N° d'ordre : 04/2002-E/CH

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene

THESE

Présentée en vue de l'obtention du grade de

Docteur Es-Sciences Physiques

Spécialité Chimie Théorique

par

Safia TAÏRI-KELLOU

Sujet

Conception Assistée par Ordinateur de Complexes ligand/récepteur de la Cholécystokinine

Soutenue le 14 Septembre 2002 devant le jury :

B.	**MAOUCHE**	Professeur	USTHB	Président
J.-L.	**RIVAIL**	Professeur	Nancy I	Directeur de thèse
B.	**MAIGRET**	Directeur de recherche CNRS	Nancy I	Co-directeur de thèse
H.-M.	**GUERMOUCHE**	Professeur	USTHB	Examinateur
M.	**HAMDI**	Professeur	USTHB	Examinateur
O.	**OUAMERALI**	Professeur	USTHB	Examinatrice
B.	**DJERDJOURI**	Maître de Conférences	USTHB	Examinatrice

REMERCIEMENTS

Les travaux présentés dans ce mémoire de thèse ont été réalisés à temps partiel au laboratoire de chimie théorique de Nancy I (Unité Mixte de Recherches, CNRS 7565), dans le cadre d'une formation à distance et grâce à une bourse franco-algérienne. Je voudrais donc remercier les autorités de l'USTHB et, en particulier, de la Faculté de Chimie, pour m'avoir fait bénéficier de cette formation. J'exprime ma sincère gratitude à Monsieur le professeur Boubekeur Maouche, pour son soutien durant ce parcours et ses encouragements constants. Je n'oublie pas l'aide qu'il m'apporta, dans une période difficile, en créant une collaboration avec un grand laboratoire de chimie théorique, celui de Nancy. Je le remercie infiniment d'avoir accepté de juger mes travaux et de présider ce jury.

J'exprime toute ma vive gratitude à Monsieur le Professeur Jean-Louis Rivail, responsable du laboratoire de chimie théorique de Nancy I, pour m'avoir accueillie au sein de son équipe de notoriété internationale et pour m'avoir donné l'opportunité de me former dans un domaine d'actualité : la biochimie théorique. Je le remercie également pour l'intérêt qu'il a porté à mes travaux de recherches en en assurant le suivi.

Je tiens à remercier tout particulièrement Bernard Maigret, directeur de recherches au CNRS, pour avoir accepté de m'initier au domaine passionnant de la biochimie théorique en dirigeant mes travaux. Je lui exprime toute ma profonde reconnaissance de m'avoir fait bénéficier de ses compétences et de son expérience en modélisation moléculaire. Son aide ainsi que ses propres travaux dont je me suis largement inspirée, m'ont toujours été d'un grand secours. La confiance qu'il m'a accordée, les moyens qu'il a mis à ma disposition pour la réalisation de ces travaux et les encouragements qu'il n'a cessé de me prodiguer ont été des catalyseurs potentiels de mes recherches.

Avoir été dirigée par ces deux personnalités scientifiques est un honneur exceptionnel : qu'ils veuillent bien trouver ici l'expression de ma profonde et respectueuse gratitude pour avoir accepté de venir à Alger afin de siéger au jury, et cela, malgré leurs multiples tâches.

Je remercie très chaleureusement Madame Ourida Ouamerali, Professeur à l'USTHB, pour sa sympathie et ses conseils amicaux. Elle a accepté de juger ce travail, j'en suis très touchée.

Je suis très sensible à l'honneur que me fait Monsieur Hassan Moulay Guermouche, Professeur à l'USTHB par sa présence dans le jury. Qu'il trouve ici l'expression de ma profonde gratitude.

Mes sincères remerciements s'adressent au professeur Mâamar Hamdi qui a bien voulu examiner ces travaux.

Madame Bahia Djerdjouri, Maître de conférences à l'USTHB, a accepté d'apporter sa critique de biochimiste en jugeant ce travail. Je la remercie très chaleureusement.

Mes remerciements vont également au Docteur Alain Cartier pour sa collaboration fructueuse ; qu'il trouve ici l'expression de ma gratitude pour sa grande disponibilité et pour sa bienveillance à mon égard.

J'ai eu maintes fois l'occasion de bénéficier de l'aide et des conseils des Docteurs Daniel Rinaldi et Manuel Ruiz-Lopez. Je leur adresse mes sincères remerciements.

Lors de mes séjours au laboratoire de Nancy, j'ai eu la chance de rencontrer un grand nombre de chercheurs, du laboratoire et de tous horizons. Je ne pourrais les citer tous, mais chacun à sa manière m'a permis de construire la trame de ce travail. Que les personnes qui m'ont patiemment communiqué une partie de leur savoir, puissent trouver dans cet écrit l'expression de ma plus profonde gratitude. Je remercie Carole Poggio pour son aide efficace et sa sympathie.

Durant la réalisation de ces travaux, j'ai eu également l'occasion d'échanger des points de vue avec des biologistes moléculaires du groupe de recherche dirigé par Christiane Garbay, Professeur à l'université René Descartes (Paris V). Je remercie sincèrement Christiane et Bruno pour leur collaboration fructueuse et la sympathie qu'ils m'ont témoignée lors de mes brèves escales dans leur laboratoire.

Je ne saurais oublier mes chères amies Amel et fifi, pour leur soutien et la sympathie qui nous unit. Je leur témoigne ici ma profonde amitié.

Que mes collègues de la Faculté de Chimie de l'USTHB et également son personnel administratif trouvent ici l'expression de mes sincères remerciements pour la sympathie et l'amitié qu'ils m'ont toujours manifestées.

Enfin, je tiens à remercier ma famille et plus particulièrement mes parents pour leur soutien constant depuis le début de mes études et pendant ces difficiles années.

Merci Toufik, Merci Lilou et Kenza pour votre patience et pour le rôle déterminant que vous avez joué dans mes choix de ces dernières années, …vers la recherche, …vers le plus beau métier du monde, …avec tout mon amour.

A mes parents

A Toufik

A lilou et à Kenza

SOMMAIRE

INTRODUCTION GENERALE 9

CHAPITRE I LA CHOLECYSTOKININE 13

I - INTRODUCTION 13

II - STRUCTURE MOLECULAIRE ET LOCALISATION 13

III - RECEPTEURS DE LA CHOLECYSTOKININE 15

1) - Identification et localisation 15
2) - Clonage 16
3) – Structure 17

IV - ACTIONS PHYSIOLOGIQUES ET THERAPEUTIQUES DE LA CHOLECYSTOKININE 21

1) - Actions sur le système gastrointestinal 21

2) - Actions sur le système nerveux central et périphérique 22
   a) - la satiété 22
   b) - l'analgésie opioïde 22
   c) - l'anxiété et les troubles psychiques 23
   d) - la mémoire 23

3) – Conclusion 23

V - DESCRIPTION DE QUELQUES LIGANDS DES RECEPTEURS CCK. 24

1) - Les agonistes 25
   a) Récepteurs CCK_1 25
   b) Récepteurs CCK_2 26

2) - Les antagonistes 28
   a) Récepteurs CCK_1 28
   b) Récepteurs CCK_2 32

BIBLIOGRAPHIE 36

CHAPITRE II LES METHODES DE LA MODELISATION MOLECULAIRE 45

I – INTRODUCTION 45

II- LA MÉCANIQUE QUANTIQUE 46

1) – Bases et Principes 46

2) - Approximations usuelles 47

   a) Approximation de Born-Oppenheimer 47
   b) Approximation orbitale 48

3) - Méthodes du champ auto cohérent ou SCF (Self Consistent Field) 49

a) Equations de Hartree-Fock 49
b) Equations de Roothaan 50
c) Méthodes de calcul ab initio 51
d) Méthodes de calcul semi-empirique 52
d1) Principe 52
d2) Avantages et inconvénients 53

4) – Méthodes de la fonctionnelle de la densité 54

III - LA MÉCANIQUE MOLÉCULAIRE 55

1) - Bases et principes 55

2) - Le champ de forces 57
a) Définition 57
b) Expression générale des différents termes d'un champ de force 59

3) - Minimisation de l'énergie potentielle 65
a) Techniques aux dérivées premières 66
b) Techniques aux dérivées secondes 68

4) - Recherche systématique de l'espace conformationnel 69

5) - La dynamique moléculaire 69
a) Définition et principes 69
b) Méthodologie 70
c) Recuit simulé 73
d) Autres méthodes 74

6) - Les logiciels de la mécanique moléculaire 74

7) - Conclusion 75

IV – EVOLUTION DES METHODES DE LA MODELISATION 76

BIBLIOGRAPHIE 77

CHAPITRE III ETUDE DE RELATIONS QUANTITATIVES STRUCTURE-ACTIVITE D'UNE SERIE D'ANTAGONISTES NON PEPTIDIQUES DE LA CHOLECYSTOKININE 81

I – INTRODUCTION 81

II –RECHERCHES DE RELATIONS QUANTITATIVES ENTRE LA STRUCTURE ET L'ACTIVITE 82

1) - Présentation des composés étudiés 82

2) - Méthodologie utilisée 84
a) Modélisation des composés étudiés 85
b) Calcul des descripteurs 86
c) Utilisation des descripteurs calculés 89

3) - Résultats 90
c) Analyse conformationnelle 90
d) Corrélations obtenues et discussion 91

III- CONCLUSION 101

BIBLIOGRAPHIE 102

CHAPITRE IV **MODÉLISATION DES RÉCEPTEURS DE LA CHOLÉCYSTOKININE. ÉTUDE DU SITE DE LIAISON DES ANTAGONISTES NON PEPTIDIQUES.** 105

I – INTRODUCTION 105

II - MODÉLISATION DU RÉCEPTEUR CCK_1 ET PRÉSENTATION DU MODÈLE CCK_2 108

1) - Méthodologie utilisée dans la construction du récepteur CCK_1 108

2) - Présentation des deux modèles de récepteurs CCK 112

3) - Discussion 115

III - MODÉLISATION DES COMPLEXES 116

1)–Introduction 116

2)- Présentation des ligands sélectionnés pour le docking 117

3) Construction et optimisation des complexes 118
a) ligand/récepteur CCK_1 118
b) ligand/récepteur CCK_2 119

4) Explorations des sites de liaison ligand/récepteur CCK_1 120

5) Explorations des sites de liaison ligand/récepteur CCK_2 124

6) Discussion 127

IV – CONCLUSION 130

BIBLIOGRAPHIE 132

CHAPITRE V **ETUDE STRUCTURALE DE NOUVEAUX AGONISTES PSEUDOPEPTIDIQUES SUR LE RECEPTEUR CCK_2. ETUDE DE LEUR MODE D'INTERACTION AVEC LE RECEPTEUR.** 135

I – INTRODUCTION 135

II – MODELISATION MOLECULAIRE D'UNE SERIE DE NOUVEAUX AGONISTES PSEUDOPEPTIDIQUES 138

1) Méthodologie utilisée 138
2) Résultats de l'analyse conformationnelle 140
3) Discussion 146

III – ETUDE DES MODES D'INTERACTIONS LIGAND/RECEPTEUR 152

1) Introduction 152

2) Résumé des connaissances actuelles sur les sites de liaison des agonistes CCK sur le récepteur CCK_2. 152

3)- Construction et modélisation des complexes ligand/récepteur 153

4) - Analyse du site de liaison PJC42/récepteur CCK2 154

5) - Analyse du site de liaison BBL454/récepteur CCK_2 155

7) - Discussion des résultats 157

IV - CONCLUSIONS 158

BIBLIOGRAPHIE 160

CONCLUSION **GENERALE ET PERSPECTIVES** 163

ANNEXE 1 : GENERALITES SUR DES ASPECTS STRUCTURAUX FONCTIONNELS ET CONCEPTUELS EN BIOCHIMIE 170

ANNEXE 2 : DEFINITION DES PARAMETRES STATISTIQUES A UTILISER DANS UNE ANALYSE DE REGRESSION MULTILINEAIRE 180

ANNEXE 3 : SURFACES MOLECULAIRES ET MODELE DE CALCUL DU POTENTIEL ELECTROSTATIQUE MOLECULAIRE 185

INTRODUCTION GENERALE

La mise au point d'un médicament représente une tâche longue et onéreuse, au cours de laquelle il est nécessaire d'étudier et de tester plusieurs milliers de molécules avant de trouver la meilleure d'entre elles. Sa commercialisation ne se fait qu'après de nombreux tests concernant son efficacité, ses effets sur le métabolisme humain et sa toxicité. Ainsi, il s'écoule souvent plus d'une dizaine d'années entre la phase de recherche de la molécule active et la disponibilité du médicament sur le marché. Dans ce long processus, la collaboration entre les différentes disciplines touchant à la pharmacologie est devenue indispensable. Chacune d'elles apporte un éclairage différent du même problème. Parmi ces dernières, la conception assistée par ordinateur ou la modélisation moléculaire s'est imposée, par son aspect rationnel, comme un partenaire de choix dans ces programmes complexes. Cette technique peut permettre d'orienter le choix des molécules à synthétiser dans la phase de recherche en modélisant le site actif de l'agent de la maladie à contrer, puis en construisant une molécule pouvant s'y fixer.

C'est vers le milieu des années 50 que l'outil informatique a été mis en œuvre dans le domaine chimique pour modéliser les structures moléculaires. Depuis cette époque, l'évolution de la puissance de calcul des ordinateurs a permis à la modélisation moléculaire de s'appliquer à des structures de plus en plus complexes et de manière de plus en plus réaliste. La modélisation moléculaire est devenue, ainsi, une nouvelle technique de compréhension des phénomènes chimiques et un outil de recherche et d'analyse, à part entière, dans le domaine de la chimie structurale. Sa mise en œuvre nécessite l'utilisation de méthodes de calcul théorique (mécanique moléculaire, dynamique moléculaire, mécanique quantique) permettant de déterminer la représentation tridimensionnelle de la structure d'une molécule et d'en évaluer les propriétés physico-chimiques. La modélisation moléculaire, associée à une représentation info-graphique de la stéréochimie, permet d'interpréter des phénomènes physico-chimiques, de suggérer de nouvelles expériences et d'analyser ainsi des résultats à l'échelle moléculaire. Cette approche théorique est complémentaire de l'approche expérimentale et les deux resteront toujours associées pour une meilleure compréhension des phénomènes étudiés.

Un des domaines d'investigation les plus fascinants de la biologie structurale aujourd'hui est la compréhension de la liaison spécifique des ligands à leurs récepteurs biologiques. Malgré ses limites, la modélisation moléculaire est un bon outil pour simuler les interactions ligand/récepteur mises en jeu dans de tels systèmes.

La notion de récepteur pharmacologique a été introduite en 1905 par John Newport Langley. Cette notion demeurait conceptuelle jusqu'au début des années 70. Aucune étude n'a

démontré jusqu'alors l'existence, ni même la nature, d'une molécule réceptrice. C'est la mise au point de techniques de radio marquage des ligands qui a permis le développement des études de caractérisation moléculaire des récepteurs et des mécanismes de transduction du signal intracellulaire. La technique de marquage au tritium, préservant l'activité biologique du ligand, s'avéra particulièrement performante. Les récepteurs canaux ioniques et, en particulier, les récepteurs cholinergiques, les récepteurs peptidiques du glucagon, de l'insuline et de l'ocytocine furent les premiers récepteurs analysés.

Il faut noter que jusqu'aux années 80, les médicaments étaient généralement développés sans prendre en compte explicitement la structure moléculaire de leurs cibles thérapeutiques. L'accès aux premières séquences et structures tri-dimensionnelles de cibles fonctionnelles a alors engendré le développement de méthodes dites "rationnelles" de conception et d'optimisation de composés actifs. Grâce à ces méthodes et aux techniques de la modélisation moléculaire, le chercheur peut ainsi simuler et visualiser son système à l'échelle atomique. La modélisation d'une protéine se fonde habituellement sur sa structure cristalline, lorsqu'elle est disponible, ou sur la structure d'une protéine homologue. Cependant, les structures tridimensionnelles de la majorité des protéines clonées sont encore inconnues. Ce fait est d'autant plus vrai lorsqu'il s'agit d'une protéine membranaire. L'obtention de cristaux adéquats, pour une analyse par diffraction aux rayons X, d'un quelconque récepteur transmembranaire est, à l'heure actuelle, une opération extrêmement difficile, voire impossible.

Les recherches menées et présentées dans ce manuscrit constituent une contribution à l'élucidation du mode d'interaction ligand/récepteur de la cholécystokinine : une hormone neuropeptidique impliquée dans de nombreux de phénomènes physiologiques et qui, de ce fait, constitue une cible de choix pour développer des médicaments potentiels. Les objectifs principaux de nos travaux consistent à caractériser d'une part, le site de liaison des antagonistes non peptidiques et, d'autre part, le site de liaison de nouveaux agonistes pseudopeptidiques.

Après une introduction générale présentant le but du travail, ce mémoire s'articule autour de cinq chapitres suivis d'une conclusion générale et de trois annexes.

Le chapitre I est consacré à la présentation de la cible moléculaire de nos recherches, à savoir la cholécystokinine (CCK). Une revue bibliographique, permettant de connaître les diverses données actuellement disponibles sur cette importante hormone neuropeptidique et sur lesquelles repose la plupart de nos travaux, y est également présentée.

Nous décrivons dans le chapitre II les deux méthodologies utilisées en chimie théorique : la mécanique quantique et la mécanique moléculaire. Nous nous sommes attachés, au cours de l'exposé, à mesurer la pertinence et les limitations de chacune de ces deux méthodes.

Dans le chapitre III, nous présentons une étude quantitative de relations structure-activité (QSAR) d'une série d'antagonistes de type benzodiazépine sur les deux types de récepteurs de la cholécystokinine : CCK_1 et CCK_2, localisés respectivement dans le système nerveux périphérique et central. Cette étude a pour but de mettre en évidence les éléments structuraux responsables de l'activité antagoniste et de la sélectivité de ces antagonistes non peptidiques sur chacun de ces récepteurs.

Devant les limitations inhérentes aux corrélations QSAR obtenues, nous avons entrepris la construction et la modélisation des structures tridimensionnelles des protéines cibles, ce qui représente une étape obligatoire pour aller plus loin dans la recherche des éléments structuraux indispensables à une activité anti-CCK. Ces protéines réceptrices appartiennent à la grande famille des récepteurs membranaires liés à la production des seconds messagers par l'intermédiaire des protéines qui fixent la guanine et qui sont appelées les protéines G. Ces récepteurs sont appelés : récepteurs couplés aux protéines G (**RCPGs)**. Cette famille de protéines réceptrices est d'une importance cruciale car elle régule une grande partie des signaux reçus par la cellule. Nous avons pris en compte toute l'information expérimentale disponible sur cette classe de récepteurs durant la réalisation de ce travail.
Le docking qui est une technique d'assemblage particulière des structures 3D du ligand et du récepteur permettant de construire un modèle de complexe et d'étudier les interactions ligand/récepteur sera utilisé pour rechercher les modes potentiels de complexation avec les antagonistes non peptidiques faisant l'objet de l'étude QSAR. Le chapitre IV concerne l'ensemble de ces travaux de modélisation et de docking.

Le chapitre V est consacré à l'étude structurale de nouveaux agonistes pseudopeptidiques sur le récepteur CCK_2. Ce travail a été mené en étroite collaboration avec le groupe de recherche dirigé par Mme C. Garbay dans l'unité du professeur B.P. Roques à l'université de Paris V. Ce groupe a effectué la synthèse des composés et leur étude pharmacologique sur le rat. Nous avons également présenté, dans ce chapitre, nos travaux de docking effectués sur les deux composés parents de chacune des deux séries qui se distinguent par leur profil pharmacologique.

Dans la conclusion générale, nous résumons les principaux apports de nos travaux. Nous essayerons ensuite de cerner les limites et les insuffisances de nos modèles. Quelques perspectives seront esquissées pour essayer d'affiner les modèles des sites de liaison proposés.

Dans les annexes 1, 2 et 3, sont reportés des compléments d'informations aux chapitres I, III et V, respectivement.

CHAPITRE I

LA CHOLECYSTOKININE

I – INTRODUCTION

L'existence d'une substance régulant les fonctions digestives a été soupçonnée déjà en 1914, puis en 1928 Ivy et Oldberg mirent en évidence, dans le tractus gastro-intestinal, une substance capable d'activer la contraction de la vésicule biliaire[1]. C'est d'ailleurs pour cette propriété que cette substance a été appelée **cholécystokinine** (**CCK**) qui signifie d'après l'éthymologie grecque : hormone (*kinine* : mouvoir, circuler) contractant la vésicule biliaire (*kustis* : vessie ou sac ; *kholé* : bile). En 1943 Harper et Raper montrèrent que des extraits d'intestins de porc contiennent une substance capable de stimuler la sécrétion d'enzymes pancréatiques qu'ils nommèrent la pancréozymine[2]. Plus tard, dans les années 70, Jorpes et Mutt[3] ont purifié et séquencé un peptide de 33 acides aminés à partir d'extraits intestinaux de porc. Ce peptide possède les propriétés de la CCK et de la pancréozymine qui se révélèrent en fait identiques. La CCK fut longtemps considérée comme une hormone régulant les fonctions digestives. En 1975 l'équipe de Vanderhaeghen mit en évidence une immunoréactivité de type gastrique dans le système nerveux central des mammifères[4]. Cette immunoréactivité se révéla très associée à la CCK[5,6]. Ceci a conduit à un impressionnant champ d'études sur le plan fondamental et sur le plan application. Beaucoup de questions se posèrent alors concernant les différents aspects de l'activité biologique de la CCK et en particulier pour savoir s'il existe une relation entre son action dans le cerveau et celle dans l'appareil digestif[7].

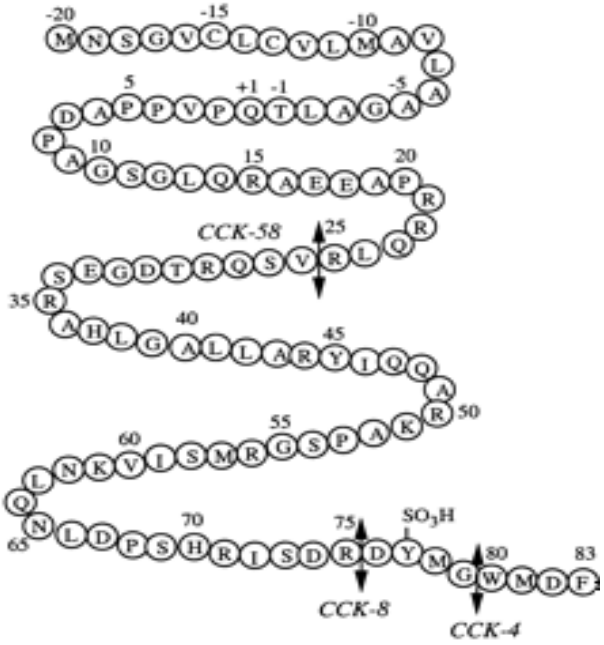
II - STRUCTURE MOLECULAIRE ET LOCALISATION

Comme la plupart des hormones la cholécystokinine est très hétérogène du fait qu'elle existe sous plusieurs formes moléculaires qui diffèrent par leur longueur de chaîne peptidique et par leur efficacité biologique. La CCK est issue d'une préprohormone de 115 acides aminés qui est ensuite clivée en de multiples sites, lors de sa maturation post-traductionnelle, pour donner naissance aux différentes formes moléculaires[8,9]. La structure de ce précurseur est donnée dans la ***figure 1***[10]. Ces sites de clivage enzymatique se trouvent au niveau de résidus basiques (arginine et lysine). Ces peptides sont généralement nommés soit par le nombre d'acides aminés qu'ils contiennent (ainsi l'octapeptide est désigné CCK-8), soit selon la numérotation des acides aminés faite par ordre croissant de l'extrémité N-terminale vers l'extrémité C-terminale avec la CCK-33 comme référence (ainsi la CCK-8 est aussi désignée CCK(26-33)). La maturation de la CCK induit deux modifications majeures de sa structure : l'α-amidation de la phénylalanine C-terminale et la sulfatation de la tyrosine en position sept à partir de l'extrémité C-terminale. Différents travaux ont montré que la tyrosine sulfatée et la

phénylalanine C-terminale amidée sont essentielles à l'activité biologique optimale de l'hormone[11-16]

Les formes biologiquement actives retrouvées dans le sang, l'appareil digestif et le cerveau sont : CCK-83, CCK-58, CCK-39, CCK-33, CCK-12, CCK-8, CCK-7, CCK-5 et CCK-4 ; les formes majeures circulantes sont : CCK-58, CCK-33 et CCK-8[17]. Il faut noter que l'octapeptide C-terminal sulfaté et amidé (CCK-8) est la structure minimale conservant l'activité complète de la CCK[18-19]. Cependant, l'heptapeptide C-terminal reproduit le spectre complet des activités biologiques de la CCK, *in vivo et in vitro* et le tétrapeptide est le plus petit fragment à posséder une activité biologique[20]. Il est important de noter que la CCK est le neuropeptide le plus abondant du cortex cérébral. En effet, un cerveau de rat contient environ 1µg de CCK alors qu'un cerveau humain en contient 1 à 2 mg[21]. Les techniques de radio immuno essai et immuno histochimie ont permis d'établir la distribution de la CCK[22-26]. Les diverses formes moléculaires de la CCK sont présentes en quantités variables selon le tissu et l'espèce considérés. Ainsi, la CCK-8 est la forme prédominante dans le cerveau de nombreuses espèces[27,28] alors que les fragments de plus haut poids moléculaire sont les plus abondants dans le tractus gastro-intestinal[29,30]. La CCK est produite par les cellules I endocrines de la muqueuse présentes dans le duodénum, le jéjunum et l'intestin proximal. Elle est également détectée chez le cobaye, au niveau des neurones présents dans le colon et l'intestin distal [31]. Elle est libérée dans le sang, en réponse à un repas riche en graisses et en protéines[32-34]. Le système nerveux central représente le site principal de production de ce neuropeptide[5]. La CCK est localisée en forte concentration dans le cortex cérébral et en quantité non négligeable dans l'hippocampe, les bulbes olfactifs, le septum, l'amygdale, l'hypothalamus et l'épine dorsale de la moelle épinière[31]. Dans le cerveau, la CCK est souvent colocalisée avec d'autres neurotransmetteurs tels que la dopamine, la substance P et les enképhalines[35].

Figure 1: Structure prédite de la préprocholécystokinine humaine[10]. Les résidus –20 à –1 concernent le signal du peptide.



III - RECEPTEURS DE LA CHOLECYSTOKININE

1) - Identification et localisation

L'utilisation de la CCK-8 radio-iodée, dans plusieurs travaux réalisés en 1980[13,36,37], a montré l'existence de sites de liaison de la CCK dans le pancréas et dans le cerveau du rat. Ces deux sites présentent des caractéristiques différentes. L'emploi d'agonistes et d'antagonistes de différentes structures chimiques a ensuite permis de réaliser la caractérisation de ces sites récepteurs. En effet, il a été observé que le récepteur CCK pancréatique lie la CCK-8 sulfatée avec une affinité environ 1000 fois plus importante que la CCK-8 non sulfatée ou la CCK-4 ; il en est de même pour l'efficacité dans la stimulation de la sécrétion d'amylase pancréatique [36,38]. En revanche, les études réalisées sur des homogénats de cortex cérébral montrent que le récepteur CCK, présent dans ce tissu, lie la CCK-8 sulfatée avec une affinité identique ou seulement 10 fois plus forte que la CCK-8 non sulfatée et la CCK-4[36]. Ces différentes études sont les premières à suggérer l'existence de deux types de récepteurs CCK. Ces derniers ont une pharmacologie différente, vis-à-vis de la sulfatation du peptide endogène ; l'un est localisé dans le système nerveux périphérique et dépendant de la présence du sulfate sur la CCK, l'autre est localisé dans le système nerveux central et ne discriminant pas (ou faiblement) les formes sulfatées et non sulfatées de la CCK. Ces deux

classes de récepteurs ont été nommés en 1987 : le récepteur de type A (Alimentary) noté CCK-A et le récepteur de type B (Brain) noté CCK-B[39].

Le clonage des ADN complémentaires (ADNc) du récepteur CCK-B et du récepteur de la gastrine identifié dans l'estomac a révélé que leurs séquences sont identiques et qu'il s'agit, en fait, du même récepteur[40]. De ce fait, le récepteur CCK-B est aussi appelé CCK-B/G.

Les deux types de récepteurs CCK-A et CCK-B/G se distinguent par des profils pharmacologiques différents pour les 2 agonistes CCK et gastrine. En effet, les récepteurs CCK-A sont fortement sélectifs vis-vis de ces agonistes alors que les récepteurs CCK-B/G lient ces agonistes CCK et gastrine avec la même forte affinité[41]. Plusieurs travaux[42-46] biochimiques et pharmacologiques tendent à montrer que les récepteurs CCK-A et CCK-B/G existeraient sous des états d'affinité différents, c'est à dire qu'ils seraient caractérisés par la présence de deux ou trois sous-sites de récepteurs possédant soit une haute soit une faible affinité pour la CCK.

Il est important de signaler que la distribution des sites de liaison est bien corrélée avec la localisation du peptide endogène (*cf* paragraphe II). Les sites de liaison CCK-A sont, en majeure partie, localisés dans le système nerveux périphérique alors que ceux de CCK-B/G prédominent dans le cortex cérébral. Cependant, l'expression du récepteur CCK-B/G prédomine par rapport au récepteur CCK-A dans le pancréas des mammifères supérieurs (porc, veau, homme). Il faut noter également que certaines cellules du système immunitaire telles que les monocytes et les lymphocytes T expriment le récepteur CCK-B/G[41].

2) – Clonage

Le récepteur CCK-A a été le premier à être cloné à partir du pancréas du rat[47]. Il a ensuite été cloné dans différentes espèces dont le cobaye, la souris, le singe et l'homme. Le clonage du récepteur CCK-A humain a été effectué à partir d'une banque d'ADNc de vésicule biliaire, en employant une technique de PCR (Polymerase Chain Reaction) avec des oligonucléotides synthétisés d'après la séquence déjà connue du récepteur CCK-A du rat[48-50].

Le premier clonage du récepteur CCK-B/G a été effectué à partir de cellules pariétales du chien et d'une banque d'ADNc d'extraits de cortex cérébral du rat[40,51]. Le récepteur CCK-B/G humain a été également cloné à partir de banques d'ADNc de cerveau[52,53].

Le gène du récepteur CCK-B/G humain est localisé sur le chromosome **11**[52,54] , alors que celui du récepteur CCK-A humain se situe sur le chromosome **4**[55-57].

3) –Structure

L'analyse des séquences primaires et des profils d'hydrophobicité de ces protéines réceptrices a permis d'en déduire leur structure secondaire. Ces récepteurs sont caractérisés

par sept domaines transmembranaires (composés de résidus hydrophobes) reliés par des boucles intra- et extracellulaires. Ils possèdent une extrémité N-terminale extracellulaire et une extrémité C-terminale intracellulaire de nature hydrophile. Ces caractéristiques sont celles de la famille des récepteurs couplés aux protéines G (**RCPGs**) (*confer* Annexe 1).

Le récepteur CCK-A est une glycoprotéine (protéine combinée aux glucides) membranaire de 85 à 95 kDa (kilodalton) dont la masse théorique de la partie protéique est de 42 à 44 kDa. Le récepteur du rat est constitué de 444 acides aminés, alors que celui de l'humain en comporte 428[41].

Le récepteur CCK-B/G est également une glycoprotéine, mais de 74 à 78 kDa, composée d'un corps protéique de masse théorique de 45 kDa. La protéine humaine est composée de 447 acides aminés et celle du rat en compte 452[41].

Les deux protéines réceptrices CCK-A et CCK-B/G présentent 50 % d'homologie au niveau de leur séquence primaire[49,51]. En revanche, à l'intérieur de chaque type de récepteur CCK-A ou CCK-B/G, une homologie supérieure à 80% est retrouvée entre les récepteurs clonés dans les différentes espèces[41,58]. Ces récepteurs ont deux cystéines conservées, situées dans la première et la deuxième boucle extracellulaires. Ces cystéines doivent former un pont disulfure requis pour la stabilisation de la structure tertiaire[10], comme cela a été observé pour certains récepteurs de la même famille : la rhodopsine, les récepteurs β-adrénergiques et muscariniques[59-61]. Une autre cystéine, présente dans l'extrémité C-terminale intracellulaire des récepteurs, est également conservée et correspond à un site de palmitoylation, servant à l'ancrage de la protéine à la membrane plasmique comme cela a été décrit pour la rhodopsine et les récepteurs β-adrénergiques[41,10].

A cause des difficultés rencontrées lors de la cristallisation de ces récepteurs membranaires, leurs structures tertiaires ne sont pas encore déterminées expérimentalement. Cependant, grâce à la modélisation moléculaire, on peut élaborer des modèles tridimensionnels qui doivent être validés sur la base de données expérimentales.

Il faut noter qu'il existe actuellement une nouvelle nomenclature pour ces récepteurs définie par le comité international sur la nomenclature des récepteurs : IUPHAR (International Union of Pharmacology)[10]. En fait, ces récepteurs ont été renommés conformément à leur ordre chronologique de mise en évidence et de clonage. Par conséquent, CCK-A et CCK-B/G s'appellent respectivement CCK_1 et CCK_2. Nous utiliserons cette nouvelle nomenclature dans la suite de notre exposé. Les structures primaires extraites de la banque de données SWISS-PROT et les structures secondaires[10] des récepteurs du rat sont données respectivement dans les ***figures 2 et 3***. Les séquences en acides aminés dans les modèles secondaires montrent les

hélices transmembranaires putatives, les sites de glycosylation (tridents) et les sites potentiels de phosphorylation des sérines par les protéines kinases C et A ($-PO_3$). Notons en particulier que les longueurs des troisièmes boucles intracellulaires distinguent bien les deux types de récepteurs.

Figure 2 : Séquences primaires des récepteurs du rat extraites de la banque de données SWISS-PROT : CCK_1 (P30551) et CCK_2 (P30553).

CCK_1 : 444 Acides aminés

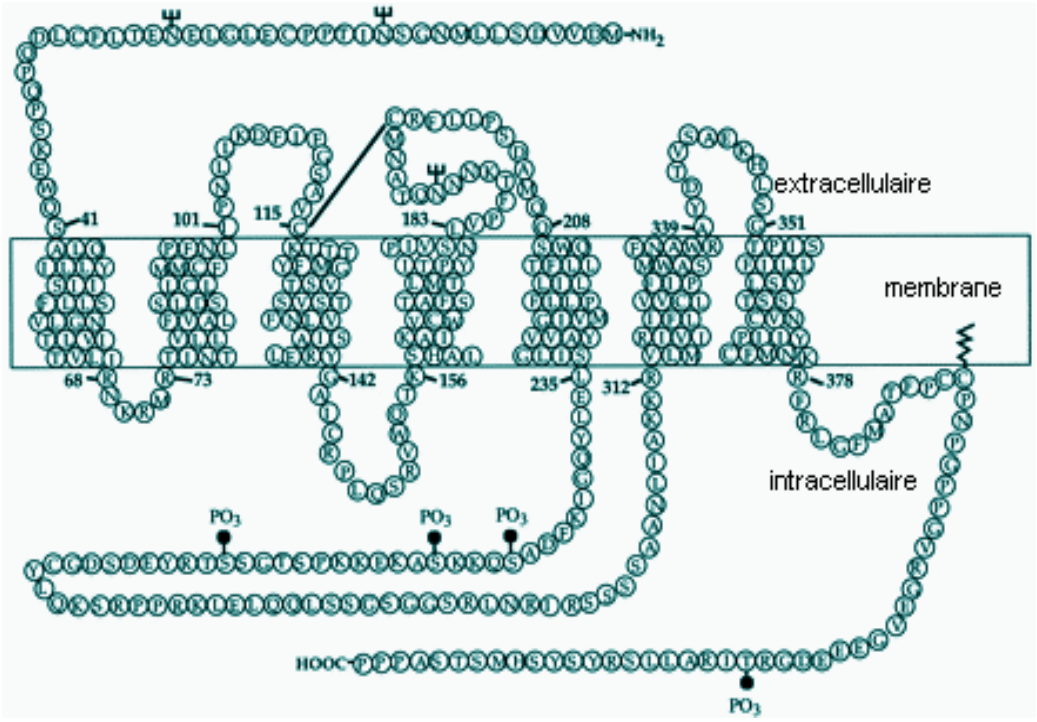
MSHSPARQHL VESSRMDVVD SLLMNGSNIT PPCELGLENE TLFCLDQPQP SKEWQSALQI
LLYSIIFLLS VLGNTLVITV LIRNKRMRTV TNIFLLSLAV SDLMLCLFCM PFNLIPNLLK
DFIFGSAVCK TTTYFMGTSV SVSTFNLVAI SLERYGAICR PLQSRVWQTK SHALKVIAAT
WCLSFTIMTP YPIYSNLVPF TKNNNQTANM CRFLLPSDAM QQSWQTFLLL ILFLLPGIVM
VVAYGLISLE LYQGIKFDAS QKKSAKEKKP STGSSTRYED SDGCYLQKSR PPRKLELQQL
SSGSGGSRLN RIRSSSSAAN LIAKKRVIRM LIVIVVLFFL CWMPIFSANA WRAYDTVSAE
KHLSGTPISF ILLLSYTSSC VNPIIYCFMN KRFRLGFMAT FPCCPNPGPP GVRGEVGEEE
DGRTIRALLS RYSYSHMSTS APPP

CCK_2 : 452 Acides aminés

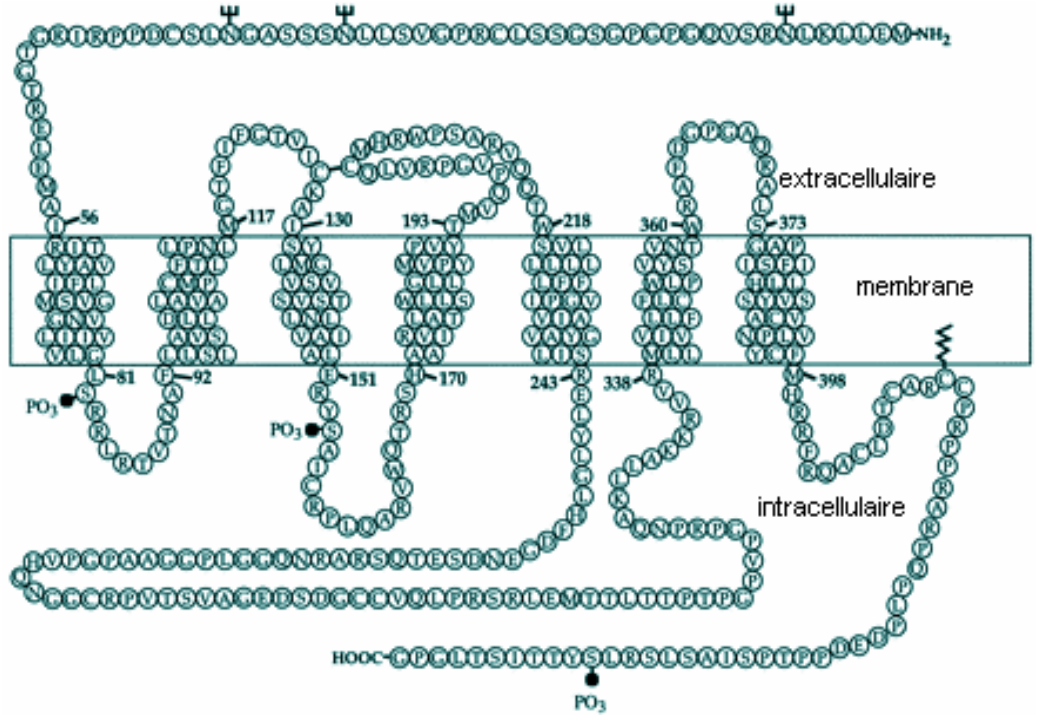
MELLKLNRSV QGPGPGSGSS LCRPGVSLLN SSSAGNLSCD PPRIRGTGTR ELEMAIRITL
YAVIFLMSVG GNVLIIVVLG LSRRLRTVTN AFLLSLAVSD LLLAVACMPF TLLPNLMGTF
IFGTVICKAI SYLMGVSVSV STLNLVAIAL ERYSAICRPL QARVWQTRSH AARVILATWL
LSGLLMVPYP VYTMVQPVGP RVLQCMHRWP SARVQQTWSV LLLLLLFFIP GVVIAVAYGL
ISRELYLGLH FDGENDSETQ SRARNQGGLP GGAAPGPVHQ NGGCRPVTSV AGEDSDGCCV
QLPRSRLEMT TLTTPTPGPV PGPRPNQAKL LAKKRVVRML LVIVLLFFLC WLPVYSVNTW
RAFDGPGAQR ALSGAPISFI HLLSYVSACV NPLVYCFMHR RFRQACLDTC ARCCPRPPRA
RPQPLPDEDP PTPSIASLSR LSYTTISTLG PG

Figure 3 : modèles schématiques des structures secondaires des récepteurs CCK_1 et CCK_2 du rat[10].

CCK_1 du rat



CCK_2 du rat



IV - ACTIONS PHYSIOLOGIQUES ET THERAPEUTIQUES DE LA CHOLECYSTOKININE

La cholécystokinine agit en tant qu'hormone dans le système gastrointestinal et en tant que neurotransmetteur-neuromodulateur dans le système nerveux central et périphérique. Le développement d'antagonistes et d'agonistes de ce peptide endogène a beaucoup contribué à essayer de comprendre ses rôles physiologiques.

1) - Actions sur le système gastrointestinal

En réponse à une prise alimentaire, la CCK est sécrétée et agit dans le tractus gastrointestinal et sur différents tissus cibles dont la vésicule biliaire, le pancréas, l'intestin et l'estomac.

La CCK régule la contraction des muscles lisses de la vésicule biliaire et relaxe, par l'intermédiaire de récepteurs situés sur des neurones, le sphincter d'Oddi exerçant ainsi une action coordonnée sur l'évacuation biliaire notamment chez l'homme[62]. De plus, certains antagonistes de la CCK inhibent la contraction de la vésicule biliaire confirmant ainsi le rôle du peptide endogène dans cet effet biologique[35,63-64].

La CCK régule également, soit par voie directe, soit par médiation nerveuse, la contraction des muscles de la paroi intestinale[64]. Elle ralentit la vidange gastrique en induisant la contraction du sphincter pylorique et le relâchement de l'estomac proximal[35,65]. Elle inhibe aussi la sécrétion d'acide gastrique[66].

L'action majeure de la CCK au niveau du pancréas exocrine est la stimulation de la sécrétion d'enzymes digestives, d'eau et de bicarbonates à partir de cellules acineuses[35]. Ces enzymes sont ensuite libérées dans le duodénum *via* le canal pancréatique. Elles interviennent dans la digestion en hydrolysant les hydrates de carbone, les protéines, les graisses et les acides nucléiques. La CCK joue également un rôle au niveau du pancréas endocrine, en stimulant selon des taux variables, la sécrétion d'insuline, de glucagon, de polypeptide pancréatique et de somatostatine[35]. De plus, il a été suggéré que cette régulation de la sécrétion des hormones pancréatiques, serait induite essentiellement par une libération de la CCK, sous forme courte (CCK-4), par des neurones situés au niveau du pancréas endocrine[67].

D'autres données de la littérature[68-70], suggèrent que la CCK est essentielle pour maintenir une croissance normale du pancréas. En effet, une administration répétée de CCK conduit à une hyperplasie et une hypertrophie du pancréas chez le rat, résultant d'une augmentation de la prolifération des cellules acineuses[68,69]. En revanche, l'administration d'un antagoniste de la CCK, inhibe les effets du peptide (endogène et exogène) et induit une atrophie du

pancréas[68]. Il faut noter aussi que le rôle de la CCK a été montré dans le développement de nodules précancéreux[71].

2) – Actions sur le système nerveux central et périphérique

En tant que modulateur, la CCK est distribuée *via* un réseau très important de neurones dans le système nerveux central et périphérique. En outre, elle coexiste avec d'autres neurotransmetteurs (sérotonine, dopamine et opiacés endogènes) et agit en modulant leurs actions. La distribution de la CCK, dans le système nerveux, étant ubiquitaire de nombreux troubles pourraient survenir à la suite d'une anomalie dans le métabolisme de ce peptide.

a) – la satiété

De nombreuses données de la littérature[35,72-77] démontrent que l'action la plus importante de la CCK, au niveau du système nerveux, est sur le contrôle de la satiété. Cependant, il n'est pas encore établi si cette action est médiée par voie centrale ou par voie périphérique, puisque l'administration centrale ou périphérique de fortes doses de CCK provoque, de la même façon, une réduction de la prise alimentaire[35]. La voie principale, par laquelle l'information de satiété atteint le cerveau, semble être les fibres afférentes du nerf vague. La CCK libérée de l'intestin, après un repas, agirait sur les récepteurs périphériques qui relaieraient ensuite l'information *via* le nerf vague aux régions du cerveau régulant les comportements alimentaires[78]. Nous donnons la représentation schématique[10] de ce mécanisme dans la ***figure 4***. Ainsi, d'un point de vue thérapeutique, les analogues de la CCK pourraient avoir des applications possibles dans le traitement de certains désordres de l'appétit comme la boulimie et l'anorexie nerveuses. D'autres études[79-81] ont suggéré que la CCK et la leptine qui est un peptide sécrété par le tissu adipeux, agiraient en synergie dans l'inhibition de la prise alimentaire notamment en agissant sur la sécrétion de l'insuline et du glucagon.

b) – l'analgésie opioïde

Crawley a montré que la CCK est un antagoniste opioïde dans l'analgésie et peut être également impliquée dans le développement des processus de tolérance aux opiacés[35]. En effet, l'administration de la CCK, chez le rat, atténue l'amplitude et la durée de la réponse analgésique à la morphine. De plus, l'injection d'anticorps dirigés contre ce peptide, permettant d'inactiver la CCK endogène, augmente l'analgésie due à la morphine. Ceci suggère bien que la CCK endogène agit sur l'analgésie opioïde. Ces données permettent donc une application thérapeutique des antagonistes CCK, dans le traitement de la douleur ou du moins dans les phénomènes de tolérance.

c) – l'anxiété et les troubles psychiques

L'injection intraveineuse de la CCK peut produire des attaques de panique et des crises d'anxiété chez l'homme et également chez l'animal. Ces effets sont confirmés par l'action de certains antagonistes de la CCK qui ont un rôle anxiolytique chez l'animal. Bien que ces mécanismes soient encore mal compris, la CCK semble exercer ses effets anxiogènes en interagissant avec les neurones dopaminergiques du mésencéphale[35,64]. L'activation des récepteurs CCK présents dans le noyau *accumbens* stimule la libération de la dopamine. Le déséquilibre de cette régulation peut entraîner des maladies neuropsychiatriques, telle que la schizophrénie[35,82].

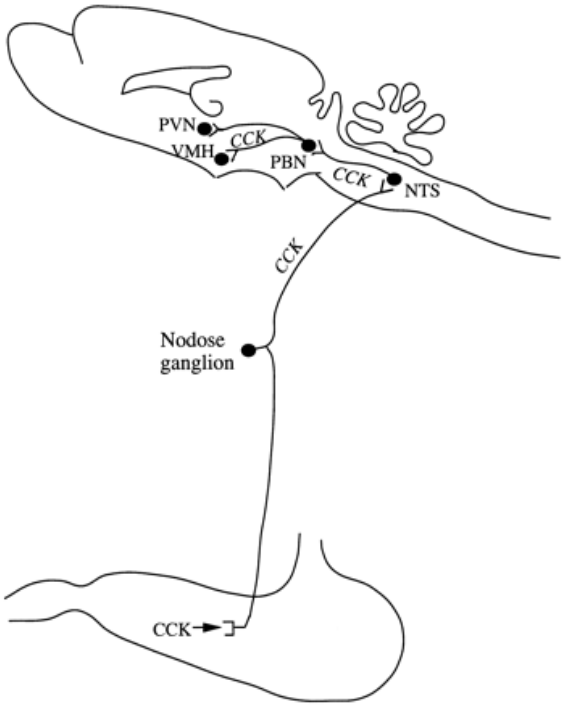
d) – la mémoire

La CCK semble également intervenir dans les processus de mémorisation et d'apprentissage comme l'ont montré de nombreux tests effectués sur l'animal[83,84]. Par ailleurs, la très forte concentration de la CCK dans l'hippocampe suggère son rôle possible dans les processus de mémorisation et d'apprentissage. En effet, les travaux de l'équipe de Lemaire[85] montrent que la CCK serait impliquée dans la régulation physiologique de ces processus et ce par l'activation des récepteurs CCK_1 et CCK_2 qui conduisent respectivement à des effets de type promnésiant et amnésiant.

3) - Conclusion

La diversité des actions biologiques de la cholécystokinine a mis en évidence la complexité de son mode d'action et explique l'intérêt considérable que lui portent les chercheurs travaillant dans des disciplines connexes. En effet, ce peptide est actif à des doses très faibles (μ g/kg) et cela qu'il soit administré par voie centrale ou périphérique, ce qui soulève le problème du type de récepteur impliqué dans la médiation des différentes réponses. Le développement de molécules, possédant un pouvoir antagoniste ou agoniste, représente depuis plusieurs années un domaine d'activité très intense. De nouvelles molécules, présentant un pouvoir meilleur que celles déjà existantes, permettraient de mieux connaître le rôle exact de la CCK notamment quand son action est concomitante à celle d'autres neurotransmetteurs.

Figure 4 : Représentation schématique du mécanisme d'action de la CCK dans la régulation du comportement alimentaire[10]. NTS (Nucleus Tractus Solitarius) ; PVN (Paraventricular Nucleus) ; PBN (Parabrachial Nucleus) ; VMH (Ventromedial Nucleus of the Hypothalamus).



V - DESCRIPTION DE QUELQUES LIGANDS DES RECEPTEURS CCK

Étant donné les nombreuses actions biologiques de la CCK de nombreux troubles de l'organisme pourraient survenir à la suite d'une anomalie dans le métabolisme de ce peptide. Par conséquent, de nombreux ligands (agoniste ou antagoniste) de ses récepteurs ont été développés jusqu'à aujourd'hui. Certains, hautement spécifiques pour un type de récepteur CCK_1 ou CCK_2, ont permis d'étudier l'implication sélective de ces deux types de récepteurs aux niveaux biochimique et pharmacologique. Toutefois, dans ce qui suit, nous nous limiterons à décrire les agonistes et les antagonistes les plus pertinents ou ceux représentant les têtes de séries. Les affinités données, dans les tableaux 1-4, sont mesurées dans le cerveau (CCK_2) et les membranes pancréatiques (CCK_1) du rat.

1)- Les agonistes

Comme l'activité biologique complète de la CCK réside dans son octapeptide C-terminal et que l'heptapeptide C-terminal reproduit le spectre complet de l'activité biologique du peptide endogène, *in vivo* et *in vitro*, les premiers analogues de la CCK ont été synthétisés à partir de ces deux peptides (CCK-7 et CCK-8) [20]. Les dérivés obtenus, en bloquant

l'extrémité N-terminale par des groupements protecteurs carbobenzoxy (z ou cbz) ou tertiobutyloxycarbonyl (Boc) ou acétyl (Ac), ont une efficacité équivalente, voire supérieure, à leurs homologues non protégés[12]. La substitution des méthionines par des norleucines (leucine non ramifiée : groupement isostère de la Met) a été effectuée dans de nombreux peptides CCK dans le but d'éliminer les sites d'oxydation de ces peptides et d'augmenter leur stabilité chimique. Ces peptides substitués sont plus résistants à la dégradation enzymatique (limitation de l'hydrolyse par les aminopeptidases). Il faut noter que cette substitution n'altère pas l'activité biologique du peptide[86].

a)- Récepteurs CCK_1

Les premiers agonistes sélectifs des récepteurs CCK_1 ont été développés par les laboratoires Abbott, à partir de la CCK-7 tel que le *A-71378* et de la CCK-4 tels que le *A-71623* et le *A-70874* (**Tableau 1**). La nature de ces ligands sélectifs pour le récepteur CCK_1 est d'autant plus remarquable que la CCK-7 n'est spécifique d'aucun récepteur CCK et que la CCK-4 a une bonne affinité uniquement au récepteur CCK_2.

Le *A-71378* contient un N-méthyl, positionné sur le lien peptidique entre l'acide aspartique et la norleucine, qui semble très critique pour la sélectivité des récepteurs CCK_1[87]. Les agonistes *A-71623* et *A-70874,* dérivés de la structure (Boc)-CCK-4, sont N-méthylés sur le lien peptidique entre l'acide aspartique et la phénylalanine C-terminaux. La méthionine a été substituée par une lysine modifiée portant différents groupements (**Tableau 1**). Ces modifications réalisées sur le peptide CCK-4 inversent la spécificité des composés pour les récepteurs CCK_1 et CCK_2. En effet, ces dérivés sont des ligands spécifiques du récepteur CCK_1 alors que la molécule mère (Boc)-CCK-4 est spécifique du récepteur CCK_2[88]. Le ligand *A-71623* se comporte comme un agoniste complet sur l'hydrolyse des inositols phosphates et la libération d'amylase. Par contre, le *A-70874* est un agoniste partiel dans sa capacité à stimuler ces mêmes activités biologiques. Cependant, la double N-méthylation sur les résidus Asp et Phe C-terminaux le rend hautement actif[87].

L'équipe de Pierson[89] a développé une série d'hexapeptides dont le plus sélectif des récepteurs CCK_1 est le *ARL-15849* (**Tableau 1**) (Hpa : 4-hydroxyphenylacetyl).

Une autre série dérivant des (1,5)-benzodiazépines a également été développée par plusieurs équipes[90-92].

Tableau 1 : Formules chimiques, activités (données en nanomoles : nM) et sélectivités des agonistes du récepteur CCK_1. La sélectivité est estimée par le rapport entre les valeurs d'IC_{50} (concentration correspondant à une inhibition de 50% de l'activité biologique) ou d'affinités relatives aux récepteurs CCK_1 et CCK_2, mesurées dans le cerveau et dans les membranes pancréatiques.

Agoniste	Formule ou nomenclature	IC_{50} (CCK_1)	Sélectivité CCK_2/CCK_1
A-71378	des-NH_2-Tyr(SO_3H)-Nle-Gly-Trp-Nle-(NMe)Asp-Phe-NH_2	0.5	1140
A-71623	Boc-Trp-Lys(o-tolylaminocarbonyl)-Asp-(NMe) Phe-NH_2	3.7	1200
A-70874	Boc-Trp-Lys(p-hydroxycinnamoyl)-Asp-(NMe) Phe-NH_2	4.2	170
ARL-15849	4-hydroxyphenylacetyl(SO_3H)-Nle-Gly-Trp-Nle-(Me)Asp-Phe-NH_2	0.03[a]	6590
GW-5823	2-[3-(1*H*-indazol-3-ylmethyl)-2,4-dioxo-5-phenyl-2,3,4,5-terahydrobenzo[*b*][1,4]diazepin-1-yl]-*N*-isopropyl-*N*-(4-methoxyphenyl) acétamide	22.9	50

[a]**:** K_i, constante d'affinité mesurée dans le cerveau et les membranes pancréatiques de rat.

b) - Récepteurs CCK_2

Différentes stratégies ont été utilisées pour développer des agonistes sélectifs des récepteurs CCK_2 dont la protection du peptide contre la dégradation enzymatique. Pour cela, l'heptapeptide Boc[$Nle^{28,31}$]CCK(27-33) noté BDNL qui est biologiquement actif a servi, par conséquent, de composé de départ pour le design de nouveaux analogues résistants et sélectifs[86]. Les modifications chimiques effectuées[93-95], sur ce composé (BDNL), sont la rétro-inversion de la liaison peptidique entre les résidus Nle28 et Gly29 et la N-méthylation du résidu Nle31. Elles ont conduit au *BC264* : un puissant agoniste sélectif des récepteurs CCK_2 (**Tableau 2**). D'autre part, les études d'analyse conformationnelle, réalisées au moyen des techniques de résonance magnétique nucléaire, de fluorescence et de modélisation moléculaire sur la CCK-8, ont permis de montrer que ce peptide existe préférentiellement sous forme repliée[96]. Par conséquent, dans le but de stabiliser certains des repliements, des composés cycliques ont été synthétisés. Parmi ces composés, on retrouve le *BC254* et le *RB204* (ou *BC197*) qui se sont avérés également sélectifs des récepteurs CCK_2 [97,98]. Un autre analogue de la CCK-8 non sulfaté, noté *SNF-8702*, a aussi été décrit[99]. Son affinité sur CCK_2 est 4000 fois supérieure à celle sur l'autre type de récepteur.

Une autre série d'agonistes très sélectifs des récepteurs CCK_2 a été développée à partir du tétrapeptide C-terminal. Plusieurs modifications ont été faites telles que la protection du tetrapeptide en Boc-CCK-4[100] ou la substitution de la Met par la Nle ou la NMeNle (norleucine N-méthylé)[101]. Par ailleurs, les études de modélisation par dynamique moléculaire et de résonance magnétique nucléaire, sur ces analogues de CCK-4, ont permis de synthétiser de nouveaux agonistes sélectifs[88,102,103]. En effet, les préférences conformationnelles de ces dérivés ont été prises en compte dans les travaux de synthèse en vue d'améliorer l'affinité, voire l'efficacité, d'un nouveau produit. De nombreux travaux ont conduit à la synthèse d'agonistes très sélectifs et prometteurs sur le plan thérapeutique[104-106].

Tableau 2 : Formules et affinités des agonistes sélectifs du récepteur CCK_2. Les constantes K_i, mesurées dans le cerveau et les membranes pancréatiques du rat, sont données en nM.

Agoniste	Formule	$K_i(CCK_2)$	Sélectivité(CCK_1/CCK_2)
BC264	Boc-Tyr(SO_3H)-gNle-mGly-Trp-(NMe)Nle-Asp-Phe-NH_2	0.1	780
BC254	Boc-D.Glu-Tyr(SO_3H)-Nle-D.Lys-Trp-Nle-Asp-Phe-NH_2	0.56	4500
BC197	Boc-D.Asp-Tyr(SO_3H)-Nle-D.Lys-Trp-Nle-Asp-Phe-NH_2	150	20
SNF-8702	Asp-Tyr-(NMe)Nle-Gly-Trp-(NMe)Nle-Asp-Phe-NH_2	0.9	4000

2)- Les antagonistes

Plusieurs antagonistes des récepteurs de la CCK, appartenant à des classes chimiques différentes, ont été développés. De manière générale, ces antagonistes peuvent être regroupés en familles de molécules que nous décrivons brièvement dans la suite.

a)- Récepteurs CCK_1

Les diverses séries d'antagonistes compétitifs de la CCK, spécifiques du récepteur CCK_1, peuvent être classées selon leur structure chimique : les dérivés de nucléotides cycliques, les dérivés d'acides aminés, les peptides CCK modifiés, les dérivés de benzodiazépine et les peptoïdes. Les formules des composés, représentant le chef de file dans chaque série, sont données dans la ***figure 5***. Leurs affinités et leurs sélectivités sont également données dans le **tableau 3**.

**) Les dérivés de nucléotides cycliques*

Ces dérivés sont les premiers antagonistes des récepteurs de la CCK qui ont été reportés[107]. L'acide dibutyryl guanosine monophosphate cyclique, noté Bt_2GMPC, est le plus puissant de ces composés. Toutefois, il s'est révélé toxique *in vivo* et il présente, ainsi que les dérivés bromés du GMPC, un faible antagonisme compétitif[108].

En 1997, la synthèse de nouveaux antagonistes spécifiques du récepteur CCK_1, dérivant de la structure (1,3)-dioxoperhydropirido[1,2-c]pyrimidine, a été décrite[109]. Le plus puissant de la série est un diastéréoisomère, noté *IQM-95,333,* qui est hautement spécifique du récepteur CCK_1.

**) Les dérivés d'acides aminés*

Les acides aminés modifiés sont initialement l'acide glutamique et le tryptophane. Les premiers antagonistes, décrits dans cette classe, sont le *proglumide* et le *benzotript* qui présentent peu de sélectivité entre les deux types de récepteurs[110]. Ensuite, des modifications ont été apportées par l'équipe de Markovec[111] à la structure du *proglumide*, notamment en allongeant les chaînes hydrophobes et en substituant les atomes d'hydrogène du cycle aromatique par des atomes de chlore, ce qui conduit respectivement au *lorglumide,* noté *CR-1409,* et au *loxiglumide,* noté *CR-1505*. Chacun de ces composés présente un gain en affinité et en sélectivité pour les récepteurs CCK_1.

En 1996, l'équipe d'Akiyama[112] a synthétisé un composé dérivé de la sérine noté *TP-680*. Son étude pharmacologique a révélé une bonne sélectivité et un antagonisme irréversible du récepteur CCK_1.

**) Les dérivés de la CCK ou antagonistes peptidiques*

La suppression du résidu phénylalanine en position C-terminale de la CCK-8, a conduit au premier antagoniste peptidique des récepteurs CCK_1[113]. La substitution du résidu L-tryptophane par un D-tryptophane, dans des analogues hepta- et octa-peptidiques de la CCK, conduit à des peptides présentant des propriétés antagonistes des récepteurs CCK. Le membre plus puissant de cette classe est le JMV179, dont la structure dérive de la Boc-CCK-7 ; [Boc-Tyr(SO_3)H-Nle-Gly-DTrp-Nle-Asp-2-O-CH_2-CH_2-Phényl] [114]. Les principales caractéristiques structurales de ce dérivé sont la substitution du résidu L-tryptophane en D-tryptophane et la substitution de la phénylalanine amidée par une fonction phényléthylester.

D'autres antagonistes ont été développés à partir de la structure de l'agoniste tétrapeptidique *A-71623* (**Tableau 1**) sélectif du récepteur CCK_1. Ils se caractérisent par la substitution du résidu lysine par des groupements L-4-aminophénylalanine et D- (ou L-) 3-aminophénylalanine.

*)*Les dérivés de benzodiazépine*

La famille des dérivés de benzodiazépine regroupe les antagonistes non peptidiques très puissants de la CCK. Ces composés ont été synthétisés sur la base de la structure de *l'asperlicine*, composé naturel non peptidique isolé à partir de la souche *Aspergillus alliaceus*. *L'asperlicine* a un effet antagoniste sur la contraction de la vésicule biliaire et de l'iléon induite par la CCK-8 chez le cobaye[115]. Le chef de file de cette famille d'antagonistes est le *dévazépide*, noté *L-364,718* ou *MK-329*. Il se lie au récepteur CCK_1 pancréatique ou de la vésicule biliaire avec une forte affinité (81 et 45 pM) et au récepteur CCK_2, exprimé dans le cerveau, avec une faible affinité[115]. De plus, son administration par voie orale ou intraveineuse inhibe les différentes actions biologiques de la CCK, chez la souris, le rat, le cobaye et l'homme[116,10].

Sur la base des études de relations structure-activité, différents dérivés tricycliques (1,4)-benzodiazépine ont été synthétisés dont le plus puissant est le *FK-480*[117-119].

*) *les peptoïdes*

Des substitutions adéquates sur le dipeptide (Boc)-Trp-Phe-NH_2 dont l'affinité est de l'ordre de 10 µM, ont donné naissance à de nouveaux antagonistes plus puissants dont le plus sélectif, développé par l'équipe de Hull, est noté 2-NAP[120-121]. Toutefois, la plupart de ces composés développés par les laboratoires Parke Davis sont des antagonistes du récepteur CCK_2 mais certains d'entre eux, par exemple le *PD-140,548*, se sont révélés de bons antagonistes du récepteur CCK_1.

*)*autres antagonistes*

De plus, il existe d'autres antagonistes non peptidiques puissants des récepteurs CCK_1 et qui n'appartiennent à aucune des familles précédentes. L'un d'eux est le *SR-27897 :* une molécule découverte par les laboratoires de Sanofi-Recherche[122]. Un autre antagoniste très sélectif : le *T-0632*, a été synthétisé par l'équipe de Taniguchi en 1996[123].

Figure 5 : Formules chimiques des antagonistes du récepteur CCK_1 et décrits dans le texte.

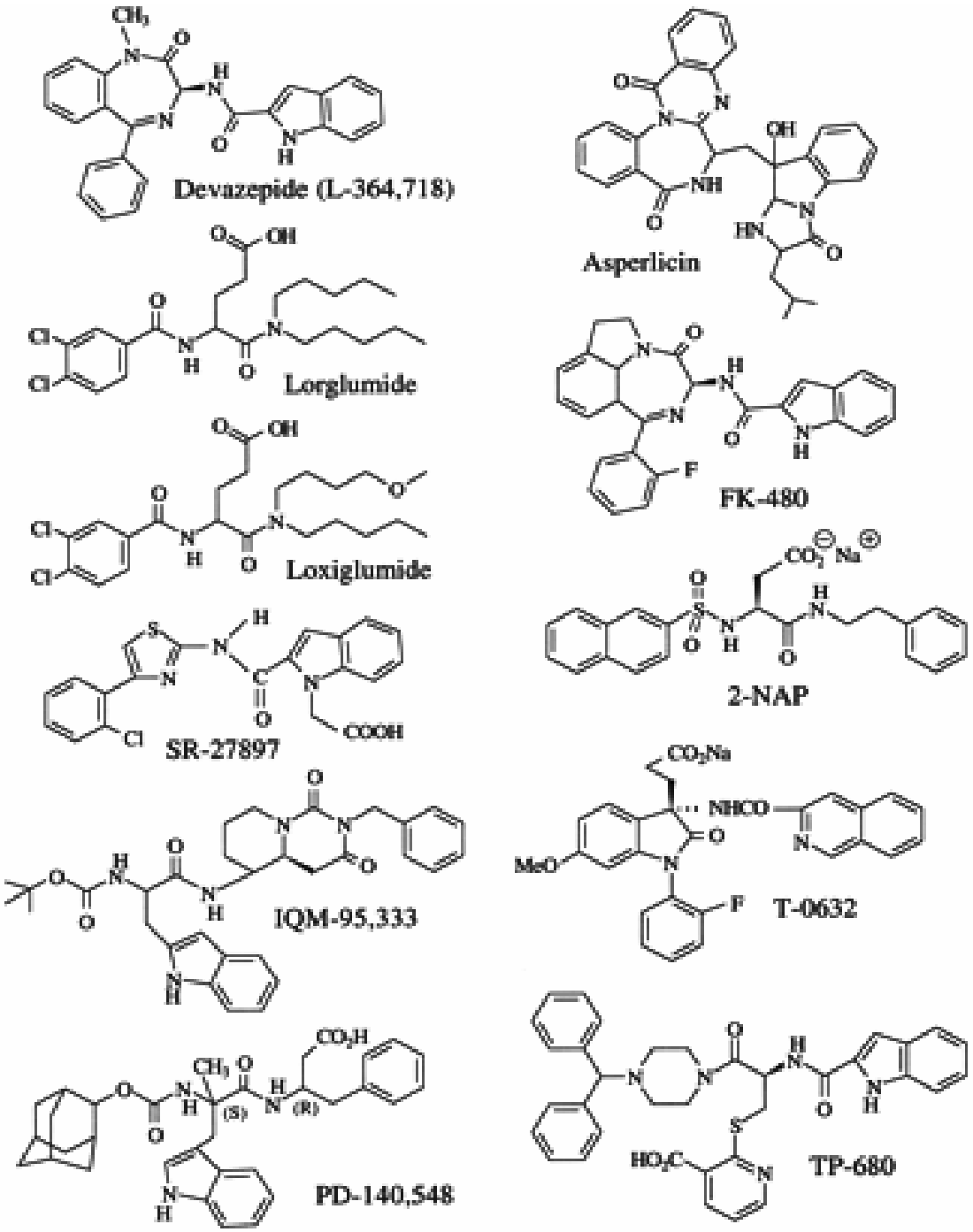


Tableau 3: Affinités et sélectivités des antagonistes du récepteur CCK_1. Les constantes d'affinité Ki mesurées dans le cerveau et les membranes pancréatiques de rat, sont données en nanomoles (nM).

Antagoniste	$K_i(CCK_1)$	Sélectivité $K_i(CCK_2)/K_i(CCK_1)$
L-364,718 ou *Dévazépide*	0.1	3750
IQM-93,333	0.6	>8000
PD-140,548	2.8[a]	93
FK-480	0.4[a]	480
2-NAP	250	300
T-0632	0.24	23000
TP-680	1.2	1510
SR-27897	0.2	800

[a]: Valeur en IC_{50}, concentration correspondant à une inhibition de 50% de l'activité biologique.

b)- Récepteurs CCK_2

Plusieurs tentatives ont été faites afin de développer des antagonistes puissants et spécifiques du récepteur CCK_2. Les composés obtenus appartiennent à différentes classes chimiques : les dérivés de benzodiazépine, les dérivés d'acides aminés, les dipeptoïdes, les dérivés de pyrazolidinone et d'uréidoacétamide. Les formules des composés, représentant dans chaque série le chef de file, sont données dans la ***figure 6***. Leurs affinités et leurs sélectivités, lorsqu'elles sont disponibles, sont également données dans le **tableau 4**.

**)les dérivés de benzodiazépine*

L'affinité modérée du *dévazépide* pour le récepteur CCK_2 (**Tableau 3**) a laissé supposer que le noyau benzodiazépine est une structure clé pour développer des composés plus 'affins'. En effet, en utilisant cette stratégie, l'équipe de Bock a développé le premier antagoniste puissant de ce récepteur[(124)]. Ce composé est le *L-365,260* (***Figure 6***). L'un des facteurs déterminant la sélectivité dans la série obtenue est la stéréochimie du carbone 3 du cycle diazépine ; l'énantiomère R est en général (statistiquement) le plus sélectif. De plus, la substitution du phényle branché en C5 de la diazépine par un cyclohexyl semble améliorer cette sélectivité[(125)]. Les inconvénients majeurs de ces dérivés sont leur biodisponibilité limitée et leur inactivité par voie orale. Les nombreux travaux entrepris en vue de surmonter ces inconvénients et, en particulier, augmenter leur absorption ou leur solubilité aqueuse[(126-

[133], ont conduit aux composés dont les plus sélectifs sont : *L-740,093*[128], *CP-212,454*[131], *YM-022*[132], *YF-476*[133].

**)les dipeptoïdes et les dérivés d'acides aminés*

Les antagonistes dipeptoïdiques dérivent de la structure Boc-Trp-Phe-NH_2, dont le *PD-134,308*[134] et le composé *19*[135] qui présente une forte affinité au récepteur CCK_2.

Une comparaison directe de la structure de ces dérivés a montré que la réduction de leur taille pourrait augmenter leur lipophilie. Sur cette base, des composés ont été synthétisés et certains d'entre eux se sont révélés puissants et sélectifs vis-à-vis du récepteur CCK_2[136]. D'autre part, de nombreuses restrictions conformationnelles ont été introduites, dans la synthèse de nouvelles structures, afin d'améliorer les propriétés du *PD-134,308*. Malheureusement, ni la cyclisation du N-terminal[137], ni la cyclisation globale[138] et ni la 'rigidification' de la liaison amide[139] n'ont conduit à des résultats positifs. Cependant, la cyclisation du coté C-terminal des dérivés du *PD-134,308*, au moyen du tetrahydronaphtyl, s'est révélée améliorer l'affinité[140]. Cette approche a été utilisée dans la synthèse de nouveaux antagonistes qui sont aussi puissants que le *RB 210*[141,142]. La faible biodisponibilité du *PD-134,308* a conduit le groupe de Parke-Davis à synthétiser des composés en réduisant le poids du composé parent. Ces modifications, permettant une meilleure absorption, ont donné de meilleurs antagonistes dont le *CI-1015*[143].

**) les dérivés de pyrazolidinone et d'uréidoacétamide*

La famille des pyrazolidinones est basée sur la structure diphényl-pyrazidinone et comprend le ligand *LY-288,513*[144]. Les études pré cliniques de toxicologie, sur cette série de composés, ont décelé des effets défavorables et par conséquent n'ont pas permis son développement.

Le groupe Rhône-Poulenc a développé des urédoacétamides dont le plus sélectif est le *RP-73,870*[145]. Malgré sa biodisponibilité orale relativement faible, ce composé est aussi un puissant antiulcéreux chez le rat.

Figure 6: Formules des antagonistes sélectifs du récepteur CCK_2

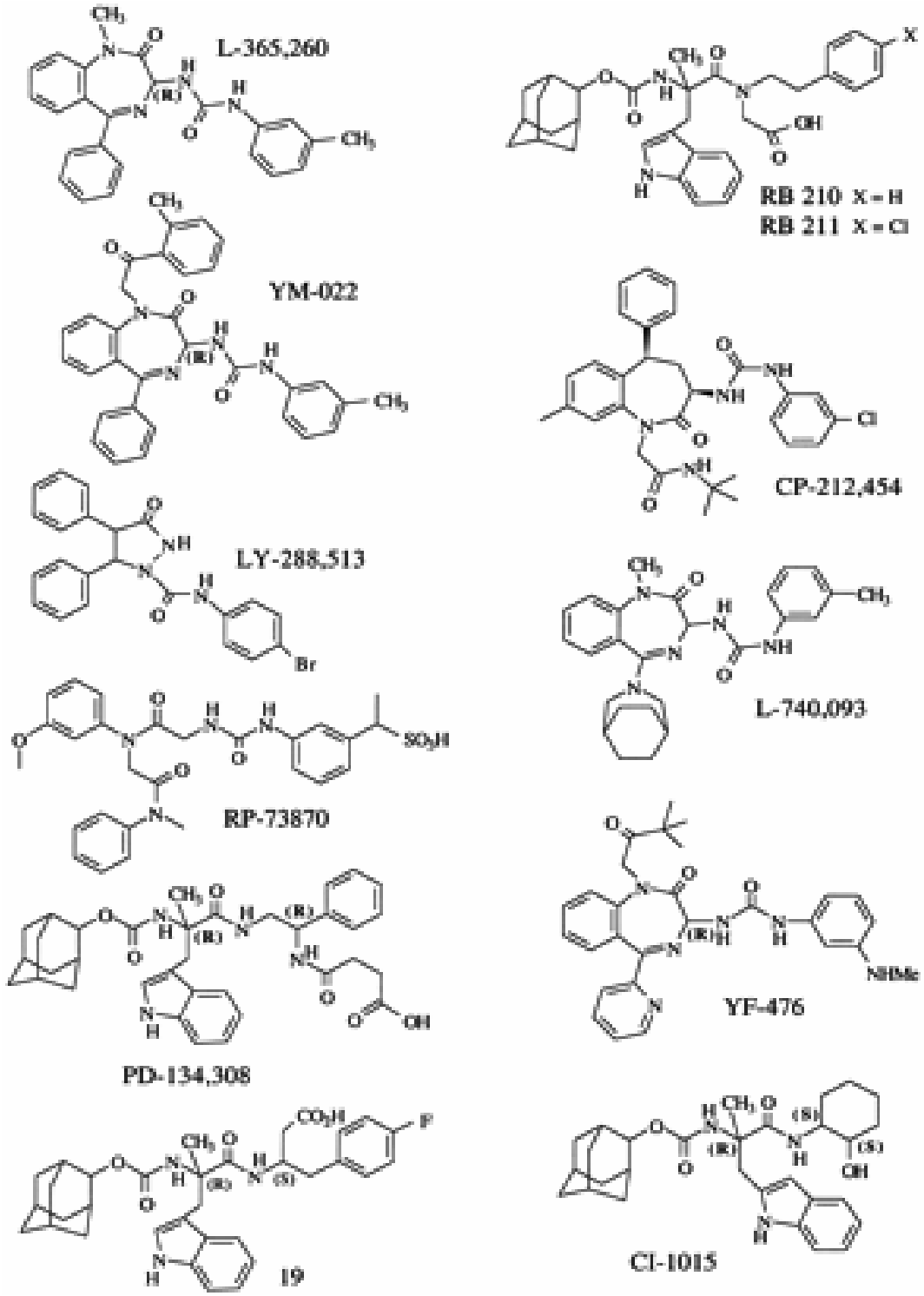


Tableau 4 : Affinités et sélectivités des antagonistes du récepteur CCK_2. Les constantes d'affinité K_i mesurées dans le cerveau et dans les membranes pancréatiques, sont données en nanomoles(nM). La sélectivité est estimée par le rapport des 2 constantes d'affinité.

Antagoniste	$K_i(CCK_2)$	Sélectivité ($K_i(CCK_1)$/ K_i (CCK_2))
L-365-260	7	115
PD-134,308	0.9	1600
LY-288,513	31	370
RP-73,870	0.5	3300
YM-022	0.1	1500
RB-210	14	110
CP-212,454	0.5	360
L-740,093	0.1	16000
YF-476	0.2[a]	565
CI-1015	3	967

[a] : Valeur en IC_{50}, concentration correspondant à une inhibition de 50% de l'activité biologique.

BIBLIOGRAPHIE

1 - Ivy, A. C. et Oldberg, E., *Am. J. Physiol.* 86, 599-613, **1928**.

2 - Harper, A. A. et Raper, H. S., *Am. J. Physiol*., 102, 115-125, **1943**.

3 - Mutt, V. et Jorpes, J. , *E. Biochem. J*., 125, 57-58, **1971.**

4 - Vanderhaeghen, J. J., Signeau, J. C. et Gepts, W., *Nature*, 257, 604-605, **1975**.

5 - Dockray, G. J., *Nature*, 264, 568-570, **1976**.

6 - Robberecht, P., Deschodt-Lanckman, M. et Vanderhaeghen, J.J., *Proc. Natl. Acad. Sci.* USA,75, 524-528, **1978**.

7 - Neuronal cholecystokinin Annals of the New York. Academy of Sciences Vol. 448, **1985**.

8 - Deschenes, R. J., Lorentz. L. J., Haun. R. S., Roos. B. A., Collier, K. J. et Divon, J. E., *Proc. Natl. Acad. Sci.* USA, 81, 726-730, **1984**.

9 - Rehfeld, J.F. et Hansen, H. F., *J. Biol .Chem.* 261, 5832-5840, **1986**.

10 – Noble, F., Wank, S. A., Crawley, J. N., Bradwejn, J., Seroogy, K. B., Hamon, M. et Roques, B .P. *Pharmacological Reviews*, 51, Issue 4, 745-781, **1999.**

11 – Jensen, S.L., Holst, J.J., Nielsen, O. V. et Rehfeld, J.F., *Digestion*, 22, 305-309, **1981**.

12 – Bodanszky, M., Natarajan, S., Hahne, W. et Gardner, J.D., *J. Med. Chem.* , 20, 1047-1050, **1977.**

13 – Jensen, R. T., Lemp, G. F. et Gardner J. D. *Proc. Natl. Acad. Sci.* USA, 77, 2079-2083, **1980.**

14 – Gardner, J. D., Knight, M., Sutliff, V. E., Tamminga, C.A. et Jensen, R.T., *Am. J. Physiol.*, 246, G292-295, **1984**.

15 – Howard, J. M., Knight, M., Jensen, R.T. et Gardner, J. D., *Am. J. Physiol.*, 247, G261-4, **1984.**

16 – Spanarkel, M., Martinez, J., Briet, C., Jensen, R.T. et Gardner, J. D., *J. Biol. Chem.,* 258, 6746-6749, **1983.**

17 – Beinfeld, M.C., *Life Sci*, 61,2359-2366, **1997**.

18 - Ondetti, M. A., Pluscec, J., Sabo, E. F., Shechan, U. T. et Williams, N. J., J. Am. Chem. Soc., 92, 195-199, **1970**.

19- Van Dijk, A., Richards, J.G., Trzeciak, A., Gillesen, D. et Möhler, H.J. *Neurosci*. 4, 1021-1033, **1984**.

20 – Jensen, R. T., Lemp, G. F. et Gardner, J. D., *J. Biol. Chem*. 257, 5554-5559, **1982**.

21 - Rehfeld, J.F. *J. Biol. Chem.* 253, 4022-4030, **1978**.

22 - Beinfeld, M.C.et Palkovits, M. *Neuropeptides* 2, 123-129, **1981**.

23 - Innis, R. B., Couea, F. M. A., Uhl, G., Schneider, B. et Synder, S.H. *Proc. Natl. Acad. Sci*. USA.76, 521-525, **1979**.

24 - Beinfeld, M. C.et Palkovits, M., *Neuropeptides* 2, 123-129, **1981**.

25 - Vanderhaeghen, J.J., Lotstra, F., Mey, J.D.et Gilles, C. *Proc .Natl. Acad .Sci.* USA 77,1190-1194, **1980**.

26 - Vanderhaeghen, J.J., Lostra, F., Vierendeels, G., Gilles, C., Deschepper, C. et Verbanck, P. *Peptides*, 81-88, **1981**.

27 - Dockray, G. J., Desmond, H., Gayton, R. J., Jonson, A. C., Raybould, H., Sharkey, K. A., Varra, A. et Williams, R. G. *Ann. N.Y .Acad. Sci.* 448, 32-43, **1985**.

28 - Miller, L. J., Jardine, I., Weissman, E. et Speicher, D. J., *Neurochem.* 43, 835-840, **1984**.

29 - Calam, J., Ellis, A. et Dockray, G. J. J. *clin. Invest*. 69, 218-225, **1982**.

30 - Eysselein, V. E., Reeve, J. R., Shiveley, J. E., Miller, C. et Walsh, J. H., *Proc. Natl. Acad. Sci.* USA, 81, 6565-6568, **1984**.

31 - Larson, L. I. et Rehfeld, J. F. *Brain Res.*165, 201-218, **1979**.

32 - Williams, J. A., Bailey, A. C., Preissler, M., Goldfine, I.D., *Diabetes*, 31, 674-682, **1982.**

33 - Walsh, J.H., Lamers, C. B. et Valenzuela, J. E. *Gastroenterology*, 82,438-444, **1982.**

34 - Liddle R.A. , *Am.J. Physiol*. 269, G319-27. Review **1995.**

35 - Crawley, J. N. et Corwin, R. L. *Peptides*, 15, 731-755, **1994.**

36 - Innis, R. B. et Synder, S. H., *Proc. Natl. Acad. Sci.* USA, 77, 6917-6921, **1980**.

37 - Saito, A., Sankaran, H., Goldfine, I. D. et Williams, J. A., *Science* 208, 1155-1156, **1980**.

38 - Sankaran, H.,Goldfine, T.D.,Deveney, C.W., Wong, K.Y. et Williams, J. A., *J. Biol. Chem.* 255, 1849-1853, **1980**.

39 - Dourish, C. T. et Hill, D.R. *Trends Pharmacol. Sci*., 8 207-208, **1987**.

40 - Wank, S. A., Pisegna, J. R. et de Weerth, A. *Proc. Natl. Acad. Sci.* USA, 89, 8691-8695 **1992b**.

41 - Wank, S. A., *Am. J. Physiol.* 269, G628-646, **1995.**

42 - Durieux, C., Coppey, M., Zajac, J. M. et Roques, B. P., *Biochem. Biophys. Res. Commun.*, 137,1167-1173, **1986b**.

43 - Menozzi, D., Vinayek, R., Jensen, R. T. et Gardner, J. D., *J. Biol. Chem.* 266, 10385-10391, **1991**.

44 - Talkad, V. D., Patto, R. J., Metz, D. C., Turner, R. J., Fortune, K. P.,Bhat, S. T. et Gardner, J. D., *Biochim. Biophys. Acta*, 1224, 103-1116, **1994**.

45 - Durieux, C., Corringer, P.J., Ruiz-Gayo, M., Bergeron, F. et Roques, B. P., *Mol. Pharmacol.,* 41,1089-1095, **1992**.

46 - Derrien, M., McCort-Tranchepain, I., Ducos, B., Roques, B.P. et Durieux, C., *Pharmacol. Biochem. Behav.*, 49, 133-141, **1994**.

47 - Wanks, S. A., Harkins, R., Jensen, R. T., Shapira, H., de Weerth, A. et Slattery, T., Proc. Natl. Acad. Sci. 89, 3125-3129, **1992a**.

48 - de Weerth, A., Pisegna, J.R., Huppil, K. et Wank, S. A., *Biochem. Biophys. Res. Commun.*194, 811-818, **1993a**.

49 - Ghanekar, D., Hadac, E. M., Holicky, E. L. et Miller, L. J., J. Pharmacol. Exp. Ther., 282, 1206-1212, **1997**.

50 - Ulrich, C.D., Ferber, I., Holicky, E., Hadac, E., Buell, G. et Miller, L.J., *Biochem. Biophys. Res. Comm.,* 196,204-211, **1993**.

51 - Kopin, A. S., Lee, Y., McBride, E. W., Miller, L. J., Lu, M., Lin, H. Y, Kolakowski, J., L. F. et Beinborn, M., *Proc. Natl. Acad. Sci. USA*, 89, 3605-3609, **1992**.

52 - Pisegna, J.R., de Weerth, A., Huppi, K. et Wank, S.A., *Biochem. Biophys . Res. Comm.*, 189, 296-303, **1992**.

53 - Lee, Y. M., Beinborn, M., McBride, E.W., Lu,M., Kolakowski, L.F. et Kopin, A.S., J. Biol. Chem ., 268,8164-8169, **1993**.

54 - Zimonjic, D. B., Popescu, N.C., Matsui, T.,Ito, M. et Chihara, K., *Cytogenet. Cell. Genet.*, 65, 184-185, **1994.**

55 - de Weerth, A., Pisegna, J.R. et Wank, S. A., *Am. J. Physiol.*, 265, G1116-G1121, **1993b**.

56 - Huppi, K., Siwarski, D., Pisegna, J. R. et Wank, S., *Genomics,* 25, 727-729, **1995**.

57 - Inoue, H., Iannotti, C. A., Welling, C. M., Veile, R., Donis-Keller, H. et Permutt, M. A., *Genomics,* 42, 331-335, **1997.**

58 – Holicky, E. L., Hadac, E. M., Ding, X. O. et Miller, L.J., *Am. J. Physiol. Gastrointest. Liver. Physiol.*, 281, G507-14, **2001**.

59 - Karnik, S. S., Sakmar, T.P., Chen, H. B. et Khorana, H.G., *Proc. Natl. Acad. Sci. USA*, 85, 8459-8463, **1988.**

60 - Dixon, R. A. F., Sigal, I. S., Candelore, M. R., Register, R. B., Scattergood, W., Rands, E. et Strader, C. D., *EMBO*, 6, 3269-3275, **1987**.

61 – Hulme, H.C., *Symp. Soc. Exp. Biol.*, 44, 39-54, **1990.**

62 – Liddle, R. A., Goldfine, I. D., Rosen, M. S., Taplitz, R. A. et Williams, *J. A., J. Clin. Invest.* 75, 1144-1152, **1985**.

63 – Gully, D., Fréhel, D., Marcy, C.,Spinazze, A., Lespy, L., Neliat, G., Maffrand, J. P. et Le Fur, G., Eur. J. Pharmacol., 232, 13-19, **1993.**

64 – Woodruff, G. N. et Hughes, J., *Annu. Rev. Pharmacol. Toxicol.*, 31, 469-501, **1991.**

65 – Van der Schaar, P. J., Bremer, Y., Lamers, C. B. et Masclee, A. A., Scand. J. Gastroenterol., 36, 361-366, **2001.**

66 – Dockray, G. J., *Dig. Dis. Sci*., 39, 82S-85S, **1994.**

67 – Rehfeld, J. F., larsson, L. I., Goltermann, N. R., Schwartz, T. W., Holst, J. J., Jensen, S. L. et Morley, J. S., *Nature*, 284, 33-38, **1980.**

68 – Folsch, U. R., Winckler, K. et Wormsley, K. G., Scand. J. Gastroenterol., 13, 663-671, **1978**.

69 – Ohlsson, B., Axelson, J., Sternby, B., Rehfeld, J.F. et Ihse, I., *Int. J. Pancreatol.,* 18, 59-66, **1995**.

70 – Logsdon, C. D., *J. Biol. Chem*., 261, 2096-2101, **1986.**

71 – Shulkes, A. et Baldwin, G. S., *Clin. Exp. Pharmacol. Physiol.,* 24, 209-216,**1997**.

72 – Gibbs, J., Young, R. C. et Smith, G. P., J. Comp. Physiol. Psychol. 84, 488-495, **1973**.

73 – Covasa, M. et Ritter, R. C., *Peptides*, 22, 1339-48, **2001.**

74 – Ebenezer, I. S., de la Riva, C. et Baldwin, B.A., Physiol. Behav., 47, 145-148, **1990**.

75 – Weatherford, S.C., Chiruzzo, F. Y. et Laughton, W. B., Am. J. Physiol., 262, R574-8, **1992**.

76 – Wolkowitz, O. M., Gertz, B., Weingartner, H., Beccaria, L., Thompson, et Liddle, R. A., *Biol. Psychiatry*., 28, 169-173,**1990.**

77 – Tamai, H., Takemura, J., Kobayashi, N., Matsubayashi, S., Matsukura, S. et Nakagawa, T., *Metabolism*, 42, 581-589, **1993.**

78 – Smith, G. P. et Gibbs, J., The development and proof of the CCK hypothesis of satiety, in Multiple Cholecystokinin Receptors in the CNS (Dourish, C.T., Cooper, S. J., Iversen, S. D. et Iversen, L. L. eds) pp 166-182, Oxford University Press, Oxford, **1992**.

79 – Bado, A., Levasseur, S., Attoub, S., Kermorgant, S., Laigneau, J. P., Bortoluzzi, M. N., Moizo, L., Lehy, T., Guerre-Millo, M., Marchand-Brustel, Y. et Lewin, M. J., *Nature*, 394, 790-793, **1998.**

80 – Matson, C. A., Wiater, M. F., Kuijper, J. L. et Weigle, D. S., *Peptides,* 18, 1275-1278, **1997.**

81 – Barrachina, M. D., Martinez, V., Wang, L., Wei, J. Y. et Tache, Y*., Proc. Natl. Acad. Sci. USA,* 94, 10455-60, **1997.**

82 - Crawley, J. N., *Trends Pharmacol. Sci.*, 12, 232-236, **1991**.

83 - Itoh, S. et Lal, H., Drug Dev. Res., 21, 257-276, **1990**.

84 – Weller, A., Tsitolovskya, L., Gispan, I. H. et Rabinovitz, S., *Peptides*, 22, 1317-23, **2001**.

85 - Lemaire, M., Piot, O., Roques, B. P., Böhme, G. A. et Blanchard, J. C., *Neuroreport.*, 3, 925-932, **1992**.

86 - Ruiz-Gayo, M., Dangé, V., Memant, I., Bégué, D., Gacel, G. et Roques, B. P., *Peptides*, 6, 415-420, **1985**.

87 – Holladay, M. W., Bennett, M. J., Tufano, M. D., Lin, C. W., Asin, K. E., Witte, D. G., Miller, T. R., Nikkel, A. L., Bednarz, L. et Nadzan, A. M., *J. Med. Chem.,*35, 2919-2928, **1992.**

88 - Shiosaki, K., Lim, C. W., Kopecha, H., Craig, R., Wagenaar, F. L., Bianchi, B., Miller, T., Witte, D. G. et Nadzan, A.M., *J. Med. Chem.,* 33, 2950-2952, **1990**.

89 – Pierson, M. E., Comstock, J. M., Simmons, R. D., Kaiser, F., Julien, R., Zongrone, J. et Rosamond, J. D., *J. Med. Chem.*, 40, 4302-4307, **1997**.

90 – Hirst, G. C., Aquino, C., Birkemo, L., Croom, D. K., Dezube, M., Dougherty, R. W., Erwin, G. N., Grizzle, M. K., Henke, B., James, M. K., Johnson, M. F., Momtahen, T., Queen, K. L., Sherrill, R. G., Szewczyk, J., Willson, T. M. et Sugg, E. E., *J. Med. Chem.,* 39, 5236-5245, **1996**.

91 – Willson, T. M., Henke, B. R., Momtahen, T., Myers, P. L., Sugg, E. E., Unwalla, R. J., Croom, D. K., Dougherty, R. W., Grizzle, M. K., Johnson, M. F., Queen, K. L., Rimele, T. J., Yingling, J. D. et James, M. K., *J. Med. Chem.,39,* 3030-3034, **1996**.

92 – Henke, B. R., Aquino, C. J., Birkemo, L., Croom, D. K., Dougherty, R. W., Erwin, G. N., Grizzle, M. K., Hirst, G. C., James, M. K., Johnson, M. F., Queen, K. L., Sherrill, R. G., Sugg, E. E., Suh, E. M., Szewczyk, J. W., Unwalla, R. J., Yingling, J. D. et Willson, T. M., *J. Med. Chem.,*40, 2706-2725, **1997**.

93 - Charpentier, B., Durieux, C., Pélaprat, D., Dor, A., Reibaud, M., Blanchard, J. C. et Roques, B. P., *Peptides*, 9,835-841, **1988a.**

94 - Ruiz-Gayo, M., Delay-Goyet, P., Durieux, C., Corringer, P.J., Baamonde, A., Gacel, G. et Roques, B. P., *J. Control. Release*, 13,147-155, **1990**.

95 - Durieux, C., Ruiz-Gayo, M. et Roques, B. P., *Eur. J. Pharmacol.,* 209, 185-193, **1991**.

96 - Fournié-Zaluski, M. C., Belleney, J., Lux, B., Durieux, C., Gérard, D., Gacel, G., Maigret, B. et Roques, B. P., *Biochemistry*, 25, 3778-3787, **1986**.

97 - Charpentier, B., Durieux, C., Menant, I. et Roques, B.P., J. Med. Chem., 30, 962-968, **1987**.

98 – Charpentier, B., Dor, A., Roy, P., England, P., Pham, H., Durieux, C. et Roques, B.P., J. Med. Chem., 32, 1184-1190, **1989**.

99 – Knapp, R. J., Vaughn, L. K., Fang, C. L., Yamamura, M. S., Hruby, V. J. et Yamamura, H. J., *J. Pharmacol. Exp. Ther.,* 255, 1278-1286, **1990**.

100 – Harhammer, R., Schafer, U., Henklein, P., Ott, T. et Repke, H., *Eur. J. Pharmacol.*, 209, 263-266, **1991**.

101 – Corringer, P. J., Weng, J. H., Ducos, B., Durieux, C., Boudeau, P., Böhme, A. et Roques, B. P., *J. Med. Chem.*, 36, 166-172, **1993.**

102 – Goudreau, N., Weng, J. H., Roques, B. P., *Biopolymers*, 34, 155-159, **1994**.

103 – Kolodziej, S. A., Nikiforovich, G. V., Skeean, R., Lignon, M. F., Martinez, J. et Marshall, G. R., *J. Med. Chem.,* 38, 137-149, **1995**.

104 – Blommaert, A. G. S., Dhôtel, H., Ducos, B., Durieux, C., Goudreau, N., Bado, A., Garbay, C. et Roques, B. P., *J. Med. Chem.,* 40, 647-658, **1997**.

105 – Weng, J. H., Bado, A., Garbay, C., et Roques, B. P., *Regul. Pept*, 65, 3-9, **1996a**.

106 – Million, M. E., Léna, I., Da Nascimento, S., Noble, F., Daugé, V., Garbay, C. et Roques, B. P., *Lett. Peptide Sci.,* 4, 407-410, **1997**.

107 - Peikins, S. J. et Withrich, K., *Biochem. Biophys. Acta*, 576, 409-423, **1979**.

108 - Barlas, N., Jensen R.T., Beinfeld, M. C. et Gardner, J. D., *Am. J. Physiol.,* 242, 161-167, **1982**.

109 – Martin-Martinez, M., Bartolomé-Nebreda, J. M., Gomez-Monterrey, I., Gonzalez-Muniz, R., Garcia-Lopez, M. T., Ballaz, S., Barber, A., Fortuno, A., Del Rio, J. et Herranz, R., *J. Med. Chem.,* 40, 3402-3407, **1997**.

110 - Hahne, W. F., Jensen, R. T., Lemp, G. F. et Gardner, J. D., *Proc Natl. Acad. Sci.* USA,78, 6304-6308, **1981**.

111 - Markovec, F., Chiste, R., Rovati, L. C. et Setnikar, I., Gastroenterology, 94, 279-283, **1988**.

112 – Akiyama, T., Tachibana, I., Hirohata, Y., Shirohara, H., Yamamoto, M. et Otsuki, M., *Br. J. Pharmacol.,*117, 1558-1564, **1996**.

113 - Fulcrand, P., Rodriguez, M., Galas, M. C., Lignon, M. F., Laur, J., Aumelas, A. et Martinez, J., *Int. J. Pept. Protein Res.,*32, 384-395, **1988**.

114 – Silvente-Poirot, S., Dufresne, M., Vaysse, N. et Fourmy, D., *Eur. J. Biochem.*, 215, 513-529, **1993a**.

115 – Chang, R. S. L., Lotti, V.J., Monaghan, R. L., Birnbaum, J., Stapley, E. O., Goetz, M. A., Albers-Schonberg, G., Patchett, A. A., Liesch, J. M., Hensens, O. D. et Springer, J. P., *Science*, 230, 177-179, **1985**.

116 - Chang, R. S. L. et Lotti, V.J., *Proc. Natl. Acad. Sci. USA,* 83, 4923-4926,**1986.**

117 – Satoh, Y., Matsuo, T., Sogabe, H., Itoh, H., Tada, T., Kinoshita, T. et Takaya, T., *Chem. Pharm. Bull. (Tokyo)*, 42, 2071-2083, **1994**.

118 – Itoh, H., Sogabe, H., Nakarai, T., Satoh, Y., Tomoi, M., Kadowaki, M., Matsuo, M., Tokoro, K. et Yoshida, K., *J. Pharmacol. Exp., Ther.,* 268, 571-575, **1994a**.

119 – Akiyama, T. et Otsuki, M., *Pancreas*, 9, 324-331, **1994**.

120 – Hull, R. A. D., Shankley, N. P., Harper, E. A., Gerskowitch, V. P., et Black, J. W., *Br. J. Pharmacol.,* 108, 734-740, **1993**.

121 - Boden, P. R., Higginbottom, M., Hill, D.R., Horwell, D. C., Hughes, J., Rees, D. C., Roberts, E., Singh, L., Suman-Chauhan, N. et Woodruff, G. N., *J. Med. Chem.*, 36, 552-565, **1993**.

122 – Gully, D., Fréhel, D., Marcy, C., Spinazze, A., Lespy, L., Neliat, G., Maffrand, J. P. et Le Fur, G., *Eur. J. Pharmacol.,*232, 13-19, **1993**.

123 - Taniguchi, H., Yazaki, N., Yomota, E., Shikano, T., Endo, T. et Nagasaki, M., *Eur. J. Pharmacol.,* 312, 227-233, **1996**.

124 – Bock, M. G., DiPardo, R. M., Evans, B. E., Rittle, K. E., Whitter, W. L., Veber, D. F., Anderson, P. S. et Freidinger, R. M., *J. Med. Chem.*, 32, 13-16, **1989**.

125 – Chambers, M. S., Hobbs, S. C., Fletcher, S. R., Matassa, V. G., Mitchell, P. J., Watt, A. P., Baker, R., Freedman, S. B., Patel, S. et Smith, A. J., *Bio. Med. Chem. Lett.*, 3, 1919-1924, **1993**.

126 – Semple, G., Ryder, H., Kendrick, D. A., Szelke, M., Ohta, M., Satoh, M., Nishida, A., Akuzawa, S. et Miyata, K., *Bioorg. Med. Chem. Lett.*, 6, 51-54, **1996a**.

127 – Semple, G., Ryder, H., Kendrick, D. A., Szelke, M., Ohta, M., Satoh, M., Nishida, A., Akuzawa, S. et Miyata, K., *Bioorg. Med. Chem. Lett.*, 6, 55-59, **1996b**.

128 – Showell, G. A., Bourrain, S., Neduvelil, J. G., Fletcher, S. R., Baker, R., Watt, A. P., Fletcher, A. E., Freedman, S. B., Kemp, J. A., Marshall, G. R., Patel, S., Smith, G. R. et Matassa, V. G., *J. Med. Chem.,* 37, 719-721, **1994**.

129 – Bock, M. G., DiPardo, R. M., Mellin, E. C., Newton, R. C., Veber, D. F., Freedman, S. B., Smith, A. J., Patel, S., Kemp, J. A., Marshall, G. R., Fletcher, A. E., Chapman, K. L., Anderson, P. S. et Freidinger, R. M., *J. Med. Chem.,* 37, 722-724, **1994**.

130 – – Chambers, M. S., Hobbs, S. C., Graham, M. I., Watt, A. P., Fletcher, S. R., Baker, R., Freedman, S. B., Patel, S., Smith, A. J. et Matassa, V. G., *Bioorg. Med. Chem. Lett.*, 5, 2303-2308, **1995**.

131 – Lowe, J. A., Drozda, S. E., Mclean, S., Bryce, D. K., Crawford, R. T., Zorn, S., Morrone, J., Appleton, T. A. et Lombardo, F. A., *Bioorg. Med. Chem. Lett.*, 5, 1933-1936, **1995**.

132 – Nishida, A., Miyata, K., Tsutsumi, R., Yuki, H., Akuzawa, S., Kobayashi, A., Kamato, T., Ito, H., Yamano, M., Katuyama, Y., Satoh, M., Ohta, M. et Honda, K., *J. Pharmacol. Exp. Ther.*, 269, 725-731, **1994**.

133 – Semple, G., Ryder, H., Rooker, D. P., Batt, A. R., Kendrick, D. A., Szelke, M., Ohta, M., Satoh, M., Nishida, A., Akuzawa, S. et Miyata, K., *J. Med. Chem.*, 40, 331-341, **1997**.

134 – Horwell, D. C., Hughes, J., Hunter, J. C., Pritchard, M. C., Richardson, R. S., Roberts, E. et Woodruff, G. N., *J. Med. Chem.,* 34, 404-414, **1991**.

135 – Augelli-Szafran, C. E., Horwell, D. C., Kneen, C., Ortwine, D. F., Pritchard, M. C., Purchase, T. S., Roth, B. D., Trivedi, B. K., Hill, D., Suman-Chauhan, N. et Webdale, L., *Bioorg. Med. Chem.,* 4, 1733-1745, **1996**.

136 – Blommaert, A. G., Weng, J. H., Dorville, A., McCort, I., Ducos, B., Durieux, C. et Roques, B. P., *J. Med. Chem.*,36, 2868-2877, **1993**.

137 – Fincham, C. I., Horwell, D. C., Ratcliffe, G. S. et Rees, D.C., *Biomed. Chem. Lett.,*2, 403-406, **1992b**.

138 –Bolton, G. L., Roth, B. D. et Trivedi, B. K., *Tetrahedron,* 49, 525-536, **1993**.

139– Fincham, C. I., Higginbottom, M., Hill, D. R., Horwell, D. C., O'Toole, J. C., Ratcliffe, G. S., Rees, D.C. et Roberts, E., *J. Med. Chem.*, 35, 1472-1484, **1992a**.

140 – Higginbottom, M., Hill, D. R., Horwell, D. C., Mostafai, E., Suman-Chauhan, N. et Roberts, E., *Bioorg. Med. Chem.,*1, 209-217, **1993**.

141 – Bellier, B., Da Nascimento, S., Meudal, H., Gincel, E., Roques, B. P. et Garbay, C., *Bioorg. Med. Chem. Lett.,*8, 1419-1424, **1998**.

142 - Bellier, B., McCort-Tranchepain, I., Ducos, B., Da Nascimento, S., Meudal, H., Noble, F., Garbay, C. et Roques, B. P., *J. Med. Chem.*, 40, 3947-3956, **1997.**

143 – Trivedi, B. K., Padia, J. K., Holmes, A., Rose, S., Wright, D. S., Hinton, J. P., Prithchard, M. C., Eden, J. M., Kneen, C., Webdale, L., Suman-Chauhan, N., Boden, P., Singh, L., Field, M. J. et Hill, D., *J. Med. Chem.*, 41, 38-45, **1998.**

144 – Howbert, J. J., Lobb, K. L., Brown, R. F., Reel, J. K., Neel, D. A., Mason, N. R., Mendelsohn, L. G., Hodgkiss, J. P. et Kelly, J. S., *Progress Toward CNS Therapeutic Targets (Dourisch C. T. et Cooper S. J. eds)* pp:28-37, Oxford University Press, London, **1992**.

145 – Pendley, C. E., Fitzpatrick, L. R., Capolino, A. J., Davis, M. A., Esterline, N. J., Jakubowska, A., Bertrand, P., Guyon, C., Dubroeucq, M. C. et Martin, G. E., *J. Pharmacol. Exp. Ther.*, 273, 1015-1022, **1995**.

CHAPITRE II

LES METHODES
DE
LA MODELISATION MOLECULAIRE

I – INTRODUCTION

Depuis l'époque de Kekulé, le chimiste fait usage de modèles mécaniques formés de boules et de tiges, pour représenter les molécules ou les cristaux, permettant de relier des concepts structuraux à des propriétés fonctionnelles ou à la réactivité.

Le premier système graphique piloté par un ordinateur, a été mis au point dès 1952[1] et s'est largement répandu dans les bureaux d'ingénierie et de conception à la fin des années 70[2]. L'écriture du premier programme de manipulation interactive de structures tridimensionnelles, en 1964, a laissé penser que l'infographie allait remplacer l'utilisation traditionnelle des modèles moléculaires CPK (Corey-Pauling-Koltum) ou Dreiding. En fait, les techniques info-graphiques ne sont utilisées couramment, dans le cadre de la recherche, qu'une vingtaine d'années plus tard, c'est à dire une fois que des progrès techniques, dans la conception des composants électroniques et de nouvelles architectures graphiques, ont permis de construire des machines dotées d'une puissance suffisante pour permettre la manipulation interactive[3,4]. Le développement important des moyens informatiques (mémoire plus importante, rapidité de calcul, info-graphie, développement de logiciels scientifiques) a permis à la chimie de s'enrichir d'outils informatiques, 'spécialement dédiés' à la représentation des molécules. Ainsi, un nouvel outil est apparu permettant à tout chimiste de mieux appréhender les problèmes liés au domaine moléculaire et qui est **la Modélisation Moléculaire.**

La modélisation moléculaire qui englobe différentes techniques de graphisme moléculaire a fait de l'ordinateur un lieu d'expériences numériques, grâce aux nombreuses possibilités de calcul théorique, qui permettent de dessiner, afficher, simuler, analyser, calculer et stocker les propriétés physico-chimiques des molécules. De ce fait, elle est considérée aujourd'hui comme une nouvelle technique de compréhension des phénomènes chimiques et une technique de travail à part entière dans le domaine de la chimie structurale.

Le point de départ de la modélisation est la construction de la structure moléculaire relative au problème posé. Cette construction peut se faire à partir d'atomes, de groupements fonctionnels et de molécules simples ou complexes qui sont fournis par des données spectroscopiques : RX, IR, RMN, Raman et UV se trouvant dans une banque de données interne au programme, ou externe telle que Cambridge Structural data (CSD) ou Protein Data Bank (PDB). Ensuite, selon le problème posé, les besoins et les objectifs de l'étude, le modélisateur utilise les méthodes théoriques afin d'améliorer le modèle de départ, en essayant

de lui 'soutirer' des informations. La comparaison de ces informations avec certaines données expérimentales permettra ensuite de valider ou d'affiner le modèle obtenu.

Les deux types d'approches théoriques de base qui existent actuellement pour aborder une étude par modélisation moléculaire sont :

La mécanique quantique et la mécanique moléculaire

Ces deux approches sont radicalement différentes. La principale caractéristique les distinguant est : contrairement à la mécanique quantique qui tient compte de la distribution électronique autour des noyaux, la mécanique moléculaire ne tient pas compte explicitement des électrons. En effet, cette dernière traite les atomes et les électrons qui leur sont associés comme des points matériels obéissant aux règles de la mécanique classique.

Dans ce chapitre, nous présentons un exposé de ces différentes méthodes théoriques dont certaines sont utilisées dans nos travaux.

II - LA MECANIQUE QUANTIQUE

Le formalisme de la mécanique quantique ayant été largement présenté dans les manuscrits de thèses soutenues antérieurement, à l'USTHB, nous nous limiterons ici, à décrire seulement les principes de base et à jalonner l'historique de la théorie quantique par les travaux qui nous ont semblé les plus marquants dans l'évolution de ses concepts.

1) - Bases et Principes

Les méthodes issues de la théorie quantique partent du principe, que la molécule est une collection de noyaux positifs et d'électrons négatifs en mouvement sous l'influence d'un potentiel coulombien.

En accord avec la mécanique classique, l'énergie des particules en interaction est la somme de leurs énergies cinétique et potentielle qui s'écrit :

$$E = T + V \qquad \textbf{(II-1)}$$

Pour rendre compte du caractère ondulatoire des particules, Schrödinger suggère de remplacer les potentiels classiques par leurs opérateurs linéaires associés :

$$T + V \rightarrow \hat{T} + \hat{V}$$

Dans ce cadre, tout système est traité en résolvant l'équation de Schrödinger[5], généralement indépendante du temps :

$$H\psi = E\psi \qquad \textbf{(II-2)}$$

H est l'opérateur hamiltonien associé à l'énergie totale E du système étudié et ψ la fonction d'onde décrivant le mouvement spatial de toutes les particules du système, dans le champ

défini par le potentiel V. Pour un système moléculaire à N électrons et à M noyaux, l'hamiltonien en unités atomiques (u.a) est donné par la relation suivante :

$$H = \sum_{i=1}^{N} -\frac{1}{2}\Delta_i + \sum_{k=1}^{M} \frac{1}{2M_K}\Delta_K + \sum_{K=1}^{M}\sum_{i=1}^{N} -\frac{Z_K}{R_{Ki}} + \sum_{J=1}^{N}\sum_{i<j}^{N} \frac{1}{r_{ij}} + \sum_{L=1}^{M}\sum_{K<L}^{M} \frac{Z_K Z_L}{R_{KL}} \qquad \textbf{(II-3)}$$

Avec Δ_i, le laplacien de l'électron i

$$\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$ en coordonnées cartésiennes.

Afin de mettre en pratique ces concepts et traiter l'équation de Schrödinger qui n'a de solutions que pour des systèmes simples il est nécessaire de mettre en œuvre une série d'approximations.

2) - Approximations usuelles

a) Approximation de Born-Oppenheimer

L'approximation fondamentale de M. Born et R. Oppenheimer[6] consiste à séparer le mouvement des électrons de celui des noyaux en considérant, lors de l'étude de la structure électronique moléculaire, que les noyaux sont immobiles du fait de leur grande masse comparée à celle de l'électron. On peut dire aussi que les électrons peuvent ajuster leur distribution lorsque les noyaux, plus lourds, changent leurs positions. Ceci nous permet d'écrire, sans entrer dans les détails de calcul:

$$H^{el}\psi^{el} = E^{et}\psi^{el} \qquad \textbf{(II-4)}$$

$$H^{el} = \sum_{i=1}^{N} -\frac{1}{2}\Delta_i + \sum_{K=1}^{M}\sum_{i=1}^{N} -\frac{Z_K}{R_{Ki}} + \sum_{j=1}^{N}\sum_{i<j}^{N} \frac{1}{r_{ij}} \qquad \textbf{(II-5)}$$

H^{el} est l'hamiltonien électronique, ψ^{el} est la fonction d'onde électronique qui dépend des coordonnées des électrons aussi bien que des coordonnées des noyaux et E^{el} est l'énergie électronique qui dépend des coordonnées des noyaux.

Les solutions de l'équation (II-4), décrivent le mouvement des électrons dans le champ des noyaux supposés fixes. Cette approximation a une grande importance pour les calculs de structure moléculaire, car elle nous autorise à optimiser **indépendamment** la structure électronique (calcul SCF, explicité plus bas) et les coordonnées nucléaires (optimisation de la géométrie).

La présence de l'opérateur répulsion électronique $r^{-1}{}_{ij}$ (éq.II-5), dépendant des coordonnées instantanées de deux électrons i et j, empêche la séparation des variables, ce qui rend insoluble l'équation de Schrödinger. Par conséquent, l'introduction d'une 'approche' de la fonction d'onde est inévitablement la prochaine étape dans les approximations.

b) Approximation orbitale

L'approximation orbitale[7,8] dont l'idée a été imaginée dès 1928 par D.R. Hartree, consiste à poser qu'une solution approchée particulière de la fonction d'espace est un produit de fonctions mono-électroniques : les orbitales. Afin de décrire la distribution électronique du système, la coordonnée de spin doit être introduite. Ceci conduit à une spin-orbitale, fonction des quatre variables de l'électron i, que l'on note :

$$\chi(i) = \psi(i)\eta(i) \qquad \textbf{(II-6)}$$

avec $\eta = \alpha \;\; ou \;\; \beta$, correspondant à ±1/2.

Dans le cadre de cette approximation et pour obéir au principe de Pauli, la fonction d'onde de l'état fondamental d'un système moléculaire à couches complètes, c'est à dire n'ayant pas d'électron célibataire, sera représentée par un déterminant dit de Slater[9], construit sur les (N/2) premières orbitales occupées.

$$\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_a(1)\ \chi_b(1) \ldots\ldots\ \chi_p(1) \\ \ldots\ldots\ldots\ \ldots\ldots\ldots\ldots \\ \chi_a(i) \ldots\ldots\ldots\ldots\ldots\ \chi_p(i) \\ \ldots\ldots\ \ldots\ldots\ \ldots\ldots\ldots \\ \chi_a(N) \ldots\ldots\ldots\ldots\ \chi_p(N) \end{vmatrix} \qquad \textbf{(II-7)}$$

A partir de l'hamiltonien électronique (éq.II-5) et de la forme de la fonction d'onde (éq.II-7), on peut trouver l'énergie électronique en utilisant la méthode des variations. Dans cette méthode, la meilleure fonction d'onde est recherchée en minimisant l'énergie par rapport aux paramètres de cette fonction.

3) - Méthodes du champ autocohérent ou SCF (Self Consistent Field)

a) Equations de Hartree Fock

L'application du principe des variations[10] à l'équation (II-4) conduit aux équations dites Hartree-Fock.

$$F\chi_k = e_k\chi_k \qquad \textbf{(II-8)}$$

Comme l'opérateur de Hartree Fock n'agit pas sur le spin électronique et la seule fois où on fait référence au spin est lors du remplissage des orbitales alors l'équation (II-8) peut s'écrire en remplaçant χ_k par l'orbitale correspondante :

$$F\psi_k = e_k \psi_k \qquad \textbf{(II-9)}$$

Avec

$$F(i) = h(i) + \sum_{l=1}^{N/2}\left(2J_l(i) - K_l(i)\right) \qquad \textbf{(II-10)}$$

$$h(i) = -\frac{1}{2}\nabla_i^{\,2} - \sum_{k=1}^{M}\frac{Z_k}{R_{Ki}} \qquad \textbf{(II-11)}$$

$$J_l(i)\psi_k(i) = \left[\int_j \psi_l(i)\frac{1}{r_{ij}}\psi_l(j)d\tau_j\right]\psi_k(i) \qquad \textbf{(II-12)}$$

$$K_l(i)\psi_k(i) = \left[\int_j \psi_l(j)\frac{1}{r_{ij}}\psi_k(j)d\tau_j\right]\psi_l(i) \qquad \textbf{(II-13)}$$

h(i) prend en compte le mouvement de l'électron et les interactions électron-noyaux. Le terme J_l est l'opérateur coulombien auquel correspond l'intégrale de Coulomb (éq.II-12). K_l est l'opérateur d'échange auquel correspond l'intégrale d'échange (éq.II-13). Ces 2 intégrales décrivent les interactions entre les électrons. Il faut noter que l'intégrale de coulomb a un analogue classique alors que l'intégrale d'échange provient de la nécessité d'antisymétriser la fonction d'onde. L'équation (II-9) de Hartree-Fock représente un jeu d'équations intégro-différentielles couplées et ne peuvent être résolues que de façon itérative. En effet, l'opérateur F qui permet d'obtenir les solutions ψ_k est exprimé en fonction de deux opérateurs *J* et *K* qui sont, eux mêmes, définis à partir des inconnues ψ_k. Dans la pratique, un jeu d'orbitales d'essai est choisi, on construit ensuite F pour résoudre l'équation II-9. La procédure est répétée avec le nouveau jeu d'orbitales jusqu'à ce que le champ électrostatique ressenti par un électron (champ provoqué par les autres électrons dans les autres orbitales) reste stationnaire ou jusqu'à ce que le système ait atteint sa cohérence interne, d'où le nom de *méthode du champ auto-cohérent* (Self Consistent Field), abrégé en utilisant les initiales anglaises SCF. Cependant, un problème existe et qui concerne la forme analytique des orbitales à utiliser.

b) Equations de Roothaan

Compte tenu de la nature des problèmes chimiques, Hall et indépendamment Roothaan[11] ont proposé de relier les orbitales moléculaires (OM) aux orbitales atomiques (OA) des atomes constituant la molécule : il s'agit de l'approximation LCAO (Linear Combination of Atomic Orbitals). Cette dernière consiste à exprimer toute OM comme une combinaison linéaire d'un nombre fini 'n' d'OA connues et centrées sur les atomes

$$\psi_k = \sum_{s=1}^{n} C_{ks}\phi_s \qquad \textbf{(II-14)}$$

Dans cette approximation, les équations de Hartree-Fock se transforment en un système d'équations non linéaires couplées : les équations de Roothaan et Hall

$$FC_k = e_k SC_k \qquad \textbf{(II-15)}$$

Dans l'expression (II-15), les inconnues sont les coefficients C_{ks} des orbitales atomiques utilisées. De plus, ces fonctions de base sont quelconques alors que les spin-orbitales sont orthogonales entre elles. Ceci conduit à l'apparition d'une matrice *S* dont chaque élément est l'intégrale de recouvrement entre 2 orbitales atomiques. La résolution du système d'équations nécessite donc une étape supplémentaire où de nouvelles fonctions, orthogonales entre elles, sont déduites des précédentes par une méthode d'orthogonalisation dont la plus employée est celle de Lowdin ; les calculs sont menés aussi de manière itérative jusqu'à convergence des énergies (seuil à fixer). Les orbitales atomiques utilisées sont en général déduites des solutions de l'équation de Schrödinger pour l'ion hydrogénoïde. Elles sont le produit d'une partie radiale et d'une partie angulaire exprimées dans un système de coordonnées sphériques (r,θ,φ) dont chaque atome est l'origine :

$$\phi_s \rightarrow \phi_{nlm}(r,\theta,\varphi) = N_{nl}(r)Y_{lm}(\theta,\varphi) \qquad \textbf{(II-16)}$$

n, l et *m* représentent les nombres quantiques qui sont à l'origine de la nomenclature de ces orbitales, à savoir : 1s pour le triplet (1,0,0), 2s pour (2,0,0), 2p pour (2,1,{-1..0..1}) etc…

Dans la pratique, c'est la forme paramétrée due à Slater qui est très employée. Les orbitales de type Slater (STO) décrivent assez bien le comportement des électrons à la périphérie de l'atome, elles conviennent donc bien au traitement des liaisons chimiques.

A ce niveau d'évolution des concepts de la théorie quantique, nous distinguons deux modalités de calcul permettant la résolution des équations de Hartree Fock-Roothaan : *ab initio* et semi-empirique.

c) Méthodes de calcul ab initio

Les calculs *ab initio* concernent tous les calculs SCF sans approximations supplémentaires à celles inhérentes à la méthode elle-même. Toutes les intégrales et tous les électrons de valence sont pris en compte dans les calculs. Cependant, les orbitales de Slater, compliquant considérablement le calcul des intégrales ont ralenti le développement de ce type de méthodes. La situation changea dans les années 60, quand ces orbitales furent remplacées par des « contractions » de gaussiennes qui sont des sommes de fonctions mathématiques dont les propriétés présentent l'avantage de permettre une intégration analytique de toutes les intégrales biélectroniques. Comme elles décrivent mal les propriétés atomiques, une combinaison linéaire de plusieurs gaussiennes est nécessaire pour obtenir une représentation

approximative convenable d'une OA de Slater. Il faut noter qu'il existe différentes bases de gaussiennes possibles pour effectuer un calcul *ab initio*[12,13]. Cependant, la taille des fichiers sur disque et le temps de calcul augmentent approximativement comme la puissance 4 du nombre de fonctions de base utilisées. En revanche, la qualité des résultats croît beaucoup moins vite et tend vers une certaine limite, compte tenu de l'approximation orbitale.

Le nombre important des intégrales biélectroniques, ainsi que la relative difficulté de leur calcul, contraignent les méthodes *ab initio,* encore de nos jours, au traitement de systèmes contenant au maximum une trentaine d'atomes lourds. Ce nombre varie évidemment avec le type de base utilisée et les caractéristiques de l'ordinateur.

En pratique, le chimiste a souvent besoin de résultats moins précis permettant cependant des comparaisons à l'intérieur de familles de composés. Son objectif est de tirer des conclusions qualitatives d'une certaine généralité. C'est cette motivation qui a suscité la mise au point de méthodes de calcul intermédiaires entre les méthodes *ab initio* et les méthodes empiriques qui s'inspirent de la méthode de Hückel[14,15].

d) Méthodes de calcul semi-empirique

d1) Principe

Ces méthodes ont toutes un objectif commun qui est de réduire dans une grande proportion le nombre d'intégrales à calculer et en particulier celui des intégrales biélectroniques. Elles se distinguent par les divers niveaux d'approximations pour le calcul de ces intégrales et sont toutes fondées sur les approximations suivantes :

* La base d'orbitales utilisée est constituée par les orbitales de Slater de la couche de valence.

* Les intégrales de recouvrement sont négligées dans la résolution des équations SCF : c'est ce qu'on appelle approximation du RDN «Recouvrement différentiel nul »

* Toutes les intégrales biélectroniques à trois ou quatre centres, sont supposées nulles. Certaines intégrales biélectroniques à un ou deux centres, sont également nulles, leur nombre et leur nature dépendent du niveau d'utilisation du RDN et sont à l'origine des méthodes de type : NDDO[16] « Neglect of Diatomic Differential Overlap », CNDO[17] « Complet Neglect of Differential Overlap » et INDO[18] « Intermediate Neglect of Differential Overlap ».

L'ensemble des méthodes semi empiriques développées antérieurement à 1969, ont constitué un réel progrès dans l'histoire des méthodes de la chimie quantique. Elles ont permis, en particulier, d'aborder dans un cadre de type Hartree-Fock semi-empirique la structure de molécules de grande taille. Malgré cela, ces méthodes n'étaient pas réellement satisfaisantes, elles étaient, en effet, incapables de prévoir quantitativement l'ensemble des propriétés

moléculaires, au moyen d'une paramétrisation unique. Le statut de ces méthodes commença à changer, lorsqu'en 1969, l'équipe de M. J. S. Dewar décida de déterminer, de façon empirique, toutes les intégrales non négligées au niveau INDO. Selon cette équipe, les raisons d'insuccès des méthodes SCF semi empiriques provenaient du fait que leur paramétrisation tendait trop souvent à simuler des calculs de type *ab initio*. Les améliorations successives de la paramétrisation ont conduit à la méthode MNDO[19] qui peut être considérée comme l'aboutissement satisfaisant de toutes les recherches initiées, dès 1960, sur les méthodes SCF semi empiriques. Cependant, MNDO décrit mal la liaison hydrogène, ce qui a conduit M.J.S. Dewar et al à développer la méthode AM1[20] (Austin Model 1), qui ne diffère de MNDO que par le procédé empirique utilisé pour le calcul des répulsions entre cœurs (noyaux entourés des électrons internes). En effet, selon cette équipe, les formules empiriques utilisées par MNDO conviennent mal aux distances de van der Waals. Dans le même axe de recherche J. J. P. Stewart a proposé, en 1989 sous le nom de PM3[21], une reparamétrisation complète de MNDO. Celle-ci s'est effectuée sur des séries de molécules test au moyen d'un procédé automatique.

Il est important de signaler que le groupe de Ruiz-Lopez du Laboratoire de chimie théorique de Nancy I a entrepris des tentatives, révélées fructueuses, afin d'améliorer la méthode PM3 en vue de l'adapter à la description des interactions intermoléculaires[22].

d2) Avantages et inconvénients

Les nombreux travaux réalisés au moyen des méthodes semi-empiriques, développées avant 1976, ont permis de mieux cerner les faiblesses de chaque méthode et de déterminer le champ d'application sûre de chacune d'elles. En effet, elles sont paramétrées au moyen d'un critère qui se réfère à une caractéristique moléculaire choisie (distribution électronique, énergie de formation ou encore énergie de transition). C'est la raison pour laquelle, ces méthodes peuvent difficilement prétendre à l'universalité et leurs résultats sont toujours un peu suspects aux yeux de la théorie : ce qui est d'ailleurs à l'origine d'une certaine disgrâce en faveur des méthodes *ab initio*. Cependant, les coûts prohibitifs des calculs *ab initio* et l'étendue des moyens de calcul à mettre en œuvre pour ces derniers, alors que les ordinateurs personnels permettent des calculs semi-empiriques sur des molécules de taille respectable, ont poussé depuis l'apparition de MNDO vers la réhabilitation de ces méthodes semi-empiriques. Par ailleurs, ces méthodes ont l'avantage par rapport aux méthodes *ab initio* dites plus rigoureuses de tenir compte de la corrélation électronique intra-atomique et cela de façon implicite. En effet, l'usage de données atomiques ou moléculaires d'origine expérimentale permet de tenir compte, au moins partiellement, de certains effets de corrélation.

En résumé, les méthodes semi-empiriques disponibles actuellement, comportent un nombre substantiel d'avantages qui en font des outils de choix, chaque fois qu'il s'agit d'étudier l'évolution d'une propriété physique ou chimique à l'intérieur d'une famille de composés.
Dans la pratique, le choix de la méthode est dicté par la taille de la molécule, les informations recherchées (charges, géométries optimisées, chaleur de formation, chemins de réaction, etc.…), la précision souhaitée pour ces grandeurs et les moyens techniques utilisés.

Parmi ces méthodes, CNDO est incontestablement la plus simple des méthodes du type SCF. Elle fournit de bonnes valeurs pour les longueurs et les angles de liaison, ainsi que pour les distributions de charges, mais conduit à des valeurs de l'énergie qui sont toujours éloignées des grandeurs expérimentales. Ce défaut peut être gênant dans la recherche des conformères d'une molécule. En outre, elle souffre d'un autre défaut qui est son inaptitude à prendre en compte les différences d'énergie d'interaction entre électrons de même spin et électrons de spin contraire (intégrales d'échange négligées). De ce fait, elle ne permet pas le calcul de certaines grandeurs liées au spin électronique. En revanche, les méthodes de type NDDO, telles que MNDO et ses versions révisées AM1 et PM3 sont nettement plus lourdes du fait d'approximations moins poussées. Néanmoins, elles approchent correctement la plupart des grandeurs intéressant un chimiste : énergie de formation, moments dipolaires et multipolaires, constantes de force et elles fournissent, en particulier, des géométries moléculaires réalistes. Ceci nous a motivé à utiliser le cadre semi-empirique dans une partie des recherches menées dans cette thèse.

4) - Méthodes de la fonctionnelle de la densité

Ces méthodes dites de la fonctionnelle de la densité (DFT) originellement développées aux problèmes de l'état solide, ont actuellement un important potentiel pour l'étude des systèmes moléculaires et des problèmes chimiques. Elles reposent sur une approche différente des méthodes du type SCF et trouvent leur fondement dans un théorème dû à Hohenberg et Kohn[23] qui ont démontré que toutes les propriétés d'un système dans l'état fondamental non dégénéré sont complètement déterminées par sa densité électronique $\rho(r)$. Ceci permet de déduire que l'énergie peut être considérée comme une fonctionnelle de la densité. Cette fonctionnelle n'est évidemment pas connue du fait que l'on ne sait pas résoudre exactement l'équation aux valeurs propres d'un système à plusieurs électrons. Cette difficulté a été contournée par Kohn et Sham qui ont étendu le principe variationnel à la densité[24]. Ils ont montré que la densité électronique exacte correspond au minimum de l'énergie qui peut être évaluée au moyen de trois termes : énergie cinétique, énergie d'interaction coulombienne et

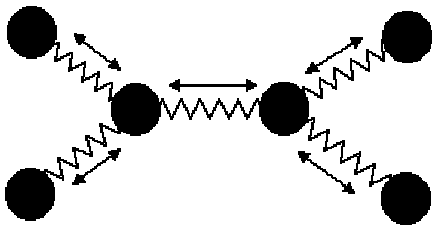
une énergie liée aux effets d'échange et de corrélation. Le problème se ramène donc à rechercher une expression approchée de ce dernier terme à partir de la densité. On associe alors à cette énergie deux fonctions des densités de spin ($\pm\frac{1}{2}$) et dont l'expression découle du traitement statistique du gaz homogène d'électrons. Les équations, à résoudre, ont une forme qui rappellent celles de Hartree-Fock mais en diffèrent par le fait qu'elles font intervenir la densité électronique totale du système dans les termes d'interaction entre électrons. Le succès, de ces méthodes DFT qui se distinguent par les techniques de calcul de la fonction d'onde et par l'expression du terme d'échange-corrélation, est dû surtout au volume des calculs qu'elles nécessitent, beaucoup plus réduit que dans les méthodes *ab initio.* Ces méthodes sont très 'attractives' en raison de leur très bon rapport précision/rapidité.

III – LA MECANIQUE MOLECULAIRE

1) - Bases et principes

La mécanique moléculaire considère la molécule comme un modèle mécanique déformable que nous illustrons dans la figure suivante :

Figure 1 : Représentation mécanique d'une structure moléculaire



La mécanique moléculaire (MM) est, d'une certaine manière, la formalisation mathématique des modèles moléculaires de boules et de tiges en termes d'atomes liés par des ressorts de raideur et de longueur variables. L'idée de base de cette méthode est que les atomes dans une molécule ont tendance à former entre eux des liaisons qui, pour un type donné, ont une longueur standard ou 'idéale', ces liaisons forment également entre elles des angles standards ou 'idéaux'. Dans la construction de la molécule, si celle-ci n'est pas très simple, toutes les tendances ne peuvent évidemment pas être satisfaites du fait qu'elles entrent toutes en conflit. La géométrie construite résulte en fait d'un compromis qui fait que certaines liaisons ou certains angles s'écartent de leurs valeurs 'idéales'. Ces écarts se traduisent par un accroissement de son énergie interne que l'on nomme 'énergie potentielle' ou 'énergie de contrainte' ou encore 'énergie stérique'. Contribuent également à cette énergie stérique des

interactions entre atomes non liés dont les plus importantes sont les interactions électrostatiques et de van der Waals. Il résulte de cette définition que les conformations sans énergie stérique (ou conformations de référence) et les valeurs optimales des paramètres géométriques des molécules sont des vues de l'esprit qui ne sont jamais observées. Cette énergie stérique correspond en fait à la différence d'énergie entre la conformation réelle de la molécule et une conformation hypothétique dans laquelle tous ces paramètres géométriques ont des valeurs optimales correspondant aux valeurs d'équilibre.

Les idées de base de la MM datent de 1930[25], mais la première application de cette méthode a eu lieu en 1946[26]. Avec l'avènement des ordinateurs dans les années 50 et au-delà l'intérêt, dans cette approche de la détermination et de la compréhension de la structure moléculaire, a cru rapidement, à tel point que la MM peut être considérée aujourd'hui comme une des méthodes standards de la chimie structurale. L'inconvénient majeur de cette méthode est qu'elle est incapable de décrire les systèmes impliqués dans la formation et/ou la rupture de liaisons. En revanche, elle s'est largement développée pour permettre de déterminer la conformation des molécules, leurs propriétés thermodynamiques et leurs spectres vibrationnels. Il faut noter que ces développements trouvent leurs fondements dans les travaux de P. Claverie[27a,b,c].

Le but de cette méthode de calcul est de déterminer la géométrie d'une structure moléculaire, par minimisation de son énergie. Elle permet également de calculer la variation d'énergie induite par la perturbation de cette structure (lors d'un changement de conformation, d'un changement de configuration, de l'introduction d'une nouvelle fonction…).

La MM repose implicitement sur l'hypothèse de Born-Oppenheimer[6]. Elle traite les interactions atomiques au moyen d'un potentiel effectif entre noyaux conformément aux lois de la mécanique classique. Ce potentiel d'interaction engendré de façon implicite par les interactions électroniques, s'exprime au moyen de variables géométriques centrées sur les noyaux des atomes formant la molécule. Cette méthode s'appuie sur une représentation de l'énergie qui est fonction de la géométrie de la molécule et du type d'atomes la constituant. Ainsi, comme on le verra dans le paragraphe suivant, les variables mises en jeu dans l'expression mathématique de l'énergie, sont les seules coordonnées internes de la structure : les longueurs de liaison entre deux atomes liés, les distances entre deux atomes non liés, les angles de valence entre trois atomes liés et les angles dièdres entre quatre atomes successifs. Ces coordonnées étant elles-mêmes définies par la seule position des atomes constituant le système moléculaire.

2) - Le champ de force

a) - Définition

Suite à ce qui vient d'être dit, la mécanique moléculaire est donc une méthode empirique résultant de l'ajustement de résultats expérimentaux sur des fonctions mathématiques simples. Cet ensemble de fonctions de potentiel ou de modèles mathématiques représentant l'énergie stérique d'une molécule est appelé '**champ de forces**'. Sous ce terme sont en fait regroupés deux éléments : d'une part, l'expression des différentes fonctions contribuant au calcul énergétique et, d'autre part, les valeurs des différentes constantes paramétrant ces fonctions. Ces paramètres ajustables (valeurs standards de longueurs, angles, constantes de force, etc.…) sont optimisés de façon à obtenir le meilleur accord entre les grandeurs calculées et grandeurs expérimentales des molécules : géométrie, énergie conformationnelle, chaleur de formation et si possible d'autres propriétés. Dans le cadre de cette mécanique, on fait toujours l'hypothèse que les paramètres ajustés, pour rendre compte de ces propriétés sur un ensemble de composés simples, sont applicables à d'autres composés. Un champ de forces qui n'est qu'une représentation empirique des aspects dynamiques et énergétiques des systèmes moléculaires est, en général, limité à une classe de molécules ayant suffisamment de ressemblances structurales et fonctionnelles. La fiabilité d'un champ de forces dépend donc de la cohérence de l'ensemble de ses paramètres. Il ne doit être utilisé que pour l'étude de systèmes analogues à ceux sur lesquels il a été calibré.

Actuellement, il existe de nombreux champs de forces qui évoluent avec la croissance rapide des applications. A l'origine, ils ont été construits, pour la plupart, sur des hydrocarbures et ensuite étendus aux molécules renfermant des hétéro atomes. Ils incluent tous, dans leur conception, cinq types de termes potentiels qui traduisent :

- l'élongation et/ou la compression de liaisons (stretching)
- la déformation angulaire entre 3 atomes liés (flexion, bending)
- la torsion de l'angle dièdre formé par 4 atomes liés
- les interactions de van der Waals
- les interactions électrostatiques.

L'énergie stérique de la molécule peut donc s'écrire :

$$E_S = \underbrace{\sum E_{stret} + \sum E_{bend} + \sum E_{tor}}_{E_{\text{cov alente}}} + \underbrace{\sum E_{vdW} + \sum E_{\acute{e}\,lec}}_{E_{\text{non cov alente}}} \qquad \textbf{(III-1)}$$

Dans cette expression, les sommes s'étendent respectivement sur toutes les liaisons, les angles de valence, les angles dièdres et les interactions entre tous les atomes non liés. Chaque terme de chacune des sommes a une expression analytique simple qui est constituée le plus souvent, d'un seul terme principal représentant une première approximation et, éventuellement, de termes correctifs. Toutes ces expressions sont suffisamment simples, pour être calculées très rapidement au moyen des ordinateurs et permettent donc le traitement de très grosses molécules telles que les protéines . Cette énergie est fonction de différents termes représentant une déformation moléculaire à partir d'une géométrie arbitraire de référence et qui varient d'un champ de force à un autre. De nombreux champs de forces sont actuellement disponibles dans différents logiciels de modélisation. Selon le champ de force utilisé, pour le calcul de l'énergie, les structures moléculaires modélisées et leurs propriétés seront exactes ou non. Le choix du champ de forces est donc à faire sur la base des résultats déjà obtenus dans la littérature concernant leurs applications aux systèmes moléculaires. Les champs de force les plus connus et utilisés sont : MM2 et MM3 développés par Allinger et al[28,29] pour les petites molécules organiques ; AMBER [30,31], GROMOS [32,33], CHARMM [34], CVFF [35,36], CFF91[37] et ECEPP[38,39] s'adaptent mieux aux systèmes biologiques polypeptidiques ou polynucléotidiques. Bien entendu, cette liste n'est pas complète et d'autres champs de force sont aussi disponibles.

b) – Expression générale des différents termes d'un champ de force

La formulation générale d'un champ de forces peut être donnée de la façon suivante :

$$V = V_{covalent} + V_{non\ covalent} \qquad \textbf{(III-2)}$$

avec

$$V_{covalent} = V_d + V_\theta + V_\phi + V_\omega + (V_{dd'} + V_{\theta\theta'} + V_{d\theta} + V_{\phi\theta\theta'} + V_{\omega\omega'}) \qquad \textbf{(III-3)}$$

$$V_{non\ covalent} = V_{vdW} + V_{élec} + V_{HB} \qquad \textbf{(III-4)}$$

Nous donnons ci-dessous la définition et l'expression mathématique de chacun des termes potentiels. Une visualisation graphique des différentes et principales déformations qui s'exercent sur une liaison d'une molécule est présentée dans la ***figure 2.***

V_d : *Potentiels de liaisons*

En général, à température ordinaire et en l'absence de réaction chimique, les déformations des liaisons sont faibles et peuvent donc être décrites par un potentiel harmonique (éq.III-5). Cette

approche correspond, en mécanique classique, à la description, à partir de la loi de Hooke, des propriétés d'un ressort.

$$\sum_{d}\frac{1}{2}\mathrm{K}_{d}\left(d-d_{0}\right)^{2} \qquad \textbf{(III-5)}$$

Il faut noter que les propriétés d'élongation et de compression d'une liaison sont mieux décrites par un potentiel de Morse (éq.III-6), mais la présence d'une fonction exponentielle, dans ce potentiel, exige un temps de calcul plus important. Ceci explique le fait que le potentiel harmonique lui soit préféré dans la majorité des cas. Dans ces équations, K_{d} *et* D_{d} sont des constantes de forces qui varient avec la nature de la liaison d, et d_0 la distance à l'équilibre.

$$\sum_{d} D_{d}\left[1-e^{-\alpha\left(d-\ d_{0}\right)^{2}}\right] \qquad \textbf{(III-6)}$$

$\mathbf{V_\theta}$: *Potentiels d'angles de valence*

Pour les mêmes raisons que précédemment les déformations des angles de valence s'expriment, en général, à l'aide d'un potentiel harmonique de Hooke(éq.III-7). K_θ est la constante de force, θ est l'angle de valence et θ_0 la valeur de cet angle à l'équilibre.

$$\sum_{\theta}\frac{1}{2}K_{\theta}\left(\theta-\theta_{0}\right)^{2} \qquad \textbf{(III-7)}$$

$\mathbf{V_\phi}$ *: Potentiels d'angles dièdres*

L'introduction d'un terme de torsion garantit la bonne représentation de la rotation de groupements autour d'une liaison donnée. A la différence des deux potentiels précédents, l'approximation harmonique ne peut être utilisée puisque les déformations sont souvent importantes. L'expression mathématique la plus utilisée et qui permet de décrire l'évolution de l'énergie d'une molécule en fonction de l'angle de torsion est un développement en série de Fourier (éq.III-8). K_ϕ est la constante de force, n la multiplicité ou périodicité, ϕ est l'angle de torsion et ϕ_0 la phase du potentiel de cet angle.

$$\sum_{\phi}\sum_{n}\ K_{\phi}\ \left[1-\cos\left(n\,\phi-\phi_{0}\right)\right] \qquad \textbf{(III-8)}$$

$\mathbf{V}_{\boldsymbol{\kappa}}$: *Potentiels hors du plan ou potentiels d'angles dièdres impropres*

La contribution des angles dièdres impropres qui représentent les déformations (généralement faibles) hors du plan d'atomes de carbone hybridés en sp^2 par exemple, est représentée par un potentiel harmonique (éq.III-9). K_ω est la constante de force, ω l'angle donnant l'écart à la planéité de l'atome en question et ω_0 l'angle optimal.

$$\sum_{\omega} \frac{1}{2} K_{\omega} (\omega - \omega_0)^2 \qquad \textbf{(III-9)}$$

$V_{élec}$: *Potentiel dû aux interactions électrostatiques*

Ce terme traduit les interactions électrostatiques entre 2 atomes porteurs de charges ponctuelles q_i et q_j , séparés par une distance r_{ij} et dans un milieu caractérisé par une constante diélectrique ε. Ce terme peut s'avérer particulièrement important dans les molécules contenant des hétéro atomes. On utilise en général, un potentiel coulombien simple avec la possibilité d'utiliser un rayon de troncature, qu'on explicitera dans la suite. ε peut être soit un entier, soit dépendant de r_{ij}.

$$\sum_{ij} \frac{q_i q_j}{\varepsilon r_{ij}} \qquad \textbf{(III-10)}$$

V_{vdW} : *Potentiel dû aux interactions de van der Waals*

Ces interactions qui permettent de décrire la répulsion des atomes à courte distance et leur attraction à longue distance sont représentées le plus souvent par un potentiel de Lennard-Jones[40] (éq.III-11). r_{ij} est la distance entre les atomes *i* et *j* et ε_{ij}, la profondeur du puits de potentiel à la distance inter atomique σ_{ij} qui correspond au minimum de la fonction.

$$\sum_{ij} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \right] \qquad \textbf{(III-11)}$$

Comme le montre la ***figure 2***, l'énergie de van der Waals qui intervient sur une petite échelle, est importante dès lors que les deux atomes sont proches, mais diminue rapidement quand les deux atomes interagissant sont séparés d'une distance de quelques angströms.

$V_{dd'}$, $V_{d\theta}$, $V_{\theta\theta'}$, $V_{\omega\omega'}$, $V_{\phi\theta\theta'}$: *Potentiels croisés*

Les potentiels croisés sont dus au couplage entre les déformations de différentes coordonnées internes (éqs III-12). Le couplage entre 2 liaisons (d et d') et l'angle formé entre elles, permet de rendre compte du phénomène d'élongation des liaisons dans le cas où l'angle de valence devient petit. Contrairement à tous les autres termes, leur présence dans la fonction énergie n'a pas une importance décisive sur le résultat des calculs, leur prise en compte est donc souvent négligée.

$$\sum_{d} \sum_{d'} K_{d\,d'} (d - d_0)(d' - d'_0) \qquad \textbf{(III-12a)}$$

$$\sum_{\theta}\sum_{\theta'} k_{\theta\theta'}\left(\theta-\theta_0\right)\left(\theta'-\theta'_0\right) \qquad \textbf{(III-12b)}$$

$$\sum_{d}\sum_{\theta} K_{d\theta}\left(d-d_0\right)\left(\theta-\theta_0\right) \qquad \textbf{(III-12c)}$$

$$\sum_{\omega}\sum_{\omega'} k_{\omega\omega'}\omega\ \omega' \qquad \textbf{(III-12d)}$$

$$\sum_{\phi} k_{\phi\theta\theta'}\cos\phi\left(\theta-\theta_0\right)\left(\theta'-\theta'_0\right) \qquad \textbf{(III-12e)}$$

V_{HB} :*Potentiel dû à la liaison hydrogène*

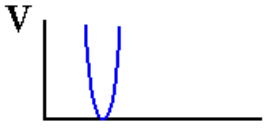
Certains champs de forces utilisent des potentiels particuliers pour décrire l'énergie de la liaison hydrogène. Cependant, une fonction du type Lennard-Jones est largement utilisée. Signalons que ce potentiel n'est pas défini dans tous les champs de forces et que sa présence dans le calcul de l'énergie potentielle est facultative.

$$\sum_{paires}\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{10}} \qquad \textbf{(III-13)}$$

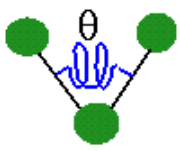
Figure 2 : Illustration graphique d'un champ de force

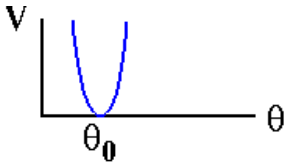
1) Elongation et compression

d

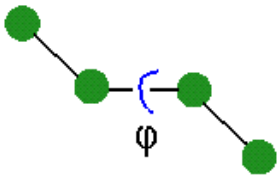


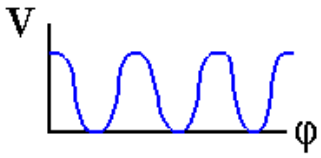
2)Flexion



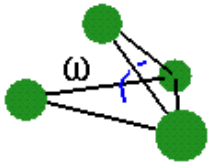


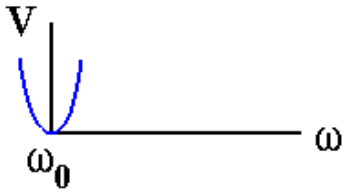
3)Torsion



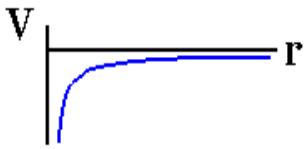


4)Angles impropres

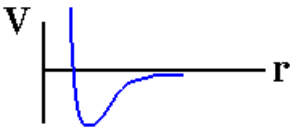




5)Électrostatiques



6) Van der Waals



Il faut noter que des modèles plus raffinés concernant le potentiel dû aux interactions électrostatiques ont été développés récemment, dans le but de prendre en compte la polarisabilité des atomes[41]. En effet, le phénomène de polarisation, ne peut plus être négligé dans le cadre des études incluant certains atomes, en particulier, ceux dotés de fortes charges, tels que les métaux. Ces modèles, basés sur le concept de fluctuations de charges dues à l'environnement des molécules, ont été développés pour les simulations des métalloprotéines où le fait de négliger la polarisation présente réellement un problème.

De plus, différentes approches ont été réalisées afin d'évaluer les propriétés diélectriques d'un système macromoléculaire dans lequel le solvant n'est pas inclus de manière implicite. Généralement, lorsque tous les atomes d'un système sont explicitement inclus dans le calcul, on utilise une constante diélectrique relative de 1, mais d'autres valeurs sont possibles selon que l'on désire simuler implicitement la présence d'un solvant, ou bien que l'on désire inclure

un effet d'écran : on pourrait prendre alors une constante diélectrique ε variant avec la distance inter atomique.

Les interactions non liantes devraient, *a priori,* être calculées pour toutes les paires d'atomes non liés du système. Cependant, afin de réduire le temps de calcul sur les systèmes de plusieurs milliers d'atomes, on a recourt à des fonctions de troncature, des interactions entre atomes non liés, qui spécifient une distance inter atomique r_{off} au-delà de laquelle les interactions non liées ne seront plus calculées. Dans le but d'éviter l'annulation brutale de ces interactions, ce qui aurait pour effet de créer artificiellement une force infinie, on utilise des fonctions analytiques dont la valeur dépend de la distance 'r' entre les atomes telles que les fonctions switch (r) ou schift (r) qui sont données par les expressions suivantes :

$$\text{Switch (r)} = \begin{cases} 1 & \text{Pour } r < r_{on} \\ \dfrac{\left(r_{off} - r\right)^2 \left(r_{off} + 2r - 3r_{on}\right)}{\left(r_{off} - r_{on}\right)^3} & \text{Pour } r_{on} \leq r \leq r_{off} \\ 0 & \text{Pour } r > r_{off} \end{cases} \qquad \textbf{(III-14)}$$

r_{on} est une distance fixée et à partir de laquelle la contribution des interactions de van der Waals diminue jusqu'à devenir nulle à partir de r_{off}.

$$\text{Schift (r)} = \begin{cases} \left[\left(\dfrac{r}{r_{off}}\right)^2 + \left(\dfrac{r}{r_{off}}\right)^4\right] & \text{Pour } r \leq r_{off} \\ 0 & \text{Pour } r > r_{off} \end{cases} \qquad \textbf{(III-15)}$$

L'expression de l'énergie stérique décrite ici est souvent élargie dans la plupart des champs de forces existant actuellement, afin d'obtenir des résultats plus conformes aux mesures expérimentales ou aux calculs quantiques poussés de type ***ab initio*** et qui représentent actuellement la principale référence pour le raffinement des champs de force. Le développement de plus en plus raffiné des champs de forces continue d'aboutir à des outils généraux qui peuvent être utilisés sur un panel de plus en plus large de molécules et ne sont plus restreints, comme auparavant, à une famille limitée de composés tels que les systèmes biologiques. En effet, la version initiale du champ de forces CVFF (Consistent Valence Force

Field) utilisé dans le cadre d'une partie de nos recherches, limitait son application principalement aux protéines, alors que les versions successives permettent l'étude d'une gamme étendue de composés[35,36].

3) - Minimisation de l'énergie potentielle

L'étape qui suit le choix du champ de force permettant de reconnaître les types d'atomes et de définir les liaisons, consiste à calculer l'énergie de la structure moléculaire. Comme nous l'avons dit précédemment, le but de la mécanique moléculaire est de déterminer la géométrie de la molécule qui aurait une énergie minimale. Cette recherche se fait au moyen de méthodes de minimisation qui consistent à trouver les valeurs des variables pour lesquelles la valeur de la fonction énergie potentielle est minimale. La non linéarité de la fonction et le nombre important de degrés de liberté compliquent considérablement le problème mathématique. Il existe en fait deux types de minima : le minimum global qui correspond à la valeur la plus faible que peut prendre la fonction énergie potentielle, et les minima locaux qui correspondent aux points où la fonction a une valeur minimale dans un voisinage fini. Il faut noter l'importance de la conformation initiale de la molécule, puisque c'est à partir de cette conformation que la méthode de minimisation détermine la direction de recherche menant à un minimum ainsi que la distance à ce minimum. La ***figure 3***, illustre bien la convergence de la fonction énergie potentielle vers un minimum global ou local selon la conformation initiale de la structure moléculaire.

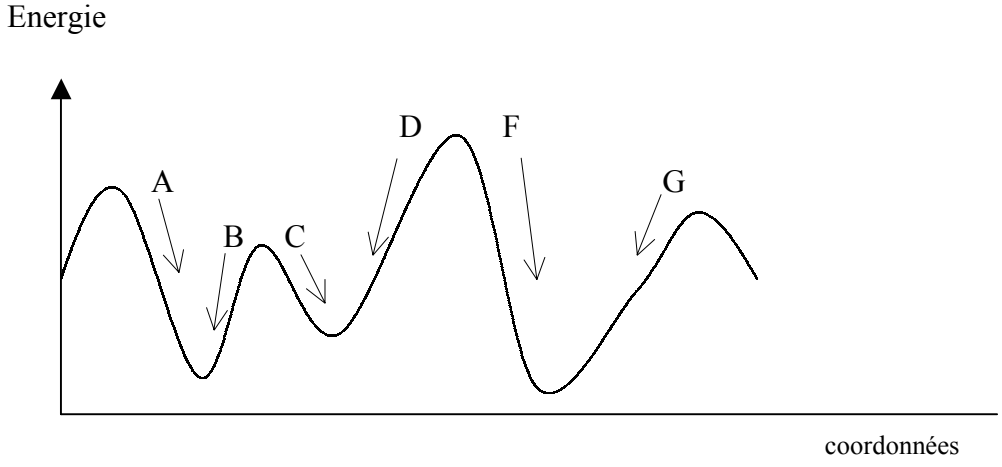


Figure 3 : Représentation d'une surface d'énergie potentielle

Les méthodes d'optimisation qui ont été développées, pour trouver les points pour lesquels le gradient de l'énergie est nul et les dérivées secondes de l'énergie sont positives, sont toutes basées sur le même principe : à chaque étape du calcul le minimum de la fonction est

déterminé suivant une direction donnée prédéfinie. Ainsi, le problème est ramené à une optimisation unidimensionnelle qui ne pose aucune difficulté de calcul. Elles sont adaptées aux différents problèmes et elles diffèrent par leur efficacité, leur convergence et le nombre d'opérations qu'elles nécessitent. Ces techniques peuvent être séparées en deux groupes que nous décrivons brièvement ci-dessous :

a) - Techniques aux dérivées premières

Ces techniques sont simples et font usage uniquement de la pente de la surface de potentiel, en utilisant les dérivées premières obtenues analytiquement ou numériquement[42,43].

**Méthode de la plus grande pente* ('Steepest descent')

Cette méthode, souvent utilisée pour les premiers pas de minimisation, démarre d'un point x_0 qui correspond à l'ensemble des coordonnées de la structure initiale.

Pour approcher du minimum d'énergie dans l'espace à (3N-6) dimensions, les atomes doivent être déplacés dans la direction de la pente descendante de la surface d'énergie potentielle définie dans le champ de forces, c'est-à-dire dans le sens opposé au gradient local de l'énergie. Ce gradient est en fait la dérivée première de la fonction $E(x)$ soit :

$$g_i = \frac{\delta E(x_i)}{\delta x} \qquad \text{au point } x_i \qquad \textbf{(III-16)}$$

La recherche linéaire du scalaire α rendant E minimum dans la direction de cette pente est alors effectuée selon la relation itérative suivante:

$$x_{i+1} = x_i - \alpha g_i \qquad \textbf{(III-17)}$$

Ce processus est répété jusqu'à ce qu'un critère de convergence fixé au préalable soit atteint. Cet algorithme suit donc une direction imposée par les forces inter atomiques dominantes, c'est pourquoi il se révèle très efficace lorsque x_i est très loin du minimum local alors qu'il devient inefficace au voisinage de ce minimum, à cause de l'orthogonalité des directions de recherche. La convergence devient très lente au delà des premiers pas là où la pente est faible et cela à cause du phénomène d'oscillations. Par conséquent, cette méthode qualifiée de descente abrupte est utilisée pour une optimisation préliminaire des géométries, ce qui permet d'éliminer les mauvais contacts de van der Waals et/ou d'ajuster les contacts inter atomiques forts d'une structure initiale.

* Méthode du gradient conjugué ('conjugate gradient')

Cette méthode fondée sur le même principe que la méthode précédente (direction opposée au gradient de l'énergie) prend également en compte les étapes précédentes et ceci afin de déterminer plus finement la direction et le pas. En effet, il s'agit de trouver une direction qui,

dans la mesure du possible, soit conjuguée avec l'ancien gradient et avec les directions précédemment suivies. Pour ce faire, le vecteur représentant la nouvelle direction à suivre en un point donné (h_{i+1}) est calculé en ajoutant au gradient en ce point (g_{i+1}) la direction précédente (h_i) multipliée par une constante γ_i telle que :

$$h_{i+1} = g_{i+1} + \gamma_i h_i \qquad \textbf{(III-18)}$$

avec

$$\gamma_i = \frac{g_{i+1} \cdot g_{i+1}}{g_i \cdot g_i} \qquad \textbf{(III-19)}$$

Le pas est ajusté à chaque cycle pour obtenir la meilleure diminution de l'énergie. Une caractéristique de cette procédure est que le gradient g_{i+1} est orthogonal à tous les gradients précédents $g_0, g_1, g_2 \ldots\ldots g_i$ et la nouvelle direction h_{i+1} est conjuguée à toutes les directions précédentes $h_0, h_1, h_2 \ldots\ldots h_i$. L'intérêt majeur de cet algorithme est d'éviter un comportement oscillatoire autour du minimum et d'accélérer la convergence tout en conservant une bonne efficacité. Cependant, il est plus coûteux en temps de calcul que le précédent et il se révèle moins efficace ou même inutilisable pour des structures qui présentent des mauvais contacts, par exemple, les structures moyennées sur la trajectoire d'une dynamique moléculaire. Dans ce cas, il est préférable d'utiliser la méthode de la plus grande pente pour les premiers pas de la minimisation et d'enchaîner ensuite avec cette méthode des gradients conjugués.

b) – Techniques aux dérivées secondes

Ces techniques sophistiquées font appel à la pente et à la courbure de la surface de potentiel, en utilisant aussi bien les dérivées premières que les dérivées secondes[44].

Ces techniques convergent très rapidement lorsque la surface de potentiel est quadratique : la fonction $E(x)$ est représentée par une approximation quadratique (du deuxième ordre), obtenue par développement en série de Taylor. Les algorithmes les plus employés sont ceux de 'Newton Raphson', nécessitant le calcul de la matrice des dérivées partielles secondes (le Hessien) qui nous permettent de prévoir la courbure locale de la surface d'énergie afin de mieux déterminer le vecteur du déplacement. Cependant, malgré leur excellente convergence, ces algorithmes requièrent un espace mémoire important et ils ne peuvent être appliqués avec succès qu'aux molécules de petite taille. De plus, la méthode peut ne pas converger si le point de départ est loin du minimum, c'est à dire à un endroit où la surface d'énergie est anharmonique.

Comme nous l'avons mentionné ci-dessus, aucune méthode de minimisation ne permet de trouver de façon absolue le minimum global. En effet, partant d'un endroit donné de l'hypersurface (énergie – coordonnées), les algorithmes connus à ce jour ne peuvent conduire

qu'au minimum local le plus proche. Ils ne poussent donc pas le système au-delà des barrières d'énergie, mais descendent plutôt vers le puits le plus proche (***figure 3***). Afin de contourner ce problème, d'autres approches ont été développées afin de mieux explorer l'espace conformationnel accessible aux molécules. Nous décrivons, dans ce qui suit, quelques unes de ces techniques.

4) - Recherche systématique de l'espace conformationnel

Les plus importantes différences structurales qui existent entre les différents conformères d'une molécule donnée, se situent d'abord au niveau des angles dièdres puis au niveau des angles de valence ou des longueurs de liaison. Ceci est dû au fait qu'il faut beaucoup plus d'énergie pour déformer une liaison ou un angle de valence qu'un angle dièdre. Par conséquent, une méthode efficace, pour explorer l'espace conformationnel accessible à une molécule, est de faire varier systématiquement chacun de ses angles dièdres, par petit incrément, tout en gardant les liaisons et les angles de valence fixes[45]. Dans ce cas, le nombre de conformères théoriques est donné par P^i où P est le nombre d'états possibles par torsion et i le nombre total d'angles dièdres contenus dans la molécule. Ainsi, pour une molécule possédant 2 angles de torsion et si chacun des liens subit une rotation par incréments de 30° alors P sera égal à 12 (P = 360° / 30°). Ce qui génère 144 conformères. Dans ces mêmes conditions, le nombre de conformères possible passe à 1728 pour une molécule possédant 3 angles de torsion et à 20736 pour une molécule possédant 4 torsions !

Il est évident que cette méthode de recherche systématique permet, en principe, d'explorer tout l'espace conformationnel accessible à une molécule. Cependant, elle devient vite longue et onéreuse pour des molécules contenant plus de 5 torsions. Pour cette raison, nous avons écarté cette méthode dans l'exploration de l'espace conformationnel, du fait de la taille relativement importante des systèmes moléculaires traités dans le cadre des travaux présentés dans ce manuscrit. Par conséquent, nous avons utilisé les méthodes de dynamique moléculaire que nous présentons dans la suite.

5) - La dynamique moléculaire

a) Définition et principes

La dynamique moléculaire est une méthode d'étude de l'évolution de la configuration spatiale des systèmes moléculaires, au moyen des principes de la mécanique classique Newtonienne. Elle consiste donc à simuler les mouvements moléculaires au cours du temps[46,47]. Ces mouvements correspondent à des oscillations autour d'un minimum énergétique ou au passage d'un minimum à un autre. D'autre part, si l'énergie fournie

virtuellement (par le biais de la température) au système est suffisamment élevée des barrières énergétiques importantes peuvent être franchies, et ainsi le système peut s'échapper des minima locaux.

Afin de mimer au mieux les conditions expérimentales au cours d'une dynamique, on peut exiger que certaines propriétés du système telles que la température T, le nombre de particules N, le volume V, la pression P et l'énergie E restent invariants. Le choix des paramètres constants détermine le type d'ensemble thermodynamique dans lequel la simulation aura lieu. Ainsi les calculs peuvent être effectués avec l'ensemble NVE, ou NPT, ou NVT. Nous avons utilisé dans nos travaux l'ensemble NVE, appelé ensemble *microcanonique*. Ce choix permet d'imposer au système une énergie globalement constante et ceci par le biais de la température désirée qui est généralement 300 K pour l'étude des biomolécules.

Le principe de la dynamique moléculaire consiste, en fait, à déterminer la trajectoire au cours du temps d'un ensemble fini de particules constituant le système, en intégrant numériquement les équations classiques du mouvement. Les trajectoires sont utilisées pour évaluer des propriétés structurales, dynamiques et thermodynamiques du système considéré[48,49].

b) Méthodologie

Le formalisme de la dynamique moléculaire utilise la seconde loi de Newton pour décrire le mouvement d'une molécule en fonction du temps. Dans le cadre de cette loi, la vitesse et le sens du mouvement atomique dépendent des forces qui s'exercent entre les atomes. En pratique, les atomes possèdent une vitesse initiale propre qui est fonction de l'énergie cinétique totale du système.

La molécule est donc perçue comme un système dynamique dont les atomes modifient leurs positions spatiales au cours du temps. Ces changements sont influencés par la résultante des forces qui s'exercent sur eux par les atomes environnants et aussi par leur énergie cinétique qui est reliée à la température de simulation T_0, au moyen de l'équation suivante :

$$\frac{3}{2}(N-6)K_bT_0=\frac{1}{2}\sum_{i=1}^{N}m_iv_i^2 \qquad \textbf{(III-20)}$$

N est le nombre d'atomes définissant le système moléculaire, K_b la constante de Boltzmann, m_i et v_i sont respectivement la masse et la vitesse de chaque atome.

Pour démarrer une dynamique, il faut connaître les positions des atomes et leurs vitesses pour t = 0. Or, si les positions sont connues (choisies initialement) il n'en est pas de même pour les vitesses. Ces dernières sont en fait générées selon une distribution statistique et sont re-échelonnées de manière à ramener le système à l'équilibre ou à retrouver des vitesses cohérentes.

L'équation de Newton suivante est donc vérifiée pour chaque atome :

$$F_i(t) = m_i a_i(t) = \frac{m_i dv_i(t)}{dt} \quad \textbf{(III-21)}$$

$a_i(t), v_i(t)\ et\ F_i(t)$ sont respectivement l'accélération, la vitesse et l'ensemble des forces exercées sur cet atome au temps t.

Cependant, les trajectoires ne sont pas déterminées par résolution de l'équation (III-21) et ceci en raison d'un défaut de solution analytique. Une approche différente est utilisée et qui consiste à calculer les grandeurs décrites ci-dessous:

*) la force F_i qui s'exerce sur un atome *i* se trouvant à l'instant t en position $r_i(t)$. Elle est déterminée par différenciation de la fonction énergie potentielle au voisinage de la position $r_i(t)$.

$$F_i(t) = \frac{\delta E r_i(t)}{\delta r_i(t)} \quad \textbf{(III-22)}$$

*) l'accélération de chaque atome à partir de cette force s'exerçant sur lui et de sa masse atomique (éq.III-21)

*) la vitesse de chaque atome à partir de son accélération

$$a_i(t) = \frac{dv_i(t)}{dt} \quad \textbf{(III-23)}$$

*) la position de chaque atome à partir de sa vitesse

$$v_i(t) = \frac{dr_i(t)}{dt} \quad \textbf{(III-24)}$$

L'intégration de ces équations de mouvements se fait en subdivisant la trajectoire en une série d'états discrets séparés par des intervalles de temps très courts dont la longueur définit le pas d'intégration Δt. Différents algorithmes ont été proposés dont le plus connu et le plus largement employé est celui de Verlet[50]. Dans cet algorithme, les coordonnées sont développées en série de Taylor au troisième ordre, aux temps $(t + \Delta t)$ *et* $(t - \Delta t)$:

$$r_i(t \pm \Delta t) = r_i(t) \pm \frac{dr_i(t)}{dt}\Delta t + \frac{d^2 r_i(t)}{dt^2}\frac{(\Delta t)^2}{2!} \pm \frac{d^3 r_i(t)}{dt^3}\frac{(\Delta t)^3}{3!} \quad \textbf{(III-25)}$$

La somme de ces deux séries fournit la position de l'atome *i* après un temps Δt.

$$r_i(t + \Delta t) = 2.r_i(t) - r_i(t - \Delta t) + \frac{d^2 r_i(t)}{dt^2}(\Delta t)^2 \quad \textbf{(III-26)}$$

Alors que leur différence permet d'obtenir la vitesse de l'atome *i* à l'instant *t*

$$v_i(t) = \frac{dr_i(t)}{dt} = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} \quad \textbf{(III-27)}$$

Ces étapes répétées à chaque pas d'intégration aboutissent à l'identification de la trajectoire. Afin d'assurer la constance des forces inter atomiques au cours du pas d'intégration (Δt), celui-ci doit être 10 à 20 fois, au moins, inférieur à la période du mouvement dont la fréquence est la plus élevée. Des pas trop grands causeraient un éclatement ("blow up") du système moléculaire. En ce qui concerne les composés étudiés dans ce travail, le mouvement dont la fréquence est la plus importante, est l'élongation des liaisons C – H (3600 cm^{-1}), ce qui correspond à une période de 10^{-14} s ou 10 fs (10 femtoseconde) . Nous avons donc, dans tous nos calculs de dynamique moléculaire, utilisé un pas d'intégration de 1 fs. Cette contrainte restreint évidemment la durée de la simulation à quelques centaines de picosecondes (souvent 100 ps) à cause du temps prohibitif de calcul. Des intégrations successives en fonction du temps conduisent à une trajectoire de l'atome en fonction du temps sous la forme d'une série de positions et de vitesses dans l'espace. Plusieurs mouvements sont alors effectués durant la période de simulation. Si la période d'évolution est plus courte (fraction de ps), la dynamique moléculaire présente peu de cohérence dans le déplacement des atomes. De plus, les mouvements moléculaires sont fréquemment interrompus par des collisions avec les groupements voisins, nuisant ainsi à la trajectoire. Enfin, les calculs nécessitent une bonne puissance de calcul (mémoire et rapidité) et des performances graphiques adéquates.

Afin d'identifier les mouvements préférentiels, on peut effectuer la simulation en imposant des contraintes expérimentales au système moléculaire. Généralement, avant de commencer une dynamique sur un système moléculaire, il est conseillé de relaxer d'abord le système afin d'éliminer les points chauds ("hot spots") de haute énergie. Pour finir, il faut faire remarquer qu'il n'est pas possible de savoir de manière absolue si cette technique de dynamique moléculaire explore l'espace conformationnel de façon représentative. Cependant, cet inconvénient peut être partiellement surmonté de deux façons distinctes:

- Réduire l'espace conformationnel, à explorer en utilisant des informations expérimentales (sous formes de contraintes) comme, par exemple, les données moléculaires issues de la RMN. Les contraintes de distances interatomiques et d'angles dièdres sont injectées dans le calcul sous forme de pseudo-forces et interviennent au niveau de l'expression du champ de force.

- Elargir l'espace conformationnel à explorer en simulant des trajectoires de dynamique moléculaire à haute température (1000 K). En effet, une simulation de l'ordre de 100 ps à 300 K ne suffit pas à conduire à des changements conformationnels importants, Par contre à température élevée, la molécule est 'secouée', elle possède donc une énergie cinétique

suffisante pour franchir les barrières d'énergie, ce qui lui permet d'osciller entre différentes conformations, puis en abaissant la température à celle désirée, la molécule adoptera alors une conformation 'naturelle'. Cette technique connue sous le nom de recuit simulé est présentée ci-dessous.

c) le recuit simulé

Cette méthode tire son nom de l'analogie avec le procédé métallurgique de recuit. Le 'protocole' du recuit simulé consiste à effectuer d'abord une dynamique moléculaire à haute température où le système possède suffisamment d'énergie thermique, pour franchir les barrières d'énergie séparant des minima locaux, et ensuite, à refroidir graduellement le système et à l'équilibrer à plus basse température (d'où le nom de recuit), afin de lui permettre de s'échapper des minima locaux de hautes énergies[51,52]. Les structures résultantes sont ensuite minimisées. Cette approche n'a évidemment pas de sens physique pour les systèmes moléculaires, mais permet de fixer un ordre de grandeur des barrières d'énergie que l'on s'autorise à franchir. On rend ainsi possible le passage entre les différents minima de l'hypersurface d'énergie potentielle (***Figure 3***). Cette technique permet d'effectuer une bonne exploration de l'espace conformationnel et permet également d'obtenir des conformères qui sont généralement mieux relaxés, que ceux issus des dynamiques réalisées à 300 K[53]. La répétition de cette procédure (chauffage/refroidissement), en partant d'une nouvelle structure, permet d'identifier les conformations énergétiquement accessibles à une température donnée. Le recuit simulé (simulated annealing) consiste, en fait, à effectuer des explorations aléatoires d'un ensemble de configurations à des températures de plus en plus basses. Sur cet organigramme de principe du recuit simulé, il faut noter que la mise en œuvre de la méthode fait intervenir six éléments principaux : une configuration de départ, une température initiale (1000 K) , une transformation (dynamique) permettant les transitions entre les configurations, la liste des températures successives, le nombre d'itérations de la transformation à chaque température et un critère d'arrêt. Nous avons utilisé cette technique pour effectuer divers échantillonnages conformationnels dans le cadre de nos travaux.

d) Autres méthodes

D'autres méthodes permettant d'échantillonner l'espace conformationnel, ont été développées et nous pouvons citer :

- La méthode de Monte carlo[54,55], qui est basée sur la génération aléatoire d'un ensemble de conformations et n'introduit pas la dimension du temps dans l'étude des molécules. Le

couplage de cette technique avec l'algorithme de Metropolis[54] rend cette méthode plus efficace en échantillonnant préférentiellement les conformations de plus basse énergie.

- Méthodes 'hybrides' couplant dynamique moléculaire et Monte carlo[56,57]
- Méthodes de 'Distance Geometry' utilisées lorsque des données structurales expérimentales sont disponibles[58].

6) - Les logiciels de la mécanique moléculaire

Le développement de plusieurs programmes de simulation de molécules et de macromolécules, avec différents champs de forces, montre qu'aucun d'eux n'est totalement satisfaisant, d'où leur évolution continue. Ces logiciels de modélisation sont d'utilisation assez générale et c'est à l'utilisateur de définir les conditions de leur mise en œuvre, dans le cadre de sa stratégie d'étude. Certains logiciels sont destinés à l'analyse de structures ; ils seront donc utilisés pour analyser les conformations jugées pertinentes sur une station graphique. Il est également possible, à l'aide de lunettes spéciales, d'avoir une vue stéréographique (3D), ce qui permet de bien percevoir les poches ou les repliements d'une protéine et également d'analyser l'ancrage du ligand dans le site du récepteur.

Parmi les principaux logiciels dont certains existent en plusieurs versions successives pouvant tourner aussi sur PC, on peut citer : MACROMODEL[59], HYPERCHEM[60], MAD[61], SYBYL [62], CHEM3D[63], INSIGHT/DISCOVER[64] ; pour une liste plus complète nous suggérons au lecteur de consulter le site du NIH[65].

Nos études de modélisation ont été entièrement réalisées avec différents logiciels de la société Accelerys[64] implantés sur des stations de travail Silicon Graphics. Le premier logiciel utilisé, INSIGHT II (Version 95.0), est un programme de modélisation moléculaire graphique qui ne tourne sous UNIX et qui permet la construction et la manipulation de différents systèmes moléculaires dans un espace à 3 dimensions. INSIGHT II est interfacé avec DISCOVER un logiciel de mécanique et de dynamique moléculaire dont l'objet est l'étude des propriétés moléculaires. Il est également interfacé avec d'autres programmes. Citons, par exemple, ceux utilisés au cours des études présentées dans ce manuscrit : ANALYSIS qui permet l'analyse conformationnelle, DELPHI qui permet d'étudier les comportements électrostatiques des molécules, DECIPHER qui permet le calcul des propriétés structurales et HOMOLOGY qui permet, de façon schématique, la construction de la structure tridimensionnelle d'une protéine connaissant seulement sa séquence en acides aminés et la structure atomique spatiale d'une protéine de référence présentant une bonne homologie.

7) - Conclusion

La mécanique moléculaire est une méthode facile à mettre en œuvre, elle ne demande pas des temps de calcul prohibitifs et peut être effectuée sur des ordinateurs personnels (PC). Cependant, elle fait apparaître quelques distorsions au niveau des résultats obtenus en comparaison à ceux obtenus par la mécanique quantique. Compte tenu de la courte durée des calculs qu'elle nécessite, la mécanique moléculaire convient bien pour l'étude de gros systèmes, comme les protéines ou les enzymes. Mais son utilisation reste, néanmoins limitée à des problèmes essentiellement structuraux. Elle permet également d'effectuer une analyse conformationnelle préliminaire sur des composés présentant de fortes flexibilités. Nous l'emploierons dans nos travaux, en particulier, pour distinguer rapidement des conformères. En outre, la mécanique moléculaire s'adapte bien à l'étude du docking moléculaire qui est l'étude des interactions intermoléculaires, puisque les interactions entre un médicament, **agissant en tant qu'agoniste ou antagoniste,** et sa cible sont de nature non covalentes. En revanche, son caractère trop empirique, fait d'elle une méthode inadaptée à l'étude des réactions chimiques et imprécise pour le calcul des grandeurs moléculaires fines dans l'absolu.

IV - EVOLUTION DES METHODES DE LA MODELISATION

Les méthodes de la chimie quantique, qui permettent d'étudier des propriétés moléculaires dépendantes de la structure électronique du système, conviennent parfaitement à l'étude des réactions chimiques qui sont le fait de réarrangements électroniques importants. Cependant, les performances des ordinateurs actuels ne permettent pas le traitement de systèmes comportant plusieurs milliers d'atomes comme les macromolécules d'intérêt biochimique que sont les enzymes par exemple. D'autre part, les méthodes de la mécanique moléculaire ne peuvent pas rendre compte des importantes réactions chimiques qui ont lieu au sein de ces enzymes. Afin de remédier aux limites de chacune d'elles, de nouvelles méthodes combinant les deux mécaniques et appelées méthodes hybrides ont été élaborées dans le but de traiter des systèmes sièges de réactions chimiques telles que les réactions enzymatiques[66-67]. Ces méthodes hybrides MQ/MM (ou mixtes) consistent à combiner le traitement quantique des atomes impliqués dans la réaction chimique au traitement par la mécanique moléculaire du reste du système. Le lecteur intéressé, peut consulter les articles de revues publiés ces dernières années[68,69].

BIBLIOGRAPHIE

1- Everett, Rev. Electron. Digital Comput., Février **1952**, 70.

2- Newman, W.N. et Sproull, R.E., " Principles of Interactive Computer Graphics", 2nd Ed., Mc Graw-Hill, **1979**.

3- Fox, J.L., CEN News, 27, **juillet 1980**.

4- Langridge, R., Ferrin,T.E., Kuntz, I.D. et Connolly, M.L., Science, 211, **1981**, p:661

5- Schrödinger, E., Ann.Physik, 79, 361, **1926**.

6- Born, M. et Oppenheiner, J.R, *Ann. Physik*, 84, 457, **1927**.

7- Hartree, D.R., Proc. Cambridge Phil. Soc., 24, 89, **1928**.

8- Fock, V., *Z. Physik*, 61, 126, ***1930***.

9- Slater, J.C. *Phys. Rev.*, 35, 509, **1930** ; ibid, 34, 1293, **1929**.

10- Rivail, J.L. Eléments de chimie quantique à l'usage des chimistes, deuxième édition.Inter Editions et CNRS Edition **1994**.pp : 299-301.

11- Roothaan, C.C.J. *Rev. Mod. Phys.*, 23, 69 (1951).

12- Toumi, A., thèse de Magister, « Etude théorique des mécanismes de réaction et des propriétés physico-chimiques de dérivés pyroniques d'intérêt biologique ». USTHB **1998**.

13- Rivail, J.L. Eléments de chimie quantique à l'usage des chimistes, deuxième édition.Inter Editions et CNRS Edition **1994**.p: 313.

14- Rivail, J.L., Eléments de chimie quantique à l'usage des chimistes, deuxième édition.Inter Editions et CNRS Edition **1994**.p: 213-230.

15- Huckel, E., *Z. Physik*, 70, 204, **1931**, 72, 310, **1932**, 76, 628, **1932**.

16- Pople, J.A., Santry, D.P. et Segal, G.A., *J. Chem. Phys.*,43, 5129, **1965**.

17- Pople, J.A. et Segal, G.A., *J. Chem. Phys.,* 43, S136-S149, **1965**, 44, 3298, **1966**.

18- Pople, J.A., Beveridge, D.L. et Dobosh, P. A. *J. Chem. Phys.,* 47, 2026, **1967**.

19- Dewar, M.J.S. et Thiel, W. *J. Amer. Chem. Soc.*, 99, 4899, **1977**.

20- Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. et Stewart, J.J.P. *J. Amer. Chem. Soc.,* 107, 3902,**1985**.

21- Stewart, J.J.P., *J. Comp. Chem.*, 10, 209 ; 221,**1989**.

22- Bernal-Uruchurta, M. I. et Ruiz-Lopez, M. F., Chem. Phys. Let., 330,118,**2000**.

23- Hohenberg, P. et Kohn, W., *Phys. Rev.*, B136, 864, **1964**.

24- Kohn, W. et Sham, L.J., *Phys. Rev.*, A140, 1133, **1965**.

25- Andrews, D. H., *Phys. Rev*.,36, 544, **1930**.

26- Hill, T. L. *J. Chem. Phys*., 14, 465, **1946**.

27- (**a**) Caillet, J. et Claverie, P., *Biopolymers,* 13, 601, **1974**. (**b**) Caillet, J. et Claverie, P., *Acta Crysta.* A31, 448, **1975**. (**c**) Claverie, P., in :B. Pullman (Ed, Intermolecular Interactions : From Diatomics to Biopolymers, John Wiley, New York, **1978**.

28- Allinger, N.L. *J. Am. Chem. Soc.* 99, 8127-8134, **1977**.

29- Allinger, N. *et al. J. Comput. Chem.* 11, 848-867, **1990**.

30- Weiner, P.J., Kollman P.A, Case D.A., Singh U.C., Ghio C., Alagona G., Profeta Jr S. et Weiner P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* 106, p:765-784, **1984**.

31- Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M. Jr., Ferguson D.M., SpellmeyerD.C., Fox T., Caldwell J.W. et Kollman P.A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 117, p:5179-5197, **1995**.

32- GROningen MOlecular Simulation (GROMOS) (**1987**) Biomos, Biomolecular software by Laboratory of physical chemistry University of Groningen Nijenborgh 16, 9747 AG Groningen Hollande Roven Press, New York.

33- Gunsteren W.F.V., Billeter S.R., Eising A.A., Hünenberger P.H., Krüger P.K.H.C., Mark A.E., Scott W.P.R. et Tironi I.G. Biomolecular simulation: The GROMOS96 Manual and User Guide. Hochschuleverlag AG.**1996**.

34- Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S. et Karplus M., CHARMM : a program for macromolecular energy, minimization and dynamics calculations, *Journal of computational Chemistry*, Vol.4, p 187-217, **1983**.

35- Hagler A.T., Dauber P., Lifson S., *J. Am. Chem. Soc*. 101, p: 5131, **1978**.

36- Hagler A.T., Lifson S., Dauber P., J. Am. Chem. Soc. 101, p: 5122, **1979**.

37- Dinur, U., et Hagler, A. T. New approaches to empirical force fields. In: *Reviews in Computational Chemistry,* Lipkowitz, K.B., and Boyd, D.B., Eds., VCH Publishers, New York, **1991**, Vol. 2, Chapter 4, pp. 99-164.

38- Scheraga et al. *PNAS* 84, 6611, **1987**.

39- Scheraga et al., *J. Biomol. Struct. & Dynamics*, 7, 391, **1989**.

40- Lennard-Jones, J.E., Proc. Roy. Soc., London, **A106**, p:463, **1924**.

41- Bref, C., Amara, P. et Field, J. *Developpement d'un champ de force de mecanique moleculaire avec la polarisation,* **1999.**

42- Courant, R., *Bull. Amer. Math. Soc.* , 49, 1, **1943.**

43- Hestenes, M.R. et Stiefel, *J. Res. N.B.S.,* 49, 409, **1952**.

44- Householder, A.S., Principles of Numerical Analysis, New York: McGraw Hill, **1953** .

45- Lipton, M. et Still, W.C., *J. Comput. Chem.*, 9,343-355, **1988**.

46- Van Gusteren, W.F. et Berendson, H.J.C. « Computer simulation of molecular dynamics : Methodology, applications, and perspectives in chemistry . » Angrew.chem, Int.Ed.Engel., 29, 992-1023, **1990**.

47- Karplus, M. et Petsko, G.A. « Molecular dynamics simulations in biology. » *Nature*, 347,631-639, **1990**.

48- Brooks, C.L., Karplus, M. et Montgomery Pettitt, B., A theoretical perspective of dynamics, structure, and thermodynamics., New York: Wiley, **1988.**

49- McCammon, A. et Harvey, J.A., Dynamics of proteins and nucleic acids.: Cambridge University Press, **1987**.

50- Verlet L., Computer experiments on classical fluids. I.Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev*. 159 :98-103, **1967**.

51- Beredsen, H.J.C., Postma, J.P.M., van Grunstern, W.F, DiNola, A. et Haak, J.R. « Molecular dynamics with coupling to an external bath. » *J. Chem. phys.,* 81,3683-3690, **1984**.

52- Kirkpatrick, S., Gebatt, Jr.C.D. et Vecchi, M.P,. « Optimization by simulated annealing. » *Science*, 220, 6719-680, **1983**.

53- Mackay, D.H.J., Cross, A.J. et Hagler, A.T., « The role of energy minimization in simulation strategies of biomolecular systems. « In Prediction of protein structure and the principles of protein conformation. G.Fasman, Ed. ;Plenum Press, N.Y., 317-358, **1989**.

54- Metropolis, N., Rosenbluth, A.W.,Rosenbluth, M.N., Teller,A.H. et Teller, E.J. « Equation of state calculations by fast computing machine. » *J. Chem. phys.*, 21,1087-1092, **1953**.

55- Zhegin, L. et sheraga, H.A. « Monte Carlo-minimisation approch to the multiple minima problem in protein folding . » *Proc. Nath. Acad. Sci.* USA, 84, 6611-6615, **1987**.

56- Morley, S.D., Jackson, D.E., Saunders, M.R. et Vinter, J.G. DMC : A multifunctional hybrid dynamics/ Monte carlo simulation algorithm for the evaluation of conformational space *J. Comp. Chem.* 13(6) : 693-703, **1992**.

57- Clamp, M.E., Baker, P.G., Stirling, C.J. et Brass, A., Hybrid Monte Carlo : An efficient algorithme for condensed matter simulation. *J. Comp. Chem.* 15(8) :838-846, **1994**.

58- Crippen, G.M., Distance geometry approach to rationalizing binding data. *J. Med. Chem.* 22(8) :988-997, **1979**.

59- Still, C., Columbia University, NY (MMOD version 5.5, **1996**)

60- Hypercube, Inc. Ontario, Canada.

61- MAD, Molecular Advanced Design. Aquitaine Systems Tour Eff2, Place de la coupole Cedex 45 92078 Paris la défense France.

62- SYBYL of TRIPOS Inc. 1699 South Hanley Road St.Louis, MO 63144 USA.

63- CS Chem Office, Ultra. For microsoft Windows, Cambridge Scientific Computing Inc., 875 Massachusetts Avenue, Suite 61, Cambridge, MA 02139, USA.

64- Accelrys/MSI. Road San Diego, CA 92121 USA.

65- http://www.nih.gov

66- Antes, I. et Thiel, W., *ACS Symp. Ser.*, 712 (Combined Quantum Mechanical and Molecular Mechanical Methods), p. 50, **1998**.

67- Reuter, N., Thèse, Méthodes Hybrides Mécanique Quantique/Mécanique Moléculaire: Développements Méthodologiques et Applications, Université Henri Poincaré-Nancy I, **1999**.

68- Gao, Reviews in Computational Chemistry, Vol. 7, Chap. 3, VCH Publishers, Inc. New York , **1996**.

69- Ruiz-Lopez, M.F. et Rivail, J. L., Combined Mechanics And Molecular Mechanics Approaches To Chemical And Biochemical Reactivity. In Encyclopedia of computational Chemistry, Vol. 1, pp 437-448, Ed. P.v.R.Schleyer, Wiley & sons **1998**.

CHAPITRE III

ETUDE DE RELATIONS QUANTITATIVES STRUCTURE-ACTIVITE D'UNE SERIE D'ANTAGONISTES NON PEPTIDIQUES DE LA CHOLECYSTOKININE

I – INTRODUCTION

Nous avons noté dans le chapitre I, que la cholécystokinine (**CCK**) est impliquée dans de nombreuses actions biologiques. Par conséquent, une anomalie dans le métabolisme de ce neuropeptide endogène pourrait entraîner de nombreux troubles pathologiques. Dans le but d'apporter des remèdes, plusieurs antagonistes des récepteurs de la CCK, appartenant à des classes chimiques différentes, ont été développés (*Confer* **chapitre I**). Du fait de leur bonne stabilité et de leur meilleure biodisponibilité par rapport aux peptides, le développement d'antagonistes non peptidiques a suscité un grand intérêt dans le domaine thérapeutique. De nombreux dérivés de benzodiazépines ont donc été synthétisés et ont été largement étudiés[1-8]. Malgré l'intérêt croissant pour ces antagonistes, que reflète une importante bibliographie, leur mécanisme d'action reste encore à élucider.

Ce présent travail[9] consiste à déterminer les éléments structuraux responsables des activités antagonistes sur les deux types de récepteurs de la CCK, d'une série de composés de type benzodiazépine issus de différentes sources[1,2,10-12]. Cette étude de relations quantitatives entre la structure et l'activité (QSAR), a été entreprise au moyen d'un grand nombre de quantités moléculaires (globales) et/ou locales calculées à l'aide des techniques de la chimie théorique. Comme ces descripteurs électroniques sont très sensibles au changement conformationnel de la molécule, notre attention a été de les calculer pour chaque composé dans sa conformation la plus stable et non dans celle qui est 'supposée' biologiquement active et qui est une conformation toujours difficile à déterminer de façon précise[8]. Les objectifs principaux de ce travail sont :

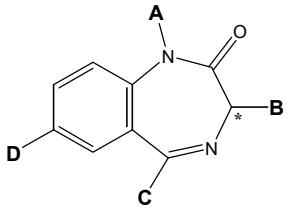
- de rationaliser la sélection des substituants les plus liés à la sélectivité du composé vis à vis de l'un ou l'autre des récepteurs de la CCK.
- d'obtenir, si possible, une prédiction de l'activité antagoniste pour un composé non encore synthétisé sur la base de la (des) meilleure(s) équation(s).
- de dégager une information qualitative sur le mécanisme d'action et les interactions entre le ligand et sa cible biologique.

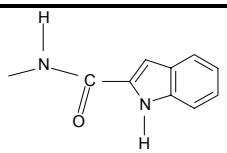
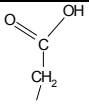
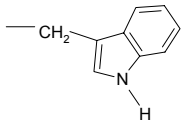
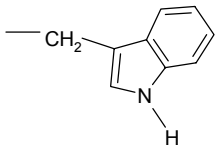
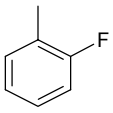
II – RECHERCHES DE RELATIONS QUANTITATIVES ENTRE LA STRUCTURE ET L'ACTIVITE

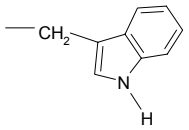
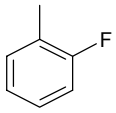
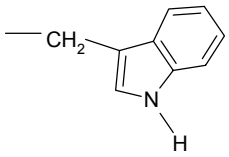
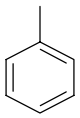
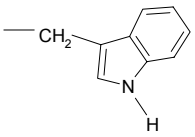
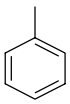
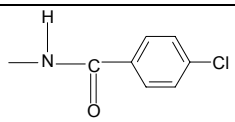
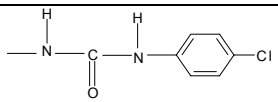
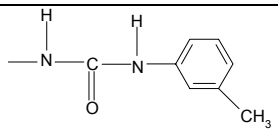
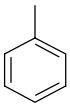
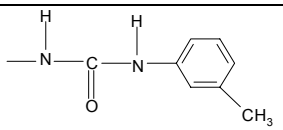
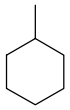
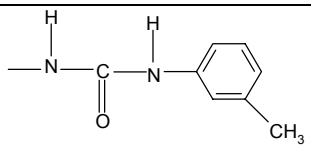
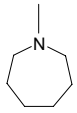
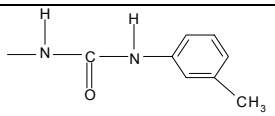
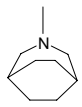
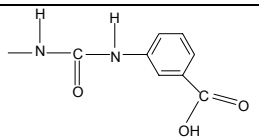
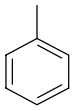
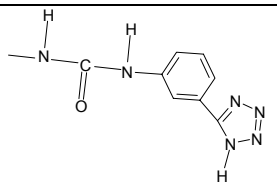
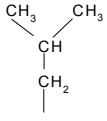
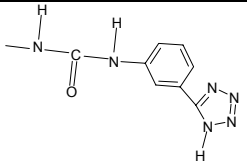
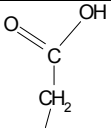
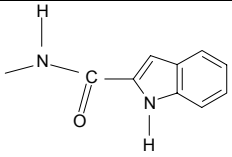
1 - Présentation des composés étudiés

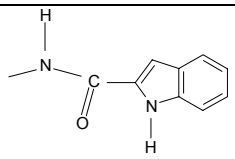
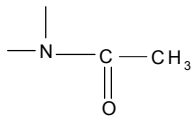
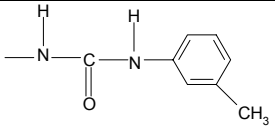
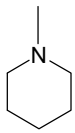
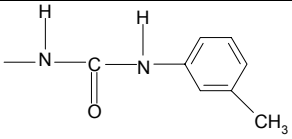
Nous avons sélectionné, à partir de différentes sources de la littérature, une série de composés dérivés des (1,4)-benzodiazépines présentant une activité antagoniste sur les deux types de récepteurs de la CCK. Ces composés, présentés avec leur formule générale dans le **tableau 1**, ont été choisis de manière à disposer d'une base de données caractérisée par une diversité structurale et biologique adéquate pour une analyse statistique. Les affinités de liaison mesurées *in vitro* sont estimées en termes d'IC_{50} qui sont des concentrations maximales d'inhibition à 50 % de la liaison des radio ligands [^{125}I]CCK-8 ou [^{125}I]CCK-33 dans les tissus pancréatiques (pour CCK_1) et dans le cortex cérébral (pour CCK_2) du rat. Comme ces mesures d'affinités, pour tous les composés, n'ont pas été effectuées dans les mêmes conditions expérimentales, nous avons donné les affinités relatives (notées A_R) évaluées par rapport à la même molécule de référence. Ces valeurs sont exprimées au moyen du cologarithme des valeurs relatives et sont reportées dans le **tableau 1** ; ainsi, une grande valeur positive de $-\log(A_R)$ correspond à un composé très actif.

Tableau 1: Formules et valeurs du cologarithme des activités antagonistes relatives des molécules étudiées. Le carbone asymétrique est marqué par un astérisque.



Composé	A	B	C	D	Stéréo	$-\log A_R$ (CCK_1)	$-\log A_R$ (CCK_2)
B1	CH_3			H	**S**	3.544	-2.088
					R	1.630	-3.699
B2				H	**R**	0.970	-3.699
B3	CH_3			H	**R**	1.015	n.d.

B4	H			H	**R**	0.748	n.d.
					S	-0.548	n.d.
B5	CH_3			Cl	**R**	0.301	n.d.
B6	H			Cl	**R**	-0.084	n.d.
					S	-1.057	n.d.
B7	CH_3			H	**R**	1.984	-3.740
					S	2.856	-3.161
B8	CH_3			H	**S**	1.032	-2.312
					R	-0.594	-0.439
B9	CH_3			H	**S**	1.970	-1.878
					R	0.000	0.000
B10	CH_3			H	**R**	-0.387	1.482
B11	CH_3			H	S	1.969	-1.229
					R	-0.610	0.799
B12	CH_3			H	**S**	2.053	-0.409
					R	-0.338	1.929
B13	CH_3			H	**R**	-0.324	0.104
B14	CH_3			H	**S**	2.077	-1.469
					R	0.105	0.920
B15				H	**S**	0.288	-0.477
					R	-0.289	1.777
B16				H	**RS**	3.300	n.d.

B17	H			H	**RS**	3.124	-2.176
B18	H			H	**RS**	-1.154	n.d.
B19	CH_3			H	**RS**	1.636	0.174
B20	CH_3			H	**RS**	0.185	-1.207

n.d.: non disponible

Il faut noter que cette famille de benzodiazépines, constituée de racémiques et/ou de couples d'énantiomères purs, présente une large gamme d'activité. Le propos de notre étude QSAR est de relier l'affinité mesurée aux descripteurs structuraux : géométriques et physico-chimiques.

2) - Méthodologie utilisée

Il est bien connu que les méthodes de la chimie quantique permettent de décrire de façon détaillée et avec une grande précision les effets électroniques d'une molécule. Ces méthodes peuvent être, par conséquent, appliquées à l'étude de relations quantitatives structure-activité, en déterminant des descripteurs électroniques directement calculés à partir de la fonction d'onde qui décrit le système moléculaire. Ces indices locaux ou globaux sont, soit des 'observables' moléculaires, soit des quantités théoriques qui caractérisent les différents aspects de la structure électronique ou de la réactivité de la molécule. Une grande variété de tels indices a été déterminée et a servi dans plusieurs études de relations structure-activité[13-15]. Ces travaux ont montré l'utilité de tels indices dans les études QSAR.

a) – Modélisation des composés étudiés

Les structures moléculaires des composés considérés sont construites à partir d'une librairie de données disponible dans le logiciel de modélisation moléculaire Insight II[16]. Ces structures préliminaires, sont ensuite soumises à un premier cycle de minimisation d'énergie composé successivement de 200 itérations par la méthode de la plus grande pente puis du nombre d'itérations nécessaire jusqu'à la convergence par la méthode du gradient conjugué :

le critère de convergence utilisé sur le gradient de l'énergie (RMSEG : Root Mean Square Energy Gradient) est inférieur à 10^{-3} kcal.mol^{-1}.Å^{-1}. La conformation obtenue pour chaque composé sera ensuite utilisée comme point de départ pour rechercher ses conformations énergétiquement accessibles. Cette recherche a été réalisée au moyen de la dynamique moléculaire à haute température (1000 K) avec la technique du recuit simulé[17]. La simulation a commencé par une thermalisation de 1 ps (picoseconde =10^{-12}s), comportant 5000 étapes d'échauffement, suivie par une période de 10 ps à 5000 étapes de dynamique à température constante de 1000 K. Cette dynamique est poursuivie en refroidissant progressivement jusqu'à 300 K, avec un pas de 100 K. La structure recueillie à 300 K est ensuite minimisée en appliquant les mêmes algorithmes, et dans les mêmes conditions, que précédemment. Cette procédure est reconduite 99 fois à partir de la structure obtenue lors de l'étape précédente ; ce qui nous a permis de générer un échantillon de 100 structures conformationnelles. Une technique de superposition moléculaire, disponible dans Insight II et basée sur le calcul de l'écart quadratique moyen entre deux conformations de l'échantillon, nous a permis de rassembler en familles structurales les conformations les plus stables et dont l'écart énergétique est au maximum de 4 kcal/mol. Le conformère le plus stable de chaque famille structurale est ensuite soumis à une optimisation complète de la géométrie dans le cadre semi-empirique. L'application du même traitement à deux énantiomères d'une même molécule montre que la surface d'énergie potentielle obtenue pour l'énantiomère **R** est l'image de celle obtenue pour le **S** ; ce qui confirme la validité de notre approche. Le plus stable des conformères obtenus est ensuite sélectionné pour le calcul des descripteurs électroniques et stériques nécessaires à l'étude QSAR. L'avantage d'utiliser la mécanique moléculaire comme préliminaire ou « pré-optimiseur » aux calculs semi-empiriques réside dans la vitesse de calcul et dans l'accès à la dynamique des molécules par le 'recuit simulé', ce qui contribue généralement à éviter les faux minima de conformation risquant de piéger la molécule.

L'optimisation complète de la géométrie a été effectuée au moyen du logiciel GEOMOS[18], qui offre le choix parmi les méthodes semi-empiriques existant au moment où ce travail a commencé. En principe, n'importe quelle méthode semi-empirique est un outil de choix chaque fois qu'il s'agit d'étudier l'évolution d'une propriété physique ou chimique à l'intérieur d'une famille de composés. Néanmoins, il est connu d'après l'expérience des quantochimistes qu'elles présentent toutes des avantages et des inconvénients (*Confer* Chapitre II, paragraphe II). Ainsi, nous avons utilisé dans cette étude, la méthode PM3[19] qui est la variante MNDO[20] la mieux paramétrée (quand cette étude a commencé), notamment pour la description de la liaison hydrogène[21,22]. Dans une première étape, le logiciel

GEOMOS recherche la géométrie correspondant au minimum de l'énergie et fournit la fonction d'onde de la molécule dans cette géométrie. Cette fonction d'onde est ensuite utilisée dans des programmes auxiliaires pour le calcul de certains descripteurs stériques et électroniques[23]. Il est important de noter que nos calculs sont effectués sur des molécules isolées. Comme les conformères les plus stables dans le vide ne sont pas forcément les plus stables dans l'eau ou dans le milieu biologique, nous avons entrepris dans un autre travail[24] l'étude de l'effet de solvant sur la stabilité relative des conformères en utilisant le modèle du champ de réaction auto-cohérent (SCRF)[25] et qui est implanté dans le logiciel GEOMOS.

b) – Calcul des descripteurs

Le problème majeur lors de l'établissement d'une relation structure-activité reste le choix, parmi les nombreuses possibilités offertes par la chimie quantique, d'une série limitée des meilleurs descripteurs pour une activité donnée et d'éviter l'utilisation simultanée de variables qui apparaîtraient trop étroitement corrélées entre elles. Comme le mécanisme d'action de ces antagonistes sur les récepteurs est inconnu, le choix des descripteurs pertinents est donc très délicat, voire difficile. Nous avons, par conséquent, calculé un large ensemble de descripteurs globaux et locaux. D'autres paramètres moléculaires ont été également déterminés à l'aide du logiciel ChemOffice[26]. Ces derniers sont soit physico-chimiques soit reliés à la forme moléculaire. Tous les descripteurs calculés dans cette étude sont présentés avec leur définition et leur abréviation dans le **tableau 2**.

Il est important de faire remarquer que tous les composés considérés ici sont de configuration R, S, ou RS (racémiques) et aucun des descripteurs, décrit dans le **tableau 2,** ne reflète la chiralité du composé. Par conséquent, afin de tenir compte de cet effet de configuration, nous avons défini un autre paramètre noté STRO, indicateur de la chiralité.

L'observation de la base de données considérée ici (**Tableau 1**), révèle que statistiquement les molécules présentant les meilleures affinités sont S, puis racémiques pour le récepteur CCK_1, et de configuration R, puis racémiques pour le récepteur CCK_2. Cependant, on devrait noter que cette observation est beaucoup plus évidente dans le cas du récepteur CCK_2. Par conséquent, nous attribuons à cet indicateur de la chiralité les valeurs 1, -1 et 0, respectivement, pour les configurations S, R et RS, dans le cas du récepteur CCK_1. Nous avons inversé ces valeurs pour les configurations S et R, dans le cas de l'autre type de récepteur.

Tableau 2: Symbole et définition des descripteurs quantiques et stériques calculés. Ceux dont l'unité n'est pas indiquée sont donnés en unités atomiques (u.a).

Symbole	Définition
HEAT	Enthalpie de formation de la molécule ($kcal.mol^{-1}$)
CHA	Somme des valeurs absolues des charges atomiques de la molécule
DIPO	Moment dipolaire de la molécule (Debye)
HOMO	Energie de la plus haute orbitale moléculaire occupée
LUMO	Energie de la plus basse orbitale moléculaire inoccupée
HMLM	Gap d'énergie entre la HOMO et la LUMO
VDW	Volume de van der Waals de la molécule ($Å^3$) [27]
FORM	Facteur de forme défini par bc/a^2 où a, b et c sont les moments principaux d'inertie obtenus en remplaçant les masses atomiques par les numéros atomiques (ce qui permet d'accroître l'influence des atomes d'hydrogène).
SAPO	Somme sur tous les atomes de l'autopolarisabilité atomique définie par[28,29] $$4\sum_i\sum_a\sum_{p=1}^{N_k}\sum_{r=1}^{N_k}\frac{C_{pi}^k C_{ri}^k C_{pa}^k C_{ra}^k}{E_i - E_a}$$ E_i et E_a sont les énergies des orbitales moléculaires i et a. N_k est le nombre d'orbitales atomiques de valence de l'atome k. La sommation est faite sur les orbitales occupées, la sommation sur a est faite sur les orbitales moléculaires inoccupées C_{pi}^k, C_{ri}^k, C_{pa}^k et C_{ra}^k sont les coefficients des orbitales atomiques p ou r de l'atome k dans l'orbitale moléculaire i ou a.
SPHL	Définie comme la SAPO, mais en ne prenant en compte que la plus haute orbitale moléculaire occupée et la plus basse orbitale moléculaire inoccupée.
SDEL	Somme sur tous les atomes de la superdélocalisibilité atomique définie par[13,28] : $$\sum_i\sum_{p=1}^{N_k}\frac{(C_{ip}^k)^2}{E_i}$$ N_k est le nombre d'orbitales de valence de l'atome k, E_i est l'énergie de l'orbitale moléculaire i, C_{ip}^k est le coefficient de l'orbitale moléculaire i. La

	somme sur i est faite sur les orbitales moléculaires occupées.
SDHO	Définie comme la SDEL mais ne concerne que la plus haute orbitale moléculaire occupée.
POLV	Polarisabilité de la molécule calculée selon une méthode variationnelle[30] et définie par: $\alpha = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$ α_{xx}, α_{yy} et α_{zz} sont les trois composantes du tenseur de polarisabilité.
ANIV	Anisotropie de la polarisabilité calculée à partir du tenseur diagonalisé et définie par: $\sqrt{\frac{1}{2}\left[(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2\right]}$
CHAX*	Somme des valeurs absolues des charges nettes sur les atomes du fragment X.
QX*	Somme des charges nettes sur les atomes du fragment X
SDEX*	Somme des superdélocalisabilités calculées sur toutes les orbitales moléculaires occupées de tous les atomes du fragment X.
SAPX*	Somme des autopolarisabilités calculées sur toutes les orbitales moléculaires de tous les atomes du fragment X.
SHLX*	Somme des autopolarisabilités calculées sur la HOMO et la LUMO de tous les atomes du fragment X.
QO	Charge nette de l'atome d'oxygène du fragment commun
QNI	Charge nette de l'atome d'azote imine dans le fragment commun
QNA	Charge nette de l'atome d'azote amide dans le fragment commun

*X représente A, B, C, D ou le fragment commun

Tableau 2 (suite**)** : Symbole et définition des descripteurs calculés avec le logiciel Chem3D[26]

Symbole	Définition
LogP	Logarithme du coefficient de partage dans le mélange n-octanol/eau
MORE	Réfractivité molaire ou indice de réfraction molaire ($cm^3.mol^{-1}$)
MOWE	Masse moléculaire (u.m.a.)
COVO	(Connolly solvent-excluded Volume): le volume duquel le solvant est

	exclu ($Å^3$) (Voir annexe 3)
COSU	(Connolly accessible surface area): Surface de contact obtenue en faisant rouler sur la surface de Van der Waals, une sphère de rayon identique à celui du solvant ($Å^2$) (Voir détails **annexe 3**)
MOAR	Surface moléculaire de Connolly ($Å^2$) : Surface décrite par le centre de la sphère simulant le solvant et roulant sur la surface de Van der Waals.

c) – Utilisation des descripteurs calculés

Après avoir déterminé les différents descripteurs structuraux à corréler avec l'activité antagoniste des molécules étudiées, nous devons ensuite effectuer une étude statistique de ces données. Pour cela, nous avons utilisé le programme REGRE[31] qui permet de réaliser un traitement statistique, avec simultanément la méthode de régression multilinéaire et la méthode de validation croisée (*Confer* **Annexe 2**). Ce programme modifié au laboratoire de chimie théorique de Nancy I, fonctionne selon certains critères fixés par l'utilisateur et qui sont : le nombre de variables à corréler **p**, le coefficient de corrélation **R** et l'erreur standard de prédiction R^2_{cv} / R^2 (validation croisée). Ces trois valeurs permettent au programme de rechercher systématiquement, à partir d'un échantillon de **n** variables, toutes les régressions à **p** variables et de sélectionner celles conduisant à une corrélation dont le coefficient est supérieur à la valeur de **R** choisie et dont l'erreur standard de prédiction est supérieure à celle choisie par l'utilisateur. Ce dernier doit ensuite analyser ces résultats et ne retenir que les corrélations satisfaisant simultanément les critères statistiques suivants :

- Le coefficient de corrélation **R** doit avoir une valeur proche de **1**.
- La matrice d'inter-corrélation des variables ne doit contenir aucune valeur supérieure à **0.5**.
- La valeur de la variable **F** (FISCHER-SNEDECOR) doit être supérieure aux valeurs tabulées[32].
- L'erreur standard de prédiction R^2_{cv} / R^2 doit être proche de **1**.
- Le nombre de variables indépendantes doit être inférieur à **5** fois le nombre de molécules soumises à la corrélation (nombre de points de repères).

Notons que la définition et la formulation de chacun de ces paramètres statistiques et d'autres non mentionnés ci-dessus sont données dans **l'annexe 2.**

3)- Résultats

a) - Analyse conformationnelle

Le but de l'analyse conformationnelle effectuée est en fait la recherche de la conformation qui correspond au minimum énergétique absolu.

L'analyse indique que les conformères présentant le substituant lié au carbone asymétrique (**noté B dans le tableau 1**) en position équatoriale sur le cycle benzodiazépine sont plus stables que ceux le présentant en position axiale. Cette stabilité relative serait due à la faiblesse des contraintes stériques dans les 'conformères' équatoriaux. Il faut noter que les calculs quantiques au moyen de GEOMOS utilisant la méthode semi-empirique PM3 conduisent au même résultat. Les différences d'énergie entre les conformations axiale et équatoriale sont de 3 à 4 kcal/mol. Ces résultats sont en accord avec ceux obtenus, sur des séries analogues, par des travaux antérieurs utilisant les méthodes AM1[6] et MNDO[8].

Nos résultats préliminaires concernant l'effet du solvant sur la stabilité des conformères, nous a permis de constater que leur stabilité relative n'est pas affectée par l'effet de la solvatation. Par conséquent, nous pouvons retenir avec confiance le conformère le plus stable, obtenu dans le vide, par l'optimisation complète de la géométrie, pour le calcul des indices électroniques, physico-chimiques et stériques nécessaires pour l'étude statistique.

b) – Corrélations obtenues et discussion

Des 47 descripteurs locaux et globaux calculés et donnés à la fin de **l'annexe 2**, très peu ont été retenus par les tests statistiques. Nous les donnons dans le **tableau 3**, avec leur abréviation et leur définition (ils sont tous en unités atomiques sauf quand c'est indiqué). Nous donnons également leurs valeurs dans le **tableau 4**. Nous donnons ci-dessous, pour chaque récepteur les équations retenues lors de notre recherche de régressions (par ordre décroissant de l'erreur standard de prédiction, R^2_{cv} / R^2). La comparaison du taux de confiance de chacune d'elles avec l'indice de Fischer-Snedecor montre qu'elles sont toutes significatives à 99 % (Cf **Annexe 2**). L'écart type de la régression est représenté par **s**. Le nombre de paramètres est donné par **n**. Les chiffres qui sont entre parenthèses sont les intervalles de confiance. Dans chaque équation, les termes paramétriques sont listés par ordre de contribution décroissante (c.a.d. décroissance du coefficient de corrélation partiel avec l'activité).

Nous avons reporté dans les **tableaux 5-6,** les valeurs prédites par chaque équation retenue lors de notre recherche de corrélations statistiques significatives. Nous avons également

reporté dans les figures 1 (CCK_1) et 2 (CCK_2), la distribution des points (obtenue par la meilleure équation) autour de la droite idéale.

- ***Récepteurs CCK_1***

L'échantillon (jeu de données) utilisé dans les régressions multivariables données ci-après, contient soit l'énantiomère le plus actif du composé (dans le cas où les 2 activités sont disponibles), soit le racémique (*Cf* **tableau 1**). Le nombre de molécules soumises à la recherche de régressions (ou points de repères) est donc 20.

$$-\log (A_R)_{CCK1} = -704.4(\pm 220.8)QB^2 + 7.5(\pm 2.1)\mathrm{Ln}(CHA) - 1964.8(\pm 827.9)LUMO^2 - 10.0 \quad \textbf{(1)}$$

n = 20 ; s = 1.015 ; R = 0.771 ; F = 7.800 ; F(tabulé) = 5.29

$$-\log(A_R)_{CCK1} = -802.4(\pm 225.4)QB^2 + 8.0(\pm 2.0)\mathrm{Ln}(CHA) - 2381.8(\pm 857.1)LUMO^2 - 6.1(\pm 4.4)\,QC - 10.2 \quad \textbf{(2)}$$

n = 20 ; s = 0.985 ; R = 0.801 ; F = 6.696 ; F(tabulé) = 4.89

L'examen des corrélations (1) et (2) respectivement à trois et à quatre variables montre une cohérence qualitative dans l'évolution des résultats. Les deux équations indiquent que l'activité antagoniste devrait dépendre essentiellement de l'effet électronique du substituant **B**, lié au carbone asymétrique (**Tableau 3**). Cet effet est exprimé par le descripteur **QB** (somme des charges nettes sur les atomes du fragment **B**) qui représente l'indice quantique de réactivité chimique décrivant le fragment **B** dans la molécule isolée. Ces équations montrent que l'activité varie de façon parabolique avec l'indice **QB**, dont la valeur optimale est proche de zéro. Un tel résultat révèle que la neutralité électronique du substituant **B** est favorable à l'activité antagoniste du composé. En effet, l'énantiomère S du composé noté **B1** (**Tableau 1**) qui a la plus faible valeur de **QB** ($3\ 10^{-5}$ u.a.) présente une bonne activité sur le récepteur CCK_1. Rappelons qu'une bonne activité correspond à une forte valeur positive de $-\log(A_R)$.

Le coefficient positif de Ln(**CHA**), qui représente le logarithme de la somme des valeurs absolues des charges nettes sur les atomes de la molécule et qui apparaît dans chaque corrélation, laisse suggérer que le caractère polaire de la molécule est en faveur de l'activité. Les corrélations paraboliques par rapport au descripteur **LUMO** qui est l'énergie de la plus basse orbitale moléculaire inoccupée, suggèrent qu'une petite valeur de **LUMO** (en valeur absolue) est en faveur d'une bonne activité. D'un point de vue physique, le descripteur **LUMO** représente la capacité qu'a une molécule à accepter des électrons. Notre résultat

reliant une bonne activité à une valeur proche de zéro de **LUMO** laisse supposer une faible interaction par transfert de charge au niveau du site de liaison avec le récepteur CCK_1. Le coefficient négatif du paramètre électronique **QC** suggère que le substituant **C** (lié au Carbone 5 de la benzodiazépine, Tableau 3) ayant un pouvoir accepteur d'électrons aura un effet positif sur l'activité. Il est important de noter que les intervalles de confiance (erreur standard) des coefficients de régression, donnés entre parenthèses, sont très grands et en particulier pour **QB** et **LUMO**. Ainsi, d'un point de vue statistique, ces équations devraient être utilisées seulement dans le cas d'une conclusion qualitative.

Une analyse de régression multiple réalisée sur l'ensemble des composés (couples d'énantiomères et racémiques, voir **Tableau 1**) conduit aux équations suivantes, avec trois puis quatre variables

$$-\log (A_R)_{CCK1} = 4.4(\pm1.2)\mathrm{Ln(CHAB)} - 2.7(\pm0.9)\mathrm{Ln\ (ANIV)} - 1926.2(\pm677.8)\mathrm{LUMO}^2 + 6.5 \quad \textbf{(3)}$$

n = 29 ; s = 1.068 ; R = 0.665 ; F = 6.876 ; F(tabulé) = 4.68

$$-\log (A_R)_{CCK1} = 4.5(\pm1.1)\mathrm{Ln(CHAB)} - 2.6(\pm1.1)\mathrm{Ln(ANIV)} - 2312.6(\pm706.8)\mathrm{LUMO}^2 + 6.6(\pm4.3)\ \mathrm{QR} + 7.9 \quad \textbf{(4)}$$

n = 29 ; s = 1.041 ; R = 0.700 ; F = 6.020 ; F(tabulé) = 4.22

Malgré leur faible qualité statistique les corrélations données ci-dessus montrent une bonne évolution qualitative dans les résultats. Elles indiquent que l'activité antagoniste sur le récepteur CCK_1 est essentiellement gouvernée par l'effet électronique du substituant B. Cet effet est exprimé par Ln(**CHAB**) : le logarithme de la somme des valeurs absolues des charges nettes sur tous les atomes du fragment **B**. Le coefficient positif du descripteur Ln(**CHAB**) suggère que l'activité antagoniste sur le récepteur CCK_1 peut être améliorée par un caractère fortement polaire de ce substituant. Ce résultat, en accord avec une étude précédente basée sur des descripteurs empiriques[5], pourrait être associé à une existence probable d'interactions électrostatiques dans le site de liaison environnant ce substituant. Le coefficient négatif du descripteur Ln(**ANIV**), qui est le logarithme de l'anisotropie de la polarisabilité signifie qu'une grande déformation du nuage électronique de la molécule affaiblirait l'activité. Ce phénomène peut arriver quand le médicament s'approche du récepteur ; ce qui conduit à supposer une poche limitée voire une petite poche au niveau du site de liaison avec le récepteur CCK_1. Les variations paraboliques de l'activité avec le

paramètre **LUMO** (Eqs 3-4) suggèrent que la valeur optimale de ce descripteur est proche de zéro. La présence de **QR** relatif au fragment commun benzodiazépine, dans l'équation (4), avec un coefficient positif signifie qu'un effet attracteur d'électrons des substituants (i.e. **A**, **C** et **D**, **Tableau 3**) est nécessaire pour une bonne activité de la molécule sur le récepteur CCK_1. Néanmoins, ces résultats dérivant des équations 3-4 devraient être considérés avec prudence à cause de leur faible signification statistique.

Il est important de noter dans toutes les équations ci-dessus, l'absence du descripteur **STRO** indiquant la configuration de la molécule. A notre avis, cette absence pourrait être due au fait que la série considérée ne met pas assez en évidence la sélectivité selon l'énantiomère et surtout que ce descripteur est mal défini avec des valeurs discrètes.

- ***Récepteurs CCK_2***

Compte tenu du petit nombre de molécules dont l'activité sur le récepteur CCK_2 est disponible, les régressions multilinéaires retenues et données ci-dessous sont basées seulement sur deux variables. Elles dérivent du jeu de données constitué soit du plus actif des 2 énantiomères de la molécule (quand les 2 sont disponibles) soit du mélange racémique (*Cf* **Tableau 1**). Le nombre de points de repères est donc seulement de 14.

$$-\log (A_R)_{CCK2} = 17.1(\pm 2.9)\mathrm{Ln}\,(\mathrm{POLV}) - 55.3(\pm 9.6)\,\mathrm{QB} - 70.8 \qquad \textbf{(5)}$$

n = 14 ; s = 0.699 ; R = 0.935 ; F = 38.264 ; F(tabulé) = 7.20

$$-\log (A_R)_{CCK2} = 15.3(\pm 2.6)\mathrm{Ln}(\mathrm{COVO}) - 51.4(\pm 10.2)\mathrm{QB} + 90.4 \qquad \textbf{(6)}$$

n = 14 ; s = 0.735 ; R = 0.928 ; F = 34.021 ; F(tabulé) = 7.20

L'examen des paramètres statistiques montre que les corrélations (5) et (6) sont très significatives. En effet, les mesures de la proximité de l'ajustement des activités prédites par les équations (5) et (6) sont respectivement de 87% et de 86% ($R^2 = 0.87$; $R^2 = 0.86$). Le coefficient positif du descripteur Ln(**POLV**) dans l'équation (5) suggère que l'activité antagoniste est améliorée par une augmentation de la polarisabilité de la molécule, qui est une mesure globale de la distorsion de la molécule quand elle est exposée à un champ électrique externe. Ce champ électrique provient évidemment du récepteur ou de son environnement. Par conséquent, cette corrélation suggère l'implication des forces d'induction et de dispersion au sein du site de liaison avec le récepteur CCK_2. L'équation (6) met en évidence le rôle

majeur de Ln(**COVO**) qui est le logarithme du volume duquel le solvant est exclu (Connoly solvent-excluded volume). L'intervention de ce descripteur stérique avec un coefficient positif semble indiquer l'existence d'une large poche au niveau du site de liaison avec le récepteur CCK_2. Dans les deux équations le coefficient du descripteur **QB** est négatif, ce qui signifie que l'activité antagoniste augmente proportionnellement avec l'augmentation en valeur absolue de la charge négative du substituant **B, lié au carbone asymétrique**. Un tel résultat reflète qu'un effet électronique attracteur de ce substituant améliore l'activité de la molécule sur le récepteur CCK_2. Ce résultat, comparé avec celui relatif au récepteur CCK_1, nous permet de trouver la sélectivité d'une nouvelle molécule benzodiazépine vis-à-vis de l'un ou l'autre des récepteurs CCK. En effet, si une molécule a un **QB** proche de zéro, elle serait plus active sur le récepteur CCK_1.

L'analyse de régressions multilinéaires réalisée sur l'ensemble des composés pour lesquels les valeurs des activités expérimentales sont disponibles (i.e. paires d'énantiomères ou mélange racémique) conduit aux équations suivantes :

$$-\log (A_R)_{CCK2} = 11.8(\pm3.3)\text{Ln (COVO)} - 47.8(\pm15.8)\text{QB} - 70.3 \quad \textbf{(7)}$$

n = 22 ; s = 1.174 ; R = 0.776 ; F = 14.401 ; F(tabulé) = 5.93

$$-\log (A_R)_{CCK2} = 12.7(\pm3.6)\text{Ln (POLV)} - 52.1(\pm15.6)\text{QB} - 53.5 \quad \textbf{(8)}$$

n = 22 ; s = 1.175 ; R = 0.775 ; F = 14.321 ; F(tabulé) = 5.93

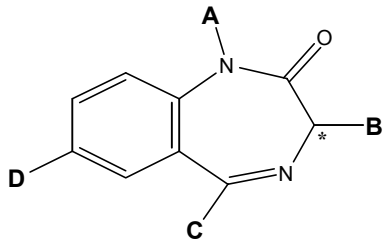
$$-\log (A_R)_{CCK2} = 12.9(\pm3.0)\text{Ln(POLV)} - 55.3(\pm12.8)\text{QB} - 0.7(\pm0.2)\text{STRO} + 54.1 \quad \textbf{(9)}$$

n = 22 ; s = 0.964 ; R = 0.864 ; F = 17.620 ; F(tabulé) = 5.09

Il est important de noter la cohérence qualitative dans l'évolution des résultats obtenus au moyen des corrélations à deux variables indépendantes (Eqs 7-8) puis trois variables (Eq. 9). En effet, nous observons la présence des mêmes descripteurs dans les équations et avec le même effet. Il est également intéressant de noter une similarité dans l'apparition des descripteurs avec les équations (5) et (6). Cependant, la signification statistique de ces dernières corrélations diminue malgré l'augmentation du nombre de molécules soumises à la corrélation (**n**). Ceci peut être dû au fait que le descripteur indicateur de la configuration soit mal défini par une valeur discrète. Néanmoins, la présence du descripteur fictif **STRO** dans l'équation (9) suggère qu'une valeur de '1' pour ce descripteur et qui correspond à la

configuration R, est en faveur d'une bonne activité sur le récepteur CCK_2. Ce résultat est en accord avec les observations statistiques.

Tableau 3 : Structure générique des composés étudiés, définition des descripteurs retenus par les meilleures corrélations et type de récepteur dans lequel le descripteur apparait :



Descripteur	Définition	Récepteur
LUMO	Energie de la plus basse orbitale moléculaire inoccupée.	CCK_1
CHA	Somme des valeurs absolues des charges atomiques de la molécule.	CCK_1
CHAB	Somme des valeurs absolues des charges nettes sur les atomes du fragment B.	CCK_1
QB	Somme des charges nettes sur les atomes du fragment B	CCK_1 CCK_2
QC	Somme des charges nettes sur les atomes du fragment C	CCK_1
QR	Somme des charges nettes sur les atomes du fragment commun noté R	CCK_1
POLV	Polarisabilité de la molécule calculée selon une méthode variationnelle et définie par: $\alpha = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$ α_{xx}, α_{yy} et α_{zz} sont les trois composantes du tenseur de polarisabilité.	CCK_2
ANIV	Anisotropie de la polarisabilité calculée à partir du tenseur diagonalisé et définie par: $\sqrt{\frac{1}{2}\left[(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2\right]}$	CCK_1
COVO	(Connolly solvent-excluded Volume), le volume duquel le solvant est exclu ($Å^3$)	CCK_2

Tableau 4: Valeurs pour chaque composé des descripteurs retenus par les meilleures corrélations, Eqs (1-9) .

Comp.	- LUMO	CHA	CHAB	QB	QC	- QR	POLV	ANIV	COVO
B1	0.0244	5.8025	2.6186	0.00003	0.0263	0.2259	58.499	15.312	324.25
B2	0.0264	6.6974	2.0894	0.05270	-0.0017	0.1925	59.388	25.284	342.31
B3	0.0179	5.2240	2.0239	0.03850	-0.0054	0.2259	57.301	18.717	324.92
B4	0.0184	5.1239	2.0189	0.04189	-0.0071	0.2315	54.889	26.049	302.57
B5	0.0228	5.2025	2.0181	0.04005	0.0192	0.2215	60.249	26.976	334.70
B6	0.0244	4.8929	1.9428	0.06050	0.0218	0.2468	57.283	25.837	311.70
B7	0.0237	5.0805	1.8860	0.00023	0.0025	0.2022	53.656	11.061	317.86
B8	0.0251	5.2578	2.0726	-0.0109	0.0266	0.2160	56.364	17.040	321.44
B9	0.0221	5.2971	2.0978	-0.0079	0.0248	0.2141	57.698	20.247	325.85
B10	0.0211	5.3193	2.0569	-0.0127	0.0883	0.2725	60.445	15.434	355.33
B11	0.0187	5.7397	2.1559	-0.0240	0.1862	0.3561	62.242	18.331	371.65
B12	0.0199	5.6950	2.1172	-0.0151	0.1272	0.3074	66.568	22.387	386.23
B13	0.0338	6.4468	3.1728	-0.0398	0.0309	0.1899	56.742	19.383	328.49
B14	0.0322	5.8909	2.8137	-0.0168	0.0286	0.2150	60.040	20.924	360.44
B15	0.0407	6.5648	2.6783	-0.0105	0.0294	0.2265	68.232	17.207	395.78
B16	0.0249	7.3750	2.8392	-0.0006	0.0318	0.1835	60.414	17.392	352.30
B17	0.0254	5.9270	2.8254	-0.0007	0.0015	0.2032	54.862	16.710	312.59
B18	0.0245	4.1995	1.0775	-0.0002	0.0011	0.2029	39.555	12.754	239.37
B19	0.0208	5.3295	2.0828	-0.0143	0.1330	0.3154	59.231	17.032	351.42
B20	0.0206	5.1738	2.1299	-0.0119	0.1465	0.3320	56.462	18.706	327.92

Tableau 5: Valeurs observées et prédites de $-\log (A_R)_{CCK1}$

Composé	Stéréo	Observée	Prédite Eq.1	Prédite Eq.2	Prédite Eq.3	Prédite Eq.4
B1	S	3.544	2.250	2.025	2.244	2.354
	R	1.630	2.250*	2.025*	2.244	2.354
B2	R	0.970	1.099	0.941	-0.294	0.033
B3	R	1.015	1.069	0.736	1.100	1.312
B4	S	-0.578	0.663*	0.364*	0.165	0.372
	R	0.748	0.663	0.364	0.165	0.372
B5	R	0.301	0.312	0.228	-0.279	-0.072
B6	S	-1.057	-2.023*	-1.826*	-0.476	-0.474
	R	-0.084	-2.023	-1.826	-0.476	-0.474

B7	S	2.856	1.412	1.098	1.740	1.944
	R	1.984	1.412*	1.098*	1.740	1.944
B8	S	1.032	1.281	1.137	0.861	1.010
	R	-0.594	1.281*	1.137*	0.861	1.010
B9	S	1.970	1.734	1.510	0.723	0.961
	R	0.000	1.734*	1.510*	0.723	0.961
B10	R	-0.387	1.402	1.557	1.450	1.284
B11	S	1.969	1.306	2.021	1.378	0.722
	R	-0.610	1.306*	2.021*	1.378	0.722
B12	S	2.053	1.774	2.117	0.671	0.342
	R	-0.338	1.774*	2.117*	0.671	0.342
B13	R	-0.324	0.490	0.622	1.400	1.586
B14	S	2.077	1.078	1.072	0.870	0.926
	R	0.105	1.078*	1.072*	0.870	0.926
B15	S	0.288	0.602	0.785	-0.014	-0.302
	R	-0.289	0.602*	0.785*	-0.014	-0.302
B16	RS	3.300	4.076	3.769	2.209	2.614
B17	RS	3.124	2.452	2.086	2.247	2.506
B18	RS	-1.154	-0.194	-0.400	-1.179	-1.039
B19	RS	1.636	1.139	1.565	1.264	0.833
B20	RS	0.185	0.889	1.404	1.126	0.602

*: cet énantiomère n'est pas utilisé dans la détermination de cette corrélation

Tableau 6: Valeurs observées (expérimentales) et prédites de $-\log (A_R)_{CCK2}$

Composé	Stéréo	Observée	Prédite Eq.5	Prédite Eq.6	Prédite Eq.7	Prédite Eq.8	Prédite Eq.9
B1	S	-2.088	-1.042	-1.694	-2.128	-1.646	-2.494
	R	-3.699	-1.042*	-1.694*	-2.128	-1.646	-1.056
B2	R	-3.699	-3.698	-3.572	-4.008	-4.197	-3.776
B3	R	n.d.	-3.525*	-3.674*	-3.943*	-3.912*	-3.449*
	S	n.d.	-3.525*	-3.674*	-3.943*	-3.912*	-4.885*
B7	S	-3.161	-2.535	-2.010	-2.372	-2.758	-3.616
	R	-3.740	-2.535*	-2.010*	-2.372	-2.758	-2.179
B8	R	-0.439	-1.075	-1.265	-1.708	-1.551	-0.930
	S	-2.312	-1.075*	-1.265*	-1.708	-1.551	-1.551
B9	R	0.000	-0.840	-1.211	-1.691	-1.409	-0.799
	S	-1.878	-0.840*	-1.211*	-1.691	-1.409	-2.233
B10	R	1.482	0.222	0.364	-0.440	-0.566	0.068
B11	R	0.799	1.350	1.634	0.630	0.935	1.069
	S	-1.229	1.350*	1.634*	0.630	0.935	-0.367
B12	R	1.929	2.010	1.767	0.658	0.788	1.441
	S	-0.497	2.010*	1.767*	0.658	0.788	0.003
B13	R	0.104	0.638	0.553	-0.070	0.039	0.753
B14	R	0.920	0.334	0.794	-0.075	-0.438	0.208
	S	-1.469	0.334*	0.794*	-0.075	-0.438	-1.229
B15	R	1.777	2.178	1.905	0.726	0.863	1.504
	S	-0.477	2.178*	1.905*	0.726	0.863	0.066

B16	R	n.d.	-0.453*	-0.390*	-1.118*	-1.264*	-0.606*
	S	n.d.	-0.453*	-0.390*	-1.118*	-1.264*	-2.042*
B17	RS	-2.176	-2.102	-2.218	-2.525	-2.427	-2.560
B18	R	n.d.	-2.101*	-6.313*	-5.671*	-2.425*	-3.277*
	S	n.d.	-2.101*	-6.313*	-5.671*	-2.425*	-1.841*
B19	RS	0.174	-0.036	0.227	-0.493	-0.741	-0.823
B20	RS	1.207	-0.990	-0.908	-1.425	-1.477	-1.571

n.d. : non disponible

*: cette molécule n'est pas utilisée dans la détermination de cette corrélation

Figure 1 : Distribution des valeurs de $-\log(A_R)$ (prédites et expérimentales) autour de la droite idéale. Les valeurs sont prédites par l'équation 2 (CCK_1).

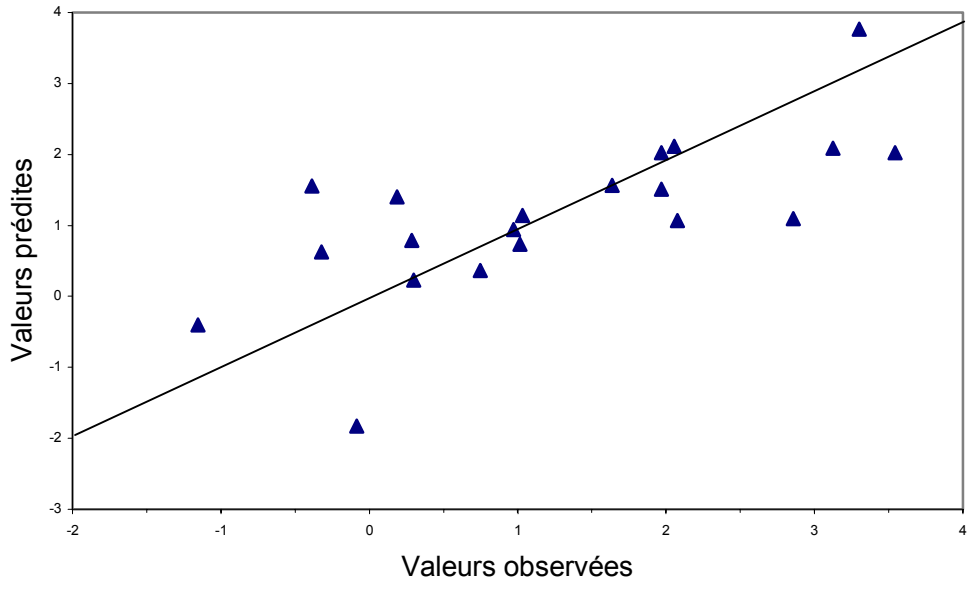
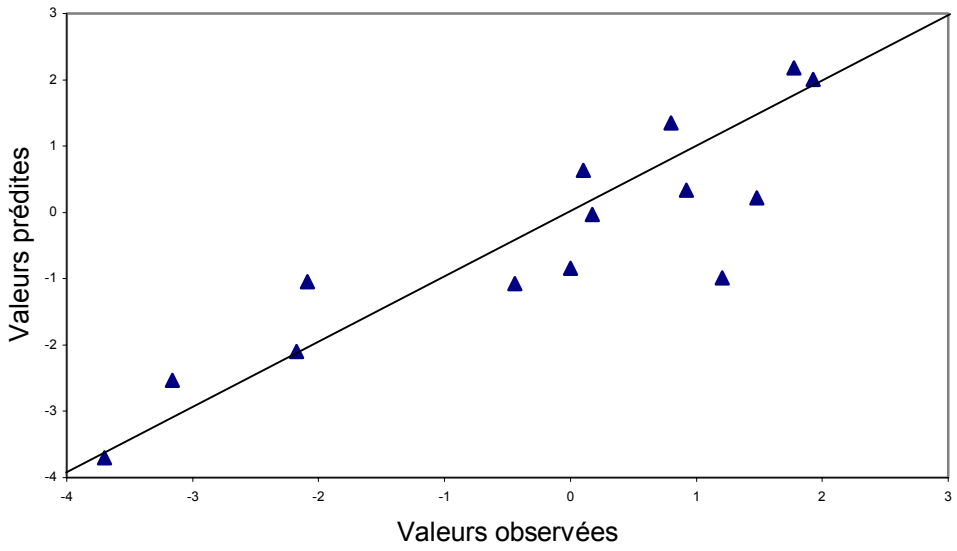


Figure 2 : Distribution des valeurs de $-\log(A_R)$ (prédites et expérimentales) autour de la droite idéale. Les valeurs sont prédites par l'équation 5 (CCK_2).



III - CONCLUSION

Les recherches de corrélations réalisées au moyen de descripteurs électroniques, sur une série de benzodiazépines agissant en tant qu'antagonistes sur les deux types de récepteurs de la cholécystokinine, ont permis de mettre en évidence les éléments structuraux gouvernant l'activité antagoniste. Les résultats ont montré en particulier le **rôle prédominant du substituant lié au carbone asymétrique de la molécule** (noté **B**, **Tableau 3**). Il apparaît, un effet électronique optimal de ce substituant selon le type de récepteur CCK. Cet effet électronique doit être neutre et polaire pour que la molécule puisse agir comme un bon antagoniste sur le récepteur CCK_1. Notons que la neutralité de ce substituant avec une valeur proche de zéro pour **LUMO** (**Tableau 3**) signifie que les interactions par transfert de charge au niveau du site de liaison CCK_1 ne sont pas importantes. En revanche, dans le cas du récepteur CCK_2, nous constatons qu'un effet attracteur de ce substituant est favorable à l'activité antagoniste de la molécule. En d'autres termes, nos résultats montrent qu'un substituant, lié au carbone asymétrique, fortement attracteur d'électrons, confère une bonne activité antagoniste à la molécule sur le récepteur CCK_2. Ce résultat est en accord avec les conclusions de précédentes études[6,8,11] conduites sur des composés analogues. Ces études ont considéré d'autres paramètres moléculaires (stériques, de forme, électrostatiques, etc…) et/ou

d'autres techniques QSAR (3D-MSA-QSAR, réseau de neurones, etc…). Il est intéressant de noter que notre résultat nous permet de déterminer la sélectivité d'une molécule non encore testée. Par ailleurs, nos résultats de calcul montrent clairement l'importance des caractéristiques stériques et électrostatiques des antagonistes CCK_2 tels que les systèmes benzodiazépines. Ceci suggère l'implication possible des forces de type van der Waals dans les interactions ligand/récepteur, alors que ce n'est pas le cas avec le récepteur CCK_1. L'analyse de nos résultats révèle également un site de liaison au sein du récepteur CCK_1, caractérisé par une poche plus petite et plus hydrophile que celle apparaissant dans l'autre type de récepteur.

Le descripteur **STRO** qui discrimine les énantiomères apparaît seulement dans les corrélations relatives au récepteur CCK_2. Ceci pourrait être associé au fait que les énantiomères sont souvent tous les deux actifs vis-à-vis du récepteur CCK_1 et sont donc mal discriminés par un descripteur représenté au moyen d'une valeur discrète. La chiralité, malgré son importance dans les applications en chimie médicinale, demeure vraiment un concept évasif, voire insaisissable quand il s'agit de l'introduire dans les méthodes QSAR. Très récemment, plusieurs auteurs[33-39] ont focalisé leur attention sur la possibilité de décrire la chiralité comme une propriété 'continue' au lieu d'une propriété 'black and white' telle que celle utilisée dans cette présente étude. De plus, l'usage simultané de variables discrètes et continues dans la même étude QSAR ne semble pas conduire à des résultats satisfaisants comme le prétendent certains auteurs[5,6]. Par conséquent, afin d'améliorer le pouvoir prédictif des modèles générés ici, nous proposons de considérer ultérieurement la chiralité comme une mesure quantitative qui dérive des indices de similarité moléculaire[39].

Il est important de signaler que contrairement aux conclusions d'autres travaux[5 et références citées] nos calculs ont mis en évidence aucune corrélation significative avec le logP, le logarithme du coefficient de partage dans un mélange n-octanol/eau qui permet d'évaluer les effets imputables à la lipophilie. Le caractère lipophile d'une molécule, probablement lié au processus de transport *in vivo,* traduit souvent des propriétés biologiques comme le métabolisme, la distribution dans les tissus. Une variation de la lipophilie entraîne ainsi une modification de la concentration de la molécule sur le site récepteur. Par conséquent, il n'est pas surprenant que de tels facteurs soient absents dans nos résultats puisque les mesures biologiques considérées ici sont effectuées *in vitro*. Signalons par ailleurs, que B.P. Roques, dans ses travaux[40] sur les récepteurs de la morphine et de l'enképhaline, souligne l'importance des facteurs stériques et électroniques de la molécule et, plus particulièrement

des substituants, mais parle très peu des processus de transport qui vraisemblablement ralentissent, dans certains cas, les effets biologiques sans les inhiber complètement.

Nos résultats, bien qu'insuffisants pour donner une conclusion définitive, ont cependant mis en évidence la validité de l'application des descripteurs quantiques dans les études QSAR. En effet, la chimie quantique offre plus d'opportunités pour décrire avec précision la structure électronique d'une molécule à travers beaucoup de paramètres dont certains sont utilisés dans notre étude, et qui sont inaccessibles au moyen d'une autre méthode théorique. En outre, l'amélioration croissante à tout point de vue (rapidité, performance, mémoire et coût) des ordinateurs fait de la chimie quantique un outil majeur. Ainsi, l'application des descripteurs quantiques à un échantillon représentatif peut alors être une étape décisive dans l'analyse du mécanisme d'interaction des complexes ligand/récepteur. Sous certaines conditions, cette étape peut être préalable dans la recherche de corrélations significatives.

Étant donné que ce travail nous a permis de proposer une conformation préférentielle pour chacun des composés étudiés et afin d'aller plus loin dans la compréhension du mode d'interaction ligand/récepteur, nous poursuivons cette étude par une analyse de docking dans le chapitre suivant.

BIBLIOGRAPHIE

1 - Evans, B. E. , Bock, M. G. , Rittle, K. E. , DiPardo, R. M. , Whitter, W. L. , Veber, D. R., Anderson P. S. et Freidinger, R. M. , *Proc. Natl. Acad. Sci.* USA 83, 4918, **1986**.

2 - Chang, R. S. L. et Lotti, V. J. , *Proc. Natl. Acad. Sci.* USA 83, 4923, **1986**.

3 - Semple, G. , Ryder, H. , Kendrick, D. A. , Szelke, M. , Ohta, M. , Satoh, M. , Nishida, A. , Muzawa, S. et Miyata, K. , *Bioorg. Med. Chem.* 6, 51, **1996**.

4 - Semple, G. , Ryder, H. , Kendrick, D. A. , Szelke, M. , Ohta, M. , Satoh, M. , Nishida, A. , Akuzawa S. et Miyata, K. , *Bioorg. Med. Chem.,* 6, 55, **1996.**

5 - Sinha, J. , Kurup, A. , Paleti A. et Gupta, S. P. , *Bioorg. Med. Chem.*7, 1127, **1999**.

6 - Huché M. et Legendre, J. L. , *Quant. Struct. -Act. Relat.* ,16, 435, **1997**.

7 - Gupta, S. P. , Mulchadani, V. , Das, S. , Subbiah, A. , Reddy, D. N. et Sinha, J. *Quant. Struct. -Act. Relat.*14, 437, **1995**.

8 - Tokarsky J. S. et Hopfinger, A. J. , *J. Med. Chem.*, 37, 3639, **1994**.

9 – Taïri-Kellou, S., Cartier, A., Maouche, B. et Maigret, B., *J. of Mol. Struct. (THEOCHEM)*, 571/1-3, 207-223, **2001**.

10 - Kerwin, J. F., Drugs of the Future, 16, 1111, **1991**.

11 - Yamazaki, Y. , Akahana, M. , Kobayashi, M., Kitazawa, M., Kurashina, Y. et K. Lizuka, *Japan J. Pharmacol.,* 63, 219, **1993**.

12 - Bock, M. G., DiPardo, R. M., Evans, B. E. , Rittle, K. E., Whitter, W. L. , Veber, D. L. Anderson, P. S. et Freidinger, R. M. , *J. Med. Chem.*, 32, 16, **1989**.

13 - Cammarata A., in L.B. Kier (Editor), Molecular Orbital Studies in Chemical Pharmacology, Springer Verlag, Berlin, New York, **1970**, pp. 156-190.

14 - Buydens, L., Massart D. L. et Geerlings, P., Journal of Chromatographic Science, 23, 304, **1985**.

15 - Cartier, A. et Rivail, J. L., Chemometrics et Intelligent Laboratory Systems, 1, 335, **1987**.

16 - Computational results obtained using software from Biosym/MSI of San Diego-dynamics calculations and minimizations were done with the Discover program, using CVFF forcefield.

17 - Kirckpatrick, S., Gelatt, C. D. et Vechi, M. P., Science, 220, 671, **1983**.

18 - Rinaldi, D., Hoggan, P. E. et Cartier, A., The GEOMOS program, QCPE 584a, QCPE Bull., 128, **1989**.

19 - Stewart, J. J. P. Journal of Computational Chemistry, 10, 209, **1989**.

20 - Dewar, M. J. S. et Thiel, W., J. Amer. Chem. Soc., 99, 4899, **1977**.

21 - Coitino, E. L., Irving, K., Rama, J., Iglesias, A., Paulino, N. et Ventura, O. N., J. Mol. Struct.(Theochem), 210, 405, **1990**.

22 - Coitino, E. L., Ventura, O. N. et Sosa, R. M., J. Mol. Struct.(Theochem), 254, 315, **1992**.

23 - Cartier, A., Catalogue de l'ANL, SFC, **1985**.

24 – Tairi-Kellou, S. et al, *Conformational equilibria of (1,4)-benzodiazepine. Study of solvent effects, (En preparation*)

25 – Rivail, J. L., Rinaldi, D. et Ruiz-Lopez, M. F., in S.J. Formosinho, I.G. Csizmadia and L. Arnaut (Eds.), Theoretical and Computational Models for Organic Chemistry, NATO ASI Ser. C, Vol. 339, Kluwer, Dordrecht, pp- 79-92, **1991.**

26 - CS Chem Office, Ultra. For Microsoft Windows, Cambridge Scientific Computing Inc., 875 Massachusetts Avenue, Suite 61, Cambridge, MA 02139, USA.

27 – Terryn, B. et Barriol, J., Journal de chimie physique,78, 207, **1981**.

28- Brown, R. E. et Simas, A. M., *Theoretica Chimica Acta*, 62, 1, **1982**.

29 - Coulson, C. A. et Longuet-Higgins, H. C. Proceedings of the Royal Society of London, Series A : Mathematical and Physical Sciences, 192, 16, **1947.**

30 - Rinaldi, D. et Rivail, J. L., *Theoretica Chimica Acta*, 32, 243, **1974**.

31- Programs from the A.D.D.A.D.'s statistical software, Association pour le Développement et la Diffusion de l'Analyse de Données, Laboratoire de statistique, Université P. et M. Curie, 4 place Jussieu, Paris 5, France.

32- Pearson, F. A. et Bennett, K. R., Statistical Methods Applied to Agricultural Economics, John Wiley & Sons, New York, **1947**.

33- Buda, A. B., Auf der Heyde, T. et Mislow, K., *Angew. Chem. Int. Ed. Engl*., 31, 989, **1992**.

34- Randic, M. et Razinger, M. J., *Chem. Inf. Comput. Sci.*, 36, 429, **1996**.

35- Harris, A. B., Kamien, R. D. et Lubienski, T. C., *Rev. Modern Phys.,* 71, 1745, **1999**.

36- M. Randic, *Acta Chim. Slov.,* 45, 239, **1998**.

37- Zabrodsky, H., Peleg, S. et Avnir, D., *J. Am. Chem. Soc*., 114, 7843, **1992**.

38- Lipkowitz, B. K., Gao, D. et Katzenelson, O., *J. Am. Chem. Soc.,* 121, 5559, **1999**.

39- Benigni, R., Cotta-Ramusino, M., Gallo, G., Giorgi, F. Giuliani, A. et Vari, M. R., *J. Med. Chem.,* 43, 3699, **2000**.

40- Roques, B. P., La vie des Sciences, Comptes rendus de l'académie des Sciences, Série Générale, Tome 4, N° 3.

CHAPITRE IV

MODELISATION DES RECEPTEURS DE LA CHOLECYSTOKININE.
ETUDE DU SITE DE LIAISON DES ANTAGONISTES NONPEPTIDIQUES

I – INTRODUCTION

Nous avons fait remarqué dans le premier chapitre l'appartenance des deux récepteurs de la cholécystokinine à la famille très étendue des récepteurs couplés aux protéines G (**RCPGs**), *cf* annexe 1. Cette famille de récepteurs comprend plus de 5000 protéines, ce qui représente 1 à 2 % de l'ensemble des gènes du génome des mammifères. De plus, 80 % des récepteurs qui reconnaissent des hormones et des neurotransmetteurs font partie de cette famille de récepteurs [(1)]. Leur grand nombre et la diversité de leurs propriétés physiologiques (vision, olfaction, médiation de l'action de nombreuses hormones …) font d'eux une classe de récepteurs très importante. L'analyse détaillée des séquences des RCPGs a montré que tous ces récepteurs dérivent d'un ancêtre commun par mutation ou délétion d'un résidu dans le gène original. La fonction de ces macromolécules est de permettre la communication cellulaire en recevant les signaux *via* le médiateur, puis de transmettre ces signaux en activant des protéines G intracellulaires. Ces systèmes ubiquitaires représentent donc des cibles pharmacologiques de première importance.

La structure secondaire de ces RCPGs, fondée sur l'analyse du profil d'hydrophobicité de la protéine et parfois sur des expériences de mutagenèse dirigée, est spéculative pour la grande majorité de ces récepteurs. Cependant, aucune expérience n'est venue infirmer la structure considérée hypothétique pour beaucoup d'entre eux. En effet, l'analyse en microscopie électronique du prototype des RCPGs : la rhodopsine (protéine photo réceptrice qui est présente dans les cellules en bâtonnet de la rétine), à l'état de cristal bidimensionnel et ayant permis de définir sa morphologie avec une résolution de 9 Å, confirme bien la présence de sept domaines transmembranaires[(2)]. Ces derniers constituent le 'corps central' de ces récepteurs et le changement de conformation de cette région transmembranaire, qui fait suite à la liaison avec le ligand, est probablement responsable de l'activation des RCPGs[(3)].

L'étude théorique des mécanismes d'activation du récepteur et de son interaction avec le ligand nécessite la connaissance de sa structure spatiale, ce qui aiderait considérablement le développement d'agents thérapeutiques originaux. L'étude en trois dimensions (3D) de la structure de ces récepteurs et de celle des complexes qu'ils sont susceptibles de former avec leur ligand, est possible grâce aux techniques de la modélisation moléculaire et de l'infographie. De manière idéale et afin de minimiser les spéculations faites sur les modes d'interaction ligand/récepteur, les modèles des RCPGs devraient être construits sur la base des structures cristallographiques à haute résolution. Cependant, l'obtention de cristaux adéquats pour une analyse par diffraction aux rayons X, d'un quelconque récepteur transmembranaire

est encore un défi à relever, étant donné les difficultés rencontrées au cours de leur purification et, en particulier, de leur cristallisation[4].

La première structure tridimensionnelle connue de récepteur à sept hélices transmembranaires est celle de la bactériorhodopsine (**BR**)[5]. Bien que cette dernière soit couplée à un transfert de proton et non à une protéine **G**, et qu'elle présente très peu d'homologie de séquence avec les RCPGs, l'analyse de son indice d'hydropathie a mis en évidence sa similarité de structure avec les RCPGs[6]. De plus, elle possède de nombreux points communs avec la rhodopsine des mammifères qui, elle, appartient à la famille des RCPGs. En effet, ces deux protéines réceptrices sont toutes deux des pompes à protons activées par la lumière et liées de manière covalente au même ligand : le rétinal. Ces similitudes ont conduit à utiliser la structure tridimensionnelle de la BR, disponible dans la Protein Data Bank[7], comme point de départ à l'élaboration de modèles qui permirent une investigation en trois dimensions (jusqu'alors impossible) des modes d'interaction des RCPGs avec leurs ligands respectifs[8-16]. Cependant, en dépit des approches similaires qui ont été utilisées pour modéliser les RCPGs, il existe une grande variabilité entre les différents alignements. Cette variabilité reflète la difficulté à générer des alignements et le degré de subjectivité qui en résulte inévitablement. La carte de densité électronique à basse résolution (9Å) de la rhodopsine de bœuf a été établie en Avril 1993[2]. La similarité entre les deux cartes à même résolution valide, en quelque sorte, les modèles fondés sur la BR[17]. Néanmoins, il existe d'autres alternatives[18-20] aux modèles basés sur la BR, par exemple, celles qui utilisent les techniques standard de la modélisation moléculaire : construction individuelle des hélices α et dynamique moléculaire.

Il est important de noter qu'actuellement on a accès essentiellement à l'organisation des sept hélices transmembranaires les unes par rapport aux autres, mais la structure des régions intracellulaires et extracellulaires qui sont mal conservées entre les RCPGs, reste encore inconnue de façon précise.

Dans ce travail, nous nous proposons d'élaborer, d'une part, le modèle du récepteur CCK_1 du rat, et d'autre part, d'établir des modèles de complexes que les récepteurs de la CCK sont susceptibles de former avec leur antagoniste de type benzodiazépine. Une série de ces antagonistes a fait l'objet d'une étude de relations quantitatives structure activité[21], présentée dans le **chapitre précédent**. Signalons que le récepteur CCK_2 n'a pas été construit dans le cadre de cette thèse car il a été modélisé et validé antérieurement par B. Maigret, dans le cadre d'une collaboration avec l'équipe de B.P. Roques du laboratoire de Pharmacochimie Moléculaire et Structurale de Paris V[22].

La stratégie de modélisation moléculaire des récepteurs utilisée, consiste à partir d'un modèle construit également par B. Maigret et qui a servi dans le cadre d'une autre étude[23 et références citées]. Il s'agit d'un modèle du récepteur β_2-adrénergique construit par homologie avec le modèle de la rhodopsine qui est lui-même construit à partir de la structure 3D de la BR. L'élaboration de ce modèle de référence a été possible grâce à l'homologie de séquences qui existe entre tous les RCPGs. Toutefois, il est important de signaler que des auteurs ont montré en appliquant la construction par homologie à des protéines de structure tertiaire connue que[23, 24] :

- les protéines dont les conformations sont similaires (RMS < 2 Å), présentent un degré d'homologie supérieur à 75 %.
- Les protéines, conservant la même topologie de structures secondaires, présentent un degré d'homologie compris en 50 et 65 %.
- Les protéines, de structures complètement différentes, ont un degré d'homologie inférieur à 50 %.

En se basant encore une fois sur l'homologie de séquences présentée par cette famille de récepteurs, un modèle du récepteur CCK_1 a été construit, à partir du modèle du récepteur β_2-adrénergique dont les nombreuses données expérimentales ont permis de l'affiner et ainsi de vérifier la cohérence de la structure tridimensionnelle proposée par la modélisation. Ce modèle du récepteur CCK_1 humain, optimisé et affiné par des résultats de travaux de mutagenèse dirigée et de pharmacologie[25,26] réalisés par l'équipe de D. Fourmy (INSERM U531- Biologie et Pathologie Digestive Toulouse), nous a donc servi de référence dans la construction du modèle du récepteur CCK_1 du rat. Cette équipe de biologie et de pathologie digestive dispose, en effet, de nombreuses données expérimentales permettant de vérifier la cohérence de la structure tridimensionnelle proposée par la modélisation et, par conséquent, de conforter le modèle du récepteur obtenu.

II- MODELISATION DU RECEPTEUR CCK_1 ET PRESENTATION DU MODELE CCK_2

1) - Méthodologie utilisée dans la construction du récepteur CCK_1

La modélisation a été entièrement réalisée avec différents logiciels de la société Accelrys[27], implantés sur des stations de travail Silicon Graphics. Dans un premier temps, nous avons utilisé l'algorithme Homology interfacé avec Insight II pour positionner et aligner les séquences primaires des deux récepteurs CCK_1[28] : celle de l'humain et celle du rat qui a

été extraite de la banque de données SWISS-PROT, référencée P30551[29]. Cette procédure consiste à aligner les deux séquences primaires de manière à maximiser les acides aminés en concordance. L'alignement utilisé commence au **47^{e}** résidu pour le récepteur du rat et au **32^{e}** pour celui de l'humain. Ces alignements de séquences présentés dans la **figure 1**, ont révélé une homologie supérieure à 93%. L'étape suivante est l'affectation des coordonnées des atomes de la protéine de référence (CCK_1 humain) à ceux de la protéine de structure inconnue (CCK_1 du rat). En ce qui concerne les chaînes latérales, les coordonnées des atomes dans la protéine de référence sont conservées lorsqu'il existe une homologie avec la protéine de structure inconnue. Dans le cas d'une mutation, il y a assignation de coordonnées à tous les atomes du résidu et dans l'espace du résidu muté. Dans le cas de 'gaps' sur la protéine de référence, les coordonnées des atomes des résidus sont assignées en utilisant l'option ` Loop Search' d'Homology qui fait appel à une banque de fragments protéiniques sélectionnés dans la banque de structures cristallographiques de protéines 'Brookhaven'[7] et sélectionne la boucle qui s'adapte le mieux aux deux extrémités reliant le gap. Une fois le modèle préliminaire de la structure 3D du récepteur CCK_1 du rat obtenu et visualisé, nous avons construit le pont cystéine entre les résidus **Cys129** et **Cys211,** localisés respectivement dans la première et la deuxième boucle extracellulaire. Afin de mimer le milieu physiologique, on ajoute les protons en considérant que le pH du système est **7**. Nous avons ensuite vérifié et, au besoin, rétabli la cohérence de notre modèle avec les données expérimentales sur les protéines naturelles et qui sont : la chiralité **S** de tous les carbones α ($\mathbf{C}\alpha$) à l'exception de ceux des cystéines, la chiralité **R** des carbones β des résidus **thréonine**, la chiralité **S** des carbones β des résidus **isoleucine** et enfin la configuration *trans* de toutes les liaisons peptidiques. Étant donné la forte homologie de séquence existant entre ces récepteurs (~ 93 %), il n'a pas été nécessaire de modifier l'orientation des hélices. Les mauvaises interactions stériques entre résidus d'acides aminés, décelées au moyen de l'option stéréo disponible dans Insight II, ont été levées en modifiant manuellement (c'est à dire sans optimisation) les angles dièdres des chaînes latérales de ces résidus.

Une relaxation en trois étapes successives a ensuite été appliquée au système obtenu. La première étape consiste à optimiser sa géométrie en maintenant la chaîne peptidique fixe. Dans la deuxième étape, des contraintes sont imposées sur la chiralité et sur les angles oméga, et enfin dans la troisième étape une optimisation complète de la géométrie et sans contraintes est appliquée au système. Il est très important d'éviter des calculs globaux où tous les atomes sont libérés en même temps dès la première étape de la relaxation car cela entraînerait un risque de changement des positions relatives des régions hyper variables et des déformations

importantes au niveau des hélices α. Tous les calculs ont été effectués dans le champ de force CVFF : Consistent Valence Force Field[30]. Afin d'« écranter » les interactions électrostatiques ou en d'autres termes de tenir compte implicitement de l'environnement du récepteur inséré dans la bicouche lipidique, nous avons utilisé une constante diélectrique dépendante de la distance inter atomique. La technique de minimisation adoptée est une succession de deux algorithmes minimiseurs : la méthode de la plus grande pente suivie de celle du gradient conjugué et avec le critère de convergence suivant : le RMSEG (Root Mean Square Energy Gradient) est inférieur à 10^{-3} kcal.mol^{-1}.Å^{-1}.

Nous présentons dans les ***figures 2*** et ***3*** les structures secondaires des récepteurs CCK_1 et CCK_2, respectivement.

Figure 1 : Alignement de séquences primaires utilisé dans la construction par homologie du récepteur CCK_1 du rat à partir de l'humain. Les résidus non conservés et le gap sont indiqués respectivement, en rouge et en violet.

```
HUMAIN    QPRPSKEWQPAVQILLYSLIFLLSVLGNTLVITVLIRNKRMRTVTNIRLLSL
RAT       QPQPSKEWQSALQILLYSIIFLLSVLGNTLVITVLIRNKRMRTVTNIFLLSL

HUMAIN    AVSDLMLCLFCMPFNLIPNLLKDFIFGSAVCKTTTYFMGTSVSVSTFNLVAI
RAT       AVSDLMLCLFCMPFNLIPNLLKDFIFGSAVCKTTTYFMGTSVSVSTFNLVAI

HUMAIN    SLERYGAICKPLQSRVWQTKSHALKVIAATWCLSFTIMTPYPIYSNLVPFTK
RAT       SLERYGAICRPLQSRVWQTKSHALKVIAATWCLSFTIMTPYPIYSNLVPFTK

HUMAIN    NNNQTANMCRFLLPNDVMQQSWHTFLLLILFLIPGIVMMVAYGLISLELYQG
RAT       NNNQTANMCRFLLPSDAMQQSWQTFLLLILFLLPGIVMVVAYGLISLELYQG

HUMAIN    IKFEASQKKSAKERKPSTTSSGKYEDSDGCYLQKTRPPRKLELRQLSTGS-S
RAT       IKFDASQKKSAKEKKPSTGSSTRYEDSDGCYLQKSRPPRKLELQQLSSGSGG

HUMAIN    SRANRIRSNSSAANLMAKKRVIRMLIVIVVLFFLCWMPIFSANAWRAYDTAS
RAT       SRLNRIRSSSSAANLIAKKRVIRMLIVIVVLFFLCWMPIFSANAWRAYDTVS

HUMAIN    AERRLSGTPISFILLLSYTSSCVNPIIYCFMNKR   (376ème résidu)
RAT       AEKHLSGTPISFILLLSYTSSCVNPIIYCFMNKR   (392ème résidu)
```

Figure 2 : Structure secondaire du récepteur CCK_1 du rat[29]

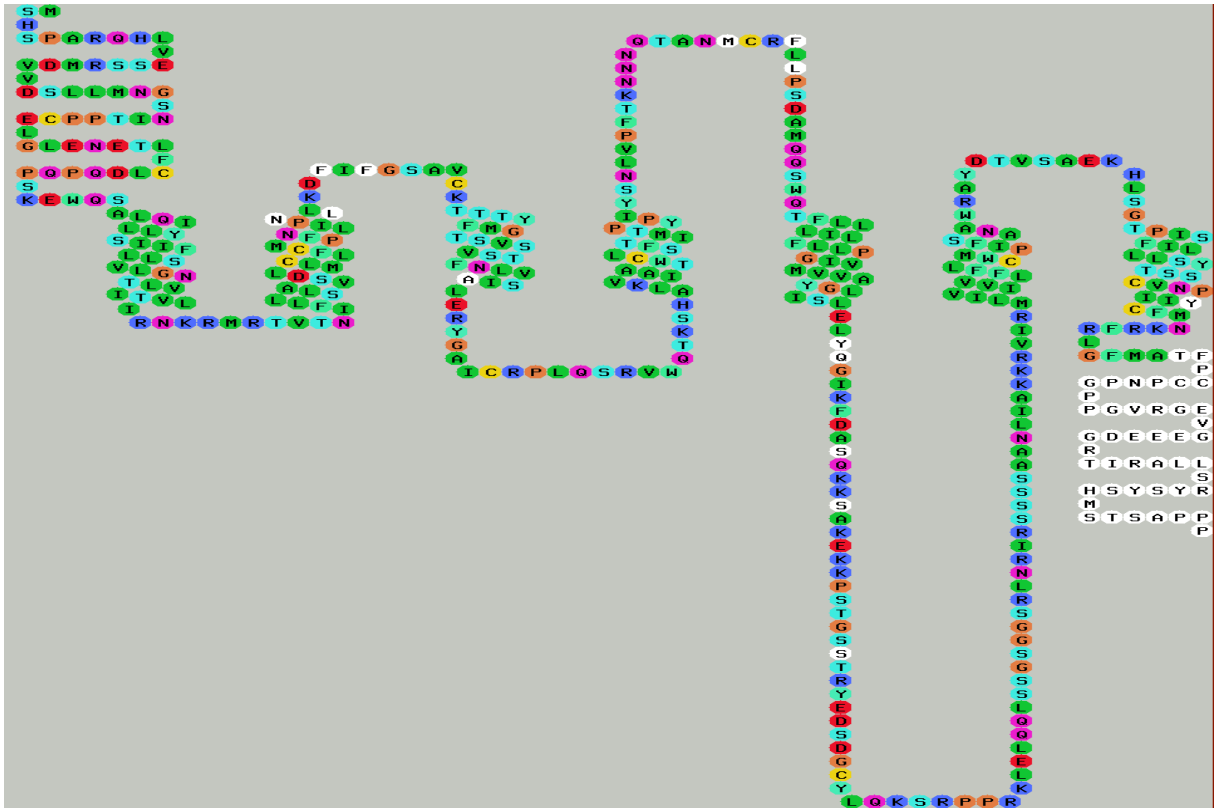
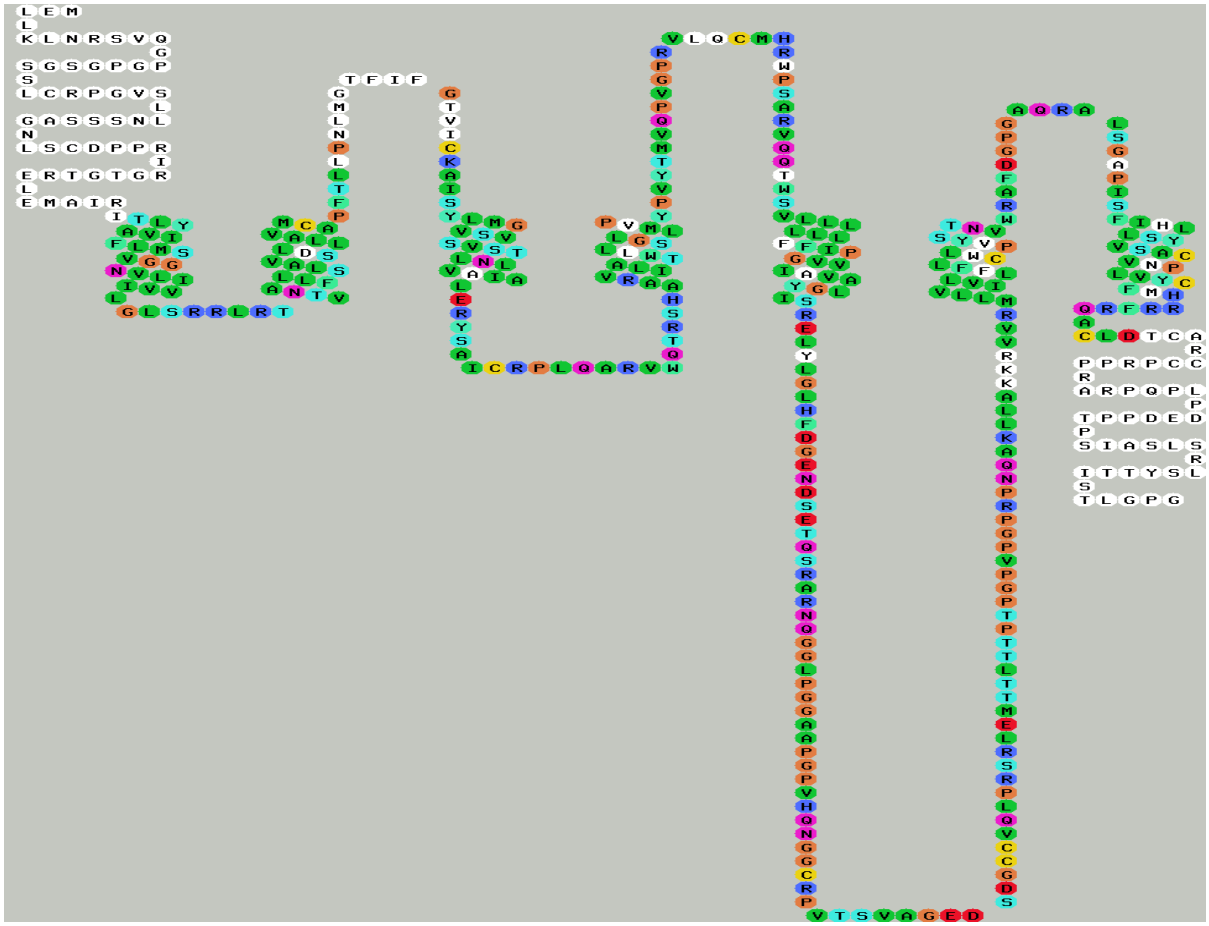


Figure 3 : Structure secondaire du récepteur CCK_2 du rat[29].



2) - Présentation des 2 modèles de récepteurs CCK

Les modèles tridimensionnels des récepteurs CCK_1 et CCK_2 sont présentés respectivement dans les ***figures 5*** et **6**. Les sept domaines transmembranaires (**TM**) prédits hélicoïdaux sont schématisés en cylindre rouge. Seul le **TMI** ne présente pas de déformations au niveau de sa structure hélicoïdale et cela dans les deux récepteurs CCK_1 et CCK_2. Les boucles extracellulaires et les deux premières boucles intracellulaires sont de longueur courte dans les deux types de récepteurs ; la plus petite contient 12 résidus et la plus longue en contient 32. En revanche, les troisièmes boucles intracellulaires sont plus longues ; celle de CCK_2 contient 105 résidus et celle de CCK_1 en contient 85. Notons que les longueurs de ces boucles compliquent considérablement leur modélisation.

Nous donnons ci-dessous (***figure 4***) la structure primaire de la séquence modélisée du récepteur CCK_2 du rat.

Figure 4 : Séquence primaire de la partie modélisée du récepteur CCK_2 du rat

PRIRGTGTRELEMAIRITLYAVIFLMSVGGNVLIIVVLGLSRRLRTVTNAFLLSLAVSDL
LLAVACMPFTLLPNLMGTFIFGTVICKAISYLMGVSVSVSTLNLVAIALERYSAICRPLQ
ARVWQTRSHAARVILATWLLSGLLMVPYPVYTMVQPVGPRVLQCMHRWPSARVQQTWSVL
LLLLLFFIPGVVIAVAYGLISRELYLGLHFDGENDSETQSRARNQGGLPGGAAPGPVHQN
GGCRPVTSVAGEDSDGCCVQLPRSRLEMTTLTTPTPGPVPGPRPNQAKLLAKKRVVRMLL
VIVLLFFLCWLPVYSVNTWRAFDGPGAQRALSGAPISFIHLLSYVSACVNPLVYC

Figure 5 : Modèle tertiaire du récepteur CCK_1 du rat. Pour des raisons de clarté, seul le squelette peptidique de la protéine est représenté. Les parties hélicoïdales sont schématisées en rouge.

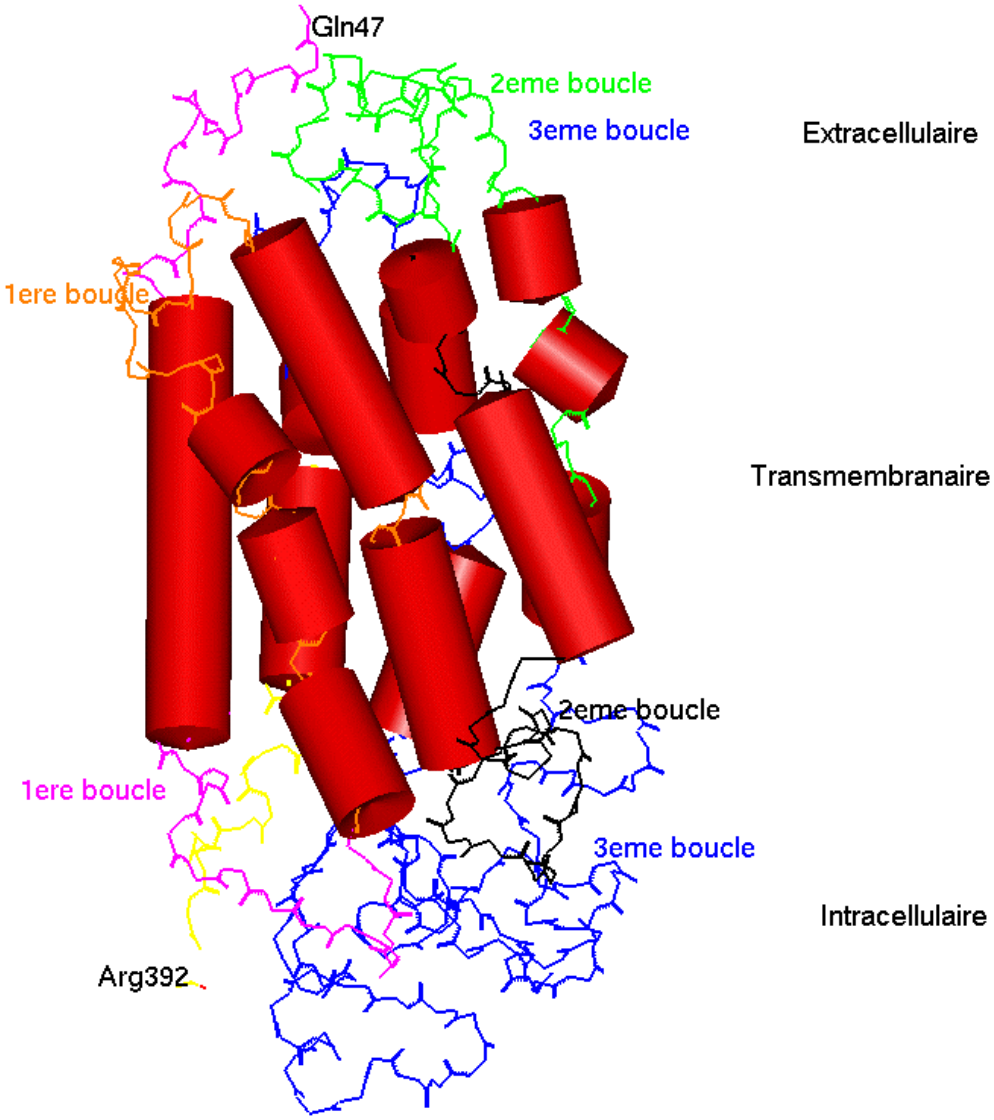
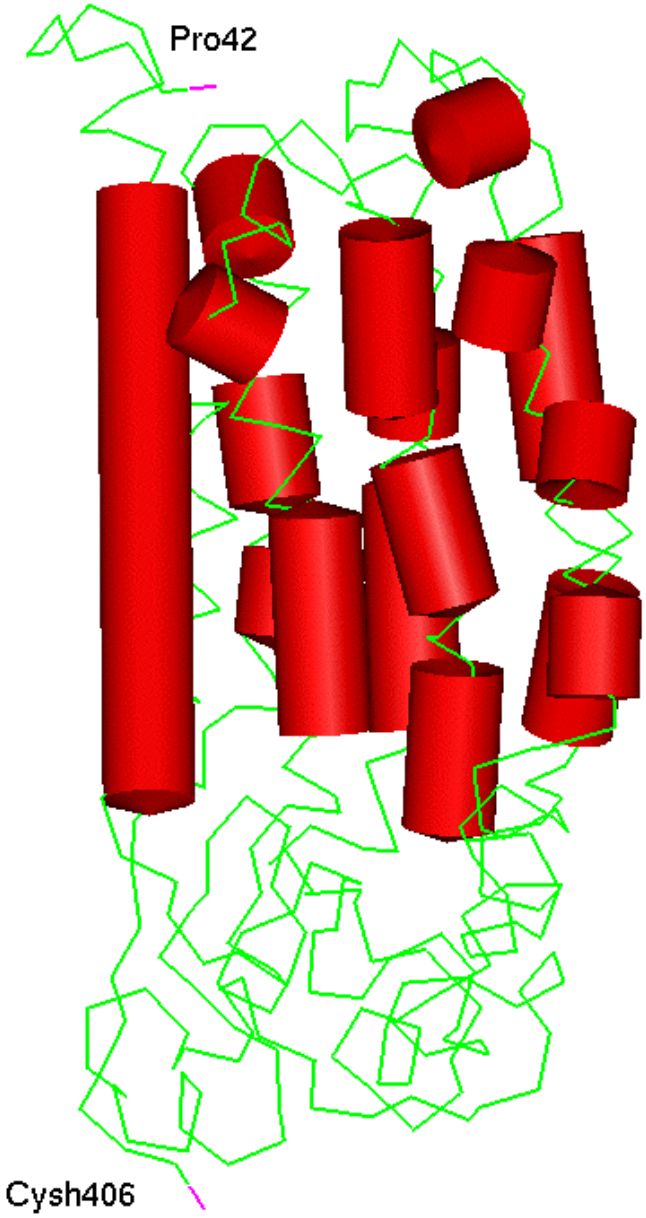


Figure 6 : Modèle tertiaire du récepteur CCK_2 du rat. Pour des raisons de clarté, seule la trace (liens entre les Carbone α) de la protéine est représentée. Les cylindres rouges représentent les domaines transmembranaires.



3) - Discussion

L'examen, au moyen de la technique stéréographique, de la structure 3D de chacun des deux modèles des récepteurs CCK_1 et CCK_2 du rat, montre qu'à l'exception de la première hélice (**TMI**), toutes les autres présentent des déformations et sont relativement plus inclinées. De tels effets, proviennent principalement des modifications conformationnelles induits par les résidus glycine et proline présents dans ces hélices (*cf* annexe 1). Notons que ces déformations se retrouvent également dans la structure RX de la rhodopsine publiée récemment[31]. La stabilité intrinsèque de ces hélices est assurée en majeure partie par des forces favorables entre les groupements aromatiques dont la distance optimale entre les centroïdes est comprise entre 5 et 6 Å. Ces forces se sont avérées importantes dans la stabilité des protéines globulaires[23,24]. Dans les deux modèles, les domaines **TMIV**, **TMV** et **TMVI** sont riches en résidus hydrophobes, alors que les autres sont constitués des deux types de résidus (polaire et apolaire). En ce qui concerne les boucles inter-hélices, nous observons de nombreux résidus chargés positivement qui sont essentiellement la lysine et l'arginine. En revanche, les résidus chargés négativement, tels que l'acide glutamique et l'acide aspartique, sont moins nombreux. Ces résidus chargés favorisent la formation de ponts salins responsables de la stabilité des boucles. Il faut noter que les résidus chargés sont plus présents dans les boucles intracellulaires. On y trouve également des résidus aromatiques dans des positions favorables à l'établissement d'interactions π-π stabilisantes avec des résidus à chaînes aromatiques des domaines transmembranaires.

En conclusion, les modèles de récepteurs dont nous disposons rendent compte des caractéristiques générales prédites par les études de profils d'hydropathie, c'est à dire que les domaines transmembranaires (**TM**) sont constitués essentiellement de résidus à caractère lipophile (ou hydrophobe) (*confer* Chapitre I).

III- MODELISATION DES COMPLEXES

1) - Introduction

La modélisation moléculaire des complexes tridimensionnels ligand/récepteur est une approche théorique, permettant de prédire les interactions mises en jeu entre les partenaires et qui peuvent ensuite être vérifiées expérimentalement. Néanmoins, l'étude de l'interaction ligand/récepteur doit toujours être précédée de la caractérisation physico-chimique des ligands et de la définition du pharmacophore.

Nous avons exposé dans le chapitre précédent une étude QSAR[21] qui nous a permis de mettre en évidence les descripteurs électroniques et stériques, gouvernant l'activité antagoniste sur les deux récepteurs de la cholécystokinine, d'une série de composés dérivant des (1,4)-benzodiazépines. Ces composés tirés de la littérature[32-36] sont tous optiquement actifs.

Les données de la littérature[37 et références citées] suggèrent des éléments communs dans les mécanismes d'activation de tous les récepteurs couplés aux protéines G. Ces données sont fondées sur des interactions critiques des agonistes au niveau des régions transmembranaires du récepteur. Elles suggèrent également que le site de liaison des ligands non peptidiques réside essentiellement dans les régions transmembranaires des récepteurs ; mais ils peuvent également impliquer des résidus situés dans les régions extracellulaires, comme il a été démontré pour le récepteur des neurokinines, NK-1[23].

Plusieurs études utilisant des approches différentes[38-44], soit la construction de récepteurs chimériques CCK_1/CCK_2 ou de récepteurs chimériques CCK_2 d'espèces différentes (chien/homme) ayant une pharmacologie divergente pour ces antagonistes, soit la mutagenèse dirigée, ont permis d'identifier des résidus, dans les 2 types de récepteurs, impliqués dans la reconnaissance et/ou la liaison sélective d'antagonistes non peptidiques. Ces études démontrent une localisation transmembranaire du site de liaison des antagonistes non peptidiques incluant les domaines **TMII**, **TMIII**, **TMV**, **TMVI** et **TMVII**.

La modélisation des sites de fixation des antagonistes non peptidiques de la cholécystokinine sur chaque type de récepteur doit être issue, en quelque sorte, de la synthèse de toutes ces informations structurales. Ensuite, elle doit se poursuivre selon les trois objectifs suivants:

- utiliser les modèles de récepteurs afin d'interpréter les données biologiques
- concevoir de nouvelles expériences permettant de vérifier ou de modifier les modèles
- générer des idées concernant la conception de nouveaux ligands ayant une meilleure activité et/ou spécificité.

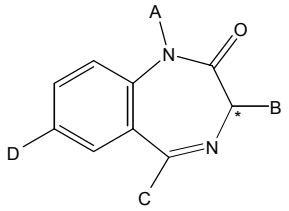
La technique du docking moléculaire (arrimage) qui est une technique d'assemblage particulière de la modélisation moléculaire permettant de construire un modèle de complexe ligand/récepteur, sera utilisée pour atteindre ces objectifs et aussi de rendre compte de certaines données biologiques.

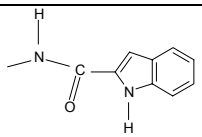
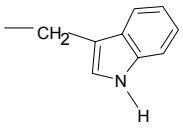
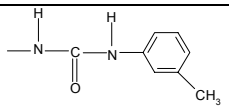
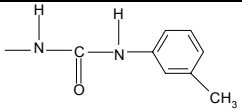
2) -Présentation des ligands sélectionnés pour le docking

Les antagonistes sélectionnés pour l'étude du docking sont présentés avec leurs activités relatives dans le **tableau 1**. La désignation des molécules est la même que celle qui est adoptée dans le chapitre précédent. Le but de la sélection est de générer de bons candidats (pertinents) pour l'étude du site de liaison ligand/récepteur. Nous avons fait cette sélection sur la base d'au moins un des critères suivants :

- activité optimale sur l'un des récepteurs (**B1** et l'un des énantiomères de **B11** et **B12)**
- bonne sélectivité vis-à-vis de l'un des récepteurs (**B1**)
- activité discriminée par la chiralité (**B11** et **B12**)
- activité minimale ou inconnue sur l'un des récepteurs (**B6**)

Tableau I : Formules et valeurs des activités antagonistes relatives des molécules dockées dans les 2 récepteurs du rat. Le carbone asymétrique est marqué par un astérisque.



Composé	A	B	C	D	C3-stereo	-logA_R (CCK_1)	-logA_R (CCK_2)
B1	CH_3			H	S R	3.544 1.630	-2.088 -3.699
B6	H			Cl	R S	-0.084 -1.057	n.d. n.d.
B11	CH_3			H	S R	1.969 -0.610	-1.229 0.799
B12	CH_3			H	S R	2.053 -0.338	-0.409 1.929

n.d. : non disponible

De façon générale, une molécule adopte plusieurs conformations préférentielles. Parmi ces conformations, il y a celles pharmacologiquement actives. L'interaction ligand/récepteur

est un phénomène de reconnaissance qui est un contact physique basé sur un principe de complémentarité de forme, similaire au système de 'clef et de serrure'. Par conséquent, nous pouvons considérer que les conformations représentatives des familles générées pour chaque composé, par l'analyse conformationnelle effectuée dans le **chapitre III**, sont de bonnes candidates pour l'étude du docking moléculaire. Du fait des restrictions conformationnelles dues à la rigidité du cycle benzodiazépine, l'analyse conformationnelle par dynamique moléculaire à haute température et recuit simulé de ces molécules a généré, au maximum, que trois familles de conformations pour chacune d'elles. La conformation représentative (la plus stable) de chaque famille est une candidate 'potentielle' pour l'étude du docking.

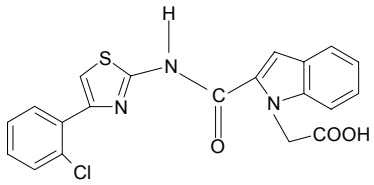
3) - Construction et optimisation des complexes

a- ligand/récepteur CCK_1

Dans le but d'introduire et de positionner chacune des conformations de l'antagoniste dans le site de fixation d'une façon la plus réaliste possible, nous nous sommes servis, comme système de référence d'un modèle 3D du complexe formé par le CCK_1 humain et le ligand antagoniste **SR-27897** (voir **figure 7**). Ce composé a été découvert par les laboratoires de Sanofi Recherche[45]. Signalons que ce modèle a été validé expérimentalement[25]. Ce type d'approche dans le docking d'un ligand a déjà été appliqué avec succès dans divers travaux[46,47].

En utilisant la technique de superposition disponible dans Insight II, nous avons donc construit les complexes entre les conformations représentatives de chaque famille des deux énantiomères de chacune des molécules **B1**, **B6**, **B11** et **B12 (Tableau 1)**. La superposition est faite sur la base de ressemblances structurales et la valeur minimale du **RMS** (critères topologiques et géométriques). L'énantiomère S de la molécule **B1** étant le plus actif, a été le premier à être positionné car il présente un fragment commun qui est le cycle indole, avec la molécule cible. L'étape suivante consiste à éliminer les mauvais contacts entre le récepteur et le ligand en modifiant manuellement la position des chaînes latérales du récepteur. La conformation spatiale de chaque complexe a été ensuite affinée par les mêmes techniques de minimisation de l'énergie, à trois étapes, utilisées précédemment dans la modélisation du récepteur. Le modèle obtenu a servi de cible dans la construction des complexes avec les autres ligands.

Figure 7 : Structure de la molécule SR-27897



b- ligand/récepteur CCK_2

Les complexes ligand/récepteur CCK_2 sont construits selon les mêmes techniques que celles utilisées dans la construction des complexes avec CCK_1. La différence est que dans ce cas, le système de référence utilisé est un modèle de complexe construit avec le peptide CCK-9, un agoniste endogène (*cf* **chapitre I**). Ce modèle de référence a été construit antérieurement par B. Maigret et a été affiné sur la base de données expérimentales[42]. Nous avons construit les complexes en tenant compte du fait que les sites de liaison **agoniste** et **antagoniste** des récepteurs CCK **se recouvrent** partiellement[42].

4) - Explorations des sites de liaison ligand/récepteur CCK_1

Dans chaque complexe optimisé, nous avons recherché à l'aide des techniques infographiques et de l'option stéréo disponibles dans Insight II, les interactions susceptibles de stabiliser le ligand dans le site de fixation proposé. Étant donné que l'étude QSAR, a révélé le rôle important joué par le substituant **lié au carbone asymétrique** (noté **B**, **tableau 1**), notre attention s'est portée en premier lieu sur l'environnement, au niveau du récepteur, de ce substituant. Rappelons que les relations de structure activité ont suggéré comme substituant optimal celui qui serait neutre et polaire.

Les complexes construits avec les paires d'énantiomères du composé **B1** (**Tableau 1**) qui remplit bien ces requis électroniques, ont été largement analysés au niveau du site d'interaction ligand/récepteur. Les résultats diffèrent selon la conformation et selon l'énantiomère. En ce qui concerne l'énantiomère S, nous observons des interactions de type liaison hydrogène entre le substituant lié au carbone asymétrique (noté **B**, **Tableau 1**) et l'oxygène peptidique de la serine 358 (**Ser358**) qui se trouve à l'interface de l'hélice 6 (**TMVI**) et de la troisième boucle extracellulaire (**e3**). Cependant, la liaison la plus forte (**2.11Å**) est celle qui s'établit avec le groupement indole. Nous présentons dans la **figure 8**, ces liaisons hydrogène. Notons aussi que des interactions stabilisantes de type liaison hydrogène et hydrophobe sont observées particulièrement avec la conformation la plus stable

parmi celles que nous avons positionnées dans le récepteur. Des interactions de type hydrophobe s'établissent entre **Phe124** et **l'indole** : les cycles aromatiques correspondant sont perpendiculaires entre eux (position en 'T'), avec une distance entre les centroïdes de l'ordre de 5 Å ; ce qui confère au système une certaine stabilité. Le résidu **Phe124** est situé dans la première boucle extracellulaire (**e1**). Le résidu **Leu61** localisé au début de l'hélice 1 (**TMI**), est perpendiculaire au cycle aromatique du groupement lié au **carbone C_4 du cycle benzodiazépine** (noté **C**, **Tableau 1**) ; ce qui est responsable d'une autre interaction hydrophobe mais moins importante. Dans le complexe formé avec l'énantiomère **R** (moins actif que le **S**, **Tableau 1**), les interactions stabilisantes ne mettent pas en jeu les mêmes atomes, aussi bien au niveau du récepteur que du côté du ligand. Ceci est dû au fait que les 2 énantiomères s'échangent le positionnement 'initial' des groupements **A** et **C** au sein du récepteur. Nous avons observé, en effet, une liaison hydrogène entre l'oxygène carboxamide du groupement **B** et l'hydroxyle de la **Ser358**. Cette liaison hydrogène semble plus **forte** que celle observée dans le complexe avec l'énantiomère **S**. Le site d'interaction est illustré dans la **figure 9.** L'oxygène peptidique du résidu **Ser358** se trouve distant de plus de **3.11 Å** des hydrogènes du ligand, susceptibles d'établir des interactions ; ce qui expliquerait leurs absences. L'environnement de nature hydrophobe est moins important dans ce complexe (formé avec l'énantiomère **R** de **B1**) puisque tous les résidus aromatiques ou aliphatiques ne sont pas dans un voisinage appréciable, pouvant rendre compte d'une poche lipophile. Le plus proche est le Trp54, distant à plus de 5 Å, il est localisé dans la partie supérieure de l'hélice 3 (**TMI**), mais son orientation ne permet pas une interaction π-π stabilisante.

Les complexes construits avec la molécule **B6** (**Tableau 1**) ont mis en évidence un autre site de fixation. En effet, nous avons observé des interactions par liaison hydrogène avec les résidus **Trp54** et **Thr132** (**TMIII**) dans les complexes construits, respectivement, avec les énantiomères **S** et **R**. Ces interactions sont cependant faibles. Les résidus aromatiques qui jouxtent de façon parallèle l'**indole** sont le **Phe213** dans le cas de l'énantiomère **S** et le **Phe135** dans l'autre énantiomère. Notons que ces résidus sont localisés, respectivement, dans la deuxième boucle extracellulaire (**e2**) et dans l'hélice 3 (**TMIII**). Le groupement **C** qui est un cycle benzénique est positionné de façon parallèle à la chaîne latérale du résidu **Thr132** dans le cas de l'énantiomère **S** (moins actif que R) et de façon perpendiculaire au cycle aromatique de **Phe122** dans le cas de **R**.

L'exploration des sites de fixation proposés pour le composé **B11** (**Tableau 1**) a mis en évidence des interactions potentielles de type liaison hydrogène entre le groupement lié au carbone asymétrique (noté **B**) et d'une part, le résidu **Ser358** (interface **TMVI** et **e3**) et,

d'autre part, l'**Arg352** (**TMVI**). Ces interactions sont observées essentiellement dans les complexes formés avec l'énantiomère **S** qui est le plus actif (Tableau 1). En revanche, les complexes formés avec l'énantiomère **R**, indiquent que le groupement **B** se positionne dans un environnement hydrophobe dû, en particulier, aux résidus **Phe213** et **Trp54**. On note l'absence de liaisons hydrogène entre le fragment **B** et le résidu **Ser358.** Les résultats du site d'interaction concernant l'énantiomère **S** de **B11** sont illustrés dans la ***figure 10***.

L'analyse des complexes modélisés et concernant les énantiomères du composé **B12** (**Tableau 1**), a mis en évidence des interactions avec d'autres partenaires au sein du récepteur, mais toujours avec le groupement **B** du ligand. Nous avons relevé une interaction de type liaison hydrogène de longueur 2.24 Å avec la **Thr133** (**TMIII**) dans le complexe formé avec l'énantiomère **S**. Des interactions **π-π** (// ou **T**) comprises entre 5.97 Å et 7.93 Å sont présentes essentiellement dans les complexes formés avec l'énantiomère **R**. Ces interactions concernent les cycles aromatiques du groupement benzodiazépine, du fragment **C** et du groupement **B**. Notons que la plus importante parmi ces interaction en 'stacking' est celle formée avec le substituant **B**.

Les figures reportées ci-dessous concernent essentiellement les résultats pertinents apparaissant dans nos modèles. Dans ces figures, les atomes d'azote et d'oxygène sont respectivement représentés en bleu et en rouge.

Figure 8 : Vue longitudinale du modèle moléculaire du site de fixation de l'énantiomère **S** du composé **B1** dans le récepteur CCK_1 du rat. Les atomes de carbone du ligand sont représentés en vert et ceux de la serine en orange. La trace du récepteur est représentée en ruban gris

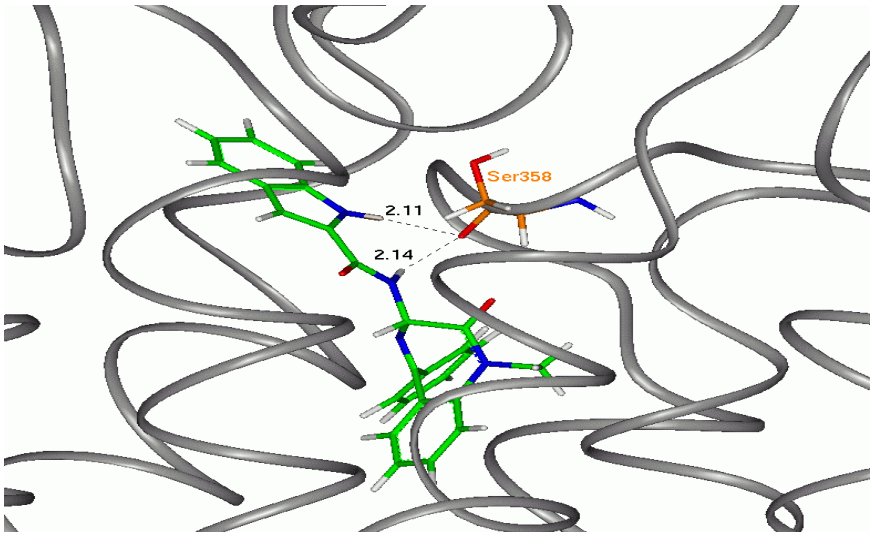


Figure 9: Vue longitudinale du modèle moléculaire du site de fixation de l'énantiomère **R** du composé B1 dans le récepteur CCK_1. Les atomes de carbone du ligand sont représentés en vert et ceux de la serine en orange. La trace du récepteur est représentée en ruban gris.

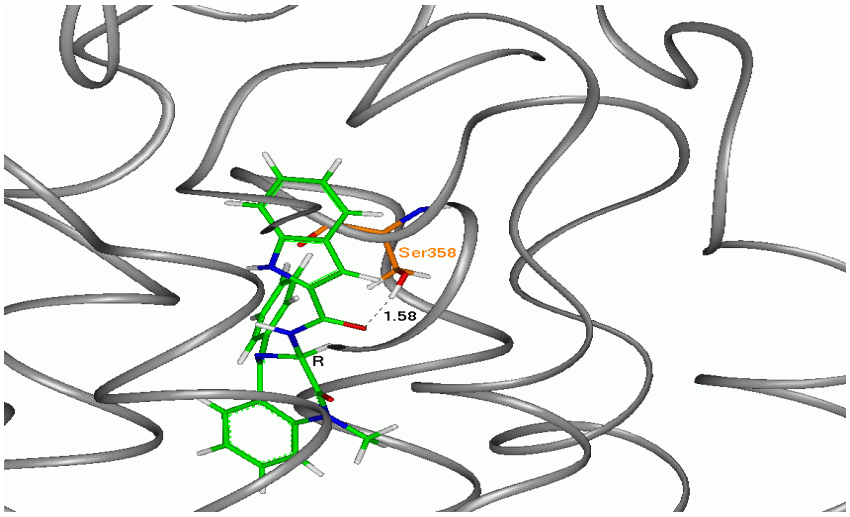
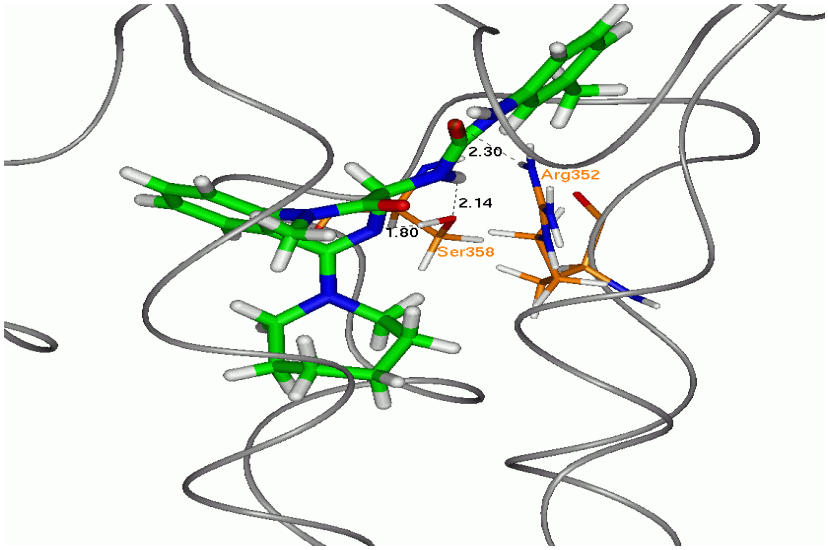


Figure 10 : Vue longitudinale du site d'interaction de l'énantiomère S de B11 dans le récepteur CCK_1. Le ligand est représenté en vert et les résidus en marron.



5) - Explorations des sites de liaison ligand/récepteur CCK_2

En utilisant les mêmes techniques que précédemment, nous avons examiné les sites de fixation des antagonistes dans les complexes optimisés construits avec CCK_2. Etant donné l'importance suggérée par notre étude QSAR du substituant **B**, **lié au carbone asymétrique** (**Tableau 1**), sur l'activité du composé, notre attention s'est portée encore une fois sur ce fragment. Nos résultats de relations structure-affinité ont montré la nécessité d'un substituant (B) ayant un pouvoir attracteur pour conférer une bonne activité antagoniste à la molécule sur le récepteur CCK_2. De plus, ces résultats QSAR suggèrent l'existence de forces de type van der Waals au niveau du site d'interaction ligand/récepteur.

Les complexes construits avec la molécule **B1** présentent des interactions par pont hydrogène entre le fragment **B** et le résidu **Ile130** situé dans la partie supérieure de l'hélice 3 (**TMIII**) (***figures 11*** **et** ***12***). Il faut noter que celle observée dans le complexe construit avec l'énantiomère **R**, est légèrement plus importante de quelques dixièmes d'angströms. Le fragment **B** du ligand se trouve également stabilisé par des interactions électroniques de type **π-π** avec le résidu **Phe120** (**e1**). Le groupement **C** de l'énantiomère **R** est, de plus, stabilisé par des interactions de même type **π-π** avec le résidu **Tyr189** situé dans la partie supérieure du **TMIV**. Ces interactions semblent être néfastes pour l'activité du composé puisque les 2 énantiomères sont peu actifs.

Les complexes construits avec les conformations préférentielles du composé **B11** ont révélé aucune interaction stabilisante par liaison hydrogène. Cependant, des interactions électroniques de type **π-π** stabilisantes sont observées, en particulier, avec l'énantiomère **S** (moins actif que R, **tableau 1**). La plus favorable est celle qui s'établit entre le fragment **B** et le **Phe120** (**e1**). Les résidus **His381** (sommet du **TMVII**) et la **Tyr385** (milieu du **TMVII**) semblent interagir de façon répulsive avec le fragment **C, lié au carbone C_5 de la benzodiazépine (Tableau 1)**. Notre modèle semble indiquer une gène stérique 'locale' au sein du récepteur. Dans les complexes formés avec l'énantiomère **R**, nous observons une interaction **π-π** entre le groupement **B** et le résidu **His207** (**TMV**) : la distance entre les centroïdes est de 5.01 Å, qui représente une distance optimale pour ce type d'interaction.

L'analyse des sites d'interaction du ligand **B12** et du récepteur CCK_2 révèle une absence de liaisons hydrogène dans les complexes construits avec l'énantiomère **R** (plus actif que S). Cependant, on note la proximité du **Phe120** (**e1**) dans l'environnement du groupement **B**. Nous avons noté une liaison hydrogène relativement faible dans les complexes construits avec l'énantiomère **S**. Cette interaction s'établit entre le cycle **C** et la chaîne latérale du résidu

Arg361 (début de **e3**). Une interaction de type **π-π** entre la **Phe120** (**e1**) et le fragment **B** a été également notée dans ces complexes (***figure 13***).

L'examen des sites d'interaction ligand/récepteur dans les complexes construits avec la molécule **B6** dont l'activité n'est pas encore connue, a révélé aucune interaction pertinente entre les partenaires. Ceci pourrait signifier que ce composé ne semble pas influer sur la conformation (ou dynamique) du récepteur. Notre modèle le prédit donc **inactif**. Nous reportons ci-dessous que les illustrations relatives au docking des composés B1 et B12 (***figures 11-12***).

Figure 11 : Vue longitudinale du site de fixation de l'énantiomère S du composé B1 dans le récepteur CCK_2. La trace du récepteur est représentée en ruban.

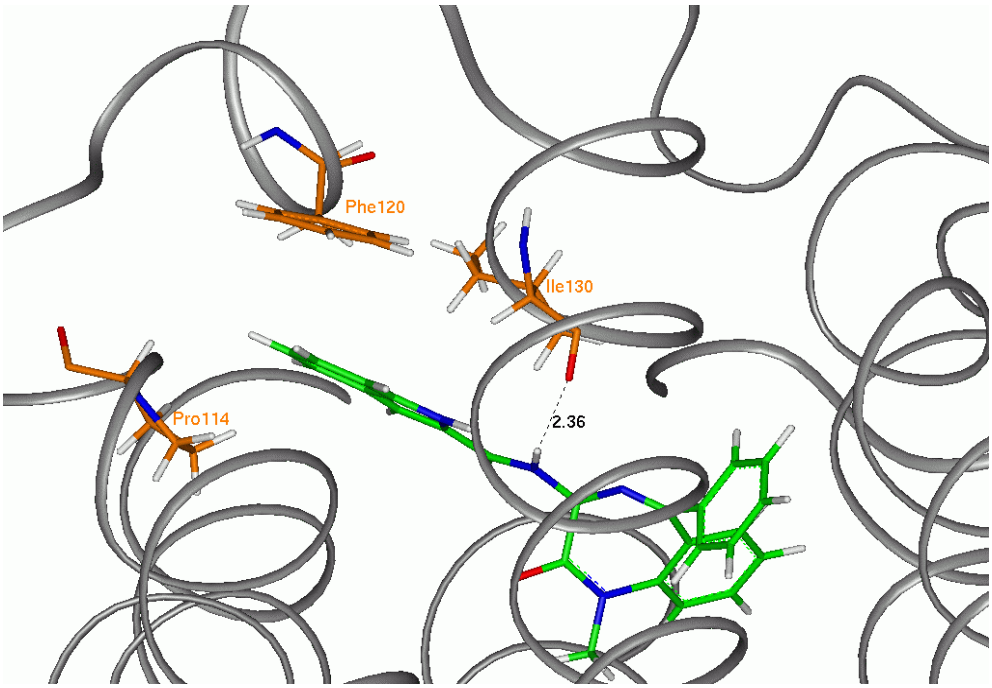


Figure 12 : Vue longitudinale du site de fixation de l'énantiomère R du composé B1 dans le récepteur CCK_2. La trace du récepteur est représentée en ruban.

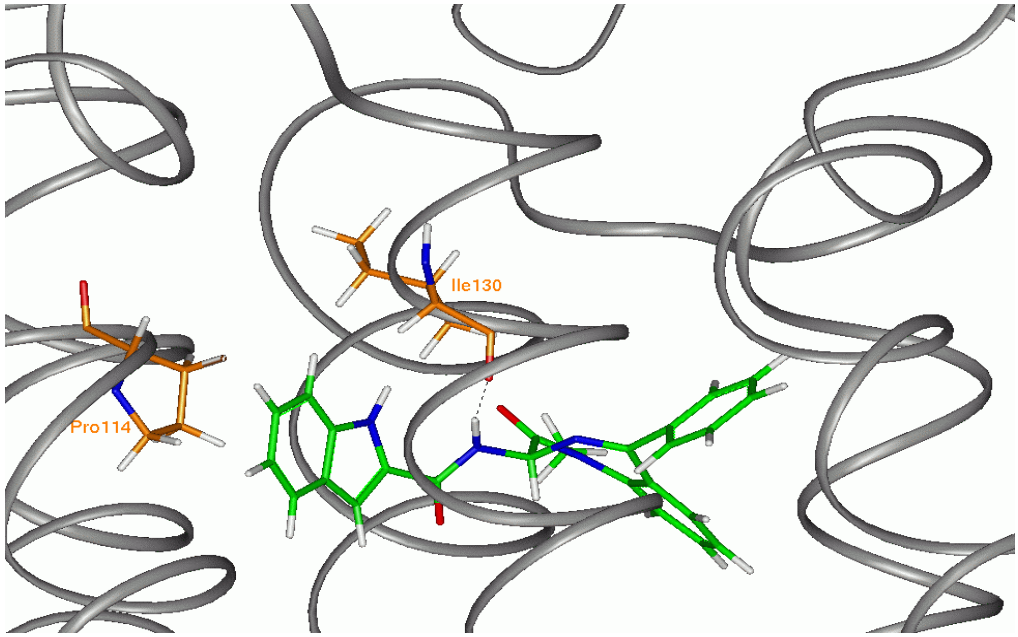
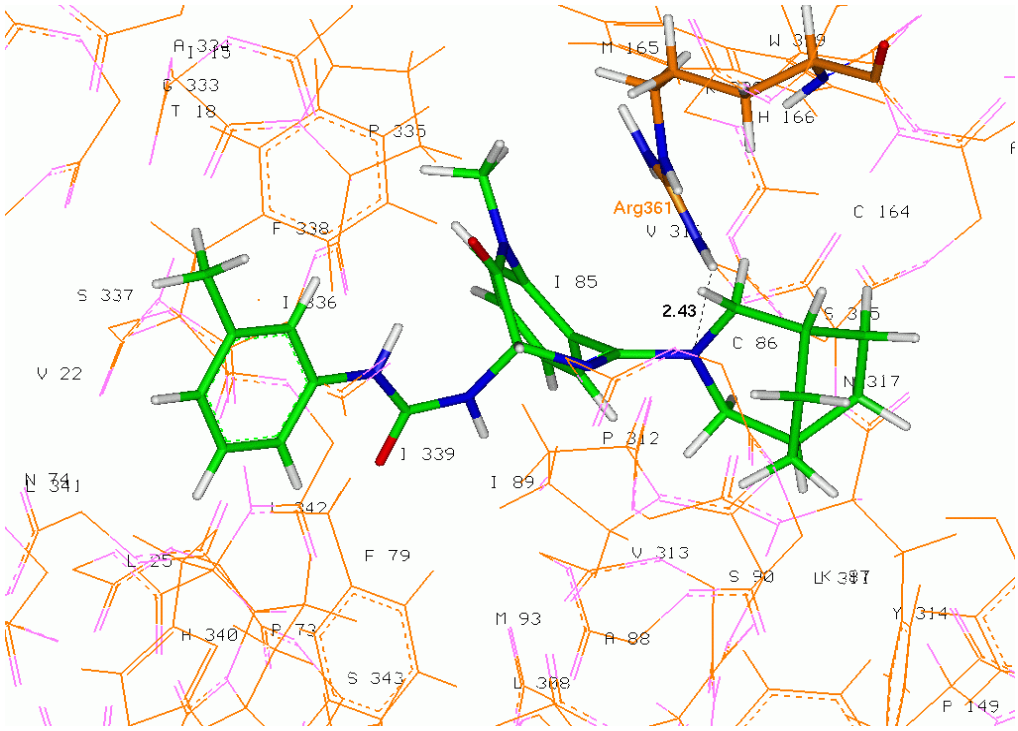


Figure 13 : Vue longitudinale du site de fixation de l'énantiomère S de B12. Le squelette peptidique, les chaînes latérales des résidus aromatiques et de l'Arg361 sont représentés en marron. Les résidus autour du ligand sont indiqués par le code à 1 lettre. La chaîne carbonée du ligand est en vert.



6) - Discussion

Nous avons évalué, pour chaque système ligand/récepteur, l'énergie de complexation. Les valeurs sont répertoriées dans le tableau II. Les résultats d'analyse des sites d'interaction ligand/récepteur sont résumés dans le tableau III.Les résultats décrits dans la littérature y sont également répertoriés.

Tableau II : Energies (en kcal/mol) : du ligand libre, du complexe et de complexation.
Avec $E_{complexation} = E_{complexe} -(E_{ligand(libre)} + E_{récepteur(libre)})$

Complexe	$E_{ligand(libre)}$	$E_{complexe}$	$E_{complexation}$	$-\log A_R$
B1S/CCK_1	206.997	3019.104	-54.664	3.544
B1S/CCK_2	**206.997**	**5127.143**	**-43.846**	**-2.088**
B1R/CCK_1	206.997	3031.446	-42.322	1.630
B1R/CCK_2	**206.997**	**5137.549**	**-33.440**	**-3.699**
B6S/CCK_1	166.693	2988.903	-44.561	-1.057
B6S/CCK_2	**166.693**	**5099.150**	**-31.535**	**n. d.**
B6R/CCK_1	166.693	3019.697	-13.767	-0.084
B6R/CCK_2	**166.693**	**5095.210**	**-35.475**	**n. d.**
B11S/CCK_1	135.777	2985.915	-16.633	1.969
B11S/CCK_2	**135.777**	**5067.345**	**-32.424**	**-1.229**
B11R/CCK_1	135.777	2974.528	-28.290	-0.610
B11R/CCK_2	**135.777**	**5063.071**	**-36.698**	**0.799**
B12S/CCK_1	175.518	3002.398	-39.891	2.053
B12S/CCK_2	**175.518**	**5096.746**	**-42.764**	**-0.409**
B12R/CCK_1	175.518	2984.546	-57.743	-0.338
B12R/CCK_2	**175.518**	**5096.479**	**-43.031**	**1.929**

R et S indiquent le type d'énantiomère du ligand.
Energie du récepteur CCK_1 libre = 2866.771 kcal/mol
Energie du récepteur CCK_2 libre = 4963.992 kcal/mol

Plusieurs 'prérequis' pour l'interaction avec le récepteur sont devenus apparents à l'issue de l'analyse des résultats QSAR[21]. Ils ont permis de refléter la nature des interactions importantes entre le ligand et chaque type de récepteur CCK. L'étude du docking moléculaire a permis de montrer que les zones polaires représentées dans le pharmacophore par des hétéroatomes d'oxygène et/ou d'azote, peuvent effectivement interagir par liaisons hydrogène

avec certains résidus du récepteur. Cependant, ces interactions ne sont pas systématiques et ne mettent pas toujours en jeu les mêmes éléments chimiques. Ceci est dû au fait que pour chaque structure nouvelle du ligand et/ou conformation, les chaînes latérales des résidus ont la possibilité de s'orienter afin de permettre des interactions optimales avec l'antagoniste. Les valeurs des énergies de complexation indiquent que le phénomène de complexation est énergétiquement favorable.

Les modèles moléculaires obtenus après docking de ces antagonistes dérivés des benzodiazépines, révèlent une orientation spatiale différente de ces ligands au sein du récepteur.

Le résidu **Arg352** qui intervient dans le site de fixation de **B11**, une molécule relativement active (Tableau 1), est l'homologue du résidu **Arg336** dans le récepteur humain et dont l'implication dans le site de liaison de **SR-27897** (***figure 7***), a été mise en évidence dans des études récentes[37, 48]. Il serait donc intéressant de voir au moyen d'expériences de mutagenèse dirigée si ce site de reconnaissance est général. Il est également intéressant de noter que les résidus qui semblent jouer un rôle important selon notre étude, appartiennent de façon générale aux mêmes domaines transmembranaires que ceux déjà décrits dans la littérature[16, 43]. En effet, dans leurs travaux de docking du devazepide, qui correspond à B1 dans notre étude, van der Bent et al[16] ont mis en évidence d'autres résidus : **Asn349** (**TMVI**), **Phe112** (**TMII**), **Tyr376** (**TMVII**), **Ser139** (**TMIII**), **Ser375** (**TMVII**) et **Ser379** (**TMVII**) qui n'interviennent pas dans notre étude et dont certains ont fait l'objet de mutations[43]. Ces dernières concernent **S139A**, **N349A** et **S379A** et affaiblissent fortement l'affinité du S-devazepide. L'examen de la position de ces résidus dans notre modèle montre qu'ils sont à proximité du site de fixation proposé. Leur modèle de complexe diffère du nôtre dans la mesure où eux, d'une part, ils ont docké la structure du S-devazepide, correspondant à l'énantiomère **S** de la molécule **B1** dans notre étude (**Tableau 1**), obtenue par rayons X et, d'autre part, ils n'ont pas inclus dans leur modèle le pont cystéine, les boucles intracellulaires et extracellulaires. En fait, ils se sont limités aux sept domaines transmembranaires dont la modélisation est évidemment sensible à l'environnement, en l'occurrence les boucles extracellulaires qui sont très proches du site de fixation. Selon notre modèle, ces dernières semblent effectivement contenir des partenaires pertinents dans l'interaction avec l'antagoniste de type benzodiazépine.

La paire d'énantiomères se distingue au sein du récepteur par le positionnement des substituants **C** et **A**. En général, avec les conformations des énantiomères **R**, le groupement **C** semble occuper une position dans un espace trop restreint pour accommoder son grand

volume. Ceci 'secouerait' fortement le récepteur et empêcherait le ligand de le bloquer dans sa conformation inactive. Il est bien connu que, par sa liaison au récepteur, un antagoniste stabilise ce récepteur dans une conformation inactive. En revanche, la liaison d'une molécule agoniste à ce même récepteur stabilise la protéine réceptrice dans une conformation active.

De l'étude de ce site de liaison ligand/CCK_1, nous proposons des expériences de mutations des résidus **Ser358, Arg352, Leu61, Phe122, Phe213 et Phe135**. Les deux premiers résidus interviennent dans des interactions de type liaison hydrogène avec des antagonistes sélectifs CCK_1. Il serait également intéressant de voir l'effet de la mutation du résidu **Thr132** sur l'affinité de la molécule **B6** qui est peu active sur le type sauvage.

En ce qui concerne les sites d'interaction ligand/CCK_2, notre étude a mis en évidence 3 résidus déjà décrits dans la littérature : **His381** (**TMVII**), **Phe120** (**e1**) et **His207** (**e2**). Cependant, les deux derniers résidus ont été décrits dans le site de fixation de l'agoniste[49]. Étant donné que le résidu **Phe120** intervient, dans notre étude, avec des molécules faiblement actives (**B1** et **B11S**) sur CCK_2, on pourrait attribuer leurs faibles activités antagonistes aux interactions avec ces résidus, induisant la conformation active du récepteur. En revanche, il serait intéressant de voir l'influence de la mutation du résidu **His207** par une valine par exemple, sur l'affinité de **B11R**. D'autre part, l'intervention de **His381** dans une interaction répulsive dans **B11S**, un bon antagoniste sur CCK_1, semble en accord avec l'étude qui décrit ce résidu comme responsable de la sélectivité CCK_1/CCK_2[42].

Il serait intéressant de voir l'effet de la mutation de **Arg361** sur l'affinité de la molécule **B12**.

La modélisation des boucles extracellulaires reliant les hélices, a attiré notre attention sur des résidus supplémentaires qui semblent participer au site de fixation du ligand dans chacun des récepteurs et qui peuvent ainsi être proposés pour de futures expériences de mutagenèse dirigée.

Cette étude de modélisation des complexes avec le récepteur CCK_2 constitue une suite à celle initiée par B. Maigret dès 1996[41]. A notre connaissance, il n'existe pas d'autres études théoriques sur ce type de récepteur.

Les modèles structuraux présentés ici, devraient être seulement des 'modèles de travail' qui peuvent être utilisés pour proposer plusieurs expériences de mutagenèse dirigée dont les résultats pourraient, à leur tour, être utiles pour affiner ces modèles tridimensionnels.

Tableau III : Récapitulation des résidus importants dans les 2 sites d'interaction : Ligand /CCK_1 et ligand/CCK_2.

Récepteur	Modèle obtenu	Modèle de littérature[16]	Expérience
CCK_1	**S**358(**TMVI,e3**), **F**124(**e1**), **L**61(**TMI**), **F**213(**e2**), **F**135(**TMIII**), **F**122(**e1**), **R**352(**TMVI**), **T**133(**TMIII**)	**N**349(**TMVI**), **F**112(**TMII**), **Y**376(**TMVII**), **S**139(**TMIII**), **S**375(**TMVII**), **S**379(**TMVII**)	**S**139**A**[43] **N**349**A**[43] **S**379**A**[43] **L**372[a], R352[b]
CCK_2	**I**130(**TMIII**), **F**120(**e1**), **Y**189(**TMIV**), **H**381(**TMVII**), **Y**385(**TMVII**), **H**207(**TMV**), **R**361(**e3**)		**V**353**L**[40] **H**381**L**[41]

[a] : la mutation de son homologue dans le récepteur CCK_1 humain a montré son implication dans l'affinité de l'agoniste[50] : L356A.
[b] : l'importance de son homologue R336 dans l'humain a été mise en évidence[37,48].

IV – CONCLUSIONS

L'étude exposée dans ce chapitre est fondée sur de récents développements dans le domaine de la chimie médicinale et de la biologie moléculaire des ligands et des deux types de récepteurs de la cholécystokinine. Ces données ont servi, en particulier, dans l'élaboration de modèles des récepteurs et de complexes. Notre approche est également basée sur une comparaison des propriétés conformationnelles des ligands ayant une forte et une faible activité. Nous avons donc analysé au préalable (*confer* **chapitre III**) les caractéristiques conformationnelles de ces ligands.

Nos résultats, concernant les sites de fixation des antagonistes de type benzodiazépine sur chacun des récepteurs CCK, ne sont pas en accord complet avec les quelques résultats expérimentaux et/ou théoriques disponibles dans la littérature. En effet, notre étude a mis en évidence de nouveaux résidus susceptibles d'interagir avec le ligand et dont la pertinence mérite d'être vérifiée par des expériences de mutagenèse dirigée. Des interactions par liaison hydrogène entre le fragment **B** qui est **lié au carbone asymétrique de l'antagoniste** et des résidus polaires représenteraient les principales forces de stabilisation du complexe, et en particulier avec le récepteur CCK_1. Un cluster de résidus aromatiques au voisinage de ce groupement serait favorable à l'affinité du composé sur les 2 récepteurs de la CCK.

Des interactions par liaison hydrogène avec le groupement lié au carbone C_4 du cycle benzodiazépine (noté **C**) semblent en faveur de la sélectivité sur le récepteur CCK_2. Notre

modèle révèle donc que l'incorporation d'hétéro atomes dans le substituant **C** améliorerait l'affinité du composé sur CCK_2.

Cependant, il ne faut pas perdre de vue les limitations actuelles du modèle telle que la non prise en compte explicite du milieu membranaire. En effet, les résultats des expériences de mutagenèse dirigée devraient affiner les modèles proposés dans cette étude. Ce qui nous permettra ainsi de mieux percevoir les caractéristiques du site de fixation des antagonistes de la cholécystokinine et de proposer au chimiste, la synthèse de molécule au profil pharmacologique choisi.

A plus long terme, en fonction des premiers résultats de mutagenèse dirigée, il est possible d'envisager des mutations 'positives'. C'est à dire qu'au lieu de rechercher la mutation qui réduit l'affinité, nous serions en mesure grâce au modèle tridimensionnel, de proposer des positions et des résidus qui pourraient augmenter, ou au moins restaurer, l'affinité du ligand pour le récepteur CCK mutant. Cette démarche est certainement la plus sûre expérience de validation d'un modèle tridimensionnel, mais nécessite une connaissance précise du site de fixation.

BIBLIOGRAPHIE

1- Teeter, M. M., Froimowitz, M., Stec, B. et DuRand, C. J., *J. Med. Chem.*, 37, 2874-2888, **1994**.

2- Schertler, G. F., Villa, C. et Henderson, R., *Nature*, 262, 770-772, **1993**.

3- Bockaert, J. et Pin, J. P., *EMBO (Eur. Mol. Biol. Organ.) J.,* 18, 1723-1729, **1999.**

4- Popot, J. L., La recherche, N° 192, Vol. 18, pp : 1170-1181, **1987**.

5- Henderson, R., Baldwin, J. M., Ceska T. A, Zemlin, F., Beckmann, E. et Dowing, K. H., *J. Mol. Biol.* 213, 899-929, **1990**.

6- Findlay, J. et Eliopoulos, E., *Trends Pharmacol. Sci.* 11, 492-499, **1990**.

7- Protein Data Bank. Chemistry Department, Bldg.555, Brookhaven National Laboratory, P.O. Box 5000,Upton, N. Y. 11973-5000, USA.

8- Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels, A. et Hibert, M., *J. Med. Chem.*, 35, 3448-3462, **1992**.

9- Ijzerman, A. P., Van Galen, P. J. M. et Jacobson, K. A., *Drug Design Discov.*, 9, 49-67, **1992**.

10- Maloney-Huss, K. et Lybrand, T. P., *J. Mol. Biol.*, 225, 859-871, **1992**.

11- Hibert, M., Trumpp-Kallmeyer, S., Bruinvels, A. et Hoflack, J., *Mol. Pharmacol.*, 40, 8-15, **1991**.

12- Livingstone, C. D., Strange, P. G. et Naylor, L. H., *Biochem. J.*, 287, 277-282, **1992**.

13- Yamamoto, Y., Kamiya, K. et Terao, S., *J. Med. Chem.*, 36, 820-825, **1993**.

14- Grötzinger, J., Engels, M., Jacoby, E., Wollmer, A. et Strabburger, W., *Protein Eng.*, 4, 767-771, **1991**.

15- Baldwin, J., *EMBO J.*, 12, 1693-1703, **1993**.

16- Van der Bent, A., Ijzerman, A. P. et Soudijn, W., *Drug Des Discov*., 12(2),129-148, **1994**.

17- Hoflack, J., Trends Pharm. Sci., 15, 7-9, **1994**.

18- Dahl, S. G., Edvarsen et Sylte, I., *Proc. Natl. Acad. Sci.* USA, 88, 8111-8115, **1991**.

19- Maloney-Huss, K. et Lybrand, T. P., *J. Mol. Biol.*, 225, 859-871, **1992**.

20- Zhang, D. et Weinstein, H., *J. Med. Chem.*, 36, 934-938, **1993**.

21- Taïri-Kellou, S., Cartier, A., Maouche, B. et Maigret, B., *J. of Mol. Struct. (THEOCHEM)*, 571/1-3, 207-223, **2001**.

22- Jagerschmidt, A., Guillaume-Rousselet, N., Vikland, M. L., Goudreau, N., Maigret, B. et Roques B. P., *Eur. J. Pharmacol.*, 296(1), 97-106, **1996**.

23- Joseph, M. P., Maigret, B., Bonnafous, J.-C., Marie, J. et Scheraga, H. A., Protein Chem., 14(5), 381-398, **1995**.

24- Luc Quéré, Thèse de Doctorat en Sciences Chimiques, Faculté des Sciences-Belgique, **1995**.

25- Kennedy, K., Gigoux, V., Escrieut, C., Maigret, B., Martinez, J., Moroder, L., Gully, D., Vaysse, N. et Fourmy, D., *J. Biol. Chem.*, 272, 2920-2926, **1997**.

26- Gigoux, V., Escrieut C., Silvente-Poirot, S., Maigret, B., Gouilleux, L., Fehrentz, J. A., Gully, D., Moroder, L., Vaysse, N. et Fourmy, D., *J. Biol. Chem.* 273, 14380-14386, **1998**.

27- Accelrys Technologies, 9685 Scranton Road, San Diego, CA92121-2777,USA.

28- Lee, R. H., *Nature*, 356, 543-544, **1992**.

29- http://www.expasy.ch/cgi-bin

30- Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M. et Hagler, A. T. , Proteins :Structure, Function and Genetics, 4(31), **1988**.

31- Palczewski, K., Kumasaka, T., Hori, T.,Behnke, CA., Motoshima, H., Fox, BA, Le Trong, I., Teller, DC., Stenkamp, RE et al, Science, 289, 739-745, **2000**.

32- Evans, B. E., Bock, M. G., Rittle, K. E., DiPardo, R. M., Whitter, W. L., Veber, D. R., Anderson, P. S. et Freidinger, R. M., *Proc. Natl. Acad. Sci. USA* 83, 4918-4922, **1986**.

33- Chang, R. S. L. et Lotti, V. J. , *Proc. Natl. Acad. Sci. USA* 83,4923-4926, **1986**.

34- Kerwin, J. F., *J. Drugs of the Future* 16, 1111-1119, **1991**.

35- Yamazaki, Y., Akahana, M., Kobayashi, M., Kitazawa, M., Kurashina, Y. et Lizuka, K., *Japan J. Pharmacol.* 63, 219, **1993**.

36- Bock, M. G., DiPardo, R. M., Evans, B. E., Rittle, K. E., Whitter, W. L., Veber, D. L., Anderson, P. S. et Freidinger, R. M., *J. Med. Chem.* 32, 16, **1989**.

37- Véronique Gigoux, Thèse de Doctorat en Sciences Biologiques, Toulouse **1999**.

38- Mantamadiotis, T. et Baldwin, G. S., *Biochem. Biophys. Res. Commun.*, 201(3), 1382-1389, **1994**.

39- Beinborn, M., Lee, Y. M., McBride, E. W., Quinn, S. M. et Kopin, A. S., *Nature*, 362(6418), 348-350, **1993**.

40- Malatynska, E., Knapp, R. J., Stropova, D., Varga, E. et Yamamura, H. I., *Neuroreport.*, 6(9), 1325-1329, **1995**.

41- Jagerschmidt, A., Guillaume-Rousselet, N., Vikland, M. L., Goudreau, N., Maigret, B. et Roques, B. P., *Eur. J. Pharmacol.*, 296(1), **1996**.

42- Jagerschmidt, A., Guillaume, N., Roques, B. P. et Noble, F., *Mol. Pharmacol*., 53(5), 878-885, **1998**.

43- Smeets, R. L., Ijzerman, A. P., Hermsen, H. P., Ophorst, O. J, Van Emst-de Vries, S. E., De Pont, J. J., Willems, P. H., *Eur. J. Pharmacol*., 325(1), 93-99, **1997**.

44- Bläker, M., Ren, Y., Gordon, M. C., Hsu, J. E., Beinborn, M. et Kopin, A. S., *Mol. Pharmacol.*, 54(5), 857-863, **1998**.

45- Gully, D., Frehel, D., Marcy, C., Spinazze, A., Lespy, L., Neliat, G., Maffr and, J. P. et Le Fur, G., *Eur. J. Pharmacol.* 232, 13-19, **1993**.

46- Folkers, G., Trumpp-Kallmeyer, S., Gutbrot, O., Krickl, S., Fetzer, J. et Keil, G. M., *J. Comp.-Aided Mol. Des.*, 5, 384-404, **1991.**

47- Siezen, R. J., Rollema, H. S., Kuipers, O. P. et de Vos, W. M., *Protein Eng.*, 8, 117-125, **1995**.

48- Gouldson, P., Legoux, P., Carillon, C., Delpech, B., Le Fur, G., Ferrara, P. et Shire, D., *Eur. J. Pharmacol*., 359, 115-124, **2000**.

49- Silvente-Poirot, S. et Wank, S.A., *J. Biol. Chem*., 271,14698-14706, **1996**.

50- Gouldson, P., Legoux, P., Carillon, C., Delpech, B., Le Fur, G., Ferrara, P. et Shire, D., *Eur. J. Pharmacol*., 383, 339-346, **1999**

CHAPITRE V

ETUDE STRUCTURALE DE NOUVEAUX AGONISTES PSEUDOPEPTIDIQUES SUR LE RECEPTEUR CCK_2.

ETUDE DE LEUR MODE D'INTERACTION AVEC LE RECEPTEUR.

I.-.INTRODUCTION

Nous avons décrit dans le chapitre I le rôle physiologique des récepteurs CCK_2 de la cholécystokinine et il semble plus complexe que celui des récepteurs CCK_1. En effet, ces derniers sont essentiellement connus et étudiés pour leur action dans le tractus gastro-intestinal et la régulation de l'appétit[1], alors que les récepteurs CCK_2 sont impliqués dans des phénomènes très importants tels que la douleur, les processus de mémorisation, l'anxiété, les désordres psychiatriques, la vigilance et l'attention. Ils représentent, par conséquent, une importante cible pour le développement de médicaments dans ces domaines de pathologie. Plusieurs molécules ont été proposées comme des antagonistes et/ou agonistes potentiels sur les récepteurs CCK_2. En ce qui concerne la recherche des agonistes sur ce type de récepteur, des investigations ont été essentiellement réalisées sur la série CCK-8, avec une analyse systématique du rôle joué par chaque acide aminé dans la séquence octapeptidique. Les dérivés du tétrapeptide CCK-4, ont été également largement étudiés dans le but d'obtenir des ligands petits et contraints. Cependant, malgré les nombreuses études de modélisation consacrées à la définition du pharmacophore CCK_2[2-4], plusieurs aspects restent encore obscurs. En effet, de nombreuses expériences pharmacologiques réalisées *in vitro* et *in vivo* ont montré que différents profils pharmacologiques peuvent être associés à différentes classes de ligands CCK_2[5-10]. Un consensus structural commun demeure encore difficile à obtenir en termes de simples relations structure-activité. Cependant, l'hétérogénéité biochimique et fonctionnelle des récepteurs CCK_2, associée aux profils pharmacologiques **CCK_{2A}** et **CCK_{2B}**, est actuellement prouvée[11].

Jusqu'à récemment, un seul composé nommé BC264 de formule : Boc-Tyr(SO_3H)-gNle-mGly-Trp-(NMe)Nle-Asp-Phe-NH_2, a été caractérisé comme un agoniste CCK_{2B}. Cet analogue de CCK-7, présentant une sélectivité et une affinité remarquables sur le récepteur CCK_2[12], se distingue des autres agonistes existants, par ses propriétés extrêmement favorables. En effet, ce composé est un bon stimulant, n'est pas anxiogène[13,14] et semble même renforcer la mémoire [15,16] ; alors que les autres agonistes dont le Boc-CCK-4 et le BC197 (Boc-D.Asp-Tyr(SO_3H)-Nle-D.Lys-Trp-Nle-Asp-Phe-NH_2), ont le profil CCK_{2A}, essentiellement caractérisé par les propriétés anxiogènes *in vivo*[13,17].

Cependant, malgré l'intérêt généré par sa pharmacologie, le composé **BC264** souffre de plusieurs inconvénients[15] liés en particulier à sa synthèse ; ce qui exclut son développement thérapeutique ultérieur. Ceci a conduit le groupe de recherche dirigé par Mme C. Garbay dans l'unité du professeur B.P. Roques à l'université de Paris V, à déployer ses efforts dans le but

de trouver de nouveaux agonistes plus simples, dépourvus d'inconvénients et possédant le profil **CCK_{2B}**. La stratégie adoptée par ce groupe a été de raccourcir le composé **BC264**, avec pour point de départ son tétrapeptide C-terminal : **Boc-Trp-(NMe)Nle-Asp-Phe-NH_2**, référencé **PJC-42** et ayant le profil **CCK_{2A}**[18]. La comparaison entre les deux structures, associées à des profils pharmacologiques 'opposés', montre que le fragment responsable de la conversion d'activité doit être situé dans la partie N-terminale **Boc-Tyr(SO_3H)-gNle-mGly** (**g** et **m** indiquent que la liaison peptidique est rétro-inversée, NHCO au lieu de CONH, voir **Annexe 1**). Une étude extensive sur le rôle de la partie N-terminale des analogues CCK-4 et CCK-5 a conduit à une nouvelle série d'agonistes CCK_2, décrivant la possibilité de l'existence de deux modes de reconnaissance au sein du récepteur CCK_2[19]. De plus, les composés représentatifs de cette nouvelle série ont été étudiés dans plusieurs expériences *in vivo*, et ont permis de les séparer en deux familles selon leur profil pharmacologique CCK_{2A} ou CCK_{2B}[20].

Ces composés possédant la structure commune **R-Trp-NMeNle-Asp-PheNH_2** où le substituant en N-terminal varie, sont présentés avec leur séquence, leur affinité et leur profil pharmacologique dans le **tableau 1**. Les études biochimiques[19] indiquent que la caractéristique cruciale d'appartenance à la famille CCK_{2B} est la présence du fragment malonyl **-CO-CH_2-CO-**. De plus, tous ces nouveaux composés qui sont actifs dans un test d'activité locomotrice 'open field', ont cette caractéristique structurale, alors que les composés inactifs ne possèdent pas un tel motif[20].

Cette caractéristique apparaît donc comme cruciale pour l'activité CCK_{2B} ; ce qui peut être considéré comme un premier critère dans la prédiction des propriétés **CCK_{2B}** et dans le design de nouveaux agonistes **CCK_{2B}**.

Les résultats de l'étude pharmacologique[20] suggèrent d'une part, que cette apparente dichotomie pourrait être reliée à la différence fonctionnelle entre les deux profils pharmacologiques, et, d'autre part, que le composé **6** (noté **BBL454, Tableau 1**) pourrait être considéré comme le nouveau agoniste **CCK_{2B}** de référence du fait de ses propriétés extrêmement favorables. Ces conclusions renforcées par les résultats d'expériences de RMN et de quelques études de modélisation préliminaires[19], nous ont incité à entreprendre une étude de modélisation moléculaire étendue à l'ensemble des composés de la série[21].

En effet, de telles études[19] réalisées sur les composés **1**, l'analogue cyclique de **1** (RB360) et le composé **4** ont suggéré que la molécule **1** serait caractérisée par une conformation repliée, alors que le composé **4** par une forme étendue (voir **figure 1**). Dans cette figure, Il apparaît clairement que les conformations étendues sont plus importantes dans le cas de **4 (**2b dans la

figure**)** tandis que les repliées sont dans le **1** (1b dans la figure) et seulement les conformations repliées apparaissent dans RB360.

Les objectifs principaux de notre présente analyse conformationnelle exhaustive sont les suivants :

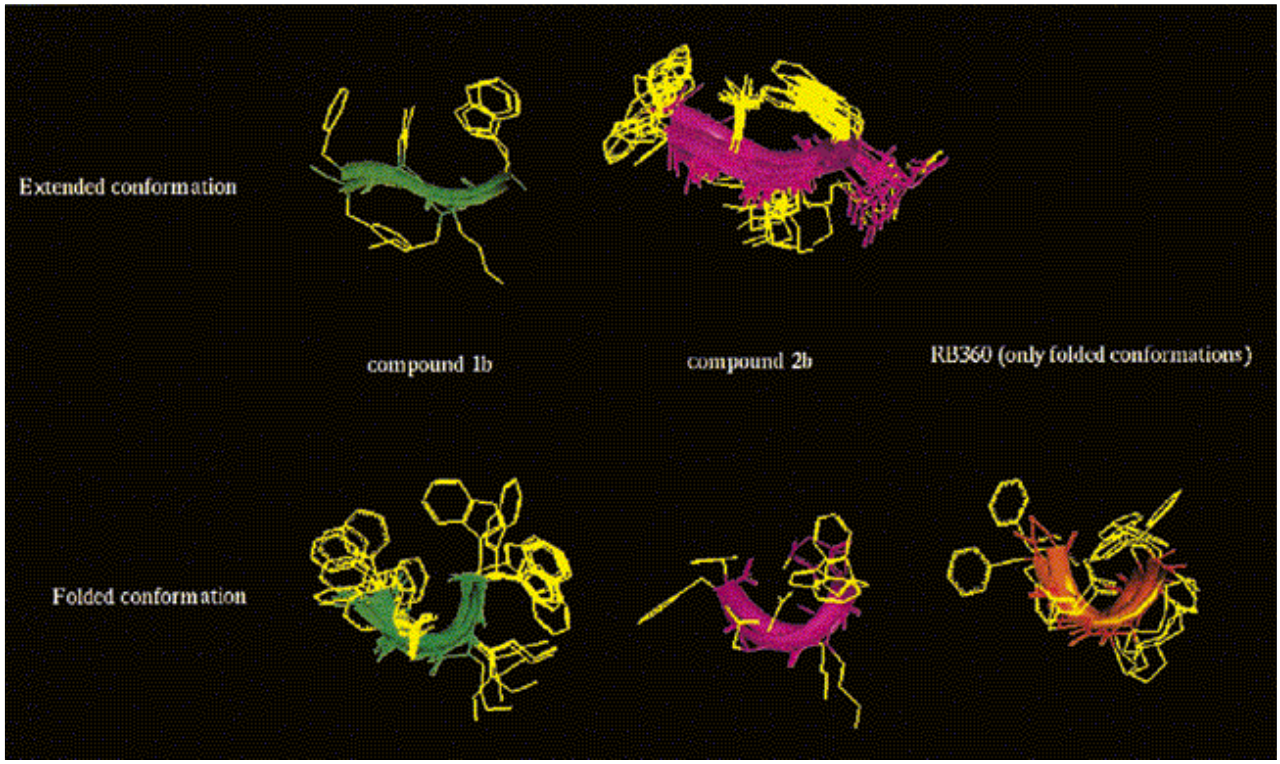
a) Associer, si possible, chaque famille de composés à une conformation représentative simple.
b) Etablir une correspondance possible entre chaque profil pharmacologique et le comportement conformationnel (ou expliquer une telle hétérogénéité des propriétés pharmacologiques en termes de comportement conformationnel).
c) Déterminer la conformation préférentielle de chaque composé et, en particulier, du **BBL454** qui peut être considéré désormais comme le nouveau agoniste CCK_{2B} de référence[20].

Tableau I : Formules, affinités et profils pharmacologiques des composés étudiés ayant pour structure commune : **R-Trp-NMeNle-Asp-Phe-NH_2**.

Composé	R	Ki(CCK_2)(nM)*	Profil Pharmacologique
1	$(CH_3)_3C$-O-CO	2.60	CCK_{2A}
2	CH_3-CH_2-CO	4.35	CCK_{2A}
3	CH_3-O-CO	23.80	CCK_{2A}
4 (RB400)	^-O_2C-CH_2-CO	0.75	CCK_{2B}
5	H_2NOC-CH_2-CO	2.05	CCK_{2B}
6 (BBL454)	CH_3-$(CH_2)_4$-NHCO-CH_2-CO	3.00	CCK_{2B}
7	^-O_2C-CH_2-CH_2-CO	3.02	CCK_{2B}

Ki est mesuré sur des cellules du récepteur CCK_2 du rat.

Figure 1 : comportement conformationnel des composés **1** (1b), **4** (2b) et du **RB360** (analogue cyclique de **1**).



II- MODELISATION MOLECULAIRE DES NOUVEAUX AGONISTES PSEUDOPEPTIDIQUES

1) - Méthodologie utilisée

La modélisation et l'analyse conformationnelle des molécules ont été entièrement réalisées dans le cadre de la mécanique moléculaire et à l'aide des différents logiciels de la société Accelrys[22], implantés sur les stations de travail Silicon Graphics. Nous avons adopté le même protocole que celui utilisé dans l'analyse conformationnelle des benzodiazépines (*cf* Chapitre III, paragraphe 2).

Chaque composé a été construit, à partir des fragments de la librairie géométrique standard disponible dans le logiciel Insight II[22]. Afin de mimer le milieu physiologique, les résidus Asp et les groupements acides en N-terminal sont pris sous forme chargée. La simulation a été effectuée dans le vide et avec une constante diélectrique égale à la distance inter atomique, ce qui permet de simuler grossièrement l'environnement électrostatique dû au solvant aqueux. Dans tous les calculs, effectués sans cut-off, nous n'avons pas tenu compte des termes croisés et les potentiels de liaison sont décrits par les fonctions harmoniques. Le champ de force

utilisé est CFF91 (Consistent Force Field 91) [23]. Le calcul de dynamique moléculaire a été effectué avec contrainte sur la chiralité et sans aucune contrainte sur les angles oméga.

Comme cette étape 'préliminaire' d'analyse de l'échantillonnage obtenu pour chaque composé a conduit à un nombre de familles structurales trop important, nous avons recherché des ressemblances entre les familles de manière à repérer les conformations les plus vraisemblables pour chaque composé. Nous avons donc déterminé certaines propriétés structurales sur le représentant de chaque famille du composé. Ainsi, le moment dipolaire, le rayon de giration, le moment d'inertie ont été calculés au moyen du module Decipher[22]. Le comportement électrostatique[24] a été déterminé à l'aide du programme Delphi[22]. Le modèle utilisé pour le calcul des potentiels électrostatiques moléculaires est présenté dans **l'annexe 3**.

Delphi permet de calculer le potentiel électrostatique sur un ensemble de points précis autour d'une molécule se trouvant dans un solvant (ici, c'est l'eau) et de présenter les résultats sous forme de contours tridimensionnels à des niveaux spécifiés. Nous avons examiné les potentiels électrostatiques positifs et négatifs à plusieurs niveaux compris entre –1.25 kT et 1.25 kT, k étant la constante de Boltzmann et T la température absolue (*Confer* annexe 3).

Etant donné que les médicaments agissant en tant qu'antagonistes ne forment pas de liaison covalente avec leurs cibles biologiques, les interactions électrostatiques sont très 'prisées' par les chercheurs en modélisation. En effet, l'importance de cet outil qui est le calcul du potentiel électrostatique d'une molécule, a été largement reconnu en biologie structurale, en biochimie et en chimie[24-26]. Il nous renseigne, en particulier, sur le comportement électrostatique de la molécule ; ce qui permet d'avoir ou de prédire des informations qualitatives sur les interactions ligand/récepteur.

Le moment d'inertie d'une molécule représente sa capacité d'opposition à pouvoir se mouvoir autour d'un axe de rotation ; plus le moment d'inertie est important, plus la molécule est compacte.

Le rayon de giration est le rayon de la sphère qui contiendrait la molécule ; il caractérise donc la répartition des atomes autour de l'axe qui passe par le centre de cette sphère.

2) - Résultats de l'analyse conformationnelle

Pour l'ensemble des pseudopeptides, la recherche conformationnelle par la méthode du recuit simulé conduit à un grand nombre de conformères se rangeant dans la gamme des structures repliées et étendues. Ce qui démontre la grande flexibilité inhérente aux peptides linéaires[2].

Nous donnons dans le **tableau II**, le nombre de structures de basse énergie extraites de l'échantillonnage, le nombre de familles, le plus petit RMSD (Root Mean square Deviation) entre deux familles distinctes et le plus grand RMSD à l'intérieur de chaque famille. Pour chaque composé étudié, nous présentons dans le **tableau III**, les familles conformationnelles classées par ordre d'énergie croissante, le poids (en % de population) de chaque famille et les propriétés structurales. Nous y avons également reporté sous forme de flèches, les liaisons hydrogènes intramoléculaires dans chaque composé. Il est intéressant de noter que la liaison peptidique en N-méthylé est *trans* dans l'ensemble des structures sélectionnées.

Composé 1 (Boc-Trp-NMeNle-Asp-Phe-NH_2)

La première étape de notre analyse a donné sept familles conformationnelles, qui se distinguent par une forme plus ou moins repliée. L'examen des propriétés structurales a révélé des similarités entre plusieurs familles réduisant ainsi leur nombre à trois. Les formes repliées qui semblent être préférées par ce composé sont favorisées par des γ-turns stabilisés par une interaction de type liaison hydrogène entre le CO du **Trp** et le NH de l'**Asp**. Ceci est cohérent avec les résultats de l'étude où l'analyse a été effectuée à l'aide de la spectroscopie RMN et la dynamique moléculaire avec contraintes géométriques[3]. Nos résultats indiquent une forte polarité du composé (moment dipolaire > 12D). Le comportement électrostatique montre que le potentiel négatif créé par le résidu aspartate est distribué sur l'ensemble de la molécule à l'exception des chaînes latérales des résidus hydrophobes : Trp, Phe, Nle et le groupement tertiobutyle (***figure 2***). Les différences dans la stabilité (~3 kcal.mol^{-1}) et dans les populations (représentées ici par le poids donné en %) (**Tableau 3**) observées entre les familles F1 et F2 nous permet de déduire que ce composé adopte une conformation présentant un rayon de giration petit, caractérisant donc une conformation repliée. Le conformère représentant la famille **F1** qui est la plus stable sera utilisé dans le modèle d'interaction avec le récepteur.

Composé 2 (CH_3CH_2CO-Trp-NMeNle-Asp-Phe-NH_2)

La première étape de l'analyse conformationnelle a généré 4 familles. Le comportement électrostatique de ces familles reflète une bonne similarité entre elles. Un γ-turn autour du résidu Norleucine stabilisé par liaison hydrogène, est présent seulement dans la famille F2. A cause de leurs basses stabilités énergétiques et de leurs populations non importantes, les familles **F3** et **F4** ne représenteraient pas de façon fiable le comportement conformationnel du composé. En revanche, il semble raisonnable de considérer **F1** et **F2**, qui sont énergétiquement similaires, comme représentant correctement le comportement conformationnel du composé. Ces conformations sont présentées dans la ***figure 4***.

Composé 3 (CH_3-OCO-Trp-NMeNle-Asp-Phe-NH_2)

Les 3 familles obtenues, caractérisées par une forme plus ou moins repliée, semblent avoir des populations similaires mais des comportements électrostatiques différents. Cependant, le 'gap' énergétique de 3 kcal.mol^{-1} entre les deux premières familles et la famille **F3** nous conduit à penser que ce composé serait mieux représenté par **F1** et **F2**, dans lesquelles le rayon de giration et le moment d'inertie sont similaires et diffèrent par leur moment dipolaire (**Tableau III**). Cependant, le représentant de **F3** n'est pas à exclure des travaux de docking.

Composé 4 (RB 400, ^-O_2C-CH_2CO-Trp-NMeNle-Asp-Phe-NH_2)

Dans la première étape de l'analyse, six familles conformationnelles ont été obtenues, dont les plus peuplées ont une forme étendue. L'examen des propriétés structurales a révélé des similarités entre quelques familles réduisant leur nombre à 3. Ces familles, caractérisées par une forme plus ou moins étendue, semblent avoir des populations différentes mais des stabilités similaires (~0.5 kcal.mol^{-1}). La forte population de la famille **F3** suggère que le composé adopte préférentiellement une conformation étendue. Il est important de noter que cette forme est due à l'absence de liaisons hydrogène entre le NH_2 en N-terminal et le CO_2^-. Le comportement électrostatique montre que le potentiel négatif créé par l'Asp et le CO_2^- est distribué sur toute la molécule à l'exception des chaînes latérales des résidus Phe et Nle.

Composé 5 (H_2N-CO-CH_2CO-Trp-NMeNle-Asp-Phe-NH_2)

La première étape de l'analyse conformationnelle a révélé une forte flexibilité du composé. Cependant, l'étude des propriétés structurales ont permis de regrouper ces conformations à seulement deux familles. Une forme étendue, favorisée par des liaisons hydrogène impliquant les groupements NH_2 terminaux, caractérisent les deux familles. Elles se distinguent selon que ces liaisons s'établissent avec les atomes du squelette peptidique (backbone) ou ceux des chaînes latérales. Elles paraissent avoir un comportement électrostatique similaire et à l'exception des chaînes latérales des résidus hydrophobes, la molécule semble se comporter comme un ion négatif. Étant donné que la famille **F1** est plus peuplée et plus stable, il est donc probable que ce composé adopte préférentiellement une conformation étendue ; ce qui est confirmé par les valeurs du rayon de giration et du moment d'inertie (**Tableau III**).

Composé 6 (BBL454, CH_3-$(CH_2)_4$-NHCO-CH_2-CO-Trp-NMeNle-Asp-Phe-NH_2)

L'analyse conformationnelle a révélé dans une première étape une forte flexibilité du composé **6**. Un examen attentif des propriétés structurales ont permis la réduction de

l'échantillonnage à 4 familles. Les comportements électrostatiques des familles **F1** et **F2** paraissent similaires. On note également qu'à l'exception des chaînes latérales des résidus hydrophobes la molécule semble se comporter comme un ion négatif ; ce qui laisse supposer qu'elle se logerait au sein du récepteur dans le même site que la molécule **5**. A cause de la forte population et de la relative stabilité de la famille F1 (plus stable que les autres de 1.5 kcal.mol^{-1}), il est clair que cette molécule BBL454 semble préférer une conformation étendue ; ce qui est confirmé par les valeurs du rayon de giration et du moment d'inertie (**Tableau III**). Par conséquent, le représentant de cette famille **F1** sera considéré comme un bon candidat dans l'étude du site d'interaction avec le récepteur.

Composé 7 ($^{-}O_2C$-CH_2CH_2CO-Trp-NMeNle-Asp-Phe-NH_2)

L'analyse préliminaire conduit à 10 familles qui seront ensuite regroupées en 3 après la deuxième étape d'analyse qui consiste à examiner d'autres propriétés structurales. Le résultat traduit toujours une forte flexibilité du composé. Toutes les familles sont caractérisées par une forte polarité et une liaison hydrogène avec le carboxylate terminal. Cependant, la forte population de la famille **F3** et la faible différence de stabilité entre **F2** et **F3** (0.7 kcal.mol^{-1}) sont en faveur de l'hypothèse que ce composé serait caractérisé par une conformation étendue, ce qui est confirmé par les valeurs du rayon de giration et du moment d'inertie (**Tableau III**).

Tableau II : Nombre de structures sélectionnées et de familles pour chaque composé. Les écarts quadratiques moyens inter-familles et intra-familles.

Composé	Nombre de conformations	Nombre de familles	RMSD[a] inter-famille	RMSD[b] intra-famille
1	13	3	5.0	4.0
2	23	4	4.5	3.5
3	13	3	4.5	3.5
4	24	3	4.5	3.2
5	18	2	5.3	2.5
6	21	4	4.8	3.4
7	26	3	4.2	3.1

[a]: plus petit RMSD entre deux structures de familles distinctes.
[b]: plus grand RMSD à l'intérieur de chaque famille.

Tableau III : Propriétés structurales des familles de composés étudiés

Les liaisons hydrogène sont représentées sous forme de flèches vers le résidu accepteur: liaisons hydrogène entre atomes des chaînes latérales sont en pointillés, liaisons hydrogène entre atomes du squelette peptidique et des chaînes latérales sont en doubles pointillés puis en flèches pleines sont représentées celles entre atomes du backbone. Le Rayon de giration, le moment d'inertie et le moment dipolaire sont donnés respectivement en Å, en $Å^2$.(u.m.a) et en Debye.

Composé 1

Famille	Poids (%)	Liaisons Hydrogène	Rayon de giration	Moment d'inertie	Moment dipolaire
F1	46	$(CH_3)_3C$-O-CO-Trp-NMeNle-Asp-PheNH_2	4.96	1591	14.0
F2	38	$(CH_3)_3C$-O-CO-Trp-NMeNle-Asp-PheNH_2	5.14	1709	13.2
F3	15	$(CH_3)_3C$-O-CO-Trp-NMeNle-Asp-PheNH_2	5.39	1881	12.6

Composé 2

F1	16	H_3C-CH_2-CO-Trp-NMeNle-Asp-PheNH_2	4.85	1413	6.7
F2	44	H_3C-CH_2-CO-Trp-NMeNle-Asp-PheNH_2	5.07	1546	12.1
F3	12	H_3C-CH_2-CO-Trp-NMeNle-Asp-PheNH_2	5.56	1856	6.5
F4	28	H_3CCH_2-CO-Trp-NMeNle-Asp-PheNH_2	5.50	1816	9.4

Tableau III: (suite)

Composé 3

F1	31	H_3C-O-CO-Trp-NMeNle-Asp-PheNH_2	4.85	1381	6.7
F2	31	H_3C-O-CO-Trp-NMeNle-Asp-PheNH_2	4.79	1349	14.4
F3	38	H_3C-O-CO-Trp-NMeNle-Asp-PheNH_2	5.39	1706	9.5

Composé 4 (RB400)

F1	4	^-O_2C-CH_2-CO-Trp-NMeNle-Asp-PheNH_2	4.95	1456	11.1
F2	9	^-O_2C-CH_2-CO-Trp-NMeNle-Asp-PheNH_2	4.95	1459	4.2
F3	87	^-O_2C-CH_2-CO-Trp-NMeNle-Asp-PheNH_2	5.23	1623	21.3

Composé 5

F1	73	H_2NCOCH_2CO-Trp-NMeNle-Asp-PheNH_2	5.46	1811	7.9
F2	27	H_2NCOCH_2CO-Trp-NMeNle-Asp-PheNH_2	5.08	1565	9.0

Tableau III : (suite)

Composé 6 (BBL454)

F1	47	CH_3-$(CH_2)_4$-NHCO-CH_2CO-Trp-NMeNle-Asp-Phe-NH_2	5.51	2145	11.6
F2	43	CH_3-$(CH_2)_4$-NHCO-CH_2CO-Trp-NMeNle-Asp-Phe-NH_2	5.25	1945	10.7
F3	5	CH_3-$(CH_2)_4$-NHCO-CH_2CO-Trp-NMeNle-Asp-Phe-NH_2	5.01	1772	13.4
F4	5	CH_3-$(CH_2)_4$-NHCO-CH_2CO-Trp-NMeNle-Asp-Phe-NH_2	5.52	2151	9.8

Composé 7

F1	23	$^-O_2CCH_2CH_2CO$-Trp-NMeNle-Asp-PheNH_2	4.75	1384	12.9
F2	19	$^-O_2CCH_2CH_2CO$-Trp-NMeNle-Asp-PheNH_2	4.90	1473	9.6
F3	58	$^-O_2CCH_2CH_2CO$-Trp-NMeNle-Asp-PheNH_2	5.13	1619	19.4

3) - Discussion

La première étape de ce travail consiste à déterminer, pour chaque composé proposé, l'échantillonnage conformationnel à l'aide des techniques de la modélisation moléculaire, impliquant une simulation de dynamique moléculaire à haute température et couplé au recuit simulé. L'échantillonnage obtenu pour chaque composé, est ensuite analysé. Il est intéressant de noter que pour chaque composé, la superposition de tous les atomes des conformères obtenus laisse suggérer une existence possible de seulement deux familles de conformations :

étendue et repliée. Les conformations repliées semblent être plus nombreuses dans quelques composés de la série, alors que les étendues sont prédominantes dans les autres.

Les résultats de la modélisation montrent, en particulier, que les composés **1-3**, supposés présenter un mode de liaison hydrophobe avec le récepteur[19], ont une préférence pour les conformations repliées. En revanche, les composés **4-7**, classés pharmacologiquement comme des ligands **CCK_{2B}** et supposés avoir un mécanisme de reconnaissance hydrophile sur le récepteur, préfèrent une forme plutôt étendue. Signalons que le même type d'analyse réalisée sur le dérivé cyclique du composé **1**, a révélé seulement des structures repliées[19], ce qui confirme la validité de notre approche. Ce composé cyclique possède un profil pharmacologique **CCK_{2A}**. Puisque les composés **1** et **6** sont considérés désormais[19,20] comme des composés de référence au profil pharmacologique **CCK_{2A}** et **CCK_{2B}**, respectivement, il est important de noter que la structure représentative repliée du composé **1** et celle étendue du composé **6** pourraient expliquer leur meilleure sélectivité pharmacologique à l'intérieur des rangées **1-3** et **4-7**, respectivement. Notons que nos résultats de calcul qui sont, bien sûr, qualitatifs, n'ont pas besoin d'être corrélés avec les résultats *in vitro* reportés ici comme des valeurs de K_i (**Tableau I**). En fait, ces valeurs d'affinités représentent une affinité 'apparente' du composé, étudié directement sur le récepteur. Brièvement, ces valeurs ont été obtenues dans des expériences de compétition, où le ligand testé 'rivalise' avec le ligand CCK-8 tritié ([^{3}H]pCCK-8) pour se lier au récepteur CCK_2. Par conséquent, le rapport entre le ligand tritié radioactif lié et libre (qui est utilisé pour déterminer l'affinité) ne sera pas représentatif d'un ''réel'' binding du composé testé.

De cet état de fait, puisque CCK-8 est généralement considéré comme un agoniste **CCK_{2A}**, il est probable que les affinités des composés **CCK_{2B} (4-7)** soient sous-estimés, et ne peuvent pas être utilisés dans des propos conformationnels[27] .

Nos résultats de calculs de mécanique moléculaire sont similaires à ceux obtenus au moyen des analyses de spectroscopie RMN[3,28]. En fait, les résultats d'études structurales par RMN conduites sur les composés **4** et **1,** pour caractériser leur spécificité potentielle, accréditent l'hypothèse de leur comportement différent[19,3] : l'absence dans la molécule **4** des pics de NOE (Nuclear Overhauser Enhancement) caractérisant la structure repliée de la molécule **1** suggère que le composé **4** adopte une conformation différente, c'est à dire non-repliée. Il est très intéressant de noter que c'est aussi le cas pour **BC264** (Boc-Tyr(SO_3H)-gNle-mGly-Trp-(NMe)Nle-Asp-Phe-NH_2) dont les conformations dominantes ne sont pas repliées[19]. Cette ressemblance est confortée par les études pharmacologiques[20], confirmant un comportement similaire des pentapeptides contenant le groupement malonate et le composé BC264.

Un grand nombre d'études conformationnelles ont été consacrées aux peptides CCK en utilisant différentes méthodes : la spectroscopie RMN, les mesures de fluorescence, les calculs théoriques, le dichroïsme circulaire et la cristallographie aux rayons X[2-4,29-37]. Les résultats de ces méthodes suggèrent une large gamme de conformations, allant des formes repliées (impliquant des β et des γ turns) pour les CCK-8 et les peptides dérivés[29-31], aux formes plus étendues pour les fragments plus petits comme le CCK-4[4,32,33], bien que des conformations repliées étaient aussi proposées[34-36]. Ce qui peut permettre dans quelques cas une adaptation conformationnelle selon le récepteur.

Il faut noter que nos résultats présentent une analogie avec ceux des précédentes études[3,37]. En effet, la conformation proposée par l'équipe de Nadzan pour l'analogue contraint de Ctp-t-3PP-Asp-Phe-NH_2, au moyen d'une combinaison de dynamique moléculaire, de recuit simulé et des techniques de 'forcing template' est en accord avec les conformations établies dans notre étude. L'analogie concerne, en particulier, l'arrangement spatial de la partie C-terminale, où la même orientation de part et d'autre du backbone des chaînes latérales du propyle (ou du Nle dans notre cas) et de l'aspartate est observée.

Malgré la flexibilité des pseudopeptides étudiés dans ce travail, la série de conformères générée pour chacun d'eux, fournit une indication de l'espace conformationnel disponible pour le composé et quelques propriétés communes (ou différentes) dans leurs caractéristiques structurales. Il est admis que les chaînes latérales aromatiques du tétrapeptide CCK-4, jouent un rôle clé dans l'interaction avec le récepteur[3]. En effet, leurs orientations spatiales et la distance entre ces résidus peuvent être critiques, en termes de liaisons avec le récepteur. Puisque des préférences conformationnelles distinctes sont observées pour ces pseudopeptides, en particulier, par rapport à leur profil pharmacologique, avec des orientations relatives différentes des chaînes latérales des résidus Trp et Phe, nous pouvons supposer que ce phénomène est responsable de la différence dans leur affinité biologique et spécialement dans leur profil pharmacologique.

Une conséquence majeure de la N-méthylation est l'existence possible d'un isomérisme *cis-trans* autour de la liaison peptidique modifiée[27,38], lequel a en effet été observé par RMN[3]. Cependant, d'un point de vue biologique, l'intérêt est bien sûr de déterminer la conformation bioactive du ligand. Toutes les structures sélectionnées par notre analyse sont trouvées *trans* par rapport à cette liaison peptidique N-méthylée. De manière intéressante, il a été proposé précédemment que l'isomère *trans* est la forme biologiquement active des analogues N-méthylés[3]. En ce qui concerne, si la conformation préférentielle déterminée ici est réellement la conformation bioactive, la question reste en suspens : en effet, l'environnement local des

récepteurs peut perturber les équilibres entre conformations. Cependant, il est important de noter que notre propos est d'abord d'établir des analogies de comportement, plus que de déterminer la conformation bioactive. Ce problème sera néanmoins important si les expérimentateurs désirent poursuivre le développement chimique à partir du BBL454 (composé 6, leader actuel) en introduisant des modifications (contraintes, cyclisations) visant justement à figer une conformation supposée bioactive.

Les potentiels électrostatiques sont calculés pour tous les représentants des familles de chaque composé, qui sont ensuite comparés sur la base de ces paramètres. Les résultats concernant seulement les composés **1** et **6**, sont illustrés respectivement dans les figures 1 et 2, qui montrent également une vue stéréoscopique de la conformation. Une comparaison visuelle de ces comportements électrostatiques montre clairement une séparation en deux groupes. En effet, les composés associés au profil pharmacologique CCK_{2A}, semblent moins enveloppés par le potentiel électrostatique négatif dû à l'aspartate que ceux associés au profil CCK_{2B}. Ce résultat pourrait être associé aux deux différents modes de liaison sur le récepteur CCK_2[19] : l'un choisi par les ligands dérivés de CCK-4 possédant un N-terminal hydrophile et l'autre par les analogues hydrophobes. Nos résultats accréditent l'hypothèse de plus d'un représentant pour les composés **4** et **7**, autorisant peut-être un processus d'adaptation particulier dans la poche de liaison du site agoniste du récepteur CCK_2. De plus, il est important de noter que les conformations représentatives des composés **1-3** présentent des interactions de type liaisons hydrogène impliquant les atomes du squelette peptidique et ce n'est pas toujours le cas pour celles représentant les composés **4-7**. Ce résultat, en accord avec ceux des expériences de *binding*[19], tend à suggérer la présence de deux types d'interaction au niveau du récepteur, soit avec une poche hydrophobe soit avec une poche hydrophile. Pris ensemble, ces résultats ont fourni une information appréciable concernant les interactions ligand/récepteur et par conséquent nous aideront dans l'étude de docking qui fait l'objet du paragraphe suivant.

Figure 2 : Comportement électrostatique et vue stéréographique de la conformation préférentielle de la molécule 1 (PJC42). Les atomes d'hydrogène ne sont pas représentés.

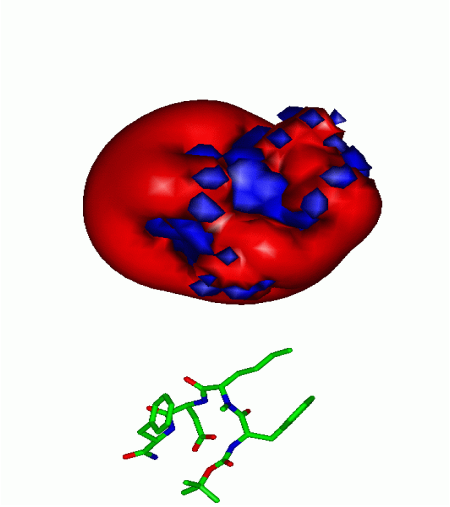


Figure 3 : Comportement électrostatique et vue stéréographique de la conformation préférentielle de la molécule 6 (BBL454). Les atomes d'hydrogène ne sont pas représentés.

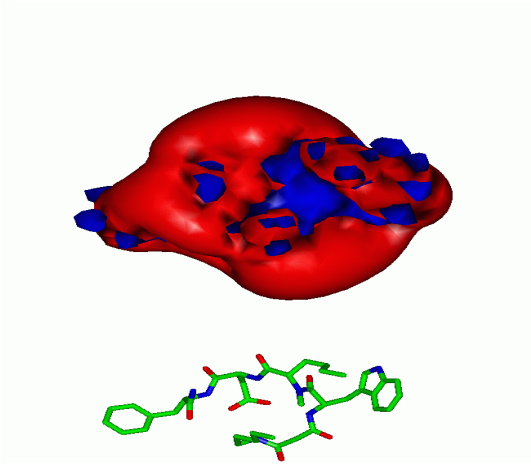
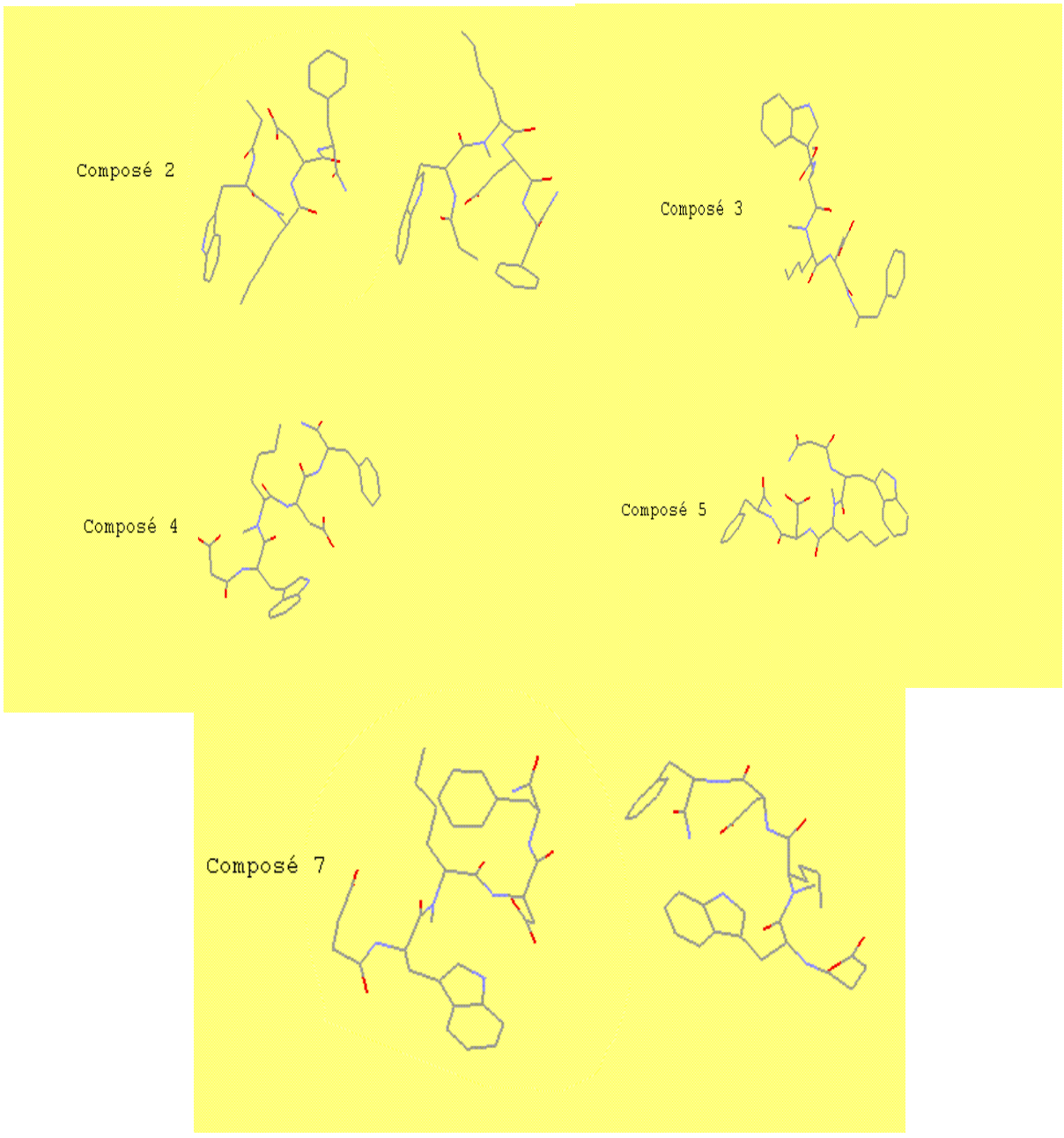


Figure 4: représentations stéréoscopiques des conformations préférentielles des composés **2, 3, 4, 5** et **7**. Les atomes d'hydrogène ne sont pas représentés.



III - ETUDE DES MODES D'INTERACTION LIGAND/RECEPTEUR

1) – Introduction

Dans ce paragraphe, nous nous proposons de docker les conformations préférentielles des composés **1** et **6** qui sont considérés désormais comme des composés de référence au profil pharmacologique $\mathbf{CCK_{2A}}$ et $\mathbf{CCK_{2B}}$, respectivement. Le but de notre travail est d'analyser les interactions ligand/récepteur et de voir les différences au niveau du N-terminal, fragment responsable de la divergence d'activité pharmacologique sur le rat.

2) – Résumé des connaissances actuelles sur les sites de liaison des agonistes CCK sur le récepteur CCK_2.

De l'ensemble des travaux de la littérature[39-41], il apparaît que la liaison des agonistes peptidiques CCK et gastrine sur le récepteur CCK_2 implique des résidus situés dans les régions extracellulaires et la partie supérieure des domaines transmembranaires du récepteur. Ces résultats sont en accord avec ceux obtenus pour d'autres récepteurs de peptides, tels que les récepteurs des neurokinines et le récepteur AT I de l'angiotensine II. Cependant, à notre connaissance, aucune étude n'a été menée, pour le moment, afin de déterminer le type d'interaction s'effectuant entre des résidus du récepteur CCK_2 et ceux des molécules agonistes non peptidiques. L'étude réalisée par l'équipe de Kopin[39], a montré l'implication faible de trois résidus transmembranaires du récepteur CCK_2 dans la liaison de l'agoniste CCK-8 : l'Arg57 (TMI), la Ser219 (TMV) et la Thr354 (TMVI). Une étude menée par le même groupe[40], a montré que l'Asn353 (TMVI) contribue pour un facteur 30 à l'affinité de liaison de la CCK-8. Au moyen d'une technique de mutagenèse dirigée par 'alanine scanning', l'équipe de Silvente-Poirot[41] a montré que l'Arg57 et quatre résidus Asn115, Leu116, Phe120 et Phe122 de la première boucle extracellulaire sont bien impliqués dans la liaison de l'agoniste CCK avec une forte contribution de la Phe120. Cette même équipe a montré également l'implication de His207 (e2) dans la reconnaissance par l'avant-dernier acide aminé (Asp) de la CCK-8. Une autre étude[42] a montré que le résidu Trp351 (TMVI) est impliqué dans la liaison de la CCK, alors que la Phe347 (TMVI) intervient dans l'activation du récepteur induite par la CCK et non dans la liaison de cet agoniste.

3) – Construction et modélisation des complexes ligand/récepteur

Un modèle du récepteur CCK_2 du rat était déjà disponible dans notre base de données quand nous avons entrepris cette étude de docking. Le positionnement de la conformation

préférentielle de chacune des 2 molécules a été réalisé en utilisant le même système de référence que celui utilisé dans le **chapitre IV**. Ce système est le complexe formé par le ligand CCK-9 et le récepteur CCK_2 du rat. Rappelons que ce modèle de référence a été construit antérieurement par B. Maigret et a été affiné sur la base de données expérimentales[42,43].

En utilisant la technique de superposition disponible dans le logiciel InsightII, nous avons donc construit les complexes avec les conformations préférées des composés **1 (***Boc-Trp-NMeNle-Asp-Phe-*NH_2) et **6** [CH_3-$(CH_2)_4$-NHCO-CH_2-CO-Trp-NMeNle-Asp-Phe-NH_2] et le modèle du récepteur CCK_2 du rat. La superposition a été faite sur la base de ressemblances structurales des ligands et la valeur minimale du RMS (critères topologiques et géométriques). Etant donné que la molécule cible (CCK-9) et le ligand à positionner possèdent les parties C-terminales communes (Trp, Asp-Phe), il nous a semblé raisonnable de superposer les backbones de ces parties. L'étape suivante consiste à éliminer les mauvais contacts entre le récepteur et le ligand en modifiant manuellement la position des chaînes latérales du récepteur. Une optimisation de la géométrie en trois étapes successives a ensuite été appliquée à chaque système obtenu. Ces étapes, consistant à relaxer graduellement le complexe, sont les mêmes que celles décrites dans le chapitre IV, paragraphe 2. Tous les calculs ont été effectués dans le champ de force CFF91[23], champ de force dans lequel ont été modélisés au préalable les ligands. Afin d'« écranter » les interactions électrostatiques dues à l'environnement hydrophobe du récepteur inséré dans la bicouche lipidique, nous avons utilisé une constante diélectrique dépendante de la distance inter atomique.

Dans chaque complexe optimisé, selon le même protocole de minimisation utilisé dans nos travaux précédents, nous avons recherché à l'aide des techniques info graphiques et de l'option stéréo disponibles dans Insight II, les interactions susceptibles de stabiliser le ligand dans le site de fixation proposé.

4) - Analyse du site de liaison PJC42/récepteur CCK_2

L'analyse du complexe tridimensionnel formé avec la molécule 1 (**PJC42 :** *Boc-Trp-NMeNle-Asp-Phe-*NH_2)), nous a permis d'observer les interactions stabilisantes potentielles. Elles sont, en général, de nature liaison hydrogène entre le ligand et les atomes des chaînes latérales de quelques résidus du récepteur. Le site d'interaction est schématisé dans la figure 5. Les interactions les plus importantes (~1.80 Å) sont celles qui s'établissent d'une part, entre l'oxygène peptidique du résidu **Phe** et la chaîne du résidu **Tyr189** et, d'autre part, entre la

chaîne latérale de l'**Asp** en tant qu'accepteur de proton et la chaîne latérale de l'**His207**. Ces 2 résidus sont localisés respectivement dans le 4^{e} domaine transmembranaire (**TMIV**) et la deuxième boucle extracellulaire (**e2**). D'autres interactions, moins fortes, concernent les atomes du squelette peptidique du ligand et les résidus **Asn358** (**TMVI**), **Ser131** (**TMIII**), **Arg361** (**e3**) et **Thr193** (**TMIV**). Nous avons relevé également une interaction stabilisante de type **π-π** entre le cycle du **Trp** et la **Phe120** (**e1**). La chaîne latérale de la norleucine (**Nle**) se positionne de façon parallèle avec le cycle aromatique de l'**His207** (**e2**), ce qui induirait une interaction stabilisante de type hydrophobe. Il est important de noter que le groupement en N-terminale (qui est ici le Boc) est partenaire d'aucune liaison hydrogène avec le récepteur ; ce qui signifierait un mode de reconnaissance hydrophobe avec le récepteur. Il est, en effet, important de noter que, selon le modèle proposé ici, ce groupement se positionne à proximité du 3^{e} domaine transmembranaire. La partie concernée de ce domaine est constituée de résidus essentiellement à caractère apolaire et qui sont: **Leu133**, **Met134**, **Gly135** et **Val136**, ce qui correspond à une poche hydrophobe au sein du récepteur. Ce résultat semble en accord avec le mode de reconnaissance hydrophobe de ce fragment sur le récepteur CCK_2[19].

5) - Analyse du site de liaison BBL454/récepteur CCK_2

L'exploration du site de fixation proposé dans le cas de la molécule **6** [BBL454 : *CH_3-$(CH_2)_4$-NHCO-CH_2-CO-Trp-NMeNle-Asp-Phe-NH_2*] a permis de mettre en évidence des interactions potentielles de nature liaison hydrogène entre le ligand et certains résidus du récepteur. Les interactions relevées et illustrées dans la **figure 6**, sont relativement fortes. Cependant, la plus importante est celle qui s'établit entre les chaînes latérales du tryptophane et du résidu **His381** (**TMVII**). Nous avons également observé d'autres interactions s'établissant, d'une part, entre l'oxygène amide de la **Nle** et l'hydrogène guanidinium de l'**Arg361** (situé dans l'interface du **TMVI** et de **e3**) et, d'autre part, entre l'hydrogène amide de la partie C-terminale du ligand et l'oxygène amide de la **Met206** (**e2**).

Au moyen de la technique stéréographique, nous avons essayé de percevoir les partenaires aussi bien du côté du récepteur que de celui du ligand, qui seraient responsables d'autres interactions stabilisantes du complexe. Des interactions de type **π-π**, semblent être présentes dans le complexe. En effet, le cycle aromatique du tryptophane se positionne de façon perpendiculaire à celui du résidu **His381** (**TMVII**), avec une distance entre les centroïdes avoisinant **6 Å**. Nous avons relevé également la proximité de la chaîne latérale hydrophobe de la **Nle** des résidus **His207** (**e2**) ; ce qui induirait un gain de stabilité au complexe formé. En ce

qui concerne le positionnement au sein du récepteur du substituant (en amont du Trp), notre modèle le présente proche du domaine extracellulaire : ce fragment s'ancre à même proximité (~6 Å) des trois boucles extracellulaires (**e1**, **e2** et **e3**). Les résidus les plus proches sont : **Ile126** (**e1**), **Cys127** (**e1**), **Met206** (**e2**), **His207** (**e2**), **Pro366** (**e3**) et **Arg361** (**e3**). Il est à noter que cette poche comprend autant de résidus polaires que de résidus non polaires.

Figure 5 : Vue longitudinale du site de fixation du composé 1 sur le récepteur CCK_2. Pour des raisons de clarté, seuls le ligand et les résidus partenaires d'interactions par liaison hydrogène avec ce ligand sont représentés.

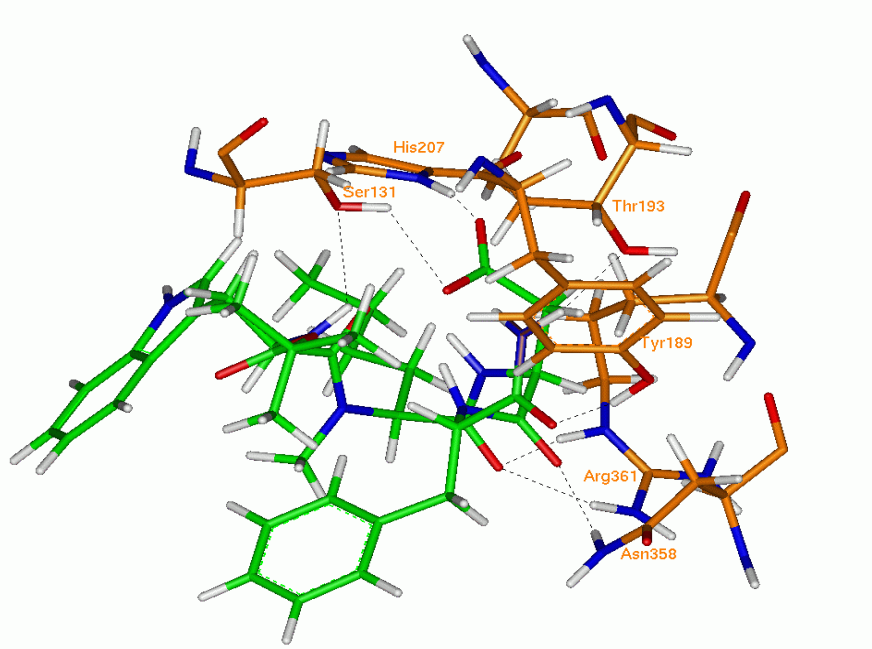
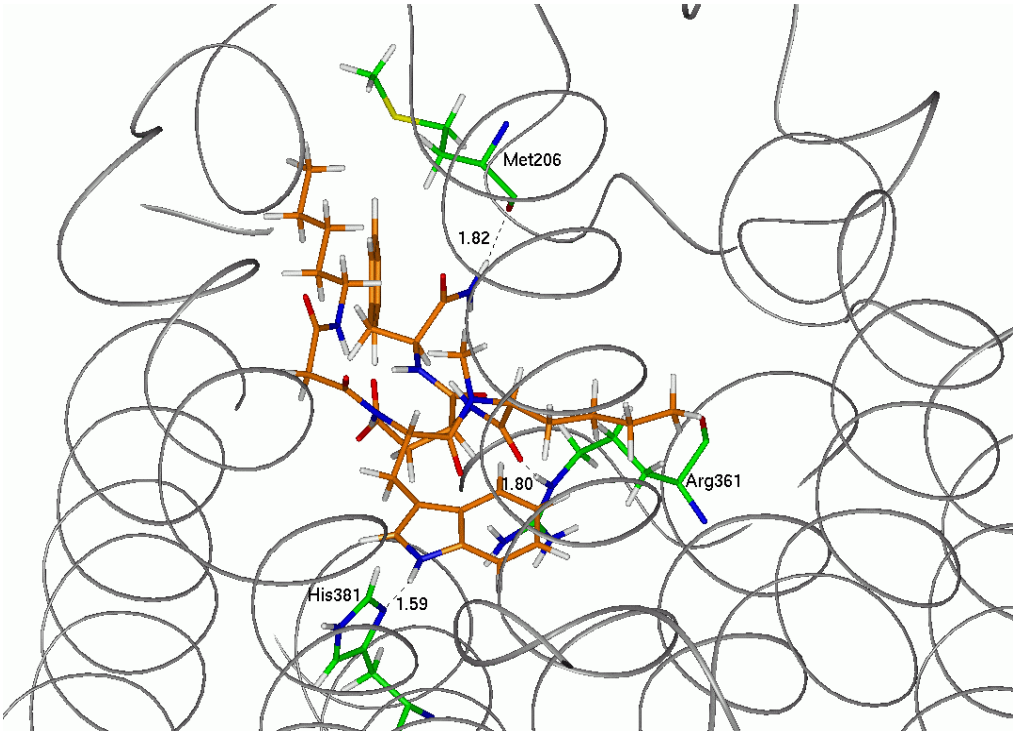


Figure 6 : Vue longitudinale du site de fixation du composé 6 sur le récepteur CCK_2. Pour des raisons de clarté, seuls le ligand et les résidus partenaires d'interactions par liaison hydrogène avec ce ligand sont représentés. Le récepteur est ici représenté sous forme de ruban.



6) - Discussion des résultats

Cette étude de docking conduite sur les composés de référence, aux deux profils pharmacologiques **CCK_{2A}** et **CCK_{2B}**, a permis de mettre en évidence deux sites de fixation. Aucun d'eux n'est en accord complet avec les données actuellement disponibles sur le site de liaison de l'agoniste CCK endogène. En effet, notre étude a mis en évidence de nouveaux résidus susceptibles d'interagir avec le ligand et dont la pertinence mérite d'être vérifiée par des expériences de mutagenèse dirigée et/ou d'affinité sur le récepteur CCK_2. Les deux composés dockés, semblent être stabilisés au sein du récepteur par des interactions avec des résidus dont certains ne sont pas décrits dans la littérature. Cependant, pour ce qui est du site de fixation du composé 1 (**CCK_{2A}**), notre étude a permis de mettre en évidence des interactions essentiellement par liaison hydrogène avec des résidus du récepteur, qui sont décrits dans la littérature comme 'fortement impliqués' dans la liaison avec l'agoniste CCK. Il s'agit, en particulier, de la **Phe120** (**e1**) et du résidu **His207** (**e2**). Il est intéressant de noter également la proximité du résidu **Asn358** (**TMVI**) des résidus **Trp351**, **Asn353** et **Thr354** rapportés par des études expérimentales[39-41] comme impliqués dans la liaison avec la CCK-8.

De plus, notre étude a montré l'intervention du résidu basique **Arg361** (**e3**) dans la stabilisation des complexes construits avec les deux molécules. Ce qui signifie que les deux sites de fixation se recouvrent partiellement et que des interactions de nature ionique pourraient y être présentes. Il est important de noter l'implication du résidu **His381** (**TMVI**) dans le site de fixation du composé **6 (CCK_{2B})**. Rappelons que Jagerschmidt et al[43] ont montré l'importance de ce résidu dans la sélectivité CCK_1/CCK_2 des antagonistes CCK et non dans la liaison des agonistes CCK. Ce point mérite d'être étayé par des travaux expérimentaux appropriés : mutations et mesures d'affinité sur le récepteur.

Selon notre étude, les fragments des composés **1** (PJC42) et **6** (BBL454), responsables de leur profil pharmacologique respectif, ne partagent pas le même espace dans le récepteur. En effet, celui du **BBL454** semble se loger dans le domaine extracellulaire alors que celui du PJC42, serait à proximité de la partie supérieure interne du 3^e domaine transmembranaire. Compte tenu de la nature chimique des résidus constituant ces poches de liaison, ces résultats ne semblent pas en accord complet avec ceux prédits par l'étude des propriétés de liaison sur le récepteur.

Cependant, il ne faut pas perdre de vue les limitations actuelles de ces modèles. En effet, les résultats des expériences de mutagenèse dirigée devraient affiner les modèles proposés dans cette étude. Ceci nous permettra ainsi de mieux percevoir les caractéristiques du site de fixation des agonistes de la cholécystokinine étudiés dans ce travail et de proposer éventuellement au chimiste de synthèse, la molécule au profil pharmacologique CCK_{2B} associé à des propriétés thérapeutiques très favorables.

Nous résumons dans le tableau suivant les résidus intervenant dans des interactions potentielles au sein des complexes ainsi que ceux décrits dans la littérature sur l'agoniste CCK endogène.

Tableau IV : Récapitulation des résidus importants dans les sites de liaison agoniste sur le récepteur CCK_2 du rat.

agoniste	**Notre modèle**		**Littérature ou Expérience**	
	ligand	récepteur	ligand	Récepteur
PJC42 (CCK_{2A})	Phe Trp, Nle Asp	Y189(**TMIV**) H207(**e2**) N358(**TMVI**) S131(**TMIII**) R361(**e3**) T193(**TMIV**)		

BBL454 (CCK_{2B})	Trp Nle	H381(**TMVII**) R361(**e3**) M206(**e2**)		
CCK-8			Asp	R57(**TMI**)[39,41], S219(**TMV**)[39] T354(**TMVI**)[39], N353(**TMVI**)[40] N115(**e1**)[41],L116(**e1**)[41],F120(**e1**)[41], F122(**e1**)[41],H207(**e2**)[42],W351(**TMVI**)[42], F347(**TMVI**)[42]

IV – CONCLUSIONS

L'étude exposée dans ce chapitre est fondée sur de récents développements de nouveaux agonistes pseudopeptidiques de la cholécystokinine sur le récepteur CCK_2. Un travail antérieur[19] à celui-ci décrit la possibilité de l'existence de deux modes de reconnaissance du récepteur CCK_2 par les ligands de cette série **R-Trp-NMeNle-Asp-PheNH$_2$** au niveau du substituant N-terminal. L'étude pharmacologique[20] sur le rat a permis de séparer ces composés en deux familles, selon le profil pharmacologique.

Les résultats de la modélisation réalisée dans la présente étude, montrent en particulier, que les composés **1-3**, supposés présenter un mode de liaison hydrophobe avec le récepteur et associés au profil CCK_{2A}, ont une préférence pour les conformations repliées. En revanche, les composés **4-7**, classés pharmacologiquement comme des ligands CCK_{2B} et supposés avoir un mécanisme de reconnaissance hydrophile sur le récepteur, préfèrent une forme plutôt étendue. Un comportement conformationnel a donc pu être associé à chaque famille de composés. Cette étude d'analyse conformationnelle approfondie, a permis de proposer la(les) conformation(s) préférentielle(s) de chaque composé.

Le docking des composés **1** et **6**, considérés désormais comme des composés de référence, a permis d'analyser les modes d'interaction au sein du récepteur. Les résultats obtenus ont reproduit quelques uns décrits dans la littérature et ont mis en évidence des interactions avec le récepteur. L'analyse des résultats montre que le site de liaison du composé 1 (associé au profil CCK_{2A}) se recouvre avec celui de l'agoniste endogène. En revanche, ceci n'est pas le cas avec le composé 6 présentant un profil CCK_{2B}.

A cause de ses propriétés thérapeutiques extrêmement favorables, le composé **6** (BBL454) pourrait servir de référence pour une optimisation future du profil CCK_{2B}. Sa structure tridimensionnelle pourrait servir, comme modèle dans le design d'agonistes CCK_{2B} non peptidiques.

En conclusion, nous pouvons dire que ce travail fournit une clé dans la compréhension des mécanismes soulignant l'hétérogénéité des 2 effets pharmacologiques associés au récepteur CCK_2.

BIBLIOGRAPHIE

1- Blundell, J. *Trends Pharmacol. Sci*., 12, 147-57, **1991**.

2- Goudreau, N., Weng, J.H., et Roques, B.P. *Arch. Pharm.(Weinheim) ,* 329, 197-204, **1996**.

3- Goudreau, N., Weng, JH., et Roques, B.P., *Biopolymers*, 3, 155-169, **1994**.

4- Pattou, D., Maigret, B., Fournié-Zaluski, M.C., et Roques, B.P. *Intl. J. Pept. Prot. Res*., 37, 440-450, **1991**.

5- Durieux, C., Coppey, M., Zajac, J.M., et Roques, B.P. *Biochem. Biophys. Res. Commun*., 137, 1167-1173, **1986**.

6- Knapp, R.J., Vaughn, L.H., Fang, S., Bogert, C.L., Yamamura, M.S., Hruby, V.J., et Yamamura, H.I. *J. Pharmacol. Exp. The.*, 255, 1278-1286, **1990.**

7- Huang, S., Fortune, K.P., Wank, S.A., Kopin, A.S., et Gardner, J.D. *J. Biol. Chem*., 269, 26121-26126, **1994**.

8- Harper, E.A., Roberts, S.P., Shankley, N.R., and Black, J.W. *Br. J. Pharmacol*., 118, 1717-1726, **1996**.

9- Bellier, B., McCort-Tranchepain, I., Ducos, B., Danascimento, S., Meudal, H., Noble, F., Garbay, C., et Roques, B.P. *J. Med. Chem*., 40, 3947-3956, **1997**.

10- Harper, E.A., Griffin, E.P., Shankley, N.P., et Black, J.W. *Br. J. Pharmacol.*, 126, 1496-1503, **1999**.

11- Daugé, V., et Léna, I. *Neurosci. Biobehav. Rev*., 22, 815-825, **1998**.

12- Charpentier, B., Durieux, C., Pélaprat, D., Dor, A., Reibaud, M., Blanchard, J.C. et Roques, B.P. *Peptides*, 9, 835-841, **1988**.

13- Derrien, M., McCort-Tranchepain, I., Ducos, B., Roques, B.P. et Durieux, C. *Pharmacol. Biochem. Behav.*, 49, 133-141, **1994**.

14- Charrier, D., Dangoumau, L., Puech, A.J., Hamon, M. et Thiébot, M.H. *Psychopharmacology (Berl)*, 121, 127-134, **1995**.

15- Léna, I., Simon, H., Roques, B.P. et Daugé, V. *Neuropharmacology*, 38, 543-553, **1999**.

16- Sebret, A., Léna, I., Crété, D., Matsui, T., Roques, B.P. et Daugé, V. *J. Neurosci.*, 19, 7230-7237, **1999**.

17- Ladurelle, N., Roques, B.P. et Daugé, V. *J. Neurosci*., 15, 3118-3127, **1995**.

18- Million, M.E., Léna, I., Danascimento, S., Noble, F., Daugé, V., Garbay, C. et Roques, B.P. *Lett. Pept. Sci.*, 4, 407-410, **1997**.

19- Bellier, B., Million, E., Danascimento, S., Meudal, H., Kellou, S., Maigret, B. et Garbay, C. *J. Med. Chem*., 43**,** 3614-3623, **2000**.

20- Bellier, B., Crété, D., Million, E., Beslot, F., Bado, A., Garbay, C. and Daugé, V., *British J. Pharmacol.* **2000**, soumis.

21- Tairi-Kellou, S., Bellier, B., Million, M.-E., Garbay, C., Cartier, A.. et Maigret, B., Computer Modelling of New Pseudopeptide CCK_2 Agonists, *Journal of Molecular Structure (THEOCHEM), **2002, (sous presse)***.

22- Accelrys/MSI. Road San Diego, CA 92121 USA.

23- Dinur, U., et Hagler, A. T. New approaches to empirical force fields. In: *Reviews in Computational Chemistry,* Lipkowitz, K.B., and Boyd, D.B., Eds., VCH Publishers, New York, **1991**, Vol. 2, Chapter 4, pp. 99-164.

24- Nicholls, A., Honig, B., *J. Comput. Chem*., 12, 435, **1991**.

25- J. S. Murray et P. Politzer, *J. Org. Chem.*, 56, 6715, **1991**.

26- M. Orozco, E. I. Canela et R. Franco, *Eur. J. Biochem.*, 188, 155, **1990**.

27- Pommier, B., Danascimento, S., Dumont, S., Bellier, B., Million, E., Garbay, C., Roques, B.P., et Noble, F., *J. Neurochem.* 73(5), 281-288, **1999**.

28- Daugé, V. et Léna, I., communication personnelle.

29- Fournie-Zalouski, M.C., Belleney, J., Lux, B., Durieux, C., Gacel, G., Maigret, B. et Roques, B. P. *Biochemistry*, 25, 3778-3787,**1986**.

30- Fang, S., Nikiforovich, G. V., Knapp, R.J., Jiao, D., Yamamura, H.I. et Hruby, V.J. *in peptides: Chemistry, Structure and Biology(Proceedings of the 11th American Peptide symposium)*, Rivier, J. E. and Smith, J. A., Eds. ESCOM, Leiden, pp. 142-143, **1992**.

31- Boteju L.W., Nikiforovich, G.V., Haskell-Luevano, C., Fang S.N., Zalewska, T., Stropova, D., Yamamura, H.I. et Hruby, V.J *J. Med. Chem.*, 39, 4120-4, **1996**.

32- Pincus, M.R., Carty, R.P., Chen, J., Lubowsky, J., Avitable, M., Shah, D., Sheraga, H. A. et Murphy, R. B. Proc. Natl. Acad. Sci. USA, 84, 4821-4825, **1987**.

33- Fournie-Zalouski, M.C., Durieux, C., Belleney, J., Pham, P., Gerard, D. et Roques, B. P. Biopolymers, 24, 1663-1681, **1985**.

34- Hughes, R. A. et Andrews, P. R. J. Pharm. Pharmacol., 39, 339-343, **1987**.

35- Kolodziej, S. A., Nikiforovich, G. V., Skeean, R., Lignon, M. F., Martinez, J. et Marshall, G. R. *J. Med. Chem*., 38, 137-149, **1995**.

36- Pham Van Chuong, P., Penke, B., De Castiglione, R. et Fromageot, P. in Hormone Receptors in Digestion and Nutrition, Elsevier/ North-Holland Biochemical Press, Amsterdam, pp. 33-44.

37- Nadzan, A.M., Garvey, D.S., Holladay, M.W., Shiosaki, M.D., Tufano, M.D., Shue, Y.K. Chung, J.Y.I., May, C.S., Lin, C.W., Miller, T.R., Witte, D.G., Bianchi, B.R., Wolfram, C.A.W., Burt, S. et Hutchins, C.W. in peptides: Chemistry, Structure and Biology (Proceedings of the 12th American Peptide symposium), Rivier, J.E. and Smith, J.A., Eds., ESCOM, Leiden, pp. 100-102, (**1992**).

38- Leff, P., Scaramellini, C., Lawe, C., et McKechnie, K. *Trends Pharmacol. Sci.*, 18(10), 355-62, **1997**.

39- Kopin, A. S., McBride, E. W., Quinn, S. M., Kolakowski, L.F. et Beinborn, M., *J. Biol. Chem.*, 270, 5019-5023, **1995**.

40- Bläker, M., Ren, Y., Gordon, M. C., Hsu, J. E., Beinborn, M. et Kopin, A. S., *Mol. Pharmacol.*, 54(5), 857-863, **1998**.

41- Silvente-Poirot, S. S., Escrieut, C. et Wank, S. A., *Mol. Pharmacol.*, 54, 364-371, **1998**.

42- Jagerschmidt, A., Guillaume, N., Roques, B. P. et Noble, F., Mol. Pharmacol., 53(5), 878-885, **1998**.

43- Jagerschmidt, A., Guillaume-Rousselet, N., Vikland, M. L., Goudreau, N., Maigret, B. et Roques, B. P., *Eur. J. Pharmacol.*, 296(1), 97-106, **1996**.

CONCLUSION GENERALE

ET PERSPECTIVES

Les recherches exposées dans ce mémoire se situent à l'interface de la chimie et de la biologie qui est un domaine fondamental en cette ère post-génomique. Elles sont basées sur de récents développements dans le domaine de la chimie médicinale et de la biologie moléculaire des ligands et des récepteurs membranaires responsables de la communication inter-cellulaire. Ces travaux qui tirent un grand profit des techniques info graphiques et des méthodes de la chimie théorique, consistent à simuler et à analyser, en visualisant, les modes d'interaction ligand/récepteur de la cholécystokinine (CCK).

La présentation de la cible moléculaire de nos recherches : la CCK, dans le chapitre I, nous a permis de situer l'état actuel des connaissances sur cette importante hormone neuropeptidique. La diversité de ses actions physiologiques et la complexité de son mode d'action ont suscité un intérêt très intense aussi bien sur le plan expérimental que sur le plan théorique.

Dans le second chapitre, nous avons présenté les deux types d'approches théoriques de base existant actuellement pour aborder une étude par modélisation moléculaire et qui sont : la mécanique quantique et la mécanique moléculaire. Leurs différences, leurs avantages, leurs limitations spécifiques et leur évolution y sont également discutés.

Le troisième chapitre concerne une étude de relations quantitatives structure-activité d'une série d'antagonistes non-peptidiques, de type benzodiazépine, de la CCK. En l'absence de données structurales sur les deux cibles biologiques de ces molécules, il est nécessaire, dans une première approche du problème, de procéder de manière indirecte sur une collection de molécules issues de différentes sources et utilisées comme système d'apprentissage. Nous avons sélectionné ces molécules en fonction de la disponibilité de données expérimentales concernant leur activité et de manière à disposer d'une base de données caractérisée par une diversité structurale et biologique adéquate pour une analyse statistique. Afin de dégager les corrélations possibles, nous avons mis en œuvre au moyen des méthodes de la chimie théorique un nombre important (47) de descripteurs géométriques et physico-chimiques. Cette étude a conduit à des corrélations tout à fait acceptables bien que susceptibles d'être améliorées et nous a permis de dégager des éléments chimiques pouvant constituer le pharmacophore. L'analyse de nos résultats révèle également un site de liaison, au sein du récepteur CCK_1, caractérisé par une poche plus petite et plus hydrophile que celle apparaissant dans l'autre type de récepteur (CCK_2).

Comme ce premier travail nous a permis de proposer des conformations préférentielles pour chacun des composés étudiés et afin d'aller plus loin dans la recherche des éléments

structuraux indispensables à une activité anti-CCK, nous avons poursuivi nos travaux en analysant, au sein des complexes, les interactions ligand/récepteur. Ces travaux font l'objet du chapitre IV. La connaissance de la structure tridimensionnelle des macromolécules réceptrices, ou des drogues interagissant, est cruciale puisque cela permet d'obtenir le meilleur ajustement possible entre la drogue et son récepteur. Nous avons modélisé les structures tridimensionnelles des récepteurs de la CCK en recherchant une homologie entre la séquence primaire de chaque récepteur et celle d'une protéine de référence dont la structure tertiaire est disponible. Nous avons ensuite procédé, sur les modèles obtenus, au docking moléculaire avec les antagonistes de la série présentant des activités limites extrêmes sur chacun des récepteurs. Les résultats d'analyse des sites de fixation modélisés ne sont pas en accord complet avec les quelques résultats théoriques disponibles dans la littérature. En effet, notre étude a mis en évidence de nouveaux résidus susceptibles d'interagir avec le ligand et dont la pertinence mérite d'être confortée par d'autres calculs et/ou vérifiée par des expériences de mutagenèse dirigée. Des interactions par liaison hydrogène, entre le fragment lié au carbone asymétrique de l'antagoniste et les résidus polaires, représenteraient les principales forces de stabilisation du complexe et particulièrement avec le récepteur CCK_1.

Bien que les deux techniques d'investigation, QSAR et docking, utilisées de manière complémentaire comportent des limites certaines elles nous ont néanmoins permis d'obtenir une image plausible de l'interaction des antagonistes CCK, de type benzodiazépine, avec les deux récepteurs. En effet, l'application des descripteurs quantiques, à un échantillon représentatif, apparaît comme une étape décisive dans l'analyse du mécanisme d'interaction des complexes ligand/récepteur. Nos résultats, malgré la complexité du problème lié à l'effet antagoniste de la CCK, ont mis en évidence les éléments structuraux responsables de l'activité antagoniste et de la sélectivité par rapport à l'un ou l'autre des récepteurs CCK. Ces études complémentaires du mode d'interaction de ces antagonistes ont permis de déboucher sur des idées de synthèse de nouveaux analogues où les caractéristiques structurales et fonctionnelles, propices à une activité et/ou sélectivité optimales, sont réunies et dont l'intérêt consiste surtout en une meilleure stabilité et une meilleure biodisponibilité.

L'étude portant sur l'analyse conformationnelle, présentée dans le chapitre V, sur la série d'agonistes pseudopeptidiques récemment synthétisée, a permis de proposer la ou les conformation(s) préférentielle(s) de chaque composé. Un comportement conformationnel a aussi pu être associé à chaque famille de composés selon son profil pharmacologique. Le docking moléculaire des composés 'têtes de file', représentatifs de chacune des deux familles,

nous a permis d'analyser les modes d'interaction au sein du récepteur CCK_2. Les résultats obtenus ont permis de reproduire quelques aspects concernant les sites de liaison de l'agoniste endogène, décrits dans la littérature et ont, en outre, mis en évidence de nouvelles interactions pertinentes avec le récepteur. En conclusion, nous pouvons dire que ce travail fournit une clé dans la compréhension des mécanismes soulignant l'hétérogénéité des deux effets pharmacologiques associés au récepteur central de la CCK.

Cependant, il ne faut pas perdre de vue les limitations actuelles des modèles proposés. En effet, les résultats des expériences de mutagenèse dirigée devraient affiner les modèles proposés dans nos travaux, ceci nous permettra d'améliorer nos modèles théoriques et ainsi de mieux percevoir les caractéristiques du site de fixation des ligands de la CCK, conduisant à proposer au chimiste la synthèse de molécule au profil pharmacologique choisi.

A plus long terme, en fonction des premiers résultats de mutagenèse dirigée, il est possible d'envisager des mutations 'positives'. En effet, au lieu de rechercher la mutation qui réduit l'affinité, nous serons en mesure, grâce au modèle tridimensionnel, de proposer des positions et des résidus qui pourraient augmenter, ou du moins restaurer, l'affinité du ligand pour le récepteur CCK mutant. Cette démarche est certainement la plus sûre expérience de validation d'un modèle tridimensionnel, mais nécessite une connaissance précise du site de fixation.

Il semble aujourd'hui qu'une connaissance intime de la structure des cibles des médicaments devrait permettre aux chimistes de synthétiser des molécules de structure complémentaire assurant une grande sélectivité de reconnaissance entre les médicaments et les récepteurs. Une telle sélectivité est essentielle pour réduire les effets secondaires parfois nocifs, de certains médicaments

En conclusion, l'ensemble de ces travaux montre que la conception assistée par ordinateur permet de contourner l'absence de données structurales cristallographiques dans le cas des récepteurs transmembranaires qui constituent une cible privilégiée pour tenter de remédier à de nombreuses pathologies. Ces nouveaux outils performants, de la modélisation moléculaire, devraient aider à mieux comprendre les mécanismes impliqués dans l'interaction ligand/récepteur et, à terme, permettre de rationaliser la conception de nouvelles molécules à visée thérapeutique. Toutefois, il faut noter que les résultats seront d'autant plus précis que le niveau conceptuel de description de la structure moléculaire sera élevé et donc riche en informations. Une description réaliste et informative de la structure moléculaire est donc un préalable nécessaire pour l'obtention de résultats significatifs et la déduction d'interprétations raisonnables. Par ailleurs, la complexité des phénomènes dynamiques qui ont lieu au sein de

la membrane lipidique et l'implication des molécules d'eau dans le processus de reconnaissance des ligands sont autant de mystères qui limitent actuellement tout triomphalisme prématuré au modélisateur de cette famille de récepteurs (RCPGs). Intégrer cette complexité, en tenant compte par exemple explicitement des molécules d'eau, dans le processus de simulation par le docking est un défi incontournable qu'il faudra relever afin de parvenir à une vision moléculaire plus complète des interactions ligand/récepteur.

D'autre part, des études très récentes ont montré que les récepteurs couplés aux protéines G ont la propriété de s'associer en homo-, voire en hétéro-diméres. La dimérisation des RCPGs et ses conséquences fonctionnelles peuvent avoir un impact considérable notamment en termes de mise au point de médicaments. C'est dans ce contexte que nous nous proposons de poursuivre nos travaux afin de déterminer si d'une part, les récepteurs CCK existent aussi sous forme homo-dimérique, et si d'autre part, ils sont capables de s'associer avec d'autres récepteurs de la même famille pour former des hétéro-dimères. Cette idée réfutée, il y a quelques années, par les scientifiques requiert aujourd'hui une attention considérable. La signification fonctionnelle de ce phénomène de dimérisation qui peut jouer un rôle décisif dans les événements moléculaires menant à l'activation des RCPGs, est toujours au centre du débat scientifique et il ne fait aucun doute, aujourd'hui, que la dimérisation des RCPGs est un mécanisme important pour l'activation et la régulation de cette classe de récepteurs. Cependant, il faut signaler, que plusieurs aspects de la dimérisation restent nébuleux. Les premières études réalisées sur certains RCPGs montrent que cette oligomérisation semble être un processus spécifique puisque toutes les combinaisons ne sont pas permises. Elles montrent également que la dimérisation peut s'effectuer par interaction des régions carboxy-terminales ou *via* la formation de ponts disulfures entre leurs domaines extracellulaires amino-terminaux. On ne sait pas encore s'il s'agit d'un mode de fonctionnement normal, ou au contraire d'une alternance, voire un équilibre, entre le mode monomérique et le mode dimérique. Mieux nous comprendrons ce fonctionnement, mieux nous pourrons agir sur son déroulement. La recherche de réponses à ces questions concernant les récepteurs CCK font partie de nos perspectives. Elles ouvriront la voie à l'élaboration éventuelle d'une nouvelle classe d'agents thérapeutiques qui pourraient inhiber ou favoriser le 'pairage' de ces récepteurs. En effet, des médicaments qui cibleraient une classe de récepteurs spécifiques en agissant sur la dimérisation pourraient s'avérer beaucoup moins toxiques et donc infiniment plus efficaces.

La faisabilité d'une telle approche est attestée par des calculs effectués, actuellement au laboratoire de chimie théorique de Nancy I, sur un dimère du récepteur de l'interleukine-8, *in*

situ dans son environnement membranaire. Ce travail montre que l'étude d'un système moléculaire comportant 150 000 atomes n'est plus une utopie et que la modélisation des systèmes moléculaires extrêmement complexes est réaliste.

ANNEXE 2

Définition des paramètres statistiques à utiliser dans une analyse de régression multilinéaire

1) Introduction

L'analyse de régression est un outil statistique pour examiner le rapport entre deux variables ou plus, permettant ainsi de calculer la valeur d'une variable en se basant sur la connaissance des autres variables. Beaucoup de systèmes physiques ne peuvent pas être décrits comme rapport simple entre 2 variables x et y. Une régression linéaire multiple de la forme $y= ax_1+bx_2+cx_3+\ldots$, s'avère souvent nécessaire pour décrire correctement le système ou pour relier l'activité biologique d'une molécule à sa structure physico-chimique définie par un ensemble de descripteurs. Les coefficients de régression (a, b, c …) pour chaque variable indépendante (x_1, x_2, x_3 …) appelées variables explicatives et un certain nombre de paramètres statistiques validant la corrélation obtenue, peuvent être déterminés au moyen de la méthode (partielle) des moindres carrés et de la validation croisée.

2) Définition des paramètres statistiques

Une corrélation est jugée significative sur la base de certains critères sur les paramètres statistiques décrits ci-dessous :

n : le nombre de molécules soumises à la régression ou nombre de points de repères ou observations. En général, **n** doit être supérieur ou égal à 5 fois le nombre de variables indépendantes.

s : l'écart type ou déviation standard. C'est une mesure d'incertitude qui donne la précision avec laquelle la fonction dérivée par l'analyse statistique prévoit l'activité biologique observée. Plus la valeur de **s** est petite, meilleure est la corrélation obtenue.

$$s=\sqrt{\frac{1}{n-k-1}}\sum_{i}^{n}(Y_{cal}-Y_{obs})_i^2$$

avec

n : nombre d'observations

k : nombre de variables indépendantes ou nombre de paramètres de la corrélation

Y_{cal} : activité calculée

Y_{obs} : activité mesurée

Erreur standard : appelée aussi intervalle de confiance. C'est l'erreur commise sur le coefficient de chaque paramètre de la corrélation. Dans le cas d'une corrélation monoparamétrique, elle rend compte de l'erreur sur la 'pente' (coefficient de direction) de la droite de régression.

r : coefficient de corrélation multiple. $\mathbf{r^2}$ est une mesure de la proximité de l'ajustement obtenu. Une explication satisfaisante des données est habituellement indiquée par un $\mathbf{r^2}$ au moins de 0.9.

$$r = \sqrt{1 - \frac{\sum_i (y_{cal} - y_{obs})^2}{\sum_i (y_{cal} - \sum_i \frac{y_{obs}}{n})^2}}$$

La significativité de la régression est obtenue par une analyse de variance de la régression $F_{résiduelle}^{régression}$. Le degré de confiance que l'on peut accorder à ce paramètre conduisant à une corrélation significative est donné par F_α selon la loi de Snedecor :

- Si $F < F_\alpha$, la corrélation est à rejeter.
- Si $F > F_\alpha$, la corrélation est significative.

La détermination de F_α est effectuée à l'aide d'une table de valeurs numériques donnant la distribution de F (Fischer-Snedecor). Pour entrer dans cette table il faut déterminer au préalable les degrés de liberté :

-de la variance expliquée par la régression notée γ

-de la variance résiduelle notée γ'

- Pour une équation monoparamétrique, γ = 1 et γ'= n (n étant le nombre de composés)
- Pour une équation multiparamétrique, γ = p-1 (p étant le nombre de paramètres) et γ' = n-p-1.

$\mathbf{F_\alpha}$ est donc déterminé en recherchant dans la table la valeur de $F_{\gamma'}^{\gamma}$.

La validation croisée est une approche qui permet d'évaluer le pouvoir prédictif associé à un modèle QSAR, ce modèle étant sélectionné parmi plusieurs autres présentant des niveaux de complexité différents. En comparaison avec les autres tests statistiques décrits ci-dessus, la validation croisée s'appuie sur des hypothèses théoriques de distribution des données et non sur des calculs longs et complexes. Cette méthode consiste à écarter un sous-ensemble du jeu de composés initial et le modèle créé avec les composés restants

est utilisé pour prédire l'activité des molécules ignorées. On répète la procédure avec un nouveau sous-ensemble choisi au hasard et ceci jusqu'à ce que l'activité de chaque élément de la population de composés ait été prédite une fois. Cette procédure de validation croisée permet de suivre l'évolution du pouvoir prédictionnel du modèle.

La prédictibilité du modèle est évaluée au moyen du coefficient de validation croisée r_{cv}^2 (crossvalidated r squared) ou ''r^2 prédictif'' défini par analogie avec le coefficient de corrélation usuel défini ci-dessus. Le rapport R_{cv}^2 / R^2 représente l'erreur standard de prédiction du modèle QSAR obtenu.

Valeurs des descripteurs utilisés dans l'étude QSAR

	HEAT	DIPO	VDW	HOMO	LUMO	FORM	SAPO	SPHL	CHA	SDEL	SDHO	SAPA	SAPB	SAPC
B1	48.694	4.493	361.64	-.3159	-.0244	.21690	-189.6	-.0032	5.8025	-254.8	-6.339	-9.212	-71.27	-40.95
B2	-50.64	6.706	407.59	-.3150	-.0264	.38810	-193.9	-.8576	6.6974	-269.1	-6.349	-19.02	-66.19	-40.52
B3	37.485	2.191	359.84	-.3096	-.0179	.27730	-184.0	-.1109	5.2240	-248.5	-6.460	-9.220	-66.12	-40.52
B4	34.353	3.915	339.34	-.3134	-.0184	.45477	-176.0	-1.396	5.1239	-237.7	-6.381	-1.223	-66.13	-40.53
B5	72.534	1.704	368.37	-.3120	-.0228	.58211	-185.1	-.9674	5.2025	-250.9	-6.411	-9.220	-66.12	-40.96
B6	70.159	2.0433	342.94	-.3005	-.0244	.25900	-177.0	-.0847	4.8929	-242.3	-6.656	-1.225	-66.06	-40.96
B7	18.310	4.6026	361.72	-.3447	-.0237	.23382	-172.5	-62.29	5.0805	-248.4	-5.803	-9.213	-54.70	-40.52
B8	24.418	5.1039	382.50	-.3182	-.0251	.1726	-178.2	-2.079	5.2578	-251.8	-6.285	-9.211	-59.87	-40.95
B9	19.948	2.9351	381.58	-.3315	-.0221	.27690	-185.4	-.0083	5.2971	-252.2	-6.033	-9.215	-67.11	-40.95
B10	-28.75	3.0743	397.99	-.3257	-.0211	.27562	-191.1	-.0062	5.3193	-263.1	-6.140	-9.215	-67.18	-46.56
B11	-18.02	4.6916	401.73	-.3179	-.0187	.31611	-196.6	-.0289	5.7397	-274.8	-6.291	-9.216	-67.20	-52.16
B12	-16.36	3.9712	442.84	-.3274	-.0199	.27516	-209.8	-.2650	5.6950	-291.2	-6.108	-9.217	-67.20	-65.33
B13	-57.20	4.3496	292.74	-.3481	-.0338	.40488	-187.2	-53.46	6.4468	-260.9	-5.745	-9.216	-68.90	-40.95
B14	116.31	7.2796	421.54	-.3361	-.0322	.24836	-200.7	-39.90	5.8909	-278.0	-5.777	-9.209	-82.47	-40.94
B15	99.735	8.6776	466.78	-.3460	-.0407	.26407	-224.7	-.0002	6.5648	-308.3	-5.779	-33.19	-82.36	-40.95
B16	-40.65	3.6980	402.21	-.3129	-.0249	.26329	-199.4	-.1072	7.3750	-278.6	-6.390	-18.83	-71.45	-40.94
B17	1.6510	.1184	361.56	-.3126	-.0254	.23779	-181.2	-.0067	5.9270	-252.2	-6.398	-1.221	-71.42	-40.52
B18	-49.45	3.3471	276.71	-.3435	-.0245	.40985	-132.2	-73.80	4.1995	-190.4	-5.822	-1.219	-22.34	-40.52
B19	-19.60	4.7236	404.15	-.3305	-.0208	.34509	-188.4	-.3596	5.3295	-263.0	-6.051	-9.215	-67.17	-43.96
B20	-13.97	4.1775	356.39	-.3207	-.0206	.23687	-180.5	-.1618	5.1738	-253.0	-6.235	-9.214	-67.26	-36.00

	SAPD	SAPR	SHLB	SHLC	SHLR	CHAA	CHAB	CHAC	CHAD	CHAR	SDEA	SDEB	SDEC	SDED
B1	-1.292	-66.84	-.0015	.00001	-.0017	.24930	2.6186	1.0447	.11150	1.7784	-11.24	-100.7	-48.91	-1.430
B2	-1.292	–66.85	-.1883	.00001	-.6686	1.5232	2.0894	1.1158	.11163	1.8574	-36.08	-82.75	-56.66	-1.426
B3	-1.292	-66.88	-.6033	.00001	-.5052	.24488	2.0239	1.0829	.11088	1.7616	-11.46	-84.18	-57.02	-1.444
B4	-1.293	-66.87	-.4653	.00001	-.9310	.08660	2.0189	1.1180	.11011	1.7903	-1.365	-83.44	-57.34	-1.444
B5	-1.786	-67.01	-.2421	.00001	-.7240	.24725	2.0181	1.0640	.08028	1.7928	-11.37	-83.73	-49.55	-12.64
B6	-1.784	-66.97	-.0242	.00001	-.0601	.08522	1.9428	1.0393	.07923	1.7463	-1.364	-85.87	-49.44	-12.61
B7	-1.292	-66.83	-.0001	.00001	-62.28	.24871	1.8860	1.0869	.11178	1.7471	-11.26	-86.50	-56.58	-1.433
B8	-1.292	-66.84	-.0020	.00001	-.0186	.25049	2.0726	1.0464	.11163	1.7767	-11.24	-97.89	-48.83	-1.429
B9	-1.292	-66.81	-.0007	.00001	-.0074	.24258	2.0978	1.0504	.11110	1.7952	-11.32	-97.32	-49.08	-1.437
B10	-1.292	-66.89	-.0002	.00001	-.0059	.24294	2.0569	1.1392	.11097	1.7691	-11.33	-97.55	-59.52	-1.435
B11	-1.292	-66.71	-.0001	.00010	-.0218	.23817	2.1559	1.3526	.11071	1.8822	-11.39	-98.60	-69.25	-1.433
B12	-1.292	-66.76	-.0025	-.0342	-.2280	.24523	2.1172	1.4114	.11096	1.8102	-11.36	-97.41	-87.20	-1.439
B13	-1.292	-66.80	-.0299	-.0003	-53.38	.24994	3.1728	1.0557	.11215	1.8561	-11.27	-107.6	-48.73	-1.408
B14	-1.291	-66.85	-.7401	.00000	-.0139	.25292	2.8137	1.0473	.11261	1.7911	-11.14	-125.5	-48.47	-1.417
B15	-1.291	-66.91	-.0001	.00000	-.0001	.88821	2.6783	1.0684	.11220	1.8177	-41.57	-124.6	-48.63	-1.423
B16	-1.291	-66.85	-.1052	.00000	-.0019	1.5705	2.8392	1.0716	.11194	1.7818	-35.10	-100.4	-49.38	-1.428
B17	-1.292	-66.78	-.0064	.00000	-.0003	.09022	2.8254	1.0912	.11212	1.8080	-1.339	-100.5	-56.54	-1.432
B18	-1.292	-66.81	.00000	.00000	-73.80	.09082	1.0775	1.1049	.11122	1.8151	-1.342	-38.49	-56.60	-1.434
B19	-1.292	-66.72	-.0014	-.0007	-.2802	.24138	2.0828	1.0440	.11156	1.8499	-11.33	-97.30	-59.47	-1.435
B20	-1.292	-66.72	-.0001	-.0003	-.1343	.24467	2.1299	.86321	.11144	1.8246	-11.31	-97.74	-48.96	-1.433

	SDER	QA	QB	QC	QD	QR	LOGP	HMLM	POLV	ANIV	COSU	MOAR	COVO	MORE
B1	-92.52	.0881	.00003	.02630	.11150	-.2259	3.8100	-.2915	58.499	15.312	686.67	376.08	324.25	118.44
B2	-92.12	.0299	.0527	-.0017	.11163	-.1925	3.5500	-.2885	59.388	25.284	664.72	377.36	342.31	121.54
B3	-94.44	.0819	.03850	-.0054	.11088	-.2259	4.2800	-.2917	57.301	18.717	636.02	359.28	324.92	115.45
B4	-94.08	.08663	.04189	-.0071	.11011	-.2315	4.0500	-.2950	54.889	26.049	612.98	341.44	302.57	110.56
B5	-93.58	.08191	.08028	-.2215	.08028	1.7928	4.6800	-.2892	60.249	26.976	657.73	370.73	334.70	120.04
B6	-93.03	.08522	.06050	.02182	.07923	-.2468	4.4500	-.2761	57.283	25.837	634.81	351.55	311.70	115.15
B7	-92.61	.08762	.00023	.00255	.11178	-.2022	4.9300	-.3210	53.656	11.061	661.25	365.10	317.86	112.50
B8	-92.38	.08869	-.0109	.02664	.11163	-.2160	4.4400	-.2931	56.364	17.040	676.52	370.94	321.44	115.41
B9	-93.03	.08608	-.0079	.02482	.11110	-.2141	4.3700	-.3094	57.698	20.247	678.49	374.86	325.85	115.65
B10	-93.26	.08588	-.0127	.08837	.11097	-.2725	4.4500	-.3046	60.445	15.434	687.83	385.00	355.33	116.68
B11	-94.11	.08317	-.0240	.18623	.11071	-.3561	4.1300	-.2992	62.242	18.331	696.55	392.94	371.65	120.59
B12	-93.75	.08433	-.0151	.12725	.11096	-.3074	4.4300	-.3075	66.568	22.387	733.05	413.84	386.23	127.73
B13	-91.99	.08663	-.0398	.03093	.11215	-.1899	3.4400	-.3143	56.742	19.383	689.30	380.70	328.49	117.37
B14	-91.42	.09050	-.0168	.02866	.11261	-.2150	3.4000	-.3039	60.040	20.924	665.89	380.63	360.44	130.05
B15	-92.06	.09525	-.0105	.02948	.11220	-.2265	4.6200	-.3053	68.232	17.207	772.32	437.62	395.78	143.79
B16	-92.28	.04030	-.0006	.03180	.11194	-.1835	3.0700	-.2880	60.414	17.392	695.30	396.43	352.30	124.53
B17	-92.31	.09022	-.0007	.00154	.11212	-.2032	3.7300	-.2872	54.862	16.710	601.19	334.08	312.59	113.76
B18	-92.58	.09082	-.0002	.00111	.11122	-.2029	2.2300	-.3190	39.555	12.754	508.36	272.34	239.37	82.630
B19	-93.51	.08519	-.0143	.13301	.11156	-.3154	3.7100	-.3097	59.231	17.032	689.25	384.45	351.42	115.99
B20	-93.54	.08591	-.0119	.14651	.11144	-.3320	3.3000	-.3001	56.462	18.706	679.69	374.58	327.92	111.39

	QNA	QNI	QO
B1	.02604	-.0985	-.3642
B2	.02990	-.0877	-.3936
B3	.00767	-.0988	-.3414
B4	.04830	-.0880	-.3528
B5	.01123	-.0879	-.3453
B6	.02796	-.0772	-.3511
B7	.02038	-.0892	-.3512
B8	.02615	-.1009	-.3565
B9	.00562	-.1327	-.3320
B10	.00765	-.1449	-.3313
B11	.00384	-.2694	-.3356
B12	.01376	-.1984	-.3465
B13	.01058	-.1408	-.3189
B14	.02958	-.1044	-.3594
B15	.03139	-.1072	-.3762
B16	.03200	-.1074	-.3543
B17	.03342	-.1274	-.3287
B18	.04293	-.1303	-.3334
B19	.00395	-.2346	-.3347
B20	.01741	-.2078	-.3579

ANNEXE 3

SURFACES MOLECULAIRES ET MODELE DE CALCUL DU POTENTIEL ELECTROSTATIQUE MOLECULAIRE

1) – Introduction

A l'échelle microscopique, les molécules sont composés d'atomes et chacun de ces atomes est formé d'un noyau et d'un nuage d'électrons gravitant autour de ce noyau. La surface moléculaire est un modèle qu'il faut utiliser avec prudence, car il dérive de considérations macroscopiques et néglige les effets quantiques. Néanmoins, ce modèle est très utile pour interpréter la taille et la forme d'une molécule, à condition d'être conscient de ses limites. En effet, une molécule n'a pas de surface au sens strict, et ses régions périphériques sont en fait représentées par une fonction à 3 dimensions de densité de charge électronique qui tend évidemment vers zéro loin de la molécule. De plus, le principe d'incertitude d'Heisenberg ($\Delta x \ \Delta p \geq \hbar$) s'applique aux électrons gravitant autour des noyaux et dans une moindre mesure aux noyaux. Même en supposant les positions des noyaux fixes, on ne peut pas dire que le nuage électronique a réellement une frontière précise.

Cependant, on peut construire des modèles de surface moléculaire en supposant que cette surface englobe la partie dite essentielle de la molécule. Différentes précisions de cette partie essentielle, qui conduisent à différentes surfaces moléculaires, sont présentées ci-dessous dans un ordre de précision croissant.

2) – Surface de van der Waals

Par analogie avec des objets macroscopiques à symétrie sphérique, on modélise chaque atome par une sphère, appelée sphère de van der Waals et dont le rayon est fonction du nombre de couches d'électrons. Lorsque les orbitales de deux atomes se recouvrent, on dit qu'ils forment une liaison covalente et leurs sphères de van der Waals ont alors une intersection non vide. Les parties des sphères qui ne se recouvrent pas forment la surface de van der Waals.

L'avantage de cette surface réside dans sa simplicité de calcul et de visualisation, en particulier pour de grandes molécules. Son inconvénient est qu'elle fait intervenir tous les atomes dont certains peuvent être cachés à l'intérieur de la molécule et de ce fait ne

participent pas directement aux interactions avec l'environnement. En fait, ces atomes forment un squelette dont la fonction est de maintenir la conformation des parties actives de la molécule.

3) – Surface de Lee-Richards ou de Connolly

B. Lee et F.M. Richards ont défini la surface décrite par le centre d'une sphère test représentant une molécule de solvant et roulant sur la surface de van der Waals de manière à lui être tangente, mais sans qu'il y ait recouvrement c'est à dire que l'intersection de la sphère test avec l'intérieur de la surface de van der Waals est toujours vide. Cette surface est exactement l'union des sphères de van der Waals agrandies du rayon de la sphère test. L'avantage de cette surface est que seules les régions des atomes, entrant en contact avec les molécules de solvant, contribuent à définir cette surface.

4) – Surface accessible au solvant

F.M. Richards a défini ensuite une surface moléculaire à partir des points de contact entre la sphère test et les atomes. Cette surface accessible au solvant est couramment appelée surface de Connolly qui est l'auteur d'une description analytique détaillée des parties composant cette surface. L'avantage de cette surface est d'être lisse sauf en d'éventuels points singuliers et aussi de donner une bonne appréciation de la forme de la molécule.

5)-Isosurfaces de la densité électronique

La densité électronique d'une molécule est naturellement reliée à sa forme et est responsable des propriétés physiques de ses régions périphériques. Cette fonction converge rapidement vers zéro (3 à 4 Å). On définit l'isosurface de la densité électronique G(**a**) de la façon suivante :

$$G(\mathbf{a}) = \{r : \rho(r) = \mathbf{a}\} \qquad (1)$$

ρ étant une fonction continue. Il est intéressant de constater que la surface obtenue de van der Waals ressemble fortement à l'isosurface de la densité électronique pour une valeur moyenne de **a**.

G(**a**) varie continûment en fonction de **a**, mais seul un nombre fini de valeurs de **a** correspondent à un changement de connectivité dans G(**a**).

- pour **a** grand, G(**a**) est complètement déconnecté et est composé de fragments sphériques autour des noyaux.

- pour **a** petit, G(**a**) est comme un grand ballon rond englobant tout.

Lorsque a augmente, les différents fragments sont progressivement connectés. Un fragment connecté pour un **a** grand représente une liaison d'énergie plus forte que pour un **a** petit.

6) – Isosurfaces du potentiel électrostatique moléculaire

Depuis la définition du potentiel électrostatique moléculaire (PEM) dans les années 1970, il est devenu un des outils les plus valables pour rationaliser et prédire la réactivité des molécules et également les interactions inter-moléculaires. Plusieurs approches ont été développées dans la littérature pour calculer le PEM.

Nous présentons brièvement ci-dessous la théorie utilisée dans le programme DELPHI et qui nous a servi dans le cadre de notre étude (Chapitre V).

a-Modèle classique

Dans le cadre de la théorie classique, les matériaux sont considérés comme des milieux diélectriques homogènes, pouvant être polarisés par des charges électriques. Plutôt que de tenir compte explicitement de la polarisation de chaque atome, une constante diélectrique est donc utilisée comme une mesure approchée de la polarisabilité du milieu et qui est appelé le modèle du *continuum.*

La description la plus simple de ce modèle consiste à utiliser la loi de Coulomb. Dans ce cadre, l'énergie électrostatique entre 2 charges i et j est donnée par :

$$G_{ij} = q_i q_j / \varepsilon\, r_{ij} \qquad (2)$$

Le potentiel sur l'atome i, dû à une charge sur l'atome j, est donné par :

$$\Phi_i = q_j / \varepsilon\, r_{ij} \qquad (3)$$

q_j est la charge sur l'atome j, ε la constante diélectrique et r_{ij} la distance entre les atomes i et j.

b) - Equation de Poisson-Boltzmann

Une molécule telle qu'une protéine a une constante diélectrique faible puisque ses groupements dipolaires sont gelés dans un réseau de liaison hydrogène assurant la structure de la protéine. Ces dipôles ne peuvent pas se réorienter suite à un champ électrostatique externe. En fait, une valeur de 2 mesure sa réponse de polarisation électronique et une valeur proche de 4 inclue quelques contributions de réorientations dipolaires. En revanche, l'eau a une très grande constante diélectrique (78) puisque ses dipôles se réorientent librement. Par

conséquent, une molécule en solution aqueuse conduit à un système avec deux milieux diélectriques très différents. Les effets de cette grande différence peuvent être considérables et devraient être pris en compte par un modèle électrostatique. L'équation de Poisson-Boltzmann que nous présentons ici, s'avère un bon modèle pour décrire les molécules en solutions aqueuses.

Le champ électrostatique (E) et le déplacement électrostatique D qui sont fonctions de r sont définis comme suit :

$$E(r) = -\nabla \Phi(r) \qquad (4)$$

$$D(r) = \varepsilon(r)\, E(r) \qquad (5)$$

$$D(r) = -\varepsilon(r) \nabla \Phi(r) \qquad (6)$$

$\Phi(r)$ est le potentiel électrostatique, $\varepsilon(r)$ la fonction diélectrique, r le vecteur position (x, y, z) et ∇ l'opérateur différentiel.

Dans un système constitué de charges et d'un milieu diélectrique inhomogène, l'équation de Poisson relie le déplacement électrostatique à la densité de charge $\rho(r)$:

$$\nabla \; D(r) = 4\pi\rho(r) \qquad (7)$$

Ce qui conduit à : $$\nabla [\varepsilon(r) \nabla \Phi(r)] + 4\pi\rho(r) = 0 \qquad (8)$$

La densité de charge $\rho(r)$ peut être séparée en 2 composantes :

$$\rho = \rho_{int} + \rho_{ext} \qquad (9)$$

ρ_{int} est une ditribution de charge intérieure des positions fixées de toutes les charges dans la molécule et ρ_{ext} est une densité de charge extérieure, qui représente la redistribution des ions en solution en réaction au potentiel électrostatique.

Cette densité de charge peut être modélisée au moyen de la distribution de Boltzmann, qui conduit à :

$$4\pi\rho_{ext} = (-\acute{\kappa}^2) \sinh [\Phi(z)] = (-\acute{\kappa}^2)\, \Phi(r)\, [1 + \Phi(r)^2/6 + \Phi(r)^4/120 + \ldots.] \qquad (10)$$

où : $\acute{\kappa} = (\varepsilon)^{1/2} \kappa = (8\pi e^2 N_A I / 1000 kT)^{1/2}$

ε est la constante diélectrique du solvant, Φ le potentiel électrostatique en unités kT/e, e la charge électronique, NA le nombre d'Avogadro, I la force ionique (moles/litre), k la constante de Boltzmann, T la température et κ l'inverse de la longueur de Debye-Huckel.

Si la densité de charge de la molécule et la force ionique de la solution sont faibles, cette densité peut être approchée par l'équation linéaire suivante

$$4\pi\rho_{ext} = -\acute{\kappa}^2\, \Phi(r) \qquad (11)$$

La substitution de ces expressions dans l'équation de Poisson (8) conduit aux formes linéaire (12) et nonlinéaire (13) de l'équation de Poisson-Boltzmann :

$$\nabla [\varepsilon(r) \nabla \Phi(r)] - \acute{\kappa}_0^2 \Phi(r) + 4\pi\rho_{int}(r) = 0 \qquad (12)$$

$$\nabla [\varepsilon(r) \nabla \Phi(r)] - \acute{\kappa}_0^2 \Phi(r) [1 + \Phi(r)^2/6 + \Phi(r)^4/120 + \ldots.] + 4\pi\rho_{int}(r) = 0 \qquad (13)$$

avec $\acute{\kappa}_0 = 0$, à l'intérieur de la molécule et $\acute{\kappa}_0 = \acute{\kappa}$, dans le solvant.

Malheureusement, cette équation de Poisson-Boltzmann peut être résolue analytiquement seulement dans le cas des systèmes ayant des formes simples, tels que les sphères ou les plans. Comme, la plupart des molécules intéressantes ont des formes complexes et leurs conformations ont un effet significatif sur les propriétés électrostatiques résultantes, une alternative aux solutions analytiques pour trouver une solution approchée est l'usage des techniques numériques.

Le programme Delphi utilise **'**the Finite Difference Approximation**'** la méthode des différences finies. Cette méthode implique le 'mappage*'* de la molécule à l'intérieur d'une grille cubique, les points de la grille sont espacés d'une longueur h et le potentiel électrostatique est calculé en chaque point de la grille.