,16	CH <sub>3</sub> : 1,15	
<b>L<sup>9</sup></b>	8,23	4,35	1,58 1,39	0,93	δ CH <sub>3</sub> : 0,88 - 0,89
<b>K<sup>10</sup></b>	8,06	4,4	1,82 1,80	1,40 1,30	δ CH <sub>2</sub> : 1,65 -1,45 ε CH <sub>2</sub> : 2,95 -2,65 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>G<sup>11</sup></b>	8,19	3,94 3,93			
<b>E<sup>12</sup></b>	8,25	4,42	2,09 1,91	2,25	
<b>T<sup>13</sup></b>	8,15	4,40	4,30	CH <sub>3</sub> : 1,2	
<b>T<sup>14</sup></b>	7,69	4,15	4,20	CH <sub>3</sub> : 1,15	

**I.2 Spectres enregistrés à 278 K tampon phosphate/TFE (75/25 v/v, 20 mM, pH 5,3)**

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>K<sup>1</sup></b>	-	3,95	1,82 1,70	1,35 1,10	δ CH <sub>2</sub> : 1,61-1,45 ε CH <sub>2</sub> : 2,90-2,82 ζ NH <sub>3</sub> <sup>+</sup> : na
<b>L<sup>2</sup></b>	8,65	4,49	1,51 1,49	0,95	δ CH <sub>3</sub> : 0,82-0,80
<b>V<sup>3</sup></b>	8,38	4,32	1,92	CH <sub>3</sub> : 0,50 CH <sub>3</sub> : 0,20	
<b>I<sup>4</sup></b>	8,37	4,19	1,72	CH <sub>2</sub> : 1,39 CH <sub>3</sub> : 1,15	δ1 CH <sub>3</sub> : 0,77
<b>N<sup>5</sup></b>	8,74	4,55	2,90 2,70		δ NH <sub>2</sub> : 6,75 - 7,52
<b>G<sup>6</sup></b>	8,52	3,93 3,70			
<b>K<sup>7</sup></b>	8,13	4,30	1,80 1,70	1,40 1,15	δ CH <sub>2</sub> : 1,63 -1,25 ε CH <sub>2</sub> : 2,92 - 2,72 ζ NH <sub>3</sub> <sup>+</sup> : 7,54
<b>T<sup>8</sup></b>	8,19	4,35	4,16	CH <sub>3</sub> : 1,13	
<b>L<sup>9</sup></b>	8,40	4,36	1,60 1,55	0,90	δ CH <sub>3</sub> : 0,85 -0,81
<b>K<sup>10</sup></b>	8,14	4,40	1,80 1,70	1,40 1,35	δ CH <sub>2</sub> : 1,65 - 1,52 ε CH <sub>2</sub> : 2,92 -2,85 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>G<sup>11</sup></b>	8,32	3,95 3,90			
<b>E<sup>12</sup></b>	8,36	4,42	2,10 1,93	2,25	
<b>T<sup>13</sup></b>	8,30	4,44	4,25	CH <sub>3</sub> : 1,15	
<b>T<sup>14</sup></b>	7,81	4,22	4,10	CH <sub>3</sub> : 1,12	

**I.3 Spectres enregistrés à 278 K tampon phosphate/TFE (75/25 v/v, 20 mM, pH 3,5)**

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>K<sup>1</sup></b>	-	3,93	1,82 1,75	1,35 1,15	δ CH <sub>2</sub> : 1,62 – 1,75 ε CH <sub>2</sub> : 2,90-2,80 ζ NH <sub>3</sub> <sup>+</sup> : na
<b>L<sup>2</sup></b>	8,61	4,42	1,51 1,47	0,95	δ CH <sub>3</sub> : 0,84 – 0,75
<b>V<sup>3</sup></b>	8,36	4,18	1,91	CH <sub>3</sub> : 0,84 CH <sub>3</sub> : 0,78	
<b>I<sup>4</sup></b>	8,32	4,2	1,73	CH <sub>2</sub> : 1,38 CH <sub>3</sub> : 1,05	δ1 CH <sub>3</sub> : 0,78
<b>N<sup>5</sup></b>	8,62	4,6	2,85 2,72		δ NH <sub>2</sub> : 6,80 – 6,55
<b>G<sup>6</sup></b>	8,53	3,87 3,76			
<b>K<sup>7</sup></b>	8,19	4,22	1,80 1,70	1,38 1,10	δ CH <sub>2</sub> : 1,60 – 1,45 ε CH <sub>2</sub> : 2,90 – 2,68 ζ NH <sub>3</sub> <sup>+</sup> : 7,54
<b>T<sup>8</sup></b>	8,1	4,3	4,15	CH <sub>3</sub> : 1,13	
<b>L<sup>9</sup></b>	8,24	4,35	1,57 1,45	0,90	δ CH <sub>3</sub> : 0,85 -0,80
<b>K<sup>10</sup></b>	8,16	4,35	1,80 1,75	1,38 1,16	δ CH <sub>2</sub> : 1,61 – 1,49 ε CH <sub>2</sub> : 2,92 -2,86 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>G<sup>11</sup></b>	8,32	3,90 3,92			
<b>E<sup>12</sup></b>	8,19	4,44	2,12 1,95	2,39	
<b>T<sup>13</sup></b>	8,25	4,5	4,21	CH <sub>3</sub> : 1,16	
<b>T<sup>14</sup></b>	7,9	4,24	4,05	CH <sub>3</sub> : 1,12	

**I.4 Spectres enregistrés à 278 K tampon phosphate/MeOH (75/25 v/v, 20 mM, pH 3,5)**

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>K<sup>1</sup></b>	-	3,98	1,80 1,75	1,25 1,10	δ CH <sub>2</sub> : 1,52 – 1,35 ε CH <sub>2</sub> : 2,85 – 2,75 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>L<sup>2</sup></b>	8,80	4,52	1,58 1,45	0,93	δ CH <sub>3</sub> : 0,84 – 0,76
<b>V<sup>3</sup></b>	8,56	4,20	2,02	CH <sub>3</sub> : 0,96 CH <sub>3</sub> : 0,90	
<b>I<sup>4</sup></b>	8,54	4,30	1,84	CH <sub>2</sub> : 1,48 CH <sub>3</sub> : 1,17	δ1 CH <sub>3</sub> : 0,95
<b>N<sup>5</sup></b>	8,85	4,70	2,94 2,80		δ NH <sub>2</sub> : 6,70-6,53
<b>G<sup>6</sup></b>	8,66	3,91 3,96			
<b>K<sup>7</sup></b>	8,33	4,45	1,90 1,82	1,60 1,45	δ CH <sub>2</sub> : 1,42 – 1,15 ε CH <sub>2</sub> : 3,05 – 2,85 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>T<sup>8</sup></b>	8,29	4,35	4,25	CH <sub>3</sub> : 1,23	
<b>L<sup>9</sup></b>	8,46	4,40	1,68 1,55	0,91	δ CH <sub>3</sub> : 0,94 – 0,91
<b>K<sup>10</sup></b>	8,44	4,30	1,89 1,79	1,47 1,35	δ CH <sub>2</sub> : 1,54 – 1,46 ε CH <sub>2</sub> : 3,01 – 2,94 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>G<sup>11</sup></b>	8,58	3,96 3,0			
<b>E<sup>12</sup></b>	8,35	4,54	2,20 2,04	2,49	
<b>T<sup>13</sup></b>	8,49	4,50	4,21	CH <sub>3</sub> : 1,26	
<b>T<sup>14</sup></b>	8,1	4,35	4,27	CH <sub>3</sub> : 1,12	

## II. Tableaux des déplacements chimiques du peptide modèle 8ADH

### II.1 Spectres enregistrés à 278 K tampon phosphate/TFE (50/50 v/v, 20 mM, pH 6,0)

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>F<sup>1</sup></b>	-	4,10	2,90 2,79		δ CH : 7,50 ε CH : 7,22 ζ CH : 7,00
<b>I<sup>2</sup></b>	8,12	4,30	1,68	1,35 CH <sub>3</sub> : 1,20	δ1 CH <sub>3</sub> : 0,85
<b>C<sup>3</sup></b>	8,80	5,35	2,90 2,75		
<b>R<sup>4</sup></b>	8,01	4,17	1,90 1,82	1,71 1,61	δ CH <sub>2</sub> : 3,17 ε NH : 7,25
<b>G<sup>5</sup></b>	7,90	4,00 3,90			
<b>K<sup>6</sup></b>	7,85	4,45	1,60 1,52	1,30 1,25	δ CH <sub>2</sub> : 1,15 – 1,10 ε CH <sub>2</sub> : 2,55 – 2,42 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>P<sup>7</sup></b>	-	4,41	2,36 1,82	2,08 2,02	δ CH <sub>2</sub> : 3,84 -3,67
<b>I<sup>8</sup></b>	8,39	4,20	1,75	CH <sub>2</sub> :1,50 CH <sub>3</sub> :1,37	δ1 CH <sub>3</sub> : 0,89

### III. Tableaux des déplacements chimiques du peptide modèle 1MOL

#### III.1 Spectres enregistrés à 278 K tampon phosphate/TFE (75/25 v/v, 20 mM, pH 5,3)

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>K<sup>1</sup></b>	-	4,20	1,79 1,70	1,50	δ CH <sub>2</sub> : 1,60 ε CH <sub>2</sub> : 3,00 – 2,85 ζ NH <sub>3</sub> <sup>+</sup> : 7,45
<b>T<sup>2</sup></b>	8,90	4,46	4,33	CH <sub>3</sub> : 1,21	
<b>I<sup>3</sup></b>	8,60	4,08	1,78	1,32 CH <sub>3</sub> : 1,11	δ1 CH <sub>3</sub> : 0,83
<b>Y<sup>4</sup></b>	8,1	4,5	3,06 2,92		δ CH : 7,16 ε CH : 6,85
<b>E<sup>5</sup></b>	8,22	4,16	2,10 2,02	2,38	
<b>N<sup>6</sup></b>	8,50	4,55	2,83 2,65		δ NH <sub>2</sub> : 6,65-6,42
<b>E<sup>7</sup></b>	8,47	4,15	2,12 2,00	2,48 2,40	
<b>R<sup>8</sup></b>	8,18	4,17	1,80 1,78	1,58 1,50	δ CH <sub>2</sub> : 3,13 ε NH : 7,04
<b>E<sup>9</sup></b>	8,13	4,21	2,46 2,33	2,07	
<b>I<sup>10</sup></b>	8,10	4,03	1,88	CH <sub>3</sub> : 1,55 CH <sub>3</sub> : 1,19	δ1 CH <sub>3</sub> : 0,84
<b>K<sup>11</sup></b>	8,33	4,26	1,74 1,63	1,37	δ CH <sub>2</sub> : 1,43 ε CH <sub>2</sub> : 2,91-2,85 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>G<sup>12</sup></b>	8,28	3,99 3,85			
<b>Y<sup>13</sup></b>	7,68	4,41	3,11 2,93		δ CH : 7,07 ε CH : 6,83

### III.2 Spectres enregistrés à 278 K tampon phosphate/MeOH (75/25 v/v, 20 mM, pH 3,5)

Résidus	H <sub>N</sub>	H <sub>α</sub>	H <sub>β</sub>	H <sub>γ</sub>	Autres
<b>K<sup>1</sup></b>	-	4,17	1,75 1,68	1,55	δ CH <sub>2</sub> : 1,68 ε CH <sub>2</sub> : 3,00 – 2,87 ζ NH <sub>3</sub> <sup>+</sup> : 7,55
<b>T<sup>2</sup></b>	8,77	4,44	4,30	CH <sub>3</sub> : 1,18	
<b>I<sup>3</sup></b>	8,58	4,13	1,80	1,36 CH <sub>3</sub> : 1,14	δ1 CH <sub>3</sub> : 0,84
<b>Y<sup>4</sup></b>	8,23	4,53	3,06 2,90		δ CH : 7,10 ε CH : 6,86
<b>E<sup>5</sup></b>	8,28	4,23	2,14 2,04	2,42	
<b>N<sup>6</sup></b>	8,50	4,59	2,87 2,64		δ NH <sub>2</sub> : 6,60
<b>E<sup>7</sup></b>	8,47	4,18	2,13 2,00	2,51 2,45	
<b>R<sup>8</sup></b>	8,23	4,18	1,85 1,78	1,58 1,46	δ CH <sub>2</sub> : 3,16 ε NH : 7,03
<b>E<sup>9</sup></b>	8,17	4,25	2,1 2,08	2,52 2,42	
<b>I<sup>10</sup></b>	8,10	4,03	1,87	CH <sub>2</sub> : 1,54 CH <sub>3</sub> : 1,17	δ1 CH <sub>3</sub> : 0,93
<b>K<sup>11</sup></b>	8,36	4,28	1,78 1,64	1,45 1,38	δ CH <sub>2</sub> : 1,30 ε CH <sub>2</sub> : 2,98 ζ NH <sub>3</sub> <sup>+</sup> : -
<b>G<sup>12</sup></b>	8,32	3,99 3,86			
<b>Y<sup>13</sup></b>	7,83	4,48	3,13 2,95		δ CH : 7,05 ε CH : 6,81

# **Bibliographie**



## ***Bibliographie***

1. Easson LH, Stedman E. Studies on the relationship between chemical constitution and physiological action. *Biochem J.* 1933;27(4):1257-1266.
2. Robinson JA. Beta-hairpin peptidomimetics: design, structures and biological activities. *Acc Chem Res.* 2008;41(10):1278-1288.
3. Beltramo M, Robert V, Galibert M, et al. Rational Design of Triazololipopeptides Analogs of Kisspeptin Inducing a Long-Lasting Increase of Gonadotropins. *J Med Chem.* 2015;58(8):3459-3470.
4. Vagner J, Qu H, Hruby VJ. Peptidomimetics, a synthetic tool of drug discovery. *Curr Opin Chem Biol.* 2008;12(3):292-296.
5. Pauling L, Corey RB, Branson HR. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A.* 1951;37(4):205-211.
6. Pauling L, Corey RB. The Pleated Sheet, A New Layer Configuration of Polypeptide Chains. *Proc Natl Acad Sci.* 1951;37(5):251-256.
7. Kaur H, Garg A, Raghava GPS. PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept Lett.* 2007;14(7):626-631.
8. Videau LL, Arendall WB, Richardson JS. The cis-Pro touch-turn: a rare motif preferred at functional sites. *Proteins.* 2004;56(2):298-309.
9. Chatenet D, Dubessy C, Leprince J, et al. Structure-activity relationships and structural conformation of a novel urotensin II-related peptide. *Peptides.* 2004;25(10):1819-1830.
10. Mori M, Sugo T, Abe M, et al. Urotensin II is the endogenous ligand of a G-protein-coupled orphan receptor, SENR (GPR14). *Biochem Biophys Res Commun.* 1999;265(1):123-129.
11. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* 1958;181(4610):662-666.
12. Linderstrøm-Lang KU, Schellman JA. *The Enzymes.* Vol 1. 2nd ed.; 1959.
13. Lehninger AL, Nelson DL, Cox MM. *Lehninger Principles of Biochemistry.* 6th ed. New York: W.H. Freeman; 2013.
14. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. *Adv Protein Chem.* 1968;23:283-438.
15. Barlow DJ, Thornton JM. Helix geometry in proteins. *J Mol Biol.* 1988;201(3):601-619.
16. Karpen ME, de Haseth PL, Neet KE. Differences in the amino acid distributions of 3(10)-helices and alpha-helices. *Protein Sci Publ Protein Soc.* 1992;1(10):1333-1342.
17. Némethy G, Phillips DC, Leach SJ, Scheraga HA. A Second Right-handed Helical Structure with the Parameters of the Pauling–Corey  $\alpha$ -helix. *Nature.* 1967;214(5086):363-365.
18. Bolin KA, Millhauser GL.  $\alpha$  and  $3_{10}$ : The Split Personality of Polypeptide Helices. *Acc Chem Res.* 1999;32(12):1027-1033.
19. Huggins ML. The Structure of Fibrous Proteins. *Chem Rev.* 1943;32(2):195-218.
20. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;22(12):2577-2637.
21. Donohue J. Hydrogen Bonded Helical Configurations of the Polypeptide Chain. *Proc Natl Acad Sci U S A.* 1953;39(6):470-478.
22. Vainshtein BK, Melik-Adamyanyan WR, Barynin VV, et al. Three-dimensional structure of catalase from *Penicillium vitale* at 2.0 Å resolution. *J Mol Biol.* 1986;188(1):49-61.

23. Low BW, Grenville-Wells HJ. Generalized Mathematical Relationships for Polypeptide Chain Helices: The Coordinates of the II Helix. *Proc Natl Acad Sci U S A*. 1953;39(8):785-801.
24. Rohl CA, Doig AJ. Models for the 3(10)-helix/coil, pi-helix/coil, and alpha-helix/3(10)-helix/coil transitions in isolated peptides. *Protein Sci Publ Protein Soc*. 1996;5(8):1687-1696.
25. Weaver TM. The pi-helix translates structure into function. *Protein Sci Publ Protein Soc*. 2000;9(1):201-206.
26. Quinkert G, Egert E, Griesinger C. *Peptides: Synthesis, Structure and Applications*. Gutte, B., Ed. Academic Press: San Diego; 1995.
27. Degrado WF. Design of peptides and proteins. *Adv Protein Chem*. 1988;39:51-124.
28. Richardson JS, Getzoff ED, Richardson DC. The beta bulge: a common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci U S A*. 1978;75(6):2574-2578.
29. Chan AW, Hutchinson EG, Harris D, Thornton JM. Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci Publ Protein Soc*. 1993;2(10):1574-1590.
30. Tyndall JDA, Pfeiffer B, Abbenante G, Fairlie DP. Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem Rev*. 2005;105(3):793-826.
31. Chou KC. Prediction of tight turns and their types in proteins. *Anal Biochem*. 2000;286(1):1-16.
32. Ruiz-Gómez G, Tyndall JDA, Pfeiffer B, Abbenante G, Fairlie DP. Update 1 of: Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem Rev*. 2010;110(4):PR1-41.
33. Némethy G, Printz MP. The  $\gamma$  Turn, a Possible Folded Conformation of the Polypeptide Chain. Comparison with the  $\beta$  Turn. *Macromolecules*. 1972;5(6):755-758.
34. Matthews BW. The  $\gamma$  Turn. Evidence for a New Folded Conformation in Proteins. *Macromolecules*. 1972;5(6):818-819.
35. Milner-White EJ. Situations of gamma-turns in proteins. Their relation to alpha-helices, beta-sheets and ligand binding sites. *J Mol Biol*. 1990;216(2):386-397.
36. Rose GD, Gierasch LM, Smith JA. Turns in peptides and proteins. *Adv Protein Chem*. 1985;37:1-109.
37. Guruprasad K, Rajkumar S. Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J Biosci*. 2000;25(2):143-156.
38. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers*. 1968;6(10):1425-1436.
39. Lewis PN, Momany FA, Scheraga HA. Chain reversals in proteins. *Biochim Biophys Acta BBA - Protein Struct*. 1973;303(2):211-229.
40. Richardson JS. The anatomy and taxonomy of protein structure. *Adv Protein Chem*. 1981;34:167-339.
41. Wilmot CM, Thornton JM. Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol*. 1988;203(1):221-232.
42. Wilmot CM, Thornton JM. Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng Des Sel*. 1990;3(6):479-493.
43. Hutchinson EG, Thornton JM. A revised set of potentials for beta-turn formation in proteins. *Protein Sci Publ Protein Soc*. 1994;3(12):2207-2216.
44. Huber R, Steigemann W. Two cis-prolines in the Bence-Jones protein Rei and the cis-pro-bend. *FEBS Lett*. 1974;48(2):235-237.

45. de Brevern AG. Extension of the classical classification of  $\beta$ -turns. *Sci Rep.* 2016;6:33191.
46. Sibanda BL, Thornton JM.  $\beta$ -Hairpin families in globular proteins. *Nature.* 1985;316(6024):170-174.
47. Chothia C. Conformation of twisted  $\beta$ -pleated sheets in proteins. *J Mol Biol.* 1973;75(2):295-302.
48. Isogai Y, Némethy G, Rackovsky S, Leach SJ, Scheraga HA. Characterization of multiple bends in proteins. *Biopolymers.* 1980;19(6):1183-1210.
49. Fuchs PFJ, Alix AJP. High accuracy prediction of  $\beta$ -turns and their types using propensities and multiple alignments. *Proteins Struct Funct Bioinforma.* 2005;59(4):828-839.
50. Crawford JL, Lipscomb WN, Schellman CG. The reverse turn as a polypeptide conformation in globular proteins. *Proc Natl Acad Sci U S A.* 1973;70(2):538-542.
51. Wolfenden R. Interaction of the peptide bond with solvent water: a vapor phase analysis. *Biochemistry.* 1978;17(1):201-204.
52. Hagler AT, Moulton J. Computer simulation of the solvent structure around biological macromolecules. *Nature.* 1978;272(5650):222-226.
53. Kuntz ID. Protein folding. *J Am Chem Soc.* 1972;94(11):4009-4012.
54. Guruprasad K, Rajkumar S.  $\beta$ - and  $\gamma$ -turns in proteins revisited: A new set of amino acid turn-type dependent positional preferences and potentials. *J Biosci.* 2000;25(2):143-156.
55. Natalello A, Ami D, Brocca S, Lotti M, Doglia SM. Secondary structure, conformational stability and glycosylation of a recombinant *Candida rugosa* lipase studied by Fourier-transform infrared spectroscopy. *Biochem J.* 2005;385(2):511-517.
56. Barth A, Zscherp C. What vibrations tell about proteins. *Q Rev Biophys.* 2002;35(4):369-430.
57. Arrondo JLR, Muga A, Castresana J, Goñi FM. Quantitative studies of the structure of proteins in solution by fourier-transform infrared spectroscopy. *Prog Biophys Mol Biol.* 1993;59(1):23-56.
58. Kong J, Yu S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochim Biophys Sin.* 2007;39(8):549-559.
59. Fabian H, Mäntele W. Infrared Spectroscopy of Proteins. In: Chalmers JM, Griffiths PR, eds. *Handbook of Vibrational Spectroscopy.* Chichester, UK: John Wiley & Sons, Ltd; 2006:s8201.
60. Johnson WC. Protein secondary structure and circular dichroism: a practical guide. *Proteins.* 1990;7(3):205-214.
61. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta.* 2005;1751(2):119-139.
62. Provencher SW, Glöckner J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry.* 1981;20(1):33-37.
63. Reed J, Reed TA. A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism. *Anal Biochem.* 1997;254(1):36-40.
64. Wagner G, Neuhaus D, Wörgötter E, Vašák M, Kägi JHR, Wüthrich K. Nuclear magnetic resonance identification of "half-turn" and 310-helix secondary structure in rabbit liver metallothionein-2. *J Mol Biol.* 1986;187(1):131-135.
65. McPherson A. *Crystallization of Biological Macromolecules.* Cold Spring Harbor Laboratory Press; 1999.
66. McPherson A, Malkin AJ, Kuznetsov YG. The science of macromolecular crystallization. *Struct Lond Engl 1993.* 1995;3(8):759-768.
67. Parker MW. Protein Structure from X-Ray Diffraction. *J Biol Phys.* 2003;29(4):341-362.

68. Williamson MP, Havel TF, Wüthrich K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol.* 1985;182(2):295-315.
69. Braun W, Go N. Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J Mol Biol.* 1985;186(3):611-626.
70. Unger VM. Electron cryomicroscopy methods. *Curr Opin Struct Biol.* 2001;11(5):548-554.
71. Zhou ZH, Chiu W. Determination Of Icosahedral Virus structures By Electron Cryomicroscopy At Subnanometer Resolution. In: *Advances in Protein Chemistry.* Vol 64. Elsevier; 2003:93-124.
72. Ludtke SJ, Chen D-H, Song J-L, Chuang DT, Chiu W. Seeing GroEL at 6 Å Resolution by Single Particle Electron Cryomicroscopy. *Structure.* 2004;12(7):1129-1136.
73. Heymann JB, Cheng N, Newcomb WW, Trus BL, Brown JC, Steven AC. Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. *Nat Struct Biol.* 2003;10(5):334-341.
74. Heymann JB, Conway JF, Steven AC. Molecular dynamics of protein complexes from four-dimensional cryo-electron microscopy. *J Struct Biol.* 2004;147(3):291-301.
75. Jones DT. Successful ab initio prediction of the tertiary structure of NK-lysin using multiple sequences and recognized supersecondary structural motifs. *Proteins.* 1997;Suppl 1:185-191.
76. Klepeis JL, Floudas CA. Ab initio prediction of helical segments in polypeptides. *J Comput Chem.* 2002;23(2):245-266.
77. Srinivasan R, Rose GD. Ab initio prediction of protein structure using LINUS. *Proteins.* 2002;47(4):489-495.
78. Ianeselli A, Orioli S, Spagnoli G, et al. Atomic Detail of Protein Folding Revealed by an Ab Initio Reappraisal of Circular Dichroism. *J Am Chem Soc.* 2018;140(10):3674-3682.
79. Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol.* 2007;5(1):17.
80. Runge E, Gross EKV. Density-Functional Theory for Time-Dependent Systems. *Phys Rev Lett.* 1984;52:997-1000.
81. Marques MAL, Ullrich C, Nogueira F, Rubio A, Burke K, Gross EKV, eds. *Time-Dependent Density Functional Theory.* Berlin Heidelberg: Springer-Verlag; 2006.
82. List NH, Jensen HJA, Kongsted J. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. *Phys Chem Chem Phys.* 2016;18(15):10070-10080.
83. Olsen JM, Aidas K, Kongsted J. Excited States in Solution through Polarizable Embedding. *J Chem Theory Comput.* 2010;6(12):3721-3734.
84. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinforma.* 2006;Chapter 5:Unit-5.6.
85. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234(3):779-815.
86. Arnold K, Kiefer F, Kopp J, et al. The Protein Model Portal. *J Struct Funct Genomics.* 2009;10(1):1-8.
87. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* 2009;4(3):363-371.
88. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc.* 2007;1(6):2876-2890.
89. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. - PubMed - NCBI. *Curr Protein Pept Sci.* 2000;1(4):349-384.

90. Johnson WC. Secondary Structure of Proteins Through Circular Dichroism Spectroscopy. *Annu Rev Biophys Biophys Chem.* 1988;17(1):145-166.
91. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc.* 2006;1(6):2876-2890.
92. Brahms S, Brahms J. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J Mol Biol.* 1980;138(2):149-178.
93. Greenfield N, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry.* 1969;8(10):4108-4116.
94. Rosenkranz H, Scholtan W. An improved method for the determination of helical protein from circular dichroism measurements. *Hoppe Seylers Z Physiol Chem.* 1971;352(7):896-904.
95. Chen YH, Yang JT, Chau KH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry.* 1974;13(16):3350-3359.
96. Lazo ND, Downing DT. A Mixture of  $\alpha$ -Helical and 310-Helical Conformations for Involucrin in the Human Epidermal Corneocyte Envelope Provides a Scaffold for the Attachment of Both Lipids and Proteins. *J Biol Chem.* 1999;274(52):37340-37344.
97. Milhauser G, Stenland CJ, Hanson P, van de Ven FJ. Estimating the relative populations of 3(10)-helix and alpha-helix in Ala-rich peptides: a... - Abstract - Europe PMC. *J Mol Biol.* 1997;267(4):963-974.
98. Woody RW. Studies of theoretical circular dichroism of polypeptides: Contributions of  $\beta$  turns. In: Blout ER, Bovey FA, Goodman M, Lotan N, eds. *Peptides, Polypeptides, and Proteins.* John Wiley & Sons, New York; 1974:338-350.
99. Sathyanarayana BK, Applequist J. Theoretical pi-pi absorption and circular dichroic spectra of beta-turn model peptides. *Int J Pept Protein Res.* 1986;27(1):86-94.
100. Chang CT, Wu CS, Yang JT. Circular dichroic analysis of protein conformation: inclusion of the beta-turns. *Anal Biochem.* 1978;91(1):13-31.
101. Compton LA, Johnson WC. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem.* 1986;155(1):155-167.
102. Brahms S, Brahms J, Spach G, Brack A. Identification of  $\beta$ , $\beta$ -turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. *Proc Natl Acad Sci.* 1977;74(8):3208-3212.
103. Gierasch LM, Deber CM, Madison V, Niu CH, Blout ER. Conformations of (X-L-Pro-Y)<sub>2</sub> cyclic hexapeptides. Preferred beta-turn conformers and implications for beta turns in proteins. *Biochemistry.* 1981;20(16):4730-4738.
104. Bush CA, Sarkar SK, Koppke KD. Circular dichroism of beta turns in peptides and proteins. *Biochemistry.* 1978;17(23):4951-4954.
105. Crisma M, Fasman GD, Balaram H, Balaram P. Peptide models for beta-turns. A circular dichroism study. *Int J Pept Protein Res.* 1984;23(4):411-419.
106. Bandekar J, Evans DJ, Krimm S, et al. Conformations of cyclo( L-alanyl-L-alanyl- c -aminocaproyl) and of cyclo( L-alanyl-D-alanyl- c -aminocaproyl) ; cyclized dipeptide models for specific types of beta-bends. *Int J Pept Protein Res.* 1982;19(2):187-205.
107. Drouillat B, Peggion C, Biondi B, et al. Heterochiral Ala/( $\alpha$ Me)Aze sequential oligopeptides: Synthesis and conformational study. *J Pept Sci Off Publ Eur Pept Soc.* 2019;25(5):e3165.
108. Perczel A, Hollósi M, Foxman BM, Fasman GD. Conformational analysis of pseudocyclic hexapeptides based on quantitative circular dichroism (CD), NOE, and x-ray data. The pure CD spectra of type I and type II .beta.-turns. *J Am Chem Soc.* 1991;113(26):9772-9784.
109. Perczel A, Hollósi M, Tusnády G, Fasman GD. Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Eng.* 1991;4(6):669-679.

110. Hollósi M, Kövér KE, Holly S, Radics L, Fasman GD. Beta-turns in bridged proline-containing cyclic peptide models. *Biopolymers*. 1987;26(9):1555-1572.
111. Perczel A, Fasman GD. Quantitative analysis of cyclic beta-turn models. *Protein Sci Publ Protein Soc*. 1992;1(3):378-395.
112. Greenfield NJ, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*. 1969;8(10):4108-4116.
113. Chen Y-H, Yang JT, Chau KH. Determination of the helix and  $\beta$  form of proteins in aqueous solution by circular dichroism. *Biochemistry*. 1974;13(16):3350-3359.
114. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol*. 1991;222(2).
115. Wishart DS, Sykes BD. The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J Biomol NMR*. 1994;4(2):171-180.
116. Wishart DS, Bigam CG, Holm RS, Sykes BD.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. 1995;5(1):67-81.
117. Wishart DS, Sykes BD. The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J Biomol NMR*. 1994;4(2):171-180.
118. Wüthrich K. *NMR of Proteins and Nucleic Acids*,. John Wiley & Sons, Inc., New York.; 1986.
119. Wüthrich K, Billeter M, Braun W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J Mol Biol*. 1984;180(3):715-740.
120. Wagner G, Kumar A, Wüthrich K. Systematic application of two-dimensional  $^1\text{H}$  nuclear-magnetic-resonance techniques for studies of proteins. 2. Combined use of correlated spectroscopy and nuclear Overhauser spectroscopy for sequential assignments of backbone resonances and elucidation of polypeptide secondary structures. *Eur J Biochem*. 1981;114(2):375-384.
121. Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J Chem Phys*. 1959;30(1):11-15.
122. Pardi A, Billeter M, Wüthrich K. Calibration of the angular dependence of the amide proton- $\text{C}$  alpha proton coupling constants,  $3\text{JHN}$  alpha, in a globular protein. Use of  $3\text{JHN}$  alpha for identification of helical secondary structure. *J Mol Biol*. 1984;180(3):741-751.
123. Piantini U, Sorensen OW, Ernst RR. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc*. 1982;104(24):6800-6801.
124. Foroozandeh M, Adams RW, Nilsson M, Morris GA. Ultrahigh-Resolution Total Correlation NMR Spectroscopy. *J Am Chem Soc*. 2014;136(34):11867-11869.
125. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963;7:95-99.
126. Lovell SC, Davis IW, Arendall WB, et al. Structure validation by  $\text{C}\alpha$  geometry: phi,psi and  $\text{C}\beta$  deviation. *Proteins*. 2003;50(3):437-450.
127. Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable  $\beta$ -hairpin in aqueous solution. *Nat Struct Biol*. 1994;1(9):584-590.
128. Maestro B, Santiveri CM, Jiménez MA, Sanz JM. Structural autonomy of a  $\beta$ -hairpin peptide derived from the pneumococcal choline-binding protein LytA. *Protein Eng Des Sel PEDS*. 2011;24(1-2):113-122.
129. Zamora-Carreras H, Maestro B, Strandberg E, Ulrich AS, Sanz JM, Jiménez MÁ. Micelle-Triggered  $\beta$ -Hairpin to  $\alpha$ -Helix Transition in a 14-Residue Peptide from a

- Choline-Binding Repeat of the Pneumococcal Autolysin LytA. *Chem Weinh Bergstr Ger.* 2015;21(22):8076-8089.
130. Searle MS, Zerella R, Williams DH, Packman LC. Native-like beta-hairpin structure in an isolated fragment from ferredoxin: NMR and CD studies of solvent effects on the N-terminal 20 residues. *Protein Eng.* 1996;9(7):559-565.
  131. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-242.
  132. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605-1612.
  133. Blanco FJ, Jimenez MA, Pineda A, Rico M, Santoro J, Nieto JL. NMR Solution Structure of the Isolated N-Terminal Fragment of Protein-G B1 Domain. Evidence of Trifluoroethanol Induced Native-Like .beta.-Hairpin Formation. *Biochemistry.* 1994;33(19):6004-6014.
  134. Roccatano D, Colombo G, Fioroni M, Mark AE. Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: A molecular dynamics study. *Proc Natl Acad Sci.* 2002;99(19):12179-12184.
  135. Myers JK, Pace CN, Scholtz JM. Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1. *Protein Sci Publ Protein Soc.* 1998;7(2):383.
  136. Luo P, Baldwin RL. Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures Back to Water. *Biochemistry.* 1997;36(27):8413-8421. doi:10.1021/bi9707133
  137. Shiraki K, Nishikawa K, Goto Y. Trifluoroethanol-induced Stabilization of the  $\alpha$ -Helical Structure of  $\beta$ -Lactoglobulin: Implication for Non-hierarchical Protein Folding. *J Mol Biol.* 1995;245(2):180-194.
  138. Nelson JW, Kallenbach NR. Stabilization of the ribonuclease S-peptide alpha-helix by trifluoroethanol. *Proteins.* 1986;1(3):211-217.
  139. Cochran AG, Tong RT, Starovasnik MA, et al. A Minimal Peptide Scaffold for  $\beta$ -Turn Display: Optimizing a Strand Position in Disulfide-Cyclized  $\beta$ -Hairpins. *J Am Chem Soc.* 2001;123(4):625-632.
  140. Santiveri CM, León E, Rico M, Jiménez MA. Context-Dependence of the Contribution of Disulfide Bonds to  $\beta$ -Hairpin Stability. *Chem - Eur J.* 2008;14(2):488-499.
  141. Russell SJ, Blandl T, Skelton NJ, Cochran AG. Stability of Cyclic  $\beta$ -Hairpins: Asymmetric Contributions from Side Chains of a Hydrogen-Bonded Cross-Strand Residue Pair. *J Am Chem Soc.* 2003;125(2):388-395.
  142. Mirassou Y, Santiveri CM, Pérez de Vega MJ, González-Muñiz R, Jiménez MA. Disulfide Bonds versus Trp $\cdots$ Trp Pairs in Irregular  $\beta$ -Hairpins: NMR Structure of Vammin Loop 3-Derived Peptides as a Case Study. *ChemBioChem.* 2009;10(5):902-910.
  143. Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks CL, Pople JA. Theoretical study of blocked glycine and alanine peptide analogs. *J Am Chem Soc.* 1991;113(16):5989-5997.
  144. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26(2):283-291.
  145. Lovell SC, Davis IW, Arendall WB, et al. Structure validation by C $\alpha$  geometry: phi,psi and C $\beta$  deviation. *Proteins.* 2003;50(3):437-450. doi:10.1002/prot.10286
  146. Brünger AT, Adams PD, Clore GM, et al. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr D Biol Crystallogr.* 1998;54(5):905-921.

147. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S a, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem.* 1983;4(2):187–217.
148. Ludvigsen S, Andersen KV, Poulsen FM. Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles. *J Mol Biol.* 1991;217(4):731-736.
149. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in <sup>15</sup>N-enriched proteins. *J Am Chem Soc.* 1993;115(17):7772-7777. doi:10.1021/ja00070a024
150. Berova N, Polavarapu P, Nakanish K, Woody R. *Comprehensive Chiroptical Spectroscopy.* 1st ed. John Wiley & Sons, Ltd; 2012.
151. Fasman GD, ed. *Circular Dichroism and the Conformational Analysis of Biomolecules.* Springer US; 1996.
152. Bakshi K, Liyanage MR, Volkin DB, Middaugh CR. Circular Dichroism of Peptides. In: Nixon AE, ed. *Therapeutic Peptides: Methods and Protocols.* Methods in Molecular Biology. Totowa, NJ: Humana Press; 2014:247-253.
153. Sreerama N, Woody RW. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.* 2004;383:318-351.
154. Woody RW. Circular Dichroism of Peptides. In: *Conformation in Biology and Drug Design.* Elsevier; 1985:15-114.
155. Kobayashi N, Muranaka A, Mack J. Circular Dichroism of Biomolecules. In: *Circular Dichroism and Magnetic Circular Dichroism Spectroscopy for Organic Chemists.* Cambridge: Royal Society of Chemistry; 2011:142-149.
156. Woody R. Studies of theoretical circular dichroism of polypeptides: contribution of  $\beta$  turns. *Int J Peptide Protein Res.* 1974;6.
157. Calais J-L. Density-functional theory of atoms and molecules. R.G. Parr and W. Yang, Oxford University Press, New York, Oxford, 1989. IX + 333 pp. Price £45.00. *Int J Quantum Chem.* 1993;47(1):101-101.
158. de Broglie L. Recherches sur la théorie des Quanta.
159. Schrödinger E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys Rev.* 1926;28(6):1049-1070.
160. Born M, Oppenheimer R. Zur quantentheorie der molekeln. *Annalen der Physik.* 1927;389(20):457-484.
161. Kohn W, Sham LJ. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev.* 1965;140(4A):A1133-A1138.
162. Perdew JP. Jacob's ladder of density functional approximations for the exchange-correlation energy. In: *AIP Conference Proceedings.* Vol 577. Antwerp (Belgium): AIP; 2001:1-20.
163. Becke AD. A new mixing of Hartree–Fock and local density-functional theories. *J Chem Phys.* 1993;98(2):1372-1377.
164. Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA. 6-31G\* basis set for third-row atoms. *J Comput Chem.* 2001;22(9):976-984.
165. Hohenberg P, Kohn W. Inhomogeneous Electron Gas. *Phys Rev.* 1964;136(3B):B864-B871.
166. Mennucci B, Tomasi J, Cammi R, et al. Polarizable Continuum Model (PCM) Calculations of Solvent Effects on Optical Rotations of Chiral Molecules. *J Phys Chem A.* 2002;106(25):6102-6113.
167. Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol.* 1976;103(2):227-249.

168. Chung LW, Sameera WMC, Ramozzi R, et al. The ONIOM Method and Its Applications. *Chem Rev.* 2015;115(12):5678-5796.
169. Caprasecca S, Jurinovich S, Viani L, Curutchet C, Mennucci B. Geometry Optimization in Polarizable QM/MM Models: The Induced Dipole Formulation. *J Chem Theory Comput.* 2014;10(4):1588-1598.
170. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput.* 2015;11(8):3696-3713.
171. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem.* 1983;4(2):187-217.
172. Oostenbrink C, Villa A, Mark AE, Gunsteren WFV. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem.* 2004;25(13):1656-1676.
173. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc.* 1996;118(45):11225-11236.
174. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem.* 2004;25(9):1157-1174.
175. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007;126(1):014101.
176. Berendsen HJC. Transport Properties Computed by Linear Response through Weak Coupling to a Bath. In: Meyer M, Pontikis V, eds. *Computer Simulation in Materials Science: Interatomic Potentials, Simulation Techniques and Applications*. NATO ASI Series. Dordrecht: Springer Netherlands; 1991:139-155.
177. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys.* 1984;81(8):3684-3690.
178. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys.* 1980;72(4):2384-2393.
179. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys.* 1984;52(2):255-268.
180. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A.* 1985;31(3):1695-1697.
181. Parrinello M. Polymorphic transitions in single crystals: A new molecular dynamics method: *Journal of Applied Physics: Journal of Applied Physics.* 1981;52(12):7182.
182. Harada N, Nakanishi K. *Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry*. Mill Valley, CA: University Science Books; 1983.
183. Koch O, Klebe G. Turns revisited: A uniform and comprehensive classification of normal, open, and reverse turn families minimizing unassigned random chain portions. *Proteins Struct Funct Bioinforma.* 2009;74(2):353-367.
184. Koch O. Use of secondary structure element information in drug design: polypharmacology and conserved motifs in protein-ligand binding and protein-protein interfaces. *Future Med Chem.* 2011;3(6):699-708.
185. Fasman GD. *Circular Dichroism and the Conformational Analysis of Biomolecules*. Plenum Press, New York; 1996.
186. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci.* 2000;1(4):349-384.
187. Zuiderweg ERP. Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy <sup>?</sup> *Biochemistry.* 2002;41(1):1-7.

188. Cavanagh J, ed. *Protein NMR Spectroscopy: Principles and Practice*. 2. ed. Amsterdam: Elsevier, Acad. Press; 2007.
189. Guéron M, Plateau P, Decors M. Solvent signal suppression in NMR. *Prog Nucl Magn Reson Spectrosc*. 1991;23(2):135-209.
190. Piotto M, Saudek V, Sklenář V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR*. 1992;2(6):661–665.
191. Hwang T-L, Shaka AJ. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J Magn Reson*. 1995;112:275-279.