



HAL
open science

**Conception de nouvelles molécules à activité
sérotoninergique par des méthodes QSAR et des études
de dynamique moléculaire de complexe ligands
/récepteur**

Christophe Marot

► **To cite this version:**

Christophe Marot. Conception de nouvelles molécules à activité sérotoninergique par des méthodes QSAR et des études de dynamique moléculaire de complexe ligands /récepteur. Autre. Université d'Orléans, 1995. Français. NNT : . tel-00148822

HAL Id: tel-00148822

<https://theses.hal.science/tel-00148822>

Submitted on 23 May 2007

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

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



THESE PRESENTEE A L'UNIVERSITE D'ORLEANS



Région Centre

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITE D'ORLEANS

Discipline : Chimie et Physicochimie des Composés d'Intérêt Biologique

PAR

Christophe MAROT

Conception de Nouvelles Molécules à Activité Sérotoninergique par des Méthodes QSAR et des Etudes de Dynamique Moléculaire de Complexes Ligand/Récepteur.

SOUTENUE PUBLIQUEMENT LE 29 Novembre 1995 à 15 heures - Salle Charles Sadron du CNRS.

MEMBRES DU JURY

<u>Président</u>	D. LESIEUR	(Professeur, Faculté de pharmacie de Lille)
<u>Rapporteurs</u>	G. GRASSY	(Professeur, UMR CNRS 9955-INSERM U414, Université de Montpellier)
	S. PEREZ	(Directeur de Recherche, INRA de Nantes)
<u>Examineurs</u>	M. GENEST	(Chargé de Recherche, CBM, CNRS Orléans)
	G. GUILLAUMET	(Professeur, ICOA URA CNRS 499, Université d'Orléans)
	L. MORIN-ALLORY	(Professeur, ICOA URA CNRS 499, Université d'Orléans)

Désireux de mettre au point de nouvelles substances pouvant présenter une activité biologique potentielle au niveau du Système Nerveux Central, nous avons envisagé d'étudier les Relations Quantitatives Structure-Activité (QSAR) sur plusieurs familles de composés 5-HT_{1A}. L'objectif était de prédire l'activité et la sélectivité biologique de nouveaux ligands 5-HT_{1A} vis-à-vis de certains neurorécepteurs transmembranaires (5-HT_{1A}, α_1 , α_2 , D₂). Pour ce faire, 382 composés issus de différentes familles chimiques (indole, tétraline, chromane, thiochromane, pipérazine, etc...) ont été modélisés et superposés aux pharmacophores 5-HT_{1A}. Nous avons ensuite, pour chaque molécule, calculé et comparé différents descripteurs moléculaires représentatifs de l'aspect électrostatique, lipophile, stérique et topologique. Les différentes tables constituées des composés et des descripteurs ont ensuite été étudiées par des analyses statistiques comme l'analyse en composantes principales, l'analyse discriminante, la régression multiple et l'analyse Partial Least Square. A partir des meilleurs modèles statistiques, l'activité biologique de nouveaux ligands a été prédite. Après la synthèse et les tests pharmacologiques, ces prédictions se sont révélées précises (écart moyen de 1 unité de pIC₅₀). En parallèle, une étude de dynamique moléculaire du récepteur sérotoninergique transmembranaire 5-HT_{1A} a été entreprise afin de rendre compte de l'importance fonctionnelle de ces ligands au niveau des interactions électrostatiques, stériques et lipophiles mises en jeu au sein du site actif du récepteur. Ces études sur un nombre limité de complexes ligand/récepteur ont aussi été menées afin de comprendre le rôle de certaines parties fonctionnelles de ces ligands, rôle mal défini jusqu'alors par les autres techniques. Nous avons aussi obtenu des informations qualitatives sur les interactions et la mobilité de certaines parties du récepteur, entre autres sur les hélices transmembranaires TMH5, 6 et 7.

INTRODUCTION GENERALE

La sérotonine (5-hydroxytryptamine, 5-HT) est une amine biogène découverte dans les cellules entérochromaffines de l'intestin par Vialli et Erspamer¹ en 1933. Elle a été mise en évidence dans le système nerveux central (SNC) des mammifères il y a une quarantaine d'années. La sérotonine est définie comme le facteur vasoconstricteur sérique relargué par les plaquettes sanguines ; son nom est alors choisi d'après sa présence dans le sérum et ses propriétés vasotoniques. Sa structure moléculaire (Figure 1) est déterminée en 1948 par Rapport², et sa synthèse chimique en 1951 par Hamlin et Fischer³ marque l'essor de l'étude intensive du système nerveux central. Cet effort d'investigation et de compréhension du système nerveux central est toujours d'actualité avec les récents travaux du prix Nobel 1994 de médecine et de physiologie décerné à Alfred Gilman et Martin Rodbeil, concernant la transmission des signaux au sein des cellules *via* les protéines G couplées aux récepteurs transmembranaires du système nerveux central.

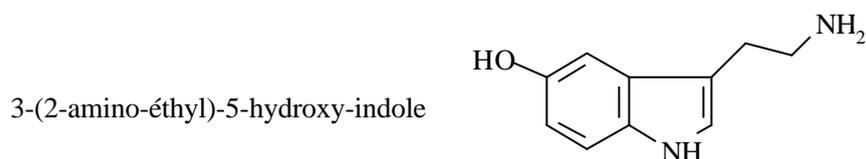


Figure 1. Représentation de la sérotonine.

Rapidement, la démonstration de l'hétérogénéité de la distribution de la 5-HT au niveau cérébral a conduit à suggérer que cette amine y jouait le rôle de neuromédiateur. Depuis, les recherches ont progressé par bonds, suivant en cela l'avènement de nouvelles techniques biochimiques et biologiques, anatomiques, électrophysiologiques, ainsi que le développement de nouveaux outils pharmacologiques et informatiques pour la modélisation moléculaire. Ce neurotransmetteur synthétisé et libéré dans des neurones spécifiques intervient dans le contrôle de nombreuses fonctions cérébrales telles que les cycles veille-sommeil, la thermorégulation, la mémoire, la migraine, les comportements de l'appétit et de la satiété, le comportement sexuel, la

nociception, etc...⁴. Et il est généralement associé à des désordres neurologiques ou psychiatriques tels que les dépressions, les démences (y compris la maladie d'Alzheimer) et l'anxiété⁵.

Cette pluralité d'interventions de la sérotonine est très probablement en rapport avec la multiplicité de ses récepteurs (plus de 15), certains d'entre eux étant couplés à plusieurs systèmes effecteurs. D'ailleurs, dès 1957, Gaddum et Picarelli⁶ apportaient les premiers arguments expérimentaux en faveur de l'existence d'au moins deux types de récepteurs pour la 5-HT. Depuis, la situation a considérablement évolué puisque l'application des techniques de la biologie moléculaire a permis le clonage et le séquençage des ADNc d'une quinzaine de récepteurs distincts (5-HT_{1A}, 5-HT_{1B/1Dβ}, 5-HT_{1Dα}, 5-HT_{1E}, 5-HT_{1F}, 5-HT_{dro2A}, 5-HT_{dro2B} pour la classe 5-HT₁ ; 5-HT_{2A}, 5-HT_{2B}, 5-HT_{2C} pour la classe 5-HT₂ ; 5-HT₄, 5-HT₆, 5-HT₇, 5-HT_{dro1} pour la classe des récepteurs stimulant la sécrétion de l'adénylate cyclase ; 5-HT_{5A}, 5-HT_{5B} pour la classe 5-HT₅ et le récepteur à canal ionique 5-HT₃). Cette classification a été révisée plusieurs fois durant les dix dernières années^{7,8,9,10,11,12,13,14}.

Comme pour d'autres neuromédiateurs (noradrénaline, dopamine, acétylcholine, histamine, prostaglandine, endorphine, etc...), la sérotonine agit sur des récepteurs qui sont couplés aux protéines G ou bien qui constituent un canal ionique (5-HT₃, 5-HT₅ ?). De cette pluralité des fonctions du système sérotoninergique central s'exprimant au travers d'une multiplicité de récepteurs est venue l'idée que la synthèse de ligands spécifiques de telle ou telle classe de ces récepteurs puisse avoir un intérêt dans le traitement d'une ou plusieurs pathologies parfaitement ciblées. Les données acquises au cours des dix dernières années semblent confirmer ce concept puisqu'il existe aujourd'hui des ligands agissant préférentiellement sur une classe donnée de récepteurs de la 5-HT et dont le domaine d'application thérapeutique s'avère plus délimité. Parmi les sous-types de récepteurs 5-HT, le site 5-HT_{1A} qui a été cloné le premier est le mieux connu grâce à l'utilisation de la 8-hydroxy-2-di-*n*-propylaminotétraline (8-OH-DPAT)¹⁵ ligand particulièrement adapté en raison de sa grande sélectivité et sa grande spécificité vis-à-vis de ce site^{16,17}.

Dans le but de trouver de nouvelles molécules radioactives pour le marquage spécifique des sites 5-HT_{1A}, Cossery *et al.*¹⁸, encouragés par les travaux antérieurs de Sarda *et al.*¹⁹, ont choisi de préparer des isostères oxygénés des dérivés hydroxyles de la di-*n*-propylaminotétraline. Parmi ces molécules, le 5-méthoxy-3,4-dihydro-3-di-*n*-propylamino-2*H*-1-benzopyrane ou 5-méthoxy-3-di-*n*-propylamino chromane (5-MeO-DPAC) s'est avéré présenter une excellente affinité pour les sites

actifs des récepteurs 5-HT_{1A} dans des conditions d'études *in vitro* et ce, sur des membranes de diverses structures cérébrales chez le rat. De surcroît les résultats obtenus avec ce ligand ont démontré que par rapport à la 8-OH-DPAT, molécule de référence, le 5-MeO-DPAC présentait une activité vis-à-vis des sites récepteurs 5-HT_{1B} beaucoup plus faible que la 8-OH-DPAT confirmant ainsi la plus grande sélectivité de ce composé¹⁸. De tels composés sélectifs de sites 5-HT_{1A} (Figure 2) sont, comme le confirment les travaux de Traber²⁰ et Glennon²¹, susceptibles de posséder une importance thérapeutique dans le traitement de l'anxiété et, de ce fait, de remplacer les benzodiazépines certes très utilisées mais aux effets secondaires bien connus²².

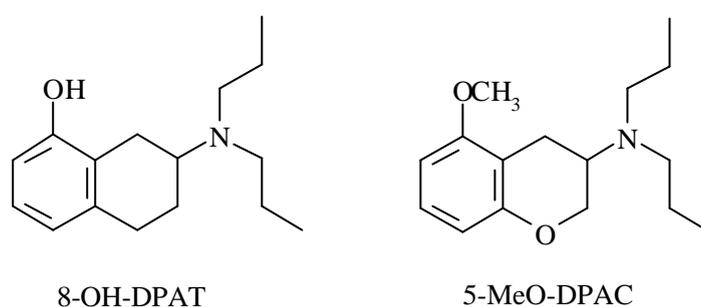


Figure 2. Représentation de deux composés très sélectifs du site 5-HT_{1A}.

De nombreux composés présentant un intérêt certain pour le récepteur 5-HT_{1A} ont été élaborés durant les dix dernières années par des méthodes de pharmacomodulation ou de modélisation moléculaire. Les méthodes les plus couramment utilisées sont la conception 3D de ligands à l'aide de pharmacophore, les études SAR et les études QSAR. Une autre approche qui devrait permettre la conception de nouvelles molécules potentiellement actives semble être l'étude des complexes ligand/récepteur 5-HT_{1A} par mécanique et dynamique moléculaire.

Le chapitre 2 de ce mémoire sera consacré, entre autres, à la description de ces méthodes et à la présentation de certains travaux publiés dans le domaine des pharmacophores et dans le domaine QSAR.

Désireux de mettre au point de nouvelles substances capables de présenter une activité au niveau du SNC, nous avons opté pour l'étude des Relations Quantitatives Structure-Activité (QSAR) de plusieurs familles de composés 5-HT_{1A} et la modélisation de ligands 5-HT_{1A} possédant la structure de base chromane et thiochromane. Ces nouveaux ligands ont été proposés à la synthèse

à partir d'études sur les pharmacophores 5-HT_{1A} et d'études QSAR entreprises avec les nombreux composés référencés au Laboratoire. Nous nous sommes efforcés de prédire l'activité dite de « binding », de même que la sélectivité vis-à-vis de plusieurs récepteurs transmembranaires (5-HT_{1A}, α_1 , α_2 , D₂). En parallèle, une étude de dynamique moléculaire du récepteur sérotoninergique transmembranaire 5-HT_{1A} a été envisagée afin de rendre compte de l'importance fonctionnelle de ces ligands au niveau des interactions de type électrostatique, lipophile et stérique pour les complexes ligand/récepteur. Ces études sur un nombre limité de complexes ligand/récepteur ont été entreprises afin d'analyser les interactions spécifiques et de comprendre la dynamique des ligands au sein du site actif et le rôle de certaines parties fonctionnelles de ces ligands, rôle mal défini jusqu'alors par les autres techniques.

Nous présenterons dans le chapitre 3 de ce mémoire les résultats des études QSAR entreprises sur les différentes familles de composés 5-HT_{1A} (indole, tétraline, chromane, thiochromane et un ensemble de 382 ligands) et dans le chapitre 4 les études de mécanique et dynamique moléculaire des complexes ligand/récepteur 5-HT_{1A} avec des ligands potentiels.

CHAPITRE 1 : LES RECEPTEURS SEROTONINERGIQUES : LEURS FONCTIONS
BIOLOGIQUES, LEURS LIGANDS

LA SEROTONINE

-

LES RECEPTEURS TRANSMEMBRANAIRES DU SYSTEME NERVEUX CENTRAL

-

POTENTIALITES THERAPEUTIQUES DES LIGANDS DES RECEPTEURS 5-HT

-

LES LIGANDS 5-HT_{1A} ET LEURS APPLICATIONS PHARMACOLOGIQUES

-

LES DONNEES PHARMACOLOGIQUES

I. LA SEROTONINE

I.1 - La sérotonine et ses effets physiologiques et pathologiques.

Un ligand peut être caractérisé par sa capacité à se lier de façon spécifique à un récepteur, qu'il s'agisse de neurorécepteurs ou d'hormones. Le site se définit comme l'endroit du récepteur où ces interactions de type électrostatique, lipophile et stérique ont lieu. La sérotonine ou 5-hydroxytryptamine (5-HT) joue un rôle capital au niveau du système nerveux central tant sur le plan physiologique que pathologique. En effet, ce neuromédiateur se trouve impliqué dans des fonctions aussi diverses que le sommeil, l'hypertension, la thermorégulation, la nociception ou l'activité sexuelle. Il a aussi pu être démontré que chez l'homme, l'activité des neurones sérotoninergiques est anormale dans certains troubles psychiatriques tels que la dépression, l'anxiété sous ses différentes formes et la démence d'Alzheimer²³.

Comme tous les neuromédiateurs, la sérotonine exerce ses effets physiologiques via des récepteurs distincts (Schéma 1). Les différents sous-type de récepteurs 5-HT₁ sont caractérisés par une haute affinité pour la sérotonine (Kd nanomolaire), par opposition aux récepteurs 5-HT₂ et 5-HT₄, dont l'affinité pour le ligand endogène est environ mille fois plus faible. Outre leurs propriétés pharmacologiques respectives, les mécanismes de transduction associés à ces deux derniers récepteurs permettent de les différencier aisément.

Les récepteurs 5-HT₂ sont couplés par l'intermédiaire d'une protéine Gq à la phospholipase C, alors que le récepteur 5-HT₄ interagit avec une protéine Gs pour activer l'adénylate cyclase. Ce récepteur avec les récepteurs 5-HT₆, 5-HT₇ et 5-HT_{dro1} provoque une stimulation qui induit une augmentation des taux intracellulaires d'AMP cyclique²⁴.

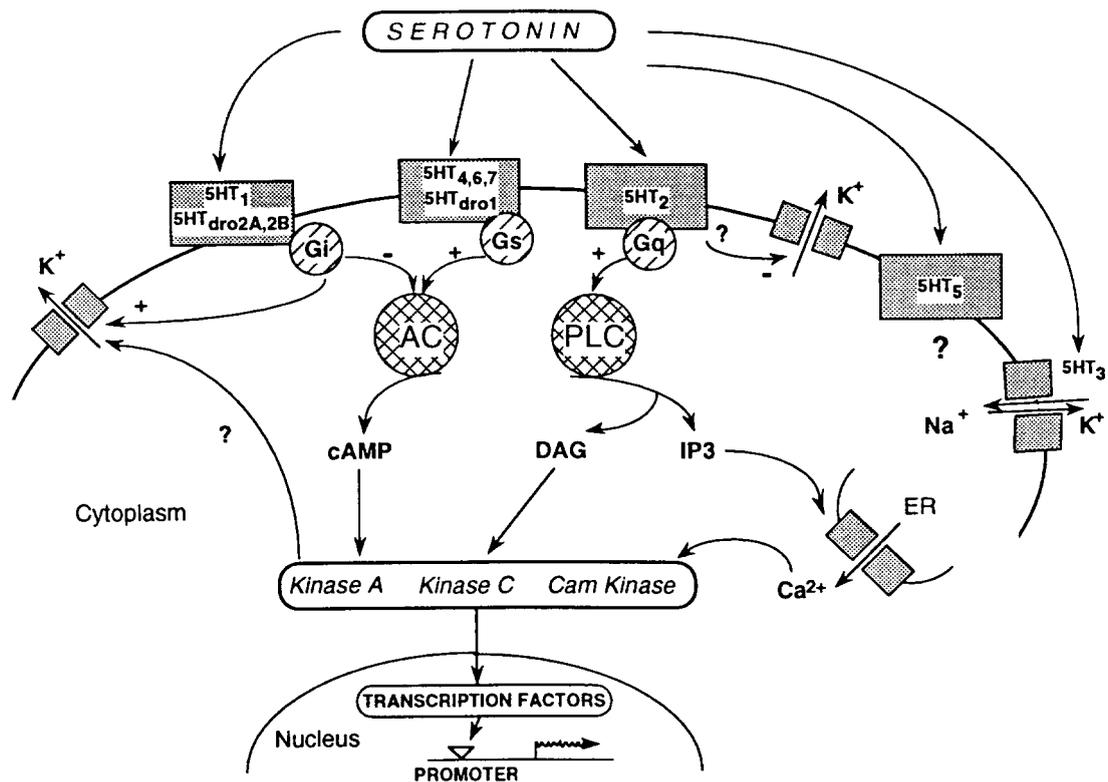


Schéma 1. Les récepteurs 5-HT et leurs systèmes effecteurs. Gi, Gq et Gs sont les protéines G. AC : adénylate cyclase, PLC : phospholipase C, cAMP : 3'-5'-adénosine monophosphate, DAG : 1-2-diacylglycérole, IP₃ : inositol 1,4,5-trisphosphate, ER : réticulum endoplasmique, Kinase A : protéine kinase cAMP dépendante, Kinase C : protéine kinase calcium-lipido dépendante, Cam : protéine kinase calcium-calmodulino dépendante²⁷.

En effet, celle des récepteurs 5-HT_{1A}, 5-HT_{1B/1Dβ}, 5-HT_{1Dα}, 5-HT_{1E}, 5-HT_{1F}, 5-HT_{dro2A} et 5-HT_{dro2B} induit au contraire une diminution de l'activité adénylate cyclase suite à l'activation d'une protéine Gi²⁵. Dans le groupe des récepteurs 5-HT₁, l'ancien type 5-HT_{1C} renommé 5-HT_{2C} se distingue de tous les autres 5-HT₁ par son système de transduction puisqu'il est couplé, comme le récepteur 5-HT_{2A}, à la phospholipase C. D'ailleurs, ces deux récepteurs 5-HT_{2C} et 5-HT_{2A} sont très proches tant au niveau de leurs propriétés pharmacologiques que de leurs caractéristiques moléculaires²⁶. A côté de ces couplages à des systèmes du type second messager, les récepteurs de la 5-HT dont la stimulation provoque l'activation de protéines G sont également couplés à des conductances ioniques (Tableau 1). C'est ainsi, par exemple, que l'activation de protéines G différentes provoque tantôt l'ouverture d'une conductance potassique (couplée au récepteur 5-HT_{1A})

tantôt au contraire la fermeture de cette même conductance spécifique du K⁺ (couplée aux récepteurs 5-HT_{2A} et 5-HT_{2C}).

Familles	Sous Types	Couplages	Localisations
5-HT ₁	5-HT _{1A}	Adénylate cyclase (-) Canal K ⁺ (+) Canal Ca ²⁺ (-)	Structures limbiques, noyaux du raphé Structures extrapyramidales
	5-HT _{1B/1Dβ}	Adénylate cyclase (-)	Hippocampe, raphé dorsal
	5-HT _{1Dα}	Adénylate cyclase (-)	Cortex
	5-HT _{1E}	Adénylate cyclase (-)	Cortex, raphé dorsal
	5-HT _{1F}	Adénylate cyclase (-)	
	5-HT _{dro2A}	Adénylate cyclase (-)	
	5-HT _{dro2B}	Adénylate cyclase (-)	
5-HT ₂	5-HT _{2A}	Phospholipase C (+) Canal K ⁺ (-)	Cortex frontal, hippocampe
	5-HT _{2B}	Phospholipase C (+) Canal K ⁺ (-)	Estomac
	5-HT _{2C}	Phospholipase C (+) Canal K ⁺ (-)	Plexus choroïdes, hippocampe
Récepteurs stimulant l'adénylate cyclase	5-HT ₄	Adénylate cyclase (+)	Collicule, Hippocampe, Substance noire
	5-HT ₆	Adénylate cyclase (+)	Hypothalamus, Thalamus
	5-HT ₇	Adénylate cyclase (+)	Coeur, intestin, Hypothalamus
	5-HT _{dro1}	Adénylate cyclase (+)	
5-HT ₅	5-HT _{5A}	Canal ionique ?	Hypothalamus, Hippocampe
	5-HT _{5B}		Hippocampe, raphé dorsal
5-HT ₃	5-HT ₃	Canal Na ⁺ /K ⁺	Cortex entorhinal, Amygdale

Tableau 1. Principales caractéristiques des récepteurs centraux de la sérotonine.

Les corps cellulaires des neurones sérotoninergiques sont essentiellement distribués dans les noyaux du raphé et en particulier dans la zone médiane du tronc cérébral (schéma 2). Les axones se

projetent dans la partie antérieure du cerveau et innervent toutes les structures centrales, aussi bien l'hippocampe, le striatum, le cortex que l'hypothalamus.

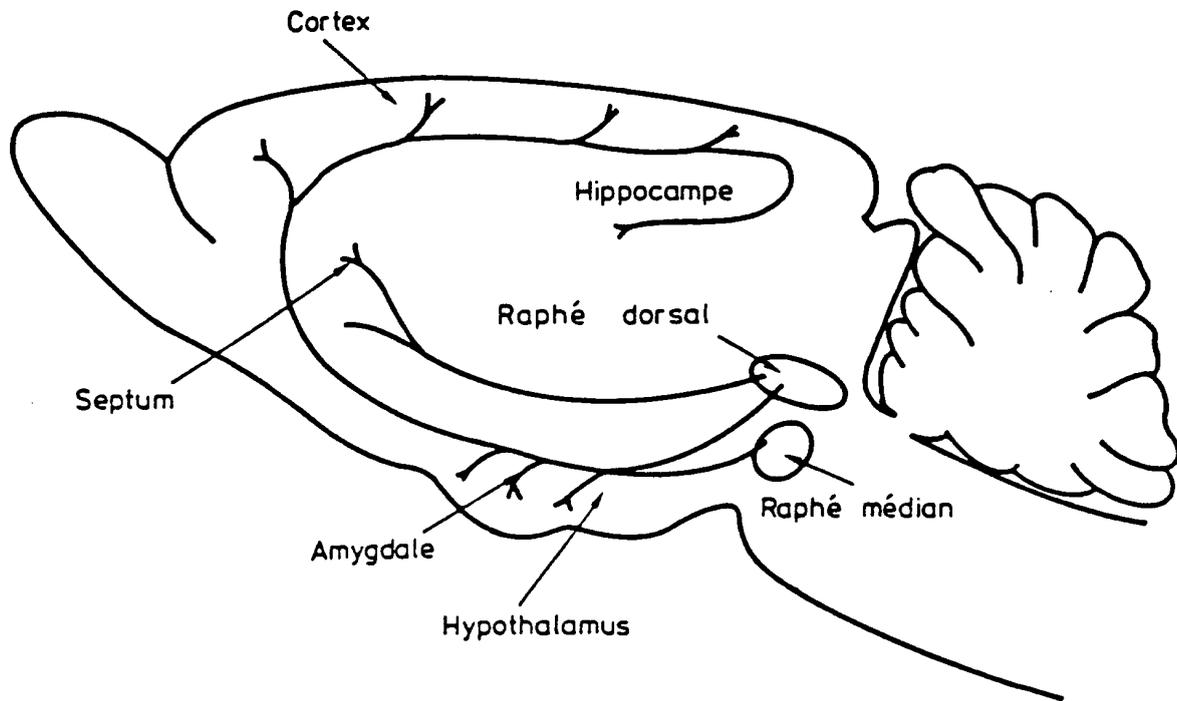


Schéma 2. Représentation schématique des plus importantes projections sérotoninergiques.

I.2 - Biosynthèse et catabolisme de la sérotonine.

La sérotonine (5-HT) est avec les catécholamines (dopamine, noradrénaline et adrénaline) une des amines biogènes la plus étudiée du système nerveux central.

Elle est synthétisée dans les neurones par hydroxylation et décarboxylation du tryptophane (Figure 3). La synthèse n'a pas lieu dans les plaquettes sanguines mais celles-ci sont capables d'activer la libération de la sérotonine à partir des granules de stockage lors de leur agrégation.

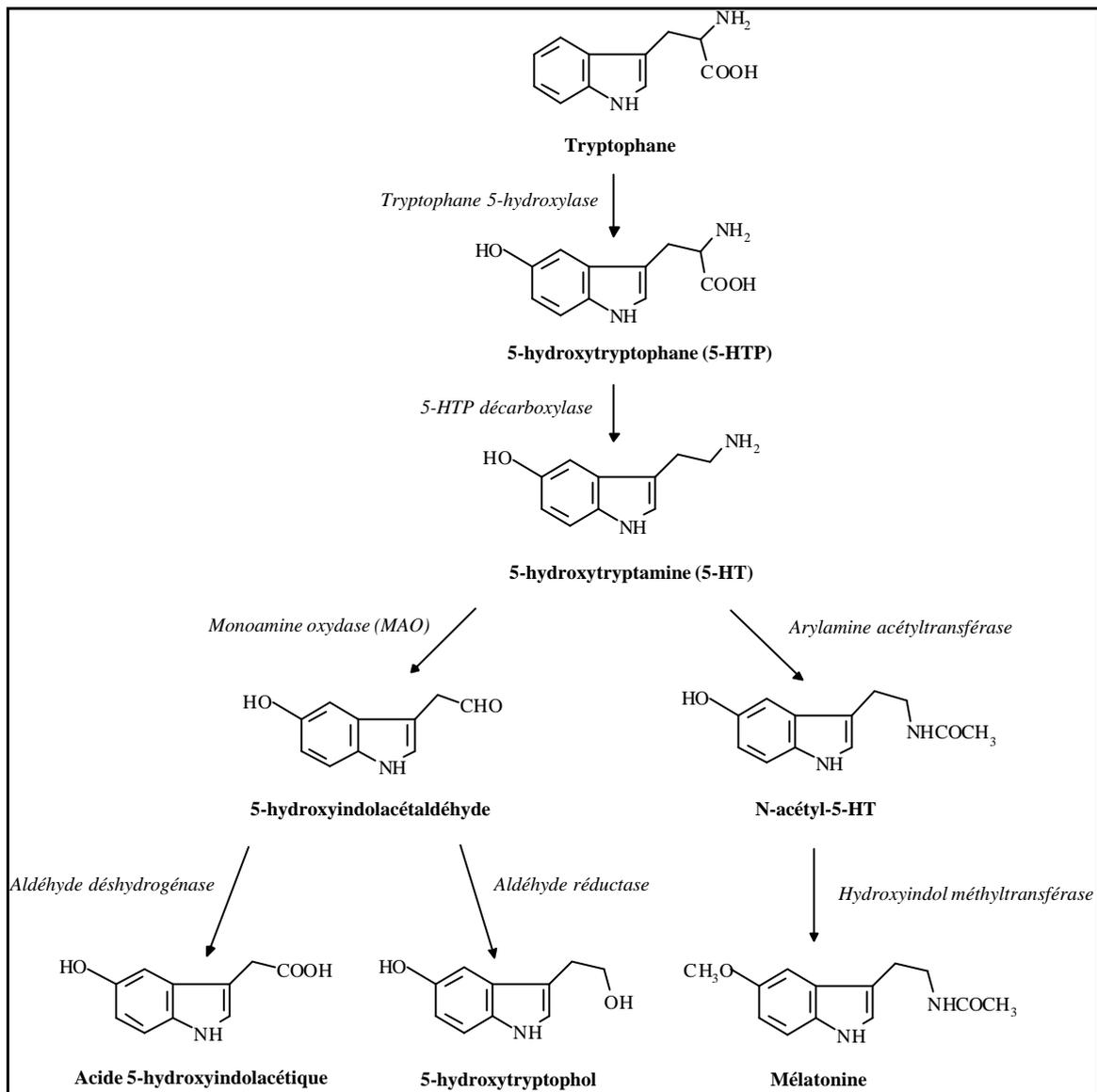


Figure 3. Biosynthèse et catabolisme de la sérotonine.

La biosynthèse de la sérotonine peut être bloquée par des composés comme la p-chlorophénylalanine, inhibiteur de la tryptophane hydroxylase. La dégradation de la sérotonine peut être bloquée par les inhibiteurs de la monoamine oxydase (M.A.O), capables de faire doubler la concentration de sérotonine dans le cerveau en moins d'une heure.

II. LES RECEPTEURS TRANSMEMBRANAIRES DU SYSTEME NERVEUX CENTRAL

L'existence de récepteurs à la sérotonine a été, pour la première fois, suggérée par Gaddum et Picarelli sur la base d'approches expérimentales pharmacologiques. Ces auteurs ont postulé l'existence de deux types de récepteurs périphériques à la 5-HT appelés D (dibenzylamine) et M (morphine) en fonction de leur sensibilité à ces antagonistes supposés. Cette classification s'est révélée assez rapidement insuffisante pour s'appliquer aux résultats des nouvelles études pharmacologiques. Parallèlement, les grands progrès effectués dans la synthèse des ligands spécifiques et la grande évolution des outils pharmacologiques et biochimiques ont permis l'identification de cinq classes de sites récepteurs à la sérotonine appelées respectivement : 5-HT₁, 5-HT₂, 5-HT₃, 5-HT₅ et de la classe des stimulateurs de l'adénylate cyclase 5-HT₄, 5-HT₆, 5-HT₇. La classe 5-HT₁ se subdivise aujourd'hui en sept sous classes¹⁴ (5-HT_{1A}, 5-HT_{1B/1Dβ}, 5-HT_{1Dα}, 5-HT_{1E}, 5-HT_{1F}, 5-HT_{dro2A} et 5-HT_{dro2B}). La classe 5-HT₂ en trois sous classes 5-HT_{2A}, 5-HT_{2B}, 5-HT_{2C} anciennement appelé 5-HT_{1C}. La classe 5-HT₅ en deux sous classes 5-HT_{5A}, 5-HT_{5B}. La classe des activateurs de la sécrétion de l'adénylate cyclase en quatre sous-types 5-HT₄, 5-HT₆, 5-HT₇, 5-HT_{dro1}. La dernière classe est celle du récepteur 5-HT₃. A l'exception du récepteur 5-HT₃ qui est un récepteur à canal ionique sodium/potassium, tous les autres récepteurs font partie de la large famille des neurorécepteurs couplés aux protéines G. Ces récepteurs transmembranaires sont tous localisés en partie avec leurs sept hélices α dans la membrane de la cellule. Le reste de la protéine se trouve sous forme de boucles intra et extra-cellulaires (schéma 3).

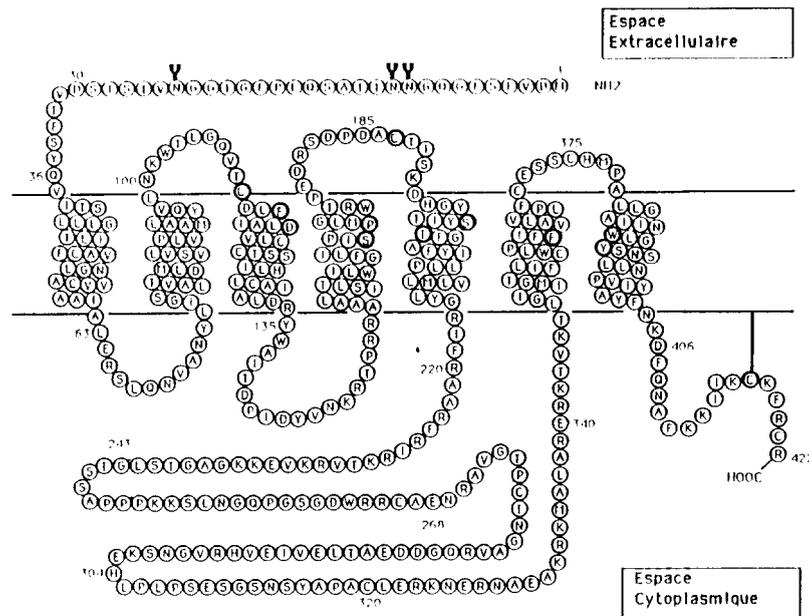


Schéma 3. Séquence en acides aminés du récepteur 5-HT_{1A} du rat. Les sites consensus pour la protéine kinase C sont situés dans les 2^{ème} et 3^{ème} boucles intra-cytoplasmiques, tandis que les asparagines glycosylées sont localisées dans la partie N-terminale extra-cellulaire.

II.1 - La famille 5-HT₁.

Les membres de la famille 5-HT₁ sont caractérisés par l'homologie de leur séquence primaire d'acides aminés et par leur effet commun d'inhibition de l'activité de l'adénylate cyclase. La comparaison des séquences révèle une haute conservation des résidus en particulier dans la région transmembranaire (Schéma 4).

		I		II			
5-HT1A	human	(39)	-SLLD	LTILFC	AVLRRN		
5-HT1B	mouse	(48)	-VAL	LALITL	ATLTSN		
5-HT1D α	human	(41)	-AVV	LSVITL	ATVLSN		
5-HT1E	human	(25)	-CMT	LAVITL	TLLN		
5-HT1F	mouse	(26)	-SLT	LPGL	ALMTT		
5-HTdro2A	(229)	-SVL	LGLM	LVYI	IQN		
5-HTdro2B	(80)	-AVV	LGLM	LVYI	IQN		
5-HTLym	(97)	-TSV	LGLF	LVCC	IIG		
5-HT5A	mouse	(43)	-LTL	LGLF	LAAT		
5-HT5B	mouse	(55)	-VTL	LVL	LIAA		
5-HT7	(87)	-GSI	LTL	ITLL	TIAG		
5-HTdro1	(165)	-SIV	LLV	ILG	TVG		
5-HT6	rat	(30)	-AA	LCV	VIVL		
5-HT2A	rat	(77)	-ALL	TVVI	ID		
5-HT2B	rat	(56)	-ALL	IFAV	IP		
5-HT2C	rat	(57)	-ALS	IVVI	IIIM		
		III		IV			
5-HT1A	human		FIAL	DV	CCTASI		
5-HT1B	mouse		WLS	DI	CCTASI		
5-HT1D α	human		WLS	DI	CCTASI		
5-HT1E	human		WLS	DI	CCTASI		
5-HT1F	mouse		WLS	DI	CCTASI		
5-HTdro2A			WLS	DI	CCTASI		
5-HTdro2B			WLS	DI	CCTASI		
5-HTLym			WLS	DI	CCTASI		
5-HT5A	mouse		WLS	DI	CCTASI		
5-HT5B	mouse		WLS	DI	CCTASI		
5-HT7	mouse		WLS	DI	CCTASI		
5-HTdro1			WLS	DI	CCTASI		
5-HT6	rat		WLS	DI	CCTASI		
5-HT2A	rat		WLS	DI	CCTASI		
5-HT2B	rat		WLS	DI	CCTASI		
5-HT2C	rat		WLS	DI	CCTASI		
		V		VI		VII	
5-HT1A	human		DPDA	CTI	SKDH	-G-	YTI
5-HT1B	mouse		EMLD	CFV	NTD	HVL	-YTV
5-HT1D α	human		EMSD	CLV	NTS	QIS	-YTI
5-HT1E	human		PPSQ	CTI	QHD	HVI	-YTI
5-HT1F	mouse		RDDE	CVI	KHD	HIV	-STI
5-HTdro2A			EQQK	CMV	SQD	-VS-	YQV
5-HTdro2B			EEQH	CMV	SQD	-VG-	YQI
5-HTLym			KTGT	CL	I	SQD	-KG-
5-HT5A	mouse		EE--	CV	S	REP	-S-
5-HT5B	mouse		QR--	CV	S	QEP	-S-
5-HT7	mouse		---	V	CL	I	SQD
5-HTdro1			GQPI	CT	V	CON	-FA-
5-HT6	rat		APGQ	CRL	-LAS	LP-	FV
5-HT2A	rat		KEGS	CL	LAD	---	NFV
5-HT2B	rat		HNIT	CEL	T	KDR	F
5-HT2C	rat		NNIT	CV	LND	P	---
5-HT1A	human		T	LC	I	MC	I
5-HT1B	mouse		T	LC	I	MC	I
5-HT1D α	human		I	LC	I	MC	I
5-HT1E	human		I	LC	I	MC	I
5-HT1F	mouse		T	LC	I	MC	I
5-HTdro2A			T	LC	I	MC	I
5-HTdro2B			T	LC	I	MC	I
5-HTLym			T	LC	I	MC	I
5-HT5A	mouse		M	Y	C	I	MC
5-HT5B	mouse		M	Y	C	I	MC
5-HT7	mouse		T	LC	I	MC	I
5-HTdro1			T	LC	I	MC	I
5-HT6	rat		T	LC	I	MC	I
5-HT2A	rat		V	LC	I	MC	I
5-HT2B	rat		V	LC	I	MC	I
5-HT2C	rat		V	LC	I	MC	I

Schéma 4. Similarité des acides aminés entre les récepteurs 5-HT. Alignement des séquences primaires des récepteurs 5-HT_{1A} humain, 5-HT_{1B} souris, 5-HT_{1D α} humain, 5-HT_{1E} humain, 5-HT_{1F} souris, 5-HT_{dro2A} et 5-HT_{dro2B} drosophile, 5-HT_{Lym} lymnea, 5-HT₇ souris, 5-HT_{dro1} drosophile, 5-HT_{5A} et 5-HT_{5B} souris, 5-HT₆ rat, 5-HT_{2A} rat, 5-HT_{2B} rat et 5HT_{2C} rat. Les sept régions en hélices ? transmembranaires sont numérotées de I à VII. Les parties en fond noir montrent les similitudes de résidus entre les différentes séquences²⁷.

II.1.A - Le récepteur 5-HT_{1A}.

Le premier gène d'un récepteur 5-HT à avoir été cloné est celui du récepteur 5-HT_{1A} humain. En criblant une banque de données génomique humaine avec une sonde β 2-adrénergique, les auteurs Kobilka et Fargin^{27,28} espéraient « pêcher » le gène du récepteur β 1-adrénergique ; en fait, ils ont ainsi isolé le gène codant pour le récepteur 5-HT_{1A}. Le gène correspondant chez le rat a été obtenu de la même manière et sa séquence est d'ailleurs très peu différente de celle du gène humain²⁹. La stimulation du récepteur 5-HT_{1A} dans les cellules transfectées conduit à une inhibition de l'activité de l'adénylate cyclase, comme dans le tissu cérébral. Cependant, dans certaines lignées cellulaires le récepteur peut être couplé à la phospholipase C. On trouve ce récepteur dans les structures limbiques comme l'hippocampe, le septum, l'amygdale, le cortex frontal, qui jouent un rôle clé dans le contrôle de l'humeur et des émotions. D'autres récepteurs appelés autorécepteurs sont situés à la surface des somas et de dendrites des neurones³⁰. C'est d'ailleurs en stimulant ces autorécepteurs somatodendritiques que certains agonistes 5-HT_{1A} exercent leurs effets anxiolytiques³¹.

Le récepteur 5-HT_{1A} se caractérise par une affinité nanomolaire pour la [³H]-5-HT et micromolaire pour les antagonistes de la sérotonine. Ce fut la spipérone un antagoniste de la 5-HT qui permit à Pedigo *et al.*³² de caractériser l'hétérogénéité des sites récepteurs 5-HT₁. Cependant l'un des grands succès dans la connaissance des données biochimiques et physiologiques revient à la découverte des qualités d'affinité et de sélectivité que possède la 8-OH-DPAT vis-à-vis des sites récepteurs 5-HT_{1A}. La stabilité et la haute activité spécifique de la [³H]-8-OH-DPAT en font une molécule plus intéressante que la [³H]-5-HT pour le marquage spécifique des sites 5-HT_{1A}. Ce ligand a permis d'explorer les distributions régionales et subcellulaires ainsi que les propriétés biochimiques et pharmacologiques de ce site³³.

Depuis lors plusieurs autres radioligands qui reconnaissent sélectivement les sites 5-HT_{1A} ont été développés : [³H]-ipsapirone³⁴, [³H]-PAPP (1-[2-(4-aminophényl)éthyl-4-(3-trifluorométhylphényl) pipérazine]³⁵, et tout récemment [³H]-spiroxatine³⁶ et [³H]-5-MeO-DPAC¹⁸. Toutefois certains de ces radioligands telle que la [³H]-spiroxatine manquent de sélectivité ; cette dernière présente une forte affinité pour d'autres sites du SNC comme les sites dopaminergiques D₂³⁷.

Grâce à ces radioligands agonistes ou agonistes partiels il s'est avéré possible de préciser la cartographie des récepteurs 5-HT_{1A}. Ainsi les études d'autoradiographie avec la [³H]-8-OH-DPAT ou la [¹²⁵I]-BH-8-MeO-N-PAT et biochimiques ont démontré que dans le cerveau les sites récepteurs 5-HT_{1A} étaient plus abondants dans des régions comme l'hippocampe et le raphé³⁸. Dans les neurones du cerveau de rat, les sites récepteurs 5-HT_{1A} sont localisés aussi bien dans les régions présynaptique que postsynaptique³⁹. Enfin, il est aujourd'hui parfaitement établi que les sites récepteurs 5-HT_{1A} sont impliqués dans les troubles de l'anxiété⁴⁰.

II.1.B - Les récepteurs 5-HT_{1B} et 5-HT_{1Dβ}.

Le récepteur 5-HT_{1B} est présent chez le rat, la souris, l'hamster mais semble absent pour toutes les autres espèces de mammifères^{41,42}. Par contre le récepteur 5-HT_{1Dβ} est absent chez les rongeurs mais a été détecté chez le chien, le porc de Guinée et l'humain avec une distribution similaire au récepteur 5-HT_{1B}. Il a été montré par des études de clonage moléculaire que la contrepartie du récepteur 5-HT_{1B} du rat, chez l'humain est le récepteur 5-HT_{1Dβ}⁴³.

Les sites récepteurs 5-HT_{1B} décrits à la suite des travaux de Pédigo *et al.*³³ ont été mis en évidence dans des expériences de déplacement de la [³H]-5-HT par la 8-OH-DPAT ou la spipérone dans le cortex et l'hippocampe de rat. Les récepteurs 5-HT_{1B} et 5-HT_{1Dβ} (96% de similarité) ont été clonés respectivement chez le rat, la souris pour le premier^{44,45,46} et l'humain pour le second^{47,48}. Les récepteurs sont constitués de 390 acides aminés pour l'humain et 386 acides aminés pour les rongeurs. Comme pour les autres récepteurs couplés à la protéine G ces récepteurs contiennent sept hélices transmembranaires. Le gène codant les récepteurs 5-HT_{1B/1Dβ} est localisé sur le chromosome 9 (position 9E) de la souris et le chromosome 6 de l'humain (6q13)⁴⁹. Les techniques de déplacement biphasique ont permis de révéler la classe de sites appelés 5-HT_{1B}, caractérisés par une affinité nanomolaire pour la [³H]-5-HT, mais seulement micromolaire pour la 8-OH-DPAT, le sumatriptan, la spipérone, la yohimbine, la kétansérine et la miansérine. Par contre, le sumatriptan composé antimigraineux est très actif envers le récepteur 5-HT_{1Dβ}. Plusieurs autres radioligands comme le [¹²⁵I]-iodocyanopindolol (ICYP), le [³H]-RU24969 et le [³H]-CGS12066B présentent une affinité appréciable vis-à-vis de ces sites mais ne peuvent malheureusement prétendre à être sélectionnés comme marqueurs en raison de leur absence de sélectivité. Jusqu'à présent la mesure de la liaison de la [³H]-5-HT en présence de quelques micromoles de 8-OH-DPAT,

permettant l'occupation totale des sites 5-HT_{1A}, reste encore l'un des meilleurs moyens d'étude des sites 5-HT_{1B} étant donné l'absence de radioligand réellement sélectif.

Les études d'autoradiographie chez le rat ont montré que les récepteurs 5-HT_{1B/1Dβ} sont localisés surtout dans les structures extrapyramidales, à savoir globus pallidus, noyau caudé-putamen, substance noire, ainsi que dans le tuberculum olfactorium dorsal. Au niveau neuronal, dans le cerveau du rat, les sites 5-HT_{1B} sont localisés à la fois dans la région présynaptique, comme autorécepteurs contrôlant la libération de la sérotonine⁵⁰, et postsynaptique comme hétérorécepteurs contrôlant la libération⁵¹ d'acétylcholine.

II.1.C - Le récepteur 5-HT_{1Dα}.

En 1989, Libert *et al*⁵², ont proposé la séquence de quatre nouveaux récepteurs faisant partie de la famille des récepteurs couplés aux protéines G. Ce récepteur a été cloné chez l'humain⁵³, le rat^{54,55} et la souris⁴⁷. De ce récepteur de 377 acides aminés se dégage une haute homologie de séquence (74-77% au niveau du domaine transmembranaire) avec les récepteurs 5-HT_{1B/1Dβ}. Il a donc été judicieusement appelé 5-HT_{1Dα}. Malgré leur haute homologie ces récepteurs n'ont pas la même localisation des gènes codant sur les chromosomes^{56,57}.

Des études de l'activité de ces récepteurs envers différents ligands potentiels montre cette hiérarchie : 5-CT > 5-HT > sumatriptan > yohimbine > RU24969 > 8-OH-DPAT > spipéronne. Une étude intéressante a montré que des ligands ayant principalement une activité 5-HT₂ antagoniste, avaient une haute affinité pour le récepteur 5-HT_{1Dα}⁵⁵

II.1.D - Le récepteur 5-HT_{1E}.

Le récepteur 5-HT_{1E} fait partie de la famille des 5-HT₁. Il est caractérisé par sa faible affinité avec la 5-CT^{58,59}. Ce récepteur a été isolé de la même façon que le récepteur 5-HT_{1A} en criblant de manière exhaustive une banque de données génomique humaine^{60,61,62,63}. Ce récepteur montre une forte homologie (64%) de séquence avec les récepteurs 5-HT_{1Dα} et 5-HT_{1Dβ}. Le récepteur 5-HT_{1E} est couplé négativement avec l'adénylate cyclase.

En ce qui concerne les études de liaisons (binding) pour ce récepteur, la classification des radioligands est : 5-HT > méthysergide > ergotamine > sumatriptan > spipérone^{59,60,62,63,64}. Le sumatriptan et la 5-CT ont une faible activité avec le récepteur 5-HT_{1E} par rapport au récepteur 5-HT_{1D}. Ces récepteurs transmembranaires sont principalement localisés dans le cortex et l'amygdale.

II.1.E - Le récepteur 5-HT_{1F}.

Le récepteur 5-HT_{1F} a été isolé⁶⁴ et cloné^{65,66}. Le pourcentage d'homologie avec le récepteur 5-HT_{1E} est de 61% et de 54% avec les récepteurs 5-HT_{1B} et 5-HT_{1D}. Des études pharmacologiques ont montré une faible activité de la 5-CT comme pour le 5-HT_{1E}. Les radioligands étudiés sur ce récepteur ont été classés dans cet ordre d'activité : 5-HT et LSD > sumatriptan > méthysergide > yohimbine > ergotamine. La kétansérine et le cyanopindolol présentent une faible affinité vis-à-vis de ce récepteur^{65,66,67}.

II.2 - La famille 5-HT₂.

Cette famille contient trois sous-types de récepteurs qui présentent une forte homologie de leur séquence primaire. Ces récepteurs sont couplés positivement à la phospholipase C. Les récepteurs 5-HT₂ semblent participer au contrôle des systèmes dopaminergiques centraux. Les antagonistes 5-HT₂ s'opposent en particulier aux effets extrapyramidaux induits par le blocage des récepteurs D₂, d'où la tendance actuelle à développer de nouveaux neuroleptiques ayant des potentialités à la fois anti-dopaminergique et anti-sérotoninergique. De fait, il semble que ces composés n'induisent pas de

mouvements anormaux chez les schizophrènes, contrairement aux neuroleptiques classiques comme l'halopéridol.

Bien que les sites 5-HT₂ aient été identifiés pour la première fois au moyen de la [³H]-spipéronne⁶⁷ c'est aujourd'hui la [³H]-kétansérine⁶⁸ qui est le radioligand le plus utilisé pour marquer ces sites car la [³H]-spipéronne s'avère également un très bon ligand des récepteurs dopaminergiques.

Sur la base d'études de liaison, Peroutka a montré que la 5-HT et la spipéronne tritiés se lient dans les homogénats de cortex du rat sur deux sites de liaison différents nommés respectivement 5-HT₁ et 5-HT₂. Il paraît tout à fait clair aujourd'hui que les sites de liaison 5-HT₂ décrits par ces auteurs sont identiques au récepteur périphérique D de Gaddum et Picarelli. Ce récepteur est répandu dans tout l'organisme, où il transmet les actions excitatrices de la 5-HT, telles que la dépolarisation de neurones ou la contraction des muscles lisses. De plus, le récepteur 5-HT₂ contrôle également l'agrégation des plaquettes sanguines.

La localisation des sites 5-HT₂ est très différente de celle des sites 5-HT₁. C'est ainsi que chez le rat, leur densité décroît au niveau du SNC dans l'ordre suivant : cortex > hippocampe > hypothalamus > tronc cérébral et cervelet. Au niveau neuronal, ces sites ne sont localisés qu'en position postsynaptique.

II.2.A - Le récepteur 5-HT_{2A}.

Le récepteur 5-HT_{2a} a été cloné la première fois chez le rat⁶⁹ par homologie avec le récepteur 5-HT_{2C} du rat⁷⁰ anciennement nommé 5-HT_{1C}. Ce récepteur contient 471 acides aminés chez le rat⁷⁰, la souris⁷¹, l'hamster⁷² et l'humain⁷³. L'analyse des gènes du récepteur 5-HT_{2A} révèle la présence de deux introns⁷⁴ conservés dans les trois récepteurs 5-HT₂. Le gène de ce récepteur est localisé pour l'humain sur le chromosome 13 et pour la souris sur le 14⁷⁵. Ce récepteur montre une haute affinité pour les radioligands [¹²⁵I]-LSD et [³H]-spipéronne de l'ordre de la nanomole. Le profil pharmacologique est : kétansérine = spipéronne = DOI > mésulergine > 5-HT. Le récepteur 5-HT_{2A} est principalement détecté dans le cortex frontal, l'hippocampe, le striatum⁷⁶.

II.2.B - Le récepteur 5-HT_{2B}.

Ce récepteur a été isolé chez le rat⁷⁷ et la souris⁷⁸. La séquence primaire est constituée de 479 acides aminés pour le rat et 504 acides aminés pour la souris. Ce récepteur montre une homologie à 45% et 51% avec les récepteurs respectifs 5-HT_{2A} et 5-HT_{2C}. Les radioligands sont [³H]-5-HT (Kd = 7.9 nM) et [¹²⁵I]-DOI (Kd = 25.8 nM). Le profil pharmacologique est : ritansérine > 5-HT > 1-(3-chlorophényl)pipérazine hydrochloride (mCPP) > kétansérine > miansérine. La miansérine qui a une faible activité avec ce récepteur a malgré tout une haute activité avec les récepteurs 5-HT_{2A} et 5-HT_{2C}. La distribution régionale comprend le coeur, l'intestin et le cerveau.

II.2.C - Le récepteur 5-HT_{2C}.

Le site 5-HT_{2C}, anciennement nommé 5-HT_{1C}, a été mis en évidence pour la première fois à l'aide de la [³H]-5-HT au niveau du plexus choroïde et du cortex chez le porc⁷⁹. La séquence primaire du récepteur 5-HT_{2C} humain comprend 458 acides aminés⁷⁵. Cette catégorie de sites a été identifiée, par la suite, au moyen de la [³H]-mésulergine, du [³H]-LSD. Comme dans le cas des sites 5-HT_{1B}, il n'existe pas de ligand aussi spécifique que la 8-OH-DPAT. Hoyer *et al.*⁸⁰ suppose l'existence d'une étroite relation entre l'affinité de certains ligands pour les récepteurs 5-HT_{2C} et leur capacité à stimuler ou à motiver le processus d'accumulation de l'inositol phosphate au niveau du plexus chloroïde du porc.

II.3 - Famille des activateurs de l'adénylate cyclase.

II.3.A - Le récepteur 5-HT₄.

A l'heure actuelle, le récepteur 5-HT₄ n'a toujours pas été cloné. Toutefois un gène codant pour un récepteur du même type, c'est à dire dont la stimulation provoque l'activation de l'adénylate cyclase a été identifié chez la Drosophile⁸¹. Il est toutefois prématuré de considérer ce récepteur comme l'homologue du récepteur 5-HT₄ chez l'humain, d'autant qu'il présente non pas sept domaines transmembranaires mais huit. Il possède en effet une longue chaîne hydrophobe dans sa partie N-terminale, que l'on ne retrouve pas chez les autres récepteurs couplés aux protéines G. Le profil pharmacologique de ce récepteur est : GR113808 > SDZ205557 > BIMU8 > cisapride > 5-HT > 5-MeOT²⁴. Un composé, le SC53116 semble être un agoniste du récepteur 5-HT₄⁸². Le composé GR113808 est un antagoniste sélectif des 5-HT₄. L'étude autoradiographique du ligand GR113808 tritié a montré une distribution régionale du récepteur au niveau du striatum, de la substance noire et de l'hippocampe⁸³.

II.3.B - Le récepteur 5-HT₆.

Ce récepteur est constitué d'une chaîne polypeptidique de 437⁸⁴ ou 436⁸⁵ acides aminés selon les résultats de deux études. La différence de ces deux séquences primaires se situe au niveau de la partie C-terminale. Ce récepteur comporte sept régions hydrophobes et la comparaison avec les autres récepteurs montre qu'il a peu de similitude avec ceux-ci (Schéma 5). La troisième boucle cytoplasmique est beaucoup plus petite que pour les autres récepteurs. Toutes ces caractéristiques moléculaires font penser que le récepteur 5-HT₆ correspond à une nouvelle classe de récepteur.

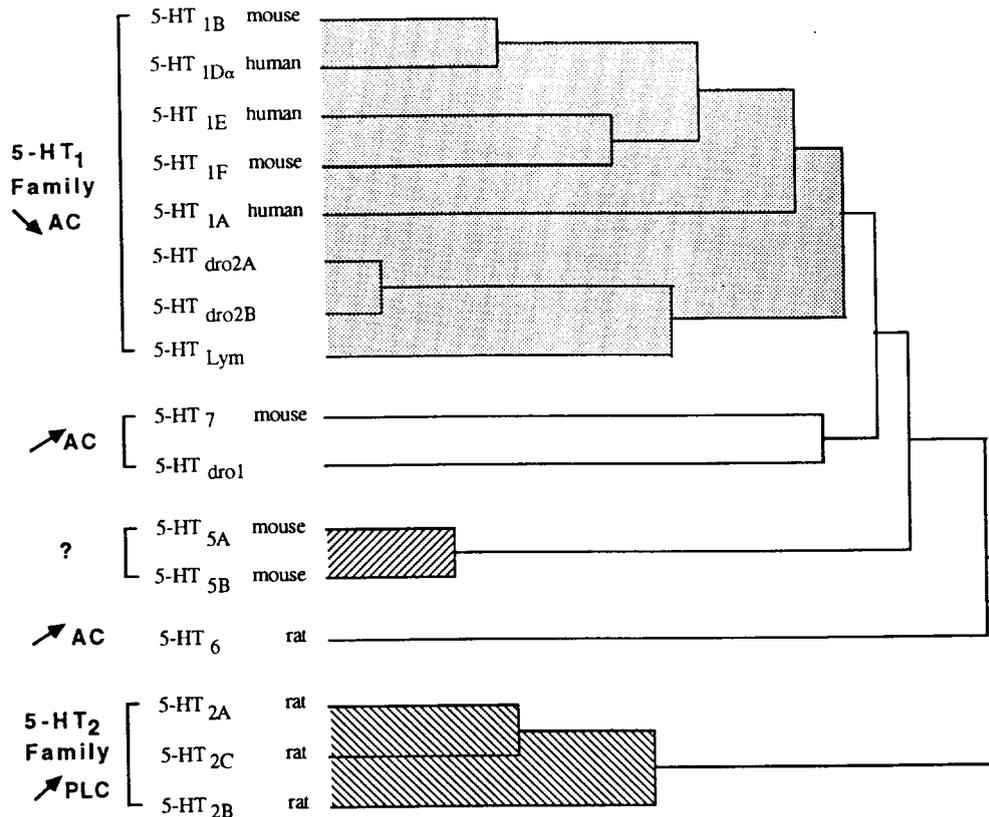


Schéma 5. Dendrogramme. Les séquences des récepteurs 5-HT sont comparées et clustérisées avec le programme CLUSTAL. La longueur des lignes horizontales est inversement proportionnelle au pourcentage d'homologie entre les récepteurs ou les groupes de récepteurs²⁷.

II.3.C - Le récepteur 5-HT₇.

Le récepteur 5-HT₇ couplé positivement à l'adénylate cyclase a été cloné chez l'humain⁸⁶, le rat⁸⁷ et la souris⁸⁸. Sa séquence primaire est constituée de 448 acides aminés. Le récepteur 5-HT₇ montre une forte homologie (Schéma 5) avec le 5-HT_{dro1}. Celui-ci est hautement réceptif à la 5-HT (Kd = 3.6 nM) et au LSD (Kd = 1.2 nM). Le profil pharmacologique est : LSD > 5-CT > méthiothépine > 5-HT > clozapine > 8-OH-DPAT.

II.4 - La famille 5-HT₅.

La famille 5-HT₅ contient deux récepteurs, le 5-HT_{5A} et le 5-HT_{5B} qui définissent une nouvelle famille de récepteurs sérotoninergiques. Au niveau de la séquence primaire, ces récepteurs sont très différents des familles 5-HT₁ et 5-HT₂.

II.4.A - Le récepteur 5-HT_{5A}.

En utilisant des dérivés oligonucléotides des domaines transmembranaires III et VI de récepteurs sérotoninergiques, Plassat et al.⁸⁹ ont isolé un nouveau récepteur. Une analyse de l'hydrophobicité des 357 acides aminés a révélée sept domaines hydrophobiques comme pour la majorité des récepteurs 5-HT. Toutefois, le faible pourcentage d'homologie a amené les auteurs à donner le nom 5-HT₅ pour cette nouvelle famille. Le LSD est très affin vis-à-vis de ce récepteur (Kd = 340 pM). Le profil pharmacologique est : LSD > ergotamine > 5-CT > méthysergide > 5-HT = RU24969 > bufoténine = yohimbine.

II.4.B - Le récepteur 5-HT_{5B}.

Ce récepteur a été isolé suivant la même méthodologie que pour le récepteur 5-HT_{5A}. Sa séquence est de 370 acides aminés⁹⁰. L'homologie de séquence avec le récepteur 5-HT_{5A} est de 77%. Comme pour le récepteur 5-HT_{5A}, ce récepteur présente très peu d'homologie de séquence avec les autres neurorécepteurs. Le LSD est très affin à ce récepteur (Kd = 470 pM). Le profil pharmacologique est le suivant : LSD > ergotamine > méthiotépine > 5-CT > méthysergide > 5-HT > bufoténine. Ce récepteur est présent au niveau de l'hippocampe et des noyaux du raphé dorsal.

II.5 - Le récepteur à canal ionique 5-HT₃.

A côté de la grande diversité des récepteurs de la 5-HT couplés aux protéines G, on n'a identifié, jusqu'à présent, qu'un seul « récepteur à canal ionique » pour ce neuromédiateur. Il s'agit du récepteur 5-HT₃ dont la stimulation entraîne l'ouverture d'un canal Na⁺/K⁺. Dans son fonctionnement et dans ses régulations, ce récepteur présente des similitudes avec le récepteur nicotinique de l'acétylcholine et le récepteur NMDA du glutamate. L'existence du récepteur 5-HT₃ nommé « récepteur M » a donc été prouvé en premier par Gaddum et Picarelli en 19576. Ce site récepteur fut mis en évidence par Hamon *et al.*¹⁷, après avoir remarqué que la 8-OH-DPAT se liait non seulement aux récepteurs 5-HT_{1A} mais également à d'autres sites présents dans certaines parties du cerveau du rat et notamment dans le striatum. Le profil pharmacologique de ces sites est nettement différent de celui des sites 5-HT_{1A}. De plus, des drogues comme la spipéronne, et certains dérivés pipéraziniques tels que la buspirone, l'ipsapirone et la gépironne qui présentent une affinité de l'ordre de la nanomole pour les récepteurs 5-HT_{1A}, sont au contraire très peu reconnues par les sites 5-HT₃. Par contre des composés comme le [³H]-zacopride, et la [³H]-quipazine sont décrits comme des puissants marqueurs sélectifs des sites 5-HT₃ dans bon nombre de tissus. Il est intéressant de souligner que le MDL72222 et son analogue MDL7314738 contrairement à tous les autres antagonistes des récepteurs 5-HT₃, montrent très peu ou pas d'activité vis-à-vis des sites 5-HT₃ présents dans l'iléon de cobaye⁹¹. Ces résultats supportent l'hypothèse selon laquelle il existerait comme dans le cas des sites 5-HT₁ d'autres sous types pour la classe 5-HT₃. Un des composés très actifs pour ce récepteur est le tropisetron. En conclusion, le rôle des sites 5-HT₃ n'est pas véritablement élucidé et il apparaît évident que les progrès réalisés dans le domaine de la modélisation moléculaire, joints à l'utilisation de nouvelles sondes comme le 5-[¹²⁵I]-iodozacopride devraient permettre de mieux connaître ce récepteur et d'envisager l'existence de sous-types 5-HT₃.

II.6 - Les récepteurs sérotoninergiques chez les invertébrés.

Les trois récepteurs de *Drosophile*⁹² 5-HT_{dro1}, 5-HT_{dro2A} et 5-HT_{dro2B} ont été isolés par Hen *et al.*. Le récepteur 5-HT_{dro1} comporte 564 acides aminés et une analyse de l'étude des zones hydrophobes a montré l'existence de huit domaines hydrophobes. Les récepteurs 5-HT_{dro2A} et 5-HT_{dro2B} comportent respectivement 834 et 645 acides aminés. Le degré d'homologie de séquence entre ces deux protéines est très élevé (84.3%). Le LSD est très affiné de ces trois récepteurs. La distribution régionale est le cerveau et le cordon ventral.

III. POTENTIALITES THERAPEUTIQUES DES LIGANDS DES RECEPTEURS 5-HT

Compte tenu de la pluralité des fonctions impliquant la sérotonine, il n'est pas surprenant *a priori* de trouver des ligands des récepteurs de ce neuromédiateur parmi des médicaments aussi divers que des anxiolytiques (notamment la buspirone, un agoniste partiel des récepteurs 5-HT_{1A}), des antimigraineux (par exemple le sumatriptan, un agoniste des récepteurs 5-HT_{1B/1Dβ}) et des antiémétiques (les antagonistes des récepteurs 5-HT₃ comme l'ondansétron et le granisetron). Les potentialités thérapeutiques des ligands des différentes classes des récepteurs de la 5-HT sont tout à fait considérables, et les industriels de la pharmacie font actuellement de gros efforts pour mettre au point des ligands de ces récepteurs qui soient encore plus sélectifs et efficaces.

Dans le domaine des anxiolytiques, plusieurs récepteurs semblent être concernés puisque des effets de type anxiolytique ont été obtenus chez l'animal avec des agonistes des récepteurs 5-HT_{1A} (beaucoup plus sélectifs que la buspirone), des antagonistes du récepteur 5-HT_{2C}, et des antagonistes des récepteurs 5-HT₃. L'intérêt de ces nouveaux composés par rapport aux benzodiazépines est multiple. Contrairement à ces dernières et à d'autres tranquillisants qui se fixent également sur les mêmes sites, les ligands des récepteurs de la 5-HT n'induisent ni accoutumance ni dépendance. De plus, ils ne sont ni sédatifs, ni myorelaxants, ni surtout, amnésiants. Au contraire, les antagonistes 5-HT, semblent stimuler la fonction mnésique dans divers tests appropriés chez l'animal. Par ailleurs, les antagonistes 5-HT₂ stimulent la fonction de récupération en augmentant les taux de sommeil lent profond (stades III et IV) aussi bien chez l'animal que chez l'homme. En effet l'intégrité des systèmes indolaminergiques centraux apparaît nécessaire à l'élaboration du processus du sommeil. L'activation de ces systèmes pendant l'éveil et au cours de la privation de sommeil suggère que la 5-HT peut jouer un rôle préparateur dans la mise en place du sommeil lent profond et du sommeil paradoxal. La sérotonine pourrait agir en tant que « neurohormone » régulant pendant l'éveil la biosynthèse et l'accumulation de substances hypnogènes qui seraient utilisées au cours du sommeil lent profond et paradoxal. Enfin, des résultats encourageants montrent qu'aussi bien les agonistes des récepteurs 5-HT_{1A}, que les antagonistes des récepteurs 5-HT₂ et 5-HT₃ peuvent diminuer la

consommation d'alcool chez les patients alcooliques, sans doute en partie via leurs effets anxiolytiques. On peut donc s'attendre à ce que des ligands de ce type soient développés pour le traitement de l'anxiété pathologique au cours des prochaines années.

Dans le domaine des antidépresseurs, l'étude des récepteurs de la 5-HT peut également déboucher sur la mise au point de molécules encore plus efficaces étant donné que la thérapeutique actuelle fait largement appel aux inhibiteurs de la recapture de ce neuromédiateur. Ces composés augmentent la concentration extra-cellulaire de la 5-HT, et entraînent donc la stimulation des diverses classes de récepteurs spécifiques par le neuromédiateur endogène. Dans ce contexte, il convient de rappeler que certains de ces récepteurs sont des autorécepteurs dont l'activation déclenche un contrôle inhibiteur de la transmission sérotoninergique centrale, c'est à dire l'inverse du but recherché avec les bloqueurs de recapture. Selon certains auteurs, la stimulation des autorécepteurs par le neuromédiateur endogène en excès s'opposerait ainsi à l'action thérapeutique des antidépresseurs. D'ailleurs, le délai d'action antidépressive des inhibiteurs de recapture de la 5-HT (supérieur à 2 semaines) correspond au temps nécessaire à la désensibilisation des autorécepteurs²⁵. Dans ces conditions, on comprend pourquoi les antagonistes des autorécepteurs (5-HT_{1A}, 5-HT_{1B/1Dβ}) constituent une famille de composés d'un grand intérêt potentiel pour le traitement des dépressions (notamment en association avec les inhibiteurs de la recapture de la 5-HT).

Enfin, on peut attendre beaucoup de ligands de tel ou tel récepteur de la 5-HT dans le traitement de désordres comportementaux tels que l'obsession-compulsion, l'anorexie ou au contraire la boulimie, l'agressivité pathologique, etc... Dès lors, plutôt que d'inonder tous les récepteurs 5-HT, peut-être vaudrait-il mieux tenter de développer des ligands sélectifs des sous-types 5-HT_{1B/1Dβ} et 5-HT_{1Dα} pour le traitement des troubles de la prise alimentaire ? Or un tel ligand existe déjà ; il s'agit du sumatriptan, dont l'efficacité dans le traitement de la crise de migraine est parfaitement établie. Mais le sumatriptan n'affecte pas la prise alimentaire. En fait, quel que soit le ligand, encore faut-il qu'il pénètre au travers de la barrière hématoencéphalique pour atteindre les récepteurs centraux, ce qui n'est pas le cas du sumatriptan ; c'est en stimulant des récepteurs extracérébraux (mais intracrâniens) que ce composé provoque la vasoconstriction de certains vaisseaux à l'origine de son action anti-migraineuse. Le récepteur sur lequel agit vraiment le sumatriptan pour exercer son action anti-migraineuse est le récepteur 5-HT_{1Dβ}. Là encore, l'étude de

ce récepteur devrait contribuer au développement de nouveaux médicaments anti-migraineux plus efficaces et entraînant le minimum d'effets secondaires.

Enfin, il convient de rappeler combien sont puissants les antagonistes des récepteurs 5-HT, pour empêcher les nausées et vomissements associés à la chimiothérapie et à la radiothérapie dans le traitement des cancers. Dans ce cas également, les récepteurs concernés semblent être surtout localisés à la périphérie, sur les fibres vagales du complexe gastro-intestinal. Jusqu'à présent, tous les ligands des différentes classes de récepteurs de la 5-HT, y compris ceux qui sont des médicaments efficaces (sumatriptan, ondansetron, etc...) sont issus de recherches chimiques et pharmacologiques « traditionnelles ». L'apport de la biologie moléculaire à la connaissance fondamentale de ces récepteurs est déjà, et sera encore pour des années, tout à fait spectaculaire. Une ère nouvelle s'ouvre actuellement avec la modélisation tridimensionnelle de l'interaction des ligands avec leurs récepteurs.

IV. LES LIGANDS 5-HT_{1A} ET LEURS APPLICATIONS PHARMACOLOGIQUES

Les cinq dernières années ont connu une multiplication importante du nombre de familles chimiques capables d'interagir avec les différents récepteurs 5-HT. Ces travaux ont facilité les tentatives de définition du modèle pharmacophore correspondant à chacun des récepteurs. Cet essor a aussi profité à la thérapeutique moderne puisque certains agonistes 5-HT_{1A} se sont avérés des anxiolytiques ne présentant pas les effets secondaires des benzodiazépines.

Au niveau des cellules cibles, la stimulation des récepteurs 5-HT entraîne une augmentation (5-HT₄, 5-HT₆, 5-HT₇) ou une diminution (5-HT_{1A}, 5-HT_{1B/1Dβ} et 5-HT_{1Dα}, 5-HT_{1E}, 5-HT_{1F}) de la production d'AMP cyclique. Outre ces différences fonctionnelles, ce sont surtout des profils pharmacologiques spécifiques qui caractérisent les diverses classes de récepteurs centraux de la 5-HT. Grâce à ces outils pharmacologiques, les biologistes et les cliniciens sont parvenus à attribuer à telle ou telle famille de ligands des potentialités thérapeutiques spécifiques. Par exemple comme cela a été déjà mentionné, il est aujourd'hui clairement établi qu'un agoniste des récepteurs 5-HT_{1A} est un anxiolytique et/ou un antidépresseur potentiel.

En dehors de la sérotonine qui est biosynthétisée, plusieurs ligands de structure analogue ou différente de cette molécule endogène ont été préparés. Plusieurs de ces ligands reconnaissent les sites récepteurs 5-HT_{1A} avec une affinité comparable ou supérieure à celle de la sérotonine. Dans ce chapitre, nous présenteront les principales familles de composés présentant une affinité pour le site 5-HT_{1A}.

IV.1 - Les dérivés indoliques.

Comme les dérivés de la 5-hydroxytryptamine (5-HT), certains composés indoliques tels que le RU24969 ($pK_i = 8.6$), le lisuride ($pK_i = 9.1$), le DP-5-CT ($pK_i = 9.7$), la 5-CT ($pK_i = 9.5$), le LSD ($pK_i = 8.4$) et les dérivés de l'ergoline sont des ligands affins du récepteur 5-HT_{1A}⁹³. A partir de ces multiples composés indoliques synthétisés et testés, plusieurs informations structurales se trouvent directement liées à l'activité et la sélectivité.

La faible activité de la tryptamine ($pIC_{50} = 6.8$) laisse penser que la fonction hydroxyle en position 5 joue un rôle important dans la reconnaissance du site. L'introduction d'un groupement fonctionnel (par exemple CONH₂) en lieu et place de l'hydroxyle augmente l'affinité. C'est le cas de la 5-carboxamidotryptamine (5-CT). La position du substituant en position 5 est donc très importante pour la reconnaissance du ligand par le récepteur transmembranaire. La 4-HT ($pIC_{50} = 7$) et la 6-HT ($pIC_{50} = 5.8$) ont une activité bien moindre que la 5-HT ($pIC_{50} = 8.5$). La substitution par un groupement méthyle sur différentes positions fait varier l'affinité et la sélectivité. C'est le cas, par exemple, de la 2-méthyle-5-hydroxytryptamine (2-Me-5-HT) ($pK_i = 5.8$). La méthylation de la chaîne latérale soit sur le carbone α soit sur le carbone β diminue l'activité. Ceci est démontré avec les composés α et β -Me-5-HT ($pK_i = 7$ et 7.1).

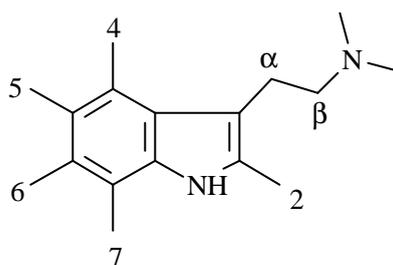


Figure 4. Représentation des différents points de substitution de la sérotonine.

IV.2 - Les arylpipérazines.

Considérés au départ comme ligands 5-HT peu actifs, les dérivés arylpipéraziniques ont connu assez rapidement un développement considérable depuis l'identification de la sous classe 5-HT_{1A} qu'ils reconnaissent préférentiellement^{94,95,96,97}. En effet, la gépirone ($pIC_{50} = 7.5$), le flésinoxan ($pIC_{50} = 8.8$), le PAPP⁹⁸ ($pK_i = 8.5$), le Br-Ac-TFMPP ($pK_i = 9.3$), le NAN190⁹⁹ ($pK_i = 8.5$), l'ipsapirone ($pIC_{50} = 8.3$) et la buspirone ($pIC_{50} = 8$) font aujourd'hui partie des meilleurs ligands 5-HT_{1A} malgré leur manque de sélectivité. Toutefois, certaines études de relations structure-activité ont montré de bonnes sélectivités^{100,101,102} dans des gammes de composés arylpipéraziniques. De surcroît, la buspirone qui a été antérieurement développée comme agent antipsychotique en série dopaminergique est actuellement considérée comme le premier composé anxiolytique non benzodiazépinique commercialisé sous le nom de « Buspar » (Figure 5).

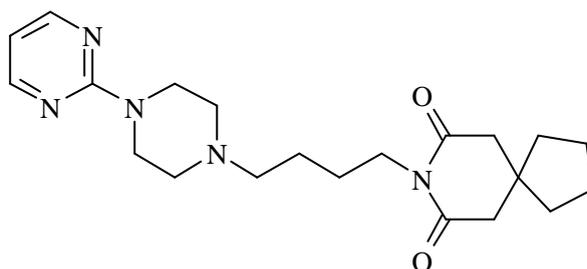


Figure 5. Représentation de la buspirone « Buspar ».

IV.3 - Les dérivés benzodioxanes.

A côté des familles de ligands précitées, les aminométhylbenzodioxanes tels que le MDL72832¹⁰³ ($pIC_{50} = 9.1$) (Figure 6) et le MDL73005EF⁴¹ ($pIC_{50} = 8.6$) possèdent également une bonne affinité pour les récepteurs 5-HT_{1A}. D'autres molécules¹⁰⁴ contenant le groupement benzodioxane (WB4101 ($pIC_{50} = 7.9$) et spiroxatrine ($pIC_{50} = 8.1$)) reconnaissent également le récepteur 5-HT_{1A} avec une affinité appréciable. Ces composés s'avèrent assez sélectifs vis-à-vis des autres récepteurs.

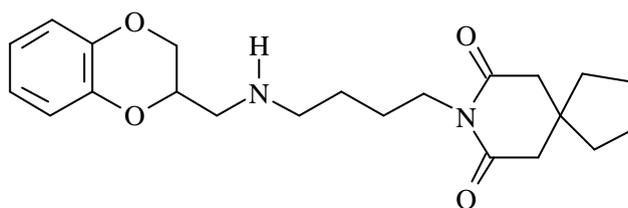


Figure 6. Représentation du MDL72832.

IV.4 - Les tétralines.

En 1981, Arvidsson et al.¹⁰⁵ ont constaté pour la première fois que la 8-hydroxy-2-di-*n*-propylaminotétraline (8-OH-DPAT) (Figure 7) avait des propriétés agonistes vis-à-vis de la 5-hydroxytryptamine au niveau du système nerveux central. Depuis lors, plusieurs équipes de chercheurs ont préparé de nombreux analogues^{106,107,108,109,110,111,112,113,114} de cette molécule, établissant ainsi des relations structure-activité entre la structure chimique de ces composés et leur action sur les récepteurs de la sérotonine.

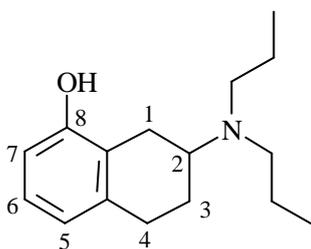


Figure 7. Représentation du 8-OH-DPAT.

Certaines informations ressortent de ces études de relation structure-activité (SAR). Une méthylation du groupement hydroxyle du 8-OH-DPAT ne change pas l'affinité de la molécule¹¹⁵. De nombreuses substitutions du groupement hydroxyle ont montré une faible perte d'activité¹¹⁶. Toute modification impliquant une substitution en position 5, 6 ou 7 du noyau aromatique conduit, soit à des agonistes dopaminergiques, soit à des molécules inactives¹¹⁷. Le raccourcissement ou la suppression d'une seule des chaînes propyles diminue peu l'affinité. Une modification d'une chaîne propyle en iodo-propényle ((*R,S*)-trans-8-OH-PIPAT)¹¹¹ donne des résultats d'activités très intéressants ($pIC_{50} = 9.5$). Par contre la suppression des deux chaînes affecte considérablement l'affinité. De nombreux composés^{118,119,120,121} obtenus à partir d'alkylation ou d'halogénéation sur le cycle saturé,

n'ont pas apporté de résultat pharmacologique suffisant pour remplacer la 8-OH-DPAT comme ligand de référence pour les études de « binding ». Une cyclisation entre l'azote et la position 1 ou 3 semble rigidifier la structure et donne des activités intéressantes selon la stéréochimie des carbones 1, 2 et 3^{110,113}. Plus récemment¹⁰⁹, une étude de l'effet de substituants aryle et hétéroaryle en position 8, a montré une faible activité *in vitro* de ces composés par rapport à la 8-OH-DPAT.

IV.5 - Les dérivés à structure benzopyranique.

En 1987, Cossery *et al.*¹⁸ a constaté suite à des études *in vitro* sur des membranes de cortex et d'hippocampe de rat que le [³H]-5-MeO-DPAC présentait une affinité aussi élevée que la 8-OH-DPAT pour les sites 5-HT_{1A}. Contrairement à la [³H]-8-OH-DPAT, le [³H]-5-MeO-DPAC ne reconnaît pas les sites présynaptiques au niveau du striatum. Ce résultat indiquait déjà la plus grande sélectivité de ce dérivé qui est en fait l'isostère oxygéné du radioligand à structure tétralinique. Les études de déplacement *in vitro* effectuées par les mêmes auteurs ont montré que le [³H]-5-MeO-DPAC ne peut pas être déplacé par des composés tels que la quipazine (ligand préférentiel des sites 5-HT_{1B/1Dβ}), la mésulergine (ligand sélectif des sites 5-HT_{2C}) et la kétansérine (antagoniste sélectif des sites 5-HT₂). Ces données confirment donc la grande sélectivité de ce ligand dérivé du benzopyrane, pour les récepteurs 5-HT_{1A}. Un travail conséquent de pharmacomodulation sur des dérivés du benzopyrane¹²² a montré l'intérêt d'une telle approche en donnant des résultats pharmacologiques très encourageants puisqu'un composé le 8-[5-[N-Propyl-N-(5-méthoxy-3,4-dihydro-2H-1-benzopyran-3-yl)amino]pentyl]-8-azaspiro[4,5]décane-7,9-dione (Figure 8) est actuellement en phase clinique.

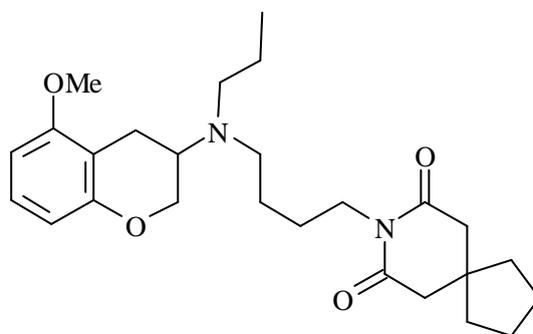


Figure 8. Représentation du 8-[5-[N-Propyl-N-(5-méthoxy-3,4-dihydro-2H-1-benzopyran-3-yl)amino]pentyl]-8-azaspiro[4,5]décane-7,9-dione.

Afin d'augmenter la sélectivité sur les sites 5-HT_{1A} par rapport aux autres récepteurs, T. Podona¹²³ et G. Guillaumet ont envisagé de rigidifier la structure de ces composés benzopyraniques. Pour obtenir cette rigidification, ils ont entrepris la synthèse de composés spiraniques (Figure 9). Avec ces composés, la sélectivité s'est accrue au détriment de l'affinité 5-HT_{1A}. Ce travail a permis d'élaborer une famille de nouveaux ligands sérotoninergiques très prometteuse.

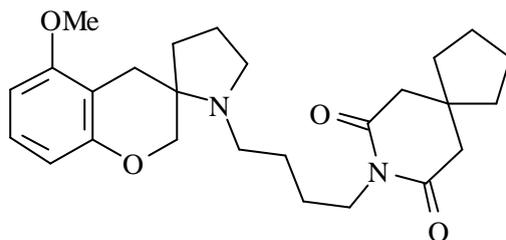


Figure 9. Représentation du composé « spiro » rigidifié ; N-[4-(8-azaspiro[4,5]décane-7,9-dione)butyl]-spiro[5-méthoxy-2*H*-1-benzopyran-3,2'-pyrrolidine).

V. LES DONNEES PHARMACOLOGIQUES

V.1 - La méthode de l'étude de liaison « binding » *in vitro*.

Les méthodes¹²⁴ de l'étude de liaison « binding » *in vitro* permettent, après marquage avec des ligands radioactifs, les sites spécifiques, de préciser leur nombre et leurs caractéristiques vis-à-vis de différents autres ligands. Ces études sont effectuées sur des homogénats de membranes cellulaires provenant de structures cérébrales bien définies. Le choix du ligand radioactif, pour les récepteurs utilisés dans ce mémoire, est donné dans le tableau suivant.

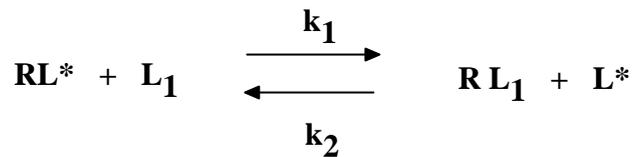
Récepteur	Ligand radioactif
5-HT _{1A}	[³ H]-8-OH-DPAT
α ₁	[³ H]-Prazosine
α ₂ (+)	[³ H]-Clonidine
α ₂ (-)	[³ H]-Rauwolscine
D ₂	[³ H]-Spipérone

Tableau 2. Etude de la liaison spécifique dans des fractions de membranes du SNC.

Les membranes des structures, préalablement disséquées, sont homogénéisées dans des solutions tampon Tris-HCL (0.05 M ; pH = 7.7) et conservées à -70°C après plusieurs lavages par ce même tampon. La sérotonine endogène est éliminée par incubation des membranes à 37°C pendant dix minutes. Dans un premier temps, la méthode de mesure consiste à incuber à des températures et pendant des temps variant selon la structure cérébrale et le ligand utilisé, les homogénats de membranes cellulaires en présence du ligand radioactif **L*** spécifique de la structure étudiée. A l'équilibre, la préparation est filtrée et la radioactivité, liée aux membranes retenues sur le filtre, est comptée en scintillation liquide.



Dans un deuxième temps, le ligand radioactif est déplacé de son site de fixation par des concentrations croissantes de la substance à étudier, \mathbf{L}_1 . Une nouvelle mesure de la radioactivité et de la connaissance de la concentration \mathbf{C} de \mathbf{L}_1 , nécessaire au déplacement total du ligand radioactif, permet d'évaluer l'affinité de la nouvelle substance pour le site étudié.



Au cours de ces essais, le radioligand \mathbf{L}^* ne se fixe pas seulement sur les sites récepteurs \mathbf{R} , mais également sur des structures non spécifiques et ce par attraction hydrophobe, électrophile ou nucléophile. La radioactivité mesurée \mathbf{X} correspond donc à la fois aux liaisons spécifiques et non spécifiques du radioligand. Pour évaluer la liaison non spécifique, la radioactivité spécifique liée aux récepteurs est déplacée par un ligand froid (5-HT pour le site 5-HT_{1A}) utilisé à une concentration suffisante pour occuper tous les sites spécifiques. La radioactivité mesurée \mathbf{Y} ne correspond donc plus qu'à la liaison non spécifique du ligand radioactif. La différence $\mathbf{X}-\mathbf{Y}$ donne donc accès à la liaison spécifique dite de « binding ».

V.2 - Quantification.

Les divers paramètres caractérisant la liaison entre le ligand étudié et le récepteur se définissent ainsi :

K_d : Constante de dissociation à l'équilibre du complexe ligand-récepteur traduisant l'affinité du ligand pour ce récepteur.

B_{max} : Capacité totale de liaison.

IC₅₀ : Concentration de ligand qui provoque 50% du déplacement de la liaison spécifique du radioligand.

$$K_i = \frac{IC_{50}}{1 + \frac{C}{K_d}}$$

K_i : Constante d'affinité.

Nous parlerons aussi de pIC₅₀ et de pK_i.

$pIC_{50} = -\text{Log } IC_{50} \text{ et } pK_i = -\text{Log } K_i$

V.3 - Interprétation des résultats.

Le pK_i et le IC_{50} constituent les mesures de l'affinité du composé testé vis-à-vis du récepteur présent dans l'homogénat de membranes cellulaires. Selon la valeur obtenue, il est possible de définir l'affinité de la substance étudiée (Tableau 3).

IC_{50} (nM)	Affinité
$< 10^{-7}$	Très active
10^{-6} à 10^{-7}	Active
10^{-5} à 10^{-6}	Peu active
$> 10^{-5}$	Inactive

Tableau 3. Classement des affinités IC_{50} .

Ce tableau 3 donne une idée généralement admise de la classification des composés selon leur valeur de «binding». Dans ce cas le seuil actif/inactif est de 6.5 unités de pIC_{50} . Dans nos études, nous n'avons pas utilisé ce classement. Le seuil actif/inactif a été choisi en fonction de l'ensemble des données d'activité. Il est de 7.5 unités de pIC_{50} . Seront considérés comme ligands inactifs, tous les composés ayant une affinité inférieure à 7.5 unité de pIC_{50} et vice versa pour les composés actifs.

Nous parlerons toujours de pIC_{50} pour exprimer l'activité vis-à-vis d'un récepteur. Nous avons donc effectué une approximation $pK_i = pIC_{50}$ pour les composés dont l'activité, donnée dans la littérature, était exprimée en K_i ou en pK_i .

Références Chapitre 1

-
- ¹ Vialli, M.; Erspamer, V. *Z.Zellforsch. Mikrosk. Anat.* **1933**, *19*, 743.
 - ² Rapport, M.M.; Green, A.A.; Page, I.H. *J. Biol. Chem.* **1948**, *176*, 1243.
 - ³ Hamlin, K.E.; Fischer, F.E. *J. Am. Chem. Soc.* **1951**, *73*, 5007-5008.
 - ⁴ Wilkinson, L.O.; Dourish, C.T. Serotonin and Animal Behavior. In Serotonin Receptor Subtypes; Basic and Clinical Aspects; Peroutka, S.; Ed.; John Wiley and Sons Ltd.: New York, **1991**; pp 147-210.
 - ⁵ Sleight, A.J.; Pierce, P.A.; Schmidt, A.W.; Hekmatpanah, C.R.; Peroutka, S.J. *The Clinical Utility of Serotonin Receptor Active Agents in Neuropsychiatric Disease. In Serotonin receptor Subtypes : Basic and Clinical Aspects*; Peroutka, S.; Ed.; John Wiley and Sons Ltd. : New York, **1991**, pp 211-227.
 - ⁶ Gaddum, J.H.; Picarelli, Z.P. *Br. J. Pharmacol.* **1957**, *12*, 323-328.
 - ⁷ Bradley, P.B.; Engel, G.; Feniuk, W.; Fozard, J.R.; Humphrey, P.P.A.; Middlemiss, D.N.; Mylecharane, E.J.; Reichardson, B.P.; Saxena, P.R. *Neuropharmacol.* **1986**, *25*, 563-576.
 - ⁸ Hartig, P.R. *Trends Pharmacol. Sci.* **1989**, *10*, 64-69.
 - ⁹ Frazer, A.; Maayani, S.; Wolfe, B.B. *Annu. Rev. Pharmacol. Toxicol.* **1990**, *30*, 307-348.
 - ¹⁰ Humphrey, P.P.A.; Hartig, P.; Hoyer, D. *Trends Pharmacol. Sci.* **1993**, *14*, 233-236.
 - ¹¹ Peroutka, S.J. *J. Neurochem.* **1993**, *60*, 408-416.
 - ¹² Martin, G.R.; Humphrey, P.P.A. *Neuropharmacol.* **1994**, *33*, 261-273.
 - ¹³ Hoyer, D.; Clarke, D.E.; Fozard, J.R.; Hartig, P.R.; Martin, G.R.; Mylecharane, E.J.; Saxena, P.R.; Humphrey, P.P.A. *Pharmacol. Rev.* **1994**, *33*, 261-273.
 - ¹⁴ Saudou, F.; Hen, R. *Med. Chem. Res.* **1994**, *4*, 16-84.
 - ¹⁵ Middlemiss, D.N.; Fozard, J.A. *Eur. J. Pharmacol.* **1983**, *90*, 151.
 - ¹⁶ Hamon, M.; Bourgoin, S.; Gozlan, H.; Hall, M.D.; Goetz, C.; Artaud, F.; Horn, A. *Eur. J. Pharmacol.* **1984**, *100*, 263.

-
- 17 Hamon, M.; Cossery, J.M.; Spampinato, U.; Gozlan, H. *Trends Pharmacol. Sci.* **1986**, 7, 336.
- 18 Cossery, J.M.; Gozlan, H.; Spampinato, U.; Perdicakis, C.; Guillaumet, G.; Pichat, L.; Hamon, M. *Eur. J. Pharmacol.* **1987**, 140, 143.
- 19 Sarda, N.; Barailler, J.; Gervais, C.; Pacheco, H. *Eur. J. Med. Chem.* **1979**, 14, 335.
- 20 Traber, J.; Gläser, T. *Trends Pharmacol. Sci.* **1987**, 8, 432.
- 21 Glennon, R.A.; Naiman, N.A.; Lyon, R.A.; Titeler, M. *J. Med. Chem.* **1988**, 31, 1968.
- 22 Williams, M. *J. Med. Chem.* **1983**, 26, 619.
- 23 Barnes, J.M.; Barnes, N.M.; Costall, B.; Naylor, R.J.; Tyers, M.B. *Nature (London)* **1989**, 338, 762.
- 24 Bockaert, J.; Fozard, J.R.; Dumuis, A.; Clarke, D.E. *Trends Pharmacol. Sci.* **1992**, 13, 141-145.
- 25 Miquel, M.C.; Hamon, M. *5-HT₁ receptor subtypes : pharmacological heterogeneity*. In : Langer, S.Z.; Brunello, N; Racagni, G.; Mendlewicz, J eds. *Serotonine Receptor Subtypes : Pharmacological Significance and Clinical Implications*. Bâle : Karger, **1992**, pp 13-30.
- 26 Julius, D.; Huang, K.N.; Livelli, T.J.; Axel, R.; Jessell, T.M. *Proc. Natl. Acad. Sci. USA* **1990**, 87, 928-932.
- 27 Kobilka, B.K.; Frielle, T.; Collin, S.; Yang Feng, T.; Kobilka, T.S.; Francke, U.; Lefkowitz, R.J.; Caron, M.G. *Nature (London)* **1987**, 329, 75-79.
- 28 Fargin, A.; Raymond, J.R.; Lohse, M.J.; Kobilka, B.K.; Caron, M.G.; Lefkowitz, R.J. *Nature* **1988**, 335, 358-360.
- 29 Albert, P.R.; Zhou, Q.Y.; Van Tol, H.H.M.; Bunzow, J.R.; Civelli, O. *J. Biol. Chem.* **1990**, 265, 5825-5832.
- 30 Vergé, D.; Daval, G.; Marcinkiewicz, M. *J. Neurosci.* **1986**, 6, 3474-3482.
- 31 Higgins, G.A.; Bradbury, A.J.; Jones, B.J.; Oakley, N.R. *Neuropharmacol.* **1988**, 27, 993-1001.
- 32 Pedigo, N.W.; Yamamura, H.I.; Nelson, D.L. *J. Neurochem.* **1981**, 36, 216.

-
- 33 Hall, M.D.; El Mestikawy, S.; Emerit, M.D.; Pichat, L., Hamon, M. Gozlan, H. *J. Neurochem.* **1985**, *44*, 1685.
- 34 Dompert, W.U.; Gläser, T.G.; Traber, J. *Arch. Pharmacol.* **1985**, *328*, 467.
- 35 Ransom, R.W.; Asarch, K.B.; Shih, J. *J. Neurochem.* **1986**, *46*, 68.
- 36 Nelson, D.L.; Taylor, E.W. *Eur. J. Pharmacol.* **1986**, *124*, 207.
- 37 Nelson, D.L.; Lambert, G.; Yamamura, J.H.; Monroe, P.J. *Life Sci.* **1987**, *41*, 1567.
- 38 Riad, M.; El Mestikawy, S.; Vergé, D., Gozlan, H.; Hamon, M. *Neurochem. Int.* **1991**, *19*, 413.
- 39 Gozlan, H; El Mestikawy, S.; Pichat, L.; Glowinski, J. Hamon, M. *Nature (London)* **1983**, *305*,140.
- 40 Moser, P.C.; Tricklebank, M.D.; Middlemiss, D.N.; Mir, A.K.; Hibert, M.F.; Fozard, J.R. *Br. J. Pharmacol.* **1990**, *99*, 343-349.
- 41 Hoyer, D.; Engel, G.; Kalkman, H.O. *Eur. J. Pharmacol.* **1985**, *118*, 1.
- 42 Waeber, C.; Dietl, M.M.; Hoyer, D. Palacios, J.M. *Naunyn Schmiedeberg's. Arch. Pharmacol.* **1989**, *340*, 486.
- 43 Hartig, P.R.; Branchek, T.A.; Weinshank, R.L. *Trends Pharmacol. Sci.* **1992**, *13*, 152.
- 44 Voigt, M.M.; Laurie, D.J.; Seeburg, P.H.; Bach, A. *EMBO J.* **1991**, *10*, 4017 .
- 45 Adham, N.; Romanienko, P. Hartig, P.; Weinshank, R.L.; Branchek, T. *Mol. Pharmacol.* **1992**, *41*, 1.
- 46 Maroteaux, L.; Saudou, F. Amlaiky, N.; Boschert, U.; Plassat, J.L.; Hen, R. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3020.
- 47 Jin, H.; Oksenberg, D.; Ashkenazi, A.; Peroutka, S.J.; Duncan, A.M.; Rozmahel, R.; Yang, Y.; Mengod, G.; Palacios, J.M.; O'Dowd, B.F. *J. Biol. Chem.* **1992**, *267*, 5735.
- 48 Weinshank, R.L.; Zgombick, J.M.; Macchi, M.J.; Branchek, T.A.; Hartig, P.R. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3630.
- 49 Simon-Chazottes, D.; Cloeztayarani, I.; Fillion, M.P., Guenet, J.L.; Fillion, G. *Mamm. Genome* **1993**, *4*, 397.

-
- 50 Engel, G.; Göthert, M. Hoyer, D.; Schlicker E.; Hillenbrand, K. *Arch. Pharmacol.* **1986**, 332, 1.
- 51 Maura, G.; Roccataglia, E.; Raiteri, M. *Arch. Pharmacol.* **1986**, 334, 323.
- 52 Libert, F.; Parmantier, M.; Lefort, A.; Dinsart, C.; Van Sande, J.; Maenhaut, C.; Simons, M.J.; Dumont, J.E.; Vassart, G. *Science* **1989**, 244, 569.
- 53 Hamblin, M.W.; Metcalf, M.A. *Mol. Pharmacol.* **1991**, 40, 143.
- 54 Hamblin, M.W.; McGuffin, R.W.; Metcalf, M.A.; Dorsa, D.M.; Merchant, K.M. *Mol. Cell. Neurosci.* **1992**, 3, 578.
- 55 Bach, A.W.J.; Unger, L.; Sprengel, R.; Mengod, G.; Palacios, J.; Seeburg, P.H.; Voigt, M.M. *J. Recept. Res.* **1993**, 13, 479.
- 56 Libert, F.; Passage, E.; Parmentier, M.; Simons, M.J.; Vassart, G.; Mattei, M.G. *Genomics* **1991**, 11, 225.
- 57 Weydert, A.; Cloez-Tayarani, I.; Fillion, M.P.; Simon-Charzottes, D.; Guenet, J.L.; Fillion, G. *C. R. Acad. Sci. III* **1992**, 314, 429.
- 58 Leonhardt, S.; Herrick Davis, K.; Titeler, M. *J. Neurochem.* **1989**, 53, 465.
- 59 Weisberg, E.; Teitler, M. *Drug. Dev. Res.* **1992**, 26, 225.
- 60 Lévy, F.O.; Gudermann, T.; Birnbaumer, M.; Kaumann, A.J.; Birnbaumer, L. *FEBS Lett.* **1992**, 296, 201.
- 61 Mc Allister, G.; Charlesworth, A.; Snodin, C.; Beer, M.S.; Noble, A.J.; Middlemiss, D.N.; Iversen, L.L.; Whiting, P. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 5517.
- 62 Zgombick, J.M.; Schechter, L.E.; Macchi, M.; Hartig, P.R.; Branchek, T.A.; Weinshank, R.L. *Mol. Pharmacol.* **1992**, 42, 180.
- 63 Gudermann, T.; Levy, F.O.; Birnbaumer, M.; Birnbaumer, L.; Kaumann, A.J. *Mol. Pharmacol.* **1993**, 43, 412.
- 64 Amlaiky, N.; Ramboz, S.; Boschert, U.; Plassat, J.L.; Hen, R. *J. Biol. Chem.* **1992**, 267, 1971.
- 65 Adham, N.; Kao, H.T.; Schechter, L.E.; Bard, J.; Olsen, M.; Urquhart, D.; Durkin, M.; Hartig, P.R.; Weinshank, R.L.; Branchek, T.A. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 408.

-
- 66 Lovenberg, T.W.; Erlander, M.G.; Baron, B.M.; Racke, M.; Slone, A.L.; Siegel, B.W.; Craft, C.M.; Burns, J.E.; Danielson, P.E.; Sutcliffe, J.G. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2184.
- 67 Leysen, J.E.; Niemegeers, C.J.E.; Tollemaere, J.P.; Laduron, P.H. *Nature (London)* **1978**, *272*, 168.
- 68 Leysen, J.E.; Niemegeers, C.J.E.; Van Nueten, J.M.; Laduron, P.H. *Mol. Pharmacol.* **1982**, *21*, 301.
- 69 Pritchett, D.B.; Bach, A.W.J.; Wozny, M.; Taleb, O.; Dal Toso, R.; Shih, J.C.; Seeburg, P.H. *EMBO J.* **1988**, *7*, 4135.
- 70 Julius, D.; MacDermott, A.B.; Axel, R.; Jessel, T.M. *Science* **1988**, *241*, 558.
- 71 Foguet, M.; Nguyen, H.; Le, H.; Lubbert, H. *NeuroReport.* **1992**, *3*, 345.
- 72 Chambard, J.C.; Van Obberghen Schilling, E.; Haslam, R.J.; Vouret, V.; Pouyssegur, J. *Nucleic Acids Res.* **1990**, *18*, 5282.
- 73 Saltzman, A.G.; Morse, B.; Whitman, M.M.; Ivanshchenko, Y.; Jaye, M.; Felder, S. *Biochem. Biophys. Res. Commun.* **1991**, *181*, 1469.
- 74 Chen, K.; Yang, W.; Grimsby, J.; Shih, J.C. *Brain Res. Mol. Brain Res.* **1992**, *14*, 20.
- 75 Sparkes, R.S.; Lan, N.; Klisak, I.; Mohandas, T.; Diep, A.; Kojis, T.; Heinzmann, C.; Shih, J.C. *Genomics* **1991**, *9*, 461.
- 76 Mengod, G.; Pompeiano, M.; Martinez Mir, M.I.; Palacios, J.M. *Brain Res.* **1990**, *524*, 139.
- 77 Foguet, M.; Hoyer, D.; Pardo, L.A.; Parekh, A.; Kluxen, F.W.; Kalkman, H.O.; Stuhmer, W.; Lubbert, H. *EMBO J.* **1992**, *11*, 3481.
- 78 Loric, S.; Launay, J.M.; Colas, J.F.; Maroteaux, L. *FEBS Lett.* **1992**, *312*, 203.
- 79 Pazos, A.; Hoyer, D.; Palacios, J.M. *Eur. J. Pharmacol.* **1984**, *106*, 539.
- 80 Hoyer, D. *Trends Pharmacol. Sci.* **1988**, *9*, 89.
- 81 Witz, P.; Amlaiky, N.; Plassat, J.L.; Maroteaux, L.; Borrelli, E.; Hen, R. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8940-8944.
- 82 Flynn, D.L.; Zabrowski, D.L.; Becker, D.P.; Nosal, R.; Villamil, C.I.; Gullikson, G.W.; Moumami, C.; Yang, D.C. *J. Med. Chem.* **1992**, *35*, 1489-1491.

-
- 83 Grossman, C.J.; Kilpatrick, G.J.; Bunce, K.T. *Br. J. Pharmacol.* **1993**, *109*, 618.
- 84 Monsma, F.J.; Shen, Y.; Ward, R.P.; Hamblin, M.W.; Sibley, D.R. *Mol. Pharmacol.* **1993**, *43*, 320.
- 85 Ruat, M.; Traiffort, E.; Arrang, J.M.; Tardivellacombe, J.; Diaz, J.; Leurs, R.; Schwartz, J.C.; *Biochem. Biophys. Res. Commun.* **1993**, *193*, 268.
- 86 Bard, J.; Zgombick, J.; Adham, N.; Branchek, T.; Weinshank, R. *Soc. Neurosci. Abstract* **1993**, *19*, 1164.
- 87 Shen, Y.; Monsma, F.J.; Metcalf, M.A.; Jose, P.A.; Hamblin, M.W.; Sibley, D.R. *J. Biol. Chem.* **1993**, *268*, 18200.
- 88 Plassat, J.L.; Amlaiky, N.; Hen, R. *Mol. Pharmacol.* **1993**, *44*, 229.
- 89 Plassat, J.L.; Boschert, U.; Amlaiky, N.; Hen, R. *EMBO J.* **1992**, *11*, 4779.
- 90 Matthes, H.; Boschert, U.; Amlaiky, N.; Grailhe, R.; Plassat, J.L.; Muscatelli, F.; Mattei, M.G.; Hen, R. *Mol. Pharmacol.* **1993**, *43*, 313.
- 91 Fozard, J.R. *Arch. Pharmacol.* **1984**, *326*, 36.
- 92 Saudou, F.; Boschert, U.; Amlaiky, N.; Plassat, J.L.; Hen, R. *EMBO J.* **1992**, *11*, 7.
- 93 Hibert, M.; Moser, P. *Drugs of the Future* **1990**, *15*, 159.
- 94 Van Steen, B.J.; Van Wijngaarden, I.; Tulp, M.T.M.; Soudijn, W. *J. Med. Chem.* **1994**, *37*, 2761-2773.
- 95 Orjales, A.; Alonso Cires, L.; Labeaga, L.; Corcostegui, R. *J. Med. Chem.* **1995**, *38*, 1273-1277.
- 96 Van Steen, B.J.; Van Wijngaarden, I.; Tulp, M.T.M.; Soudijn, W. *J. Med. Chem.* **1993**, *36*, 2751-2760.
- 97 Zhuang, Z.P.; Kung, M.P.; Kung, H.F. *J. Med. Chem.* **1994**, *37*, 1406-1407.
- 98 Ransom, R.W.; Asarch, K.B.; Shih, J. *J. Neurochem.* **1986**, *46*, 68.
- 99 Raghupathi, R.K.; Rydelek Fitzgerald, L.; Teiler, M.; Glennon, R.A. *J. Med. Chem.* **1991**, *34*, 2633-2638.
- 100 Orjales, A.; Alonso Cires, L.; Labeaga, L.; Corcostegui, R. *J. Med. Chem.* **1995**, *38*, 1273-1277.

-
- ¹⁰¹ Mokrosz, J.L.; Paluchowska, M.H.; Chojnacka Wojcik, E.; Filip, M.; Harakchieva Minol, S.; Deren Wesolek, A.; Mokrosz, M.J. *J. Med. Chem.* **1994**, *37*, 2754-2760.
- ¹⁰² Chilmonczyk, Z.; Les, A.; Wozniakowska, A.; Cybulski, J.; Koziol, A.E.; Gdaniec, M. *J. Med. Chem.* **1995**, *38*, 1701-1710.
- ¹⁰³ Van Wijngaarden, I.; Tulp, M.T.M.; Soudijn, W. *Mol. Pharmacol. Section* **1990**, *188*, 301-312.
- ¹⁰⁴ Ennis, M.D.; Baze, M.E.; Smith, M.W.; Lawson, C.F.; McCall, R.B.; Lahti, R.A.; Piercey, M.F. *J. Med. Chem.* **1992**, *35*, 3058-3066.
- ¹⁰⁵ Arvidsson, L.E.; Hacksell, U.; Nilsson, J.L.G.; Hjorth, S.; Carlsson, A.; Lindberg, P.; Sanchez, D.; Wilkström, J. *J. Med. Chem.* **1981**, *24*, 921-935.
- ¹⁰⁶ Vallgarda, J.; Arvidsson, L.E.; Svensson, B.E.; Fowler, C.J.; Hacksell, U. *Eur. J. Med. Chem.* **1993**, *28*, 399-406.
- ¹⁰⁷ Sonesson, C.; Barf, T.; Nilsson, J.; Dijkstra, D.; Carlsson, A.; Svensson, K.; Smith, M.W.; Martin, I.J.; Duncan, J.N.; King, L.J.; Wikström, H. *J. Med. Chem.* **1995**, *38*, 1319-1329.
- ¹⁰⁸ Jain, A.N.; Harris, N.L.; Park, J.Y. *J. Med. Chem.* **1995**, *38*, 1295-1308.
- ¹⁰⁹ Liu, Y.; Cortizo, L.; Yu, H.; Svensson, B.E.; Lewander, T.; Hacksell, U. *Eur. J. Med. Chem.* **1995**, *30*, 277-286.
- ¹¹⁰ Chidester, C.G.; Lin, C.H.; Lahti, R.A.; Haadsma Svensson, S.R.; Smith, M.W. *J. Med. Chem.* **1993**, *36*, 1301-1315.
- ¹¹¹ Zhuang, Z.P.; Kung, M.P.; Kung, H.F. *J. Med. Chem.* **1993**, *36*, 3161-3165.
- ¹¹² Romero, A.G.; Leiby, J.A.; Mc Call, R.B.; Piercey, M.F.; Smith, M.W.; Han, F. *J. Med. Chem.* **1993**, *36*, 2066-2074.
- ¹¹³ Lin, C.H.; Haadsma Svensson, S.R.; Phillips, G.; Mc Call, R.B.; Piercey, M.F.; Smith, M.W.; Svensson, K.; Carlsson, A.; Chidester, C.G.; Von Voigtlander, P.F. *J. Med. Chem.* **1993**, *36*, 2208-2218.
- ¹¹⁴ Stjernlof, P.; Gullme, M.; Elebring, T.; Andersson, B.; Wikstrom, H.; Lagerquist, S.; Svensson, K.; Ekman, A.; Carlsson, A.; Sundell, S. *J. Med. Chem.* **1993**, *36*, 2059-2065.
- ¹¹⁵ Naiman, N.; Lyon, R.A.; Bullock, A.E.; Rydelek, L.T.; Titeler, M.; Glennon, R.A. *J. Med. Chem.* **1989**, *32*, 253.
- ¹¹⁶ Liu, Y.; Björn, E.; Svensson, K.; Yu, H.; Cortizo, L.; Ross, S.B.; Lewander, T.; Hacksell, U. *Bio. Med. Chem. Lett.* **1991**, *5*, 257-262.

-
- ¹¹⁷ Wikström, H.; Andersson, B.; Sanchez, D.; Lindberg, P.; Arvidsson, L.E.; Johansson, A.; Nilsson, J.L.G.; Svensson, K.; Hjorth, S.; Carlsson, A. *J. Med. Chem.* **1985**, *28*, 215.
- ¹¹⁸ Mellin, C.; Liu, Y.; Hacksell, U.; Björk, L.; Anden, N.E. *Acta Pharm.* **1987**, *24*, 153.
- ¹¹⁹ Mellin, C.; Björk, L.; Karlen, A.; Johansson, A.M.; Sundell, S.; Kenne, L.; Nelson, D.L.; Anden, N.E.; Hacksell, U. *J. Med. Chem.* **1988**, *31*, 1130.
- ¹²⁰ Wikström, H.; Andersson, B.; Elebring, T.; Jacyno, J.; Allinger, N.L.; Swensson, K.; Carlsson, A.; Sundell, S. *J. Med. Chem.* **1987**, *30*, 1567.
- ¹²¹ Arvidson, L.E.; Johansson, A.M.; Hacksell, U.; Nilsson, J.L.G.; Svensson, K.; Hjorth, S.; Magnusson, T.; Carlsson, A.; Andersson, B.; Wikström, H. *J. Med. Chem.* **1987**, *30*, 2105.
- ¹²² Podona, T.; Guardiola Lemaître, B.; Caignard, D.H.; Adam, G.; Pfeiffer, B.; Renard, P.; Guillaumet, G. *J. Med. Chem.* **1994**, *37*, 1779-1793.
- ¹²³ Podona, T. Thèse soutenue le 2/12/1992 à l'université d'Orléans. 3-amino-3,4-dihydro-2H-1-benzopyranes diversement substitués : Synthèses et propriétés pharmacologiques.
- ¹²⁴ Peroutka, S.J. *Brain Res.* **1985**, *344*, 167.

CHAPITRE 2 : LES TECHNIQUES DE MODELISATION MOLECULAIRE

LES OUTILS POUR OBTENIR OU DECRIRE UNE STRUCTURE MOLECULAIRE

-

APPLICATION AUX ETUDES STRUCTURE-ACTIVITE

Dans l'industrie pharmaceutique, pour rechercher de nouveaux composés chimiques ayant un intérêt thérapeutique potentiel, la méthode classiquement utilisée consiste à synthétiser de nombreuses séries de molécules et à les soumettre à des tests pharmacologiques pour déceler une éventuelle activité biologique. Cette méthode de « screening » présente un défaut évident : son caractère aléatoire lui confère une faible rentabilité, une importante mobilisation humaine et un coût financier important. Cette méthode empirique est donc longue et onéreuse. Pourtant, la quasi-totalité des médicaments actuellement commercialisés est encore issue de ce type d'approche. L'activité biologique d'une molécule dépend en outre de sa constitution chimique et de sa structure tridimensionnelle qui déterminent ses interactions avec un récepteur. Une meilleure connaissance des relations structure-activité permet de rationaliser la mise au point d'agents thérapeutiques nouveaux. Celle-ci repose sur le développement et l'utilisation de techniques expérimentales d'analyse structurale telles que la radiocristallographie (RX) et la résonance magnétique nucléaire (RMN) d'une part, et de techniques théoriques permettant de visualiser, de manipuler les structures et d'effectuer des calculs pour caractériser leurs propriétés conformationnelles et électroniques d'autre part. La modélisation moléculaire est un outil informatique maintenant indispensable aux chimistes pour concevoir de nouvelles molécules et essayer de faire des prédictions sur l'activité biologique de composés nouveaux. Dans ce chapitre, nous présenterons les principales méthodes de modélisation moléculaire permettant la conception et l'étude de nouvelles molécules, avec pour objectif final la prédiction d'activité biologique. Ces différentes méthodes sont l'approche par étude des pharmacophores «fitting», les méthodes de relations quantitatives structure-activité (QSAR) et les études de dynamique moléculaire des complexes ligand/récepteur. Auparavant, nous présenterons les différents outils de simulation moléculaire permettant d'obtenir et/ou de décrire les structures moléculaires. Ces outils sont les méthodes statiques comprenant les méthodes quantiques et la mécanique moléculaire, les méthodes dynamiques et les recherches de coordonnées de molécules soit sur des bases de données cristallographiques soit à partir d'informations structurales issue de la résonance magnétique nucléaire (RMN).

I. LES OUTILS POUR OBTENIR OU DECRIRE UNE STRUCTURE MOLECULAIRE

I.1 - Les méthodes statiques.

I.1.A - Les méthodes quantiques.

Les méthodes *ab initio*

La mécanique quantique est la seule approche exacte -si tant est qu'il en existe une- de description d'une molécule. La distribution des particules (noyau et électron) définit la totalité de son comportement, y compris sa géométrie, sa stabilité et sa réactivité. Cette information est obtenue par un jeu d'équations complexes impliquant la résolution d'un nombre important d'intégrales. Dans le contexte de la théorie MO, *ab initio* signifie qu'aucun paramètre empirique n'est utilisé pour la résolution de l'équation de Schrödinger. Seules quelques approximations sont utilisées (approche non relativiste, théorie CLOA, approche de Born-Oppenheimer). Les calculs *ab initio* effectués sur ordinateur sont très coûteux en temps de calcul (CPU) puisque tous les électrons sont considérés. De plus cette méthode ne peut être appliquée qu'à des molécules au nombre limité d'atomes. Ils présentent toutefois l'avantage de pouvoir être appliqués à des molécules ou des fragments moléculaires pour lesquels aucune donnée expérimentale n'est accessible.

Les méthodes semi-empiriques

Les méthodes semi-empiriques¹, bien que fondamentalement méthodes quantiques, sont considérées comme intermédiaires entre la mécanique moléculaire et les méthodes *ab initio*. Dans les méthodes semi-empiriques, les électrons de coeur, c'est à dire ceux qui ne sont pas impliqués dans des liaisons et qui sont les plus proches des noyaux atomiques, sont négligés.

Dans les différentes méthodes semi-empiriques du type CNDO/2 (Complete Neglect of Differential Overlap)², INDO (Intermediate Neglect of Differential Overlap)³, NDDO (Neglect of Diatomic Differential Overlap)⁴, MNDO (Modified Neglected of Differential Overlap)⁵, AM1 (Austin Model 1)⁶, PM3 (Parametric Method 3)⁷, seuls les électrons de valence sont pris en considération. Ces méthodes rendent compte des phénomènes de répulsion électron-électron, mais négligent complètement les orbitales de recouvrement orbitalaire. Pour palier cet inconvénient, ces méthodes font intervenir différents niveaux d'approximation dans le calcul des éléments de la matrice de Fock. L'avantage d'un programme de calcul semi-empirique réside dans sa rapidité de calcul.

Par ces méthodes, nous avons accès à différentes propriétés moléculaires comme les charges atomiques, les énergies des orbitales (HOMO, LUMO) entre autres, le potentiel d'ionisation, l'enthalpie de formation, la répartition électronique, etc.... Dans le travail de cette thèse, certaines de ces propriétés ont été déterminées avec le logiciel MOPAC (V5 et V6)⁸.

I.1.B - La mécanique moléculaire.

En mécanique moléculaire, une molécule est décrite comme un édifice polyatomique statique dont la cohésion est le fruit des interactions entre les différents atomes constitutifs. La mécanique moléculaire a pour buts principaux de décrire et de prédire la structure et la stabilité du système étudié, de même que l'énergie des différents états conformationnels.

Pour ce faire, les logiciels de modélisation moléculaire utilisent des champs de forces appropriés. Certains champs de forces sont adaptés aux macromolécules (GROMOS, DISCOVER, CHARMM, X-PLOR, AMBER), d'autres sont plus généralistes (MM2, MM3, COSMIC).

I.1.B.1 - Les champs de forces.

Comme la résolution mathématique rigoureuse de l'équation du système n'est pas possible actuellement, une approche empirique de représentation de l'énergie a été développée : c'est le champ de forces ou fonction d'interaction, ou plus simplement fonction d'énergie potentielle. L'expression mathématique du potentiel devra donner l'énergie du système en fonction des positions atomiques. La méthode de détermination des paramètres qui y entrent en jeu est fondée sur des calculs *ab initio*⁹, des expériences spectroscopiques (IR, Raman, ...) et des données de structures cristallines de petites molécules.

Il existe plusieurs champs de forces qui sont couramment utilisés. En voici quelques uns :

- COSMIC¹⁰, MM2¹¹, AMBER^{12,13}, CHARMM¹⁴, GROMOS¹⁵.

L'expression des potentiels que nous avons utilisée, lors de nos travaux, est celle de COSMIC pour les molécules de petite taille (nombre d'atome inférieur à 60) et GROMOS pour les macromolécules (Schéma 6).

GROMOS

$$V = E_b + E_a + E_t + E_{di} + [E_{vdw} + E_c] S(R_{ij})$$

$E_b = \sum 1/2 k_b (b-b_0)^2$ pour les liaisons

$E_a = \sum 1/2 k_\theta (\theta-\theta_0)^2$ pour les angles de valence

$E_t = \sum k_\phi (1 + \cos(n\phi - \delta))$ pour les angles diédres

$E_{di} = \sum 1/2 k_\xi (\xi-\xi_0)^2$ pour les diédres impropres

$E_{vdw} = \sum [(C_{12}(i,j)/R_{ij}^{12}) - (C_6(i,j)/R_{ij}^6)]$ pour les interactions Lennard-Jones

$E_c = \sum [(q_i q_j)/(4\pi\epsilon_0\epsilon_r R_{ij})]$ pour les interactions de Coulomb

$S(R_{ij})$ est la fonction « switch »

COSMIC

$$V = E_b + E_a + E_t + E_{vdw} + E_c$$

$E_b = \sum 1/2 k_b (b-b_0)^2$ pour les liaisons

$E_a = \sum 1/2 k_\theta (\theta-\theta_0)^2$ pour les angles de valence

$E_t = \sum V (1 + \text{sign}(\#) \cos \# \delta)/2$ pour les angles diédres

V = barrière énergétique de rotation

δ = angle de torsion

$\#$ = rotation

$\text{sign} = +1$ si $\# \geq 0$ ou -1 si $\# < 0$

$E_{vdw} = -E_{\min} (z^2 - 2z)$ pour les interactions Lennard-Jones

E_{\min} = profondeur du puit de potentiel (vdW)

$z = \exp(b(1-R))$

$b = 5.643$

$R = r/\text{somme des rayons de vdW}$

r = distance interatomique

$E_c = (q_i q_j / D r)$ pour les interactions de Coulomb

q_i et q_j = charges atomiques

D = constante diélectrique

r = distance interatomique

Schéma 6. Equations des champs de force COSMIC et GROMOS.

Toutes les constantes de force k , les paramètres indicés 0 ainsi que ceux qui dépendent des indices i et j sont des données du champ de forces considéré, déterminées comme nous l'avons indiqué précédemment.

Chacun des termes de cette fonction est assez simple. On distingue les termes qui décrivent la géométrie interne (longueurs des liaisons, angles de valence et angles dièdres) et ceux qui traduisent les interactions entre atomes non liés que sont les interactions répulsives ($1/R^{12}$) et attractives ($1/R^6$) de van der Waals (Lennard-Jones) et les interactions électrostatiques (Coulomb). Ces dernières diffèrent pour les deux champs de forces considérés COSMIC et GROMOS.

Les champs de forces se décomposent donc en termes internes et externes.

Les termes internes.

Les fluctuations sur les longueurs de liaison et les angles de liaison existent à température ambiante. Pour exprimer la contribution des liaisons covalentes, on considère simplement chaque liaison comme un ressort de constante de force k_b dont la position d'équilibre est b_0 . On utilise le même type de contrainte harmonique pour décrire les angles de valence avec cette fois-ci θ_0 et k_θ . Les valeurs de b_0 et θ_0 sont des valeurs théoriques d'équilibre déduites des études cristallographiques à haute résolution sur des petites molécules. Elles sont obtenues à partir d'analyses vibrationnelles.

Les angles dièdres sont des angles qui ont la possibilité d'avoir plusieurs valeurs préférentielles. Les variations de l'énergie potentielle liées aux rotations autour des liaisons simples peuvent être représentées par des fonctions sinusoïdales. Chaque terme est caractérisé dans le champ de force GROMOS, par une constante k_ϕ , sa périodicité n et une phase δ .

Pour le champ de forces GROMOS, les dièdres impropres sont des angles dièdres spéciaux, définis pour maintenir les configurations tétraédrales, la planéité des cycles, ou celle de la liaison peptidique par exemple. Il est essentiel pour maintenir la planéité de groupes comme ceux contenant des carbones sp^2 . On leur fait correspondre une constante de force car il est important que la structure qu'ils décrivent soit correcte.

Les termes externes (interactions électrostatiques et de van der Waals).

Les forces non covalentes ont été découvertes en 1873 par van der Waals alors qu'il tentait d'expliquer la déviation d'un gaz réel par rapport à la loi des gaz parfaits. Elles sont très importantes pour les systèmes biologiques et leur forme analytique peut être déduite de considérations de mécanique quantique. Les forces agissant sur chaque noyau dans une molécule sont obtenues à partir des lois de la théorie de l'électrostatique classique en considérant le système de charges formé par les noyaux et la répartition de charges électroniques la plus probable. Ainsi, toutes les interactions externes entre les molécules représentent un effet d'origine électrostatique. Ce terme externe est des plus importants, car il exprime les interactions qui se font à travers l'espace, qu'elles soient attractives ou répulsives. Leur expression regroupe un potentiel de type Lennard-Jones (6-12) et un potentiel de Coulomb.

Pour le potentiel de type Lennard-Jones, la représentation d'un atome doit rendre compte des interactions de dispersion et de répulsion. Des forces de dispersion apparaissent entre les atomes, même s'ils sont apolaires. En effet, à chaque instant un atome peut être considéré comme un dipôle instantané généré par l'asymétrie de la densité électronique autour du noyau. Ces dipôles donnent naissance à un champ électrique fluctuant qui polarise les autres atomes. Le moment dipolaire, induit par le champ créé par un atome i sur un atome j , dépend de R_{ij}^{-3} et de la polarisabilité atomique de l'atome j . L'énergie d'interaction de ce dipôle dans le champ électrique, moyennée sur toutes les orientations possibles du dipôle, est proportionnelle au produit scalaire de ces deux vecteurs et dépend donc de R_{ij}^{-6} . Les forces attractives entre les paires d'atomes sont limitées par la répulsion électronique des atomes lorsqu'ils viennent en contact. Lennard-Jones a représenté le terme de répulsion par une fonction variant comme R_{ij}^{-12} et Buckingham l'a décrite par une fonction exponentielle utilisée entre autres dans le logiciel COSMIC. Les paramètres $C_{12}(ij)$ et $C_6(ij)$ peuvent être évalués à partir des polarisabilités atomiques. Actuellement, ces paramètres sont ajustés à partir de résultats de spectrométrie hertzienne ou de la cristallographie par diffraction des rayons X^{16,17}. Ces termes ne sont évalués que pour les atomes non liés.

Pour la loi de Coulomb, la plupart des atomes liés de façon covalente portent une charge partielle, même si la molécule est neutre, car des liaisons covalentes entre des atomes de types différents conduisent à une distribution électronique asymétrique. Ces charges partielles peuvent être

obtenues à partir de calculs de mécanique quantique. En supposant un modèle de charge ponctuelle pour les atomes, le potentiel d'interaction correspondant peut être exprimé selon la loi de Coulomb où q_i et q_j sont les charges partielles des atomes i et j . R_{ij} est la distance entre ces deux atomes, ϵ_0 la permittivité du vide et ϵ_r la permittivité relative. Le rôle de ce dernier terme est souvent de simuler les effets du solvant autour de la molécule si celui-ci n'est pas introduit explicitement.

Le calcul de ces termes énergétiques (Lennard-Jones et Coulomb) pour toutes les paires d'atomes non liés d'une macromolécule prendrait un temps assez considérable car le nombre d'opérations à effectuer est proportionnel à N^2 , N étant le nombre d'atomes de la molécule. Ces derniers termes sont donc multipliés par une fonction $S(R_{ij})$ dite « switch » pour le champ de forces GROMOS. En effet, l'évaluation numérique de ce dernier terme est celui qui nécessite le plus de temps CPU qui peut être très conséquent pour les macromolécules. On tente donc toujours de réduire ce temps en prenant en compte le moins d'interactions possibles. Cela se traduit par une distance limite « cut-off » comprise généralement entre 7 et 12 Å au-delà de laquelle la fonction potentielle n'est plus évaluée. De manière à assurer la continuité du potentiel et de ses dérivées, la coupure ne se fait pas brusquement¹⁸ mais en utilisant une fonction polynomiale « switch » (Figure 10)¹⁹.

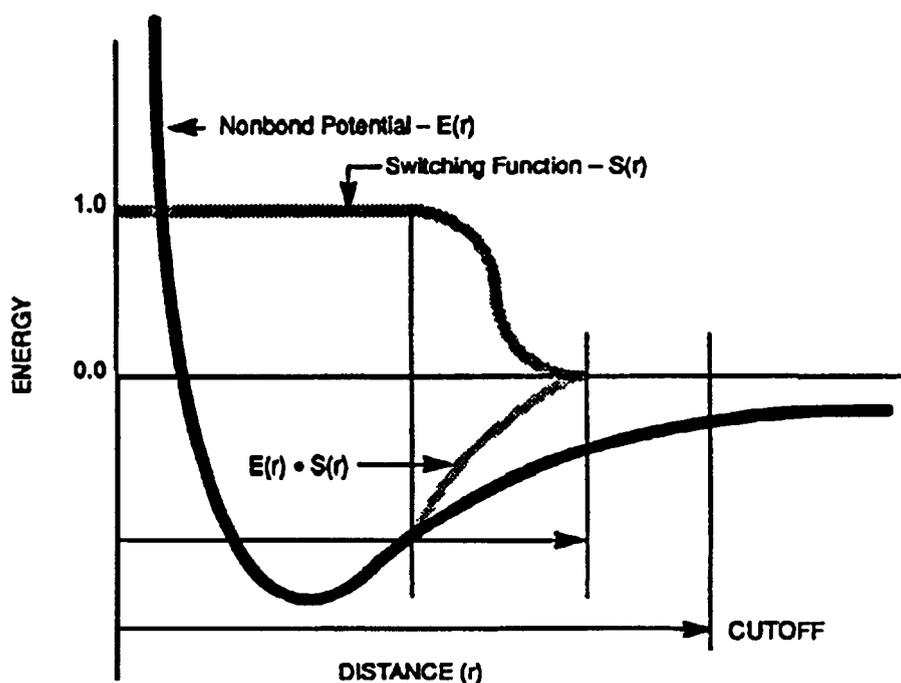


Figure 10. Fonction d'interaction entre atomes non-liés, avec la distance limite « cut-off » et la fonction de « switch ». On donne également la résultante.

Comme il est possible de l'imaginer, le principal défaut des champs de forces actuels vient de leur traitement très approximatif des charges ponctuelles, de l'absence d'effet de polarisabilité pourtant important²⁰ ainsi que d'effet à longue distance. Mais on considère que ce type de modèle est suffisant pour parcourir l'espace conformationnel des macromolécules et valider certains modèles²¹.

Ces champs de forces sont utilisés dans les méthodes de minimisation d'énergie, les méthodes de recherche conformationnelle, la dynamique moléculaire et les analyses des modes normaux de vibration. On les utilise de plus en plus pour l'affinement des structures obtenues par diffraction X ou par RMN²², mais aussi pour l'obtention des conformères de plus basse énergie des molécules selon des protocoles bien précis.

I.1.B.2 - Les charges atomiques.

Les charges partielles atomiques interviennent dans l'expression de l'énergie potentielle par l'intermédiaire du calcul du terme de Coulomb précédemment explicité. Ces charges sont aussi considérées dans le calcul de certains descripteurs moléculaires et dans les calculs de similarités moléculaires. Dans toutes les études QSAR (Quantitative Structure-Activity Relationship), les charges partielles atomiques de toutes les molécules ont été calculées avec le programme CHARGE-2 suivant une procédure empirique. Ce programme prend en compte les effets inductifs pour les systèmes saturés conjugués de liaison (σ) et le calcul des orbitales moléculaires de Hückel (HMO) pour les systèmes insaturés non conjugués (π). Les atomes sont caractérisés par leur électronégativité orbitale.

Pour la détermination de σ , trois paramètres sont requis. Pour le calcul de la charge atomique d'un atome A, le calcul prend en considération les effets de l'atome B en α , de l'atome C en β et de l'atome D en δ (enchaînement des atomes : A-B-C-D). La charge atomique de l'atome A est la somme de ces trois effets ($q^\alpha, q^\beta, q^\delta$).

$$q^\alpha A = \frac{(E_B - E_A)}{\alpha_{AB}}$$

E_A et E_B sont l'électronégativité des atomes A et B, α_{AB} est une constante.

L'effet q^β prend en compte l'électronégativité (E_C) de l'atome C et la polarisabilité de l'atome A (P_A).

$$q^\beta A = \frac{(E_C - E_H)P_A}{b}$$

E_H est l'électronégativité de l'atome d'hydrogène H, b est une constante.

L'effet q^δ est directement proportionnel à l'effet q^β .

$$q^{\delta} A = \frac{\beta_{AD}}{5}$$

En ce qui concerne l'évaluation des charges π , la méthode est basée sur la théorie HMO. Dans ce cas, la charge atomique partielle est la somme de deux termes α_r et β_{rs} .

$$\alpha_r = \alpha_0 + h_r \beta_0$$

$$\beta_{rs} = k_{rs} \beta_0$$

α_0 et β_0 sont les intégrales de Coulomb et de résonance pour les orbitales $2p_z$ des atomes de carbone. h_r et k_{rs} sont des facteurs utilisés pour modifier ces intégrales pour les orbitales autres que celles des carbones sp^2 .

La charge partielle atomique totale de la molécule est donc la somme des charges σ et π .

I.1.B.3 - La minimisation d'énergie.

Dès lors, nous disposons d'une fonction d'énergie potentielle clairement définie ; nous pouvons donc envisager ses diverses applications. La plus simple consiste à trouver le minimum de l'énergie potentielle. Notons que ce problème est plus complexe qu'il n'y paraît car il s'agit alors de trouver le minimum absolu d'une fonction non linéaire de $3N$ variables. Différentes méthodes de minimisation existent qui varient par leur complexité et leur aptitude à atteindre rapidement la convergence²³. Dans notre travail, nous n'avons utilisé que des méthodes dites du premier ordre, c'est-à-dire utilisant le gradient de la fonction d'énergie. Elles sont rapides et relativement simples à mettre en oeuvre. Les méthodes du deuxième ordre (Newton-Raphson)^{24,25} qui utilisent les dérivées secondes de la même fonction sont plus lentes et plus complexes.

Dans les méthodes du premier ordre, on calcule le déplacement des atomes de la molécule dans la direction de gradient le plus négatif. Les modalités de l'implémentation d'un tel principe varient avec la méthode. Ainsi, dans la méthode de la plus grande pente ou « Steepest Descent (SD) »^{26,27}, on utilise un pas variable de déplacement ajusté à chaque itération. Lorsque l'énergie calculée pour la nouvelle configuration est plus faible que la précédente, ce pas augmente. Il diminue dans le cas contraire. Cette méthode est très efficace quand on est loin d'un minimum mais ne converge que lentement par la suite.

La seconde méthode employée est celle du gradient conjugué ou « Conjugated Gradient (CG) »²⁸. Cet algorithme permet de prendre en compte les valeurs du gradient mais aussi la direction de déplacement de l'étape précédente. On obtient en général une convergence meilleure que dans le cadre de la méthode de la plus grande pente. L'algorithme utilise simultanément les résultats de deux étapes successives pour déterminer l'étape suivante.

Nous voyons donc que ces deux méthodes sont complémentaires. En général, on utilise le plus souvent une première minimisation par SD suivie de plusieurs étapes de CG pour affiner la convergence vers le potentiel minimum. Ces méthodes, aussi complexes soient elles, ne peuvent explorer en totalité l'espace conformationnel disponible pour une macromolécule ; par contre elles sont largement suffisantes pour une petite molécule si la conformation de départ est proche du potentiel minimum. Il est donc *a priori* impossible d'être certain d'avoir atteint le minimum absolu de l'hypersurface avec cette méthode de minimisation.

Les méthodes que nous avons décrites ci-dessus présentent l'inconvénient de converger vers des minima locaux d'énergie potentielle puisqu'elles ne permettent pas de franchir des barrières de potentiel, même de faible amplitude. Pour explorer davantage l'espace conformationnel, on a recours à d'autres méthodes. Ces méthodes sont soit la dynamique moléculaire (MD) (Figure 11), soit la méthode de Monte Carlo. Ces approches seront discutées plus tard dans ce chapitre.

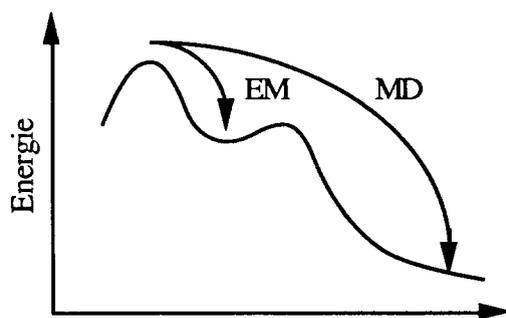


Figure 11. Chemin énergétique.

I.1.B.4 - L'analyse conformationnelle.

Bien que les prémices de l'analyse conformationnelle soient dues aux chimistes van't Hoff²⁹ et Le Bel³⁰ au 19^{ème} siècle, l'analyse conformationnelle dite « moderne » est attribuée à Barton³¹, et aux travaux de Hassel³². L'analyse conformationnelle consiste à explorer et à étudier, grâce à diverses stratégies, les différentes conformations que peut adopter une molécule dans des conditions données. Chaque conformation peut être comparée à toutes les autres par la valeur de l'énergie lui correspondant. L'objectif principal de la recherche conformationnelle est de déterminer le conformère ayant l'énergie potentielle minimum et de classer les conformères en familles structurales.

Dans ce paragraphe, nous allons décrire les méthodes les plus couramment employées pour procéder à une analyse conformationnelle. Cette approche est nécessaire ; il est clairement défini que l'activité biologique d'une molécule est étroitement liée à sa conformation. Toutefois, il est évident que ce n'est pas forcément le conformère de plus basse énergie qui interagit avec le site actif d'un récepteur donné ; un transfert d'énergie étant envisageable entre le ligand et le récepteur afin de compenser le coût énergétique nécessaire au « binding ».

La recherche systématique

La seule solution exacte permettant de trouver la conformation la plus stable d'une molécule, donc celle d'énergie la plus basse, est de procéder au calcul systématique de toutes les conformations que peut adopter une molécule. Si par exemple, une molécule n'a qu'une seule liaison simple autour de laquelle un groupement peut tourner par rapport au reste de la molécule, il suffit de calculer toutes les géométries obtenues par rotations successives de, par exemple, 30° autour de cette liaison jusqu'à avoir fait un tour complet, soit, dans ce cas 12 conformations différentes. La limitation de cette approche tient à son caractère combinatoire, la plupart de molécules, même petites, ayant très souvent bien plus d'un degré de liberté. Pour contourner ce problème, une approche a été imaginée ; la méthode de Monte Carlo.

La méthode de Monte Carlo

Quand le nombre de degrés de liberté est élevé, le seul recours est la méthode de Monte Carlo, méthode statistique ainsi nommée parce qu'elle fait appel au hasard. L'idée est de décider par tirage aléatoire quelle est la valeur de l'angle de rotation que l'on va modifier, et de combien de degrés va tourner le fragment connecté. Cette opération est répétée un grand nombre de fois, afin d'obtenir un échantillonnage représentatif des conformations possibles. Bien que parfois remarquablement efficace, cette méthode ne permet jamais d'être sûr de la validité de l'hypothèse selon laquelle la conformation la plus basse est bien l'une des conformations explorées.

Il existe de multiples méthodes appliquant le principe de Monte Carlo^{33,34}. Pour notre part, nous avons utilisé une méthode qui utilise les techniques de l'intelligence artificielle. Cette méthode, COBRA^{35,36}, est un nouveau système d'analyse automatique des conformations et de génération de structures moléculaires 3D. Ce programme recherche rapidement dans l'espace conformationnel les conformations de plus basse énergie et les classe en fonction de leur stabilité relative. La molécule est décomposée en unités moléculaires (cycles saturés, cycles aromatiques, chaînes alkyles, allènes, alcool, etc...). Toutes les combinaisons possibles des conformations de ces unités sont étudiées afin d'explorer l'espace conformationnel de la molécule en question. Si la construction de la molécule occasionne des violations, comme par exemple, des énergies de van der Waals très positives dues à

des contacts stériques entre des atomes, alors la construction est abandonnée et une autre solution est proposée. Lorsque toutes les unités de la molécule sont assemblées, l'énergie de la conformation est calculée en utilisant le champ de force COSMIC et le conformère est alors stocké. Les avantages d'une telle méthode sont multiples, comme par exemple la rapidité d'exécution de la recherche conformationnelle par rapport à une étude Monte Carlo classique.

I.2 - La dynamique moléculaire.

Les premières applications de la dynamique moléculaire Newtonienne ont été effectuées sur des fluides^{37,38}. Des études plus complexes sur les liquides et les premières simulations de biomacromolécules ont été effectuées vers les années 1970^{39,40}. Aujourd'hui, les simulations de dynamique moléculaire sont appliquées à la détermination de la structure et la dynamique interne des systèmes moléculaires et macromoléculaires. La dynamique moléculaire est une méthode utilisée pour échantillonner des régions de l'espace conformationnel qui ne peuvent pas être atteinte par d'autres méthodes. Cette recherche conformationnelle est souvent effectuée à haute température et l'énergie cinétique importante qui en résulte permet de franchir des barrières d'énergie potentielle séparant des régions intéressantes.

La méthode utilise la fonction d'énergie potentielle des atomes constitutifs de la molécule fonction de leurs coordonnées relatives.

Les équations classiques du mouvement sont résolues numériquement dans un intervalle de temps Δt de manière à produire la trajectoire des atomes.

La force

$$(F_{ij} = m_i d^2r_{ij}/dt^2)$$

est la résultante des forces agissant à chaque instant sur l'atome i due aux interactions avec les autres atomes (m_i est la masse de l'atome i). Cette force est calculée à partir de la fonction d'énergie potentielle.

$$F_{ij} = -\delta V(r_i, \dots, r_n) / \delta r_{ij}$$

L'expression de la force est généralement complexe et l'équation de Newton est résolue dans un intervalle de temps δt très petit ($\approx 1\text{fs}$) pour que toutes les composantes de la force soient considérées constantes pendant cet intervalle.

Plusieurs algorithmes sont utilisés pour résoudre les équations du mouvement. Ils diffèrent par les ordres dans les approximations utilisées lors du développement de fonctions en série de Taylor. L'algorithme le plus utilisé est l'algorithme de Verlet⁴¹ où les développements en série de Taylor de $r(t+\delta t)$ et $r(t-\delta t)$ conduisent aux relations suivantes :

$$r_{ij}(t + \delta t) = r_{ij}(t) + \delta t V_{ij}(t) + 1/2 \delta t^2 a_{ij}(t) + \dots$$

$$r_{ij}(t - \delta t) = r_{ij}(t) - \delta t V_{ij}(t) + 1/2 \delta t^2 a_{ij}(t) - \dots$$

d'où

$$r_{ij}(t + \delta t) = 2r_{ij}(t) - r_{ij}(t - \delta t) + \delta t^2 a_{ij}(t)$$

V et a sont respectivement les composantes de la vitesse et de l'accélération.

L'énergie cinétique est calculée à partir des vitesses V_{ij} .

$$V_{ij}(t) = (r_{ij}(t+\delta t) - r_{ij}(t-\delta t)) / 2\delta t$$

Un autre algorithme souvent utilisé est l'algorithme de «Leap-Frog⁴²». Celui-ci est utilisé dans le logiciel GROMOS.

La position et les vitesses sont calculées suivant les relations :

$$V_{ij}(t+(\delta t/2)) = V_{ij}(t-(\delta t/2)) + a_{ij}(t) \cdot \delta t$$

$$r_{ij}(t+(\delta t/2)) = r_{ij}(t-(\delta t/2)) + \delta t V_{ij}(t+(\delta t/2))$$

Ces relations permettent d'effectuer les calculs avec un couplage à un bain thermique.

Tous les calculs doivent être faits à chaque étape et pour chaque coordonnée atomique. Le nombre d'opérations nécessaires pour une macromolécule est considérable. Pour limiter le temps de calcul (CPU), un des moyens utilisés est de modifier la valeur de l'incrément de temps δt en introduisant des contraintes sur les mouvements de plus haute fréquence. Des contraintes peuvent donc être utilisées sur les longueurs de liaison : en effet, la valeur δt est théoriquement imposée par la valeur de la fréquence la plus élevée dans le système étudié. Elle doit vérifier la relation ci après.

$$\delta t \ll v_{\max}^{-1}$$

Si des mouvements de fréquence relativement faible ne sont pas corrélés aux mouvements de plus haute fréquence, alors des contraintes sont applicables sur ces hautes fréquences. C'est le cas des vibrations de longueurs de liaisons par rapport aux mouvements atomiques de la molécule. Un facteur de 2 à 4 peut être gagné en temps de calculs. La méthode couramment utilisée pour appliquer de telles contraintes est celle dite de SHAKE⁴³. Elle consiste à ne pas calculer les termes correspondant aux contraintes imposées dans la fonction potentielle. Ceci permet de gagner un facteur 3 si toutes les liaisons covalentes sont contraintes. La méthode de contrainte SHAKE a été utilisée dans tous nos calculs effectués sur les complexes ligand/récepteur 5-HT_{1A} (environ 1500 atomes).

Les études de dynamique moléculaire sont souvent effectuées à température constante, en considérant le système canonique (le nombre de particules, le volume et la température constant le long de la trajectoire). On sait que la température est calculée à partir de la valeur de l'énergie cinétique suivant le théorème d'équipartition d'énergie.

$$E_{\text{cin}} = \frac{1}{2} N_d K T(t) \quad N_d : \text{nombre de degrés de liberté}$$

Pour maintenir la valeur de cette température $T(t)$ à chaque instant égale à une constante T_0 , la méthode la plus adaptée et préférée consiste à coupler le système à un bain thermique⁴⁴. Dans ce cas, les équations du mouvement sont modifiées. A chaque étape, la vitesse V d'un atome est ajustée à une nouvelle valeur λV qui est fonction de la température de référence T_0 .

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau} \left(\frac{T_0}{T} - 1 \right)}$$

τ est la constante de couplage au bain thermique. En fait, la température étant calculée en fonction des vitesses, celles-ci sont remises à jour à chaque étape de la simulation. Les coordonnées atomiques sont alors modifiées en conséquence. Cette méthode de couplage à un bain thermique a été utilisée dans notre protocole de simulation présenté dans le chapitre 4.

I.3 - Les données expérimentales de cristallographie et de RMN.

Les données de la cristallographie aux rayons X et la diffraction électronique sont des informations fondamentales pour la construction des structures moléculaires. Ces données sont obtenues, soit par recherches bibliographiques soit par interrogation de banques de données structurales ; Cambridge Structural Database (CSDB)⁴⁵ qui répertorie les dizaines de milliers de structures de petites molécules (>120000) obtenues par diffraction des rayons X et Protein Data Bank (PDB)⁴⁶ pour les macromolécules répertoriées. En sortie de ces banques de données, les structures peuvent présenter des anomalies nécessitant des manipulations graphiques et obligeant par conséquent une relaxation énergétique du système. En effet, les contraintes conformationnelles dues aux forces d'empilement dans la maille cristalline en phase solide (RX) sont différentes de celles existantes en milieu liquide ou dans le vide qui sont les milieux couramment utilisés en modélisation moléculaire.

On peut également utiliser des données issues d'autres méthodes d'analyse spectroscopique, en particulier de la résonance magnétique nucléaire (RMN), qui permet sur la base de calculs de distances inter-protons de décrire les structures moléculaires en solution. Ces paramètres, les constantes de couplage 3J et les intensités nOes (nuclear Overhauser effect), sont classiquement incorporés dans des calculs de minimisation d'énergie et sont utilisés pour définir une conformation satisfaisante. Dans le cas des macromolécules, la RMN est très employée, d'une part parce qu'il est souvent difficile d'obtenir un cristal et d'avoir une résolution rapide, d'autre part parce qu'elle permet d'accéder à certains phénomènes de dynamique interne que la cristallographie ignore, tels que certains mouvements et relaxations.

Pour l'étude des complexes ligand/récepteur, la construction du récepteur 5-HT_{1A} est basée sur les coordonnées cartésiennes de la Bactériorhodopsine extraite de la Protein Data Bank (PDB).

II. APPLICATION AUX ETUDES STRUCTURE-ACTIVITE

II.1 - L'approche par pharmacophores.

II.1.A - Présentation de la méthode.

Dans l'étude des systèmes drogues/récepteur, deux méthodes se sont développées durant ces dernières années. Lorsque le récepteur est caractérisé structuralement soit par RX, RMN ou tout simplement par des études d'homologie, on utilise l'approche du « receptor fitting » où le récepteur est visualisé sur écran graphique et où l'on cherche à déterminer la molécule présentant la meilleure complémentarité au niveau du site actif soit par des méthodes de mécanique moléculaire soit par des études de dynamique moléculaire des complexes drogue/récepteur. Cette approche sera traitée dans le chapitre 4. Mais en général, si les récepteurs sont pharmacologiquement bien définis, leurs structures chimiques peuvent être complètement inconnues. On essaye alors de concevoir un modèle tridimensionnel appelé pharmacophore (Figure 12) qui rassemble les propriétés topographiques et physico-chimiques communes aux molécules agissant sur le récepteur considéré et essentielles à leur activité. Pour cette approche, il s'agit de la méthode du « receptor mapping ». Si l'on n'a aucune donnée de pharmacophore, la méthode souvent exploitée, consiste en une comparaison des molécules par une méthode de superposition (« fitting ») entre celles-ci. Cette approche tridimensionnelle est la méthode AAA (« Active Analogue Approach »).

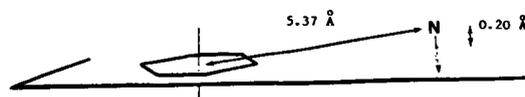


Figure 12. Exemple d'un pharmacophore du site agoniste 5-HT_{1A}.

La comparaison de deux molécules (AAA) ou d'une molécule avec un pharmacophore est un problème classique en modélisation moléculaire. Cette méthode est classiquement utilisée dans l'industrie pharmaceutique surtout pour des raisons de temps de calculs. En effet ces méthodes

demandent beaucoup moins de temps CPU que les méthodes QSAR ou les simulations dynamiques des récepteurs. Il faut donc, sur la base de deux modèles structuraux préalablement optimisés, rechercher la ou les superpositions moléculaires mettant en évidence le plus grand nombre de caractéristiques géométriques et structurales communes. La technique généralement adoptée consiste à comparer deux molécules (une flexible et une rigide) ou une molécule flexible et un pharmacophore de structure rigide. La conformation de la molécule flexible est alors ajustée de façon à obtenir la superposition la plus intéressante (Figure 13). La superposition est quantifiée par une valeur de RMS (« Root Mean Squares »).

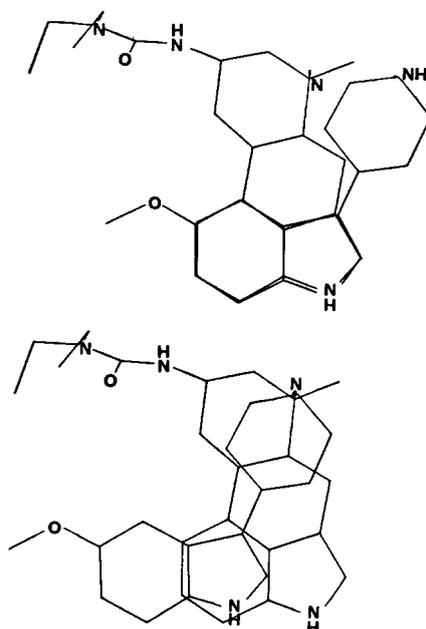


Figure 13. Deux superpositions possibles pour le RU24969 et le lisuride.

Il existe plusieurs méthodes de superposition :

- **La superposition rigide** : Les longueurs de liaisons, les angles de valence et les angles de torsion des deux molécules ne varient pas, l'algorithme de superposition rigide généralement utilisé est celui de la méthode des moindres carrés « Least Squares Fitting Method » qui consiste à déterminer la superposition géométrique optimale obtenue en minimisant la somme des carrés des distances entre chaque couple d'atomes spécifiés au préalable. Cette approche est valable avec deux molécules ou avec une molécule et un modèle pharmacophore.

- **La superposition flexible** : L'algorithme de superposition flexible explore l'espace conformationnel accessible des molécules ou des fragments moléculaires afin de trouver les conformations donnant la meilleure superposition. En général, seuls les angles de torsion varient. Le « fitting » flexible est idéal lorsqu'il s'agit de superposer des molécules présentant un squelette identique et ne différant que par l'arrangement spatial de certains de leurs substituants. A l'issue d'une telle superposition, l'écart énergétique entre la géométrie initiale et celle obtenue relaxée doit se situer aux alentours de quelque KJ/mole.

Nous avons souvent utilisé ces méthodes de « fitting », d'une part pour tester nos ligands sur les pharmacophores 5-HT_{1A} et d'autre part pour effectuer un alignement de nos ligands avant les études de similarité moléculaire QSAR. D'après les résultats obtenus lors des superpositions avec les pharmacophores 5-HT_{1A}, il apparaît que ces méthodes de « fitting » semblent nécessaires mais non suffisantes pour une prédiction judicieuse de l'affinité d'un nouveau composé vis-à-vis d'un récepteur donné. Nous montrerons par la suite dans le chapitre 4 que certaines molécules présentent une bonne superposition avec certains pharmacophores 5-HT_{1A} alors que leurs activités biologiques sont très faibles (de l'ordre de pIC₅₀ < 4).

II.1.B - Les modèles de pharmacophores.

Lorsqu'il est impossible d'avoir accès à toute information structurale concernant la protéine hôte (récepteur, enzyme, etc...), une des méthodes les plus utilisées en analyse structurale est celle de l'approche par pharmacophore, introduite par Humblet et al. en 1980⁴⁷. Hibert *et al.*⁴⁸ furent les premiers à proposer un pharmacophore 5-HT_{1A} défini sur les bases d'études de relations conformation-activité. Des zones accessibles et interdites du récepteur 5-HT_{1A} ont été définies à l'aide de la modélisation moléculaire, conduisant à l'élaboration d'une représentation tridimensionnelle du site de reconnaissance qui rend compte de l'activité des composés agonistes 5-HT_{1A}. Les classes de ligands sélectionnés pour définir ce pharmacophore sont des 2-aminotétramines, des 1-arylpipérazines et le Lisuride. Le pharmacophore (Figure 14) présenté par ces auteurs donne plusieurs informations structurales.

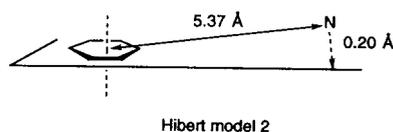


Figure 14. Pharmacophore de Hibert.

Tout d'abord, on obtient une information précise sur la distance entre le centre du noyau aromatique à l'atome d'azote possédant des électrons libres. Cette distance est de 5,37 Å. L'autre information concerne la position spatiale de l'atome d'azote. Celui-ci doit être placé à 0.2 Å par rapport au plan cycle aromatique. Le doublet d'électrons est perpendiculaire à ce plan.

Un autre modèle (Figure 15) a ensuite été proposé en 1991 par Mellin *et al.*⁴⁹ Ce pharmacophore 5-HT_{1A} basé sur une étude conformationnelle d'analogues de la 8-OH-DPAT prend toujours en compte le cycle aromatique mais cette fois les auteurs proposent un emplacement théorique de l'ion carboxylate du récepteur qui doit avoir une interaction électrostatique avec le doublet d'électrons de l'azote.

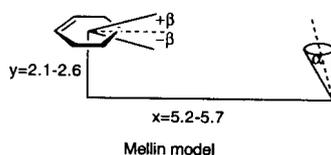


Figure 15. Pharmacophe de Mellin.

La distance entre un atome fictif passant par le centre de l'aromatique distant de 2.1 à 2.5 Å et la position fictive de l'ion carboxylate du récepteur est comprise entre 5.2 et 5.7 Å. Cette valeur ne prend donc plus en compte la position de l'atome d'azote et son doublet d'électrons mais la position fictive de l'ion carboxylate du récepteur. Ceci donne donc une information nouvelle sur la direction du doublet par rapport au cycle aromatique. De plus, on peut remarquer la présence d'angles cruciaux (α et β). Les valeurs de ces angles sont en relation directe avec les positions relatives aux énantiomères.

Toutes ces données ne permettent pas d'expliquer les différences importantes d'affinité vis-à-vis du récepteur 5-HT_{1A}, pour des composés plus volumineux comme par exemple des 4-(ω -substitué)alkyl-1-arylpiperazines.

Les modèles de Glennon *et al.*⁵⁰ et Chidester *et al.*⁵¹ (Figures 16 et 17) essayent d'apporter une réponse plus précise concernant l'environnement du ligand dans le site actif. Ces auteurs présentent des modèles comprenant les éléments de base similaires à celui de Mellin, mais leur modèle est basé non pas sur des conformations de molécules issues de minimisation mais sur des conformations extraites d'études RX.

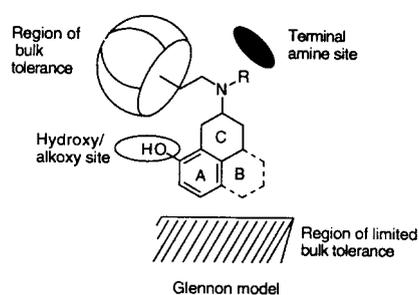


Figure 16. Pharmacophore de Glennon.

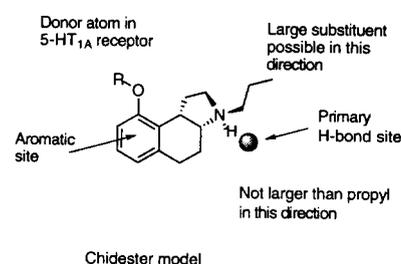


Figure 17. Pharmacophore de Chidester.

Ces deux modèles de pharmacophore donnent des informations très précises sur les zones d'interactions potentielles et sur les régions sensibles aux encombrements stériques. On voit donc que la fonction OH ou OR doit être en vis-à-vis d'un résidu du récepteur pouvant accepter des électrons de l'oxygène. On remarque aussi qu'une partie des cycles aromatiques (modèle de Glennon) est soumise à une restriction stérique. Et une information très intéressante montre que les substituants de l'azote sont soumis à des tolérances stériques. Un des substituants ne peut raisonnablement être supérieur à une fonction propyl alors que l'autre substituant peut être une longue chaîne alkyle puisque dans cette direction le ligand ne rencontre pas de gêne stérique.

Le dernier pharmacophore (Figure 18) présenté est celui issu d'une étude réalisée en 1995 par Mokrosz *et al.*⁵² sur des dérivés spiro[pipéridine-4',1-(1,2,3,4-tétrahydro- β -carboline)] et le LSD.

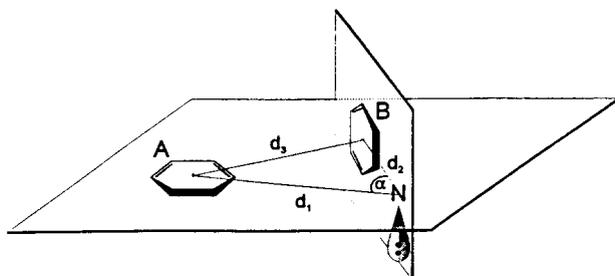


Figure 18. Pharmacophore de Mokrosz: $d_1 = 5.0-5.4 \text{ \AA}$, $d_2 = 4.3-6.6 \text{ \AA}$, $d_3 = 7.1-9.1 \text{ \AA}$, $\alpha = 98.6^\circ$, les déviations de l'atome d'azote par rapport aux cycles A et B sont respectivement $h_1 = 0.9-1.6 \text{ \AA}$ et $h_2 = 0.1-1.1 \text{ \AA}$.

Ce modèle est le seul à proposer un système d'ancrage, au niveau du récepteur, à trois points simulés par deux cycles aromatiques A et B et l'azote. Ce pharmacophore est donc un modèle adaptable aux ligands comportant des chaînes latérales de longueur plus ou moins importantes sur l'azote. Ce modèle devrait permettre de mieux comprendre des différences d'activités entre des composés présentant des chaînes alkyle de longueurs différentes. De plus, avec un modèle à trois points, des différences d'activité entre des énantiomères devraient pouvoir être explicitées.

En conclusion, nous pouvons dire que les modèles de pharmacophores sont au cours du temps de plus en plus précis tout en étant limités à certaines familles de ligands bien précises avec lesquelles ils ont été conçus. Les conditions structurales des pharmacophores sont en générales nécessaires mais non suffisantes pour la prédiction de nouveaux ligands. Ces études peuvent être complétées par des études QSAR où des études de « docking » des ligands lorsque la modélisation du récepteur est envisageable.

II.2 - Les méthodes QSAR.

II.2.A - Historique et développement du QSAR⁵³.

En 1868, Crum-Brown et Fraser⁵⁴ publiaient une équation qui est considérée depuis comme la première formulation générale en QSAR. Leurs travaux sur différents alcaloïdes ont montrés que l'alkylation sur un atome d'azote basique, donnant un ammonium quaternaire, faisait varier l'effet biologique de façon significative. Ils en ont déduit que l'activité physiologique Φ était fonction de la structure moléculaire chimique C.

$$\Phi = f(C)$$

Plus tard, Richet⁵⁵ a découvert que la toxicité de certains composés organiques était inversement proportionnelle à leur solubilité dans l'eau.

$$\Delta\Phi = f(\Delta C)$$

A la fin du 19^{ème} siècle, Meyer⁵⁶ et Overton⁵⁷, ont indépendamment observé une relation linéaire entre la lipophilie et l'activité biologique des narcotiques⁵⁸.

Fühner⁵⁹ a montré par la suite que l'activité, pour une série de narcotiques homologues, augmentait en fonction de la progression géométrique de la série de composés, ceci montrant l'importance de la contribution d'additivité de groupements fonctionnels pour l'activité biologique.

Les méthodes QSAR ont ensuite rapidement évolué dans les années 1950 avec Bruice, Kharasch et Winzler⁶⁰. En 1962 Hansen⁶¹ a donné une relation de type Hammett, montrant une corrélation entre la toxicité des acides benzoïques substitués et les constantes électroniques σ des substituants.

$$\text{Log } 1/C = 1.454 (\pm 0.42) \sigma + 1.787 (\pm 0.17)$$

L'année 1964 est considérée comme le début des méthodes QSAR modernes. Des travaux importants, de Hansch *et al.* et Free *et al.* apportent indépendamment deux nouvelles méthodes

QSAR, qui seront appelées par la suite l'analyse de Hansch⁶² et l'analyse de Free Wilson⁶³. La réelle percée de la méthode QSAR résultera de la combinaison des différents paramètres physico-chimiques de manière linéaire comme par exemple cette équation :

$$\text{Log } 1/C = a \text{ Log } P + b \sigma + \dots + \text{constante}$$

Depuis 30 ans, on dénombre environ mille équations de Hansch et quelques 500 équations de Free Wilson. Ces méthodes ont connu un succès certain tout en étant limitées par le nombre de variables utilisées pour décrire les composés.

Par la suite, la méthode « partial least squares » (PLS)⁶⁴ a permis de traiter de nombreux descripteurs moléculaires. Les méthodes de régression classiques telle que la méthode de régression linéaire doivent tenir compte du fait que le nombre de variables doit être inférieur au nombre d'individus.

Maintenant, des méthodes 3D comme l'étude CoMFA (Comparative Molecular Field Analysis)⁶⁵ ou l'analyse des indices de similarité moléculaire avec le logiciel ASP (Automated Similarity Package)⁶⁶ permettent de traiter les relations structure-activité en trois dimensions ; c'est du 3D QSAR.

Les molécules biologiquement actives sur un récepteur donné se ressemblent, on parle même de mimétiques pour les dénommer. Il est donc fort utile de pouvoir comparer de façon quantitative à quel point des molécules peuvent se ressembler ou différer, tant au point de vue stérique, que lipophile et électrostatique. En effet l'adéquation d'une molécule avec un site récepteur dépend du fait que géométriquement la molécule puisse s'insérer dans ce récepteur et qu'électrostatiquement elle vienne le compléter afin de s'y lier. Nous reviendrons de façon plus approfondie sur ces méthodes 3D QSAR par la suite dans ce chapitre.

II.2.B - Les descripteurs moléculaires.

Pour corréler les activités biologiques (pK_i , IC_{50} , K_d , etc...), les méthodes QSAR ont besoin de paramètres physico-chimiques. Ces paramètres doivent représenter les phénomènes agissant au niveau biologique, tels que, les effets stériques, les paramètres liés à la lipophilie et à l'électrostatique. Il existe une multitude de paramètres pouvant décrire une molécule, notre présentation ne sera donc pas exhaustive. Dans ce chapitre, nous ne présenterons que les descripteurs moléculaires qui ont été utilisés dans notre travail QSAR avec le logiciel TSAR⁶⁷ développé par G. GRASSY et R. LAHANA.

II.2.B.1 - Les indices de connectivité.

Hall et Kier⁶⁸ ont développé les indices (χ et χ^v) de connectivité moléculaire qui reflètent l'identité des atomes, l'environnement des liaisons et le nombre de liaisons hydrogènes. Le graphe moléculaire des structures est décomposé en quatre catégories (Path, Cluster, Path/Cluster, Ring) (Schéma 7). L'identité des atomes est spécifiée par des valeurs de connectivité moléculaire^{69,70} δ et δ^v . Le δ représente le nombre de voisins d'un atome. Cette valeur de δ est directement interprétée de manière électronique^{71,72}. δ est donné par l'équation qui suit.

$$\delta_i = \delta_i - \eta_i = \text{nombre de voisins de l'atome } i$$

δ_i = nombre d'électrons de valence de l'atome i dans les orbitales δ

η_i = nombre d'atomes d'hydrogène liés à l'atome i

Pour augmenter la quantité des informations électroniques, les auteurs ont développé le δ^v où tous les électrons de valences sont pris en compte.

$$\delta^v_i = Z^v_i - \eta_i \quad \text{pour les atomes du premier rang}$$

$$\delta^v_i = (Z^v_i - \eta_i)/(Z_i - Z^v_i - 1) \quad \text{pour les autres atomes}$$

Z^v_i = nombre total d'électrons de valence pour l'atome i

Z_i = numéro atomique de l'atomes i

η_i = nombre d'atomes d'hydrogène liés à l'atome i

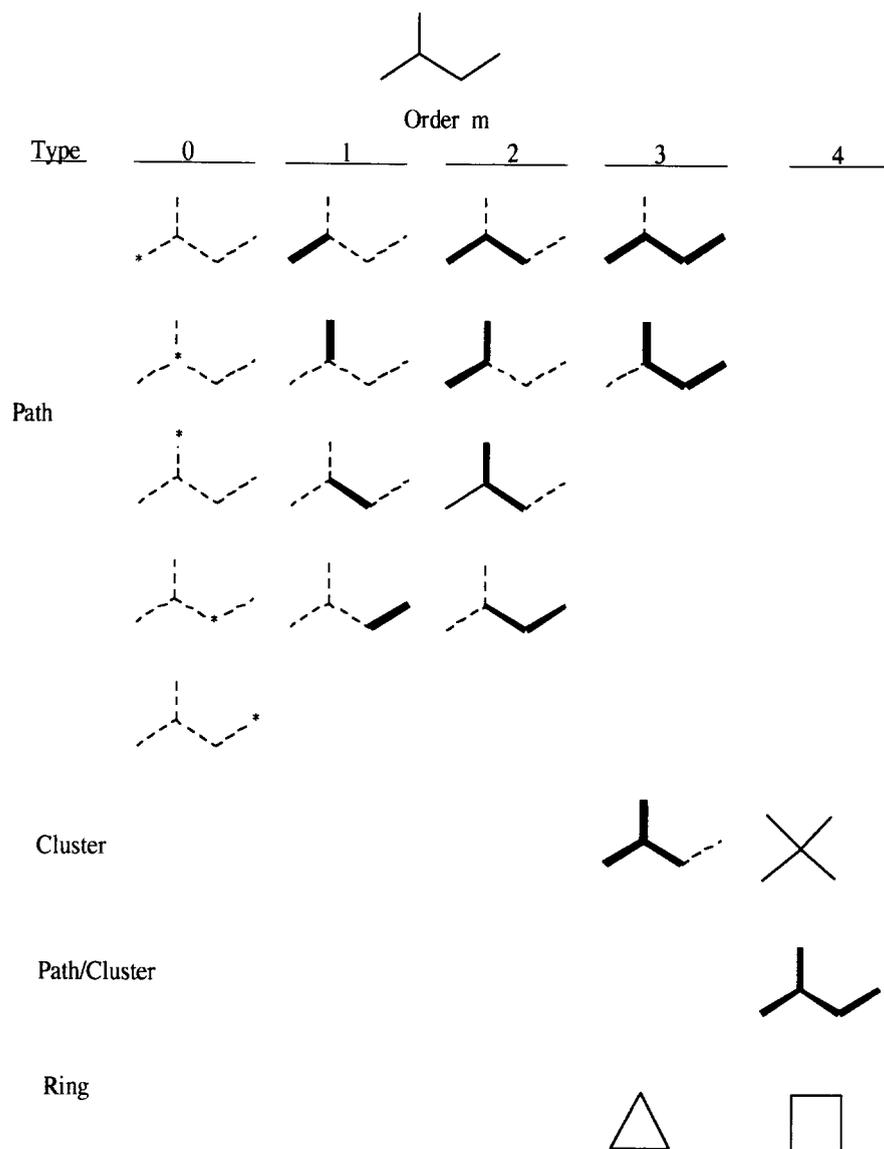


Schéma 7. Exemple de décomposition topologique de l'isopentane.

Ces deux indices δ et δ^v ont la particularité d'être très corrélés ($r = 0.99$) avec le volume de van der Waals selon l'équation⁷³ :

$$V_{\text{vdw}} = 17.03 - 1.59 (\delta + \delta^v)$$

Le graphe moléculaire est donc décrit avec les indices χ et χ_V . L'indice χ_0 , qui représente les liaisons d'ordre 0, est la somme des contributions de δ pour chaque atome.

$$\chi_0 = \sum_{i=1}^A (\delta_i)^{-1/2}$$

L'indice χ_1 , qui représente les liaisons d'ordre 1, est la somme des contributions de δ pour chaque fragment T contenant une liaison.

$$\chi_1 = \sum_{i=1}^T \sum_{j=1}^T (\delta_i \delta_j)^{-1/2}$$

En général, les indices de connectivité d'ordre m se calculent selon l'équation:

$$\chi_t^m = \sum_{s=1}^T \binom{m}{s} C_s \quad \text{avec} \quad C_s = \prod_{j=1}^N [(\delta_j)^s]^{-1/2}$$

Les méthodes de calculs des indices de connectivité sont identiques avec le δ_V . La somme de tous ces indices représente le graphe moléculaire entier de la molécule. Ces indices ont été largement utilisés pour corréler des indices de rétention chromatographique⁷⁴, des calculs de volume moléculaire⁷⁵ et des calculs de potentiel d'ionisation⁷⁶.

II.2.B.2 - Les indices de forme et de flexibilité.

Hall et Kier ont aussi introduit des indices de forme moléculaire (Kappa) en s'appuyant sur l'hypothèse que la forme d'une molécule est fonction du nombre d'atomes et des liaisons de celle-ci. Ils ont trouvé que pour une molécule ayant A atomes, on pouvait définir un ensemble de variables dites de chemin entre ces atomes (mP).

L'ensemble des atomes et des liaisons représente le squelette moléculaire ou le graphe moléculaire sans les atomes d'hydrogènes.

mP = nombre total de chemins adjacents à travers le graphe moléculaire de longueur de liaison m pour chaque molécule.

Si m = 1 pour 1K (nombre de liaisons comprises entre deux atomes adjacents)

Si m = 2 pour 2K (nombre de liaisons comprises entre trois atomes adjacents)

Si m = 3 pour 3K (nombre de liaisons comprises entre quatre atomes adjacents)

$$\text{Kappa 1} = {}^1K = (A(A-1)^2) / ({}^1P_1)^2$$

L'indice Kappa 1 met en évidence le degré de complexité de l'ensemble des liaisons et donne une indication sur l'aspect cyclique ou acyclique des molécules.

${}^1K = A$ pour les molécules acycliques

${}^1K > A$ pour les molécules cycliques

$$\text{Kappa 2} = {}^2K = (A-1)(A-2)^2 / ({}^2P_1)^2$$

L'indice Kappa 2 montre le degré de linéarité de l'ensemble des liaisons.

$$\text{Kappa 3} = {}^3K = (A-1)(A-3)^2 / ({}^3P_1)^2 \text{ si } A \text{ est impair}$$

$$\text{Kappa 3 : } {}^3K = (A-2)(A-3)^2 / ({}^3P_1)^2 \text{ si } A \text{ est paire}$$

L'indice Kappa 3 indique le degré de ramification des molécules.

Les auteurs considèrent que tous les atomes ne sont pas équivalents et créent les indices Kappa alpha ($K\alpha$). La valeur de ? est calculée pour tous les atomes en fonction des rayons covalents.

$$\alpha_x = (R_x/R_{C(sp3)})^{-1}$$

$$\text{Kappa alpha 1 : } {}^1K\alpha = (A+\alpha) (A+\alpha-1)^2 / ({}^1P_{i+\alpha})^2$$

$$\text{Kappa alpha 2 : } {}^2K\alpha = (A+\alpha-1) (A+\alpha-2)^2 / ({}^2P_{i+\alpha})^2$$

$$\text{Kappa alpha 3 : } {}^3K\alpha = (A+\alpha-1) (A+\alpha-3)^2 / ({}^3P_{i+\alpha})^2 \text{ si A est impaire}$$

$$\text{Kappa alpha 3 : } {}^3K\alpha = (A+\alpha-2) (A+\alpha-3)^2 / ({}^3P_{i+\alpha})^2 \text{ si A est paire}$$

La flexibilité (ϕ) d'une molécule est directement reliée au degré de linéarité et à la présence de cycles ou de ramifications. Cet indice est donné par l'équation suivante :

$$\phi = ({}^1K\alpha \ {}^2K\alpha) / A$$

Tous ces indices permettent de modéliser des phénomènes de ramification, d'encombrement stérique et de flexibilité des molécules. Ces paramètres ont été corrélés ($r = 0.94$) à une activité biologique *in vitro*, le IC_{50} , pour expliquer la toxicité de benzènes substitués⁷⁷.

II.2.B.3 - Les propriétés moléculaires.

Par propriétés moléculaires, on distingue les paramètres électroniques comme, les constantes d'Hammett⁷⁸, les pKa⁷⁹, les constantes de transfert de charge^{80,81}, les moments dipolaires⁸², les paramètres dérivés de la mécanique quantique^{83,84} (énergie des orbitales HOMO et LUMO, charges atomiques, densité électronique, etc...). Tous ces paramètres décrivent les distributions de densité électronique de certains groupes ou substituants.

Il y a aussi les paramètres de lipophilie⁸⁵ comme le Log P. La lipophile (synonyme d'hydrophobicité) est quantifiée par la mesure physique du coefficient de partage entre un solvant organique, l'octanol le plus souvent, et l'eau. Le Log P peut être calculé à partir de tables fragmentales⁸⁶ ou de contributions lipophiliques atomiques⁸⁷.

D'autres descripteurs sont les paramètres d'ordre stérique⁸⁸ comme, la surface moléculaire, le volume molaire, le rayon de van der Waals⁸⁹, le volume ellipsoïdal et des paramètres de polarisabilité comme la réfractivité molaire. La réfractivité moléculaire dépend de la lipophilie et de la polarisabilité de la molécule. Elle donne des informations relatives au volume et à la cohésion de la molécule.

Plusieurs de ces paramètres moléculaires ont été calculés par TSAR comme la réfractivité molaire, la surface moléculaire, le volume moléculaire, le Log P calculé⁹⁰, le moment dipolaire, la masse moléculaire, le volume ellipsoïdal et les indices du moment d'inertie.

Les valeurs de réfractivité molaire calculées par TSAR sont des valeurs prédites, en utilisant les valeurs de réfractivité molaire atomiques déterminées par Viswanadhan *et al*⁹¹. Les valeurs atomiques du Log P sont issues aussi des travaux de Viswanadhan *et al*.

Le calcul des moments dipolaires utilise les charges partielles atomiques évaluées par le programme CHARGE-2. Le logiciel TSAR permet d'obtenir le moment dipolaire total et le moment dipolaire de liaison. Le moment dipolaire total est calculé en utilisant le centre de gravité des charges.

$$\mu = e \sum_i r_i q_i \quad \text{avec } r_i = \text{distance entre l'atome } i \text{ et l'origine}$$

$q_i = \text{la charge de l'atome } i$

Le moment dipolaire de liaison peut être calculé seulement pour un substituant. Le calcul est effectué en plaçant la liaison suivant un axe x passant par l'origine. Le moment dipolaire selon l'axe x donne le moment dipolaire de liaison (Schéma 8).

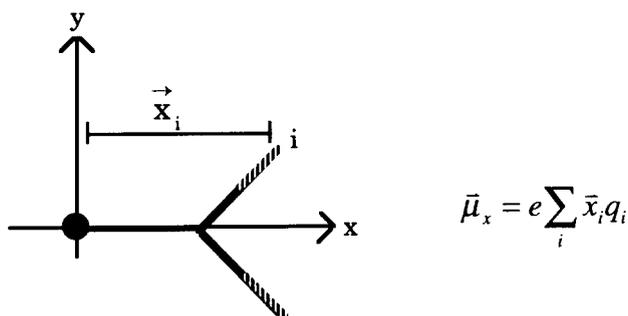


Schéma 8. Le moment dipolaire de liaison.

II.2.B.4 - Les vecteurs d'autocorrélation.

Les autocorrélogrammes⁹², dérivés d'une fonction mathématique (fonction d'autocorrélation), génèrent des indices de similarité moléculaire basés, soit sur la connectivité pour les autocorrélogrammes à deux dimensions (2D), soit sur les distances interatomiques pour les autocorrélogrammes à trois dimensions (3D).

$$2D \text{ Bin } (N) = \sum_{inj} (X_i)(X_j)$$

$inj = \text{Toutes les paires } ij \text{ d'atomes séparés par } N \text{ atomes.}$

$$3D \text{ Bin } (R) = \sum_{inj} (X_i)(X_j)$$

$inj = \text{Toutes les paires } ij \text{ d'atomes compris entre } (R-1) \text{ et } R \text{ Å.}$

Cette technique effectue donc une transformation simple d'une matrice de connectivité en un vecteur structural. Elle permet d'y inclure plusieurs caractéristiques atomiques. La molécule est donc représentée par un ensemble de propriétés distribuées sur le graphe moléculaire de celle-ci. Ces composantes peuvent être utilisées directement, dans ce cas $X=1$, ou bien être affectées d'un coefficient de pondération. Les valeurs de X_i et X_j peuvent donc être pondérées par certaines propriétés physiques comme :

Charge	$X =$ charge atomique partielle
van der Waals	$X =$ rayon de van der Waals
Lipophilie	$X =$ contributions atomiques au Log P calculé
Electronégativité	$X =$ valeur d'électronégativité de Pauling
Valence	$X =$ contribution de valence de Kier et Hall
Numéro atomique	$X =$ numéro atomique

Tous ces vecteurs permettent des études de reconnaissance de forme en quantifiant les structures topologiques et tridimensionnelles.

Dans notre cas, toutes les molécules étudiées n'ont pas le même nombre d'atomes, donc pas le même nombre de composantes 2D ou 3D. Pour comparer ces molécules, nous avons éliminé les composantes qui n'étaient pas communes à toutes les molécules, c'est à dire, les composantes d'indice élevé.

Des études de relation structure-activité (RSA), utilisant les autocorrélogrammes ont donné des résultats intéressants sur une série de quinolones⁹³, sur des composés apparentés à la clonidine⁹⁴ et sur des hydrocarbures aromatiques polybenzéniques⁹⁵ aux pouvoirs carcinogènes.

II.2.B.5 - Les indices de topologie.

En général, les différentes formes de graphes moléculaires sont très exploitées en chimie. Ceci est dû au fait que les représentations graphiques des molécules peuvent servir comme objet mathématique dans la théorie des graphes. Une de ces méthodes, basée sur la théorie des graphes, et couramment utilisée dans les approches QSAR, est la méthode utilisant les indices de topologie^{96,97}. Ces indices de topologie utilisés sont ceux de Wiener⁹⁸, Randic⁹⁹ et Balaban¹⁰⁰. L'indice de Wiener est obtenu en faisant la somme des distances topologiques (définies comme étant le plus petit nombre de liaisons séparant 2 noeuds) entre toutes les paires d'atome des molécules. Cet indice caractérise le degré de ramification ainsi que le volume moléculaire. Il est généralement plus élevé pour les molécules allongées que pour les composés compacts. L'indice de Randic est un indice de connectivité. Pour calculer cet indice, chaque atome du graphe prend pour valeur l'inverse de la racine carré de sa connectivité. Chaque liaison reçoit une valeur égale au produit des valeurs des atomes qu'elle relie. L'indice est obtenu en faisant la somme de toutes les liaisons. Il est un bon descripteur de la forme moléculaire. Cet indice a été à la base de la mise au point de descripteurs plus sophistiqués comme les indices de Kier. La détermination de l'indice de Balaban suppose, dans une première étape, de compter tous les atomes dont la connectivité est égale à un ; la valeur obtenue est élevée au carré. Dans une deuxième étape, ces atomes sont éliminés du graphe et on réitère ce procédé jusqu'à la disparition de tous les atomes. L'indice de Balaban est obtenu en additionnant les valeurs issues de chacune des étapes. Cet indice décrit très bien le degré de ramification des molécules non cycliques. Tous ces indices ont surtout une bonne corrélation avec des propriétés physiques dans le cas des alcanes saturés¹⁰¹.

II.2.C - Les outils statistiques.

Les méthodes statistiques sont aujourd'hui utilisées dans presque tous les secteurs de l'activité humaine. Selon la définition de l'Encyclopedia Universalis «Le mot statistique désigne à la fois un ensemble de données d'observations et l'activité qui consiste dans leur recueil, leur traitement et leur interprétation». Faire de la statistique suppose que l'on étudie un ensemble d'objets équivalents sur lesquels on observe des caractéristiques appelées «variables». Dans notre cas, les objets (ou individus) sont les molécules proprement dites et les variables sont les descripteurs précédemment cités dans ce chapitre. Après le recueil des descripteurs, la démarche statistique consiste à traiter et interpréter les informations recueillies sur ces molécules. Cette démarche comporte deux grands aspects : l'aspect descriptif et l'aspect décisionnel ou prédictif.

La statistique descriptive, synthétise, résume et structure l'information contenue dans les données (descripteurs). Elle utilise pour cela des représentations des données sous forme de tableaux, de graphiques ou d'indicateurs numériques. L'analyse de données est une des branches les plus vivantes de la discipline statistique. Elle comprend les méthodes de classification, l'analyse des composantes, l'analyse des clusters et l'analyse des correspondances. L'objectif principal de ces méthodes vise à réduire la taille de l'ensemble des variables (descripteurs) ou des individus pour les méthodes de classifications. Pour notre part, nous avons principalement utilisé l'analyse en composantes principales (ACP).

La prévision statistique consiste généralement à rechercher une relation approximative entre une variable et plusieurs autres, la forme de cette relation étant le plus souvent linéaire. Dans notre cas, on cherche à relier une activité biologique (pIC_{50} ou pKi) à des variables communes (descripteurs) pour un ensemble de composés. Lorsque la variable à expliquer ou à prévoir est numérique comme par exemple une valeur de pIC_{50} , nous avons utilisé la régression linéaire multiple (RM) ou la méthode Partial Least Squares (PLS). Lorsque l'on a cherché à prévoir une variable qualitative comme par exemple ACTIF ou INACTIF, nous avons utilisé l'analyse discriminante (AD).

Dans ce chapitre, nous présenterons les méthodes statistiques que nous avons utilisées. Ces méthodes sont l'analyse en composantes principales (ACP), l'analyse discriminante (AD), la régression linéaire multiple (RM) et le Partial Least Squares (PLS).

Toutes ces méthodes d'analyses sont accessibles avec le logiciel d'Oxford Molecular, TSAR (Tools for Structure Activity Relationships).

II.2.C.1 - L'analyse en composantes principales.

Cette méthode^{102,103} est utilisée pour réduire un nombre important de descripteurs (variables initiales) en un nombre plus faible de nouvelles variables sans toutefois perdre trop de l'information initiale. Les n individus (pour nous les n composés étudiés) sont initialement représentés dans un espace à m dimensions (pour nous les m descripteurs calculés). Le principe de la méthode est d'obtenir une représentation approchée du nuage des n individus dans un sous-espace de plus faible dimensionnalité. Ces nouvelles dimensions sont une combinaison linéaire des dimensions initiales. Elles sont créées de telle façon que le maximum d'information (généralement la variance) soit présent sur la première dimension (premier axe factoriel). Les axes factoriels suivants sont choisis de la même façon en leur imposant comme contrainte d'être orthogonaux aux autres axes. L'analyse en composantes principales construit donc de nouvelles variables artificielles.

Des représentations graphiques 2D et 3D permettent de visualiser les relations entre les descripteurs, ainsi que l'existence éventuelle de groupes d'individus (cluster) et de groupes de variables. L'analyse en composantes principales permet aussi par l'étude de la matrice des corrélations des descripteurs, de réduire le nombre de descripteurs, en éliminant ceux qui présentent une corrélation importante. Ceci est très utile pour réduire le nombre de variables avant d'effectuer une régression linéaire. L'analyse nous donne aussi une information quantitative sur la composition des axes factoriels en fonction des descripteurs.

Nous avons donc utilisé cette méthode statistique dans toutes nos études. Soit pour étudier graphiquement la dispersion des variables et des individus, soit pour réduire le nombre de variables dans nos tableaux de données en vue d'effectuer des régressions linéaires avec le moins de variables possibles.

II.2.C.2 - L'analyse discriminante^{104,105,106}.

Par rapport à l'analyse en composantes principales, l'analyse discriminante prend en compte une variable qualitative (dans notre cas l'activité biologique pIC_{50}) en plus des autres variables (descripteurs). L'objectif de cette méthode consiste à combiner l'ensemble de ces variables afin de discriminer les individus (pour nous les composés) suivant deux classes (active et inactive dans notre cas). Une discrimination est optimum si les barycentres des classes sont les plus éloignés possible et si la totalité des individus est bien classée. Une des particularités de l'analyse discriminante est la possibilité de prédiction qualitative à partir du modèle statistique. On distingue donc classiquement deux aspects en analyse discriminante ; l'aspect descriptif où l'on cherche quelles sont les combinaisons linéaires de variables qui permettent de séparer le mieux possible les différentes classes, et l'aspect prédictif qui cherche à classer un nouvel individu dans une des classes. Dans notre cas, nous avons opté en général pour l'attribution de deux classes ; une classe dite « INACTIF » pour les composés ayant un pK_i ou un $pIC_{50} < 7.5$ et une classe dite « ACTIF » pour les molécules ayant ces mêmes activités biologiques supérieures ou égales à 7.5.

La validité du modèle est estimée à partir du nombre d'individus bien classés, du nombre de variables discriminantes choisies, et en fonction de la valeur du « confidente estimate » qui est considéré comme l'équivalent du « cross validation » pour les méthodes de régression. TSAR utilise donc une méthode de resubstitution qui consiste à réaffecter tous les individus selon les fonctions discriminantes trouvées afin de prédire leur activité. Il semble toutefois que cette méthode sous-estime le taux d'erreur puisqu'on utilise pour la prédiction les mêmes individus que ceux qui ont servi à trouver les fonctions discriminantes. La règle, optimale pour l'échantillon, doit donner *a priori* de bons résultats si on la réapplique sur l'échantillon en question. Pour échapper à ce problème, il suffit de partager l'échantillon en deux sous-échantillons ; l'un servira à l'élaboration du modèle et l'autre à la validité du modèle. Ceci ne peut être appliqué que si l'on a assez d'individus pour scinder son échantillon en « training set » et « test set ».

Nous avons donc utilisé cette méthode statistique dans toutes nos études pour discriminer nos échantillons suivant deux classes (ACTIF, INACTIF).

II.2.C.3 - L'analyse de régression linéaire multiple.

L'analyse de régression linéaire multiple^{107,108,109} (RM) corrèle des variables indépendantes (descripteurs) avec des variables dépendantes (activité biologique). Pour les études de régression multiple, le nombre de variables doit être inférieur ou égal au nombre d'individus. Si le nombre de variables est élevé, il est plus judicieux d'utiliser la méthode du Partial Least Squares (PLS). Cette méthode sera présentée par la suite. La RM donne donc une équation du type :

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n + C$$

où a_1, a_2, \dots, a_n et C sont des constantes.

La régression multiple utilisée par le logiciel TSAR est une méthode « stepwise ». Les méthodes pas à pas procèdent par élimination successive (méthode descendante) ou ajout successif de variables (méthode ascendante). La méthode dite « stepwise » est un perfectionnement de l'algorithme des méthodes descendantes et ascendantes. En effet par des tests du type F (Fisher), elle ajoute ou élimine des variables dans l'équation. L'algorithme s'arrête quand on ne peut plus, ajouter ou retrancher des variables sans dépasser les seuils statistiques choisis. Ces méthodes ne donnent pas forcément les meilleures régressions, mais elles sont très pratiques d'emploi ; la méthode « stepwise » semblant être la meilleure. Ces méthodes ne sont cependant pas à l'abri de l'élimination intempestive de variables réellement significatives, ce qui risque de biaiser les résultats.

Pour valider les modèles, il existe plusieurs critères de choix. Ces critères dépendent des usages que l'on fait de la régression. Le critère r^2 qui est le coefficient de corrélation, est bien adapté pour la reconstitution des variables, il n'est cependant pas à l'abri de critiques.

$$r^2 = \text{Variance expliquée par la régression} / \text{Variance totale}$$

En effet, il varie de façon monotone avec le nombre de variables ; il ne peut qu'augmenter si on rajoute un prédicteur même peu corrélé. Le coefficient de corrélation ne mesure que le caractère linéaire d'un nuage de points et son usage doit être réservé à des nuages où les points sont répartis aléatoirement de part et d'autre d'une tendance linéaire.

Si l'objectif est de minimiser l'erreur de prévision, le r^2 n'est pas adapté et on préférera des critères tels que le s (écart type pour les prédictions) ou le « cross validation » $r(\text{CV})^2$. Pour le calcul du « cross validation », le pouvoir prédictif du modèle est déterminé de la façon suivante. Chaque individu (composé) est enlevé du crible initial et son activité est prédite à partir du modèle statistique.

$$S^2 = (1 - r^2) (\text{variance totale}) / \text{DL}$$

$$\text{DL} = n - k - 1 \text{ (degré de liberté)}$$

n = nombre d'individus

k = nombre de variables

Un autre critère qui mesure la signification statistique pour le modèle, est le F (test de Fischer).

$$F = r^2 (\text{DL}) / k (1 - r^2)$$

En général, l'équation d'une régression linéaire sera acceptée si :

- le nombre de variables introduites dans l'équation est faible par rapport au nombre d'individus,
- le r et le r^2 sont élevés (seuil variable),
- le « standard deviation » s est faible,
- le F exprime une forte probabilité (plus que 95%),
- le $r(\text{CV})^2$ exprime un fort pouvoir prédictif du modèle et s'il n'est pas très différent du r^2 .

Lorsque qu'une équation valide un modèle, alors de nouveaux composés peuvent être injectés dans l'équation afin de prédire leur activité biologique.

II.2.C.4 - L'analyse Partial Least Square^{110,111,112}.

L'analyse Partial Least Squares (PLS) est dans son objectif identique à la régression linéaire mais toutes les variables sont prises en compte dans l'équation grâce à une procédure proche de l'analyse factorielle. Cette méthode d'analyse est aujourd'hui la plus utilisée, puisqu'elle permet de prendre en compte un grand nombre de descripteurs. Elle est particulièrement adaptée pour les descripteurs 3D issus d'analyses CoMFA ou d'analyses de similarité moléculaire. L'analyse PLS est une méthode qui utilise des «vecteurs» représentant l'ensemble des variables. Plusieurs critères permettent de valider un modèle. Le premier de ces critères est la «statistical significance». Le second critère est la valeur du PRESS qui doit être la plus faible possible pour le vecteur considéré.

$$\text{PRESS} = \sum \left(y_{\text{théorique}} - y_{\text{prédite}} \right)^2$$

Le calcul de ce critère est directement en relation avec la méthode qui détermine un autre critère qui est le «cross validation» $r(\text{CV})^2$. Le dernier critère est le r^2 qui donne l'aspect de la linéarité du modèle.

Nous avons utilisé cette méthode statistique dans toutes nos études pour corréler les valeurs biologiques avec les différents descripteurs.

II.2.D - Les outils 3D.

Les outils tridimensionnels (3D), comme les analyses de similarité moléculaire ou les études CoMFA (Comparative Molecular Field Analysis) sont aujourd'hui des méthodes très employées pour les études QSAR. Elles permettent de comparer les molécules à travers leur aspect tridimensionnel. Les comparaisons se font sur des critères stériques, électrostatiques et/ou lipophiliques, phénomènes responsables des interactions dans les récepteurs.

Un certain nombre d'autres approches 3D QSAR existent, comme l'approche « molecular shape » développé indépendamment par Simon et al.^{113,114} et Hopfinger^{115,116}, ou une comme la méthode CoMSIA¹¹⁷.

II.2.D.1 - L'analyse des similarités moléculaires.

Richards et al.¹¹⁸ ont récemment développé une nouvelle méthode 3D QSAR, basée sur l'analyse des similarités moléculaires. L'objectif de cette étude est de comparer toutes les molécules du crible une à une afin d'obtenir une matrice à NxN dimensions. Cette méthode préconise un alignement des molécules avant d'effectuer les calculs de similarités moléculaires. Les auteurs proposent pour comparer les molécules, d'utiliser les indices de Carbo¹¹⁹ (R_{AB}) et de Hodgkin¹²⁰ (H_{AB}).

$$R_{AB} = \frac{\sum_{i=1}^N (P_A P_B)}{\left(\sum_{i=1}^N P_A^2 \right)^{1/2} \left(\sum_{i=1}^N P_B^2 \right)^{1/2}}$$

N = Nombre total de points du Grid

$$H_{AB} = \frac{2 \sum_{i=1}^N (P_A P_B)}{\sum_{i=1}^N (P_A^2) + \sum_{i=1}^N (P_B^2)}$$

P_A et P_B sont des propriétés des molécules A et B. Les propriétés peuvent être des potentiels ou des champs électrostatiques et/ou lipophiliques.

$$P_R = \sum_{j=1}^n \frac{q_j}{|r - r_j|}$$

Le potentiel électrostatique P_R d'un atome j au point r du Grid tient compte de la charge de l'atome j (q_j) et de la position de cet atome j (r_j).

$$F_R = \sum_{j=1}^n \frac{q_j (r - r_j)}{|r - r_j|^3}$$

Le champ électrostatique F_R d'un atome j au point r du Grid prend en compte la charge de l'atome j (q_j) et la position de l'atome j (r_j). Ces valeurs de P_R et F_R sont ensuite à introduire dans les équations de Carbo et Hodgkin respectivement à la place de P_A pour une molécule A et P_B pour une molécule B.

Le potentiel P_R donne une idée de la localisation et de la force des interactions ioniques. Le champ F_R donne les mêmes informations mais pour des interactions dipolaires.

Pour l'analyse des formes, dite de « shape », des molécules, l'équation proposée par Meyer *et al.*¹²¹, est une version modifiée des équations de Carbo et de Hodgkin,

$$S_c = \frac{B}{(T_A T_B)^{1/2}}$$

$$S_h = \frac{2B}{T_A + T_B}$$

où B est le nombre de points du Grid commun aux deux molécules, T_A est le nombre de points du Grid pour la molécule A plus l'ensemble des points commun aux deux molécules, et T_B est le nombre de points du Grid pour la molécule B plus l'ensemble des points communs aux deux molécules.

Une modification du programme ASP d'Oxford Molecular, permet d'utiliser des fonctions Gaussiennes, ceci seulement pour les calculs des potentiels et de forme. Dans le cas de P_R , le calcul de $1/r$ est remplacé par une fonction Gaussienne. L'équation P_R devient :

$$P_R = \sum_{j=1}^n q_j (\chi_1 e^{-\alpha_1 r_j^2} + \dots + \chi_n e^{-\alpha_n r_j^2})$$

Cette méthode a le mérite d'être beaucoup plus rapide en temps de calcul¹²². De plus, il n'y a pas de singularité quand la distance par rapport à la charge s'approche de zéro.

Pour la comparaison des molécules, plusieurs méthodes sont proposées par le logiciel ASP. Les composés peuvent être comparés selon leurs positions relatives, issus d'une étude de superposition, sans aucune optimisation géométrique. On peut aussi demander une optimisation de la superposition, par des rotations et des translations moléculaires. La dernière méthode consiste, suivant l'algorithme de Nelder *et al.*¹²³, à effectuer une optimisation en modifiant les angles dièdres de certaines parties des molécules tout en gérant le coût énergétique de cette rotation des liaisons. Cette dernière approche est très coûteuse en temps de calcul.

Tous ces indices représentent la similarité moléculaire et ont des valeurs comprises entre 0 et 1, la valeur 1 exprimant une totale similarité entre deux molécules. Pour les indices électrostatiques, ces valeurs vont de -1 à 1.

Ces matrices peuvent ensuite être réduites par une analyse en composantes principales, et étudiées par des méthodes statistiques (AD, RM, PLS).

II.2.D.2 - L'analyse CoMFA.

L'analyse CoMFA (Comparative Molecular Field Analysis) est considérée comme une approche de corrélation structure-activité. Les caractéristiques de cette méthode sont :

- la représentation des molécules (ligands) par des champs stériques et électrostatiques, les champs ayant l'aspect de treillis tridimensionnel.
- les molécules sont alignées selon une méthode « field fit » et le RMS entre ces molécules est minimisé.
- les données sont analysées par la méthode PLS.
- des représentations graphiques des parties communes des molécules en terme stérique et électrostatique sont données.

La méthode CoMFA génère une relation linéaire entre les différentes valeurs des champs électrostatiques et stériques environnant les molécules et leurs activités biologiques. Dans la méthodologie de la méthode CoMFA, l'alignement des molécules est très important et il influence beaucoup le résultat de l'analyse. L'analyse CoMFA est beaucoup moins indulgente que la méthode ASP vis-à-vis de l'alignement des molécules au départ¹²⁴.

Lors de la comparaison des champs, une valeur de 1 est donnée pour des champs similaires, de -1 pour des champs complémentaires et de 0 pour des champs dissimilaires.

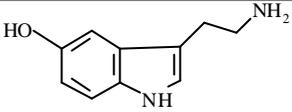
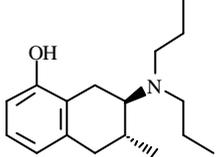
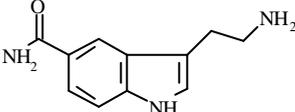
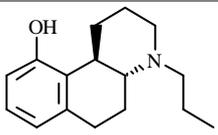
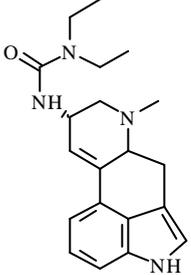
Un champ tenant compte des interactions hydrophobes¹²⁵ (hydrophobic interaction HINT) est aussi disponible pour l'analyse CoMFA.

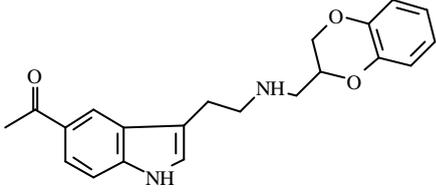
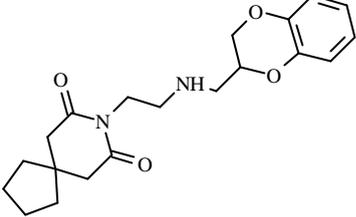
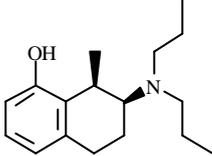
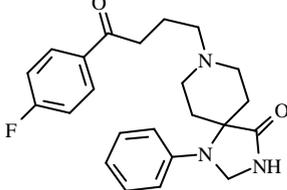
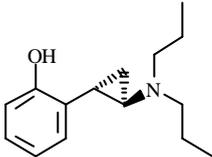
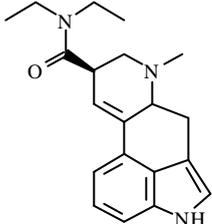
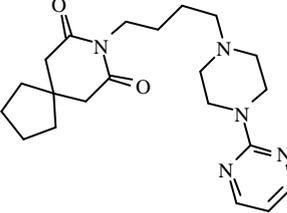
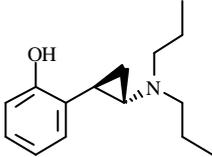
De nombreuses applications de cette méthode CoMFA publiées dans la littérature^{126,127,128,129,130} montrent l'importance des champs moléculaires pour expliquer les activités biologiques.

II.2.E - Présentation d'une étude QSAR sur des ligands 5-HT_{1A}.

Dans ce chapitre, nous allons présenter une étude QSAR effectuées sur des ligands 5-HT_{1A}, utilisant l'analyse CoMFA et la méthode d'analyse statistique du Partial Least Square.

Cette étude effectuée par Atul Agarwal et Ethan Will Taylor¹³¹, concerne une série de composés agonistes, agonistes partiels et antagonistes du récepteur 5-HT_{1A}. L'objectif de ce travail était de corrélérer l'activité intrinsèque des 20 composés, (voir tableau suivant) à des variables d'ordre stériques et électrostatiques, par une étude CoMFA. Pour ce faire les auteurs ont attribué un coefficient d'activité intrinsèque IA compris entre 0 et 1 (Tableau 4). La valeur IA = 1 est donnée à la 5-HT (sérotonine). Les valeurs IA = 0 sont attribuées aux antagonistes et $0 < IA < 1$ pour les agonistes partiels.

Composés	Structure	IA	IA
		Expérimentale	Prédiction
5-HT		1	1
(2R,3R)CM-12		0	0
5-CT		0.98	0.99
(4aR,10bR)JV-26		0.28	0.29
Lisuride		0.89	0.86

BDT		0.55	0.51
MDL73005EF		0.72	0.68
(1S,2R)ALK-3		0.22	0.18
Spipérone		0	-0.04
(1R,2S)LEA-146		0.9	0.95
d-LSD		0.94	0.99
Buspirone		0.54	0.60
(1S,2R)LEA-146		0	0.07

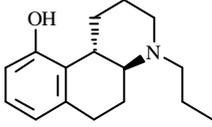
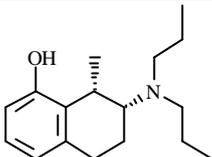
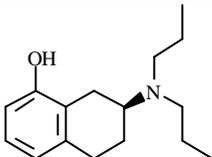
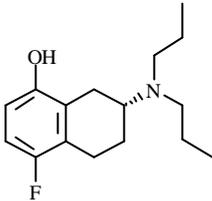
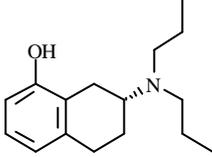
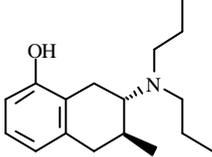
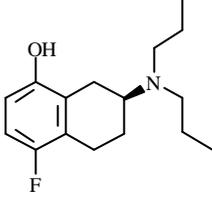
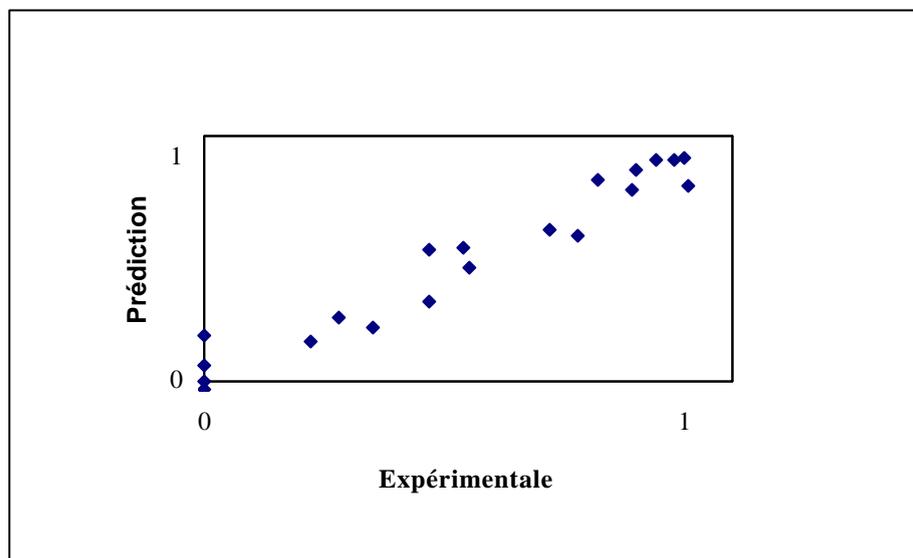
(4a <i>S</i> ,10 <i>bS</i>)JV-26		0.82	0.90
(1 <i>R</i> ,2 <i>S</i>)ALK-3		0.35	0.24
(<i>S</i>)8-OHDPAT		0.47	0.36
(<i>R</i>)UH-301		0.47	0.59
(<i>R</i>)8-OH-DPAT		1.01	0.88
(2 <i>S</i> ,3 <i>S</i>)CM-12		0.78	0.65
(<i>S</i>)UH-301		0	0.21

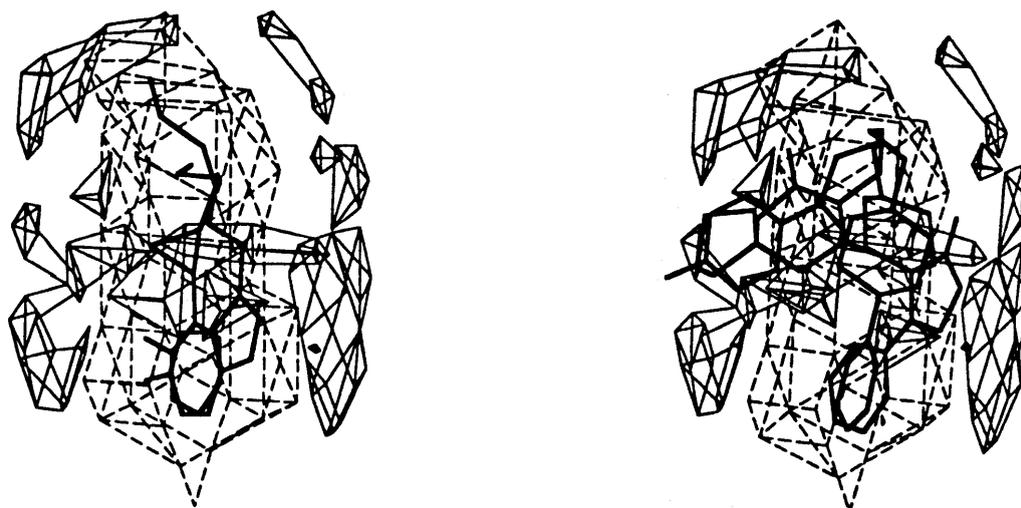
Tableau 4.

L'étude de ces composés par une analyse CoMFA suivie d'une analyse statistique PLS a donné les résultats suivants. Le coefficient r^2 est de 0.943 en prenant en compte la totalité des vecteurs. Le « standard deviation » est $s = 0.103$ (Graphe 1).



Graphe 1. Valeurs prédites en fonctions des valeurs expérimentales (IA).

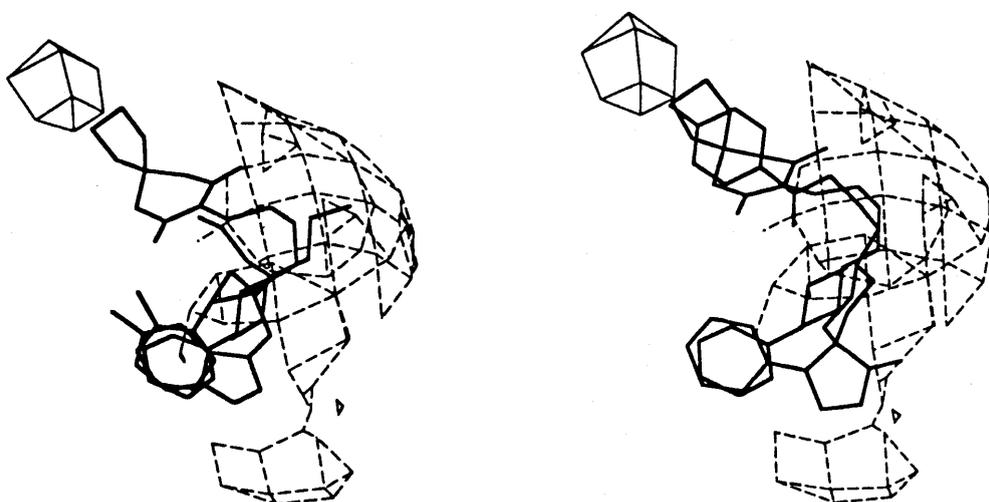
Les graphes CoMFA des champs stériques et électrostatiques montrent les différentes zones relatives aux sites agonistes et antagonistes et sont présentés ci-dessous.



Graphes du champ stérique.

Pour le graphe de gauche, les zones en pointillés montrent le volume occupé par les composés agonistes 5-HT et (2R,2S)LEA-146. Les zones en trait continu montrent les espaces qui ne doivent pas être occupé par les composés pour ne pas diminuer l'effet agoniste. A travers ce résultat, les molécules agonistes sont donc planes et peu volumineuses.

Pour le graphe de droite, les molécules antagonistes (Spipérone) et agonistes partiels (Buspirone) présentent des parties fonctionnelles qui sortent du volume agoniste (en pointillés). De plus, ces parties fonctionnelles des composés sont en contact avec une zone (en trait continu) susceptible d'affecter l'effet agoniste. Les molécules antagonistes et agonistes partiels sont donc moins planes et plus volumineuses que les molécules agonistes.



Graphes du champ électrostatique.

Au niveau des champs électrostatiques, on remarque une zone à potentiel positif importante (pointillés) pour tous les composés. Cette zone est centrée sur l'azote extracyclique. Ceci suggère que cette partie, commune aux molécules, présentant un potentiel électrostatique positif a une importance pour le site agoniste. Sur les deux graphes, on remarque aussi une zone (trait continu) accessible seulement pour les antagonistes.

En conclusion, une telle étude a montré qu'il était possible, avec ces 20 composés, de corréler des champs stériques et électrostatiques à une activité intrinsèque du type agoniste, agoniste partiel et antagoniste. De plus, l'analyse CoMFA a mis en évidence par l'intermédiaire des graphes des champs stériques et électrostatiques, des zones d'exclusions pour le site agoniste et une zone supplémentaire pour le site antagoniste ; cette dernière zone étant liée à la présence d'une chaîne alkyle mobile que l'on retrouve chez la plupart des composés antagonistes. Les composés antagonistes ou agonistes partiels ont le même site actif que les agonistes, mais ont soit un potentiel électrostatique plus faible aux alentours de l'azote extracyclique, soit un second site de contact par l'intermédiaire de la chaîne alkyle latérale.

II.3 - Dynamique des complexes ligand/récepteur.

Dans une première partie, nous présenterons la méthode de construction des complexes ligand/récepteur. Ensuite nous discuterons de la nature et de l'intensité des interactions mises en jeu dans les systèmes drogue/récepteur.

II.3.A - La construction des complexes ligand/récepteur.

Les banques de données tridimensionnelles telles que la PDB, contiennent des coordonnées de structures de protéines déterminées par cristallographie. Ces modèles sont donc accessibles directement par leurs coordonnées cartésiennes et peuvent être visualisés. Lorsque la protéine n'a pas été cristallisée, étudiée en RX et ses coordonnées cartésiennes stockées dans une base de donnée, l'établissement de la structure secondaire uniquement sur la base de la séquence primaire est difficile et d'un succès très relatif. C'est le cas du récepteur transmembranaire 5-HT_{1A}. De surcroît, il existe peu de sondes chimiques ou biochimiques permettant de cartographier l'accessibilité relative des résidus constituant la protéine native. La modélisation tridimensionnelle du récepteur 5-HT_{1A}, s'est donc orientée vers un décalquage de la structure 3D de la Bactériorhodopsine^{132,133} (Figure 19) connue, étudiée^{134,135} et dont les coordonnées sont accessibles par la PDB, par remplacement de l'ancienne séquence primaire de la Bactériorhodopsine par la séquence primaire¹³⁶ du récepteur sérotoninergique 5-HT_{1A}.

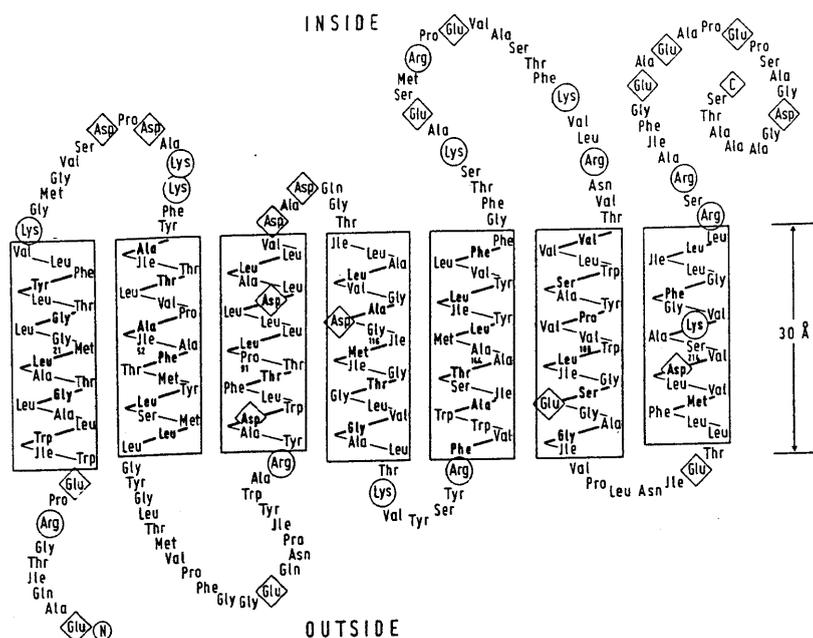


Figure 19. Modèle de la Bactériorhodopsine. Les cercles et les losanges représentent les résidus chargés¹³⁷.

Cette approche est basée sur l'homologie de séquences, concept ancré dans l'évolution biologique Darwinienne. Des séquences primaires sont dites homologues lorsqu'elles divergent à partir d'un ancêtre protéique commun. Elles sont, par contre, dites analogues lorsque au cours de l'évolution, elles ont convergé vers une structure globalement similaire avec une fonction biologique identique. La Bactériorhodopsine et le récepteur 5-HT_{1A}, présentent une certaine homologie de séquence. Le degré de succès d'une modélisation par homologie dépendra en premier lieu de l'étendue de la similitude, mais aussi du nombre de séquences homologues comparées. En effet, les effets biologiques sont le produit de l'évolution et donc de l'héritage historique. Plus le nombre de séquences comparées sera grand, plus il sera possible de distinguer les résidus et les conservations structurales ou fonctionnelles importantes.

Dans le domaine de la modélisation moléculaire des récepteurs transmembranaires, les travaux de Trumpp Kallmeyer *et al.*¹³⁸ font référence. Ces travaux scientifiques présentent une étude complète d'homologie de 42 séquences et une analyse d'hydrophobicité qui a permis d'élucider la structure secondaire de plusieurs récepteurs et entre autres celui du récepteur 5-HT_{1A}.

Le principe de construction du récepteur sérotoninergique 5-HT_{1A} est le suivant. A partir de plusieurs séquences primaires dont celle de la Bactériorhodopsine, présentant une homologie

certaine, une étude d'alignement est effectuée suivant la méthode de Needleman et Wunsch¹³⁹, en utilisant la table de Dayhoff¹⁴⁰ des similarités pour les acides aminés. Ensuite, il faut identifier les régions hydrophobes correspondant aux structures secondaires de forme hélice. Cette étude d'hydrophobicité utilise les paramètres de Kyte et Doolittle¹⁴¹ ou de Engelman et al.¹⁴² (Schéma 9).

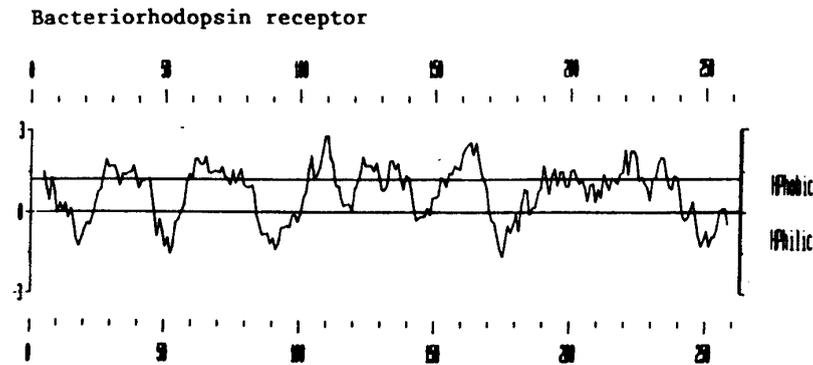


Schéma 9. Profil d'hydrophobicité de la Bactériorhodopsine. Le profil est obtenu avec les paramètres de Kyte et Doolittle.

Pour l'étude d'hydrophobicité de la Bactériorhodopsine par Kyte et Doolittle, on remarque très bien les sept domaines représentant les sept hélices α . La séparation des domaines six et sept est arbitraire et considérée à partir du résidu 200.

Ces profils d'hydrophobicité permettent donc de repérer les régions hydrophobes qui optent pour une configuration en hélice α .

Ces domaines en hélices α de la séquence 5-HT_{1A} sont ensuite superposés aux hélices α de la Bactériorhodopsine. Le récepteur est relaxé, en minimisant l'énergie potentielle de l'ensemble des atomes.

Le ligand peut ensuite être intégré de manière graphique à la protéine en vue d'études de dynamique moléculaire du complexe.

II.3.B - Nature des interactions drogue/récepteur.

Le complexe drogue/récepteur est le siège d'un grand nombre de forces attractives ou répulsives de diverses natures. L'état stable d'un tel complexe dans un environnement donné correspond au minimum de son énergie libre totale $G = H - TS$. Le plus souvent, dans les modélisations, on ne considère que le complexe drogue/récepteur dont l'état stable correspond au minimum de l'énergie d'interaction entre les deux partenaires. Cette énergie potentielle est définie de façon empirique et dépend de la constitution des molécules ainsi que de leur distance. Les interactions non covalentes sont représentées par différents termes additifs correspondant aux interactions électrostatiques, aux interactions de van der Waals et aux interactions par liaison hydrogène.

A ces termes, on peut ajouter des termes correspondant aux perturbations conformationnelles des deux molécules : rotations autour des liaisons, déformation des liaisons (longueur et angle de valence). L'ensemble de ces termes définit l'énergie d'interaction intermoléculaire, calculée et minimisée par les méthodes de mécanique moléculaire. Dans certains cas, il est nécessaire d'ajouter des termes de transfert de charge et de polarisation.

En dehors des termes d'interactions électrostatiques qui sont généralement plus élevés, les termes d'interaction non covalente énumérés ci-dessus correspondent à des énergies de quelques kJ/mole à quelques dizaines de kJ/mole. La liaison covalente met en jeu des énergies beaucoup plus élevées. L'un des objectifs de la modélisation est de déterminer l'importance relative de ces interactions dans la stabilisation du complexe ligand/récepteur. Dans ce paragraphe, nous discuterons la nature et l'importance relative de ces interactions.

II.3.B.1 - Les liaisons covalentes.

Certaines drogues présentant des fonctions HS-, HO-, et HN- peuvent donner naissance à ce type de liaison avec les récepteurs. Parmi ces drogues, on dénote des agents antibactériens¹⁴³, des diurétiques mercuriels, certains inhibiteurs de l'enzyme cholinestérase¹⁴⁴ et des agents adrénergiques bloquants. Les ponts disulfures intervenant dans la structure tertiaire des protéines sont également des liaisons covalentes.

II.3.B.2 - Les interactions électrostatiques.

Pour calculer l'énergie électrostatique d'interaction E_E de deux molécules, c'est à dire de deux distributions de charges, on utilise préférentiellement l'approximation monopolaire. L'énergie d'interaction de deux charges ponctuelles q_1 et q_2 est inversement proportionnelle à la distance r_{12} les séparant.

$$E_E = (q_1 q_2) / 4\pi\epsilon r_{12}$$

où r_{12} est la distance séparant les deux charges q_1 et q_2 et $\epsilon = \epsilon_0\epsilon_r$ la constante diélectrique du milieu avec ϵ_0 permittivité du vide et ϵ_r permittivité relative. Dans le vide ($\epsilon_r = 1$), l'énergie d'interaction de deux ions, un anion et un cation monovalents, situés à 5 Å l'un de l'autre, est de l'ordre de -277.1 kJ/mole. En milieu aqueux ($\epsilon_r = 80$), cette énergie n'est que de -3.5 kJ/mole, donc à peine supérieure à celle de l'agitation thermique. Les interactions dépendent donc fortement des propriétés du milieu. A longue distance, l'énergie électrostatique intermoléculaire diminue très vite en milieu polaire. Cette énergie est surestimée si ϵ_r reste constant. Il faudrait en toute rigueur tenir compte des molécules de solvant pouvant venir s'intercaler entre les différentes espèces en interaction. La constante diélectrique ϵ_r varie en effet en fonction de la distance et n'est pas la même pour toutes les paires d'atomes en interaction. Pour simuler la situation dans un solvant organique apolaire, on prend souvent $\epsilon_r = 2$ à 4, pour le milieu physiologique $\epsilon_r = 28$ ¹⁴⁵.

Rein *et al.*¹⁴⁶ ont montré que l'approximation monopolaire généralement utilisée pour évaluer l'énergie électrostatique était très contestable. En effet, pour des distances $< 5 \text{ \AA}$, le développement limité de l'expression de l'énergie électrostatique fait apparaître des termes multipolaires non négligeables. Ces termes multipolaires correspondent principalement à des interactions de type dipôle-dipôle Edip,

$$E_{\text{DIP}} = (\mu_1\mu_2)/4\pi\epsilon r_{12}^3$$

ou μ_1 et μ_2 sont les moments dipolaires des deux liaisons, r_{12} la distance entre ces dipôles.

L'ensemble des forces d'origine électrostatiques joue un rôle important dans les mécanismes d'association d'une drogue à un récepteur. Une grande majorité de drogues et de sites récepteurs comportent des groupements polaires et peuvent, aux pH physiologiques, se retrouver sous forme d'espèces chargées ou protonées. Par exemple, les groupements ammonium du récepteur transmembranaire 5-HT_{1A} généreront de fortes interactions électrostatiques.

II.3.B.3 - Les liaisons hydrogènes.

Les liaisons hydrogènes maintiennent et stabilisent la conformation des protéines, des acides nucléiques, l'eau et certaines molécules organiques. Une liaison hydrogène résulte de l'attraction de deux ou plusieurs atomes électronégatifs voisins et de la mise en commun d'un atome d'hydrogène lié de manière covalente à l'un de ces atomes (Figure 20).

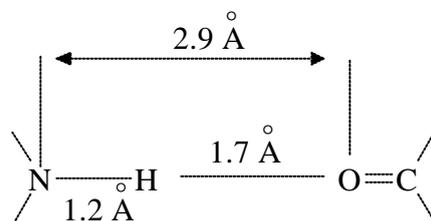


Figure 20. Géométrie d'une liaison hydrogène entre unités peptidiques.

Les atomes d'une molécule susceptibles d'accepter un proton doivent posséder au moins un doublet libre sur leur dernière couche électronique. Les atomes d'oxygène, d'azote et de fluor sont de bons accepteurs de protons, et à un degré moindre, le soufre et les autres halogènes. L'énergie de ces liaisons est comprise entre -12 et -33 KJ/mole. De très nombreuses fonctions ont été proposées pour représenter la liaison hydrogène^{147,148,149,150}. Ces fonctions doivent rendre compte de la géométrie de la liaison.

Pour une liaison hydrogène linéaire X-H...Y, la distance H...Y varie généralement entre 1.6 et 2.5 Å. A courte distance, un transfert de charge peut se produire au niveau du doublet libre de l'atome électronégatif accepteur Y. Bien qu'une liaison hydrogène linéaire soit le cas idéal d'un point de vue énergétique, il n'en demeure pas moins qu'une grande flexibilité est observée. L'angle défini par les trois protagonistes X, H et Y peut varier de 180° à 125°. Les plus faibles valeurs de cet angle sont rencontrées dans le cas particulier de liaisons hydrogènes mettant en jeu deux accepteurs. La liaison hydrogène se distingue des autres interactions dites faibles par son caractère directionnel.

De nombreuses drogues (ligands) contiennent des groupements hydroxyles, carbonyles, amides, amines et imines pouvant être impliqués dans la formation d'une liaison hydrogène, soit en tant que donneurs, soit en tant qu'accepteurs de protons. Les macromolécules biologiques où se situent les sites actifs des récepteurs contiennent également de tels groupes. Des liaisons hydrogènes peuvent donc intervenir dans la formation du système ligand/récepteur.

II.3.B.4 - Les transferts de charges.

Le complexe formé entre deux molécules en présence d'une liaison hydrogène est un cas particulier du phénomène plus généralement rencontré des complexes donneur-accepteur. Ces complexes résultent de l'attraction électrostatique à courte distance s'établissant entre une molécule pouvant céder une charge électronique, et une autre molécule pouvant accepter cette charge. Un électron de l'orbitale HOMO (Highest Occupied Molecular Orbital) de la molécule « donneur » peut ainsi être excité et passer dans l'orbitale LUMO (Last Unoccupied Molecular Orbital) de la molécule « accepteur ». En fait, dans les systèmes biologiques, le transfert de charge n'est pas complet, seule une fraction de charge est transférée.

Cette notion de complexe par transfert de charge développée par Mulliken¹⁵¹ a permis d'expliquer bon nombre d'interactions au niveau de divers systèmes biologiques. Les macromolécules biologiques pouvant jouer le rôle de sites récepteurs, possèdent des groupements protéiques qui peuvent participer soit en tant que donneur, soit en tant qu'accepteur, soit simultanément en tant que donneur et accepteur. L'interaction par transfert de charge s'applique plus particulièrement aux molécules à électrons π , c'est à dire à électrons délocalisés. Les systèmes π déficitaires jouent le rôle d'accepteurs et les systèmes π excédentaires le rôle de donneurs.

II.3.B.5 - Les interactions de van der Waals.

Les interactions de van der Waals-London ou interactions dispersives résultent de la polarisation instantanée induite lors du rapprochement de deux atomes non liés. Cette polarisation entraîne une déformation des nuages électroniques, les électrons de chaque atome étant respectivement attirés par le noyau positif de l'autre. A l'équilibre, ces forces attractives (forces de van der Waals) sont contrebalancées par des forces de répulsion électrostatique entre les nuages électroniques. Les forces dispersives de van der Waals-London s'appliquent aussi bien aux molécules polaires qu'aux molécules non polaires¹⁵². Pour exprimer l'énergie potentielle de ces interactions, on groupe généralement le terme attractif de van der Waals en $1/r^6$ et un terme répulsif en $1/r^{12}$.

$$V(r) = -A/r^6 + B/r^{12}$$

où r est la distance interatomique, et A et B des constantes caractéristiques de la nature électronique des atomes, ajustées d'après les résultats de spectrométrie hertzienne et de cristallographie. L'énergie potentielle d'interaction de van der Waals est minimale lorsque les deux atomes se trouvent à une distance d , telle que l'équilibre des forces attractives et répulsives soit atteint. La distance $d_0/2$ pour deux atomes identiques est communément appelée rayon de van der Waals (Figure 21).

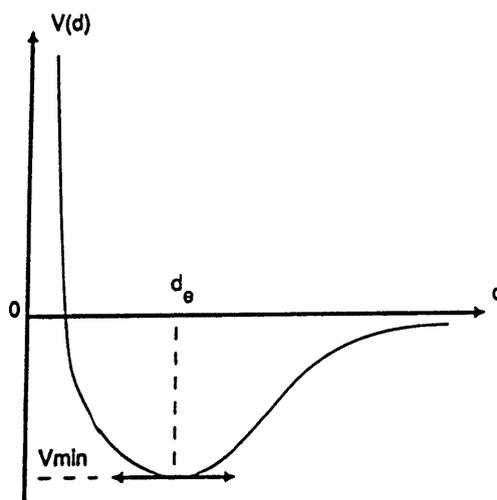


Figure 21. Energie potentielle $V(d)$ en fonction de la distance d .

Bien que de faible intensité, ces forces à courte distance, lorsqu'elles s'appliquent à de longues chaînes carbonées ou divers cycles aromatiques, peuvent contribuer significativement aux interactions ligand/récepteur.

II.3.B.6 - Les interactions hydrophobes.

L'eau est un solvant qui a la particularité d'avoir ses molécules connectées par un réseau interne de liaisons hydrogènes. L'introduction de molécules présentant des zones hydrophobes, c'est à dire non polaires, entraîne la rupture d'un certain nombre de ces liaisons. Chaque molécule de soluté prise séparément, exposée aux molécules d'eau, contribuerait à une augmentation défavorable de l'énergie libre du système soluté/solvant (facteur entropique). Si les zones non polaires des molécules s'auto-associent, l'effet perturbateur sur le solvant est moins important que les effets combinés pris séparément, l'association des zones non polaires est donc thermodynamiquement plus favorable. Cette association est appelée interaction hydrophobe ou lipophile, et résulte de la combinaison des interactions de van der Waals-London et des interactions par liaisons hydrogènes. L'énergie libre libérée lors de l'interaction hydrophobe de deux groupements $-CH_2-$ appartenant à deux chaînes alkyles voisines est de l'ordre de 3 kJ/mole. Les parties (chaînes) alkyles des ligands donneront donc lieu à d'importantes interactions de type hydrophobe avec le récepteur.

De nombreuses corrélations ont été mises en évidence entre l'activité biologique d'une drogue et sa lipophilie. La lipophilie globale d'une molécule est généralement évaluée à partir de son

coefficient de partage *n*-octanol/eau ($\text{Log } P_{\text{ow}}$), le couple *n*-octanol/eau ayant été choisi pour ses propriétés de partage proches de celles du milieu biologique¹⁵³. La lipophilie globale d'une drogue intervient essentiellement lors du franchissement des membranes biologiques rencontrées durant l'acheminement de celle-ci vers son site récepteur. Une drogue diffusera d'autant mieux à travers la bicouche phospholipidique des membranes que son coefficient de partage sera élevé en faveur de la phase non aqueuse, mais un trop fort coefficient gênera son transport dans les milieux aqueux.

L'ensemble de ces forces est en général représenté, par des composantes, dans les champs de forces.

Références Chapitre 2

-
- ¹ Lipkowitz, K.B.; Boyd, D.B. *Reviews in Computational Chemistry II*, VCH Publishers, New York, **1991**.
 - ² Pople, J.A.; Segal, G.A. *J. Chem. Phys.* **1965**, *43*, 136.
 - ³ Pople, J.A.; Beveridge, D.L.; Dobosh, P.A. *J. Chem. Phys.* **1967**, *47*, 2026.
 - ⁴ Dewar, M.J.S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
 - ⁵ Dewar, M.J.S.; Haselbach, E. *J. Am. Chem. Soc.* **1970**, *92*, 590.
 - ⁶ Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
 - ⁷ Stewart, J.J.P. *J. Comput. Chem.* **1989**, *10*, 209.
 - ⁸ Stewart, J.J.P. *J. Comput. Aided Mol. Design* **1990**, *4*, 1-105.
 - ⁹ Hehre, W.J.; Radom, L.; Schleyer, P.V.R.; Pople, J.A. *Ab Initio Molecular Orbital Theory*, Ed. J. Wiley, New York, **1986**.
 - ¹⁰ Vinter, J.G.; Davis, A.; Saunders, M.R. *J. Comput. Aided Mol. Design*, **1987**, *1*, 31-51.
 - ¹¹ Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
 - ¹² Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem.* **1986**, *7*, 230.
 - ¹³ Pearlman, D.A.; Case, D.A.; Caldwell, J.C.; Seibel, G.L.; Singh, U.C.; Weiner, P.; Kollman, P.A. AMBER 4.0 **1991** University of California.
 - ¹⁴ Brooks, C.L.; Brucoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.
 - ¹⁵ Van Gunsteren, W.F.; Berendsen, H.J.C. *Gronigen Molecular Simulation (GROMOS) Library manual* **1987**, Groningen : Biomos.
 - ¹⁶ Hagler, A.T.; Dauber, P.; Lifson, S. *J. Am. Chem. Soc.* **1979**, *101*, 5131-5141.
 - ¹⁷ Jorgensen, W.L.; Tirado Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.

-
- 18 Loncharich, R.J.; Brooks, B.R. *Proteins* **1989**, *6*, 32-45.
- 19 Van Gunsteren, W.F.; Berendsen, H.J.C. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992-1023.
- 20 Van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. *J. Mol. Biol.* **1987**, *198*, 721-735.
- 21 Van Gunsteren, W. *Eur. J. Biochem.* **1992**, *204*, 947-961.
- 22 Karplus, M.; Petsko, G.A. *Nature* **1990**, *347*, 631-639.
- 23 Press, W.H.; Flannery, B.P.; Teukolsky, S.A.; Vetterling, W.T. Ed. *Numerical Recipes In Pascal : The Art of Scientific Computing* **1989**, Cambridge University Press : Cambridge 759.
- 24 Beeman, D. *J. Comp. Phys.* **1976**, *20*, 130-139.
- 25 Fletcher, R. *Practical Methods of Optimisation*, Vol 1, Wiley, New York, **1980**.
- 26 Wiberg, K.B. *Comput. Chem.* **1977**, *1*, 221.
- 27 Wiberg, K.B. *J. Am. Chem. Soc.* **1965**, *87*, 1070.
- 28 Gelin, B.R.; Karplus, M. *J. Am. Chem. Soc.* **1975**, *97*, 6996.
- 29 Van't Hoff, J.H. *Bull. Soc. Chim. Paris* **1875**, *23*, 295.
- 30 Le Bel, J.A. *Bull. Soc. Chim. Paris* **1874**, *22*, 337.
- 31 Barton, D.H.R. *Experientia* **1950**, *6*, 316.
- 32 Hassel, O. *Q. Rev.* **1953**, *7*, 221.
- 33 Valleau, J.P. *J. Comp. Phys.* **1991**, *96*, 193-216.
- 34 Noguti, T.; Go, N. *Biopolymers* **1985**, *24*, 527-546.
- 35 Cobra V3.2, Oxford Molecular Ltd, Magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford, OX4 4GA, England.
- 36 Leach, A.R.; Prout, C.K. *J. Comp. Chem.* **1990**, *11*, 1193-1205.
- 37 Alder, B.J.; Wainwright, T.E. *J. Chem. Phys.* **1957**, *27*, 1208.
- 38 Alder, B.J.; Wainwright, T.E. *J. Chem. Phys.* **1959**, *31*, 459.

-
- 39 Rahman, A.; Stillinger, F.H. *J. Chem. Phys.* **1974**, *60*, 1545.
- 40 Mc Cammon, J.A.; Gelin, B.R.; Karplus, M. *Nature (London)* **1977**, *267*, 585.
- 41 Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.
- 42 Hockney, R.W.; Eastwood, J.W. *Computer Simulation Using Particles*, Mc Graw Hill, New York, **1981**.
- 43 Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. *J. Comput. Phys.* **1977**, *23*, 327-341.
- 44 Berendsen, H.J.C.; Postma, J.P.M.; Van Gunsteren, W.F.; Dinola, A.; Haak, J.R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- 45 Kennard, O.; Watson, D.G.; Allen, F.H.; Isaacs, N.W.; Motherwell, W.D.S.; Pettersen, R.C.; Town, W.G. *Molecular Structures and Dimensions*, Crystallographic Data Centre, Cambridge, U.K., Vol 1A, **1972**.
- 46 Abola, E.E.; Bernstein, F.C.; Koetzle, T.F., in *The Role of Data in Scientific Progress*. Glaeser, P.S., Ed., Elsevier, New York, **1985**. The Protein Data Bank.
- 47 a) Humblet, C.; Marshall, G. *Annu. Rep. Med. Chem.* **1980**, *15*, 267.
b) Humblet, C.; Marshall, G. *Drug. Dev. Res.* **1981**, *1*, 409.
- 48 a) Hibert, M.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K., Fozard, J. R. *J. Med. Chem.* **1988**, *31*, 1087.
b) Hibert, M.; McDermott, I.; Middlemiss, D. N.; Mir, A. K., Fozard, J. R. *Eur. J. Med. Chem.* **1989**, *24*, 31-37.
- 49 Mellin, C.; Vallgarda, J.; Nelson, D. L.; Björk, L.; Yu, H.; Anden, N. E.; Csöregi, I.; Arvidsson, L. E.; Hacksell, U. *J. Med. Chem.* **1991**, *34*, 497-510.
- 50 Glennon, R. A., Westkaemper, R. B.; Bartyzel, P. *Serotonin Receptor Subtypes: basic and Clinical Aspects*, pp. 19-64, Peroutka, S. J., Ed.; Wiley-Liss; NY, **1991**.
- 51 Chidester, C. G.; Lin, C. H.; Lahti, R. A.; Haadsma-Svensson, S. R.; Smith, M. W. *J. Med. Chem.* **1993**, *36*, 1301-1315.
- 52 Mokrosz, M. J.; Duszynska, B.; Bojarski, A. J.; Mokrosz, J. L. *Bio. Med. Chem.* **1995**, *3*, 533-538.
- 53 Kubinyi, H. *QSAR : Hansch Analysis and Related Approaches*, ed. Mannhold, R.; Krogsgaard Larsen, P.; Timmerman, VCH, New York, **1993**.

-
- 54 Crum Brown, A.; Fraser, T.R. *Trans. Roy. Soc. Edinburgh* **1868-69**, 151-203, 693-739.
- 55 Richet, M.C. *Compt. Rend. Soc. Biol. (Paris)* **1893**, 45, 775-776.
- 56 Meyer, H. *Arch. Exp. Path. Pharm.* **1899**, 42, 109-118.
- 57 Overton, E. *Studien über die Narkose, zugleich ein Beitrag zur allgemeinen Pharmakologie*, Fischer, G., Jena, **1901**.
- 58 a) Lipnick, R. L. *Trends Pharmacol. Sci.* **1986**, 7,161-164.
b) Lipnick, R. L. *Trends Pharmacol. Sci.* **1989**, 10, 265-269.
- 59 Fühner, H.; Neubauer, E. *Arch. Exp. Path. Pharm.* **1907**, 56, 333-345.
- 60 Bruce, T.C.; Kharasch, N.; Winzler, R.J. *Arch. Biochem. Biophys.* **1956**, 62, 305-317.
- 61 Hansen, O.R. *Acta Chem. Scand.* **1962**, 16, 1593-1600.
- 62 Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, 86, 1616-1626.
- 63 Free Jr, S.M.; Wilson, J.W. *J. Med. Chem.* **1964**, 7, 395-399.
- 64 Wold, S.; Ruhe, A.; Wold, H.; Dunn III, W.J. *SIAM J. Sci. Stat. Comput.* **1984**, 5, 735-743.
- 65 Cramer III, R. D.; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.* **1988**, 110, 5959-5967.
- 66 ASP V3.0 Oxford Molecular Ltd, The magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kingdom.
- 67 TSAR V2.22 Oxford Molecular Ltd, The magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kingdom.
- 68 Hall, L.H.; Kier, L.B. *The Molecular Connectivity chi Indexes and Kappa Shape Indexes in Structure-Property Modeling, Reviews in Computational Chemistry* ed. Lipkowitz, K.B.; Boyd, D.B., VCH Publishers, New York, **1991**.
- 69 Kier, L.B.; Hall, L.H. *J. Pharm. Sci.* **1981**, 70, 583.
- 70 Kier, L.B.; Hall, L.H. *J. Pharm. Sci.* **1983**, 72, 1170.
- 71 Kier, L.B.; Hall, L.H. *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, Wiley, Letchworth, England, **1986**.

-
- 72 Kier, L.B.; Hall, L.H. *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, **1976**.
- 73 Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- 74 Sabljic, A. *J. Chromatogr.* **1985**, *319*, 1.
- 75 Edward, J.T.; Farrell, P.G.; Shahidi, F. *J. Phys. Chem.* **1978**, *82*, 2310.
- 76 Watanabe, K.; Nakayama, T.; Mottle, J. *J. Quant. Spectrosc. Radiat. Transfer.* **1962**, *2*, 369.
- 77 Hall, L.H.; Maynard, E.L.; Kier, L.B. *Environ. Toxicol. Chem.* **1989**, *8*, 783.
- 78 Hansch, O.R. *Acta Chem. Scand.* **1962**, *16*, 1593-1600.
- 79 Perrin, D.D.; Dempsey, B.; Serjeant, E.P. *pKa Prediction for Organic Acids and Bases*, Chapman and Hall, London, **1981**.
- 80 Hetnarski, B.; O'Brien, R.D. *J. Med. Chem.* **1975**, *18*, 29-33.
- 81 Livingstone, D.J.; Hyde, R.M.; Foster, R. *Eur. J. Med. Chem.* **1979**, *14*, 393-397.
- 82 Lien, E.J.; Liao, R.C.H.; Shinouda, H.G. *J. Pharm. Sci.* **1979**, *68*, 463-465.
- 83 Dearden, J.C.; Nicholson, R.M. *Pestic. Sci.* **1986**, *17*, 305-310.
- 84 Debnath, A.K.; Lopez de Compadre, R.L.; Debnath, G.; Shusterman, A.J.; Hansch, C. *J. Med. Chem.* **1991**, *34*, 786-797.
- 85 Young, R.C.; Mitchell, R.C.; Brown, T.H.; Ganellin, C.R.; Griffiths, R.; Jones, M.; Rana, K.K.; Saunders, D.; Smith, I.R.; Sore, N.E.; Wilks, T.J. *J. Med. Chem.* **1988**, *31*, 656-671.
- 86 Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity : The Hydrophobic Fragmental Constant Approach*, VCH, New York, **1992**.
- 87 Ghose, A.K.; Crippen, G.M. *J. Comput. Chem.* **1986**, *7*, 565-577.
- 88 Taft Jr, R.W. *Steric Effects in Organic Chemistry*, Newman, M.S.; ed., Wiley, New York, **1956**.
- 89 Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, **1979**.

-
- 90 Ghose, A.K.; Crippen, G.M. *J. Chem. Inf. Comput. Sci.*, **1987**, 27, 21-35
- 91 Viswanadhan, V.N.; Ghose, A.K.; Revankar, G.R.; Robins, R.K. *J. Chem. Inf. Comput. Sci.* **1989**, 29, 163.
- 92 Broto, P.; Moreau, G.; Van Dycke, C. *Eur. J. Med. Chem.* **1984**, 19, 66-70.
- 93 Calas, M.; Bompart, J.; Giral, L.; Grassy, G. *Eur. J. Med. Chem.* **1991**, 26, 279.
- 94 Montagut, M.; Saux, M.; Carpy, A; Grassy, G. *Eur. J. Med. Chem.* **1990**, 25, 387-395.
- 95 Stigliani, J.L.; Bonnafous, M.; Grassy, G. *Eur. J. Med. Chem.* **1990**, 25, 157-170.
- 96 Balaban, A.T. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 334.
- 97 Balaban, A.T. *Chemical Applications of Graph Theory*, Academic Press, London, **1976**.
- 98 Wiener, H.J. *J. Am. Chem. Soc.*, **1972**, 15, 1006.
- 99 Randic, M.J. *J. Am. Chem. Soc.*, **1975**, 97, 6609.
- 100 Balaban, A.T. *Chem. Phys. Lett.*, **1982**, 89, 399.
- 101 Mihalic, Z.; Trinajstic, N.J. *J. Med. Chem.*, **1992**, 69, 701.
- 102 Chatfield, C.; Collins, A.J. *Introduction to Multivariate Analysis*, Chapman et Hall, **1980**.
- 103 Manly, B.F.J. *Multivariate Statistical Methods A Primer*, Chapman et Hall, **1986**.
- 104 Dunn III, W. J.; Wold, S. *Quantitative Drug Design*. Hansh, C.; Sammes, P. G.; Taylor, J. B. ed. Pergamon Press, Oxford, **1990**.
- 105 Martin, Y. C.; Holland, J. B.; Jarboe, C. H.; Plotnikoff, N. *J. Med. Chem.* **1974**, 17, 409-413.
- 106 Prakash, G.; Hodnett, E. M. *J. Med. Chem.* **1978**, 21, 369-374.
- 107 Daniel, C.; Wood, F.S. *Fitting Equation to Data*, Wiley, New York, **1980**.
- 108 Draper, N.R.; Smith, H. *Applied Regression Analysis*, Wiley, New York, **1981**.

-
- 109 Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, The Iowa State University Press, Ames, **1973**.
- 110 Stahle, L.; Wold, S. *Progress in Medicinal Chemistry*, ed. Ellis, G.P.; West, G.B., Elsevier, **1988**.
- 111 Kubinyi, H. *3D QSAR in Drug Design. Theory, Methods, and Applications*, ESCOM Science Publishers, Leiden, **1993**.
- 112 Dunn III, W.J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. *J. Quant. Struct.-Act. Relat.* **1984**, 3, 131-137.
- 113 Simon, Z.; Badileuscu, I.; Racovitan, T. *J. Theor. Biol.* **1977**, 66, 485.
- 114 Simon, Z.; Dragomir, N.; Plauchithiu, M.G.; Holban, S.; Glatt, H.; Kerek, F. *Eur. J. Med. Chem.* **1980**, 15, 521.
- 115 Hopfinger, A.J. *J. Am. Chem. Soc.* **1980**, 102, 7196-7206.
- 116 Hopfinger, A.J. *J. Med. Chem.* **1983**, 26, 990-996.
- 117 Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, 37, 4130-4146.
- 118 Good, A.C.; So, S.S.; Richards, W.G. *J. Med. Chem.* **1993**, 36, 433-438.
- 119 Carbo, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, 17, 1185-1189.
- 120 Hodgkin, E.E.; Richards, W.G. *Int. Quantum Chem. Quantum Biol. Symp.* **1987**, 14, 105-110.
- 121 Meyer, A.Y.; Richards, W.G. *J. Computer Aided Mol. Design* **1991**, 5, 427.
- 122 Good, A. C.; Hodgkin, E. E.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 188-191.
- 123 Nelder, J.A.; Mead, R. *Comput. J.* **1965**, 7, 308-313.
- 124 Richards, W. G. *The dominant Role of Shape Similarity and Dissimilarity in QSAR*, 10th European Symposium on SAR: QSAR and Molecular Modelling. Barcelone, 4-9 Septembre **1994**.
- 125 Kellogg, G.E.; Semus, S.F.; Abraham, D.J. *J. Comput Aided Mol. Design* **1991**, 5, 545-552.
- 126 Gantchev, T.G.; Ali, H.; Van Lier, J.E. *J. Med. Chem.* **1994**, 37, 4164-4176.

-
- 127 Nicklaus, M.C.; Milne, G.W.; Burke Jr, T.R. *J. Comput Aided Mol. Design* **1992**, *6*, 487-504.
- 128 Van Steen, B. J.; Van Wijngaarden, I.; Tulp, M.T.; Soudijn, W. *J. Med. Chem.* **1994**, *37*, 2761-2773.
- 129 Langlois, M.; Bremont, B.; Rousselle, D.; Gaudy, F. *Eur. J. Pharmacol.* **1993**, *244*, 77-87.
- 130 Agarwal, A.; Pearson, P.P.; Taylor, E.W.; Li, H.B.; Dahlgren, T.; Herslof, M.; Yang, Y.; Lambert, G.; Nelson, D.L.; Regan, J.W.; Martin, A.R. *J. Med. Chem.* **1993**, *36*, 4006-4014.
- 131 Agarwal, A.; Taylor, E. W. *J. Comput. Chem.* **1993**, *14*, 237-245.
- 132 Henderson, R.; Baldwin, J.; Ceska, T.H.; Zemlin, F.; Beckmann, E.; Downing, K. *J. Mol. Biol.* **1990**, *213*, 899-929.
- 133 Hargrave, P.A. *Prog. Retinal Res.* **1982**, *1*, 1-51.
- 134 Popot, J.L.; Engelman, D.M.; Gurel, O.; Zaccai, G. *J. Mol. Biol.* **1989**, *210*, 829-847.
- 135 Tuffery, P.; Popot, J.L.; Lavery, R. *Structures and Functions of Retinal Proteins*. Ed., Rigaud, J.L. Colloque INSERM, Libbey, J.; Eurotext Ltd, Vol. 221, pp. 13-16, **1992**.
- 136 Kobilka, B.K.; Frielle, T.; Collins, S.; Yang Feng, T.L.; Kobilka, T.S.; Francke, U.; Lefkowitz, R.J.; Caron, M.G. *Nature* **1987**, *329*, 75-79.
- 137 Jähnig, F.; Edholm, O. *J. Mol. Biol.* **1992**, *226*, 837-850.
- 138 Trumpp kallmeyer, S.; Hoflack, J.; Bruinvels, A.; Hibert, M. *J. Med. Chem.* **1992**, *35*, 3448-3462.
- 139 Needleman, S.B.; Wunsch, C.D.A. *J. Mol. Biol.* **1970**, *48*, 443-453.
- 140 Dayhoff, H.O.; Schwarz, R.M.; Orcutt, B.C. *In Atlas of Protein Sequence and Structure*; Dayhoff, M.O.; ed., NBFRR, Washington, DC, **1978**.
- 141 Kyte, J.; Doolittle, R.F. *J. Mol. Biol.* **1982**, *157*, 105-132.

-
- 142 Engelman, D.M.; Goldman, A.; Steitz, T.A. *Annu. Rev. Biophys. Chem.* **1986**, *15*, 321-353.
- 143 Cavallito, C.; Bailey, J.H.; Mc Cormick, J.R. Warner, W.F. *J. Bact.* **1945**, *50*, 61-69.
- 144 Ban, T.; Nagata, C. *Jap. J. Pharmacol.* **1965**, *15*, 32.
- 145 Gill, E.W. *Prog. Med. Chem.* **1965**, *4*, 39.
- 146 Rein, R.; Rabinowitz, J.R.; Swissler, T.J. *J. Theor. Biol.* **1972**, *34*, 215.
- 147 Stockmayer, W.H. *J. Chem. Phys.* **1941**, *9*, 398.
- 148 Lippincott, E.R.; Schröder, R. *J. Chem. Phys.* **1955**, *23*, 1099.
- 149 Poland, D.; Scheraga, H.A. *Biochemistry* **1967**, *6*, 3791.
- 150 Mc Guire, R.F.; Momany, F.A.; Scheraga, H.A. *J. Phys. Chem.* **1972**, *76*, 375.
- 151 Mulliken, R.S. *J. Am. Chem. Soc.* **1952**, *74*, 811.
- 152 Pauling, L. *General Chemistry*, Freeman, San Francisco, **1947**.
- 153 Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 526.

CHAPITRE 3 : ETUDES QSAR

FAMILLE DES INDOLES

-

FAMILLE DES TETRALINES

-

FAMILLE DES CHROMANES

-

FAMILLE DES THIOCHROMANES

-

ENSEMBLE DES 382 LIGANDS

-

CONCLUSION ; BILAN DES PREDICTIONS

Ce chapitre sera consacré à la présentation d'une grande partie des travaux réalisés en analyse QSAR (Quantitative Structure-Activity Relationships) sur plusieurs familles de composés (tétraline, indole, chromane et thiochromane). Les données sur les composés utilisés pour ces études sont issues soit de la littérature, soit de la banque de données des composés synthétisés au Laboratoire. L'objectif de ce travail QSAR était multiple et nous cherchions :

- à vérifier la faisabilité d'une approche QSAR sur des composés à usage sérotoninergique,
- dans l'affirmative, à choisir parmi les méthodes utilisables, que ce soit d'un point de vue descripteur physico-chimique ou méthodes statistiques celle ou celles présentant le meilleur pouvoir prédictif,
- à utiliser ces méthodes dans un but prédictif pour déterminer *a priori* l'affinité ou la sélectivité de composés issus ou destinés à l'équipe de synthèse du Laboratoire.

Les 382 composés utilisés dans ces études QSAR ont été modélisés par mécanique moléculaire. Après une construction graphique, ces ligands ont fait l'objet d'une recherche conformationnelle, et d'une étude de superposition au pharmacophore 5-HT_{1A} de Mellin afin de sélectionner le conformère qui sera utilisé ultérieurement dans les études QSAR. Pour tous les composés, nous avons ensuite recensé certains descripteurs classiques comme les propriétés moléculaires (masse moléculaire, Log P calculé, moment dipolaire, etc...), les indices de topologie et de connectivité, mais aussi les indices de similarité moléculaire simulant des propriétés électrostatiques, lipophiliques et stériques. A partir de ces ensembles constitués de descripteurs moléculaires, plusieurs méthodes statistiques ont été utilisées pour essayer de corréler ces variables à l'activité biologique (pIC₅₀, pK_i). Ces méthodes sont l'analyse en composantes principales (ACP), l'analyse discriminante (AD), la régression multiple (RM) et le Partial Least Square (PLS).

Après validation statistique des modèles, nous avons testé la prédiction *a posteriori* de composés appartenant aux familles des tétralines et des indoles. Le crible constitué des composés de type chromanes a fait aussi l'objet de prédictions *a priori* de l'affinité 5-HT_{1A} sur plusieurs molécules qui ont ensuite été proposées à la synthèse. Il en a été de même pour la famille des thiochromanes avec en plus une étude de prédiction de la sélectivité entre différents récepteurs (5-HT_{1A}, α_1 , α_2 , D₂).

Nous avons enfin étudié un échantillon hétérogène au niveau des structures, comprenant la totalité de 382 ligands ayant une activité connue pour le neurorécepteur 5-HT_{1A}. Cet échantillon comprend toutes les familles précitées de même que d'autres composés issus soit de la banque de données des composés synthétisés au Laboratoire, soit de la littérature comme des dérivés à motif pipérazine, benzodioxane, etc...

I. FAMILLE DES INDOLES

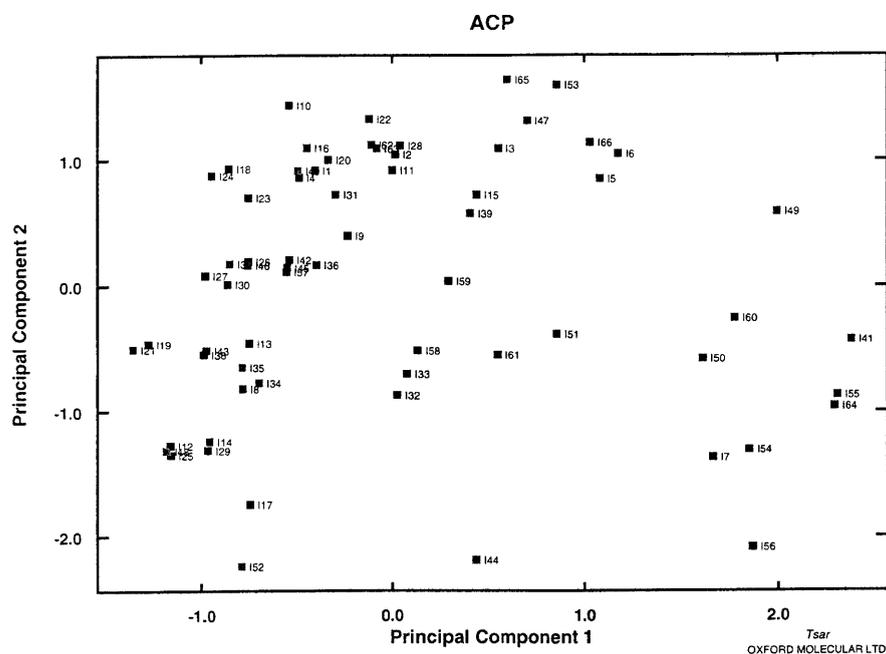
La composition de l'échantillon (annexe 1) est de 66 composés. Cet échantillon présente une gamme d'activité biologique de $pIC_{50} = 4$ à 9.7. Cet échantillon est séparé en trois sous-groupes formés respectivement de 30 composés inactifs (**I1** à **I30**) ayant un $pIC_{50} < 7.5$, de 22 composés actifs (**I31** à **I52**) avec un $pIC_{50} \geq 7.5$ et de 14 composés à prédire *a posteriori* (**I53** à **I66**). Ces 14 composés (7 actifs et 7 inactifs) pris au hasard des publications serviront de test pour évaluer le pouvoir prédictif des modèles proposés. Cet échantillon, pas totalement représentatif de la famille indole a fait l'objet d'une étude des descripteurs moléculaires classiques et des indices de similarité moléculaire. Pour l'étude des indices de similarité moléculaire, les composés ont été alignés selon les critères du pharmacophore 5-HT_{1A}. Les indices de Carbo ont été calculés par optimisation (rotation, translation et déformation des angles dièdres) des molécules. Les trois critères choisis sont la charge, la lipophilie et la forme des molécules. Différentes méthodes d'analyses statistiques ont été testées comme l'analyse en composantes principales (ACP), l'analyse discriminante (AD), la régression multiple (RM) et le Partial Least Square (PLS). Cet échantillon présente une gamme étendue d'activité avec une bonne disparité entre les composés inactifs et actifs. L'échantillon possède une diversité structurale (bicycle, tricycle et tétracycle) importante et de nombreuses substitutions fonctionnelles. De plus la présence de nombreux isomères de position présentant une différence d'activité semble intéressante pour tester le pouvoir de différenciation des descripteurs utilisés. Tous les composés pris en compte dans cette étude QSAR sont extraits de publications diverses¹⁻¹¹.

Etude des descripteurs classiques

Les descripteurs utilisés sont ceux proposés par le logiciel TSAR ; ils comprennent les indices de connectivité, de forme et de flexibilité, les indices topologiques, les propriétés moléculaires et les autocorrélogrammes utilisés avec un pas maximum de 11 pour le 2D et de 10 pour le 3D. Tous ces descripteurs forment un crible de 395 variables.

Une étude d'analyse en composantes principales (Graphe 2) sur ce crible a d'abord été envisagée pour essayer de séparer les deux sous-groupes de composés en fonction d'un ou plusieurs axes factoriels. Les résultats de cette ACP sont les suivants.

- les trois premiers axes contiennent 81% de l'information.
- l'axe principal N°1 qui est un axe de taille, discrimine les composés en différents « clusters ».



Graphe 2.

Une analyse discriminante classe les composés de l'échantillon à 96% où 100% des composés actifs sont bien classés et 92% des composés inactifs sont classés correctement. Les variables discriminant ces composés sont principalement des variables d'autocorrélogrammes 2D et 3D, de forme (indice de Kappa) et d'inertie (Tableau 5).

Variables	Total	Inactifs bien classés	Actifs bien classés
3D bin 9 (charge Q positive)	71 %	83 %	59 %
2D bin 5 (charge Q)	77 %	80 %	74 %
3D bin 0 (charge Q positive)	79 %	83 %	75 %
3D bin 2 (charge Q)	87 %	88 %	86 %
Volume ellipsoïdal	89 %	90 %	88 %
2D bin 6 (Log P)	94 %	93 %	95 %
Kappa 3	96 %	92 %	100 %
$r(CV)^2$	80 %	87 %	90 %

Tableau 5.

Les indices 2D et 3D donnent des informations sur la répartition électronique et lipophile le long du graphe moléculaire des composés. L'indice Kappa 3 donne lui une information sur le degré de ramification des molécules. Le volume ellipsoïdal montre l'importance de la taille des ligands dans la discrimination des composés. Les seuls composés inactifs mal classés sont le **I11** et **I13**. Ces ligands sont les seuls composés ayant une substitution en alpha de l'azote indolique et un substituant fonctionnel sur le cycle aromatique du motif indolique. Leurs homologues **I4** et **I14**, ne comportant pas de substituants sur le noyau aromatique sont, eux, classés inactifs comme le prédisent les études de « binding ». Le résultat de cette analyse discriminante est donc très intéressant de part sa validité statistique (pouvoir discriminant de 96 % avec sept variables). Les descripteurs classiques sélectionnés permettent de différencier d'une part la totalité des isomères de

position de cet échantillon et d'autre part les diverses familles de composés (bicyclique, tricyclique et tétracyclique).

Pour les composés à prédire (Tableau 6), les résultats donnent 5 composés actifs bien prédits et 2 molécules mal prédites. Tous les composés inactifs sont prédits correctement.

Composés	Prédiction	Expérimental
I53	Actif	Actif
I54	Actif	Actif
I55	Actif	Actif
I56	Inactif	Actif
I57	Inactif	Actif
I58	Inactif	Inactif
I59	Inactif	Inactif
I60	Inactif	Inactif
I61	Inactif	Inactif
I62	Inactif	Inactif
I63	Inactif	Inactif
I64	Actif	Actif
I65	Inactif	Inactif
I66	Actif	Actif

Tableau 6.

L'étude des prédictions donne des résultats en accord avec la validité statistique du modèle. En effet, la quasi-totalité des 14 composés à prédire le sont correctement. En ce qui concerne les deux ligands (**I56** et **I57**) dont la prédiction est mauvaise, nous pouvons spécifier que leur valeur expérimentale de binding (7.8 et 7.85) se trouve très proche du seuil de 7.5 choisi pour distinguer les composés actifs et inactifs.

Une analyse discriminante effectuée sur les trois premiers axes de l'analyse en composantes principales donne un coefficient de discrimination de 0.65% avec les trois axes. Ce résultat faible ne permet pas d'envisager une étude de prédiction pour valider ce modèle.

remarque

Pour l'analyse discriminante, il a fallu que nous déterminions un seuil séparant les composés inactifs des composés actifs. La valeur de ce seuil, arbitrairement choisie, est de 7.5 unités de pIC_{50} . Ce choix pose certaines questions :

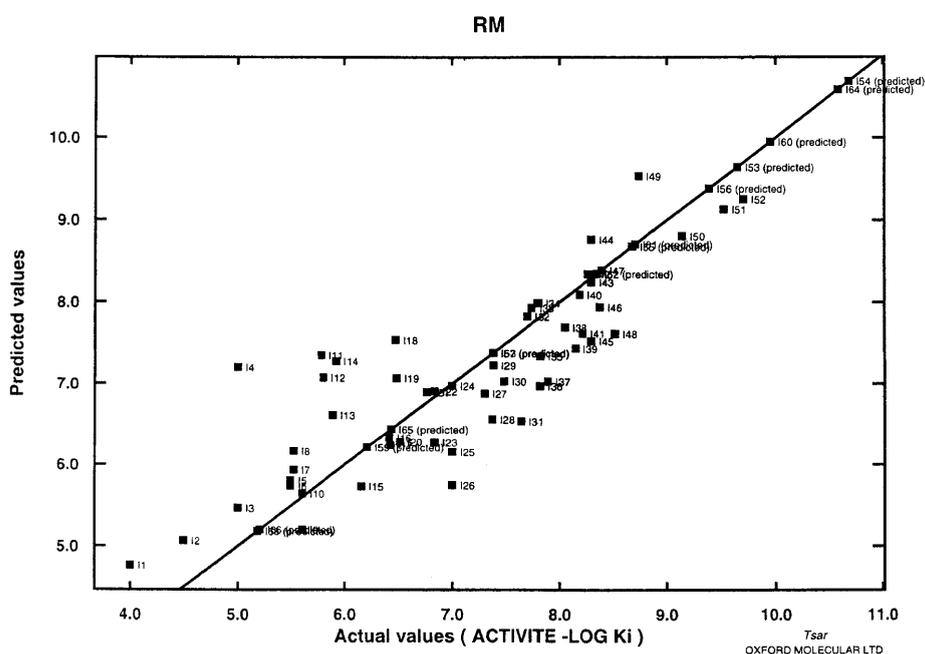
- Pourquoi le choix de cette valeur ?
- Quelle sera la validité des prédictions pour ces produits ayant une activité comprises entre 7 et 8 unités de pIC_{50} ?
- Pourquoi ne pas éliminer des échantillons toutes les molécules ayant une activité comprise entre 7 et 8 de façon à créer des classes plus différenciées ?

Pour répondre à la première question, nous dirons que cette valeur correspond à la valeur moyenne des activités extrêmes de l'ensemble des composés de notre échantillon. Le composé le plus actif à une activité de 11 et le plus faible de 4. Cette valeur de 7.5 semble donc une valeur moyenne de l'échelle d'activité de notre échantillon. En ce qui concerne la validité des prédictions, il est certain que ce choix implique un effet frontière : certains composés prédits inactifs peuvent se révéler faiblement actif et certains ligands prédits actifs seront faiblement inactifs. Pour répondre à la dernière question, nous dirons que le fait d'éliminer certains produits diminue la taille de l'échantillon et de plus supprime arbitrairement des composés qui apportent une information importante. Nous pensons que cette solution serait plus pénalisante pour nos modèles, ceci par une perte de composés et d'information.

La méthode d'analyse de régression multiple (Graphe 3) donne les résultats suivants ; le coefficient de régression est $r^2 = 0.70$, le « cross validation » qui donne le

pouvoir prédictif de l'équation $r(CV)^2$ est de 0.60. La probabilité de F est $5 \cdot 10^{-10}$, la valeur de s est 0.745.

L'équation utilise sept variables qui sont principalement des variables d'autocorrélogrammes à 2 et 3 dimensions et une variable d'inertie. Les variables sont V1 = 3D bin 9 (charge Q positive), V2 = 2D bin 6 (Log P), V3 = 3D bin 5 (charge Q négative), V4 = 2D bin 11 (Log P), V5 = 2D bin 4 (charge Q négative), V6 = RX et V7 = 2D bin 4 (charge Q). Cette équation prend donc en compte des effets stériques, lipophiliques et électrostatiques.



Graphe 3.

Equation

$$Y = -0.372 (V6) - 0.343 (V7) + 0.979 (V5) - 0.370 (V2) + 0.291 (V4) + 0.565 (V1) - 1.222 (V3) + 7.090$$

Les indices pris en compte sont des indices de connectivité pour le 2D et de forme pour le 3D et un indice d'inertie des molécules. Ces descripteurs sont pondérés soit par la charge partielle des atomes, soit par la lipophilie. Ceci est logique puisque ces descripteurs quantifient les trois types d'interactions existant dans le récepteur.

Composés	Prédiction	Expérimental	Différence
I53	9.5	8.5	1
I54	10.7	9.7	1
I55	8.6	7.6	1
I56	9.4	7.8	1.6
I57	7.45	7.85	0.4
I58	5.2	5	0.2
I59	6.2	5	1.2
I60	9.9	6.5	3.4
I61	8.6	6.6	2
I62	8.3	5.7	2.6
I63	7.4	5.9	1.5
I64	10.5	8.9	1.6
I65	6.4	6.2	0.2
I66	5.2	8.9	3.7

Tableau 7.

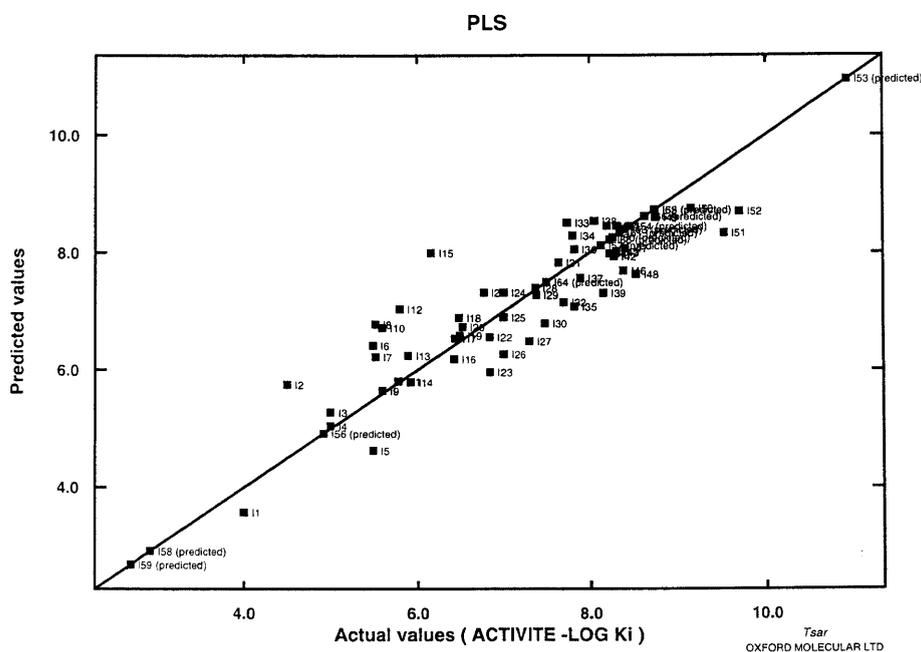
Sur les 14 composés (Tableau 7), 6 sont prédits avec un écart de pIC_{50} inférieur ou égal à 1, 5 sont prédits avec un écart de pIC_{50} inférieur ou égal à 2 et 3 composés ont un écart de prédiction compris entre 2.6 et 3.7. Parmi ces trois composés, deux sont des composés inactifs.

La dernière méthode utilisée est le Partial Least Square (PLS). Les résultats de cette étude sont regroupés dans le tableau 8 et dans le graphe 4.

Vecteurs	1	2	3	4	5	6	7	8	9	10
Signification statistique	0.760	0.851	0.956	1.001	1.241	1.289	1.445	1.989	2.019	2.453
PRESS	12.05	10.72	9.920	8.256	6.555	9.258	13.05	11.31	12.34	13.93
$r(CV)^2$	0.056	0.123	0.359	0.447	0.548	0.319	0.207	0.386	0.362	0.259
r^2	0.173	0.336	0.599	0.747	0.813	0.889	0.909	0.933	0.950	0.964

Tableau 8.

Le nombre de vecteurs que nous avons sélectionnés est de 5. Ce vecteur correspond à la première valeur de PRESS faible (6.555). Le coefficient de régression est de 0.81 avec un « cross validation » de 0.55.



Graphe 4.

Ce graphe 4, comme le graphe 3, montre que les composés à forte activité se trouvent plus proche de l'axe de la régression que ne le sont les composés à faible activité. En résumé, les composés actifs sont mieux prédits dans nos modèles que ne le sont les composés inactifs. Ce phénomène est pratiquement retrouvé dans toutes les études QSAR effectuées sur nos familles de composés. Nous attribuons ceci au fait que les composés actifs sont électrostatiquement, stériquement et lipophiliquement plus proche les uns des autres que ne le sont les composés inactifs entre eux. Ce qui fait qu'il existe une meilleure corrélation pour les composés actifs du fait de leurs faibles différences stériques, électrostatiques ou lipophiliques

Composés	Prédiction	Expérimental	Différence
I53	10.9	8.5	2.4
I54	8.5	9.7	1.2
I55	8.7	7.6	1.1
I56	5	7.8	2.8
I57	8.1	7.85	0.25
I58	3	5	2
I59	3	5	2
I60	8.25	6.5	1.75
I61	8.4	6.6	1.8
I62	8.7	5.7	3
I63	8.4	5.9	2.5
I64	7.5	8.9	1.4
I65	8.2	6.2	2
I66	8.6	8.9	0.3

Tableau 9.

Sur les 14 composés (Tableau 9), 2 sont prédits avec un écart de pIC_{50} inférieur ou égal à 1, 8 sont prédits avec un écart de pIC_{50} inférieur ou égal à 2 et 4 composés ont un écart de prédiction compris entre 2.4 et 3. Parmi ces quatre composés, deux sont des composés inactifs. Pour les composés **I59** et **I58**, l'écart est de 2 unités de pIC_{50} mais les composés restent toujours prédits très inactifs (3). On peut également noter que, dans le cas de cette analyse comme dans le cas précédent de la régression multiple, les composés qui sont prédits en dehors de la gamme de départ présentent une différence entre la valeur expérimentale et la valeur prédite plus importante, ce qui n'est que la vérification des principes de base des statistiques.

Etude des indices de similarité moléculaire

Potentiel de charge électrostatique

L'analyse en composantes principales donne un pourcentage d'information sur les trois premiers axes de 91%. La visualisation des graphes 2D et 3D n'apporte aucune information.

L'analyse discriminante sépare les composés du crible à 90%, avec 84% des composés inactifs bien classés et 96% des composés actifs bien classés (Tableau 10).

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo I15	85 %	83 %	87 %
Carbo I45	86 %	82 %	90 %
Carbo I6	87 %	84 %	90 %
Carbo I17	88 %	84 %	92 %
Carbo I39	90 %	84 %	96 %
r(CV) ²	65 %	73 %	54 %

Tableau 10.

Le pourcentage de discrimination est de 90 % avec une validité statistique de 65%. Parmi les composés inactifs, les ligands classés actifs sont les **I3**, **I7**, **I8**, **I28** et **I29**. Pour les ligands **I28** et **I29**, la valeur de binding (7.4) est proche de la valeur limite séparant les inactifs des actifs. En ce qui concerne les composés actifs, un seul composé le **I37** est mal classé. Ce ligand a une activité expérimentale de 7.9 unité de pIC₅₀ et est le seul composé du crible ayant un atome de chlore en position 5.

Pour les composés à prédire (Tableau 11), les résultats sont : 6 composés actifs sont bien prédits et 1 le **I64** est mal prédit. Un seul composé inactif est mal prédit, le **I63**.

Composés	Prédiction	Expérimental
I53	Actif	Actif
I54	Actif	Actif
I55	Actif	Actif
I56	Actif	Actif
I57	Actif	Actif
I58	Inactif	Inactif
I59	Inactif	Inactif
I60	Inactif	Inactif
I61	Inactif	Inactif
I62	Inactif	Inactif
I63	Actif	Inactif
I64	Inactif	Actif
I65	Inactif	Inactif
I66	Actif	Actif

Tableau 11.

L'étude des prédictions donne des résultats en accord avec la validité statistique du modèle. En effet, la quasi-totalité des 14 composés à prédire le sont correctement.

Une analyse discriminante effectuée sur les trois premiers axes de l'analyse en composantes principales donne un coefficient de discrimination de 0.45% avec l'axe 2. Ce résultat faible n'est pas suffisant pour envisager une étude de prédiction pour valider ce modèle.

La régression multiple a donné les résultats suivants : un coefficient de régression r^2 est de 0.46 avec sept variables. Ce coefficient est très faible et le nombre de variables est élevé. Ces données ne permettent pas d'exploiter cette régression multiple.

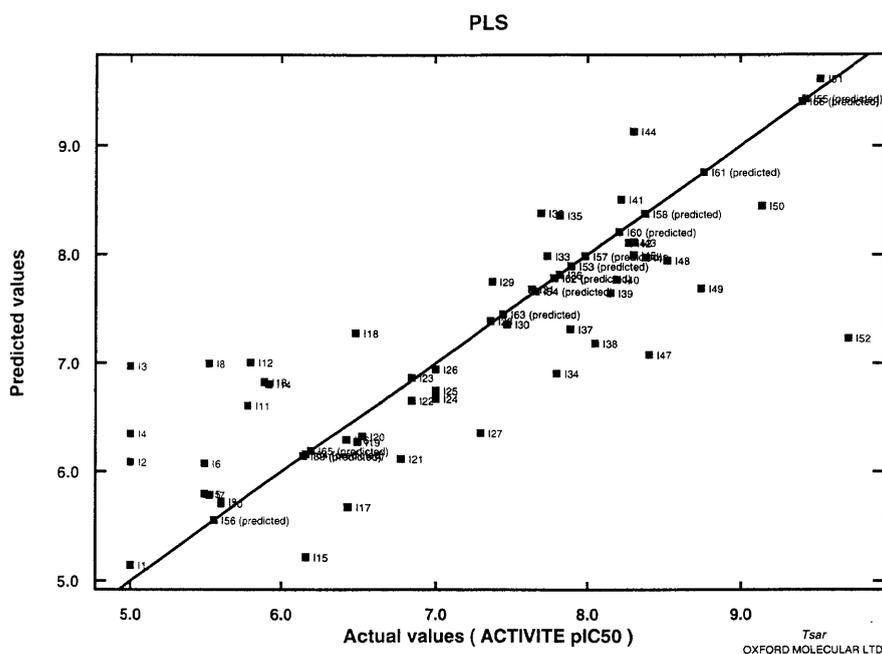
L'analyse Partial Least Square donne les résultats statistiques suivants (Graphe 5). Avec huit vecteurs, le coefficient de régression est de 0.61% avec un pouvoir prédictif de 0.65%. Les résultats de prédiction sont regroupés dans le tableau 12.

Composés	Prédiction	Expérimental	Différence
I53	8	8.5	0.5
I54	7.7	9.7	2
I55	9.4	7.6	1.8
I56	5.5	7.8	2.3
I57	7.9	7.85	0.05
I58	8.4	5	3.4
I59	6.1	5	1.1
I60	8.2	6.5	1.7
I61	8.6	6.6	2
I62	7.7	5.7	2
I63	7.5	5.9	1.6
I64	6.2	8.9	2.7
I65	6.2	6.2	0
I66	9.4	8.9	0.5

Tableau 12.

Sur les 14 composés, 4 ligands sont prédits avec un écart de pIC_{50} inférieur ou égal à 1, 7 sont prédits avec un écart de pIC_{50} inférieur ou égal à 2 et 3 composés ont

un écart de prédiction compris entre 2.3 et 3.4. Ce résultat montre que l'on peut avoir des prédictions plutôt correctes alors que la validité statistique du modèle est faible ($r^2 = 0.61\%$).



Graphe 5.

Potentiel de lipophilie

L'analyse en composantes principales donne un pourcentage d'information sur les trois premiers axes de 93%. L'analyse des graphes 2D n'apporte aucune information sur la formation éventuelle de « clusters ».

L'analyse discriminante (Tableau 13) sépare les composés du crible en deux groupes (actif et inactif) avec une probabilité de 98% et en utilisant 13 variables.

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo I37	73 %	87 %	59 %
Carbo I41	75 %	89 %	61 %
Carbo I14	77 %	80 %	73 %
Carbo I6	79 %	87 %	68 %
Carbo I43	83 %	87 %	77 %
Carbo I32	87 %	93 %	77 %
Carbo I2	89 %	93 %	82 %
Carbo I47	90 %	93 %	86 %
Carbo I48	90 %	93 %	86 %
Carbo I27	90 %	93 %	86 %
Carbo I16	90 %	90 %	91 %
Carbo I46	94 %	97 %	91 %
Carbo I50	98 %	97 %	100 %
r(CV) ²	50 %	67 %	27 %

Tableau 13.

Le pourcentage de discrimination de 98 % est très bon mais le nombre de variables utilisées pour atteindre ce pourcentage est assez élevé. Dans ce cas précis, on pourrait sélectionner seulement huit variables ce qui donnerait un pourcentage de discrimination correct de 90 % avec cette fois-ci un $r(CV)^2$ de 67 % (70% pour les composés inactifs et 64 % pour les actifs). Nous comparerons les résultats de ces deux analyses. Pour l'analyse utilisant 13 variables, tous les ligands inactifs sont classés inactifs sauf le **I29** qui a une activité expérimentale de 7.4. Tous les composés actifs sont bien classés. Etant donné l'activité de 7.4 du composé mal classé et la valeur discriminante choisie arbitrairement de 7.5 nous pouvons considérer que cette analyse donne un résultat optimum. En ce qui concerne l'analyse utilisant 8 variables, cinq composés sont mal classés, les **I19**, **I29**, **I34**, **I37** et **I49**. Le nombre de composés mal prédits est donc plus élevé tout en tenant compte que trois des produits (**I29-34-37**) ont des valeurs très proche de 7.5.

Les résultats des prédictions sont présentés dans le tableau 14.

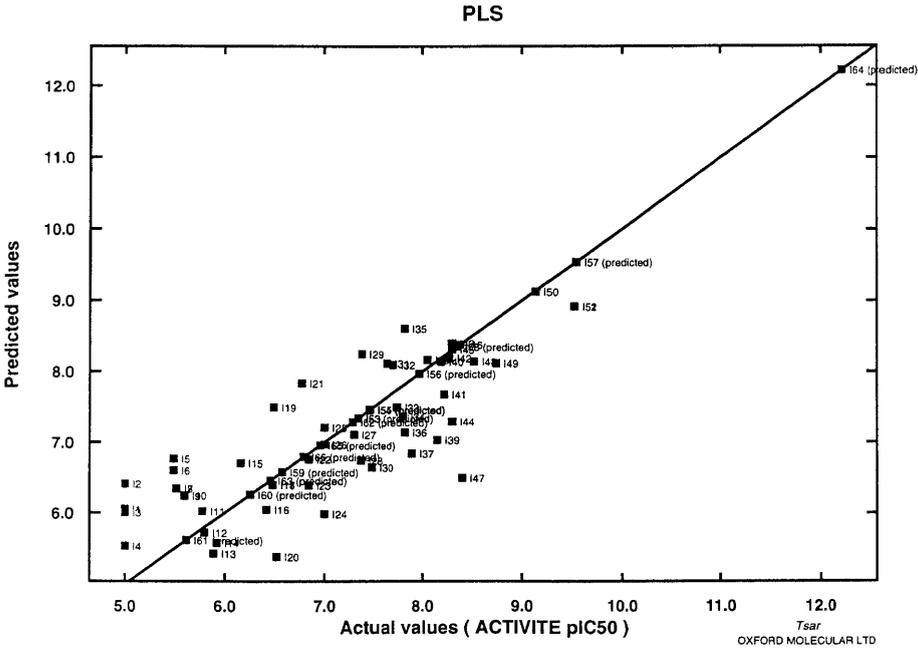
Composés	Analyse avec 13 variables		Analyse avec 8 variables	
	Prédiction	Expérimental	Prédiction	Expérimental
I53	Inactif	Actif	Inactif	Actif
I54	Inactif	Actif	Inactif	Actif
I55	Actif	Actif	Actif	Actif
I56	Actif	Actif	Actif	Actif
I57	Actif	Actif	Actif	Actif
I58	Actif	Inactif	Actif	Inactif
I59	Inactif	Inactif	Inactif	Inactif
I60	Inactif	Inactif	Inactif	Inactif
I61	Actif	Inactif	Inactif	Inactif
I62	Actif	Inactif	Inactif	Inactif
I63	Inactif	Inactif	Inactif	Inactif
I64	Actif	Actif	Actif	Actif
I65	Inactif	Inactif	Inactif	Inactif
I66	Inactif	Actif	Inactif	Actif

Tableau 14.

La première remarque qui ressort de cette analyse des prédictions est que l'analyse avec 8 variables donne de meilleures prédictions. En effet les composés **I61** et **I62** sont classés correctement. Ceci montre qu'il ne suffit pas *a priori* d'avoir une très bonne discrimination pour avoir un bon pouvoir prédictif. Dans ce cas de figure, l'analyse avec 8 variables offre un pouvoir de prédiction supérieur à l'analyse avec 13 variables (50% pour 67%).

La régression multiple donne des résultats très faibles. Le coefficient de régression r^2 est de 0.08 avec une variable. La valeur de s est 1.204. Ces données ne permettent pas d'exploiter cette régression multiple. Tous les composés à prédire le sont avec des valeurs comprises entre 7 et 7.4.

L'analyse Partial Least Square comme la régression multiple donne des résultats statistiques faibles. Nous avons retenu une équation composée de six vecteurs dont le coefficient de linéarité est de 0.66% avec un coefficient de prédiction de 0.41% (Graphe 6).



Graphe 6.

Les prédictions sont données dans le tableau 15.

Composés	Prédiction	Expérimental	Différence
I53	7.5	8.5	1
I54	7.5	9.7	2.2
I55	7.5	7.6	0.1
I56	8	7.8	0.2
I57	9.5	7.85	1.65
I58	8.3	5	3.3
I59	6.5	5	1.5
I60	6.3	6.5	0.2
I61	5.6	6.6	1
I62	7.3	5.7	1.6
I63	6.4	5.9	0.5
I64	12.2	8.9	3.3
I65	6.9	6.2	0.7
I66	6.8	8.9	2.1

Tableau 15.

Sur les 14 composés, 7 ligands sont prédits avec un écart de pIC_{50} inférieur ou égal à 1, 3 sont prédits avec un écart de pIC_{50} inférieur ou égal à 2 et 4 composés ont un écart de prédiction compris entre 2.1 et 3.3. Ce résultat montre que l'on peut avoir des prédictions intéressantes alors que la validité statistique du modèle, entre autres le coefficient de linéarité, est très faible ($r^2 = 0.66\%$). Là encore, un composé se trouve très éloigné de la gamme et l'écart entre la valeur expérimentale et la valeur de la prédiction est important (3.3).

Les indices de forme

L'analyse en composantes principales donne un pourcentage d'information sur les trois premiers axes de 70%. L'analyse des graphes 2D n'apporte aucune information concrète.

L'analyse discriminante sépare les composés du crible avec une probabilité de 96% en combinant 6 variables (Tableau 16).

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo I45	73 %	73 %	73 %
Carbo I18	79 %	83 %	75 %
Carbo I32	83 %	87 %	79 %
Carbo I13	87 %	90 %	84 %
Carbo I1	94 %	93 %	95 %
Carbo I12	96 %	97 %	95 %
$r(CV)^2$	65 %	70 %	60 %

Tableau 16.

Le pourcentage de discrimination de 96 % est très bon et le nombre de variables utilisées pour atteindre ce pourcentage est seulement de six. Pour cette analyse, tous les ligands inactifs sont classés inactifs sauf le **I28** qui a une activité expérimentale de 7.4. Tous les composés actifs sont bien classés sauf le **I49**. Etant donné l'activité de 7.4 du composé inactif mal classé et la valeur discriminante choisie arbitrairement de 7.5, nous pouvons considérer que cette analyse donne un résultat optimal.

Pour les composés à prédire les résultats sont présentés dans le tableau 17.

Composés	Prédiction	Expérimental
I53	Actif	Actif
I54	Actif	Actif
I55	Inactif	Actif
I56	Actif	Actif
I57	Actif	Actif
I58	Inactif	Inactif
I59	Actif	Inactif
I60	Actif	Inactif
I61	Actif	Inactif
I62	Actif	Inactif
I63	Actif	Inactif
I64	Actif	Actif
I65	Inactif	Inactif
I66	Actif	Actif

Tableau 17.

La première remarque qui ressort de cette analyse des prédictions est que la discrimination de ces composés par leur aspect stérique semble peu précise. En effet les composés **I59** à **I63** sont classés parmi les ligands actifs. Ceci est sûrement dû à leur forme se rapprochant de certains composés actifs. Ces descripteurs ne tiennent compte que de la forme des molécules. La différence d'activité entre ces composés n'est donc pas explicable seulement par l'aspect stérique des molécules même si le test statistique était bon.

L'analyse discriminante sur les trois premiers axes de l'ACP donne une discrimination de 79% avec les trois axes. Les prédictions sont plutôt mauvaises puisque sur les 14 composés à prédire seulement 5 sont bien classés.

La régression multiple donne des résultats non exploitables. Le coefficient de régression r^2 est de 0.31 avec deux variables.

L'analyse statistique Partial Least Square donne des résultats statistiques suivants. Nous avons retenu une équation composée de 8 vecteurs dont le coefficient de linéarité est de 0.76% avec un coefficient de prédiction de 0.34% (Graphe 7).

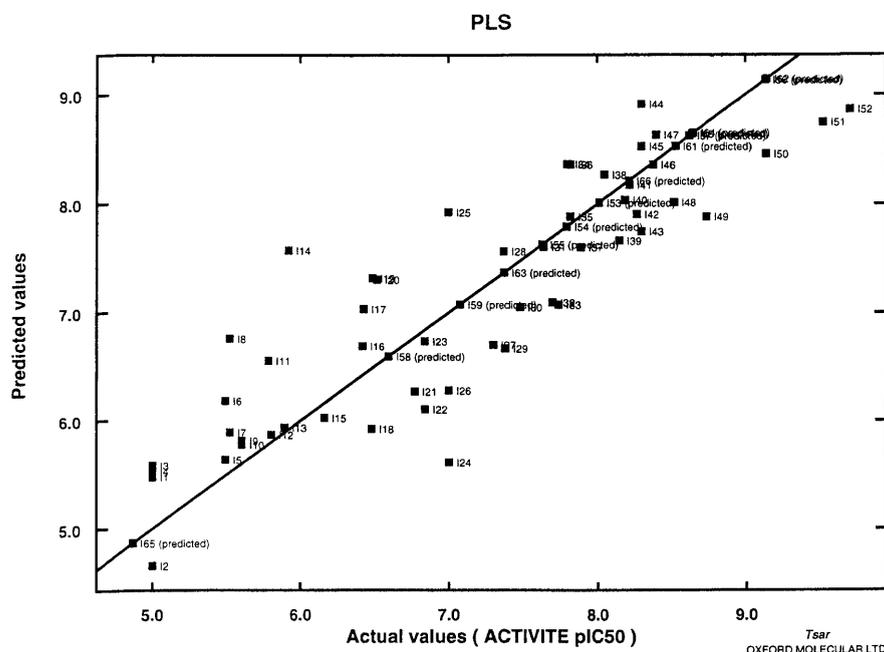
Les prédictions sont données dans le tableau 18.

Composés	Prédiction	Expérimental	Différence
I53	8.1	8.5	0.4
I54	7.8	9.7	1.9
I55	7.6	7.6	0
I56	9.1	7.8	1.3
I57	8.6	7.85	0.75
I58	6.5	5	1.5
I59	7	5	2
I60	8.5	6.5	2
I61	8.5	6.6	1.9
I62	9.1	5.7	3.6
I63	7.4	5.9	1.5
I64	8.7	8.9	0.2
I65	4.9	6.2	1.3
I66	8.3	8.9	0.6

Tableau 18.

Sur les 14 composés, 5 ligands sont prédits avec un écart de pIC_{50} inférieur ou égal à 1, 8 sont prédits avec un écart de pIC_{50} inférieur ou égal à 2 et 1 seul composé

est prédit avec un écart de prédiction de 3.6. Ce résultat montre que 13 composés sur 14 ont un écart de prédiction inférieur à deux unités de pIC_{50} avec seulement deux composés le **I61** et le **I60** qui passent du « cluster » inactif à actif.



Graphe 7.

Le champ électrostatique

Les résultats de l'analyse en composantes principales sont les suivants :

- Les trois premiers axes factoriels contiennent 45% de l'information des descripteurs.
- L'analyse des combinaisons axiales 2D et 3D n'apporte aucun renseignement sur la répartition des composés suivant ces repères 2D et 3D.

L'analyse discriminante (Tableau 19) sépare les composés du crible avec une probabilité de 85% en combinant 4 variables.

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo I35	73 %	73 %	73 %
Carbo I1	80 %	83 %	77 %
Carbo I2	83 %	87 %	79 %
Carbo I43	85 %	90 %	80 %
$r(CV)^2$	26 %	30 %	20 %

Tableau 19.

Le pourcentage de discrimination de 85 % est bon et le nombre de variables utilisées pour atteindre ce pourcentage est seulement de 4 variables. Toutefois le pouvoir prédictif est très faible (26%).

L'analyse de régression multiple et l'analyse Partial Least Square donnent des résultats de linéarité faibles ($r^2 = 0.25$ et 0.42) avec des pouvoirs descriptifs faibles ($r(CV)^2 = 0.17$ et 0.32) et des prédictions très éloignées de la valeur expérimentale.

Le champ lipophile

Les résultats de l'analyse en composantes principales sont les suivants :

- Les trois premiers axes factoriels contiennent 25% de l'information des descripteurs. L'analyse des combinaisons axiales 2D et 3D n'apporte aucun renseignement sur la répartition des composés suivant ces repères 2D et 3D.

L'analyse discriminante prévoit un pourcentage de discrimination de 75% avec 15 variables. Ce pouvoir discriminant est très faible et ne permet pas une étude approfondie de la prédiction.

L'analyse de régression multiple et l'analyse Partial Least Square donnent des résultats de linéarité faibles ($r^2 = 0.18$ et 0.57).

Conclusions

Dans cette étude, il semblait intéressant de constater si un échantillon, aussi diversifié au niveau structural, pouvait donner des résultats statistiques et prédictifs intéressants.

Les analyses statistiques effectuées avec les descripteurs classiques apportent des résultats satisfaisants pour l'analyse discriminante (discrimination de 96% avec $r(\text{CV})^2 = 80\%$), la régression multiple ($r^2 = 0.70$ et $r(\text{CV})^2 = 0.60$) et l'analyse PLS ($r^2 = 0.81$ et $r(\text{CV})^2 = 0.55$). En ce qui concerne, les analyses statistiques utilisant les indices de similarité moléculaire, les résultats intéressants sont issus :

- de l'analyse discriminante avec l'étude des potentiels de charge qui donne un coefficient de discrimination de 90% et un $r(\text{CV})^2$ de 65%,
- de l'étude PLS utilisant les potentiels de charge ($r^2 = 0.61$ et $r(\text{CV})^2 = 0.65$),
- de l'analyse discriminante avec les indices de potentiels de lipophilie (discrimination de 90% et $r(\text{CV})^2 = 67\%$),
- de l'étude PLS utilisant les potentiels de lipophilie ($r^2 = 0.66$ et $r(\text{CV})^2 = 0.41$),
- de l'analyse discriminante avec les indices de forme (discrimination de 96% et $r(\text{CV})^2 = 65\%$),
- de l'étude PLS avec les indices de forme où le coefficient de régression linéaire est de 0.76 et un $r(\text{CV})^2$ de 0.34.

L'ensemble des autres modèles présente des résultats statistiques non exploitables de part les critères de validation.

En ce qui concerne les prédictions des composés **I53** à **I66**, les résultats pour l'ensemble des analyses sont regroupés dans le tableau 20. Nous séparerons les résultats qualitatifs issus de l'analyse discriminante, des résultats quantitatifs donnés par la régression multiple et le Partial Least Square. Cette vue d'ensemble devrait permettre de synthétiser et de valider nos résultats. Pour cela nous effectuerons des valeurs moyennes quantitatives et qualitatives. Pour les quatre analyses discriminantes, un ligand prédit 3 fois actif et une fois inactif sera considéré

comme prédit actif en moyenne. Pour les autres analyses (RM et PLS), nous calculerons la valeur moyenne, et nous considérerons comme bon résultat, toutes les prédictions dont l'écart entre la valeur expérimentale et la valeur moyenne est inférieur à une unité de pIC₅₀.

Cps	Valeur Qualitative						Valeur Quantitative						
	AD DC	AD PE	AD PL	AD S	Moy.	Exp.	RM DC	PLS DC	PLS PE	PLS PL	PLS S	Moy.	Exp.
I53	A	A	I	A	A	A	9.5	10.9	8.0	7.5	8.1	8.8	8.5
I54	A	A	I	A	A	A	10.7	8.5	7.7	7.5	7.8	8.4	9.7
I55	A	A	A	I	A	A	8.6	8.7	9.4	7.5	7.6	8.4	7.6
I56	I	A	A	A	A	A	9.4	5.0	5.5	8	9.1	7.4	7.8
I57	I	A	A	A	A	A	7.45	8.1	7.9	9.5	8.6	8.3	7.85
I58	I	I	A	I	I	I	5.2	3.0	8.4	8.3	6.5	6.3	5.0
I59	I	I	I	A	I	I	6.2	3.0	6.1	6.5	7.0	5.8	5.0
I60	I	I	I	A	I	I	9.9	8.25	8.2	6.3	8.5	8.2	6.5
I61	I	I	I	A	I	I	8.6	8.4	8.6	5.6	8.5	7.9	6.6
I62	I	I	I	A	I	I	8.3	8.7	7.7	7.3	9.1	8.2	5.7
I63	I	A	I	A	A/I	I	7.4	8.4	7.5	6.4	7.4	7.4	5.9
I64	A	I	A	A	A	A	10.5	7.5	6.2	12.2	8.7	9.0	8.9
I65	I	I	I	I	I	I	6.4	8.2	6.2	6.9	4.9	6.5	6.2
I66	A	A	I	A	A	A	5.2	8.6	9.4	6.8	8.3	7.7	8.9

Tableau 20. Cps = composés ; AD = Analyse Discriminante ; RM = Régression Multiple ; PLS = Partial Least Square ; A = Actif ; I = Inactif ; DC = Descripteur Classique ; PE = Potentiel Electrostatique ; PL = Potentiel de Lipophilie ; S = Stérique ; Moy. = Moyenne ; Exp. = Expérimentale.

Pour les prédictions qualitatives moyennes, presque tous les composés sont prédits correctement. Un seul composé se trouve prédit deux fois actif et deux fois inactif et ce résultat ne nous autorise donc pas à conclure. Nous pouvons donc affirmer que l'ensemble de nos modèles discriminants permet une prédiction correcte à 93% (ou 100% si on ne conclue pas pour le **I63**). Ces modèles montrent l'intérêt d'une telle approche QSAR et devraient donc pouvoir servir à la prédiction

de nouveaux composés. Pour les méthodes dites quantitatives, les résultats sont moins probants. Toutefois, 50% des composés sont prédits avec un écart inférieur à 1 unité de pIC_{50} et 93% avec un écart inférieur à 2 unités. Et si l'on prend en compte le seuil de 7.5, seulement 4 ligands sur 14 sont prédits dans la mauvaise classe (active, inactive).

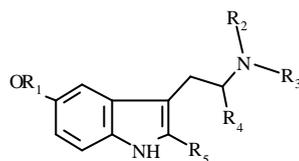
En conclusion, nous supposons que la prédiction d'un nouveau composé de cette famille indole, en conjuguant l'ensemble de ces analyses, pourrait être envisagée avec un faible risque de mauvaise prédiction de son activité biologique 5-HT_{1A}. Ceci est valable que si l'on tient compte du fait qu'une approche QSAR n'apporte aucune information pour la création des nouveaux composés mais qu'elle donne seulement une information sur la prédiction de l'activité. De plus, les composés à prédire doivent être relativement proches de ceux de l'échantillon de départ.

On peut essayer d'analyser l'intérêt d'une telle approche dans une situation où les composés testés par le modèle QSAR, ne seraient pas encore synthétisés. En se basant sur les résultats qualitatifs précédents, on aurait synthétisé que les 8 ligands prédits actifs ou douteux au lieu des 14. Le gain aurait été de plus de 40% en temps et coût de synthèse. Sur ces composés 7 sont réellement actifs (moyenne des 8 composés : 8.2). On n'aurait perdu aucun composé actif.

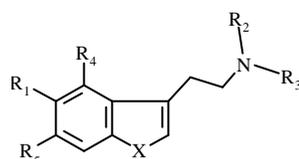
En se basant sur les résultats quantitatifs, 9 composés sur 14 auraient été synthétisés. Le gain aurait été de plus de 35% en temps et coût de synthèse. Sur ces 9 composés 6 sont réellement actifs et on aurait perdu un seul composé actif le **I56**.

Ces modèles QSAR pour la famille des indoles sont donc fiables et permettent de réduire le nombre de composés à synthétiser sans trop de mauvaises surprises. Ils ne sont pas utilisables sans l'imagination créative du chimiste, mais par contre ils optimisent son travail en donnant une information rapide et plutôt fiable de l'activité potentielle des composés.

ANNEXE 1

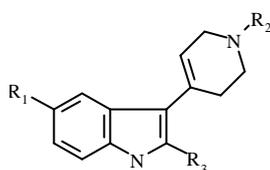


Composés	R ₁	R ₂	R ₃	R ₄	R ₅	pIC ₅₀ (M)
I13	Me	H	H	H	Me	5.9
I14	H	H	H	H	Me	5.9
I28	Me	Me	<i>i</i> -Pr	H	H	7.4
I29	H	H	H	Me	H	7.4
I36	H			H	H	7.8
I38	Me	H	H	H	H	8.05
I39	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	8.15
I40	Me	Me	Me	H	H	8.2
I43	H	Me	H	H	H	8.3
I46	H	Me	Me	H	H	8.5
I47	Me	<i>n</i> -Pr	<i>n</i> -Pr	H	H	8.4
I48	H	H	H	H	H	8.5

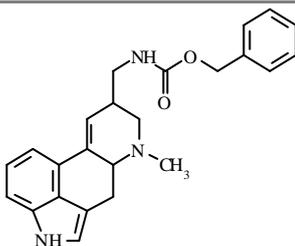


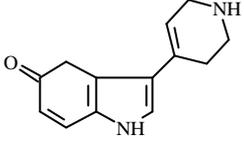
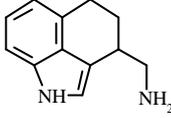
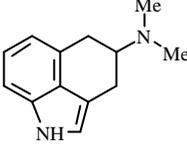
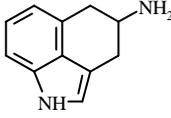
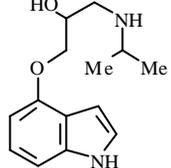
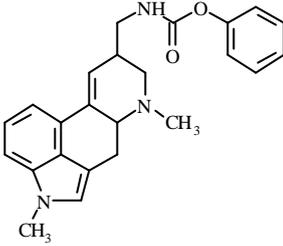
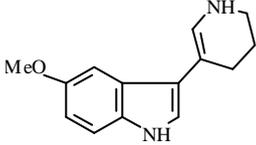
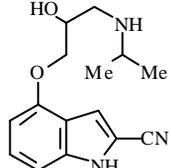
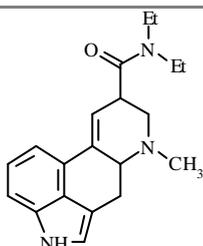
Composés	R ₁	R ₂	R ₃	R ₄	R ₅	X	pIC ₅₀ (M)
I1	H			H	H	NH	4
I2	H	Me	<i>i</i> -Pr	H	OMe	NH	4.5
I3	OMe	Me	<i>i</i> -Pr	H	OMe	NH	5
I5	H	<i>i</i> -Bu	<i>i</i> -Bu	OH	H	NH	5.5
I6	H	<i>sec</i> -Bu	<i>sec</i> -Bu	OH	H	NH	5.5
I9	H	Me	<i>i</i> -Pr	OH	H	NH	5.6
I10	H	Me	<i>i</i> -Pr	H	H	O	5.8
I12	H	H	H	H	OH	NH	5.8
I15	H	Et	Et	OH	H	NH	6.2

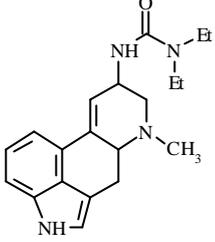
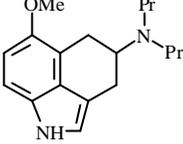
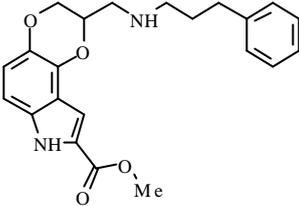
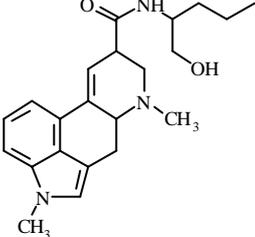
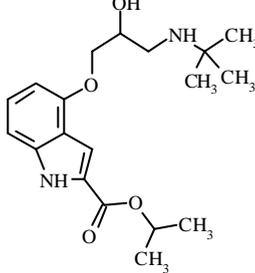
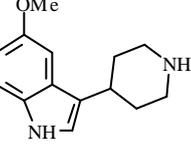
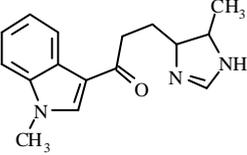
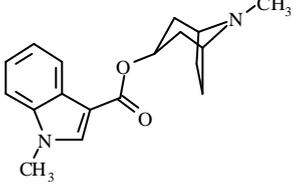
I16	H	Me	<i>i</i> -Pr	H	H	NH	6.4
I20	H			H	H	NH	6.5
I21	H	H	H	H	H	NH	8.8
I22	Me	Me	<i>i</i> -Pr	H	H	NH	6.8
I24	H	Me	Me	H	H	NH	7
I25	H	H	H	OH	H	NH	7.1
I26	H	Me	Me	OH	H	NH	7.2
I51	CONH ₂	<i>n</i> -Pr	<i>n</i> -Pr	H	H	NH	9.5
I52	CONH ₂	H	H	H	H	NH	9.7

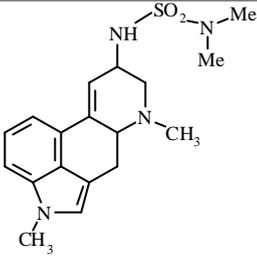
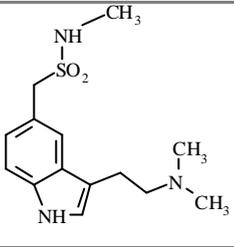
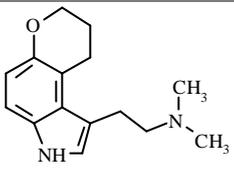
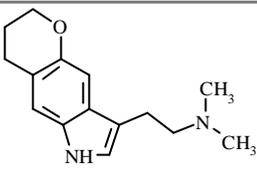
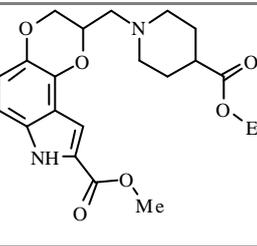
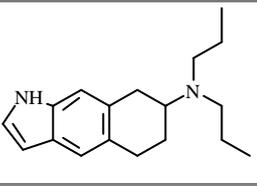
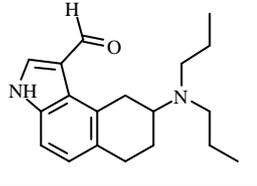


Composés	R ₁	R ₂	R ₃	pIC ₅₀ (M)
I4	H	Me	Me	5
I11	OMe	Me	Me	5.8
I23	H	Me	H	6.9
I27	H	H	H	7.3
I30	F	H	H	7.5
I31	OMe	Me	H	7.6
I33	NO ₂	H	H	7.7
I34	CN	H	H	7.8
I35	OH	H	H	7.8
I37	Cl	H	H	7.9
I45	OMe	H	H	8.3

Composés	Structures	pIC ₅₀ (M)
I7		5.5

I8		5.5
I17		6.4
I18		6.5
I19		6.5
I32		7.7
I41		8.2
I42		8.3
I44		8.3
I49		8.7

I50		9.1
I53		(8.5)
I54		(9.7)
I55		(7.6)
I56		(7.8)
I57		(7.85)
I58		(5)
I59		(5)

I60		(6.5)
I61		(6.6)
I62		(5.7)
I63		(5.9)
I64		(8.9)
I65		(6.2)
I66		(8.9)

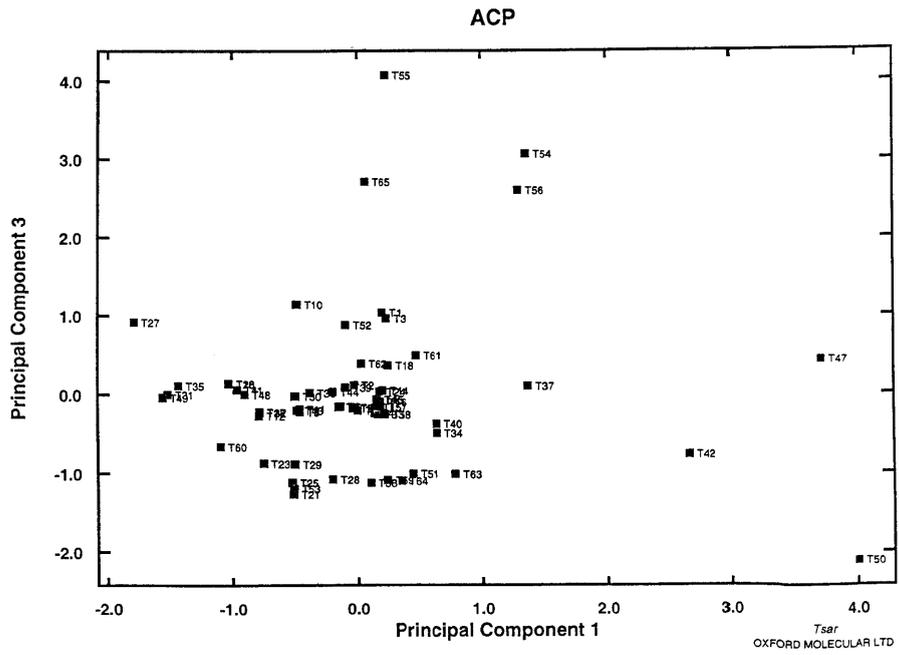
II. FAMILLE DES TETRALINES

La composition de cet échantillon (annexe 2) est de 65 composés. Cet échantillon présente une série de ligands avec des activités biologiques comprises entre 4 et 9.7 unités de pIC_{50} . Cet ensemble est séparé en trois sous-groupes, un sous-groupe de 32 ligands inactifs (**T1** à **T32**) ayant un $pIC_{50} < 7.5$, un de 21 composés actifs (**T33** à **T53**) avec un $pIC_{50} \geq 7.5$ et de 12 composés à prédire *a posteriori* (**T54** à **T65**). Dans cette étude, l'activité 5-HT_{1A} de ces 12 derniers composés pris au hasard des publications récentes sera évaluée pour vérifier le pouvoir prédictif des modèles proposés. Cet échantillon a fait l'objet d'une étude des descripteurs moléculaires classiques et des indices de similarité moléculaire. Comme pour la première famille indole, différentes méthodes d'analyses statistiques ont été testées. Ces méthodes sont l'analyse en composantes principales, l'analyse discriminante, la régression multiple et le partial least square. La présence de nombreux stéréoisomères présentant une différence d'activité biologique semble intéressante pour tester le pouvoir de différenciation des descripteurs utilisés. Tous les composés pris en compte dans cette étude QSAR sont extraits de publications diverses^{1,2,11,12-33}.

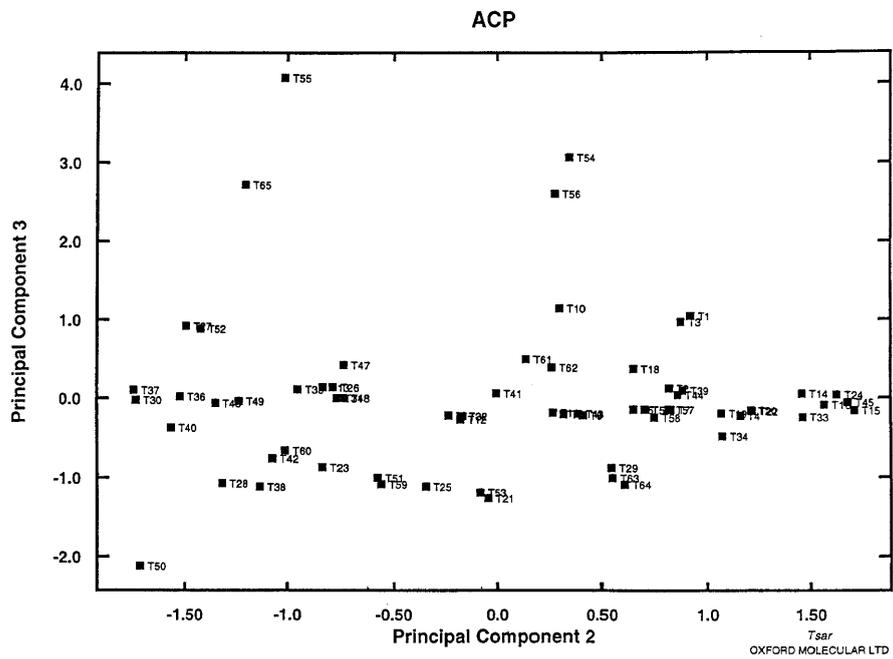
Etude des descripteurs classiques

Les descripteurs utilisés sont ceux proposés par le logiciel TSAR ; ils comprennent les indices de forme, de connectivité et de flexibilité, certaines propriétés moléculaires (moment d'inertie, surface moléculaire, Log P, etc...), les indices topologiques et certains autocorrélogrammes 2D et 3D. Pour les autocorrélogrammes, le nombre de vecteurs maximum utilisé est de 12. Pour les indices de Kier le nombre maximum de cycles utilisé est 7 et le pas maximum est de 12. Tous ces descripteurs forment un crible de 295 variables.

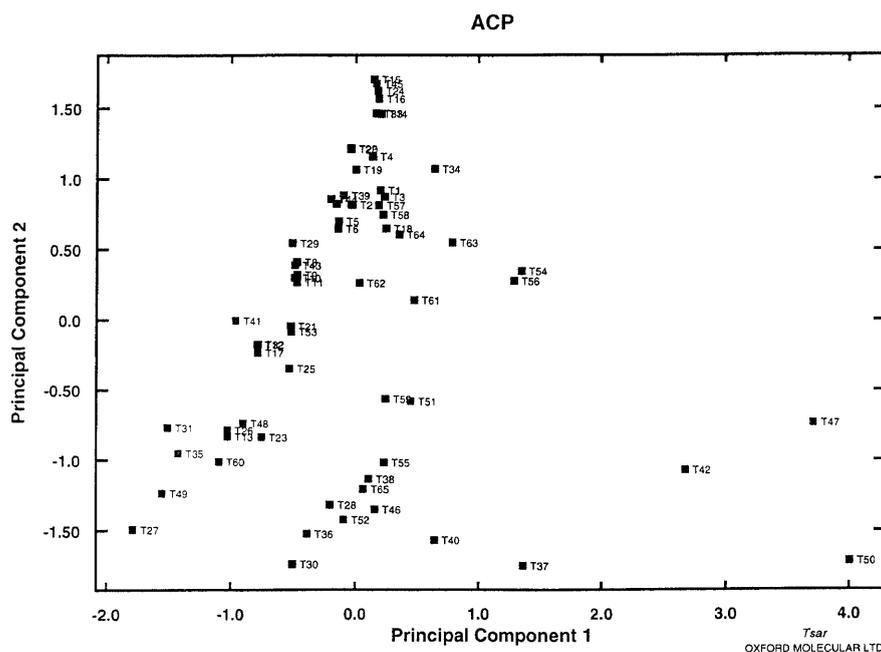
La première analyse envisagée, est une analyse en composantes principales. L'objectif est de pouvoir étudier notre ensemble de composés dans un espace à deux ou trois dimensions afin de repérer d'éventuels « clusters » de molécules. Les trois premiers axes factoriels contiennent 67% de l'information initiale. Les représentations 2D (Graphe 8, 9, 10) montrent que l'axe N°1, qui est un axe surtout de taille, discrimine les composés **T37**, **T42**, **T47** et **T50**. Ces quatre composés ont un volume et une inertie importants. L'autre axe qui apporte une information intéressante est l'axe N°3. En effet celui-ci permet de différencier les composés à prédire **T54**, **T55**, **T56** qui comportent des fonctions triflates et le composé **T65**. Il sera intéressant par la suite de voir si cet ensemble (« training set ») permet de classer ces composés (« test set ») correctement sachant que ceux-ci contiennent des fonctions moléculaires triflates absentes du « training set ».



Graphe 8.



Graphe 9.



Graphe 10.

Une étude d'analyse discriminante effectuée sur l'ensemble de ces descripteurs classique donne les résultats suivants présentés dans le tableau 21.

Variables	Total	Inactifs bien classés	Actifs bien classés
2D bin 11 (Log P)	77 %	97 %	57 %
2D bin 4 (électronégativité)	81 %	91 %	71 %
2D bin 5 (charge Q positive)	81 %	94 %	68 %
Kappa 3	86 %	91 %	81 %
3D bin 7 (Charge Q)	89 %	94 %	84 %
3D bin 5 (Log P)	92 %	94 %	90 %
$r(CV)^2$	66 %	75 %	52 %

Tableau 21.

Le pourcentage de discrimination est de 92 % avec une validité de 66%. Parmi les composés, deux ligands inactifs sont classés actifs (**T27** et **T30**). Ces composés sont très proche de la valeur limite (7.3 et 7.35) que nous nous sommes fixés (7.5) pour séparer les composés actifs et inactifs. En ce qui concerne la classe des

composés actifs, le **T43**, le **T45** et le **T53** sont mal classés. Le **T43** et le **T53** diffèrent seulement d'un CH₂ et ont la même configuration *trans* au niveau des carbones 1 et 2. Ces composés sont classés parmi les inactifs ; notre modèle ne peut donc pas différencier les diastéréoisomères **T43** et **T11**, **T53** et **T21**. Le composé **T45** quant à lui est le seul composé actif du crible avec les **T43** et **T53** à avoir une substitution sur le carbone 1 tous les autres composés substitués en position 1 ont une activité biologique faible. Les trois composés actifs, qui sont classés parmi inactifs, ont tous une substitution en position 1.

Pour les composés à prédire (Tableau 22), les résultats sont : 7 composés actifs sont bien prédits et 1 seul est mal prédit. Deux composés inactifs sont bien prédits et 2 sont mal classés.

Composés	Prédiction	Expérimental
T54	Actif	Actif
T55	Actif	Actif
T56	Inactif	Inactif
T57	Actif	Inactif
T58	Actif	Actif
T59	Inactif	Inactif
T60	Inactif	Actif
T61	Actif	Actif
T62	Actif	Actif
T63	Actif	Actif
T64	Actif	Actif
T65	Actif	Inactif

Tableau 22.

Les descripteurs sélectionnés permettent de différencier des isomères de position comme les composés **T54** et **T56**. Par contre ils ne différencient pas les diastéréoisomères **T57** et **T58**.

On remarque aussi que les composés **T54**, **T55** et **T56** sont bien classés alors que ceux-ci contiennent des fonctions triflates qui ne sont pas représentées dans le « training set ». Le composé **T60**, qui est mal prédit par notre modèle, est substitué en position 1.

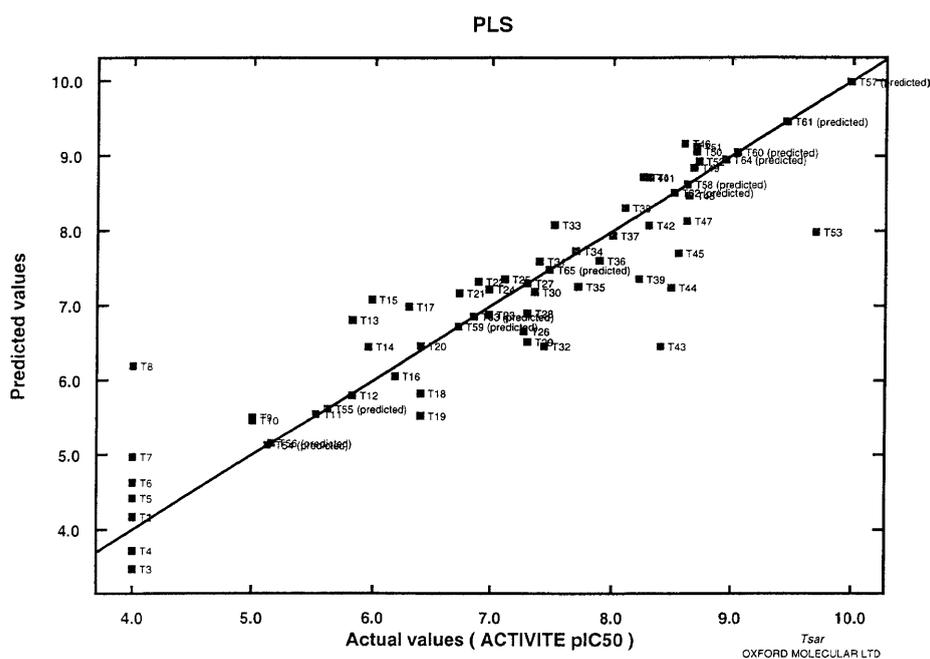
La régression multiple donne un coefficient de régression r^2 de 0.63. Une équation est proposée avec six variables. La valeur de s est 1.031 et le $r(\text{CV})^2$ est de 0.13. Toutes ces données ne permettent pas d'exploiter cette régression multiple.

L'analyse Partial Least Square (Graphe 11) donne des résultats assez intéressants (Tableau 23).

Vecteurs	1	2	3	4	5	6	7	8	9	10
Signification statistique	0.750	0.853	0.909	0.968	1.12	1.38	1.61	1.78	1.95	2.01
PRESS	54.50	42.96	39.09	38.54	35.20	30.21	29.56	37.25	41.20	55.61
$r(\text{CV})^2$	0.121	0.282	0.325	0.421	0.451	0.562	0.658	0.452	0.298	0.178
r^2	0.239	0.412	0.590	0.678	0.717	0.761	0.808	0.847	0.873	0.901

Tableau 23.

Le nombre de vecteurs sélectionnés est 7. Ce vecteur correspond en effet à la première valeur de PRESS faible (29.56), et avec la combinaison de ces vecteurs le coefficient de régression linéaire est de 0.81. Ce résultat pourrait être optimisé en diminuant le nombre de composés de l'échantillon ayant une activité de $\text{pIC}_{50} = 4$. En effet, notre échantillon comprend 7 composés dont l'activité est annoncée à 4 unités de pIC_{50} . Ces ligands sont tous très inactifs et les résultats pharmacologiques donnent toujours un $\text{IC}_{50} < 10^{-4}$. Nous avons donc attribué une valeur arbitraire de 4 à ces composés mais ceci introduit inévitablement un biais.



Grphe 11.

Pour les composés à prédire, les résultats (Tableau 24) sont assez encourageants puisque 9 composés sur 12 sont prédits avec un écart d'unité de pIC₅₀ inférieur ou égal à 1 unité de pIC₅₀. La prédiction moyenne pour ces 9 composés est de 0.7 unité de pIC₅₀. Parmi les trois molécules dont la prédiction est mauvaise, les **T54** et **T55** comportent des fonctions triflates qui ne sont pas présentes dans le « training set » et le **T57** est prédit hors échelle. Ce résultat montre les limites d'une étude QSAR. Lorsque l'on travaille sur des familles homogènes, il semble difficile de prédire avec justesse l'activité de composés présentant des différences fonctionnelles importantes. Toutefois, ces produits **T54** et **T55** avaient été prédits actif qualitativement avec l'analyse discriminante. A noter que notre modèle ne différencie pas le composé **T56** (isomère de position) des ligands **T54-55**.

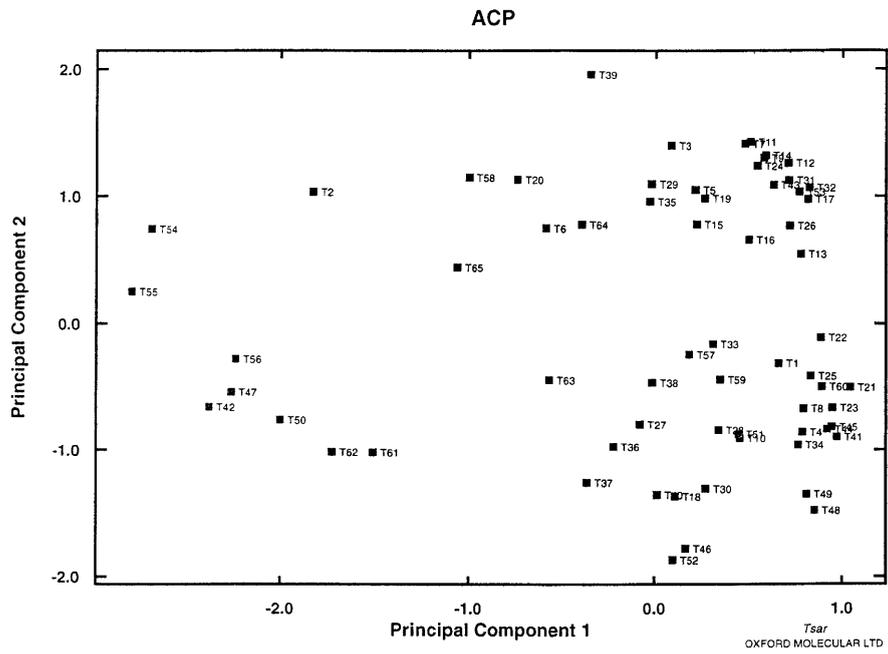
Composés	Prédiction	Expérimental	Différence
T54	5.1	9.1	4
T55	5.6	8.6	3
T56	5.2	6	0.8
T57	9.9	6.9	3
T58	8.6	8.3	0.3
T59	6.7	6.3	0.4
T60	9.0	8.2	0.8
T61	9.4	8.6	0.8
T62	8.5	7.4	0.9
T63	6.9	7.9	1
T64	8.9	8.2	0.7
T65	7.5	6.9	0.6

Tableau 24.

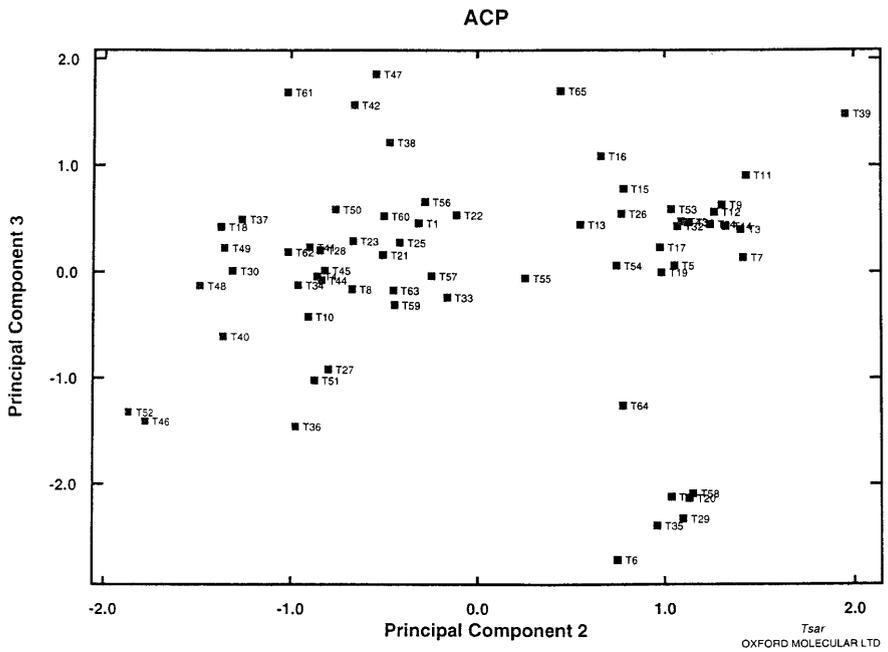
Etude des indices de similarité moléculaire

Potentiel de charge électrostatique

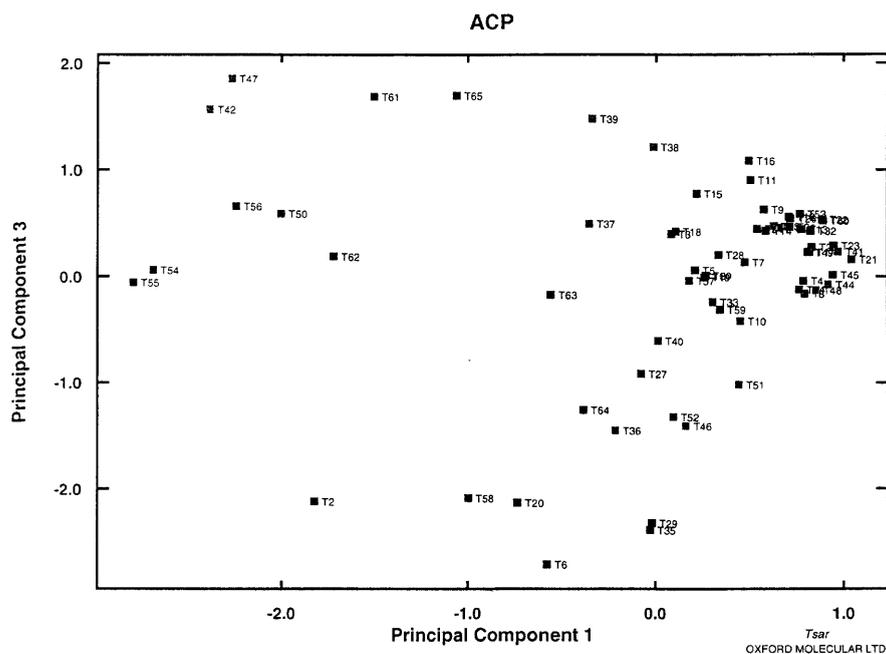
L'analyse en composantes principales effectuée sur cette matrice prévoit un pourcentage d'information sur les trois premiers axes de 80%. L'analyse des graphes 2D (Graphes 12, 13, 14) et 3D montre que l'axe 1 représenté par 47 descripteurs sur 53, sépare les composés ayant les fonctions triflates (**T54**, **T55**, **T56**), les composés à forte inertie (**T42**, **T50**, **T47**), les composés tricycliques contenant une fonction amide (**T61**, **T62**) des autres molécules.



Graphe 12.



Graphe 13.



Grphe 14.

L'étude d'analyse discriminante sépare la totalité des composés du crible avec un pourcentage de 88%. Pour la classe des composés inactifs, 100% sont bien classés et pour les ligands actifs 76% (Tableau 25).

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo T20	74 %	78 %	70 %
Carbo T51	81 %	78 %	84 %
Carbo T23	84 %	94 %	74 %
Carbo T44	88 %	100 %	76 %
$r(CV)^2$	68 %	81 %	48 %

Tableau 25.

Le pourcentage de discrimination est donc de 88% avec une validité statistique de 68%. Parmi les composés, tous les ligands inactifs sont bien classés. En ce qui concerne les composés actifs, six composés les **T34**, **T35**, **T39**, **T43**, **T44** et le **T53** sont mal classés. Les ligands **T34** et **T35** ($pIC_{50} = 7.7$) sont des composés proches de la valeur discriminatoire des deux classes. Il est donc difficile d'incriminer le

résultat prédictif donné. Le produit **T43** et le composé **T53** sont une fois encore classés parmi les ligands inactifs. Ces composés qui diffèrent seulement d'un CH₂ et ont la même configuration au niveau des carbones 1 et 2, ne sont pas différenciés des ligands inactifs **T11** et **T21** qui sont par ailleurs des diastéréoisomères.

Pour la prédiction des composés **T54** à **T65**, les résultats sont donnés dans le tableau 26. Six composés actifs sont bien prédits et 2 sont mal prédits. Un seul composé inactif est bien prédit et 3 sont mal classés.

Composés	Prédiction	Expérimental
T54	Actif	Actif
T55	Actif	Actif
T56	Actif	Inactif
T57	Actif	Inactif
T58	Inactif	Actif
T59	Inactif	Inactif
T60	Actif	Actif
T61	Actif	Actif
T62	Actif	Actif
T63	Actif	Actif
T64	Inactif	Actif
T65	Actif	Inactif

Tableau 26.

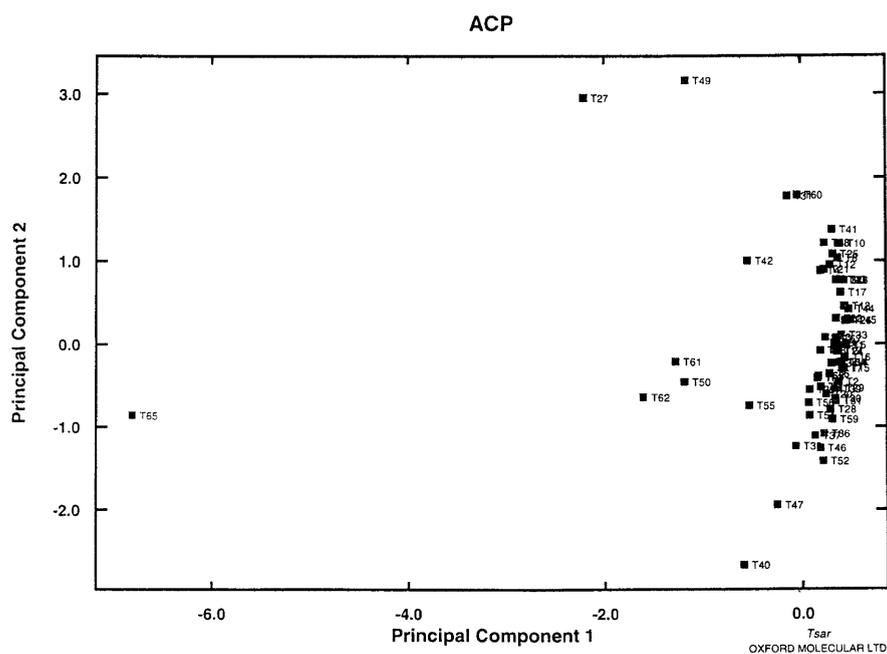
Cette approche butte encore sur le problème des isomères de position (**T54** et **T56**) et des diastéréoisomères (**T57** et **T58**).

L'analyse discriminante sur les trois premiers axes de l'ACP donne une discrimination de 75% sans pour autant donner de meilleures prédictions.

Les résultats obtenus avec les méthodes de régression multiple et de PLS ne seront pas exploités ; les coefficients de linéarité et les pouvoirs prédictifs étant faibles.

Potentiel de lipophilie

L'analyse en composantes principales prévoit un pourcentage d'information sur les trois premiers axes de 94%. L'analyse des graphes 2D (Graphe 15) montre que l'axe 1 sépare très nettement le composé **T65** des autres composés. L'aspect lipophile de ce composé semble donc assez différent des autres molécules. Nous verrons par la suite si ce composé est bien prédit par nos modèles.



Grphe 15.

L'analyse discriminante sépare les composés du crible à 83%. Le pourcentage de composés inactifs bien classés est de 94% et 72% des composés actifs sont bien classés (Tableau 27).

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo T21	81 %	94 %	68 %
Carbo T25	83 %	94 %	72 %
$r(CV)^2$	77 %	91 %	57 %

Tableau 27.

Le pourcentage de discrimination est acceptable de même que la validité statistique de 77%. Les variables discriminantes (Carbo T21 et T25) sont des indices de similarités de composés inactifs proches du seuil des classes. Parmi les composés, tous les ligands inactifs sont classés inactifs sauf le **T3** et le **T27**. En ce qui concerne les composés actifs, sept composés le **T33**, **T39**, **T41**, **T44**, **T45**, **T48** et le **T51** sont mal classés. Le **T33** ($pIC_{50} = 7.5$) est un composé proche de la valeur discriminatoire des deux classes. On peut donc comprendre le mauvais résultat prédictif donné par l'analyse.

Pour les composés à prédire, les résultats sont : 5 composés actifs sont bien prédits et 3 sont prédits inactifs. Deux composés inactifs sont bien prédits et 2 sont classés actifs (Tableau 28).

Composés	Prédiction	Expérimental
T54	Actif	Actif
T55	Actif	Actif
T56	Actif	Inactif
T57	Inactif	Inactif
T58	Inactif	Actif
T59	Inactif	Inactif
T60	Inactif	Actif
T61	Actif	Actif
T62	Actif	Actif
T63	Actif	Actif
T64	Inactif	Actif
T65	Actif	Inactif

Tableau 28.

Les descripteurs sélectionnés ne permettent pas de différencier des isomères de position comme les composés **T54** et **T56**. Ils ne différencient pas non plus les diastéréoisomères **T57** et **T58**.

L'analyse discriminante sur les trois premiers axes de l'ACP donne une discrimination de 79% et les résultats de prédiction ne sont pas meilleurs que les précédents.

L'analyse Partial Least Square comme la régression multiple donne des résultats statistiques inexploitable.

Les indices de forme

Une analyse en composantes principales effectuée sur l'ensemble des descripteurs de forme montre que la quantité d'information retenue sur trois axes est

de 66%. L'analyse des combinaisons axiales 2D et 3D n'apportent aucune information.

L'analyse discriminante discrimine le crible à 85% avec cinq variables (Tableau 29).

Variables	Total	Inactifs bien classés	Actifs bien classés
Carbo T53	74 %	72 %	76 %
Carbo T32	79 %	84 %	72 %
Carbo T14	81 %	88 %	74 %
Carbo T2	83 %	94 %	72 %
Carbo T31	85 %	94 %	76 %
$r(\text{CV})^2$	62 %	69 %	52 %

Tableau 29.

Parmi l'ensemble des composés, tous les ligands inactifs sont classés inactifs sauf le **T5** et le **T30**. Pour les composés actifs, six composés le **T39**, **T41**, **T43**, **T44**, **T45** et le **T53** sont mal classés. Le **T30** ($\text{pIC}_{50} = 7.35$) est un composé proche de la valeur discriminatoire des deux classes. Le **T43** et le **T53** diffèrent seulement d'un CH_2 et ont la même configuration au niveau des carbones 1 et 2. Ces deux composés sont classés inactifs comme les ligands **T11** et **T21**.

Pour les composés à prédire, les résultats sont : 5 composés actifs sont bien prédits et 3 sont prédits inactifs. Deux composés inactifs sont bien prédits et 2 sont classés actifs (Tableau 30).

Composés	Prédiction	Expérimental
T54	Inactif	Actif
T55	Actif	Actif
T56	Actif	Inactif
T57	Inactif	Inactif
T58	Actif	Actif
T59	Inactif	Inactif
T60	Inactif	Actif
T61	Inactif	Actif
T62	Actif	Actif
T63	Actif	Actif
T64	Actif	Actif
T65	Actif	Inactif

Tableau 30.

Les descripteurs sélectionnés ne permettent pas de différencier des isomères de position comme les composés **T54** et **T56**. Par contre ils différencient les diastéréoisomères **T57** et **T58**.

La méthode de régression multiple donne des résultats inexploitable. Le coefficient de régression r^2 est de 0.21 avec deux variables. La valeur de s est 1.450 et le $r(CV)^2$ est de -0.24. Toutes ces données ne permettent pas d'exploiter cette régression multiple.

L'analyse Partial Least Square comme la régression multiple donne des résultats statistiques très peu fiables et des prédictions très mauvaises des composés à prédire.

Le champ électrostatique

Les résultats de l'analyse en composantes principales sont les suivants :

- Les trois premiers axes factoriels contiennent 11% de l'information des descripteurs.
- L'analyse des combinaisons axiales 2D et 3D n'apporte aucun renseignement sur la répartition des composés suivant ces repères 2D et 3D.

L'analyse discriminante prévoit un pourcentage de discrimination de 65% avec deux variables. Ce pouvoir discriminant est très faible et ne permet pas une étude approfondie de la prédiction d'activité.

L'analyse de régression multiple et l'analyse Partial Least Square donnent des résultats de linéarité faibles ($r^2 = 0.25$ et 0.32) et des prédictions très éloignées de la valeur expérimentale.

Le champ lipophile

Les résultats de l'analyse en composantes principales sont les suivants ; les trois premiers axes factoriels contiennent 25% de l'information des descripteurs et l'analyse des combinaisons axiales 2D et 3D n'apporte aucun renseignement sur la répartition des composés suivant ces repères 2D et 3D.

L'analyse discriminante prévoit un pourcentage de discrimination de 45% avec deux variables. Ce pouvoir discriminant est très faible et ne permet pas une étude approfondie de la prédiction.

L'analyse de régression multiple et l'analyse Partial Least Square donnent des résultats de linéarité faibles ($r^2 = 0.28$ et 0.47) et des prédictions très éloignées de la valeur expérimentale.

Conclusions

Dans cette étude, il semblait intéressant de constater si avec cet échantillon de composé de type tétraline, l'on obtenait des résultats statistiques corrects et si l'on pouvait correctement prédire l'activité d'isomères de position ou de diastéréoisomères.

Les analyses statistiques effectuées avec les descripteurs classiques apportent des résultats satisfaisants pour l'analyse discriminante (discrimination de 92% avec $r(\text{CV})^2 = 66\%$) et l'analyse PLS ($r^2 = 0.81$ et $r(\text{CV})^2 = 0.66$). En ce qui concerne, les analyses statistiques utilisant les indices de similarité moléculaire, les résultats intéressants sont issus :

- de l'analyse discriminante avec l'étude des potentiels de charge qui donne un coefficient de discrimination de 88% et un $r(\text{CV})^2$ de 68%,
- de l'analyse discriminante avec les indices de potentiels de lipophilie (discrimination de 83% et $r(\text{CV})^2 = 77\%$),
- de l'analyse discriminante avec les indices de forme (discrimination de 85% et $r(\text{CV})^2 = 62\%$),

L'ensemble des autres modèles présente des résultats statistiques non exploitables de part les critères de validation.

En ce qui concerne les prédictions des composés **T54** à **T65**, les résultats pour l'ensemble des analyses sont regroupés dans le tableau 31. Nous utiliserons le même principe de calcul des valeurs moyennes que dans le cas des indoles.

Cps	Valeur Qualitative						Valeur Quantitative	
	AD DC	AD PE	AD PL	AD S	Moy.	Exp.	PLS DC	Exp.
T54	A	A	A	I	A	A	5.1	9.1
T55	A	A	A	A	A	A	5.6	8.6
T56	I	A	A	A	A	I	5.2	6.0
T57	A	A	I	I	I/A	I	9.9	6.9
T58	A	I	I	A	I/A	A	8.6	8.3
T59	I	I	I	I	I	I	6.7	6.3
T60	I	A	I	I	I	A	9.0	8.2
T61	A	A	A	I	A	A	9.4	8.6
T62	A	A	A	A	A	A	8.5	7.4
T63	A	A	A	A	A	A	6.9	7.9
T64	A	I	I	A	I/A	A	8.9	8.2
T65	A	A	A	A	A	I	7.4	6.9

Tableau 31. Cps = composés A; AD = Analyse Discriminante ; RM = Régression Multiple ; PLS = Partial Least Square ; A = Actif A; I = Inactif ; DC = Descripteur Classique ; PE = Potentiel Electrostatique ; PL = Potentiel de Lipophilie ; S = Stérique ; Moy. = Moyenne ; Exp. = Expérimentale.

Pour les prédictions qualitatives moyennes, seulement six composés sont prédits correctement. Les autres composés sont soit mal prédits (**T56**, **T60**, **T65**) soit prédits deux fois actif et deux fois inactif (**T57**, **T58**, **T64**). Nos modèles ne différencient pas les isomères de position **T54** et **T56**, et ne tranchent pas sur l'activité des diastéréoisomères **T57** et **T58**. Toutefois, ces derniers composés **T57-58** sont la preuve parfaite que les composés à prédire doivent être relativement homologues à ceux de l'échantillon. En effet, ces deux ligands n'ont plus l'entité tétralinique et aucun composé ayant cette structure n'est présent dans l'échantillon. Nous pouvons donc affirmer que l'ensemble de nos modèles discriminants ne permet pas une prédiction qualitative correcte de nouveaux composés. De plus cette étude qualitative montre bien les limites d'une étude QSAR sur une famille donnée.

Pour les méthodes dites quantitatives, les résultats sont limités puisqu'un seul modèle est retenu. Ce modèle donne 75% des composés prédits avec un écart

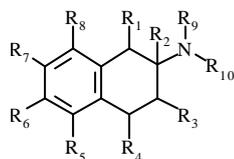
inférieur à 1 unité de pIC_{50} et 25% avec un écart compris entre 3 et 4 unités. Si l'on prend en compte le seuil de 7.5, seulement 5 ligands sur 12 sont prédits dans la mauvaise classe (active, inactive).

Effectuons maintenant la même démarche que dans l'étude précédente sur les indoles, pour étudier l'intérêt de la prédiction. On peut d'abord remarquer que les résultats des tests statistiques sont relativement médiocres et que la prudence est de rigueur. A partir des résultats qualitatifs, nous étions amenés à synthétiser 10 produits dont 3 douteux, sur les 12. On est donc en présence d'un faible gain de travail. De plus on perd, en plus, un composé prédit inactif qui présente une activité biologique de 8.2 unités de pIC_{50} . Sur les 10 composés synthétisés, 6 se seraient révélés actifs. Le résultat relatif est donc médiocre et n'apporte aucun gain réel. Le résultat de la même approche, appliquée aux données quantitatives, n'est pas meilleur : 6 composés synthétisés dont 4 sur 6 s'avèrent actifs et on perd trois ligands actifs à 8, 8.6 et 9.1.

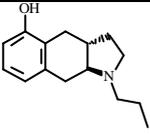
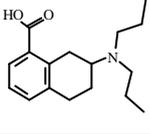
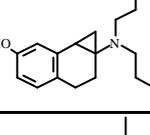
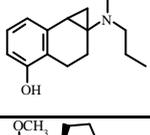
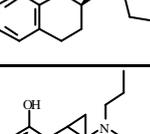
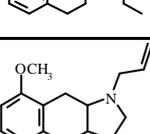
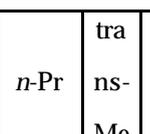
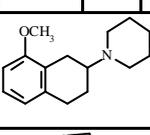
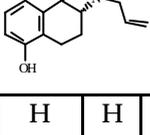
En conclusion, une approche QSAR sur cet échantillon semble vraiment peu rentable. La différence avec la famille des indoles est importante et il semble intéressant d'essayer d'en tirer des règles.

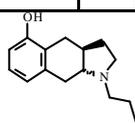
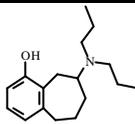
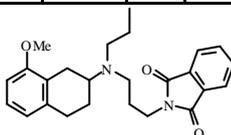
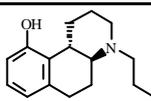
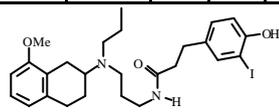
L'homogénéité interne de l'échantillon et la bonne corrélation entre les composés de cet échantillon et les composés du « test set », est clairement indispensable. Cette approche QSAR ne crée pas l'information mais ne fait que l'extraire. L'approche semble peu fiable pour les isomères de position et les diastéréoisomères. Si ceci est presque normal pour l'approche avec les descripteurs classiques, c'est un peu décevant pour l'approche avec les indices de similarité moléculaire. Une règle claire semble enfin se dégager : la qualité des prédictions est inférieure à celle des tests statistiques sur le « training set ». Des tests statistiques très bons à bons conduiront à des prévisions bonnes à moyennes. Des tests simplement moyens ne produiront que de médiocres prévisions.

ANNEXE 2

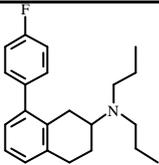
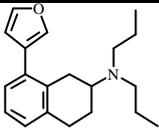
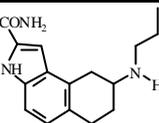


Composés	R ₁	R ₉	R ₁₀	R ₃	R ₂	R ₄	R ₅	R ₆	R ₇	R ₈	pIC ₅₀ (M)
T1	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	OH	OH	4
T2	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	I	OH	4
T3	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	OH	OH	H	H	4
T4	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	OH	F	H	H	4
T5	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	OH	H	H	4
T6	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	OH	H	4
T7	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	OH	H	H	H	4
T8											5.5
T9											5.6
T10	H	H	<i>n</i> -Pr	2 Me	H	H	H	H	H	OH	5.7
T11											5.8
T12											6
T13											6
T14	H	<i>n</i> -Pr	<i>n</i> -Pr	H	Me	H	H	H	H	OH	6
T15	trans-Me	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	H	OH	6

T16	H	<i>n</i> -Pr	<i>n</i> -Pr	cis -	H	H	H	H	H	OH	6.2
T17											6.3
T18											6.5
T19											6.6
T20											6.6
T21											6.7
T22											6.9
T23											7
T24	H	<i>n</i> -Pr	<i>n</i> -Pr	tra ns- Me	H	H	H	H	H	OH	7
T25											7.1
T26											7.25
T27	H	H	H	H	H	H	H	H	H	OMe	7.3
T28	H	-CH ₂ - Ph	Me	H	H	H	H	H	H	OMe	7.3
T29	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	H	H	7.3
T30	H	-O-CH ₂ - Ph	H	H	H	H	H	H	H	OMe	7.35

T31	H	Me	Me	H	H	H	H	H	H	OH	7.4
T32											7.4
T33											7.5
T34	H	Bu	Bu	H	H	H	H	H	H	OH	7.7
T35	H	<i>n</i> -Pr	H	H	H	H	H	H	H	H	7.7
T36	H	-(CH ₂) ₃ - Ph	H	H	H	H	H	H	H	H	7.9
T37	H	-(CH ₂) ₃ - Ph	H	H	H	H	H	H	H	-O- CH ₂ - Ph	8
T38	H	-(CH ₂) ₂ - Ph	Me	H	H	H	H	H	H	OMe	8.1
T39	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	F	H	H	OH	8.2
T40	H	-(CH ₂) ₄ - Ph	H	H	H	H	H	H	H	OMe	8.25
T41	H	Et	Et	H	H	H	H	H	H	OH	8.3
T42											8.3
T43											8.4
T44	H	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	H	OH	8.49
T45	Cis-Me	<i>n</i> -Pr	<i>n</i> -Pr	H	H	H	H	H	H	OH	8.55
T46	H	-(CH ₂) ₃ - Ph	H	H	H	H	H	H	H	OMe	8.6
T47											8.6
T48	H	H	<i>n</i> -Pr	H	H	H	H	H	H	OMe	8.6
T49	H	H	Me	H	H	H	H	H	H	OMe	8.7

T50											8.7
T51	H	-(CH ₂) ₃ - Ph	Me	H	H	H	H	H	H	OMe	8.7
T52	H	-(CH ₂) ₃ - Ph	H	H	H	H	H	H	H	OH	8.7
T53											9.7
T54											(9.1)
T55											(8.6)
T56											(6)
T57											(6.9)
T58											(8.3)
T59											(6.3)
T60											(8.2)
T61											(8.6)
T62											(7.8)

<p>T63</p>		<p>(7.9)</p>
<p>T64</p>		<p>(8.2)</p>
<p>T65</p>		<p>(6.9)</p>

III. FAMILLE DES CHROMANES

A partir des résultats vus précédemment, nous avons cherché à appliquer ces approches sur des familles de composés du Laboratoire. La composition de l'échantillon (annexe 3) est de 57 composés. Cet échantillon présente diverses activités biologiques pour le récepteur sérotoninergique 5-HT_{1A}. Les valeurs de pIC₅₀ vont de 4.1 à 10. Cet échantillon est séparé en trois sous-groupes, de 19 composés inactifs (**C1** à **C19**) ayant un pIC₅₀ < 7.5 pour le récepteur 5-HT_{1A}, de 29 composés actifs (**C20** à **C48**) avec un pIC₅₀ ≥ 7.5 pour le récepteur 5-HT_{1A} et de 9 composés à prédire avant ou après la synthèse mais toujours avant les études pharmacologiques (**C49** à **C57**). Cet échantillon présente une gamme étendue d'activité de 4.1 à 10 pIC₅₀ avec une bonne disparité entre les composés inactifs et actifs. De plus, l'échantillon est représentatif de plusieurs structures chimiques comme les amino-chromanes, les amino-méthyl-chromanes, les spiro-chromanes en C-5 et les spiro-chromanes en C-6 avec 1 ou 2 azotes .

A partir de cet ensemble de composés, l'objectif était de prédire *a priori* l'affinité et/ou de proposer à la synthèse de nouveaux composés dont l'affinité pour le récepteur transmembranaire 5-HT_{1A} serait bonne et dont la sélectivité par rapport aux autres récepteurs (5-HT_{1B}, 5-HT₂, 5-HT₃, D₁ et D₂) du SNC serait accrue. Tous les composés pris en compte dans cette étude QSAR sont extraits de travaux de synthèse effectués au Laboratoire³⁴⁻⁴⁰.

Etude 3D avec le pharmacophore de Mellin

Pour ce faire, nous avons opté pour une étude 3D des composés à partir du pharmacophore 5-HT_{1A} de Mellin. Les structures initiales des 57 composés ont été générées par recherche conformationnelle avec le logiciel COBRA d'Oxford Molecular et relaxées avec le champ de forces COSMIC. Nous avons sélectionné les conformères ayant un écart maximum de 40 KJ/mol par rapport au conformère de plus basse énergie. La superposition avec le pharmacophore 5-HT_{1A} de Mellin a été effectuée avec les atomes du noyau aromatique et l'atome d'azote situé en alpha du squelette chromane. Nous avons ensuite sélectionné le conformère qui donnait un Root Mean Square (RMS) le plus faible après superposition rigide sur le pharmacophore. D'une façon générale, l'ensemble des composés **C49** à **C57** proposés se superpose correctement au pharmacophore 5-HT_{1A}.

De plus cette étude a montré que les composés rigidifiés par une structure spiro en C-5 ou C-6, étaient tout à fait conformes aux normes du pharmacophore (distance noyau aromatique-azote, distance de l'azote par rapport au plan noyau aromatique-azote, direction du doublet électronique de l'azote). En effet, ces structures moléculaires ont la particularité de présenter des conformères de très basse énergie superposables au pharmacophore. Toutefois, nous avons pu remarquer que la forme spiro en C-6 (**C49**) présentait une meilleure superposition que son homologue en C-5 (**C10**) (Photos 1 et 2).

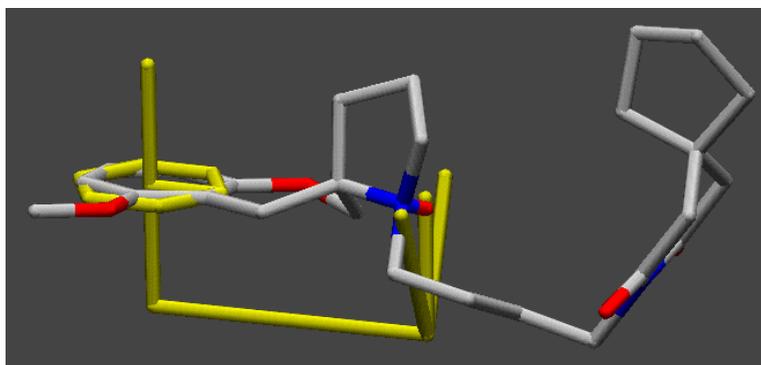


Photo 1. Composé **C10** superposé au pharmacophore de Mellin.

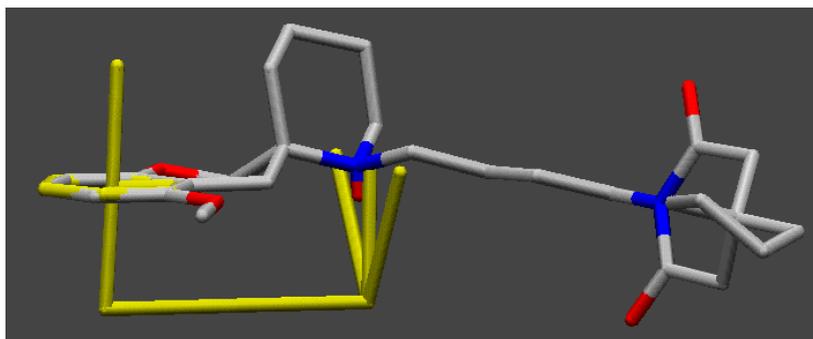


Photo 2. Composé **C49** superposé au pharmacophore de Mellin.

Ce résultat était donc très prometteur, et l'on pouvait envisager une activité biologique 5-HT_{1A} du ligand **C49** supérieure à celles des composés **C10** et **C44**.

En ce qui concerne les composés non spiraniques, le premier conformère superposable au pharmacophore 5-HT_{1A} est rarement un conformère ayant une des plus basses valeurs énergétiques. Mais dans leur totalité, ces composés se superposent très bien, et avec un faible coût énergétique, avec le pharmacophore qui est considéré comme une empreinte du site actif du récepteur.

Analyses QSAR

Pour parfaire nos prédictions qualitatives issues de l'étude de superposition, par des approches quantitatives QSAR, notre échantillon a fait l'objet d'une analyse des descripteurs de similarité moléculaire électrostatique, lipophile et stérique. La méthode statistique retenue et utilisée pour analyser ces indices est le Partial Least Squares (PLS). Les indices de similarité moléculaire et la méthode statistique PLS ont été choisis parce qu'ils présentaient de meilleurs résultats que les descripteurs classiques et les autres méthodes statistiques.

Les résultats statistiques des méthodes PLS ont servi pour la prédiction de l'affinité 5-HT_{1A} des composés **C40** à **C57**.

Une étude des indices de similarité moléculaires a donc été entreprise sur l'ensemble des 57 composés. Les tables QSAR ainsi obtenues ont été analysées par la méthode du Partial Least Square en utilisant la totalité des composés pour calculer le « cross validation $r(CV)^2$ ». Les principaux résultats statistiques sont donnés dans le tableau 32.

Méthodes d'évaluation	Résultats du PLS
	r²/r(CV)²/nombre de vecteurs
Potentiel Electrostatique	0.82 / 0.46 / 7
Potentiel Lipophile	0.76 / 0.32 / 8
Champ Electrostatique	0.89 / 0.45 / 6
Champ Lipophile	0.83 / 0.39 / 6
Forme	0.73 / 0.30 / 5

Tableau 32.

Pour le choix des modèles, nous avons sélectionné le nombre de vecteurs en fonction de la valeur de PRESS la plus faible. Pour étudier les prédictions des composés **C49** à **C57**, nous avons choisi les deux modèles (potentiel et champ) prenant en compte l'effet électrostatique. En effet, ces deux modèles présentent les meilleurs résultats statistiques tout en donnant des valeurs statistiques très proches. Les méthodes d'évaluation des indices de potentiel lipophile et de forme ont été écartées de part leur plus faible coefficient de régression, et la méthode du champ lipophile à cause de la valeur relativement faible du « cross validation ».

Les résultats des prédictions des 57 composés issus des modèles utilisés sont présentés dans le tableau 33 qui suit.

Composés	Valeurs Expérimentales	Potentiel	Champ	Valeurs Moyennes
		Electrostatique	Electrostatique	
		Valeurs Prédites	Valeurs Prédites	
C1	4.10	5.59	4.51	5.05
C2	4.84	4.36	4.64	4.50
C3	5.35	5.85	5.57	5.71
C4	5.88	5.74	6.19	5.97
C5	6.52	6.02	5.70	5.86
C6	6.74	7.22	7.06	7.14

C7	6.99	6.20	6.96	6.58
C8	7.13	6.34	6.06	6.20
C9	7.41	6.89	6.24	6.56
C10	7.97	6.44	6.89	6.66
C11	4.69	6.82	7.68	7.25
C12	6.09	7.02	7.30	7.16
C13	6.22	7.13	6.72	6.93
C14	6.39	6.96	7.02	6.99
C15	7.00	6.89	6.77	6.83
C16	7.04	8.27	7.77	8.02
C17	7.30	7.68	7.09	7.38
C18	7.39	7.48	7.15	7.32
C19	7.70	5.96	6.38	6.17
C20	7.70	7.46	7.41	7.43
C21	7.70	8.04	7.72	7.88
C22	8.10	7.13	7.10	7.12
C23	8.15	7.85	8.24	8.05
C24	8.30	8.39	8.38	8.38
C25	8.40	7.40	7.68	7.54
C26	8.40	9.03	8.92	8.98
C27	8.40	8.40	7.57	7.98
C28	8.50	8.54	8.15	8.34
C29	8.70	7.77	8.29	8.03
C30	8.70	8.59	8.38	8.48
C31	8.70	8.38	7.77	8.07
C32	8.70	8.62	8.77	8.69
C33	8.70	8.39	8.39	8.39
C34	8.70	9.05	8.47	8.76
C35	9.00	8.52	8.25	8.39
C36	9.00	8.43	8.66	8.54
C37	9.00	8.66	8.82	8.74
C38	9.00	9.00	9.36	9.18
C39	9.00	8.40	9.18	8.79

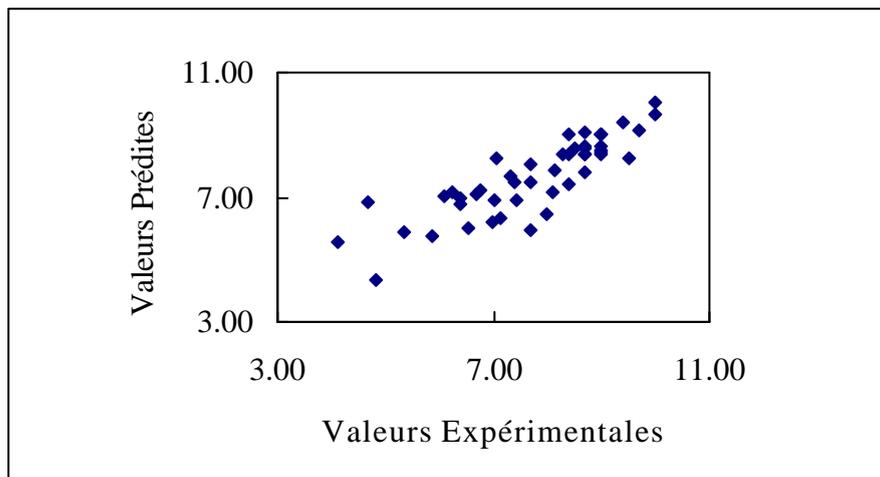
C40	9.00	9.04	9.00	9.02
C41	9.00	9.02	9.27	9.14
C42	9.40	9.39	8.74	9.07
C43	9.50	8.24	8.92	8.58
C44	9.70	9.13	9.27	9.20
C45	10.00	9.64	9.82	9.73
C46	6.40	6.80	7.20	7.00
C47	6.70	7.10	6.50	6.80
C48	10.00	10.02	10.14	10.08

Composés à prédire

C49	8.05	7.96	8.09	8.03
C50	6.40	6.59	7.06	6.82
C51	7.15	7.09	7.00	7.05
C52	7.20	7.24	6.60	6.92
C53	8.20	7.65	8.95	8.30
C54	5.38	7.85	7.27	7.56
C55	5.52	7.10	6.03	6.57
C56	4.80	6.80	5.77	6.29
C57	4.80	5.21	5.27	5.24

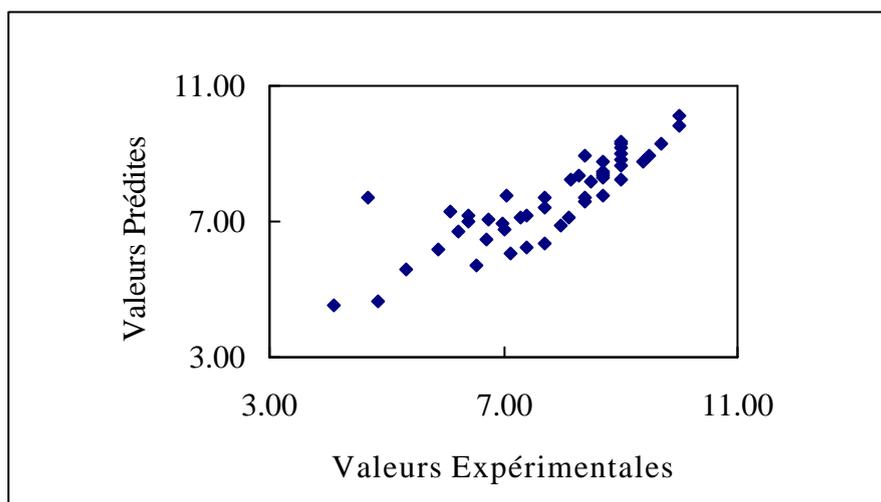
Tableau 33.

Les résultats graphiques des corrélations sont présentés dans les graphes 16, 17 et 18 qui suivent.



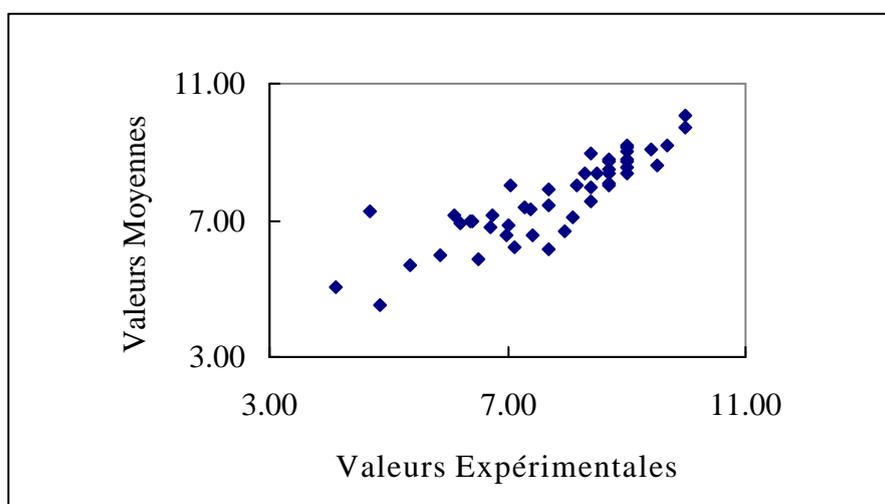
Graphe 16.

Le graphe 16 montre la corrélation entre les valeurs expérimentales et les valeurs prédites pour le modèle utilisant les indices de Carbo de potentiel électrostatique. On peut remarquer que le nuage de point s'affine pour les valeurs élevées d'activité. Nous essayerons d'expliquer ce phénomène par la suite.



Graphe 17.

Le graphe 17 montre la corrélation entre les valeurs expérimentales et les valeurs prédites pour le modèle utilisant les indices de Carbo de champ électrostatique. Nous remarquons le même phénomène que précédemment, à savoir que les composés à forte activité sont mieux corrélés.



Graphe 18.

Le graphe 18 montre la corrélation entre les valeurs expérimentales et la moyenne des valeurs prédites par le modèle utilisant les indices de Carbo de potentiel et de champ électrostatique.

Discussion

Les résultats statistiques montrent très clairement que les méthodes utilisant les indices de similarité électrostatiques donnent des meilleures conditions de linéarité et de prédiction. Nous devons donc expliquer ce phénomène et le justifier. Un examen approfondi des graphes 16, 17 et 18, montrant les corrélations entre les valeurs expérimentales et les valeurs prédites, révèle que les corrélations sont meilleures pour les ligands à forte activité que pour les composés à faible activité. Le calcul des différences entre les valeurs expérimentales et les valeurs moyennes prédites des dix premiers composés à faible activité et dix derniers composés à forte activité, donne un écart moyen de 0.29 unité de pIC_{50} pour les hautes valeurs et 0.79 pour les faibles valeurs de « binding ». De plus, on peut préciser que parmi les composés inactifs ayant un $pIC_{50} < 7.5$, 7 composés (**C1**, **C8-12**, **C16**) sur 19 ont des écarts entre les valeurs prédites et expérimentales supérieures à 1, alors que pour les composés dit actifs seulement deux composés (**C19**, **C43**) ont un écart supérieur à 1.

Ce phénomène semble être normal pour une telle analyse QSAR, et pour l'expliquer, nous appuierons notre réflexion sur le fait qu'il peut y avoir plusieurs raisons (stériques, électrostatiques et lipophiliques) expliquant un mauvais « binding » ; ces raisons peuvent être des mauvaises distributions électrostatiques, ou des problèmes liés à la lipophilie ou à la forme de la molécule. Il est donc clair, à partir de ceci, que les composés actifs ont très peu de différence entre eux alors que les ligands à faible activité peuvent se différencier par plusieurs de ces phénomènes. Nous arrivons là à une des limites des méthodes statistiques.

A partir des connaissances actuelles sur le récepteur sérotoninergique 5-HT_{1A}, nous pouvons expliquer facilement le choix des descripteurs électrostatiques. En effet, le modèle de site actif 5-HT_{1A} répond à trois points précis des composés. Ces

points correspondent à deux sites électrostatiques et un site lipophile présent chez tous les composés et qui est le cycle aromatique. Les principales différences de ces composés homogènes sont donc dues à la différence de distribution électronique induite par la variété de substituants. Les différences stériques sont surtout dues aux parties flexibles des structures et leur rôle est minimisé.

En ce qui concerne les prédictions des composés **C49** à **C57**, le tableau récapitulatif des prédictions montre que les composés **C49**, **C50**, **C51**, et **C52** sont prédits avec une erreur toujours inférieure à 1 unité de pIC_{50} (l'écart moyen est de 0.2). Pour le composé **C53**, la différence de prédiction est aussi inférieure à 1 unité de pIC_{50} . Tous ces composés contiennent un seul atome d'azote pouvant participer à une interaction de type électrostatique avec un acide aminé (résidu aspartate) du récepteur sérotoninergique 5-HT_{1A}. Pour les composés **C54**, **C55**, **C56** et **C57**, comportant dans leur graphe moléculaire deux atomes d'azotes potentiellement interactifs avec des résidus, les prédictions sont beaucoup plus aléatoires avec des différences pouvant atteindre plus de 2 unités de pIC_{50} (environ 2 pour le **C54**). Toutefois, ces composés ayant des valeurs expérimentales très inactives, sont prédits inactifs par nos modèles. Les modèles statistiques utilisés prédisent donc les neuf composés avec un faible écart pour les composés **C49** à **C53** proches de l'échantillon et avec un écart important mais tout en restant dans la gamme d'activité pour les composés **C54** à **C57** présentant une structure sensiblement différente à celle de l'ensemble des composés étudiés servant de modèle.

Pour cet ensemble de composés, les prédictions de l'affinité 5-HT_{1A} sont excellentes pour les composés proches de l'échantillon avec un écart moyen de 0.2 unité de pIC_{50} . Pour les autres composés l'écart moyen est de 1.2 mais ces ligands sont tout de même prédits inactifs par nos modèles.

Ceci montre l'intérêt d'une telle approche QSAR utilisant les indices de similarité moléculaire, avec une limite quant à la prédiction de composés ayant une structure moléculaire très différente de celles constituant le modèle. Ce modèle est

donc maintenant applicable pour de nouvelles molécules potentielles de cette famille.

Remarques

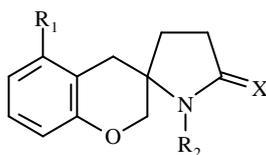
Bien que ne faisant pas partie de notre travail (notre modèle n'aborde pas cet aspect), on peut revenir au problème de la sélectivité soulevé en introduction. L'objectif était d'augmenter la sélectivité 5-HT_{1A} en rigidifiant la structure des composés. Dans le tableau 34, récapitulant les différentes activités sérotoninergiques et dopaminergiques, on remarque que les composés **C10** et **C49**, dont la structure a été rigidifiée par la forme spiro, présentent une meilleure sélectivité par rapport au récepteur D₂, ce qui n'est pas le cas pour le ligand **C44**. On voit donc ici l'intérêt d'une telle approche de rigidification de la structure.

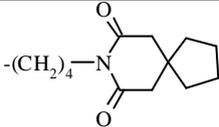
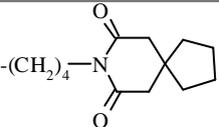
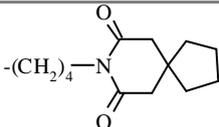
Composés	Activités					
	5-HT _{1A}	5-HT _{1B}	5-HT ₂	5-HT ₃	D ₁	D ₂
C10	7.97	5.42	5.18	4.52	4.42	6.03
C44	9.70	5.30	6.00	-	4.40	8.00
C49	8.05	5.60	5.50	4.30	4.60	5.50

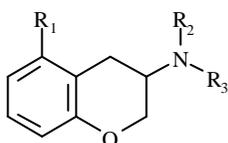
Tableau 34.

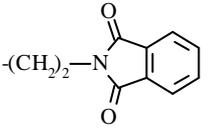
En ce qui concerne les composés **C54** à **C57**, la présence de deux azotes potentiellement interactifs avec un résidu aspartate du récepteur perturbe par évidence les interactions ligand/récepteur puisque ces composés ont tous une activité très faible (~5 unités de pIC₅₀).

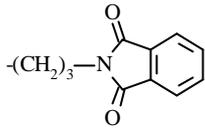
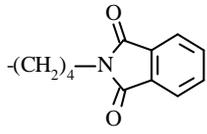
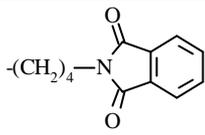
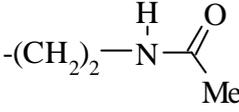
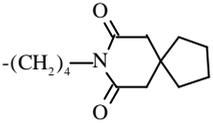
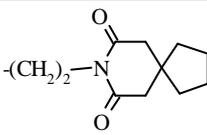
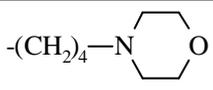
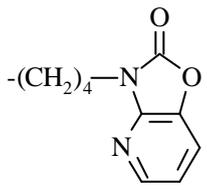
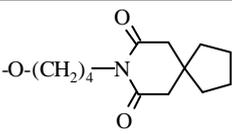
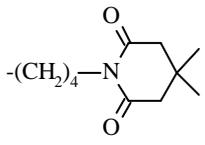
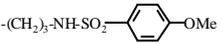
ANNEXE 3

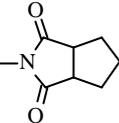
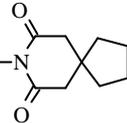
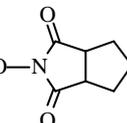
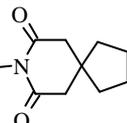
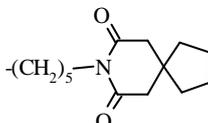
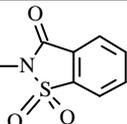
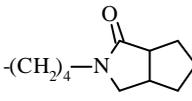
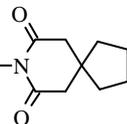
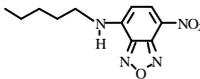


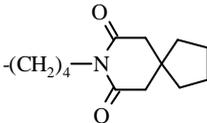
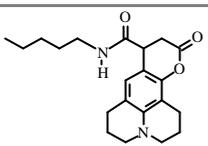
Composés	R ₁	R ₂	X	pIC ₅₀ (M)
C1	OMe	COMe	H ₂	4.10
C2	OMe	<i>n</i> -Pr	O	4.70
C3	OMe	H	O	5.50
C4	H	<i>n</i> -Pr	H ₂	6.00
C5	OMe	H	H ₂	6.60
C6	H		H ₂	6.80
C7		<i>n</i> -Pr	H ₂	7.10
C8	OMe	<i>n</i> -Pr	H ₂	7.20
C9	OH	<i>n</i> -Pr	H ₂	7.50
C10	OMe		H ₂	8.10

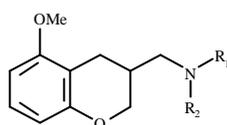


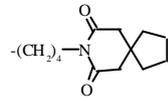
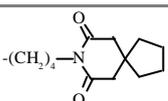
Composés	R ₁	R ₂	R ₃	pIC ₅₀ (M)
C11	C ₄ H ₄	<i>n</i> -Pr	<i>n</i> -Pr	4.70
C12	OMe		<i>n</i> -Pr	6.10
C13	OMe	H	H	6.20

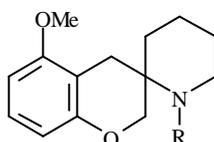
C14	H	<i>n</i> -Pr	<i>n</i> -Pr	6.40
C15	OMe	<i>n</i> -Pr		7
C16	OMe		H	7.05
C17	OMe	<i>n</i> -Pr		7.30
C18	OMe		<i>n</i> -Pr	7.40
C19	OMe	<i>n</i> -Pr	<i>n</i> -Pr	7.70
C20	OMe	H		7.70
C21	OMe		<i>n</i> -Pr	7.70
C22	OMe	<i>n</i> -Pr		8.10
C23	OMe	C ₄ H ₄		8.15
C24	OMe		<i>n</i> -Pr	8.30
C25	OH	<i>n</i> -Pr	<i>n</i> -Pr	8.40
C26		<i>n</i> -Pr	<i>n</i> -Pr	8.40
C27	OMe	<i>n</i> -Pr		8.40
C28	OMe		<i>n</i> -Pr	8.50

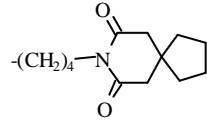
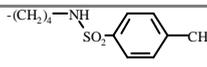
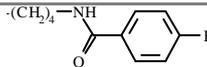
C29	OMe	<i>n</i> -Pr	$-(\text{CH}_2)_4-\text{N}$ 	8.65
C30	OMe	$-(\text{CH}_2)_4-\text{N}$ 	<i>n</i> -Pr	8.70
C31	H	$-(\text{CH}_2)_4-\text{N}$ 	<i>n</i> -Pr	8.70
C32	OMe	<i>n</i> -Pr	$-(\text{CH}_2)_4-\text{NH}-\text{SO}_2-$ 	8.70
C33	OMe	$-(\text{CH}_2)_4-\text{NH}-\text{SO}_2-$ 	<i>n</i> -Pr	8.70
C34	OMe	<i>n</i> -Pr	$-(\text{CH}_2)_4-\text{NH}-\text{SO}_2-$ 	8.70
C35	OMe	$-(\text{CH}_2)_3-\text{O}-\text{N}$ 	<i>n</i> -Pr	8.90
C36	OMe	$-(\text{CH}_2)_5-\text{N}$ 	<i>n</i> -Pr	8.90
C37	OMe	<i>n</i> -Pr	$-(\text{CH}_2)_5-\text{N}$ 	9.00
C38	OMe	$-(\text{CH}_2)_4-\text{N}$ 	<i>n</i> -Pr	9.00
C39	OMe	<i>n</i> -Pr	$-(\text{CH}_2)_4-\text{N}$ 	9.00
C40	OH	$-(\text{CH}_2)_4-\text{N}$ 	<i>n</i> -Pr	9.00
C41	OMe	<i>n</i> -Pr	$-(\text{CH}_2)_4-\text{NH}-\text{CO}-$ 	9.00
C42	OMe	$-(\text{CH}_2)_4-\text{NH}-\text{CO}-$ 	<i>n</i> -Pr	9.40
C43	OMe	<i>n</i> -Pr		9.50

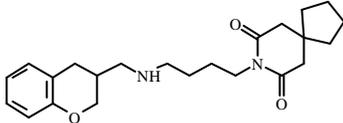
C44	OMe		<i>n</i> -Pr	9.70
C45	OMe	<i>n</i> -Pr		10.0

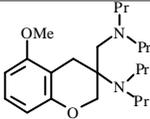
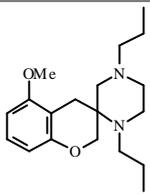
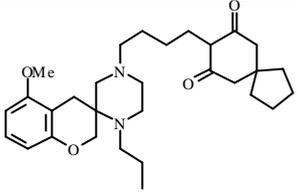
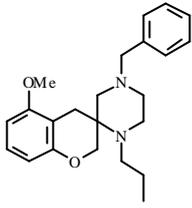


Composés	R ₁	R ₁	pIC ₅₀ (M)
C46	<i>n</i> -Pr		6.4
C47	<i>n</i> -Pr	<i>n</i> -Pr	6.7
C48		H	10



Composés	R ₁	pIC ₅₀ (M)
C49		(8.05)
C50	<i>n</i> -Pr	(6.40)
C51		(7.15)
C52		(7.20)

Composés	Structures	pIC ₅₀ (M)
C53		(8.20)

C54		(5.40)
C55		(5.52)
C56		(4.80)
C57		(4.80)

IV. FAMILLE DES THIOCHROMANES

La composition de l'échantillon (annexe 4) est de 44 composés. Cet échantillon sera étudié vis-à-vis des récepteurs sérotoninergiques 5-HT_{1A}, adrénérgiques α_1 , α_2 et dopaminérgiques D₂. Les valeurs de pIC₅₀ vont de 4 à 10. Cet échantillon est séparé en trois sous-groupes, de 19 composés inactifs (**TH1** à **TH19**) ayant un pIC₅₀ < 7.5 pour le récepteur 5-HT_{1A}, de 20 composés actifs (**TH20** à **TH39**) avec un pIC₅₀ ≥ 7.5 pour le récepteur 5-HT_{1A} et de 5 composés à prédire avant les études de pharmacologie (**TH40** à **TH44**). Dans ce chapitre, l'échantillon a fait l'objet d'une étude des descripteurs moléculaires classiques et des indices de similarité moléculaire. Différentes méthodes d'analyses statistiques ont été testées ; ce sont l'analyse discriminante (AD), la régression multiple (RM) et le Partial Least Squares (PLS). Cet échantillon présente une gamme étendue d'activité avec une bonne disparité entre les composés inactifs et actifs. Dans cette étude, nous avons testé et comparé les différents descripteurs et les différentes méthodes d'analyse statistique. Tous ces résultats ont servis pour la prédiction de l'affinité 5-HT_{1A} des composés **TH40** à **TH44**, mais aussi pour la prédiction de la sélectivité 5-HT_{1A} de ces composés envers les récepteurs α_1 , α_2 et D₂. Tous les composés pris en compte dans cette étude QSAR sont extraits de travaux de synthèse du Laboratoire^{41,42}.

Analyses QSAR pour l'affinité 5-HT_{1A}

Les descripteurs utilisés sont ceux proposés par le logiciel TSAR ; ils comprennent les indices de connectivité, de forme et de flexibilité, les indices topologiques, les propriétés moléculaires et les autocorrélogrammes utilisés avec un pas maximum de 11 pour le 2D et de 10 pour le 3D. Tous ces descripteurs forment un crible de 395 variables.

Pour l'étude des indices de similarité, les composés ont été alignés selon les critères du pharmacophore 5-HT_{1A}. Les indices de Carbo ont été calculés par optimisation (rotation, translation et déformation des angles dièdres) des molécules. Les trois critères choisis sont la charge, la lipophilie et la forme des molécules. Les méthodes utilisées sont le calcul des potentiels de charge et de lipophilie, le calcul des champs électrostatiques, lipophiliques et de forme.

Analyse discriminante

Les résultats de l'analyse sont regroupés dans le tableau 35 qui suit.

Méthode d'évaluation	Activité	Nombre de variables	r(CV) ² Inactif/Actif	Composés bien classés	Inactif	Actif
Descripteurs Classiques	5-HT _{1A}	6	0.62/0.79	93 %	86 %	100 %
	α_1	5	0.71/0.75	90 %	88 %	92 %
	α_2	2	0.90/1	97 %	95 %	100 %
	D ₂	7	0.70/0.33	100 %	100 %	100 %
Potentiel Electrostatique	5-HT _{1A}	5	0.58/0.45	82 %	80 %	85 %
	α_1	5	0.65/0.58	93 %	94 %	92 %
	α_2	4	1/0.70	97 %	100 %	93 %
	D ₂	4	0.83/0.33	93 %	100 %	67 %
Potentiel Lipophile	5-HT _{1A}	3	0.53/0.65	74 %	74 %	74 %
	α_1	2	0.77/0.75	83 %	88 %	75 %
	α_2	6	0.79/0.70	93 %	100 %	80 %
	D ₂	5	0.61/0.33	83 %	83 %	83 %
Champ Electrostatique	5-HT _{1A}	10	0.37/0.45	90 %	90 %	90 %
	α_1	7	0.77/0.50	93 %	100 %	83 %
	α_2	5	0.84/0.60	100 %	100 %	100 %
	D ₂	8	1/0.17	97 %	100 %	83 %
Champ Lipophile	5-HT _{1A}	3	0.42/0.65	77 %	79 %	75 %
	α_1	5	0.53/0.67	93 %	94 %	92 %
	α_2	4	0.74/0.60	93 %	100 %	80 %
	D ₂	5	0.91/0.33	83 %	91 %	50 %
Forme	5-HT _{1A}	6	0.47/0.60	80 %	75 %	85 %
	α_1	5	0.71/0.75	80 %	78 %	82 %
	α_2	6	0.95/0.70	97 %	100 %	90 %
	D ₂	5	0.70/0.33	83 %	83 %	83 %

Tableau 35.

Pour les descripteurs moléculaires classiques l'indice prédictif $r(CV)^2$ varie de 0.33 pour le récepteur D_2 à 1 pour α_2 . Le nombre de variables utilisées varie de 2 à 7 et le nombre de composés bien classés de 90% à 100%. Pour le récepteur 5-HT_{1A}, la discrimination de l'affinité est reliée à six variables qui sont des indices de Kier et de Kappa. Pour le récepteur α_1 , la valeur du «binding» est proportionnelle à un ensemble de cinq variables composé d'un vecteur 3D d'autocorrélation chromatisé par la charge électrostatique, et de quatre indices de Randic, de Balaban et de Kappa. Pour le récepteur α_2 , la discrimination des composés est due à deux vecteurs d'autocorrélation 3D pondérés par la charge électrostatique. Le récepteur dopaminergique D_2 a, quant à lui, une affinité proportionnelle à sept variables composées d'un indice de Kappa, d'un vecteur d'autocorrélation 3D pondéré par la charge électrostatique, de deux indices de Kier, de la réfraction molaire, du volume ellipsoïdal et de la valeur IX composante de l'inertie des molécules.

L'ensemble de cette étude de l'affinité des composés vis-à-vis des récepteurs montre des résultats qui sont dans l'ensemble statistiquement corrects. Le choix de telle ou telle méthode d'évaluation pour un récepteur donné est souvent orienté par le nombre de variables utilisé et par le pourcentage total de discrimination. De ce fait, la méthode retenue pour les récepteurs 5-HT_{1A}, α_2 et D_2 est celle utilisant les descripteurs moléculaires classiques. Pour le récepteur α_1 , nous retiendrons le potentiel électrostatique.

L'ensemble des méthodes sélectionnées, pour cette étude de l'affinité, est en relation directe avec des variables de forme et de type électrostatique. Aucune variable de lipophilie n'a été prise en compte dans les modèles sélectionnés.

La régression multiple

Les résultats de l'analyse de régression multiple pour l'étude de l'affinité des composés sont donnés dans le tableau 36.

Méthode d'évaluation	Activité	r(CV) ²	Probabilité de F	Valeur de S	r ² /variables
Descripteurs Classiques	5-HT _{1A}	0.13	1E-5	1.308	0.40/1
	α_1	0.09	2.E-5	1.101	0.64/4
	α_2	-0.25	5E-7	0.898	0.72/4
	D ₂	0.27	2E-7	0.685	0.74/4
Potentiel Electrostatique	5-HT _{1A}	0.46	6E-6	1.109	0.64/7
	α_1	0.19	9E-4	1.355	0.41/2
	α_2	0.44	5E-8	0.774	0.81/6
	D ₂	0.40	5E-7	0.631	0.83/8
Potentiel Lipophile	5-HT _{1A}	-0.16	0.1	1.637	0.10/1
	α_1	0.21	2E-3	1.448	0.30/1
	α_2	0.19	6E-4	1.217	0.49/3
	D ₂	0.10	9E-3	1.048	0.41/4
Champ Electrostatique	5-HT _{1A}	0.03	1E-6	1.191	0.59/8
	α_1	0.41	6E-7	0.903	0.79/6
	α_2	-0.08	1E-5	1.047	0.62/3
	D ₂	0.06	6E-4	0.961	0.49/3
Champ Lipophile	5-HT _{1A}	0.19	1E-3	1.380	0.40/4
	α_1	0.25	6E-5	1.094	0.67/6
	α_2	0.50	1E-4	1.091	0.62/5
	D ₂	0.29	2E-3	0.983	0.49/4
Forme	5-HT _{1A}	0.05	2E-3	1.377	0.41/5
	α_1	-0.91	3E-6	0.913	0.79/8
	α_2	0.44	1E-5	0.963	0.72/6
	D ₂	0.45	6E-5	0.830	0.64/5

Tableau 36.

Pour les descripteurs moléculaires classiques l'indice prédictif r(CV)² donne une valeur de -0.25 pour le récepteur α_2 à 0.27 pour D₂. Le nombre de variables utilisées varie de 1 à 4 et le coefficient de régression linéaire r² de 0.40 pour le

récepteur 5-HT_{1A} à 0.74 pour D₂. Pour le récepteur 5-HT_{1A}, l'affinité est reliée à une seule variable qui est un indice d'autocorrélogramme 3D pondéré par la charge électrostatique. Pour le récepteur α_1 , la valeur du « binding » est proportionnelle à un ensemble de quatre variables composé de deux vecteurs 3D d'autocorrélation chromatés par la charge électrostatique, d'un indice de Balaban et du moment dipolaire total. Pour le récepteur α_2 , l'ensemble des variables de la régression multiple est composé de trois vecteurs d'autocorrélation 3D chromatés par la charge électrostatique et d'un indice de Balaban. Le récepteur dopaminergique D₂ a une affinité proportionnelle à quatre variables composées d'un indice de Randic, d'un indice de Balaban et de deux vecteurs d'autocorrélation 3D pondérés par la charge électrostatique.

L'ensemble de cette étude de l'affinité des composés vis-à-vis des récepteurs par des analyses de régressions multiples montre des résultats qui sont dans l'ensemble statistiquement différents. Nous pouvons remarquer des $r(CV)^2$ négatifs, valeur de F faibles, des valeurs de S élevées et des coefficients de régression r^2 assez faibles. Le choix de telle ou telle méthode d'évaluation pour un récepteur donné est souvent orienté par la valeur du «cross validation», par les valeurs de F et S, et surtout par le coefficient de régression. De ce fait, la méthode devant être retenue pour le récepteur 5-HT_{1A} apparaît être celle utilisant les indices de similarité moléculaire des potentiels électrostatiques. Cette méthode est la mieux adaptée tout en étant relativement discutable de part son faible coefficient de régression 0.64, son nombre de variables élevé (7) et sa valeur de S élevée (1.109). Pour le récepteur α_1 , nous retiendrons principalement la méthode d'évaluation du champ électrostatique. Pour les récepteurs α_2 et D₂, les méthodes d'évaluations retenues sont celles utilisant les indices de similarités du potentiel électrostatique. Nous pouvons remarquer que tous les modèles sélectionnés sont en relation directe avec des indices électrostatiques.

Le Partial Least Square (PLS)

Les résultats de l'analyse PLS pour l'étude de l'affinité des composés sont proposés dans le tableau 37 suivant.

Méthode d'Evaluation	Activité	r(CV) ²	PRESS	r ² /vecteurs
Descripteurs Classiques	5-HT _{1A}	0.28	28.24	0.54/2
	α_1	0.41	16.58	0.87/5
	α_2	0.51	13.71	0.75/2
	D ₂	0.39	17.11	0.61/2
Potentiel Electrostatique	5-HT _{1A}	0.14	32.82	0.56/2
	α_1	0.53	16	0.84/7
	α_2	0.55	15.35	0.80/2
	D ₂	0.19	22.67	0.60/2
Potentiel Lipophile	5-HT _{1A}	-0.07	40.65	0.30/4
	α_1	0.27	20.38	0.70/6
	α_2	0.59	17.06	0.73/6
	D ₂	-0.06	29.56	0.12/1
Champ Electrostatique	5-HT _{1A}	0.44	32.74	0.94/4
	α_1	0.40	16.92	0.99/7
	α_2	0.50	14.05	0.99/8
	D ₂	0.47	14.94	0.99/10
Champ Lipophile	5-HT _{1A}	-0.01	38.21	0.42/2
	α_1	0.17	23.23	0.89/5
	α_2	0.14	24	0.91/6
	D ₂	0.10	25.24	0.97/10
Forme	5-HT _{1A}	-0.16	44.22	0.59/3
	α_1	0.16	23.64	0.38/1
	α_2	0.41	16.58	0.94/7
	D ₂	0.61	11.08	0.91/7

Tableau 37.

Pour les descripteurs moléculaires classiques l'indice prédictif r(CV)² donne une valeur de 0.28 pour le récepteur 5-HT_{1A} à 0.51 pour α_2 . Le nombre de vecteurs utilisés varie de 2 à 5 et le coefficient de régression linéaire r² de 0.54 pour le

récepteur 5-HT_{1A} à 0.87 pour α_1 . Les valeurs de PRESS données dans le tableau 37, sont les premières valeurs les plus faibles obtenues par l'analyse. Pour le récepteur 5-HT_{1A}, l'affinité est reliée à deux vecteurs exprimant principalement des indices d'autocorrélogramme 3D pondérés par la charge électrostatique et la lipophilie. Pour le récepteur α_1 , la valeur du « binding » est proportionnelle à un ensemble de cinq vecteurs exprimant principalement des indices de Kier, de Balaban et des autocorrélogrammes 3D chromatisés par la charge électrostatique et le Log P. Pour le récepteur α_2 , les principales variables des vecteurs sont des vecteurs d'autocorrélation 3D pondérés par la charge électrostatique. Le récepteur dopaminergique D₂ a une affinité proportionnelle à deux vecteurs exprimant principalement des variables composées de vecteurs d'autocorrélation 3D pondérés par la charge électrostatique.

Comme pour l'approche par régression multiple, les analyses PLS donnent des résultats différents selon les récepteurs et les méthodes d'évaluations utilisées. Nous pouvons remarquer des $r(\text{CV})^2$ négatifs, un nombre important de vecteurs utilisés et des coefficients de régression r^2 assez faibles. Le choix de telle ou telle méthode d'évaluation pour un récepteur donné est souvent orienté par la valeur du « cross validation », par le nombre de vecteurs utilisés et surtout par le coefficient de régression linéaire. La méthode retenue pour le récepteur 5-HT_{1A} sera celle utilisant les indices de similarité moléculaire des champs électrostatiques. Pour les récepteurs α_1 et α_2 , nous retiendrons principalement la méthode d'évaluation des potentiels électrostatiques, ceci pour des raisons de « cross validation » et du nombre de vecteurs utilisés. Pour le récepteur dopaminergique D₂, la méthode d'évaluation retenue est sans aucune ambiguïté celles utilisant les indices de similarités de forme.

Discussion

Le tableau 38 qui suit montre les résultats statistiques choisis pour les trois méthodes d'analyses avec les deux types de descripteurs pour l'étude de l'affinité relative au récepteur 5-HT_{1A}.

Méthodes d'évaluations	AD	RM	PLS
Descripteurs Classiques	93% $r(\text{CV})^2 = 0.62/0.79$	$r^2 = 0.40$ $r(\text{CV})^2 = 0.13$	$r^2 = 0.54$ $r(\text{CV})^2 = 0.28$
Indices de Similarités	82% $r(\text{CV})^2 = 0.58/0.45$ Potentiel Electrostatique	$r^2 = 0.64$ $r(\text{CV})^2 = 0.46$ Potentiel Electrostatique	$r^2 = 0.94$ $r(\text{CV})^2 = 0.44$ Champ Electrostatique

Tableau 38.

Il en ressort globalement que les indices de similarité moléculaires donnent de meilleurs résultats statistiques ($r^2 > 0.6$ et $r(\text{CV})^2 > 0.44$) que les descripteurs moléculaires classiques pour les méthodes RM et PLS. Toutefois la méthode d'analyse discriminante donne aussi de bons résultats avec les descripteurs classiques (0.93%). La méthode la plus fiable d'après les résultats statistiques serait la méthode du PLS utilisant les indices de similarités moléculaires de type champ électrostatique. Cependant, pour la prédiction de nouveaux composés, il serait bon de tester toutes ces méthodes afin d'obtenir une valeur moyenne de la prédiction.

On remarque aussi qu'en général les indices de similarités sont plutôt de type électrostatique que lipophile ou de forme. Cette remarque est en accord avec les multiples travaux de biologie moléculaire qui suggèrent que les principales interactions ligand/récepteur soient des interactions de type électrostatique entre le ligand et des résidus aspartate (ASP) et serine (SER) du récepteur transmembranaire 5-HT_{1A}.

Analyses QSAR pour la sélectivité 5-HT_{1A}

Nous avons sur cet échantillon voulu vérifier si un calcul direct donnant une prévision de la sélectivité était envisageable. L'avantage d'une telle approche sera de diviser par deux les temps de calcul, dans le cas où seule une sélectivité nous intéresserait. De plus, la prise en compte de la seule différence entre ces affinités pouvait *a priori* permettre d'effacer des caractéristiques communes et ainsi améliorer les prévisions. Les descripteurs utilisés sont ceux proposés par le logiciel TSAR ; ils comprennent les indices de connectivité, de forme et de flexibilité, les indices topologiques, les propriétés moléculaires et les autocorrélogrammes utilisés avec un pas maximum de 11 pour le 2D et de 10 pour le 3D. Tous ces descripteurs forment un crible de 395 variables.

Pour l'étude des indices de similarité, les composés ont été alignés selon les critères du pharmacophore 5-HT_{1A}. Les indices de Carbo ont été calculés par optimisation (rotation, translation et déformation des angles dièdres) des molécules. Les trois critères choisis sont la charge, la lipophilie et l'aspect stérique. Les méthodes utilisées sont le calcul des potentiels de charge et de lipophilie, le calcul des champs électrostatiques, lipophiliques et de forme.

Cette étude a pour objet de créer des modèles statistiques afin de prévoir la sélectivité pour le récepteur sérotoninergique 5-HT_{1A} des composés du crible vis-à-vis des récepteurs adrénérgiques α_1 , α_2 et dopaminergique D₂. Pour attribuer une valeur de la sélectivité à tous les composés, nous avons soustrait les valeurs de « binding » correspondant aux récepteurs donnés. Tous les composés n'avaient pas de valeur d'activité pour les récepteurs adrénérgiques et dopaminergiques. L'échantillon a donc été réduit à 9 composés inactifs, 20 composés actifs et 5 composés à prédire. Dans cette étude, l'analyse discriminante n'a pas été utilisée car la création de deux classes fait perdre trop d'information.

La régression multiple

Les résultats de l'analyse de régression multiple pour l'étude de sélectivité des composés sont donnés dans le tableau 39.

Méthode d'évaluation	Sélectivité	F	S	r(CV) ²	r ² /variables
Descripteurs Classiques	5-HT _{1A} / α_1	1E-6	0.859	-0.18	0.70/4
	5-HT _{1A} / α_2	5E-6	0.926	-0.07	0.49/1
	5-HT _{1A} / D ₂	2E-5	0.592	-0.51	0.61/3
Potentiel Electrostatique	5-HT _{1A} / α_1	5.E-8	0.785	0.46	0.74/3
	5-HT _{1A} / α_2	8E-6	0.862	0.59	0.64/3
	5-HT _{1A} / D ₂	6E-5	0.616	0.12	0.56/3
Potentiel Lipophile	5-HT _{1A} / α_1	5E-5	0.986	0.05	0.62/4
	5-HT _{1A} / α_2	1E-2	1.204	-0.34	0.22/1
	5-HT _{1A} / D ₂	2E-2	0.798	-0.92	0.25/2
Champ Electrostatique	5-HT _{1A} / α_1	6E-6	0.937	0.43	0.64/3
	5-HT _{1A} / α_2	2E-3	1.049	0.19	0.48/4
	5-HT _{1A} / D ₂	6E-3	0.739	0.26	0.37/3
Champ Lipophile	5-HT _{1A} / α_1	2E-8	0.686	0.47	0.83/6
	5-HT _{1A} / α_2	2E-6	0.727	0.57	0.77/7
	5-HT _{1A} / D ₂	2E-4	0.585	0.36	0.67/7
Forme	5-HT _{1A} / α_1	9E-7	0.820	0.59	0.74/5
	5-HT _{1A} / α_2	2E-8	0.591	0.23	0.86/8
	5-HT _{1A} / D ₂	1E-3	0.710	0.02	0.40/2

Tableau 39.

L'ensemble de cette étude de sélectivité des composés vis-à-vis des récepteurs par des analyses de régressions multiples montre des résultats qui sont dans l'ensemble statistiquement différents. Nous pouvons remarquer des r(CV)² négatifs pour tous les modèles utilisant les descripteurs classiques et des coefficients de régression linéaire r² assez faibles pour les potentiels. Les modèles statistiques retenus pour la sélectivité du récepteur 5-HT_{1A} par rapport aux récepteurs adrénergiques α_1 et α_2 et dopaminergique D₂ apparaissent être ceux

utilisant les indices de similarité moléculaires des champs lipophiliques. Tous les modèles sélectionnés sont en relation directe avec des indices de lipophilie.

Le Partial Least Square (PLS)

Les résultats de l'analyse PLS pour l'étude de sélectivité des composés sont proposés dans le tableau 40.

Méthode d'Evaluation	Sélectivité	r(CV) ²	PRESS	r ² /vecteurs
Descripteurs Classiques	5-HT _{1A} / α_1	0.31	19.30	0.75/3
	5-HT _{1A} / α_2	0.54	12.80	0.72/2
	5-HT _{1A} / D ₂	0.42	16.21	0.87/4
Potentiel Electrostatique	5-HT _{1A} / α_1	0.19	22.74	0.62/2
	5-HT _{1A} / α_2	0.47	14.84	0.96/7
	5-HT _{1A} / D ₂	0.42	16.21	0.94/7
Potentiel Lipophile	5-HT _{1A} / α_1	0.29	19.96	0.46/2
	5-HT _{1A} / α_2	0.16	23.61	0.26/1
	5-HT _{1A} / D ₂	0.10	25.09	0.21/1
Champ Electrostatique	5-HT _{1A} / α_1	0.40	16.78	0.93/10
	5-HT _{1A} / α_2	0.48	14.65	0.99/10
	5-HT _{1A} / D ₂	0.43	15.96	0.99/8
Champ Lipophile	5-HT _{1A} / α_1	0.53	13.14	0.92/9
	5-HT _{1A} / α_2	0.30	19.71	0.97/8
	5-HT _{1A} / D ₂	0.41	16.65	0.94/5
Forme	5-HT _{1A} / α_1	0.34	18.51	0.51/1
	5-HT _{1A} / α_2	0.50	22.04	0.93/6
	5-HT _{1A} / D ₂	0.24	21.31	0.84/6

Tableau 40.

Pour les descripteurs moléculaires classiques l'indice prédictif r(CV)² donne une valeur de 0.31 pour la sélectivité du récepteur 5-HT_{1A} par rapport au récepteur α_1 à 0.54 pour α_2 . Le nombre de vecteurs utilisés varie de 2 à 4 pour des coefficients de régression r² compris entre 0.72 et 0.87. Pour la sélectivité du récepteur 5-HT_{1A} comparé au récepteur α_1 , les résultats biologiques sont reliés à des vecteurs

exprimant principalement des indices d'autocorrélogramme 3D pondérés par la charge électrostatique et la lipophilie et un indice de Kier. Pour la sélectivité par rapport au récepteur α_2 , la valeur du « binding » est proportionnelle à un ensemble de vecteurs exprimant principalement des indices d'autocorrélogrammes 3D chromatisés par la charge électrostatique et le Log P. Pour la sélectivité vis-à-vis du récepteur D_2 , les principales variables des vecteurs sont des vecteurs d'autocorrélation 3D pondérés par la charge électrostatique.

Pour l'ensemble de cette étude de sélectivité des composés vis-à-vis des récepteurs par des analyses de PLS, les résultats diffèrent selon les sélectivités étudiées et les méthodes d'évaluations utilisées. Nous pouvons remarquer des $r(CV)^2$ faibles (0.10), un nombre important de vecteurs utilisés (9 voir 10) et des coefficients de régression r^2 assez faibles (0.21). Le choix de telle ou telle méthode d'évaluation pour une sélectivité donnée est souvent orientée par la valeur du « cross validation », par le nombre de vecteurs utilisés et surtout par le coefficient de régression. De ce fait, la méthode retenue pour la sélectivité du récepteur 5-HT_{1A} par rapport au récepteur α_1 apparaît être celle utilisant les indices de similarité moléculaires des champs lipophiliques où le coefficient de régression est de 0.92 pour 9 vecteurs avec un « cross validation » de 0.53. Pour la sélectivité vis-à-vis du récepteur α_2 , nous retiendrons principalement la méthode d'évaluation des indices de formes, ceci pour des raisons de nombre de vecteurs utilisés. Pour la sélectivité par rapport au récepteur dopaminergique D_2 , la méthode d'évaluation retenue est sans aucune ambiguïté celles utilisant les indices de similarités des champs lipophiliques. Comme pour la régression multiple, tous ces modèles sélectionnés sont en relation directe avec des indices lipophiliques et de forme.

Discussion

Les tableaux 41, 42, 43 suivants montrent les résultats statistiques choisis pour les deux méthodes d'analyses (RM, PLS) avec les deux types de descripteurs pour les trois études de sélectivité relative au récepteur 5-HT_{1A}.

Méthodes d'évaluations	RM	PLS
Descripteurs Classiques	$r^2 = 0.70$ $r(\text{CV})^2 = 0.18$	$r^2 = 0.75$ $r(\text{CV})^2 = 0.31$
Indices de Similarités	$r^2 = 0.83$ $r(\text{CV})^2 = 0.47$ Champ Lipophile	$r^2 = 0.92$ $r(\text{CV})^2 = 0.53$ Champ Lipophile

Tableau 41. Sélectivité 5-HT_{1A} par rapport à α_1 .

Méthodes d'évaluations	RM	PLS
Descripteurs Classiques	$r^2 = 0.49$ $r(\text{CV})^2 = -0.07$	$r^2 = 0.72$ $r(\text{CV})^2 = 0.54$
Indices de Similarités	$r^2 = 0.77$ $r(\text{CV})^2 = 0.57$ Champ Lipophile	$r^2 = 0.93$ $r(\text{CV})^2 = 0.50$ Forme

Tableau 42. Sélectivité 5-HT_{1A} par rapport à α_2 .

Méthodes d'évaluations	RM	PLS
Descripteurs Classiques	$r^2 = 0.61$ $r(\text{CV})^2 = -0.51$	$r^2 = 0.87$ $r(\text{CV})^2 = 0.42$
Indices de Similarités	$r^2 = 0.67$ $r(\text{CV})^2 = 0.36$ Champ Lipophile	$r^2 = 0.94$ $r(\text{CV})^2 = 0.41$ Champ Lipophile

Tableau 43. Sélectivité 5-HT_{1A} par rapport à D₂.

Il en ressort globalement que les indices de similarité moléculaire donnent là encore de meilleurs résultats statistiques que les descripteurs moléculaires classiques pour les méthodes RM et PLS. L'analyse de régression multiple utilisant les descripteurs classiques donne des « cross validation » négatifs (-0.18, -0.07 et -0.51) et des coefficients de régression faibles (0.49, 0.61 et 0.70). Cette même méthode d'analyse avec les indices de similarité moléculaire des champs lipophiliques donne des résultats statistiquement corrects ($r^2 > 0.67$ et $r(\text{CV})^2 > 0.36$). La méthode la plus intéressante d'après les résultats statistiques serait la méthode du PLS utilisant les

indices de similarités moléculaires, de type champ lipophile pour les sélectivités α_1 et D_2 et de type forme pour α_2 .

Pour cette étude de la sélectivité, à l'inverse de l'étude de l'affinité traitée précédemment, les descripteurs retenus sont cette fois représentatifs de la lipophilie ou de la forme des molécules. Cette information est très intéressante puisqu'elle montre clairement pour toutes les sélectivités traitées, que la différenciation entre les récepteurs se fait selon l'aspect lipophile et stérique.

Analyses des prédictions pour l'affinité et la sélectivité

Prédictions de l'affinité

Dans cette étude des composés de type thiochromane, cinq ligands (**TH40** à **TH44**) ont été prédits par les méthodes QSAR avant d'être testés par les méthodes *in vitro* dites de « binding ». Nous présenterons donc ici les résultats obtenus avec les méthodes statistiques choisies précédemment selon les critères statistiques et nous les comparerons aux résultats expérimentaux trouvés après l'étude. Nous donnerons les résultats de la prédiction des affinités 5-HT_{1A}, α_1 , α_2 et D_2 et discuterons de l'intérêt ou non de ceux issus des études de sélectivité entre ces récepteurs.

Les résultats de prédiction de l'affinité avec la méthode d'analyse discriminante sont regroupés dans le tableau 44.

Composés	5-HT _{1A}		α_1		α_2		D_2	
	Descripteurs Classiques		Potentiel Electrostatique		Descripteurs Classiques		Descripteurs Classiques	
	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.
TH40	A	A	I	I	I	I	I	I
TH41	A	I	I	I	I	I	I	I
TH42	A	A	I	I	I	I	I	I
TH43	A	A	I	A	I	I	I	A
TH44	I	A	I	I	I	I	I	I

Tableau 44. A = Actif, I = Inactif, Exp. = Expérimental, Préd. = Prédiction.

Pour les prédictions de l'affinité 5-HT_{1A}, α_1 , α_2 et D₂ des composés **TH40** à **TH44**, sur les vingt prédictions seize sont justes. Ces prédictions sont donc encourageantes.

Les prédictions des valeurs de « binding » exprimées en pIC₅₀ avec la méthode de régression multiple donnent les résultats suivants (Tableau 45).

Composés	5-HT _{1A}		α_1		α_2		D ₂	
	Potentiel Electrostatique		Champ Electrostatique		Potentiel Electrostatique		Potentiel Electrostatique	
	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.
TH40	8.2	8.5	6.4	6.5	7.3	6	6.8	5.8
TH41	8	6.5	7	5.5	6.3	5.6	6.4	5.8
TH42	7.8	6	6.4	6.9	5.7	5.6	6	6
TH43	7.9	8	6.4	8.1	6	6.7	6.4	7.1
TH44	6.8	7.3	7.2	7.1	7.2	7.4	6.3	6.2

Tableau 45. Exp. = Expérimental, Préd. = Prédiction.

Pour l'ensemble des prédictions de cette méthode, l'écart entre les valeurs expérimentales et les valeurs des prédictions est de 0.7 unité de pIC₅₀. Pour les prédictions de l'activité 5-HT_{1A}, l'écart est de 0.84 unité. Cet écart pour les autres récepteurs α_1 , α_2 et D₂ est respectivement de 0.58, 0.6 et 0.48. Sur les 20 prédictions, seulement 3 (gras) prédisent les composés dans la mauvaise classe (active, inactive).

Les prédictions des valeurs de « binding » exprimées en pIC₅₀ avec la méthode du PLS donnent les résultats suivants (Tableau 46).

Composés	5-HT _{1A}		α_1		α_2		D ₂	
	Potentiel Electrostatique		Champ Electrostatique		Potentiel Electrostatique		Potentiel Electrostatique	
	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.
TH40	8.2	7.3	6.4	6.3	7.3	6	6.8	6.9
TH41	8	7.8	7	7.2	6.3	6.1	6.4	6.4
TH42	7.8	7.8	6.4	7.1	5.7	5.8	6	7.2
TH43	7.9	8	6.4	7.3	6	6.8	6.4	5.9
TH44	6.8	7.2	7.2	7.2	7.2	7	6.3	6.2

Tableau 46. Exp. = Expérimental, Préd. = Prédiction.

Pour l'ensemble des prédictions de cette méthode, l'écart entre les valeurs expérimentales et les valeurs des prédictions est de 0.5 unité de pIC₅₀. Cette valeur est plus faible que celle pour l'analyse de régression multiple, ce qui est normal à la vu des résultats statistiques. Pour les prédictions de l'activité 5-HT_{1A}, l'écart est de 0.32 unité. Cet écart pour les autres récepteurs α_1 , α_2 et D₂ est respectivement de 0.38, 0.52 et 0.38. Sur les 20 prédictions, une seule (gras) prédit le composé dans la mauvaise classe (active, inactive).

Cette méthode du PLS donne donc des résultats de prédictions assez fiables et très proche des valeurs de « binding » expérimentales.

Prédictions de la sélectivité

Pour les études de sélectivité entre les récepteurs sérotoninergique, adrénergiques et dopaminergique, nous avons quantifié la sélectivité par la différence des valeurs de pIC₅₀. Par exemple, pour le composé **TH40**, les valeurs expérimentales 5-HT_{1A} et α_1 sont respectivement 8.2 et 6.4 unité de pIC₅₀. La sélectivité 5-HT_{1A} par rapport à α_1 de ce composé sera de 1.8.

Nous avons utilisé les méthodes de régression multiple et de partial least square. Les meilleurs résultats des prédictions sont recensés dans les tableaux 47 et 48.

RM	5-HT _{1A} par rapport à α_1		5-HT _{1A} par rapport à α_2		5-HT _{1A} par rapport à D ₂	
Composés	Champ Lipophile		Champ Lipophile		Champ Lipophile	
	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.
TH40	1.8	1.1	0.9	1	1.4	1.1
TH41	1	0.9	1.7	1	1.6	1.2
TH42	1.4	1.7	2.1	0.8	1.8	1.1
TH43	1.5	1.4	1.9	1.1	1.5	1.3
TH44	-0.4	0.2	-0.4	0.7	0.5	1.1

Tableau 47. Exp. = Expérimental, Préd. = Prédiction.

PLS	5-HT _{1A} par rapport à α_1		5-HT _{1A} par rapport à α_2		5-HT _{1A} par rapport à D ₂	
Composés	Champ Lipophile		Champ Lipophile		Champ Lipophile	
	Exp.	Préd.	Exp.	Préd.	Exp.	Préd.
TH40	1.8	0.7	0.9	1.8	1.4	1.4
TH41	1	-0.4	1.7	1.4	1.6	1.8
TH42	1.4	0.3	2.1	2.2	1.8	0.9
TH43	1.5	0.5	1.9	1	1.5	1.4
TH44	-0.4	0.6	-0.4	1	0.5	1.5

Tableau 48. Exp. = Expérimental, Préd. = Prédiction.

Pour l'ensemble des prédictions, l'écart entre les valeurs expérimentales de sélectivité et les valeurs des prédictions est d'environ 0.7 unité de pIC₅₀. L'échelle des sélectivités est faible (environ -0.4 à 2.2) et en général les erreurs de prédiction sont de l'ordre de 0.7 (valeur du s). Ces données ne sont donc pas exploitables quantitativement. Pour prédire une sélectivité, il faudra donc déterminer les prédictions de l'affinité pour les deux récepteurs et à partir de ces deux valeurs en déduire la sélectivité.

Conclusion

De cette étude de l'affinité (activité) et de la sélectivité des ligands thiochromanes, ressortent plusieurs informations. On remarque tout d'abord, pour une étude sur un récepteur donné, que plusieurs modèles donnent des résultats statistiques corrects et utilisables. On voit aussi, pour les études de l'activité ou de la sélectivité, qu'en général les méthodes QSAR utilisant les indices de similarité moléculaire donnent des meilleurs résultats statistiques que les descripteurs classiques et qu'en général la méthode statistique PLS semble mieux adaptée que la régression multiple pour traiter les données. De plus, les modèles sélectionnés pour mettre en évidence la différence d'activité des composés thiochromanes sont toujours ceux utilisant des indices électrostatiques. Par contre les modèles qui expriment le mieux les différences de sélectivité entre les récepteurs sont eux, soit de type lipophile soit de type stérique. Cette information laisse donc suggérer que la différence d'activité des composés de cet échantillon soit liée à des phénomènes électrostatiques et que la sélectivité de ces mêmes composés vis-à-vis des récepteurs 5-HT_{1A}, α_1 , α_2 et D₂ est elle, en relation directe avec des phénomènes stériques et/ou lipophiles.

Les résultats des prédictions de l'activité sont résumés dans les tableaux 49 et 50.

Cps	Activité 5-HT _{1A}					Activité α_1				
	AD	RM	PLS	Moy.	Exp.	AD	RM	PLS	Moy.	Exp.
TH40	A	8.5	7.3	A(7.9)	8.2	I	6.5	6.3	I(6.4)	6.4
TH41	A	6.5	7.8	A(7.15)	8	I	5.5	7.2	I(6.35)	7
TH42	A	6	7.8	A(6.9)	7.8	I	6.9	7.1	I(7)	6.4
TH43	A	8	8	A(8)	7.9	I	8.1	7.3	I(7.7)	6.4
TH44	I	7.3	7.2	I(7.25)	6.8	I	7.1	7.2	I(7.15)	7.2

Tableau 49. Cps = Composés ; A = Actif ; I = Inactif ; AD = analyse discriminante ; RM = régression multiple ; PLS = Partial Least Square ; Moy. = Moyenne ; Exp. = Expérimentale

Cps	Activité α_2					Activité D ₂				
	AD	RM	PLS	Moy.	Exp.	AD	RM	PLS	Moy.	Exp.
TH40	I	6	6	I(6)	7.3	I	5.8	6.9	I(6.35)	6.8
TH41	I	5.6	6.1	I(5.85)	6.3	I	5.8	6.4	I(6.1)	6.4
TH42	I	5.6	5.8	I(5.7)	5.7	I	6	7.2	I(6.6)	6
TH43	I	6.7	6.8	I(6.75)	6	I	7.1	5.9	I(6.5)	6.4
TH44	I	7.4	7	I(7.2)	7.2	I	6.2	6.2	I(6.2)	6.3

Tableau 50. Cps = Composés ; A = Actif ; I = Inactif ; AD = analyse discriminante ; RM = régression multiple ; PLS = Partial Least Square ; Moy. = Moyenne ; Exp. = Expérimentale

Pour le calcul de la moyenne, la valeur numérique est la moyenne des deux méthodes RM et PLS et la donnée qualitative correspond au classement moyen sur les trois méthodes (AD, RM, PLS) du composé, dans les classes actives ou inactives. En analysant ces deux tableaux, on voit que dans la majorité des cas les prédictions sont correctes et justes. Les seuls écarts importants sont pour les composés **TH41** et **TH42** dans la prédiction de l'activité 5-HT_{1A} et pour le composé **TH43** dans la prédiction de l'activité α_1 . Il semble donc que la prédiction de nouveaux composés soit tout à fait envisageable avec ces modèles.

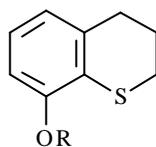
En conclusion, nous pouvons dire que cette étude sur la famille des thiochromanes a montré un réalisme certain des modèles avec un écart moyen de 0.5 unité de pIC_{50} pour l'ensemble des prédictions de l'activité 5-HT_{1A}, α_1 , α_2 et D₂. Par contre l'étude des prédictions de la sélectivité apporte seulement une information qualitative en sélectionnant statistiquement que des descripteurs de type lipophile ou stérique. Les résultats quantitatifs (prédictions) quant à eux ne sont pas exploitables. D'un point de vue qualitatif, il est intéressant aussi de rappeler que nos modèles ont montré une corrélation de l'affinité avec les descripteurs de type électrostatique et une corrélation de la sélectivité avec les descripteurs de type lipophiles et/ou stérique. Pour parfaire nos résultats, il faudrait continuer cette étude des prédictions avec d'autres composés pour avoir un « test set » plus conséquent qui apporterait plus d'information sur l'aspect prédictif de nos modèles.

Remarques

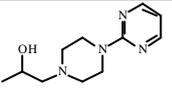
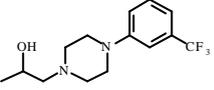
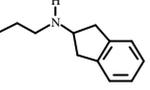
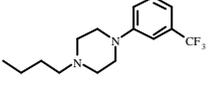
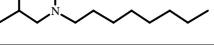
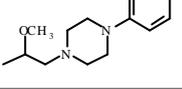
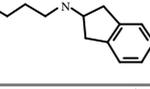
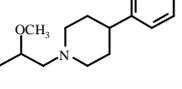
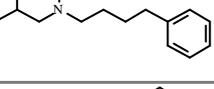
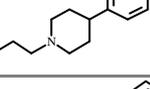
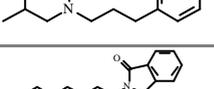
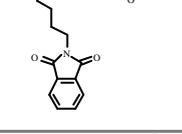
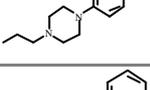
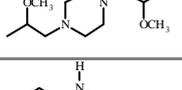
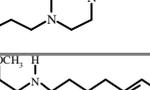
On peut sur cet ensemble faire un parallèle entre une approche du type SAR et la nôtre. Lors du calcul des prédictions de l'affinité effectués avec la régression multiple, en comparant les structures des composés **TH16**, **TH20** et **TH40**, nous pouvions nous attendre à une activité 5-HT_{1A} du ligand **TH40** plus faible que celle du **TH16**. Cette remarque prend en compte la diminution du bras espaceur aliphatique composé de CH₂. En effet, le composé **TH20** a une activité de 7.5 avec 4 CH₂, le composé **TH16** a une activité de 7.1 avec 3 CH₂ et le ligand **TH40** à 2 CH₂. On aurait donc pu s'attendre à une baisse d'activité en relation avec une diminution du bras espaceur entre les deux entités bicycliques. Ceci n'est pas le cas et les résultats de l'analyse QSAR vont dans ce sens avec une prédiction d'affinité 5-HT_{1A} de 7.9 unités de pIC_{50} pour une activité biologique de 8.2. Ce résultat est donc encourageant et vérifie que notre modèle QSAR ne réagit pas de façon simpliste. Nous retrouvons le même cas de figure avec les composés **TH42**, **TH37** et **TH33** mais cette fois dans le sens inverse. Une diminution du bras espaceur devait amener

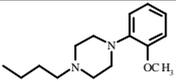
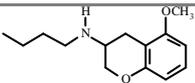
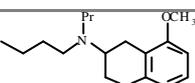
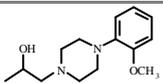
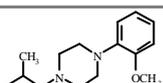
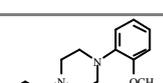
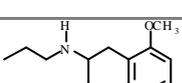
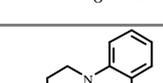
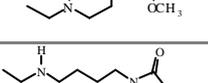
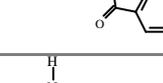
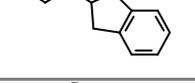
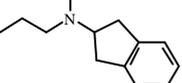
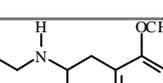
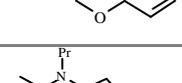
à une augmentation de l'affinité comme pour le cas précédent. Ce ne fut pas le cas puisque l'activité du ligand **TH42** qui est de 7.8 est plus faible que celles des composés **TH33** et **TH37** qui sont respectivement de 8.2 et 8.7. Là encore, notre modèle a correctement positionné le **TH42** dans la série avec une prédiction moyenne de 6.9. C'est l'exemple de l'intérêt des approches statistiques qui prennent en compte la totalité de l'information et ne restent pas conditionnée par le phénomène le plus visible.

ANNEXE 4



Composés	R	pIC ₅₀ (M)	pIC ₅₀ (M)	pIC ₅₀ (M)	pIC ₅₀ (M)
		5-HT _{1a}	a ₁	a ₂	D ₂
TH1		4.0	nd	nd	nd
TH2		4.0	4.0	4.0	4.0
TH3		4.0	nd	nd	nd
TH4		4.0	4.9	4.0	4.0
TH5		4.0	4.0	4.0	4.0
TH6		4.0	4.0	4.0	4.0
TH7		4.0	nd	nd	nd
TH8		4.4	nd	nd	nd
TH9		5.4	nd	nd	nd
TH10		5.5	nd	nd	nd
TH11		5.5	10.0	5.6	4.9
TH12		5.6	nd	nd	nd

TH13		6.0	nd	nd	nd
TH14		6.8	nd	nd	nd
TH15		6.9	nd	nd	nd
TH16		7.1	6.2	5.7	5.7
TH17		7.1	7.1	6.1	6.5
TH18		7.4	5.5	5.3	5.3
TH19		7.4	7.2	7.7	6.6
TH20		7.5	6.9	6.0	5.6
TH21		7.6	7.5	8.0	6.7
TH22		7.6	6.2	5.9	5.5
TH23		7.7	7.6	8.3	6.7
TH24		7.7	6.1	5.7	5.4
TH25		7.7	8.4	4.0	6.3
TH26		7.8	7.8	7.6	6.1
TH27		7.9	8.5	7.9	7.7
TH28		7.9	5.2	5.4	5.4
TH29		8.1	8.2	8.3	6.7
TH30		8.2	7.0	6.0	6.2

TH31		8.2	9.0	7.9	7.8
TH32		8.2	6.7	5.3	5.9
TH33		8.2	5.6	4.8	5.7
TH34		8.3	5.2	7.1	7.5
TH35		8.4	8.2	8.0	8.5
TH36		8.6	9.5	9.0	8.0
TH37		8.7	6.6	5.2	6.1
TH38		8.9	9.7	9.0	8.0
TH39		9.0	8.4	6.6	7.2
TH40		(8.2)	(6.4)	(7.3)	(6.8)
TH41		(8)	(7)	(6.3)	(6.4)
TH42		(7.8)	(6.4)	(5.7)	(6)
TH43		(7.9)	(6.4)	(6)	(6.4)
TH44		(6.8)	(7.2)	(7.2)	(6.3)

nd = non déterminé

V. ENSEMBLE DES 382 LIGANDS

La composition de l'échantillon (annexe 5) est de 382 composés (**E1 à E382**) avec une gamme d'activité biologique (pIC_{50}) comprise entre 4 et 11. Cet échantillon est séparé en trois sous groupes ; un groupe de 166 composés inactifs ayant un $pIC_{50} < 7.5$, un groupe de 161 composés actifs avec un $pIC_{50} \geq 7.5$ (« training set ») et un groupe de 55 composés à prédire (« test set »). Cet ensemble de ligands regroupe la totalité des composés des familles précédemment étudiées que sont les familles indoles, tétralines, chromanes et thiochromanes. Des nouveaux composés (agonistes, antagonistes) ont été ajoutés à cet ensemble soit dans le « training set » soit dans le « test set ». Pour les familles des indoles et des tétralines tous les ligands ont été introduits dans le « training set ». Pour les familles des chromanes et des thiochromanes les composés prédits précédemment feront partie avec d'autres ligands du « test set » de cette étude sur l'ensemble des composés. Cet échantillon a seulement fait l'objet d'une étude des descripteurs classiques et d'une étude des méthodes d'analyses statistiques que sont l'analyse discriminante, la régression multiple et le Partial Least Squares. L'intérêt de cet échantillon est multiple. Il présente une gamme très diversifiée de composés avec de nombreuses différences structurales. Dans cette étude, après avoir sélectionné les meilleures conditions statistiques de chaque méthode, nous nous sommes surtout intéressés à l'aspect prédictif des modèles afin de comparer les résultats avec les études menées précédemment sur les différentes familles. Les composés utilisés dans cette étude, en plus des composés précédemment étudiés dans les études sur les familles tétralines, indoles, chromanes et thiochromanes, sont issus de publications diverses⁴³⁻⁵⁵.

Etude des descripteurs classiques

Les descripteurs utilisés sont ceux proposés par le logiciel TSAR ; ils comprennent les indices de connectivité, de forme et de flexibilité, les indices topologiques, les propriétés moléculaires et les autocorrélogrammes utilisés avec un pas maximum de 12 pour les vecteurs 2D et 3D. Tous ces descripteurs forment un crible de 305 variables.

Une étude d'analyse discriminante sur l'ensemble des 382 ligands a permis de séparer les deux classes de cet échantillon à 83% avec treize variables, où 90% des composés inactifs sont bien classés et 75% des composés actifs sont classés correctement (Tableau 51).

Variables	Total	Inactifs	Actifs
V1 : 2D bin 11 (Electronégativité)	71 %	80 %	60 %
V2 : Kier chi3 Cluster	73 %	82 %	63 %
V3 : 2D bin 7 (Charges Q négatives)	74 %	83 %	64 %
V4 : 3D bin 9 (Charges Q positives)	76 %	84 %	66 %
V5 : 2D bin 11 (Charges Q positives)	77 %	86 %	67 %
V6 : Kier chi4 Path	78 %	86 %	69 %
V7 : 2D bin 1 (Rayon de Van der Waals)	79 %	88 %	69 %
V8 : 3D bin 8 (non pondéré)	80 %	86 %	72 %
V9 : 2D bin 4 (Charges Q négatives)	80 %	86 %	73 %
V10 : 2D bin 4 (Valence)	80 %	86 %	74 %
V11 : 2D bin 6 (Charges Q positives)	81 %	87%	75 %
V12 : 2D bin 7 (Valence)	82 %	88 %	74 %
V13 : 2D bin 11 (Charges Q négatives)	83 %	90 %	75 %
$r(CV)^2$	67 %	77 %	57 %

Tableau 51.

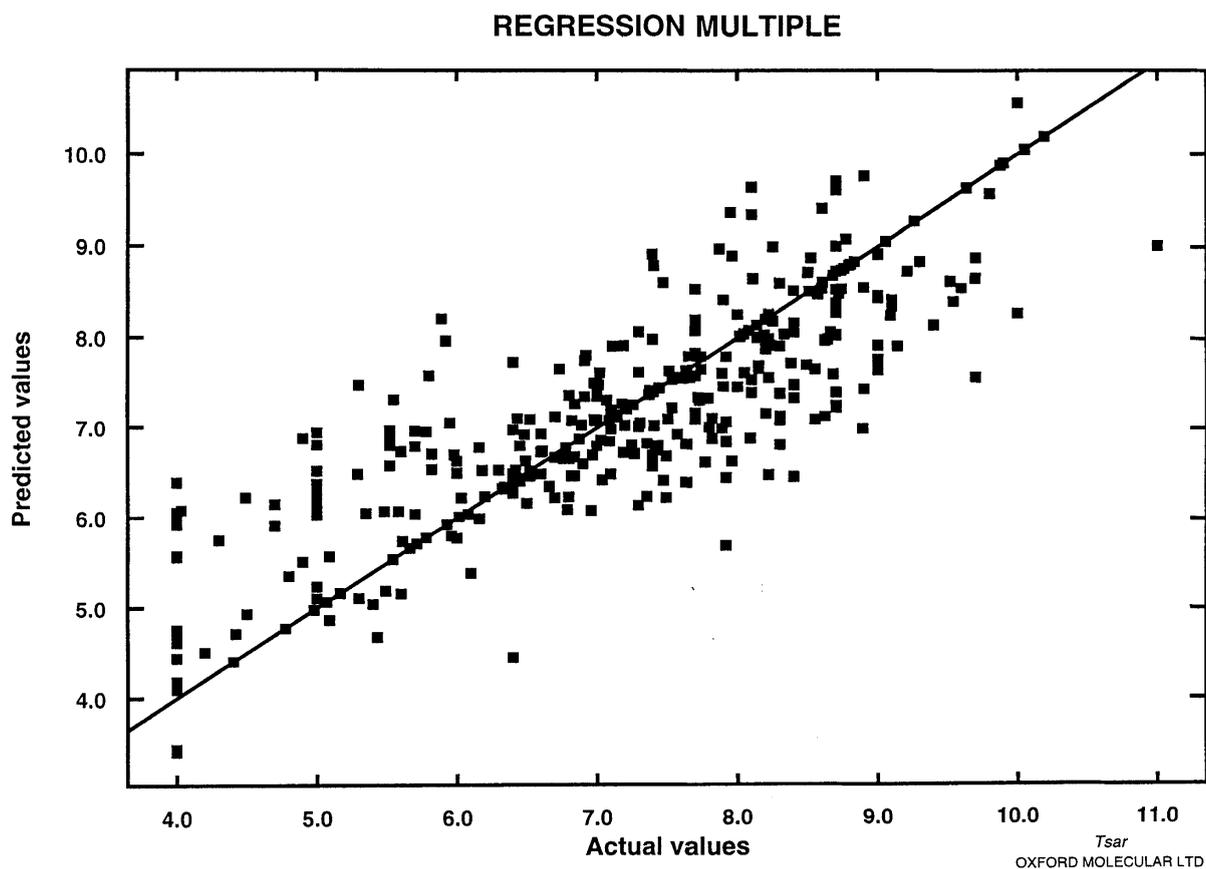
Le résultat de cette étude d'analyse discriminante semble assez encourageant avec une discrimination totale de 83 % et une prédictivité de 67 %.

La méthode d'analyse de régression multiple sur les 382 composés (Graphe 19) donne les résultats suivant ; le coefficient de régression est $r^2 = 0.72$ et $r = 0.83$, le « cross validation » qui donne le pouvoir prédictif de l'équation est de 0.44. La probabilité de F est $2.5 \cdot 10^{-40}$ et la valeur de s est 0.927. Cette valeur de s donne un ordre d'idée de l'erreur de prédiction des ligands lors de cette analyse de régression multiple. L'équation utilise vingt variables qui sont des variables d'autocorrélogrammes à 2 et 3 dimensions, des indices de Kier et de Wiener. Les variables sont V1 = 2D bin 1 (charges Q négatives), V2 = 3D bin 4 (Numéro atomique), V3 = 3D bin 4 (charges Q positives), V4 = 3D bin 9 (charges Q positives), V5 = 2D bin 4 (Log P partie hydrophile), V6 = 2D bin 3 (Log P partie hydrophile), V7 = 2D bin 3 (Log P), V8 = Kier chi3 cluster, V9 = Wiener H inclus , V10 = 2D bin 11 (Log P), V11 = 2D bin 11 (Charges Q), V12 = 2D bin 7 (Log P partie lipophile), V13 = 2D bin 7 (Valence), V14 = 2D bin 10 (Log P), V15 = 3D bin 11 (Valence), V16 = 3D bin 10 (Charges Q négatives), V17 = 3D bin 5 (Charges Q positives), V18 = 2D bin 7 (Charges Q), V19 = 2D bin 6 (Charges Q négatives) et V20 = 2D bin 6 (Charges Q).

Equation

$$Y = 0.727 (V8) - 1.898 (V9) + 0.769 (V3) - 0.419 (V17) + 0.922 (V4) - 0.368 (V16) - 0.916 (V2) - 0.267 (V11) + 1.410 (V20) + 0.449 (V18) + 0.941 (V1) - 0.339 (V19) - 1.879 (V13) + 1.083 (V15) - 1.166 (V7) + 0.424 (V10) + 2.480 (V6) - 2.362 (V5) + 0.509 (V12) - 0.289 (V14) + 7.144$$

L'ensembles des variables constituant cette équation est représentatif d'effets stériques, électrostatiques et lipophiliques.



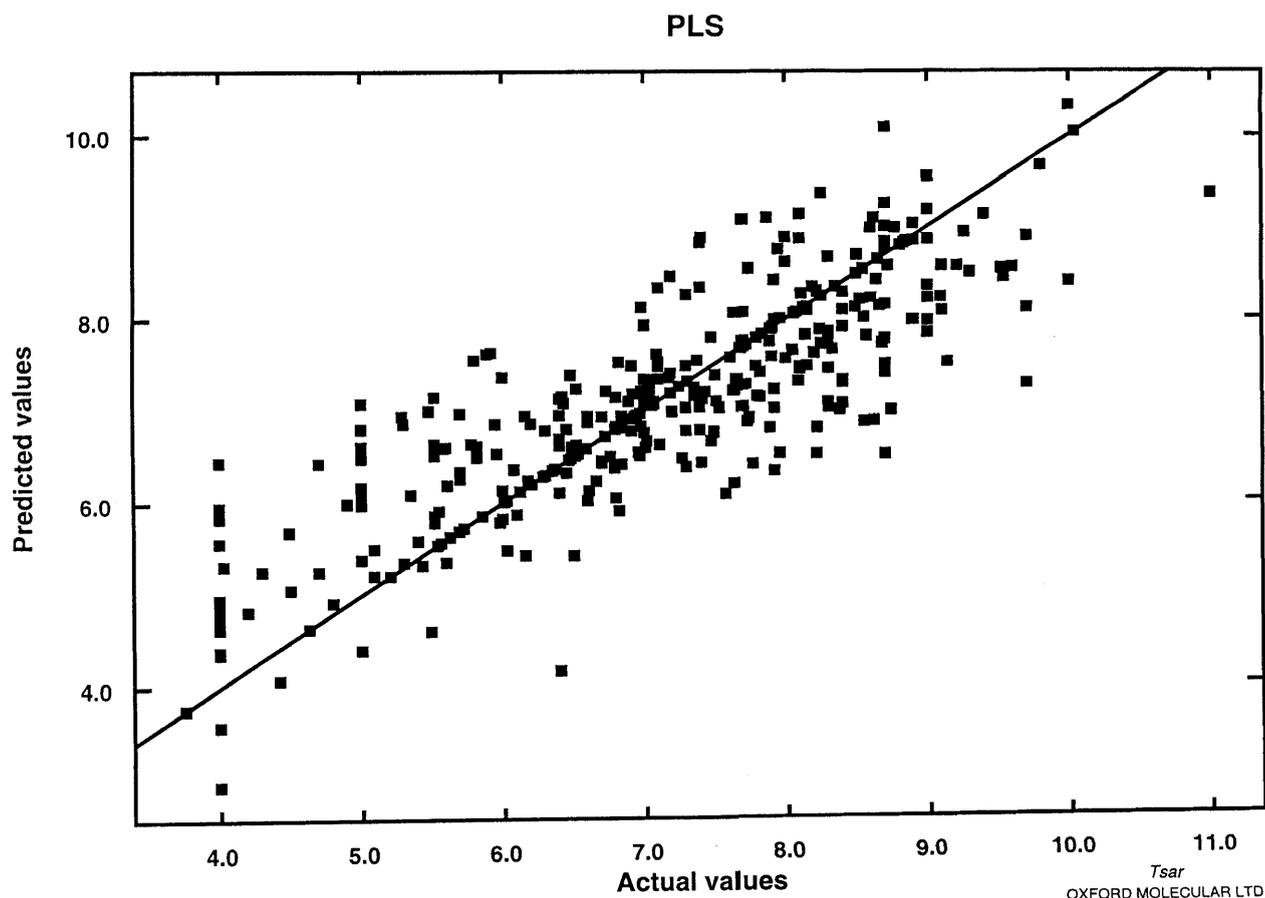
Grphe 19.

L'analyse Partial Least Square effectuée sur la totalité des 382 ligands donne les résultats suivants (Tableau 52) (Grphe 20).

Vecteurs	1	2	3	4	5	6	7	8	9	10
Signification statistique	0.970	1.016	1.131	1.152	1.250	1.402	1.502	1.609	1.692	1.778
PRESS	291.1	283.4	278.4	253.6	244.5	221.0	212.0	225.8	232.9	243.9
$r(CV)^2$	0.060	0.052	0.069	0.152	0.250	0.362	0.423	0.290	0.113	0.050
r^2	0.181	0.372	0.461	0.534	0.686	0.757	0.819	0.891	0.916	0.963

Tableau 52.

Le nombre de vecteurs que nous avons sélectionnés est de 7. Ces vecteurs correspondent à la première valeur de PRESS faible (212.0). Le coefficient de régression est de 0.82 et le « cross validation » est de 0.42.



Graph 20.

Une remarque peut être faite sur les graphes 18 et 19. Pour ces analyses RM et PLS, les graphes ne présentent pas le phénomène précédemment vu dans les études QSAR effectuées sur les familles. C'est à dire une meilleure corrélation des ligands à haute activité par rapport aux composés à faible activité. Nous pouvons expliquer ceci par le fait que, dans cet ensemble de 382 ligands, à la différence des échantillons des familles, les composés à haute activité diffèrent beaucoup plus suivant différents critères électrostatiques, stériques et lipophiliques. Cette remarque est à relier à l'existence potentielle de plusieurs sites actifs au niveau du récepteur, en partie ceux spécifiques des agonistes, agonistes partiels et antagonistes.

Ceci montre que cet ensemble de 382 ligands apporte beaucoup plus d'informations pour les composés actifs que ne le fait un échantillon sur une famille donnée. L'ensemble de ces trois méthodes (AD, RM, PLS) donne des résultats statistiques fiables qui permettent d'envisager une étude de prédiction de composés dont on connaît l'activité afin de vérifier la validité des modèles. A partir de ces

modèles, nous avons donc voulu prédire l'activité 5-HT_{1A} de composés divers comme par exemple le sumatriptan⁹ et des 14 composés chromanes (**C49** à **C57**) et thiochromanes (**TH40** à **TH44**) précédemment étudiés.

Etude des prédictions 5-HT_{1A}

Les résultats des prédictions 5-HT_{1A} pour les 55 composés sont présentés dans le tableau 53.

Composés	pIC₅₀	Prédiction Analyse Discriminante	Prédiction Régression Multiple	Prédiction Partial Least Square	Moyenne Qualitative
E111	8.60	Inactif	8.02	7.89	A
E112	7.50	Inactif	8.20	7.79	A
E113	9.69	Actif	8.76	8.85	A
E114	8.62	Actif	6.80	6.35	I
E115	8.88	Actif	7.21	7.83	A
E116	8.25	Actif	8.74	8.84	A
E117	7.85	Actif	6.87	7.07	I
E118	7.80	Inactif	7.09	7.91	I
E119	8.50	Inactif	9.13	8.65	A
E120	8.18	Actif	8.83	8.49	A
E121	7.72	Actif	6.02	6.53	I
E122	9.30	Actif	8.68	6.45	A
E123	7.60	Inactif	10.05	8.12	A
E124	7.78	Inactif	7.55	7.72	A
E125	7.90	Actif	7.57	7.68	A
E126	8.00	Actif	7.66	8.51	A
E127	9.03	Actif	7.78	6.29	A
E128	9.10	Inactif	8.14	8.82	A
E129	8.78	Actif	7.74	7.79	A
E130	8.70	Actif	8.08	7.06	A
E131	9.23	Actif	8.07	5.20	A
E132	9.05	Actif	7.26	9.21	A

E133	9.52	Actif	5.54	5.55	I
E134	10.0	Actif	8.56	5.56	A
E135	5.48	Inactif	6.43	6.59	I
E136	5.70	Inactif	5.93	6.95	I
E137	5.85	Inactif	3.71	5.69	I
E138	6.04	Inactif	8.01	6.12	I
E139	6.08	Inactif	6.32	3.75	I
E140	6.15	Inactif	7.41	6.35	I
E141	6.52	Inactif	7.51	6.97	I
E142	6.62	Inactif	6.89	8.06	I
E143	6.64	Actif	6.40	5.72	I
E144	6.67	Inactif	9.87	7.19	I
E145	6.69	Inactif	7.11	6.93	I
E146	6.71	Inactif	8.20	6.72	I
E147	7.02	Inactif	5.17	6.37	I
E148	7.03	Inactif	6.34	6.81	I
E149	7.14	Inactif	6.44	6.01	I
E150	7.20	Inactif	6.52	6.20	I
E151	7.40	Inactif	6.77	7.68	I
C49	8.05	Actif	7.15	6.89	I
C50	6.40	Inactif	5.21	7.56	I
C51	7.15	Actif	8.51	7.01	A
C52	7.20	Actif	7.56	7.98	A
C53	8.20	Actif	6.95	7.01	I
C54	5.38	Inactif	5.71	7.98	I
C55	5.52	Inactif	7.52	7.62	A
C56	4.80	Inactif	6.20	7.86	I
C57	4.80	Inactif	7.01	5.26	I
TH40	8.20	Actif	7.05	6.95	I
TH41	8.0	Actif	6.10	7.29	I
TH42	7.80	Inactif	7.81	7.78	A
TH43	7.90	Actif	8.96	5.20	A
TH44	6.80	Inactif	8.69	8.26	A

Tableau 53.

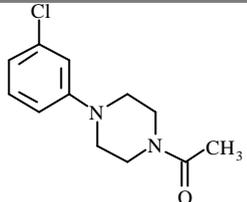
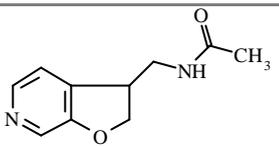
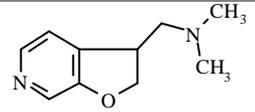
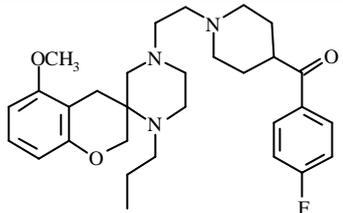
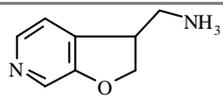
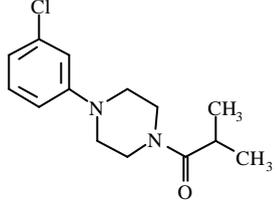
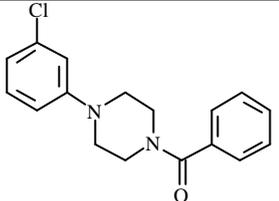
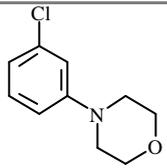
Pour l'analyse discriminante, seulement 10 composés sur 55 sont mal prédits. Pour l'analyse de régression multiple 29 composés sont prédits avec un écart inférieur à 1 unité de pIC_{50} , 21 composés le sont avec un écart compris entre 1 et 2 et 5 composés avec un écart supérieur à 2 unités. Pour l'analyse PLS, 32 composés sont prédits avec un écart inférieur à 1 unité de pIC_{50} , 12 composés le sont avec un écart compris entre 1 et 2 et 11 composés avec un écart supérieur à 2 unités. L'ensemble de ces prédictions (RM et PLS) est donc relativement moyen puisque environ 50% des composés à prédire le sont avec un écart de prédiction supérieur à 1 unité de pIC_{50} . Par contre pour l'analyse discriminante le pourcentage de composés bien prédit est de 82%. Si l'on fait une moyenne qualitative des prédictions de ces trois méthodes (voir tableau 53), le pourcentage chute à 76%.

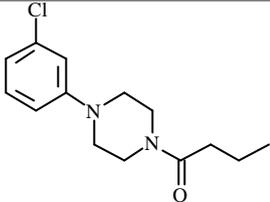
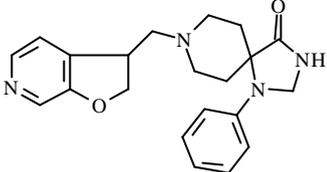
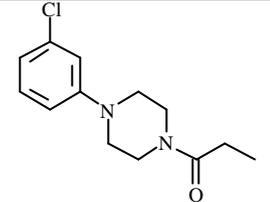
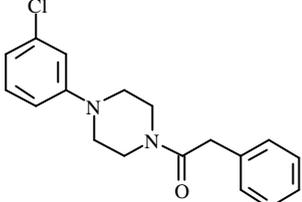
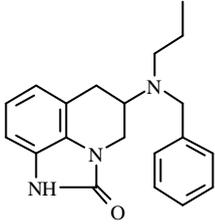
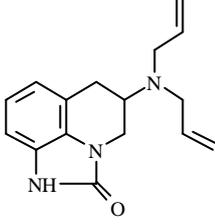
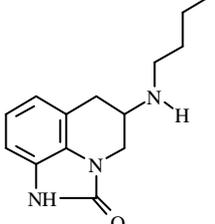
En admettant que ces 55 composés ne soient pas encore synthétisé et testé, et qu'à partir de cette étude QSAR on décide de synthétiser tel ou tel composés. Sur les 55 ligands, 25 prédits actifs auraient été synthétisés dont seulement 4 inactifs (**C51**, **C52**, **C55** et **TH44**). Nous aurions perdu dans les 30 produits non synthétisés, 9 ligands actifs.

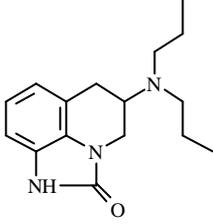
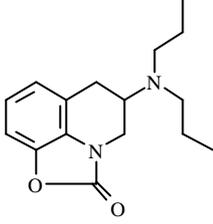
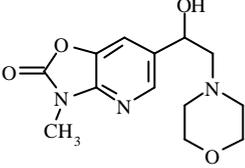
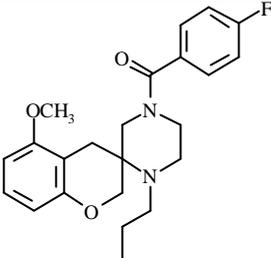
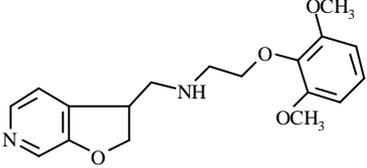
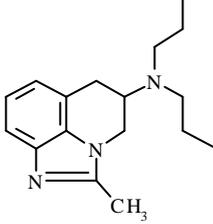
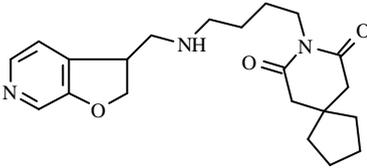
Maintenant, si on compare seulement les résultats de prédiction pour les composés **C49-C57** et **TH40-TH44**, obtenus dans cette étude sur l'ensemble des ligands et dans les études sur les familles séparées, nous pouvons dire que ces résultats sont plutôt meilleurs pour les études sur les familles étudiées seules. Pour les composés de type chromane (**C49-C57**), la moyenne des écarts de prédictions pour l'étude sur la famille seule est de 0.6 unité de pIC_{50} , alors que pour l'étude sur la totalité des ligands, la moyenne de ces écarts est d'environ 1.2 (RM) à 1.4 (PLS) unités. Pour les composés, de type thiochromane, cette différence existe aussi avec comme valeurs moyennes 0.8 et 0.3 pour l'étude sur la famille seule, et comme valeurs moyennes 1.2 (RM et PLS) unités de pIC_{50} pour l'étude sur l'ensemble des composés. On retrouve donc ici un résultat classique ; les prédictions sont donc plus proches de la réalité expérimentale lorsque l'on effectue les études QSAR sur les familles plutôt que sur l'ensemble hétérogène.

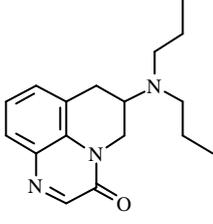
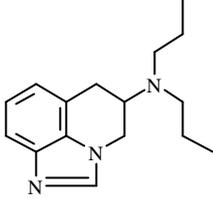
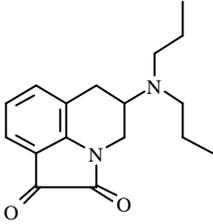
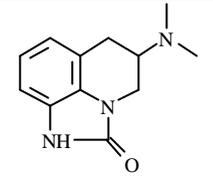
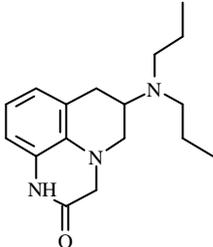
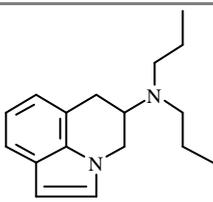
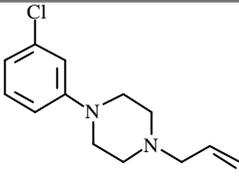
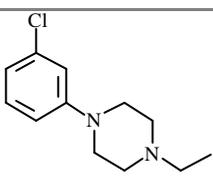
Les études effectuées, comme pour ce modèle, sur un ensemble important et hétérogène ont l'inconvénient d'une moins bonne précision de prédiction. Par contre, elles ont l'avantage majeur d'être plus largement exploitables et ainsi de permettre l'apparition de nouvelles classes de composés (« lead compound »).

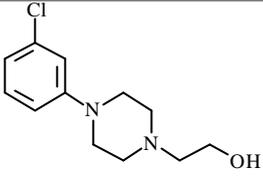
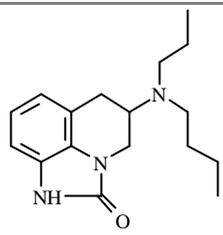
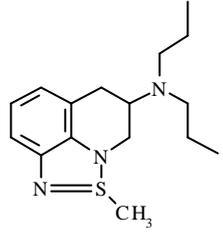
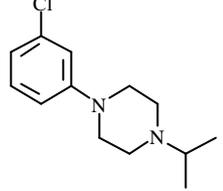
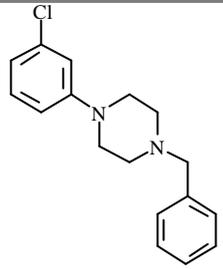
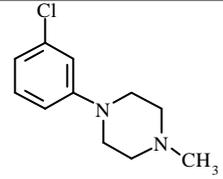
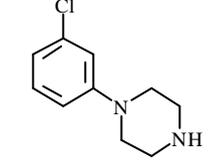
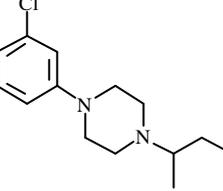
ANNEXE 5

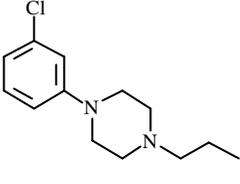
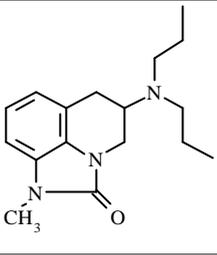
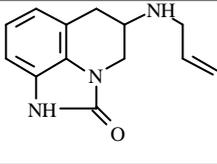
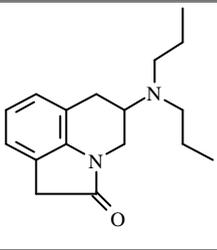
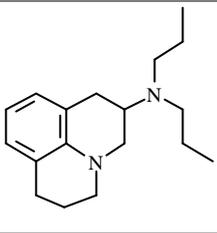
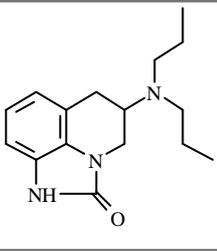
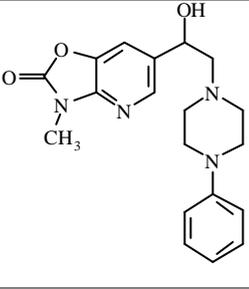
Composés	Structure	pIC ₅₀ (M)
E1		4.0
E2		4.0
E3		4.0
E4		4.0
E5		4.0
E6		4.0
E7		4.0
E8		4.0

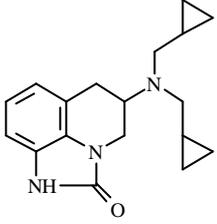
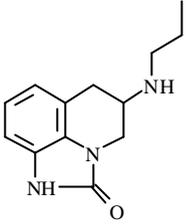
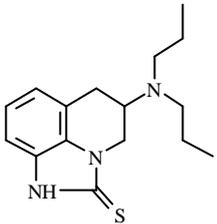
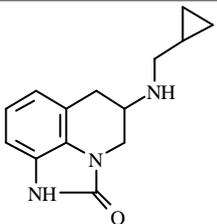
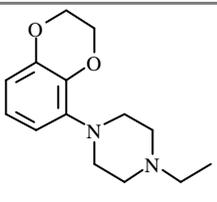
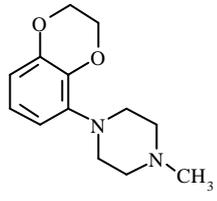
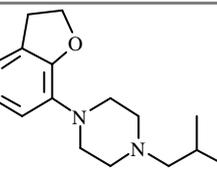
E9		4.30
E10		4.50
E11		4.50
E12		4.80
E13		5.0
E14		5.0
E15		5.0

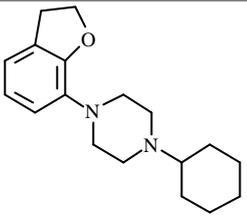
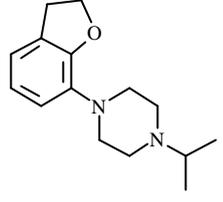
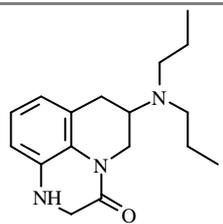
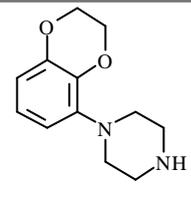
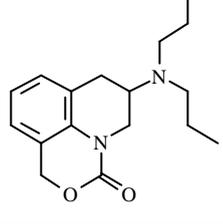
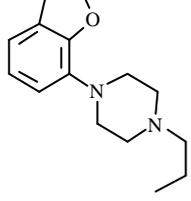
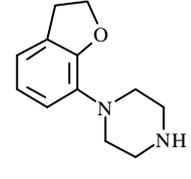
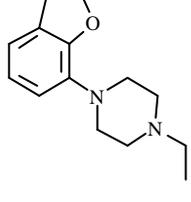
E16		5.0
E17		5.0
E18		5.10
E19		5.10
E20		5.30
E21		5.35
E22		5.55

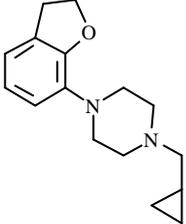
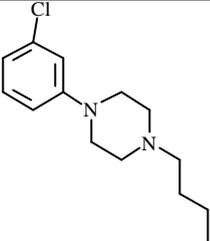
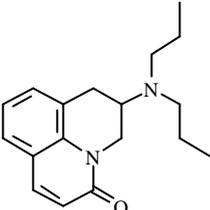
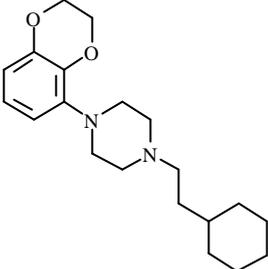
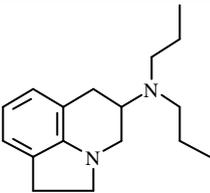
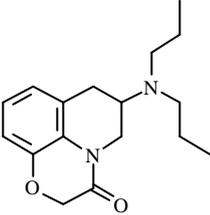
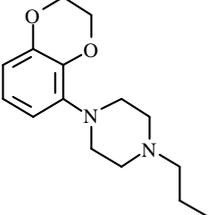
E23		6.0
E24		6.10
E25		6.20
E26		6.45
E27		6.50
E28		6.60
E29		6.60
E30		6.70

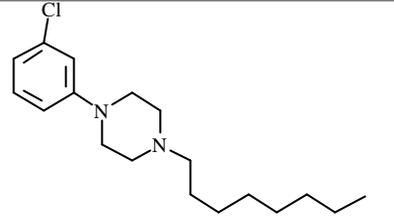
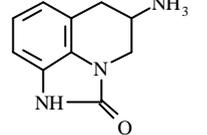
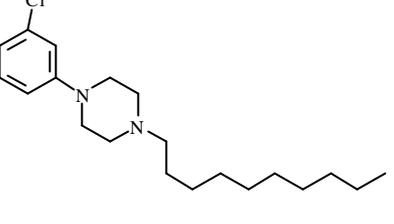
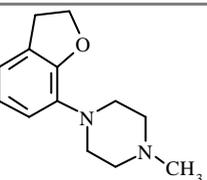
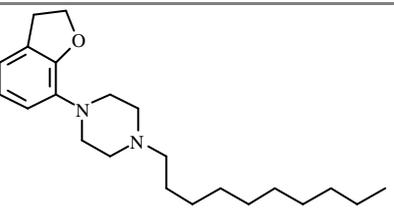
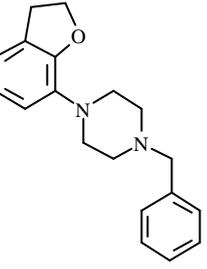
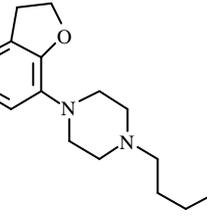
E31		6.70
E32		6.70
E33		6.80
E34		6.80
E35		6.80
E36		6.80
E37		6.85
E38		6.90

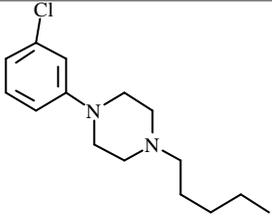
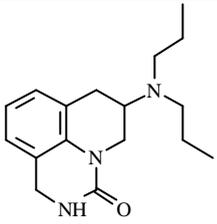
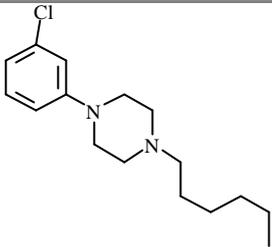
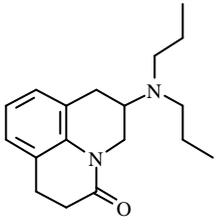
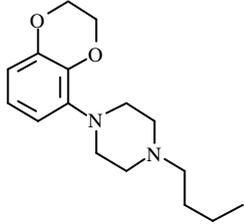
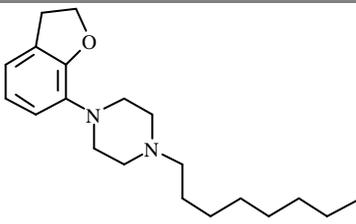
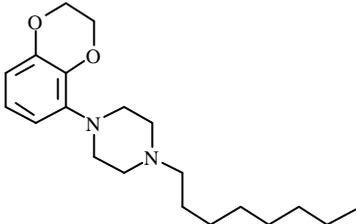
E39		6.90
E40		6.90
E41		6.96
E42		6.97
E43		6.98
E44		7.0
E45		7.01

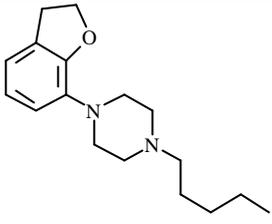
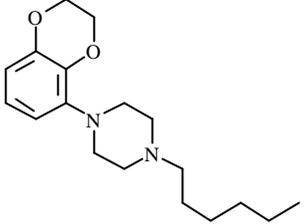
E46		7.02
E47		7.03
E48		7.04
E49		7.07
E50		7.10
E51		7.18
E52		7.19

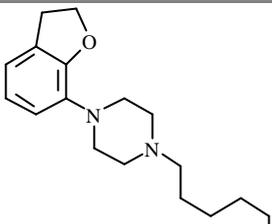
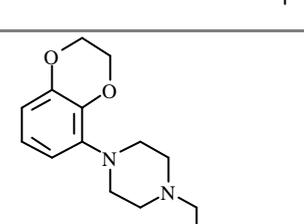
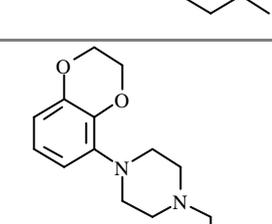
E53		7.20
E54		7.25
E55		7.31
E56		7.40
E57		7.51
E58		7.54
E59		7.62
E60		7.64

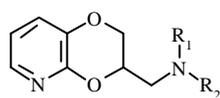
E61		7.70
E62		7.73
E63		7.77
E64		7.87
E65		7.90
E66		7.92
E67		7.93

E68		7.95
E69		7.96
E70		7.97
E71		8.09
E72		8.10
E73		8.11
E74		8.23

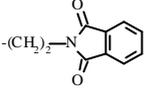
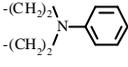
E75		8.25
E76		8.49
E77		8.57
E78		8.62
E79		8.66
E80		8.77
E81		9.0

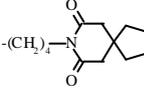
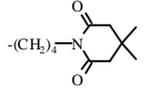
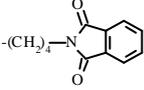
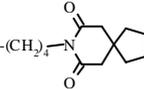
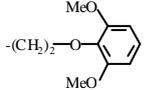
E82		9.10
E83		9.21

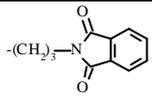
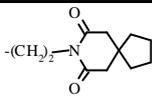
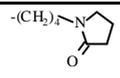
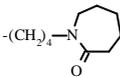
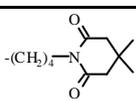
E84		9.26
E85		9.30
E86		9.60

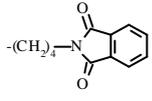
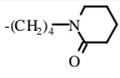
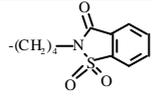
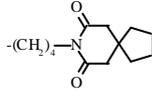


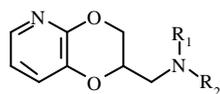
Composés	R₁	R₂	pIC₅₀ (M)
E87	H	Ac	5

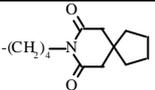
E88	H		5.7
E89	<i>n</i> -Pr	<i>n</i> -Pr	5.7
E90	Et	Et	5.7
E91	-(CH ₂) ₅ -		6
E92			6.4
E93	Me	Ac	6.4

E94	H	H	6.5
E95	H	Me	6.6
E96	H		7.4
E97	H		7.5
E98	Me		7.7
E99	Me		7.9
E100	H		8

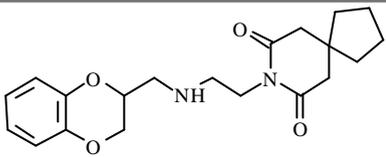
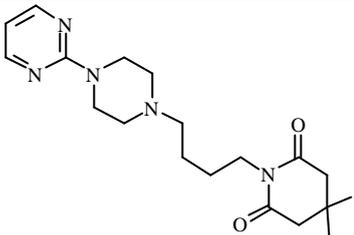
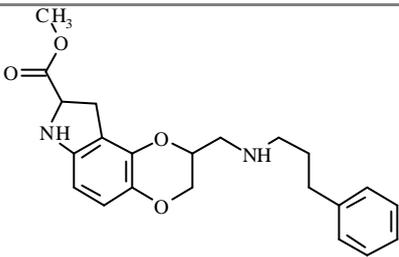
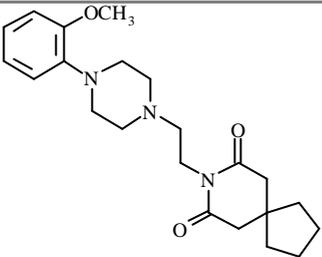
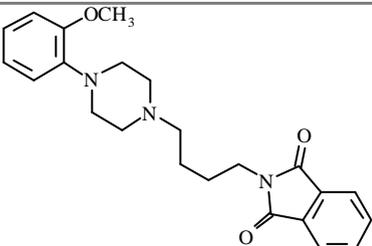
E101	H		8
E102	H		8.4
E103	H		8.6
E104	H		8.7
E105	H		8.9

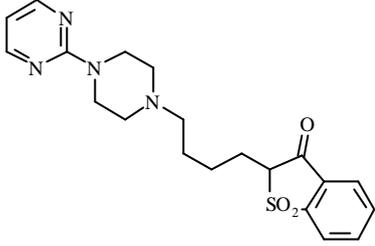
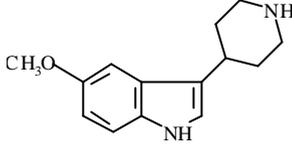
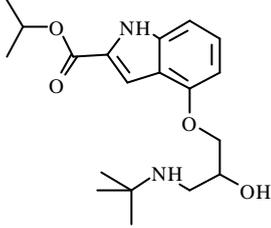
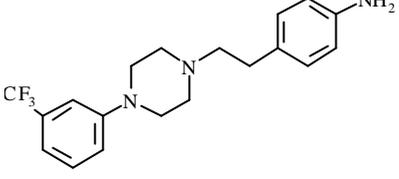
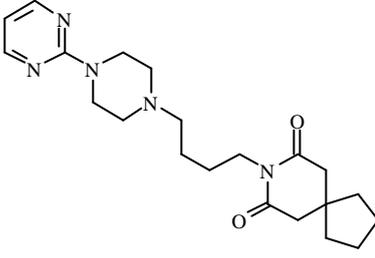
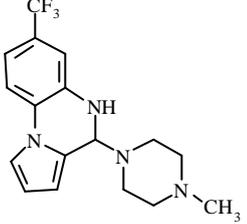
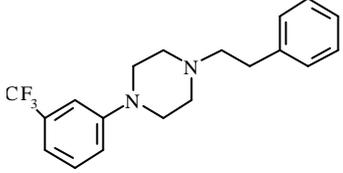
E106	H		9.1
E107	H		9.1
E108	H		9.8
E109	H		11

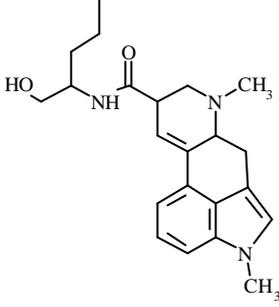
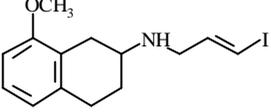
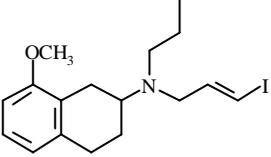
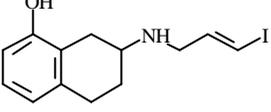
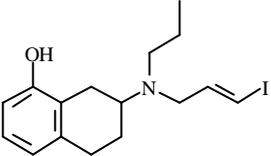
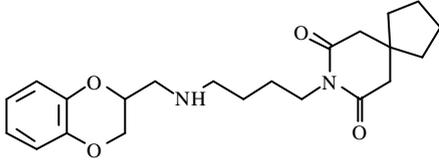
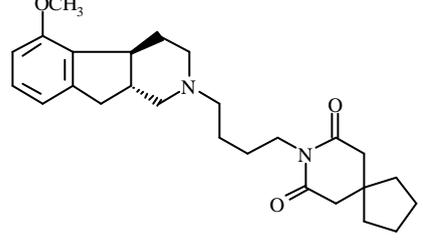
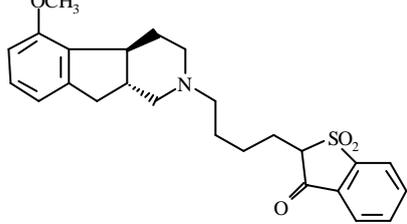


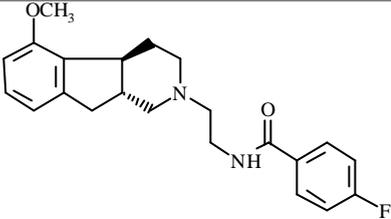
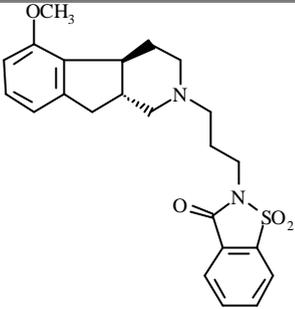
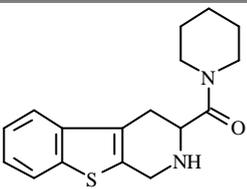
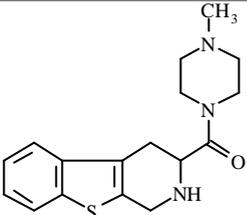
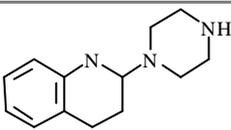
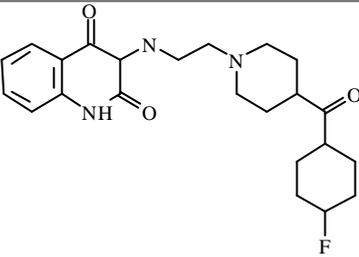
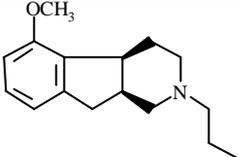
E110	H		7.4
-------------	---	---	-----

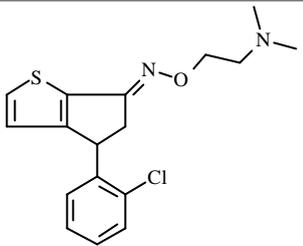
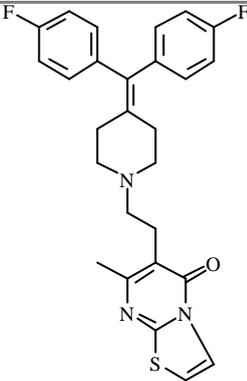
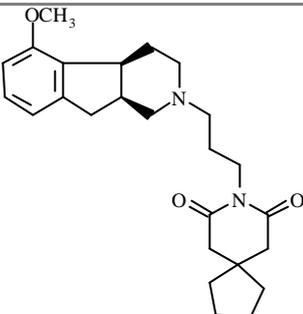
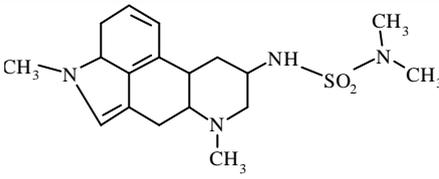
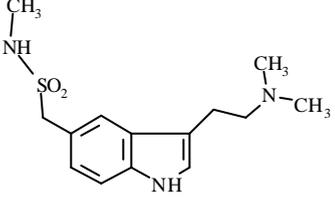
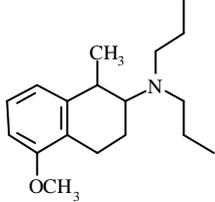
Composés à prédire

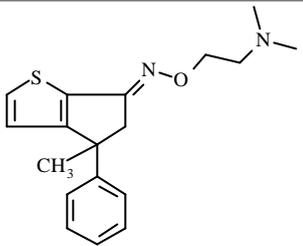
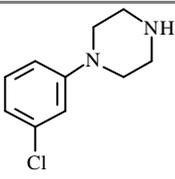
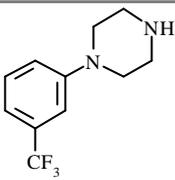
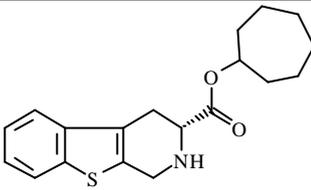
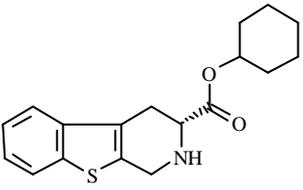
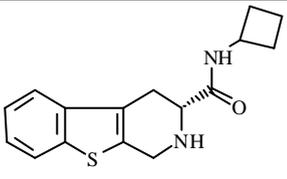
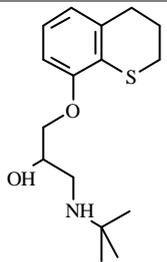
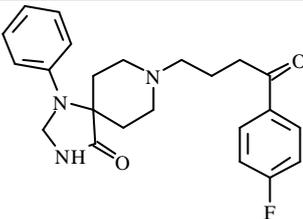
E111		(8.60)
E112		(7.50)
E113		(9.70)
E114		(8.62)
E115		(8.88)

E116		(8.25)
E117		(7.85)
E118		(7.80)
E119		(8.50)
E120		(8.18)
E121		(7.72)
E122		(9.30)

E123		(7.60)
E124		(7.78)
E125		(7.90)
E126		(8.0)
E127		(9.03)
E128		(9.10)
E129		(8.78)
E130		(8.70)

E131		(9.23)
E132		(9.05)
E133		(9.52)
E134		(10.0)
E135		(5.48)
E136		(5.70)
E137		(5.85)

E138		(6.04)
E139		(6.08)
E140		(6.15)
E141		(6.52)
E142		(6.62)
E143		(6.64)

E144		(6.67)
E145		(6.69)
E146		(6.71)
E147		(7.02)
E148		(7.03)
E149		(7.14)
E150		(7.20)
E151		(7.40)

VI. CONCLUSION ; BILAN DES PREDICTIONS

A la fin de cette étude, on peut tirer les conclusions suivantes :

- En comparant les résultats obtenus avec les échantillons et avec l'ensemble hétérogène de 382 ligands, on peut dire d'après les résultats obtenus que les prédictions sont plus précises sur les échantillons. Par contre, une telle étude QSAR sur un ensemble hétérogène a l'avantage majeur d'être plus largement exploitable et ainsi de permettre l'apparition de nouvelles classes de composés (« lead compound »).

- Les études sur les échantillons de familles ont donné des résultats statistiques et prédictifs correct pour les indoles, les chromanes et les thiochromanes. Pour contre l'étude QSAR sur la famille des tétralines n'a pas donné les résultats escomptés. Ceci montre l'importance que l'on doit accorder à l'homogénéité relative qu'il doit exister entre l'échantillon (« training set ») et les composés à prédire (« test set »). La principale limite d'une étude QSAR est donc évidente : le mécanisme de la conception est limité seulement à la combinaison de ce qui existe.

- En ce qui concerne les différents descripteurs utilisés dans les études QSAR, les indices de similarité moléculaire tridimensionnels donnent en général de meilleurs résultats statistiques que les descripteurs classiques.

- Si l'on compare les méthodes utilisées pour analyser les données, nous dirons que la méthode la mieux adaptée et qui donne en général les meilleurs résultats statistiques, est le Partial Least Square. Cette méthode semble plus adaptée à traiter un nombre de variables (descripteurs) important. L'analyse discriminante donne aussi de bons résultats en général mais n'apporte qu'une information qualitative sur la prédiction de nouveaux composés.

Ces remarques concernent l'aspect méthodologique de notre étude. Notre travail pose aussi des questions d'ordre fondamentales et pratiques.

- Les méthodes statistiques ne donnent pas toujours de bons résultats. Ceci montre qu'une analyse statistique n'est qu'un outil de visualisation, et que par ce fait l'échantillon de départ joue un rôle prépondérant dans la qualité et la validité des résultats obtenus. Nous sommes donc limités par la composition de l'échantillon de départ.

- Sommes-nous limités par la qualité des données biologiques?

Lorsque l'on prend en compte les données biologiques (IC_{50}) dans une étude QSAR, il n'est pas rare de trouver dans la littérature plusieurs valeurs pour un composé donné. Cette approximation, si petite soit-elle, se répercute obligatoirement sur la qualité des prédictions données par les modèles QSAR.

- Les indices de similarité moléculaire ne donnent pas d'informations qualitatives sur la position tridimensionnelle des parties communes ou des parties différentes entre les composés actifs et inactifs. Cet outil n'est donc pas pour l'instant une aide à la conception de nouveaux composés à activité potentielle. Le travail et l'imagination du chimiste sont donc prépondérants dans la conception de nouveaux ligands.

- Doit-on accorder une importance particulière au choix du seuil d'activité pour les analyses discriminantes? Et quel poids statistique accorder aux molécules dont l'activité est proche de ce seuil?

Le choix du seuil est délicat. Pour nos études, nous avons fixé ce seuil comme étant la valeur moyenne des activités extrêmes. Si l'on élimine les composés proches du seuil, on perd une information et surtout on diminue l'effectif de l'échantillon, ce qui n'est forcément avantageux pour une étude QSAR.

- Lorsque des composés sont totalement inactifs vis-à-vis d'un récepteur, la valeur d'activité donnée par les pharmacologues est $IC_{50} < 10^{-4}$. Dans un échantillon, si l'on a plusieurs composés ayant cette valeur d'activité $pIC_{50} = 4$, un biais important est introduit.

- Le temps de calcul (machine) limite le nombre de composés pouvant être étudiés par les indices de similarité moléculaire. La détermination de ces indices n'a pu être effectuée par exemple pour l'ensemble des 382 ligands.

- En ce qui concerne les prédictions, si l'on considère l'ensemble des prédictions effectuées lors de ces études, l'écart moyen entre les valeurs prédites et les valeurs expérimentales est d'environ 1 unité de pIC_{50} . Ce résultat est très encourageant et apporte une information quantitative importante au chimiste.

Les prédictions sur la famille des indoles ont montré que ce modèle était exploitable pour la prédiction *a priori* de nouveaux composés. Les prédictions obtenues sur les composés chromanes sont excellentes (écart moyen 0.7 unité). Pour les ligands thiochromanes, les prévisions d'activité sont correctes. Les prédictions des composés chromanes et thiochromanes sur l'ensemble des 382 ligands sont moins précises que sur les échantillons respectifs.

Nous voyons donc qu'une étude QSAR est soumise à beaucoup de paramètres et que l'on atteint rapidement les limites d'une telle étude. Si l'on veut améliorer les prévisions, la seule solution consiste à introduire dans les calculs des informations supplémentaires permettant de différencier les composés actifs et inactifs. La seule voie envisageable est l'étude des interactions entre ces ligands et le récepteur. Dans le prochain chapitre, nous présenterons une étude de dynamique moléculaire des complexes ligand/récepteur, avec certains ligands de la famille des chromanes.

Références Chapitre 3

-
- ¹ Hibert, M. F.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. *J. Med. Chem.* **1988**, *31*, 1087-1093.
 - ² Hibert, M. F.; McDermott, I.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. *Eur. J. Med. Chem.* **1989**, *24*, 31-37.
 - ³ Glennon, R. A. *J. Med. Chem.* **1987**, *30*, 1-12.
 - ⁴ Glennon, R. A.; Naiman, N. A.; Pierson, M. E.; Titeler, M.; Lyon, R. A.; Weisberg, E. *Eur. J. Pharmacol.* **1988**, *154*, 339-341.
 - ⁵ Schirlin, D.; Gerhart, F.; Hornsperger, J. M.; Hamon, M.; Wagner, J.; Jung, M. J. *J. Med. Chem.* **1986**, *31*, 30-36.
 - ⁶ Macor, J. E.; Burkhart, C. A.; Heym, J. H.; Ives, J. L.; Lebel, L. A.; Newman, M. E.; Nielsen, J. A.; Ryan, K.; Schulz, D. W.; Torgersen, L. K.; Koe, B. K. *J. Med. Chem.* **1990**, *33*, 2087-2093.
 - ⁷ Macor, J. E.; Fox, C. B.; Johnson, C.; Koe, B. K.; Lebet, L. A.; Zorn, S. H. *J. Med. Chem.* **1992**, *35*, 3625-3632.
 - ⁸ Taylor, E. W.; Nikam, S.; Weck, B.; Martin, A.; Nelson, D. *Life Sci.* **1987**, *41*, 1961-1969.
 - ⁹ Van Wijngaarden, I.; Tulp, M. T. M.; Soudijn, W. *Eur. J. Pharmacol. -Molecular Pharmacology Section* **1990**, *188*, 301-312.
 - ¹⁰ Ennis, M. D.; Baze, M. E.; Smith, M. W.; Lawson, C. F.; Mc Call, R. B.; Lahti, R. A.; Piercey, M. F. *J. Med. Chem.* **1992**, *35*, 3058-3066.
 - ¹¹ Sternlöf, P.; Gullme, M.; Elebring, T.; Andersson, B.; Wikström, H.; Lagerquist, S.; Svensson, K.; Ekman, A.; Carlsson, A.; Sundell, S. *J. Med. Chem.* **1993**, *36*, 2059-2065.
 - ¹² Hillver, S. E.; Björk, L.; Li, Y. L.; Svensson, B.; Ross, S.; Anden, N. E.; Hacksell, U. *J. Med. Chem.* **1990**, *33*, 1541-1544.
 - ¹³ Wikström, H.; Andersson, B.; Svensson, A.; Humber, L. G.; Asselin, A. A.; Svensson, K.; Ekman, A.; Carlsson, A.; Nilsson, I.; Chidester, C. *J. Med. Chem.* **1989**, *32*, 2273-2276.

-
- ¹⁴ Wikström, H.; Andersson, B.; Elebring, T.; Facyno, J.; Allinger, N. L.; Svensson, K.; Carlsson, A.; Sundell, S. *J. Med. Chem.* **1987**, *30*, 1567-1573.
- ¹⁵ Liu, Y.; Mellin, C.; Björk, L.; Svensson, B.; Csöreg, I.; Helander, A.; Kenne, L.; Anden, N. E.; Hacksell, U. *J. Med. Chem.* **1989**, *32*, 2311-2318.
- ¹⁶ Liu, Y.; Svensson, B. E.; Yu, H.; Cortizo, L.; Ross, S. B.; Lewander, T.; Hacksell, U. *Bio. Med. Chem. Lett.* **1991**, *1*, 257-262.
- ¹⁷ Liu, Y.; Cortizo, L.; Yu, H.; Svensson, B. E.; Lewander, T.; Hacksell, U. *Eur. J. Med. Chem.* **1995**, *30*, 277-286.
- ¹⁸ Jain, A. N.; Harris, N. L.; Park, J. Y. *J. Med. Chem.* **1995**, *38*, 1295-1308.
- ¹⁹ Chidester, C. G.; Lin, C. H.; Lahti, R. A.; Haadsma-Svensson, S. R.; Smith, M. W. *J. Med. Chem.* **1993**, *36*, 1301-1315.
- ²⁰ Zhuang, Z. P.; Kung, M. P.; Kung, H. F. *J. Med. Chem.* **1993**, *36*, 3161-3165.
- ²¹ Arvidsson, L. E.; Hacksell, U.; Johansson, A. M.; Nilsson, J. L. G.; Lindberg, P.; Sanchez, D.; Wikström, H.; Svensson, K.; Hjorth, S.; Carlsson, A. *J. Med. Chem.* **1984**, *27*, 45-51.
- ²² Arvidsson, L. E.; Hacksell, U.; Johansson, A. M.; Nilsson, J. L. G.; Svensson, K.; Hjorth, S.; Magnusson, T.; Carlsson, A.; Andersson, B.; Wikström, H. *J. Med. Chem.* **1987**, *30*, 2105-2109.
- ²³ Arvidsson, L. E.; Karlen, A.; Norinder, U.; Kenne, L.; Sundell, S.; Hacksell, U. *J. Med. Chem.* **1988**, *31*, 212-221.
- ²⁴ Lin, C. H.; Haadsma-Svensson, S. R.; Phillips, G.; Mc Call, R. B.; Piercey, M. F.; Smith, M. W.; Svensson, K.; Carlsson, A.; Chidester, C. G.; Von Voigtlander, P. F. *J. Med. Chem.* **1993**, *36*, 2208-2218.
- ²⁵ Romero, A. G.; Leiby, J. A.; Mc Call, R. B.; Piercey, M. F.; Smith, M. W.; Han, F. *J. Med. Chem.* **1993**, *36*, 2066-2074.
- ²⁶ Sonesson, C.; Barf, T.; Nilsson, J.; Dijkstra, D.; Carlsson, A.; Svensson, K.; Smith, M. W.; Martin, I. J.; Duncan, J. N.; King, L. J.; Wikström, H. *J. Med. Chem.* **1995**, *38*, 1319-1329.
- ²⁷ Vallgarda, J.; Arvidsson, L. E.; Svensson, B. E.; Fowler, C. J.; Hacksell, U. *Eur. J. Med. Chem.* **1993**, *28*, 399-406.
- ²⁸ Gozlan, H.; Ponchant, M.; Daval, G.; Verge, D.; Menard, F.; Vanhove, A.; Beaucourt, J. P.; Hamon, M. *J. Pharmacol. Exp. Ther.* **1988**, *244*, 751-759.

-
- 29 Naiman, N.; Lyon, R.; Bullock, A.; Rydelek, L.; Titeler, L.; Glennon, R. A. *J. Med. Chem.* **1989**, *32*, 253-256.
- 30 Björk, L.; Backlund Höök, B.; Nelson, D. L.; Anden, N. E.; Hacksell, U. *J. Med. Chem.* **1989**, *32*, 779-783.
- 31 Mellin, C.; Björk, L.; Karlen, A.; Johansson, A. M.; Sundell, S.; Kenne, L.; Nelson, D. L.; Anden, N. E.; Hacksell, U. *J. Med. Chem.* **1988**, *31*, 1130-1140.
- 32 Mellin, C.; Liu, Y.; Hacksell, U.; Björk, L.; Anden, N. E. *Acta Pharm. Sued.* **1987**, *24*, 153-160.
- 33 Mellin, C.; Vallgarda, J.; Nelson, D. L.; Björk, L.; Yu, H.; Anden, N. E.; Csöreg, I.; Arvidsson, L. E.; Hacksell, U. *J. Med. Chem.* **1991**, *34*, 497-510.
- 34 Cossery, J.M.; Gozlan, H.; Spampinato, U.; Perdicakis, C.; Guillaumet, G.; Pichat, L.; Hamon, M. *Eur. J. Pharmacol.* **1987**, *140*, 143-155.
- 35 Besson, T.; Podona, T.; Baudin, M.L.; Coudert, G.; Guillaumet, G. *Bio. Med. Chem. Lett.* **1993**, *3*, 1935-1940.
- 36 Cossery, J.M.; Perdicakis, C.; Coudert, G.; Guillaumet, G.; Pichat, L. *J. Labelled Comp. Radiopharm.* **1988**, *25*, 833-854.
- 37 Podona, T.; Guardiola-Lemaître, B.; Caignard, D.H.; Adam, G.; Pfeiffer, B.; Renard, P.; Guillaumet, G. *J. Med. Chem.* **1994**, *37*, 1779-1793.
- 38 Guillaumet, G.; Guardiola-Lemaître, B. *Eur. Pat. Appl.* EP 452, 204 (octobre 1991); *Chem. Abstr.* **1992**, *116*, 41 037n.
- 39 Comoy, C.; Marot, C.; Podona, T.; Baudin, M.L.; Morin Allory, L.; Guillaumet, G.; Pfeiffer, B.; Caignard, D.H.; Renard, P.; Rettori, M.C.; Guardiola Lemaître, B.; Adam, G. *J. Med. Chem.* soumis pour publication.
- 40 Marot, C.; Comoy, C.; Viaud, M.C.; Rettori, M.C.; Pfeiffer, B.; Morin Allory, L.; Guillaumet, G. *Bio. Med. Chem. Lett.* soumis pour publication.
- 41 Baudin, M. L.; Coudert, G.; Pfeiffer, B.; Renard, P.; Guardiola Lemaître, B.; Viaud, M.C.; Guillaumet, G. *J. Med. Chem.* soumis pour publication.
- 42 Marot, C.; Morin Allory, L. *J. Mol. Mod.* soumis pour publication.
- 43 Thèse de Benarab, A. H. *1,4-dioxino [2,3-b] pyridines substituées sur l'hétérocycle oxygène : synthèse et propriétés pharmacologiques* **1992**, Université d'Orléans.
- 44 Dabiré, H.; Bajjou, R.; Chaouche Teyara, K.; Fournier, B.; De Nanteil, G.; Laubie, M.; Safar, M.; Schmitt, H. *Eur. J. Pharmacol.* **1991**, *203*, 323-324.

-
- 45 Zhuang, Z. P.; Kung, M. P.; Kung, H. F. *J. Med. Chem.* **1994**, *37*, 1406-1407.
- 46 Raghupathi, R. K.; Rydelek Fitzgerald, L.; Teitler, M.; Glennon, R. A. *J. Med. Chem.* **1991**, *34*, 2633-2638.
- 47 Hutchison, A.; Williams, M.; De Jesus, R.; Stone, G. A.; Sylvester, L.; Clarke, F. H.; Sills, M. A. *J. Med. Chem.* **1989**, *32*, 720-727.
- 48 Moser, P. C.; Tricklebank, M. D.; Middlemiss, D. N.; Mir, A. K.; Hibert, M. F.; Fozard, J. R. *Br. J. Pharmacol.* **1990**, *99*, 343-349.
- 49 Moon, M. W.; Morris, J. K.; Heier, R. F.; Chidester, C. G.; Hoffmann, W. E.; Piercey, M. F.; Althaus, J. S.; Von Voigtlander, P. F.; Evans, D. L.; Figur, L. M.; Lahti, R. A. *J. Med. Chem.* **1992**, *35*, 1076-1092.
- 50 Mokrosz, J. L.; Pietrasiewicz, M.; Duszynska, B.; Cegła, M. T. *J. Med. Chem.* **1992**, *35*, 2369-2374.
- 51 Mokrosz, J. L.; Duszynska, B. *Pol. J. Pharmacol. Pharm.* **1992**, *44*, 527-538.
- 52 Mokrosz, J. L.; Paluchowska, M. H.; Chojnacka Wojcik, E.; Filip, M.; Charakchieva Minol, S.; Deren Wesolek, A.; Mokrosz, M. J. *J. Med. Chem.* **1994**, *37*, 2754-2760.
- 53 Kawakubo, H.; Takagi, S.; Yamaura, Y.; Katoh, S.; Ishimoto, Y.; Nagatani, T.; Mochizuki, D.; Kamata, T.; Sasaki, Y. *J. Med. Chem.* **1993**, *36*, 3526-3532.
- 54 Van Steen, B. J.; Van Wijngaarden, I.; Tulp, M. T. M.; Soudijn, W. *J. Med. Chem.* **1994**, *37*, 2761-2773.
- 55 Meyer, M. D.; De Bernardis, J. F.; Hancock, A. A. *J. Med. Chem.* **1994**, *37*, 105-112.

**CHAPITRE 4 : ETUDE DE DYNAMIQUE MOLECULAIRE DES COMPLEXES
LIGAND/RECEPTEUR**

MODELISATION DES COMPLEXES LIGAND/RECEPTEUR

-

RESULTATS

-

CONCLUSION ET PERSPECTIVES

Dans ce chapitre, nous présenterons les études de mécanique et dynamique moléculaire effectuées sur des complexes ligand/récepteur 5-HT_{1A}. Un travail similaire a déjà été effectué par quelques équipes^{1,2,3} soit dans un but cognitif de la dynamique de ces complexes, soit pour trouver une relation entre l'activité biologique des composés (agonistes, antagonistes, agonistes partiels) et les interactions mises en évidence. Ces choix impliquaient des ensembles hétérogènes de composés tests présentant une affinité notable pour le récepteur. Pour notre part, nous avons choisi d'effectuer ces études sur un ensemble homogène de ligands que nous avons étudiés en QSAR précédemment : les chromanes.

Ces études avaient comme objectifs les réponses à quelques questions :

- Les informations obtenues avec un ensemble homogène de ligands sont-elles similaires à celles obtenues avec des ensembles hétérogènes? Dans la négative, met-on en évidence des interactions spécifiques à cette famille de composés?

- Met-on en évidence une différence de comportement du complexe liée à l'affinité de ces ligands?

- Peut-on par dynamique moléculaire obtenir des informations qualitatives et/ou quantitatives permettant de mieux prédire l'activité potentielle des ligands?

- Si oui, ces informations peuvent-elles être introduites dans une approche QSAR afin de dépasser les limitations que nous avons trouvées précédemment?

Nous avons donc choisi un sous ensemble des composés chromanes utilisés précédemment (**C11**, **C13**, **C18**, **C30**, **C41**, **C44**, **C49**, **C50**). Ces ligands présentent une gamme d'activité allant de 4.70 à 9.70 unités de pIC₅₀ et diffèrent d'un point de vue structural par trois aspects : la présence ou non d'une fonction méthoxy (OMe) sur le cycle aromatique du squelette chromane, la position de l'atome d'azote de la fonction amine dans un cycle ou non et enfin la présence ou non d'une longue chaîne comme substituant sur l'atome d'azote.

Nous avons construit et modélisé le récepteur transmembranaire 5-HT_{1A} en utilisant sa séquence primaire et les données cristallographiques de la Bactériorhodopsine (BR) et en se servant du protocole de construction des récepteurs couplés aux protéines G mis au point par Hibert et al.^{4,5}. Les huit ligands

agonistes 5-HT_{1A} ont été placés graphiquement au niveau du site actif du récepteur. Nous avons ensuite analysé les interactions de ces ligands dans le complexe et l'influence de ces ligands sur la dynamique des hélices transmembranaires du récepteur. Pour chacun de ces complexes, des trajectoires d'une durée de 100ps ont été générées.

I. MODELISATION DES COMPLEXES LIGAND/RECEPTEUR

I.1 - Construction du récepteur 5-HT_{1A}.

Le récepteur 5-HT_{1A} est un récepteur transmembranaire couplé à la protéine G et composé de 422 acides aminés. Cette protéine a été isolée et séquencée par Kobilka et al. en 1987⁶. A l'heure actuelle, la structure tridimensionnelle du récepteur 5-HT_{1A} n'est pas connue. Récemment, Hibert *et al.*⁴ et Trumpp-Kallmeyer *et al.*⁵ ont décrit une méthode de construction des récepteurs couplés à la protéine G dont fait partie entre autres le récepteur 5-HT_{1A}. La construction de ce modèle tridimensionnel du récepteur 5-HT_{1A} est basé sur les coordonnées atomiques de la Bactériorhodopsine (BR) extraites de la banque de donnée Protein Data Bank⁷ (PDB) (fichier de coordonnées 1BRD.BRK).

La Bactériorhodopsine (Schéma 10), protéine de la membrane plasmique de l'archéobactérie *Halobacterium halobium*, est une pompe à protons du Rétinal. Sa chaîne polypeptidique traverse sept fois la membrane plasmique de la bactérie sous forme d'hélice α . Le chromophore rétinol est situé au coeur du faisceau d'hélices, attaché sur la dernière hélice (TMH7) par une base de Schiff. Le modèle le plus récent, dû à Henderson *et al.*⁸, est basé sur une carte de densité électronique.

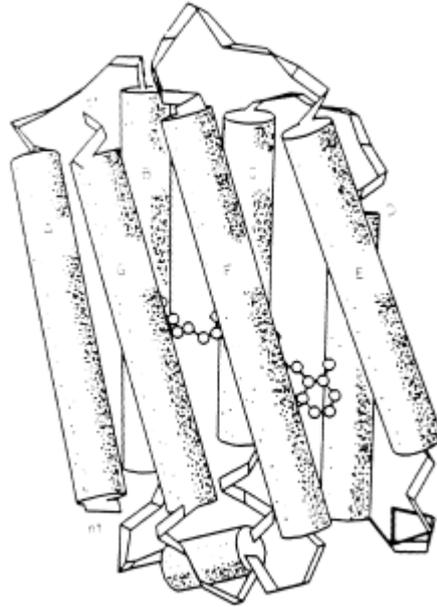


Schéma 10. Représentation 3D de la Bactériorhodopsine.

Pour établir la structure 3D du récepteur 5-HT_{1A}, nous avons utilisé la méthode de construction de Hibert *et al.*^{4,5}, généralisée aux récepteurs transmembranaires. Cette modélisation du récepteur 5-HT_{1A} est basée sur plusieurs hypothèses :

- On considère que les neurotransmetteurs couplés à la protéine G, dont fait partie le récepteur 5-HT_{1A}, ont une structure tridimensionnelle (3D) semblable à la Bactériorhodopsine. Cette structure 3D prévoit sept hélices α ^{9,10} transmembranaires et six boucles cytoplasmiques et extra-cellulaires. On considère que l'orientation des hélices dans la membrane est la même pour la Bactériorhodopsine que pour le récepteur 5-HT_{1A}. La plupart des domaines cytoplasmiques (boucles) semblent être impliqués dans le couplage avec les protéines G (sous unité alpha). Ainsi, les régions de la troisième boucle et de la partie C-terminale proche de la membrane¹¹ sont absolument nécessaires au couplage avec les protéines G. Les acides aminés de ces régions constitueraient une hélice amphiphile qui reconnaîtrait un élément structural de la sous unité alpha. Cette hypothèse est confortée par le fait que des peptides synthétiques ou extraits de venin d'abeille (mastoparan) qui possèdent cette structure amphipatique sont capables de stimuler l'adénylate cyclase sans passer par la reconnaissance du récepteur¹².

- Les indices d'hydropathie (d'hydrophobicité) calculés le long des chaînes peptidiques montrent que les récepteurs couplés à la protéine G ont sept domaines hydrophobes représentés par des hélices α présentes dans l'espace membranaire^{13,14,15}. Ces calculs d'indices permettent de localiser les résidus extrêmes des hélices α . La partie N-terminale est située dans la région extra-cellulaire alors que la partie C-terminale est localisée dans la région cytoplasmique.

- La longueur moyenne des hélices α est de 27 résidus pour la Bactériorhodopsine. Cette longueur moyenne se retrouve dans d'autres protéines comme *Rhodopseudomonas viridis*^{16,17} et dans celles de certaines plantes¹⁸. Cette longueur moyenne doit se retrouver approximativement dans les récepteurs transmembranaires.

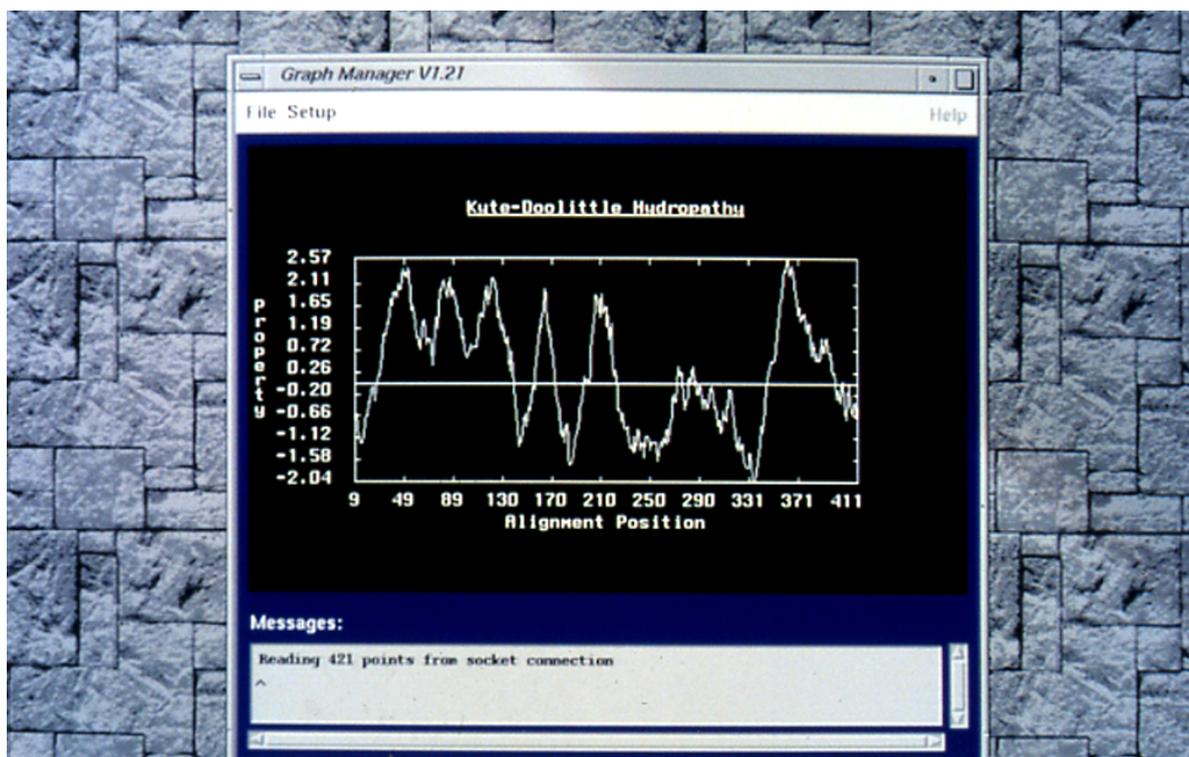
- Les études de microscopie électronique effectuées sur la Bactériorhodopsine montrent que les sept hélices sont perpendiculaires à la membrane de la cellule et qu'elles forment une structure ovale^{8,9,10}. L'orientation de ces hélices est antiparallèle. Cette orientation est la même pour les récepteurs β_2 -adrénergique¹⁹, D₂ dopaminergique²⁰ et les récepteurs sérotoninergiques 5-HT²¹ en général. Baldwin *et al.*²² ont récemment mis en évidence que l'orientation des hélices pour la Bactériorhodopsine, la rhodopsine (récepteur couplé à la protéine G) et les autres récepteurs transmembranaires couplés à la protéine G, était similaire.

- Divers travaux ont montré que le site d'interaction du ligand est localisé au niveau transmembranaire^{19,23}. Ainsi, la déglycosylation ne modifie pas l'affinité du ligand, ni le couplage avec la protéine G. De même, il est possible d'éliminer les boucles extra-cellulaires sans pour autant perdre les caractéristiques essentielles de ces récepteurs^{24,25}.

- La grande majorité des récepteurs couplés à la protéine G inclue dans leur séquence des résidus aspartiques sur les hélices 2 (TMH2) et 3 (TMH3)⁵. Ces résidus sont Asp82 et Asp116 pour le récepteur 5-HT_{1A}. Une substitution des résidus Asp82, Asp116, Ser199 (TMH5), Ser393 (TMH7) et Asn396 (TMH7) fait décroître fortement l'affinité pour le récepteur 5-HT_{1A}^{26,27,28}.

Le récepteur 5-HT_{1A} a donc été construit et modélisé sur ces bases. Il ne s'agit là que d'un modèle spéculatif puisqu'il a été construit en se basant sur de très nombreuses hypothèses qui, même si elles sont raisonnables, ne sont pas toutes vérifiables dans l'état actuel de nos connaissances.

Tout d'abord, la séquence des 422 acides aminés a été assemblée avec le logiciel PIMMS V1.44²⁹ d'Oxford Molecular. A partir de cette séquence, nous avons effectué une étude des indices d'hydropathie^{30,31}. Cette étude montre les zones hydrophobes correspondant aux hélices α (Graphe 21).



Graphe 21. Profil d'hydropathie avec les paramètres de Kyte-Doolittle (fenêtre de 7 résidus).

Dans cette étude, la modélisation des boucles intra et extra-cellulaires n'est pas prise en compte. Leur structure tridimensionnelle pourrait être modélisée à partir d'études d'homologie de séquences. Notre choix d'exclure les boucles externes est basé sur le fait que les caractéristiques de ces récepteurs sont maintenues en leur absence. Seules les sept hélices α sont modélisées suivant les critères énoncés précédemment. Les séquences de ces hélices sont celles utilisées par Trumpp-Kallmeyer *et al.*⁴ (Schéma 11). Les extrémités N-terminales et C-terminales de chacune des hélices transmembranaires (TMH) sont décrites par des fonctions acétamides et N-méthylamides respectivement. La topologie de ces sept hélices décrivant l'ensemble des atomes (leur type, leur masse, leur charge et leur géométrie interne) est construite en utilisant les dictionnaires de données RT37D inclus dans le logiciel PROSIMULATE²⁹. Les paramètres du champ de force sont issus du fichier IFP37D également inclus dans le logiciel.

TMH1 : ³⁷VITSLLLGLIFCAVLGNACVVA
 TMH2 : ⁷⁶GSLAVTDLMVSVLVLPMAALYQV
 TMH3 : ¹⁰⁶QVTCDLFIALDVLCCTSSILH
 TMH4 : ¹⁵⁹L TWLIGFLISIPPILG
 TMH5 : ¹⁹⁶TIYSTFGAFYIPLLLMLVLYG
 TMH6 : ³⁴⁶TLGIIMGTFILCWLPFFIVALV
 TMH7 : ³⁷⁶HMPDLLGAIINWLGYSNSLLNPVI

Schéma 11. Séquences en acides aminés des sept hélices?? et numérotation des résidus.

Les sept hélices sont construites initialement en structure α standard décrite par l'enchaînement des angles de torsion du squelette $\phi = -57^\circ$ et $\psi = -47^\circ$. Les résidus ionisables des hélices sont considérés chargés.

L'énergie de chaque hélice est ensuite minimisée en utilisant le champ de force GROMOS³² du logiciel PROSIMULATE V1.3. Le protocole de minimisation est le suivant : 500 pas de «steepest-descent» suivi de 1000 cycles «conjugate-gradient». Les énergies intramoléculaires des hélices après cette étape de minimisation sont données dans le tableau 54.

Hélices TMH	Energie (KJ/mol)
TMH1	-941
TMH2	-788
TMH3	-1311
TMH4	-334
TMH5	-401
TMH6	-510
TMH7	-897

Tableau 54. Energies des TMH après relaxation.

Les hélices ainsi minimisées séparément, sont superposées à celles de la Bactériorhodopsine dans le but de les orienter de manière analogue et les organiser en structure ovale. L'écart quadratique moyen (RMSD : Root Mean Square Deviation) entre les coordonnées atomiques des atomes C α des hélices initiales du récepteur 5-HT_{1A} et des hélices de la Bactériorhodopsine est indiqué dans tableau 55.

Hélices TMH	RMS
TMH1	1.05
TMH2	1.08
TMH3	1.38
TMH4	1.36
TMH5	0.71
TMH6	1.26
TMH7	1.13

Tableau 55. Valeurs des RMSD pour chaque hélice.

La valeur du RMSD est donnée par :

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (x_{a_i} - x_{b_i})^2}{N}}$$

où x_{a_i} et x_{b_i} sont les coordonnées cartésiennes (x, y, z) d'un même atome dans deux structures différentes et N étant le nombre d'atomes sur lequel est effectué la superposition.

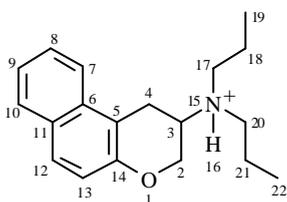
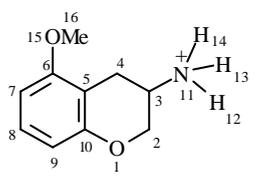
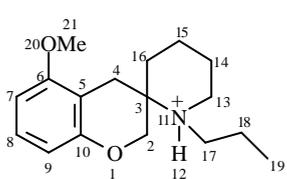
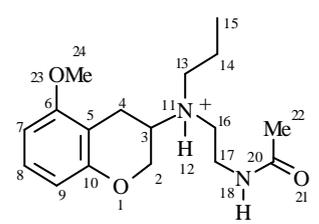
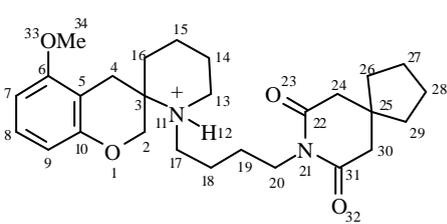
Cet ensemble des sept hélices est ensuite relaxé suivant un protocole de 1000 pas de «steepest-descent» (SD) puis 3000 cycles de «conjugate-gradient» (CG) avec le champ de force GROMOS. L'énergie en fin de cycle de minimisation est de l'ordre -6000 KJ/mol.

Remarques

Le champ de force GROMOS ne prend pas en compte explicitement les atomes d'hydrogène aliphatiques (CH, CH₂, CH₃). C'est le modèle étendu dans lequel la paramétrisation des atomes de carbone auxquels ils sont attachés est modifiée. Par contre les atomes d'hydrogène polaires sont pris en compte explicitement.

I.2 - Construction des ligands sélectionnés.

Huit ligands de type chromane, générés sous forme protonée, ont été sélectionnés pour étudier les complexes ligand/récepteur (Tableau 56). Le conformère retenu pour l'analyse est le même qui a servi antérieurement pour l'étude QSAR suivant le protocole : recherche conformationnelle avec COBRA²⁹ V3.2 suivie d'une étude de superposition avec le pharmacophore de Mellin. Le conformère est protoné sur l'azote et la géométrie interne est alors optimisée avec le programme semi-empirique MOPAC³³ V5.0 (AM1, CHARGE=1). La topologie des ligands est décrite avec les mêmes fichiers utilisés pour les résidus du récepteur. Les types d'atomes des ligands sont données dans le tableau 56.

Cps	Structures	Types d'atomes	Activité (pIC ₅₀) (M)
C11		1 : OS 2, 4, 17, 18, 20, 21 : CH ₂ 3 : CH ₁ 5, 14 : CB 6, 7, 8, 9, 10, 11, 12, 13 : CR61 15 : NL 16 : H 19, 22 : CH ₃	4.70 Inactif
C13		1, 15 : OS 2, 4 : CH ₂ 3 : CH ₁ 5, 6, 10 : CB 7, 8, 9 : CR61 11 : NL 12, 13, 14 : H 16 : CH ₃	6.20 Inactif
C50		1, 20 : OS 2, 4, 13, 14, 15, 16, 17, 18 : CH ₂ 3 : C 5, 6, 10 : CB 7, 8, 9 : CR61 11 : NL 12 : H 19, 21 : CH ₃	6.40 Inactif
C18		1, 23 : OS 2, 4, 13, 14, 16, 17 : CH ₂ 3 : CH ₁ 5, 6, 10 : CB 7, 8, 9 : CR61 11 : NL 12, 19 : H 15, 22, 24 : CH ₃ 18 : N 20 : C 21 : O	7.40 Inactif
C49		1, 33 : OS 2, 4, 13, 14, 15, 16, 17, 18, 19, 20, 24, 26, 27, 28, 29, 30 : CH ₂ 3, 25 : C 5, 6, 10 : CB 7, 8, 9 : CR61 11 : NL 12 : H 21 : N	8.05 Actif

		22, 31 : C 23, 32 : O 34 : CH ₃	
C30		1, 30 : OS 2, 4, 13, 14, 16, 17, 18, 19, 24, 25, 26 : CH ₂ 3, 23, 27 : CH ₁ 5, 6, 10 : CB 7, 8, 9 : CR61 11 : NL 12 : H 20 : N 21, 28 : C 22, 29 : O 31 : CH ₃	8.70 Actif
C41		1, 26, 32 : OS 2, 4, 13, 14, 16, 17, 18, 19 : CH ₂ 3 : CH ₁ 5, 6, 10, 24, 25 : CB 7, 8, 9, 28, 29, 30, 31 : CR61 11 : NL 12, 21 : H 20 : N 22 : C 23 : O 27, 33 : CH ₃	9.00 Actif
C44		1, 32 : OS 2, 4, 13, 14, 16, 17, 18, 19, 23, 25, 26, 27, 28, 29 : CH ₂ 3 : CH ₁ 5, 6, 10 : CB 7, 8, 9 : CR61 11 : NL 12 : H 20 : N 21, 24, 30 : C 22, 31 : O 33 : CH ₃	9.70 Actif

Tableau 56. Types d'atomes et activité 5-HT_{1A} des ligands.

I.3 - Choix du site actif.

Nous avons orienté notre choix du site actif d'après les récents travaux de biologie moléculaire qui suggèrent que les résidus Asp82 (TMH2)²⁸ et Asp116 (TMH3)²⁷ jouent un rôle capital dans les interactions ligand-récepteur 5-HT_{1A}. Ces résidus, qui sont conservés pour tous les neurorécepteurs cationiques, interagissent avec des atomes d'azote protonés des ligands. Les travaux récents de Chanda et al.²⁸, portant sur des études de mutation, montrent que les résidus Asp82 (TMH2), Ser393 (TMH7) et Asn396 (TMH7) sont nécessaires pour le «binding» des ligands agonistes du récepteur 5-HT_{1A}. Ces résultats ont guidés notre choix du site actif utilisé dans notre étude. Par ailleurs, des travaux de dynamique moléculaire effectués par Sylte et al.¹ sur le récepteur 5-HT_{1A} et le ligand agoniste 5-HT montrent des interactions spécifiques de ce ligand avec le résidu Asp82 de l'hélice transmembranaire 2 (Schéma 12). De plus, ce site actif spécifique des agonistes du récepteur humain 5-HT_{1A} a été très récemment proposé dans l'article de Hen et Saudou³⁴.

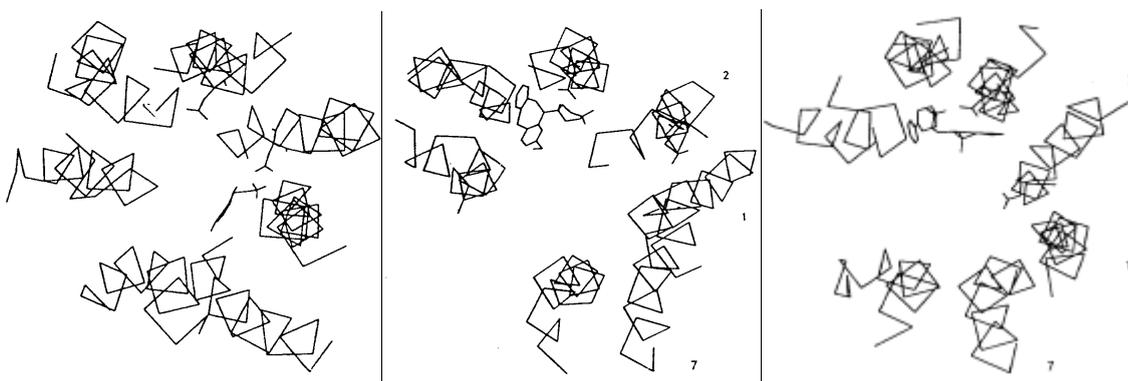


Schéma 12. A gauche : Site agoniste avec la Sérotonine (5-HT) ; au centre : Site antagoniste avec la S-Méthiothépine ; à droite : Site agoniste partiel avec l'Ipsapirone.

D'autres sites sont proposés par Sylte et al.¹ pour les agonistes partiels (Ipsapirone) et les antagonistes (Méthiothépine) (Schéma 12). Ces études de «docking» montrent que ces ligands (agoniste partiel et antagoniste) interagissent de préférence avec le résidu Asp116 de l'hélice 3 (TMH3).

Des travaux de dynamique moléculaire de complexes ligand/récepteur 5-HT_{1A} très récents de Kuipers et al.² mettent en évidence deux sites actifs différents pour des composés 5-HT_{1A} agonistes (5-HT, 8-OH-DPAT) et antagonistes (Pindolol) (Schéma 13).

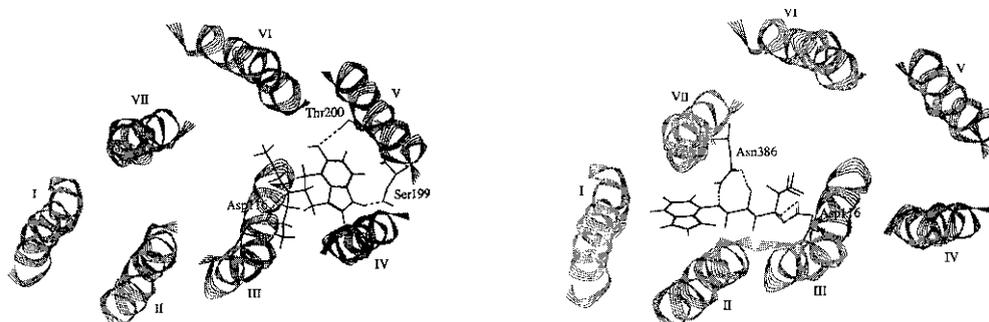


Schéma 13. Site agoniste à gauche, site antagoniste à droite.

Cette fois, le site agoniste est constitué du résidu Asp116 de l'hélice 3 et des résidus Ser199 et Thr200 de l'hélice 5. Les résidus du site antagoniste sont Asn386 (TMH7) et Asp116 (TMH3). Tous ces résultats laissent supposer l'existence de plusieurs sites actifs (agoniste-agoniste partiel-antagoniste).

Toutes ces études de dynamique moléculaire avec différents ligands agonistes, agonistes partiels et antagonistes du récepteur 5-HT_{1A}, montrent la complexité qu'il existe dans le choix du site actif. Pour notre étude, nous nous sommes appuyés sur les récents travaux de biologie moléculaire de Chanda et sur les travaux de dynamique moléculaire de Sylte. Le site actif choisi, pour nos ligands agonistes du récepteur 5-HT_{1A} est donc constitué principalement des résidus Asp82 de l'hélice 2 et Ser393 de l'hélice 7.

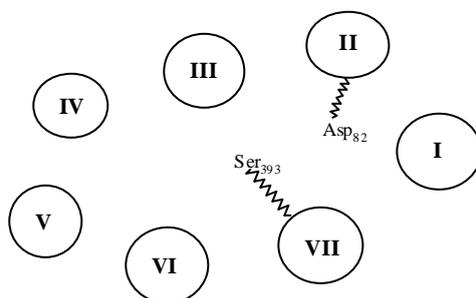


Schéma 14. Représentation schématique des principaux résidus du site actif.

I.4 - Protocoles de dynamique moléculaire pour l'étude des complexes ligand/récepteur 5-HT_{1A}.

HT_{1A}.

Les ligands sont placés graphiquement au niveau du site actif du récepteur, dans l'espace libre compris entre les hélices 1-2-7. Les groupements amino (NH^+) des ligands sont orientés vers le résidu Asp82 de l'hélice 2. Les groupements méthoxy (OMe) des ligands sont placés proche de l'atome d'hydrogène de la fonction alcool (OH) du résidu Ser393 (TMH7). Les distances initiales moyennes, entre les fonctions polaires des ligands (OMe, NH^+) et les fonctions alcools et carboxylates respectives des résidus Ser et Asp, sont comprises entre 1.5 Å et 2 Å. A l'état initial, tous ces ligands sont en interactions avec ces deux résidus (Schéma 15), sauf le composé **C11** qui n'a pas de fonction méthoxy.

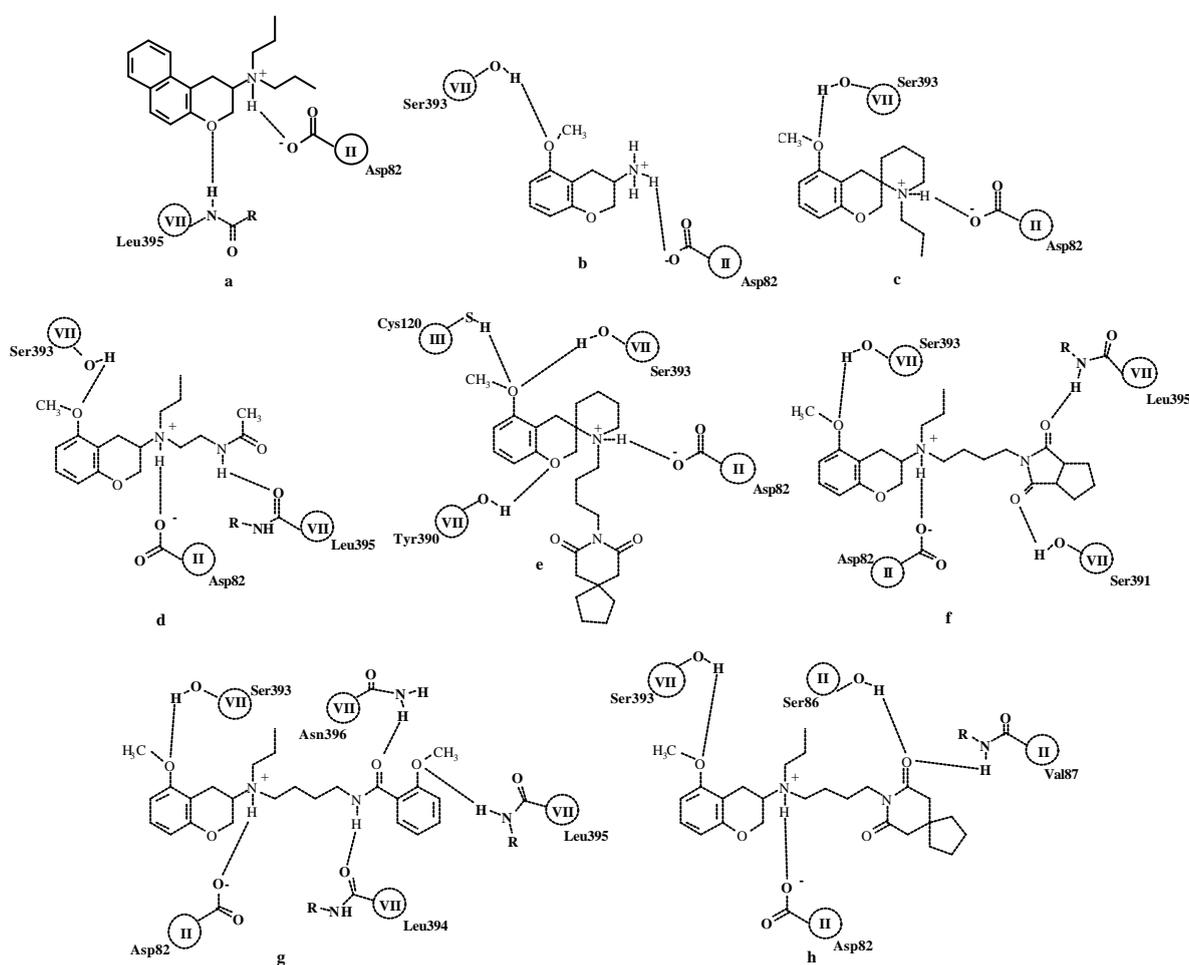


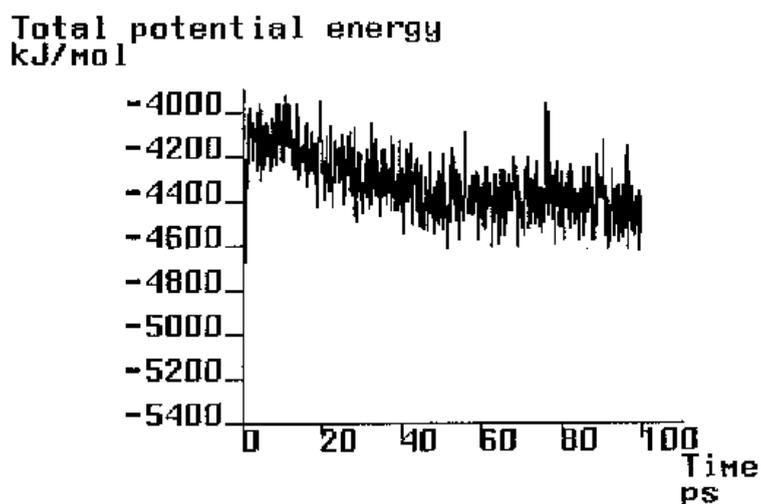
Schéma 15. Vue schématique des interactions dans les conformations initiales. (a: composé **C11**, b: composé **C13**, c: composé **C50**, d: composé **C18**, e: composé **C49**, f: composé **C30**, g: composé **C41** et h: composé **C44**).

Pour les composés ayant une chaîne latérale relativement mobile, c'est à dire pour les composés **C18**, **C30**, **C41**, **C44**, **C49**, l'orientation de cette chaîne latérale est fonction de la conformation initiale du ligand, choisie lors de la superposition avec le pharmacophore de Mellin. Pour les composés **C18**, **C30** et **C44**, cette chaîne latérale est orientée vers le milieu du récepteur et entre les hélices TMH1, TMH2 et TMH7. Pour les molécules **C41** et **C49**, la chaîne est dirigée vers l'intérieur de la cellule (espace cytoplasmique) et se situe entre les hélices TMH2 et TMH7 pour le **C49** et entre les hélices TMH1 et TMH7 pour le **C41**.

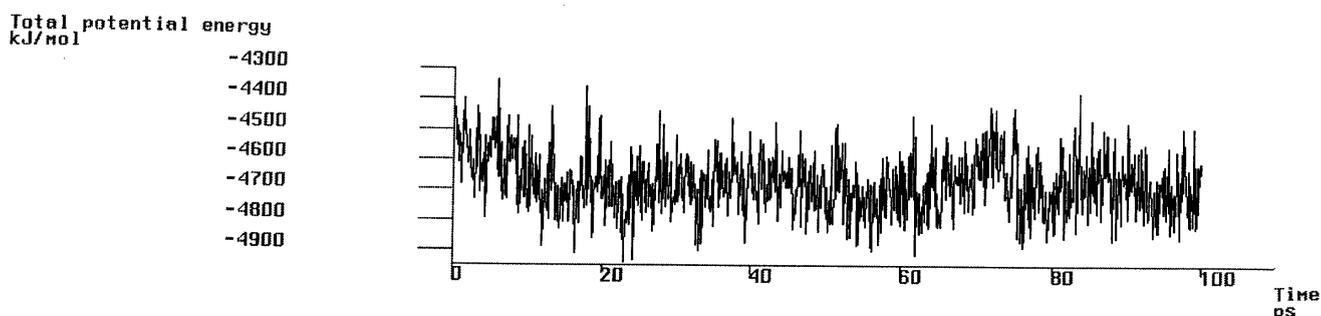
Les complexes ligand/récepteur ainsi formés (environ 1500 atomes) sont ensuite relaxés par une double étape de minimisation (SD puis CG) sans produire de changements structuraux très importants. Ceci a permis de minimiser les principaux défauts de structure comme les violations de contraintes stériques, les corrections des liaisons et des angles de valence. Ensuite, les structures de ces complexes relaxés sont utilisées comme système initial dans une étude de dynamique moléculaire. Le protocole est le suivant :

- 1- Thermalisation de 0 K à 300 K par paliers de 50 K d'une durée de 4ps.
- 2- Période d'équilibre de 100ps à 300 K.
- 3- Période de production de 100ps avec génération de 1000 configurations stockées le long de la trajectoire.

Afin d'obtenir un système moléculaire correctement équilibré, la phase de thermalisation doit être menée avec soin. Il est souvent préférable de « chauffer » le système progressivement à la température de 300 K par paliers successifs de 50 K plutôt que de passer brutalement de 0 K à 300 K. A chaque palier de température des vitesses aléatoires différentes sont attribuées à chaque atome à partir d'une distribution de Maxwell pour une température donnée. Chaque montée en température s'effectue en 4ps. L'algorithme de « Leap Frog » utilisé dans GROMOS permet le couplage à un bain thermique. Un couplage fort (constante de relaxation de 0.01ps) est utilisé pendant toute la période de thermalisation. A 300 K, le couplage au bain thermique est diminué (couplage faible 0.1ps) et une période d'équilibration d'une durée de 100ps est effectuée (Graphe 22). L'étape finale est la phase de production pendant laquelle les différentes configurations du système sont stockées afin d'en effectuer une analyse détaillée.



Grphe 22. Energie potentielle en fonction du temps pour la période d'équilibre du complexe C44/récepteur.



Grphe 23. Energie potentielle en fonction du temps pour la période de production du complexe C44/récepteur.

Le graphe 22 relatif au composé **C44**, montre que l'équilibre du système est atteint à partir de 60ps. Le graphe 23 indique les fluctuations énergétiques relatives à la période de production. Ces fluctuations sont relativement faibles (environ 300 KJ/mol). La température du système est stabilisée à 300 K \pm 5 K.

Pour les autres ligands, les représentations graphiques sont sensiblement identiques et montrent que pour chacun des complexes l'équilibre est atteint entre 60 et 80ps.

Remarques

La constante diélectrique est $\epsilon = 1$ (dans le vide). La distance de coupure ou « cutoff » choisie est de 8 Å pour les interactions non liées. Les interactions sont calculées en se basant sur une liste de groupements d'atomes remise à jour à chaque étape pour tous les groupements d'atomes distants de moins de 8 Å. Pour les autres paires, la liste est calculée toutes les 10 étapes. Toutes les longueurs de liaisons sont maintenues constantes en utilisant la méthode SHAKE³⁵, et cela pour toute la durée de la simulation. Ceci nous a permis de prendre un incrément de temps de 1fs.

Il est à noter qu'il existe des contraintes importantes en temps de calcul pour mener des simulations de dynamique moléculaire. La taille du système ligand/récepteur est considérablement réduite : les boucles extra et intra-cellulaires ne sont pas prises en compte, l'environnement membranaire et la présence éventuelle de molécules d'eau sont exclues de notre modèle. Néanmoins, c'est environ 15 jours de calcul qui sont nécessaires sur une station Silicon Graphics Challenge pour mener à bien l'étude de dynamique complète d'un ensemble protéine-ligand.

En parallèle à cette étude des récepteurs complexés, une étude de dynamique du récepteur 5-HT_{1A} seul, sans le ligand, a été menée. L'objectif était d'obtenir la structure moyenne du système des hélices transmembranaires dans laquelle le ligand pourrait être introduit de manière systématique. Une simulation de dynamique moléculaire, dans les mêmes conditions que celles décrites précédemment, a été menée. La durée des phases d'équilibre et de production est de 840ps. Il est observé que le système s'équilibre moins rapidement que dans le cas du récepteur complexé. Le résultat majeur est le compactage des hélices après 840ps de simulation (Photo 3). Le calcul de l'écart quadratique moyen (RMSD) calculé pour les atomes C α entre la structure initiale et la structure obtenue en fin de période de production est de 4.1 Å.

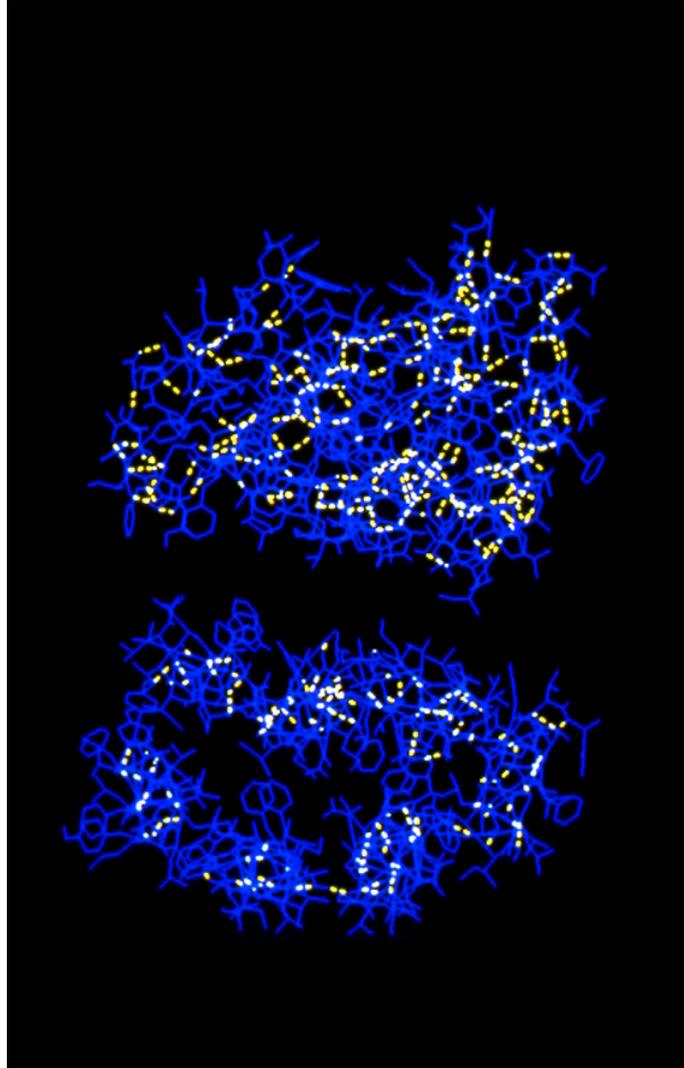


Photo 3. Représentation du récepteur à l'état initial (en bas) et en fin de production (en haut) (Trait jaune = liaison hydrogène).

Cette photo montre le nombre important de liaisons hydrogène (traits jaune) formées entre les hélices transmembranaires dans la structure obtenue en phase finale de production. Le compactage des hélices conduisant à un emplacement du site actif pratiquement occulté ne permettant pas d'introduire de ligands entre les hélices TMH2 et TMH7 est clairement mis en évidence. Ce résultat pouvait être attendu et montre l'importance, dans nos simulations, de la présence du ligand au niveau du site actif pour maintenir la structure ovale des sept hélices transmembranaires. La structure moyenne du récepteur seul ainsi modélisé ne sera pas utilisée dans notre étude et le protocole choisi qui consiste à placer le ligand, dans le site actif préalablement choisi, dans la structure initiale minimisée semble donc plus judicieux.

II. RESULTATS

Dans ce paragraphe, il nous faut aborder les limitations de ces méthodes de simulation. La principale limite est que les mouvements observables sont liés au temps de simulation qu'il est possible de produire. Les mouvements accessibles sont ceux de faible amplitude et il est par exemple impossible de simuler des mouvements tels que le déplacement de ligands dans des parties différentes du récepteur ou des changements conformationnels importants des hélices transmembranaires du récepteur. Les résultats des simulations sont donc tributaires de la position initiale du ligand mais aussi de la distribution spatiale des hélices les unes par rapport aux autres.

Nos simulations n'ont pas l'ambition de rendre compte d'un changement conformationnel important des hélices transmembranaires, ni de prétendre examiner plusieurs sites de liaisons du ligand dans le récepteur. Par contre, elles permettent de déterminer la structure stable des différents complexes ligand/récepteur étudiés, de donner des informations qualitatives et quantitatives sur les interactions de ces ligands dans le site actif présumé et de rendre compte d'une certaine mobilité des ligands dans le récepteur et des hélices transmembranaires.

Les complexes ligand actif/récepteur et ligand inactif/récepteur ainsi modélisés permettent de comprendre et justifier les différences d'activité déterminées par les études pharmacologiques, entre les huit ligands sélectionnés.

Dans un premier temps, nous avons comparé la stabilité des complexes (interactions de van der Waals et électrostatiques) évaluées dans la structure finale minimisée, obtenue en fin de période de production.

L'analyse des interactions ligand/récepteur est détaillée dans la structure moyenne calculée pour les dernières 50ps de production (Photos 4, 5, 6, 7, 8, 9, 10, 11) et ces interactions sont comparées à celles établies initialement dans le site actif (Schéma 15). La dynamique des ligands dans le site actif est suivie durant les dernières 50ps de trajectoire en analysant le pourcentage de formation de liaison hydrogène. Cette étude a été effectuée avec le module proAHB du logiciel PROSIMULATE.

Pour étudier la mobilité relative des hélices transmembranaires du récepteur, les fluctuations atomiques (RMS) du récepteur durant les 100ps de la période de production et les valeurs de RMSD des C α des hélices transmembranaires sont calculées.

Enfin, nous avons aussi superposé les structures des complexes initiaux aux structures moyennes déterminées sur les dernières 50ps de production, afin de visualiser les déplacements relatifs aux ligands et aux hélices.

Etude des interactions

Dans un premier temps, nous avons comparé les énergies d'interactions des ligands avec le récepteur (Tableau 57).

Composés	Interactions VDW/Coulomb (KJ/mol)
C11	-265/-33
C13	-147/-72
C50	-237/-42
C18	-244/-53
C49	-320/-104
C30	-312/-247
C41	-315/-196
C44	-293/-183

Tableau 57. Energies d'interactions ligand/récepteur après 100ps de période de production.

La première information est que les énergies d'interaction de van der Waals et électrostatiques des composés actifs (**C49**, **C30**, **C41** et **C44**) sont supérieures à celles observées dans le cas des composés inactifs (**C11**, **C13**, **C50**, **C18**). Cette différence de stabilité s'explique par le nombre plus important de groupements polaires des ligands actifs susceptibles d'interactions fortes par liaison hydrogène.

La seconde étude est une analyse détaillée des interactions ligand/récepteur présentes dans les structures moyennes des complexes déterminées sur les dernières 50ps de production. Ces interactions sont comparées à celles établies initialement dans le site actif. Les résultats de cette étude seront présentés sous une forme récapitulative. Pour chaque composé, nous indiquerons la liste des interactions initiales (se référer au schéma 15) et celle des interactions dans la structure moyenne (voir photos 4, 5, 6, 7, 8, 9, 10, 11 et tableau 58).

Pour chaque site d'interaction présent dans la structure moyenne nous donnerons le pourcentage de temps, sur les dernières 50ps de la période de production, durant lequel les liaisons hydrogènes sont formées (se référer au tableau 58). Enfin nous indiquerons quelques remarques relatives à cette étude.

Le complexe ligand C11/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu *Asp82* (TMH2) avec le proton du ligand.
- Proton amide de la liaison peptidique du résidu *Leu395* (TMH7) avec l'oxygène du squelette chromane du ligand.

Interactions dans la structure moyenne

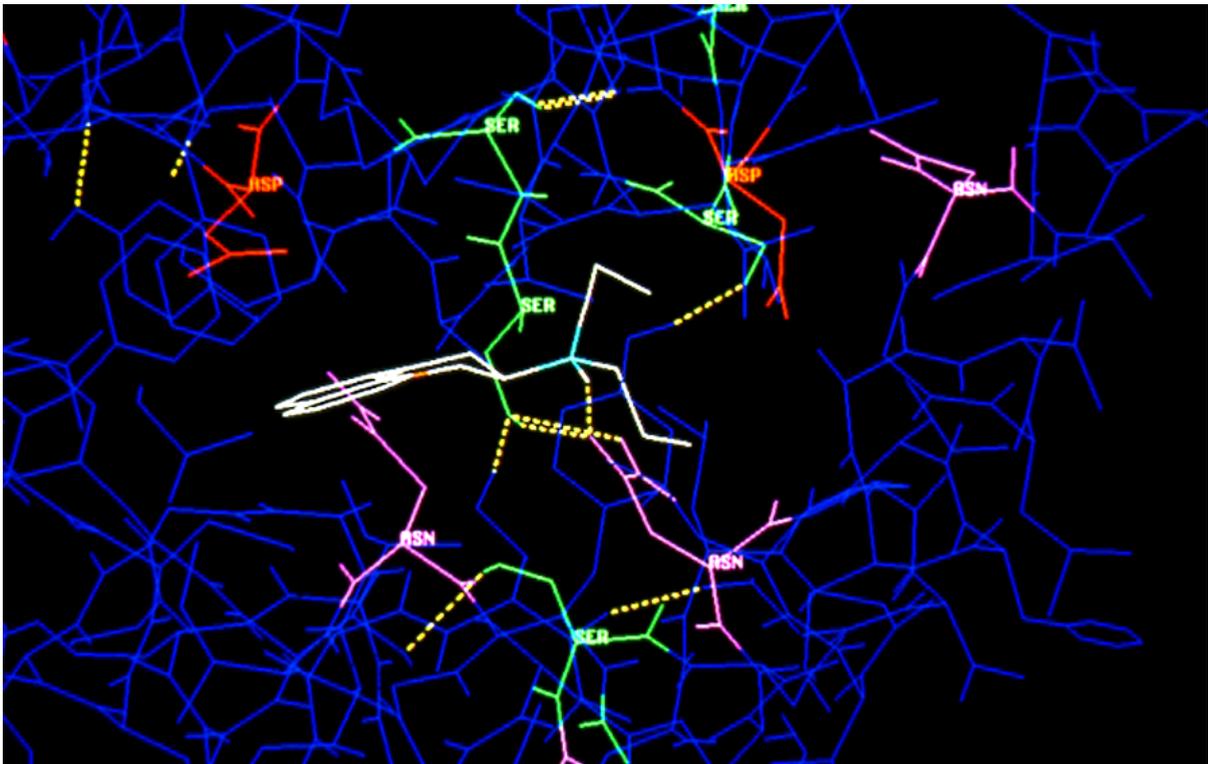


Photo 4. Interactions du complexe C11-récepteur dans la structure moyenne.

- Oxygène de la fonction amide du résidu **Asn396 (TMH7)** (50%) avec le proton du ligand.
- Proton de la fonction alcool du résidu **Ser123 (TMH3)** (10.6%) avec l'oxygène du squelette chromane du ligand.

Remarques

Le ligand perd très rapidement (dès le début de la période de thermalisation) l'interaction avec le résidu Asp82 pour une interaction préférentielle avec le résidu Asn396. La distance entre le proton amine du ligand et l'oxygène carboxylate du résidu Asp82 est de 4.35 Å. La distance entre l'oxygène chromanique et le proton de la liaison peptique de Leu395 est de 9.9 Å. L'interaction par liaison hydrogène de cet oxygène du squelette chromane avec le résidu Ser123 a un pourcentage d'existence faible (10.6%) et n'est pas représentée sur la photo 4.

Le complexe ligand C13/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** avec un proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** avec l'oxygène de la fonction méthoxy du ligand.

Interactions dans la structure moyenne

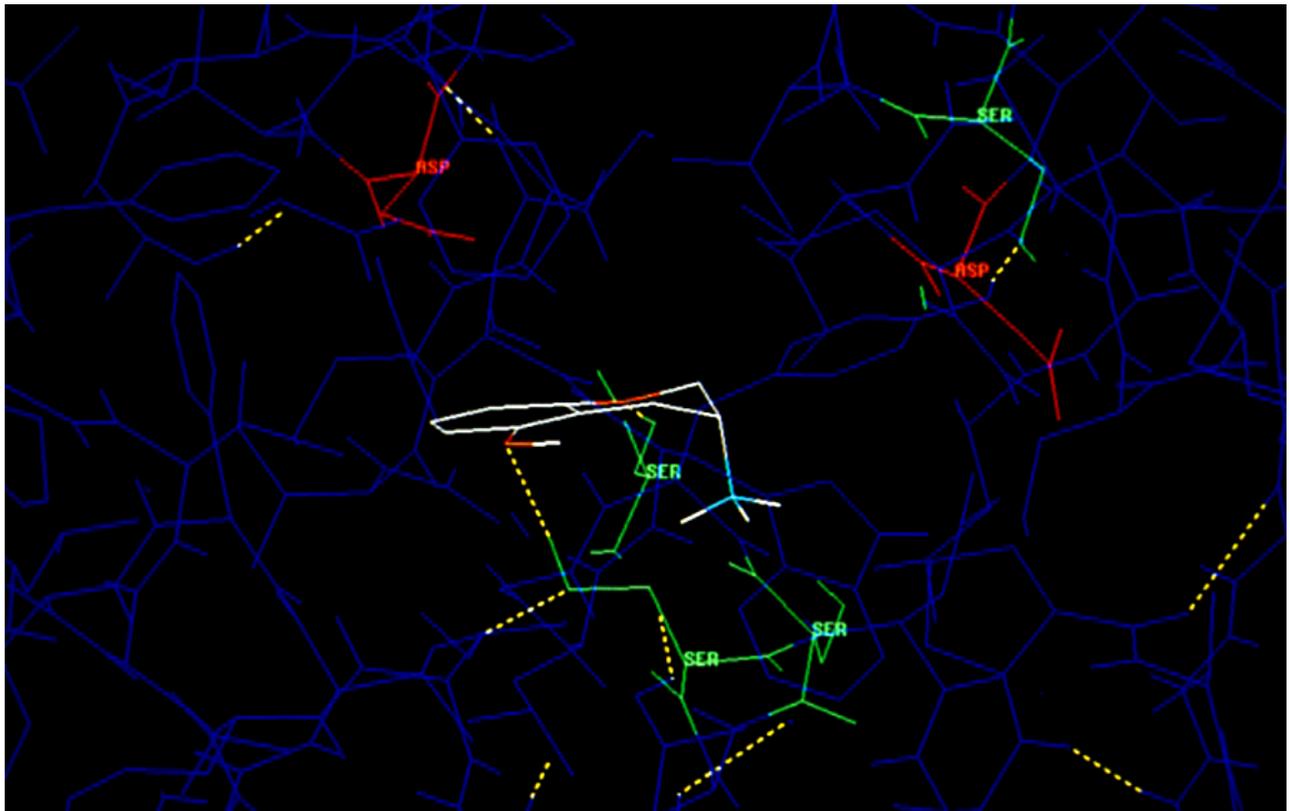


Photo 5. Interactions du complexe **C18**-récepteur dans la structure moyenne.

- Proton de la fonction alcool du résidu **Ser396 (TMH7)** (98.8%) avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction alcool du résidu **Ser123 (TMH3)** (97.6%) avec l'oxygène du squelette chromane du ligand.

Remarques

La liaison extra-cyclique entre le cycle saturé et la fonction amine du ligand passe d'une position équatoriale à axiale. Cette fonction amine n'a pas d'interaction avec le récepteur et se trouve à environ 5 Å du résidu Asp82. Cette fonction se trouve approximativement au centre des hélices transmembranaires et est orientée vers l'espace intracytoplasmique ; le résidu le plus proche étant le résidu Ser393 (≈ 3 Å).

Le complexe ligand C50/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu Asp82 (TMH2) avec le proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu Ser393 (TMH7) avec l'oxygène de la fonction méthoxy du ligand.

Interactions dans la structure moyenne

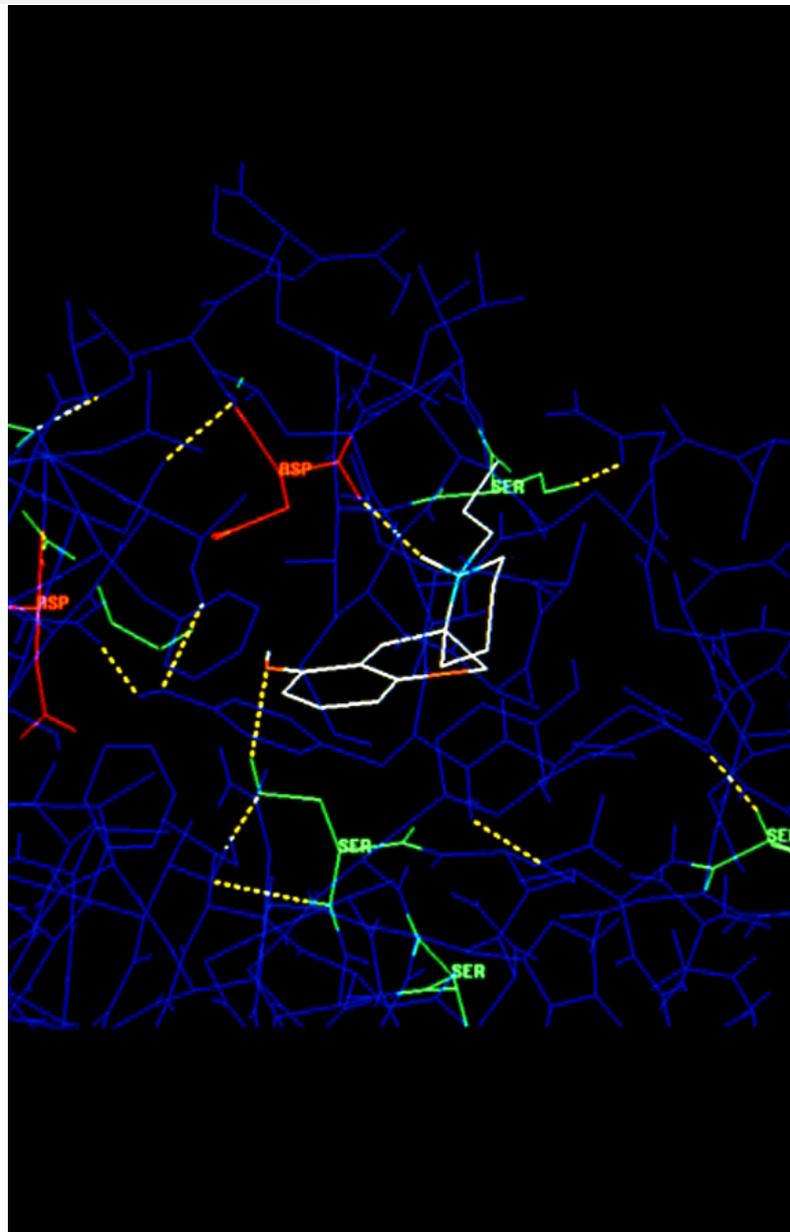


Photo 6. Interactions du complexe C50-récepteur dans la structure moyenne.

- Oxygène de la fonction peptidique du résidu **Asp82 (TMH2)** (93.5%) avec l'atome d'hydrogène de la fonction amine du ligand.
- Proton de la fonction alcool des résidus **Ser123 (TMH3)** (1.9%) et **Ser393 (TMH7)** (64%) avec l'oxygène de la fonction méthoxy du ligand.

Remarques

L'atome d'oxygène du squelette chromane ne participe à aucune liaison hydrogène durant la simulation.

Le complexe ligand C18/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** avec le proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** avec l'oxygène de la fonction méthoxy du ligand.
- Oxygène de la liaison peptidique du résidu **Leu395 (TMH7)** avec l'hydrogène de la fonction amide du ligand.

Interactions dans la structure moyenne

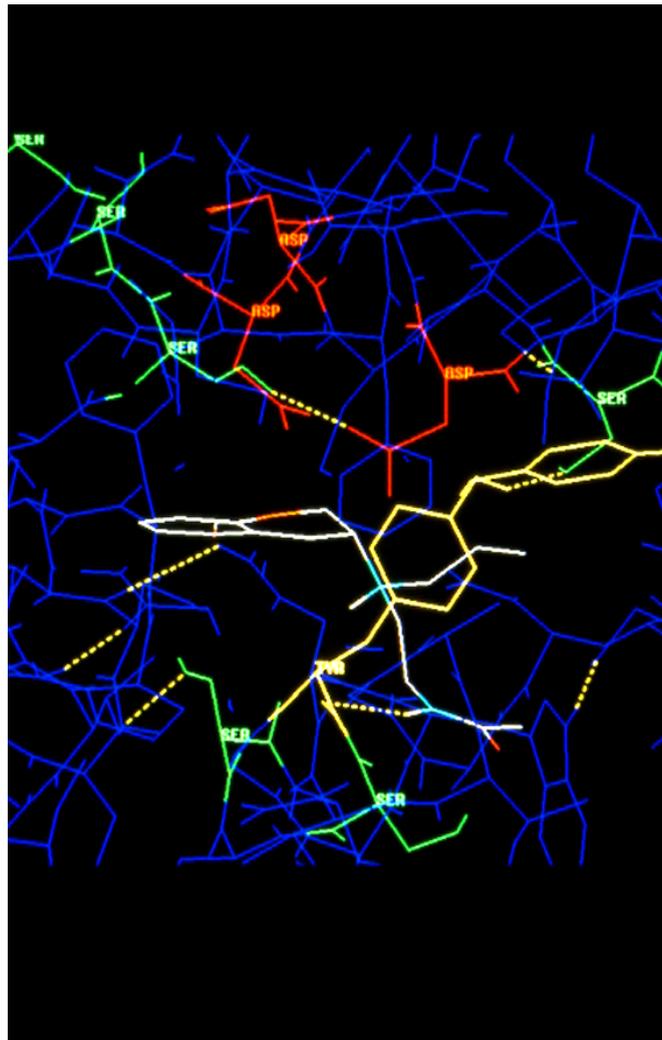


Photo 7. Interactions du complexe C18-récepteur dans la structure moyenne.

- Oxygène de la fonction peptidique du résidu *Tyr390* (TMH7) (59.6%) avec l'atome d'hydrogène de la fonction amide du ligand.
- Proton de la fonction thiol du résidu *Cys120* (TMH3) (23.6%) avec l'oxygène du squelette chromane du ligand.
- Proton de la fonction alcool du résidu *Ser391* (TMH7) (25.6%) avec l'oxygène de la fonction amide du ligand.
- Proton de la fonction alcool du résidu *Ser393* (TMH7) (29.2%) avec l'oxygène de la fonction méthoxy du ligand.

Remarques

La liaison extra-cyclique entre le cycle saturé et la fonction amine du ligand passe d'une position équatoriale à axiale. Cette fonction amine n'a qu'un pourcentage d'interaction de 2.1% avec le résidu Asp82. Dans la structure moyenne, l'hydrogène de la fonction amine se trouve à environ 6 Å du résidu Asp82. L'atome d'oxygène du squelette chromane présente une très faible interaction par liaison hydrogène avec la chaîne latérale du résidu Ser123 de l'hélice 2 (pourcentage d'existence 0.4%).

Le complexe ligand C49/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** avec le proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** et de la fonction thiol du résidu **Cys120 (TMH3)** avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction alcool du résidu **Tyr390 (TMH7)** avec l'oxygène du squelette chromane du ligand.

Interactions dans la structure moyenne

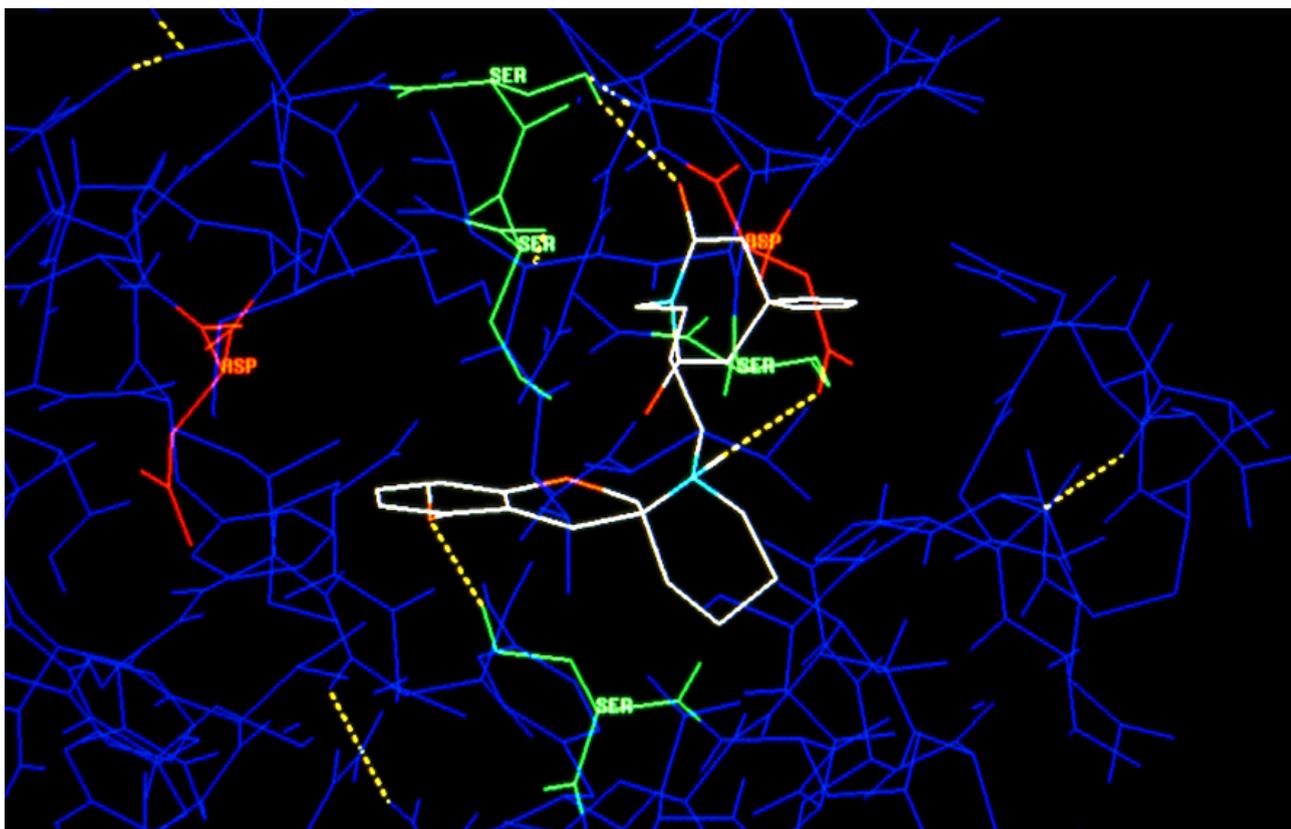


Photo 8. Interactions du complexe **C49**-récepteur dans la structure moyenne.

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** (76%) avec l'atome d'hydrogène de la fonction amine du ligand.
- Proton de la liaison peptidique du résidu **Thr353 (TMH6)** (32%) avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction alcool du résidu **Ser122 (TMH2)** (77.8%) avec un oxygène de la fonction imide du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** (77%) avec l'oxygène de la fonction méthoxy du ligand.

Remarques

La chaîne latérale du ligand n'a aucune interaction avec des résidus de l'hélice 7. L'atome d'oxygène du squelette chromane montre un pourcentage d'interaction faible (8.6%) avec le résidu Ser123 et (5.8%) avec le résidu Tyr390 (distance 2.6 Å).

Le complexe ligand C30/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** avec le proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction alcool du résidu **Ser391 (TMH7)** avec un oxygène de la fonction imide de la chaîne latérale du ligand.
- Proton de la liaison peptidique du résidu **Leu395 (TMH7)** avec un oxygène de la fonction imide du ligand.

Interactions dans la structure moyenne

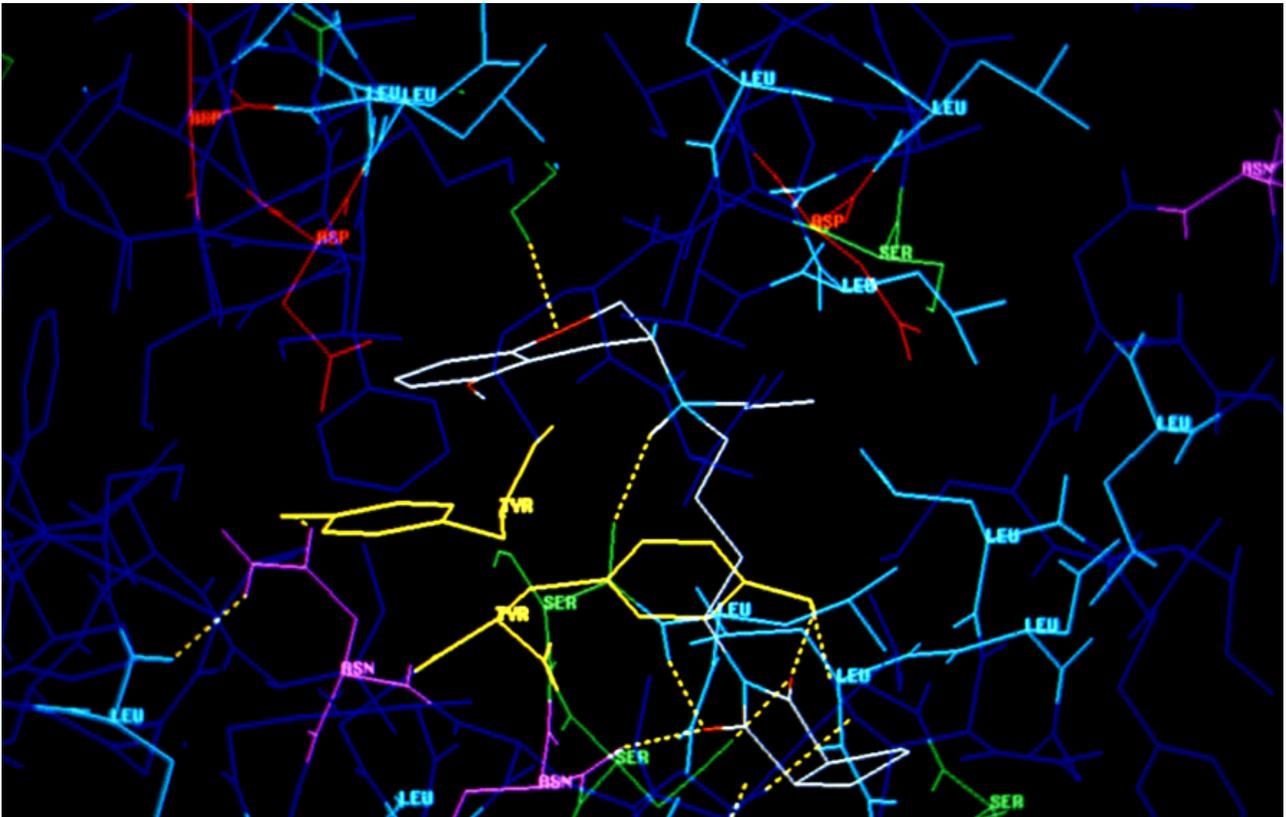


Photo 9. Interactions du complexe **C30**-récepteur dans la structure moyenne.

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** (12%) avec l'atome d'hydrogène de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser123 (TMH2)** (79.8%) avec l'oxygène du squelette chromane du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** (14%) avec l'oxygène de la fonction méthoxy du ligand. Ce résidu interagit aussi à 42% avec l'atome d'hydrogène de la fonction amine du ligand, mais l'interaction se fait entre l'oxygène de la liaison peptidique et le proton du ligand.

- Proton de la fonction alcool du résidu **Tyr390 (TMH7)** (25%) avec un oxygène de la fonction imide du ligand.
- Proton de la fonction alcool du résidu **Ser391 (TMH7)** (7.6%) avec un oxygène de la fonction imide du ligand.
- Proton amide de la liaison peptidique du résidu **Asn396 (TMH7)** (86.1%) avec un oxygène de la fonction imide du ligand.
- Proton amide de la liaison peptidique du résidu **Leu394 (TMH7)** (9.1%) avec un oxygène de la fonction imide du ligand.

Remarques

La liaison extra-cyclique entre le cycle saturé et la fonction amine du ligand passe d'une position équatoriale à axiale. Cette fonction amine n'a qu'un pourcentage d'interaction de 12% avec le résidu Asp82. Dans la structure moyenne, la fonction amine du ligand se trouve à environ 5.5 Å du résidu Asp82. L'atome d'hydrogène de la fonction alcool du résidu Ser393 se trouve, dans la structure moyenne à 3.4 Å de l'oxygène de la fonction méthoxy du ligand.

Le complexe ligand C41/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** avec le proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction amide du résidu **Asn396 (TMH7)** avec l'oxygène de la fonction amide du ligand.

- Oxygène carbonyle de la liaison peptidique du résidu *Leu394* (TMH7) avec le proton de la fonction amide du ligand.
- Proton amide de la liaison peptidique du résidu *Leu395* (TMH7) avec l'oxygène de la fonction méthoxy de la chaîne latérale du ligand.

Interactions dans la structure moyenne

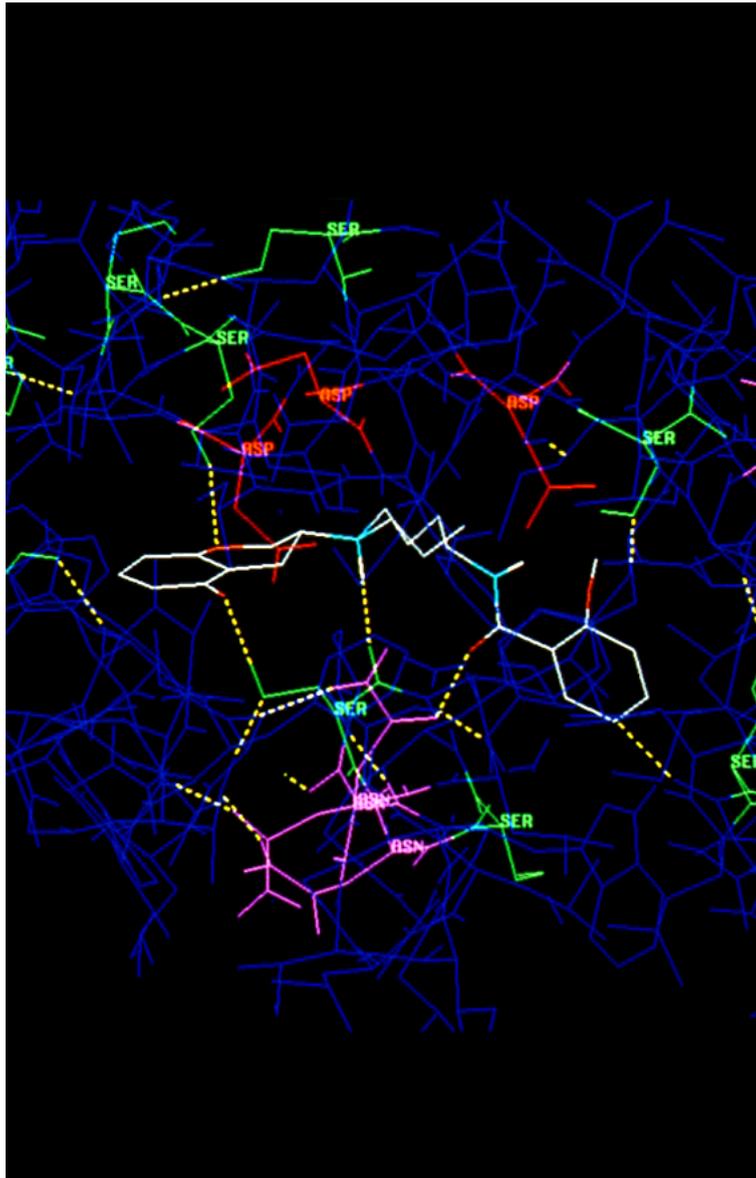


Photo 10. Interactions du complexe C41-récepteur dans la structure moyenne.

- Oxygène de la liaison peptidique du résidu *Ser393* (TMH7) (11%) avec l'atome d'hydrogène de la fonction amine du ligand.

- Proton de la fonction alcool du résidu **Ser123 (TMH2)** (78%) avec l'oxygène du squelette chromane du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** (87%) avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction amide du résidu **Asn396 (TMH7)** (7.6%) avec l'atome d'oxygène de la fonction amine du ligand.

Remarques

Dans la structure moyenne, la fonction amine du ligand se trouve à environ 5 Å du résidu Asp82, et les oxygènes de la chaîne latérale du ligand sont à une distance de 4 Å de l'atome d'hydrogène de la liaison peptidique du résidu Leu395 et 6 Å de l'atome d'oxygène de la liaison peptidique du résidu Leu394.

Le complexe ligand C44/récepteur

Interactions initiales

- Oxygène de la fonction carboxylate du résidu **Asp82 (TMH2)** avec le proton de la fonction amine du ligand.
- Proton de la fonction alcool du résidu **Ser393 (TMH7)** avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction alcool du résidu **Ser86 (TMH2)** avec un oxygène de la fonction imide du ligand.
- Proton amide de la liaison peptidique du résidu **Val87 (TMH2)** avec un oxygène de la fonction imide du ligand.

Interactions dans la structure moyenne

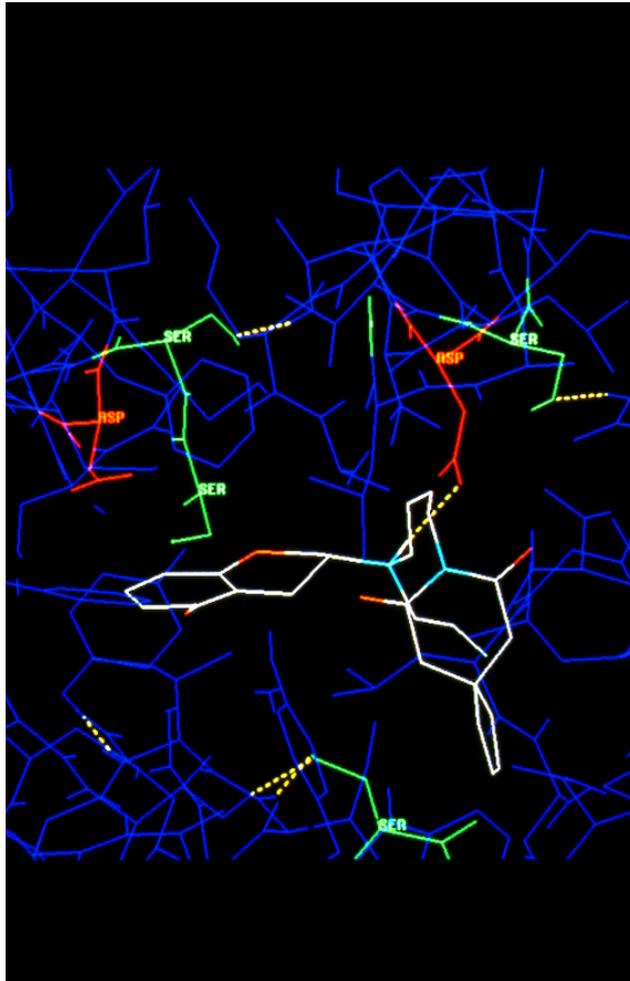


Photo 11. Interactions du complexe C44-récepteur dans la structure moyenne.

- Oxygène de la fonction carboxylate du résidu *Asp82* (TMH2) (90.8%) avec l'atome d'hydrogène de la fonction amine du ligand.
- Proton de la fonction alcool du résidu *Ser393* (TMH7) (11.2%) avec l'oxygène de la fonction méthoxy du ligand.
- Proton de la fonction amide du résidu *Asn396* (TMH7) (28.5%) avec un atome d'oxygène de la fonction imide du ligand.
- Proton de la fonction alcool du résidu *Tyr390* (TMH7) (18.6%) avec un atome d'oxygène de la fonction imide du ligand.

Remarques

Dans la structure moyenne, l'atome d'oxygène du squelette chromane du ligand se trouve à environ 6 Å du résidu Ser123. Le pourcentage d'existence de liaison hydrogène avec le résidu Cys119 n'est que de 5.6% durant les dernière 50ps de la période de production. Les atomes d'oxygène de la fonction imide du ligand se trouvent à environ 4 Å du résidu Ser86, 5 Å du résidu Val87, 3 Å du résidu Tyr390 et 7 Å du résidu Asn396.

Nous résumons les pourcentages d'existence d'interaction par liaison hydrogène entre les ligands et les résidus des hélices transmembranaires dans le tableau 58 suivant.

	I					II					III					VI		VII				
Cps	Cys 49	Ala 50	Ser 77	Leu 78	Thr 81	Asp 82	Ser 86	Cys 119	Ser 120	Ser 122	Ser 123	Thr 353	Cys 357	Trp 387	Tyr 390	Ser 391	Ser 393	Leu 394	Asn 396			
C11										10.6									50			
C13						0.6				97.6							98.8					
C50						93.5				1.9							64					
C18	3.8					2.1	6.6	23.6		0.4				11.6	59.6	25.6	29.2		0.4			
C49		0.4	4.4	1.4	2.6	76		10	77.8	8.6	32				5.8		77					
C30						12				79.8					25	7.6	56	9.1	86.1			
C41								3.6		78					5.4		96		7.6			
C44		20.8				90.8	2.2	5.6				0.6		18.6			11.2		28.5			

Tableau 58. Pourcentage de temps d'existence des liaisons hydrogènes durant les dernières 50ps de production.

Les résultats intéressants de ces études sont :

- dans la grande majorité des cas, les groupements polaires des ligands ont des interactions avec les résidus Asp82 de l'hélice 2 et Ser393 de l'hélice 7. Ces résultats sont en accord avec les travaux de biologie moléculaire de Chanda *et al.*²⁸.

- les composés les plus actifs **C18, C49, C30, C41, C44** montrent des sites d'interactions avec le récepteur plus nombreux que les composés inactifs (voir Tableau 58). La majorité de ces composés actifs interagissent avec les résidus Asp82 (TMH2), Ser123 (TMH3), et surtout avec des résidus de l'hélice 7, en particulier les résidus Tyr390 (TMH7), Ser393 (TMH7) et Asn396 (TMH7) ; tous ces résidus sont conservés et présents dans les autres récepteurs transmembranaires couplés aux protéines G,

- la fonction alcool du résidu Ser123 de l'hélice 3 se trouve être très souvent en interaction avec l'atome d'oxygène du squelette chromane, et révèle son importance, dans la stabilité des composés chromanes dans le site actif. En examinant le tableau 59, on peut noter que pour les parties communes à tous les ligands (fonction méthoxy, oxygène chromanique et fonction amine protonée), ce sont les composés les plus actifs qui présentent le plus d'interactions avec les résidus des hélices transmembranaires. Il semble donc raisonnable d'envisager un lien qualitatif entre l'affinité des ligands et le nombre d'interactions avec le récepteur, d'autant plus que les composés les plus actifs possèdent d'autres interactions avec les résidus via leur chaîne latérale.

Composés	Interactions		
	Fonction OMe du ligand	Oxygène du squelette chromane	Fonction amine NH ⁺ du ligand
C11		Ser123	Asn396
C13	Ser393	Ser123	
C50	Ser393 et Ser123		Asp82
C18	Ser393	Cys120	
C49	Ser393 et Thr353	Ser123	Asp82
C30	Ser393	Ser123 et Ser393	Asp82
C41	Ser393	Ser123	Ser393
C44	Ser393	Cys119	Asp82

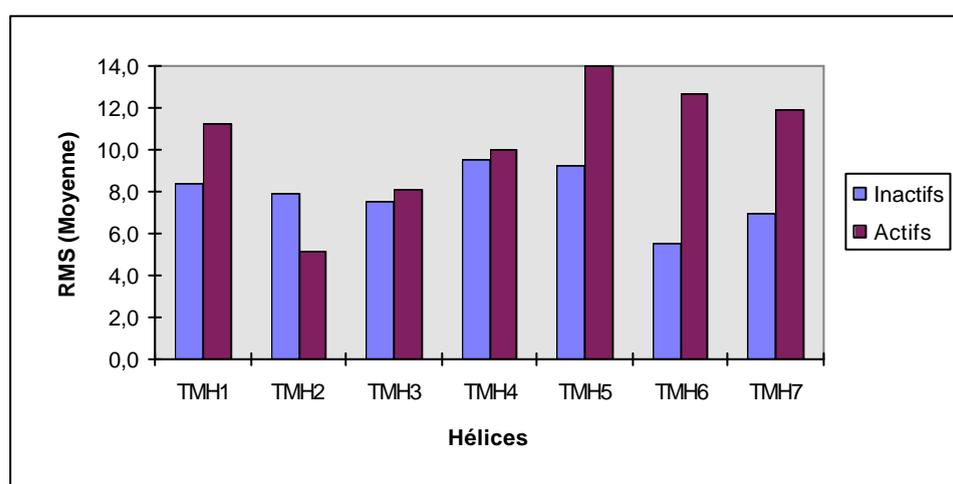
Tableau 59. Récapitulation des interactions de certaines parties fonctionnelles des ligands établies avec les résidus des hélices.

Influence des ligands sur les hélices transmembranaires

Pour analyser l'influence des ligands sur la position des hélices du récepteur au cours de la simulation, nous avons comparé les structures des hélices en évaluant le RMSD par superposition des carbones α de ces sept hélices, entre la structure initiale et la structure finale. Les résultats sont donnés dans le tableau 60 et le graphe 24.

Cps TMH	C1	C13	C50	C18	C49	C30	C41	C44	Moyenne des RMS pour les 4 composés inactifs (C11, C13, C50, C18)	Moyenne des RMS pour les 4 composés actifs (C49, C30, C41, C44)
TMH1	8.1	6.3	9.5	9.5	11.8	10.5	10.2	12.5	8.4	11.2
TMH2	8.8	9.0	5.5	8.2	5.4	4.5	4.2	6.3	7.9	5.1
TMH3	6.0	9.6	7.1	7.2	8.2	8.0	7.0	9.1	7.5	8.1
TMH4	5.9	12.9	10.2	9.2	9.8	10.4	8.2	11.6	9.5	10.0
TMH5	9.7	9.8	6.1	11.2	10.2	12.5	14.1	19.1	9.2	14.0
TMH6	5.0	4.9	6.0	6.2	8.9	13	12.6	16.1	5.5	12.7
TMH7	8.4	8.9	5.7	4.9	5.5	12.8	13.8	15.4	7.0	11.9

Tableau 60. Valeurs des RMSD (\AA) des $C\alpha$ entre les structures initiales et finales des 7 hélices.



Graphe 24. Valeurs moyennes des RMSD des 7 hélices complexées avec les ligands inactifs et actifs.

Les résultats obtenus dans cette étude montrent que l'influence des ligands sur la position des hélices est différente suivant les composés actifs ou inactifs. Pour le groupe des ligands inactifs, les valeurs moyennes des RMSD pour les hélices 1 à 5 sont d'environ 8 Å et pour les hélices 6 et 7 d'environ 6 Å. Pour le groupe des composés actifs, les valeurs moyennes pour les hélices 3 et 4 sont sensiblement similaires à celles du groupe inactif (≈ 9 Å). Par contre, pour les hélices transmembranaires TMH1-5-6-7, ces valeurs moyennes du RMSD sont plus importantes (≈ 12 Å). Pour l'hélice 2, la valeur est plus faible (≈ 5 Å).

Comme on le voit sur le graphe 24, les principales différences des valeurs de RMSD entre le groupe des composés inactifs et celui des produits actifs se situent au niveau des hélices 5, 6 et 7. Donc en moyenne, les composés actifs induisent un changement de position plus important des hélices 5, 6 et 7 du récepteur lors de nos simulations de dynamique. Ce résultat peut être mis en relation avec une des hypothèses données précédemment. Cette hypothèse suggère que la région cytoplasmique, en relation directe avec la troisième boucle intracellulaire reliant les hélices 5 et 6, est importante pour la transmission du message biologique. Les acides aminés de ces régions constitueraient une hélice amphiphile qui reconnaîtrait un élément structural de la protéine G. Ce postulat est conforté par le fait que des peptides synthétiques¹² qui possèdent cette structure amphipatique sont capables de stimuler l'adénylate cyclase par l'intermédiaire de la protéine G sans être reconnu par le récepteur lui même.

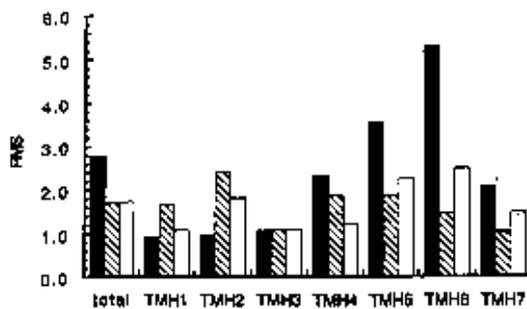
Ces valeurs de RMSD des hélices montrent que celles (TMH5 et TMH6) liées à la troisième boucle intra-cellulaire et l'hélice 7 liée à la partie C-terminale, subissent un réarrangement structural plus important avec les composés les plus actifs. Ce résultat est intéressant et soulève certaines questions :

- les fonctions de la boucle intracellulaire et de la partie C-terminale sont elles en relation avec un changement de la position des hélices 5, 6 ou 7 ?

- un changement de position des hélices 5 et 6 est-il relié aux propriétés pharmacologiques des ligands ?

Pour répondre à ces questions, nous nous appuyerons sur les travaux de D. Zhang et H. Weinstein³.

Ces travaux portent sur l'étude par dynamique moléculaire du récepteur 5-HT₂ avec des ligands agonistes soit fort (5-HT), soit faible (tryptamine) et un composé antagoniste (analogue de la 5-hydroxygramine). Des simulations effectuées sur les complexes ligand agoniste/récepteur, ligand antagoniste/récepteur ont montré, d'après une étude des RMSD des C α des hélices, que le changement de position des hélices 5, 6 et 7 était différent selon les ligands étudiés (Graphe 25).



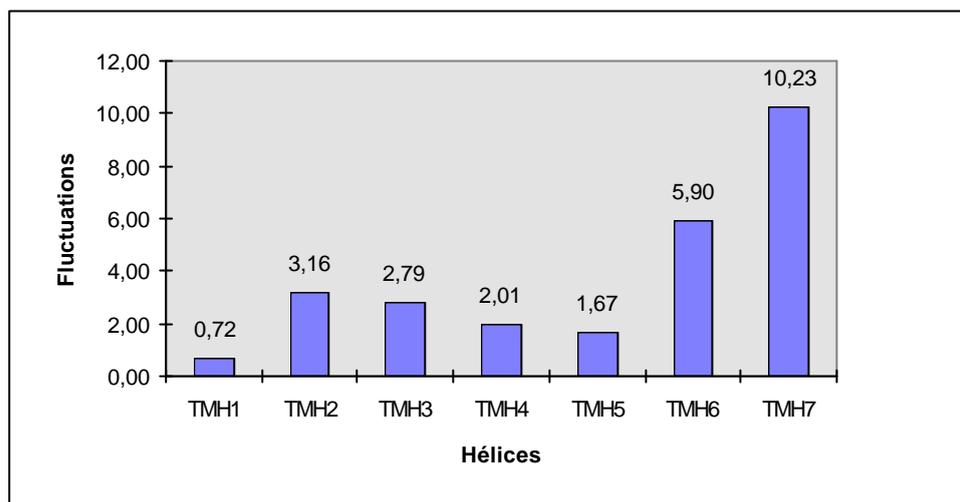
Graphe 25. Valeurs des RMSD (en angströms) des hélices α du récepteur 5-HT₂ (noir : avec agoniste fort (5-HT), hachuré : avec antagoniste, blanc : avec agoniste faible (tryptamine))³.

Comme l'indique le graphe 25, seul le complexe ligand agoniste fort/récepteur induit un changement de position important des hélices 5 et 6. A partir de ce résultat les auteurs suggèrent, entre autres, que les composés agonistes faibles et antagonistes occupent le même site actif que les ligands agonistes forts mais cela sans provoquer d'importants changements de position des hélices 5 et 6. Ils suggèrent donc qu'il existe une corrélation entre le changement structural de la protéine³⁶ au niveau de ces hélices et les propriétés pharmacologiques des ligands *via* les interactions du récepteur avec la protéine G³⁷.

Dans notre étude avec les ligands chromane, nous trouvons un résultat analogue à celui trouvé avec le récepteur 5-HT₂. Dans notre modèle, il semble que les ligands inactifs peuvent occuper le site actif mais que les interactions sont trop faibles pour provoquer un changement important des positions des hélices 5, 6 et/ou 7.

Remarques

En parallèle, une étude comparative des fluctuations atomiques (RMS) calculées sur tous les atomes du récepteur sur la période de production montre une différence majeure au niveau des hélices 6 et 7 (Graphe 26) suivant les complexes inactifs et actifs étudiés.



Grphe 26. Différences des fluctuations atomiques (RMS) des hélices α entre les composés du groupe inactif et ceux du groupe actif (durant les 100ps de période de production).

On définit l'écart quadratique moyen (RMS) d'une grandeur fluctuante x définie en N points par :

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

Cette étude révèle une différence importante de mobilité des hélices TMH6-7 suivant l'activité du ligand placé dans le site actif du récepteur ainsi que des changements structuraux. Cette différence de mobilité de ces hélices peut donc être corrélée, comme pour l'étude des RMSD des $C\alpha$, à l'activité intrinsèque des ligands.

Comparaison graphique des structures initiales et finales

Dans cette étude, nous avons superposé les structures initiales des complexes (couleur bleu sur les photos) aux structures moyennes des complexes (couleur rouge sur les photos), afin de visualiser le déplacement des ligands et des hélices durant les simulations.

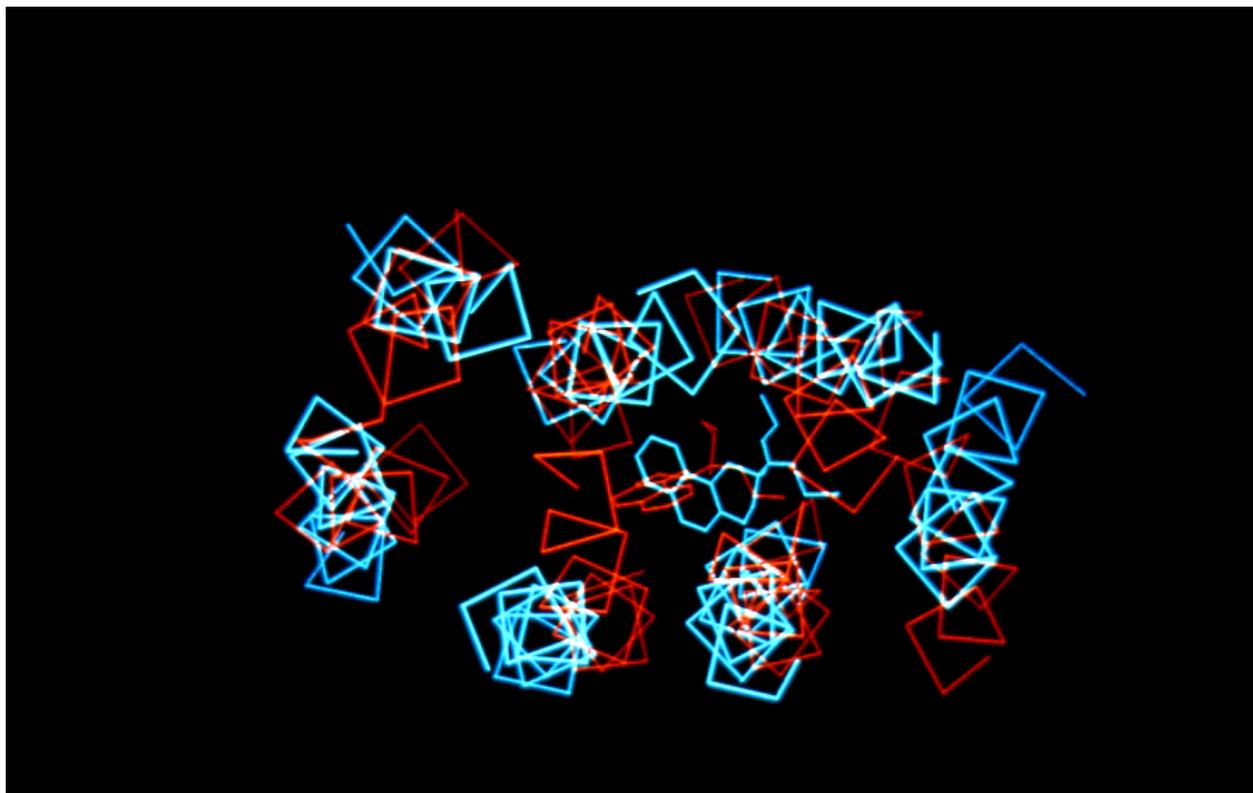


Photo 12. Complexe C11/récepteur.

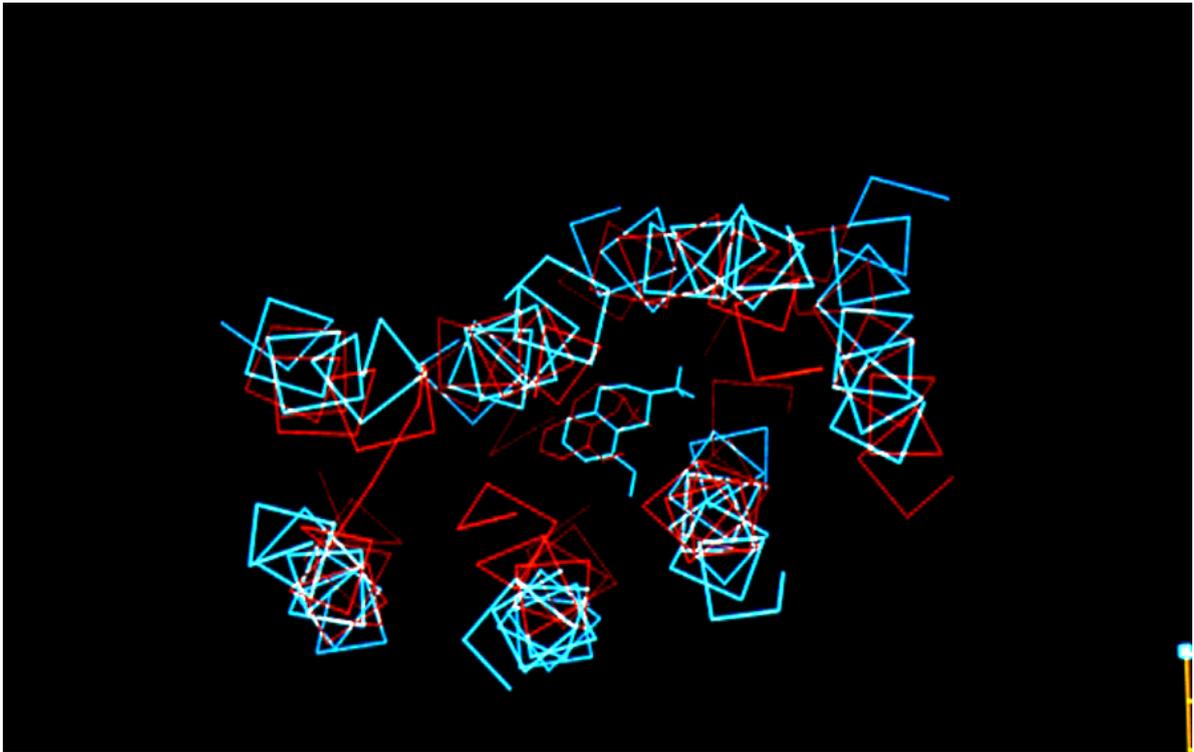


Photo 13. Complexe C13/récepteur.

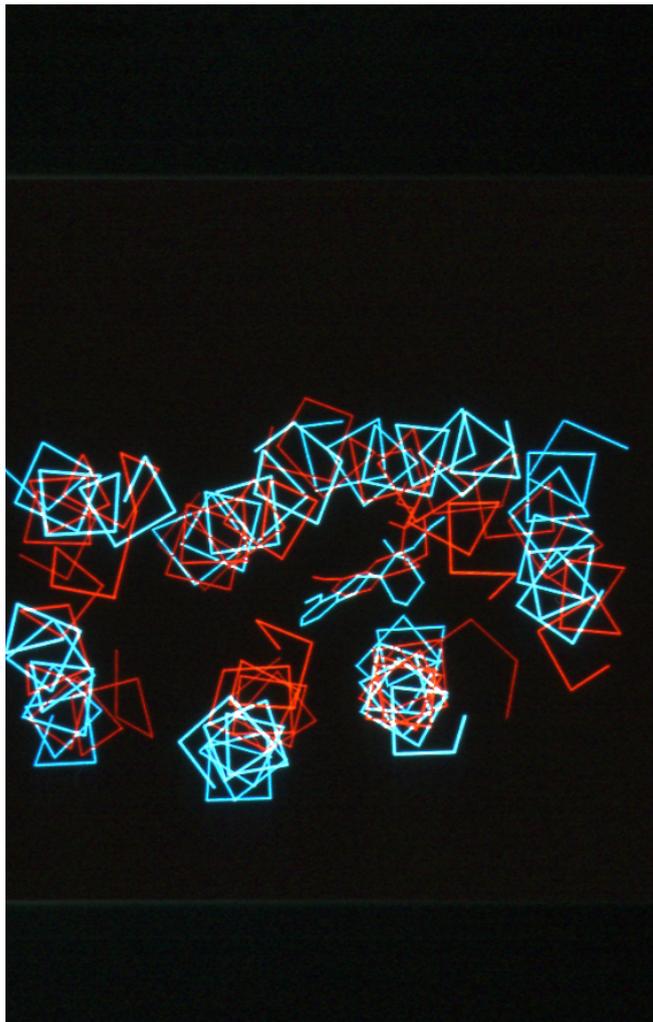


Photo 14. Complexe C50/récepteur.

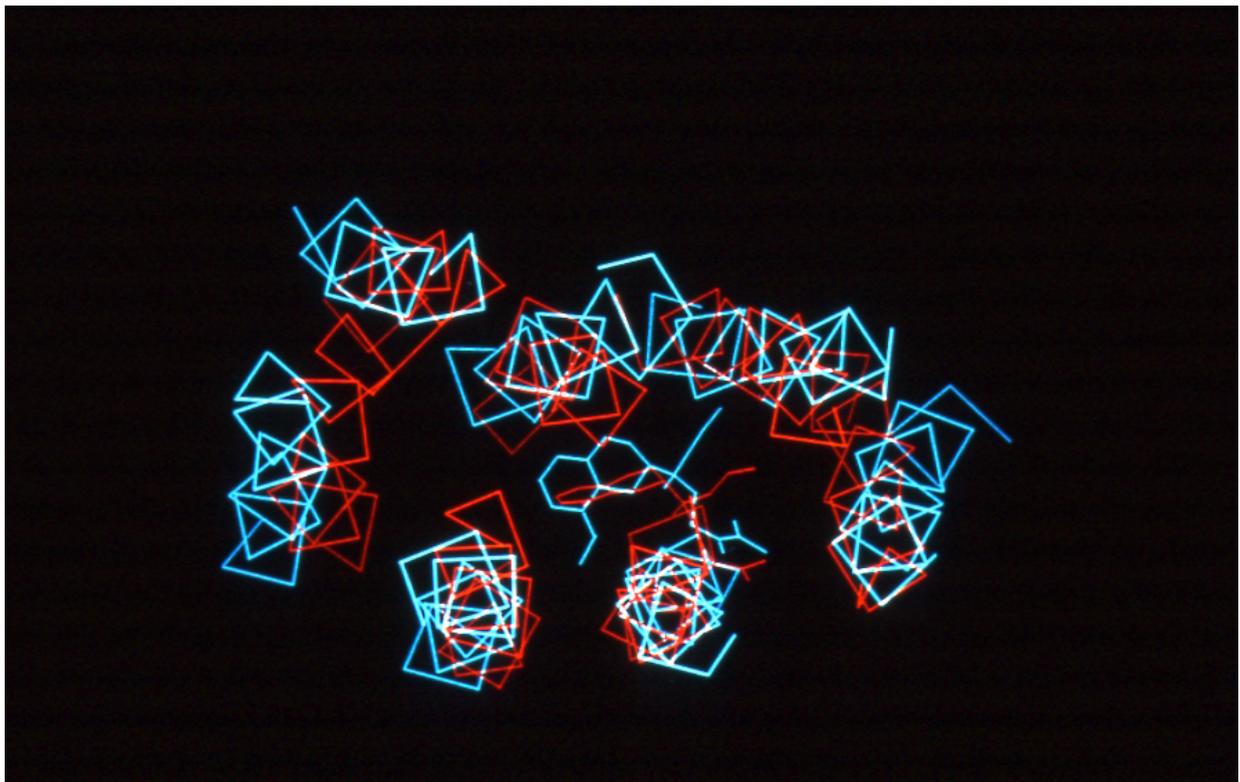


Photo 15. Complexe **C18**/récepteur.

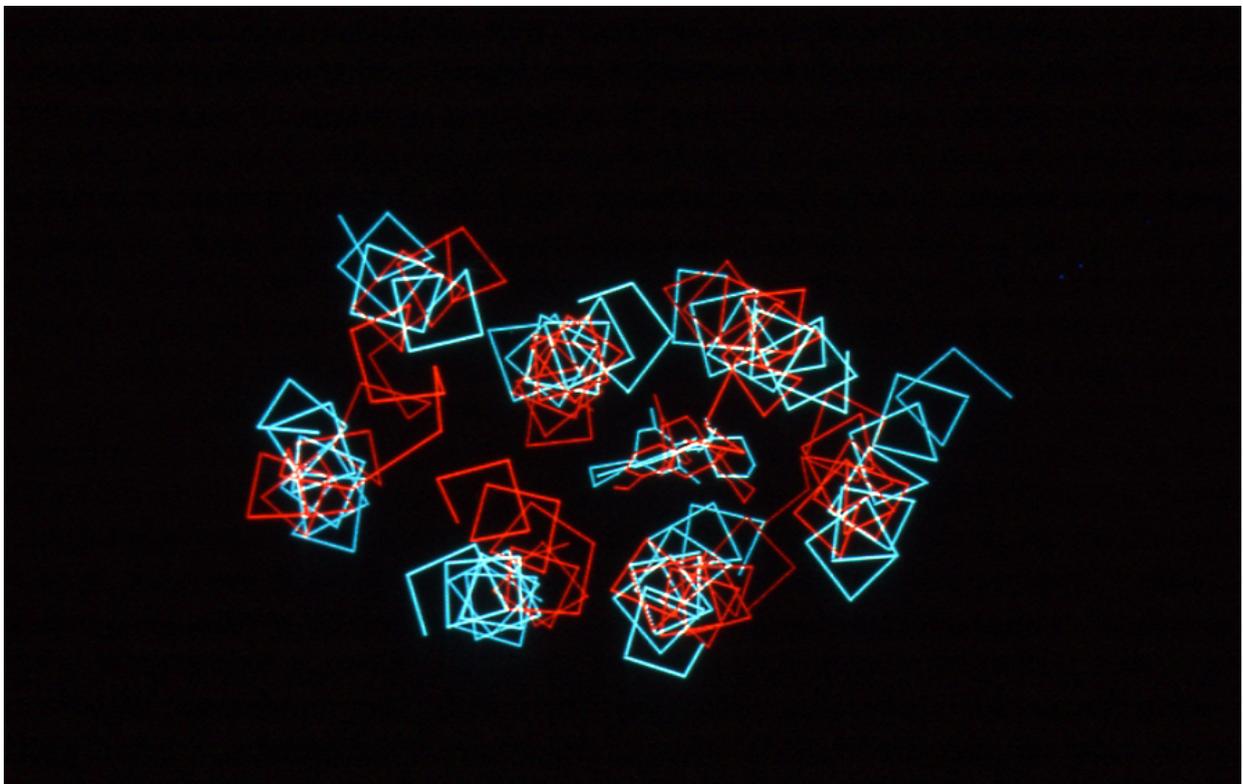


Photo 16. Complexe **C49**/récepteur.

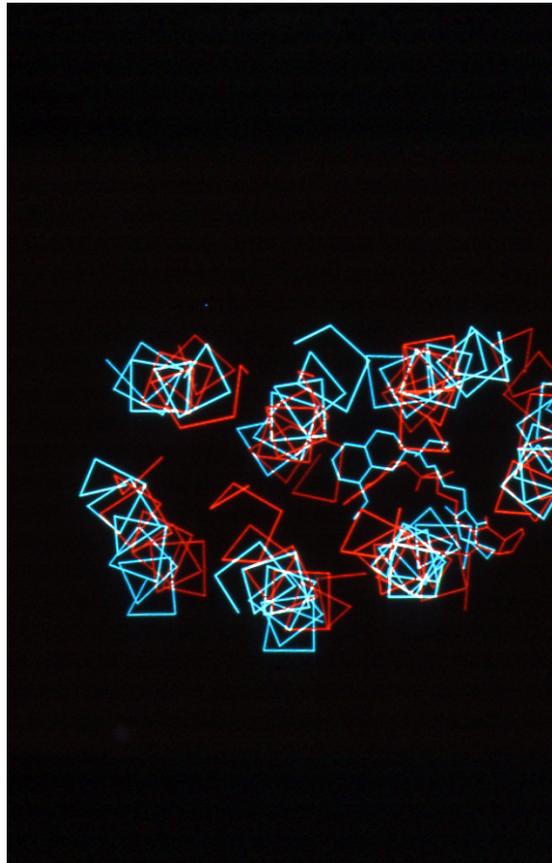


Photo 17. Complexe C30/récepteur.

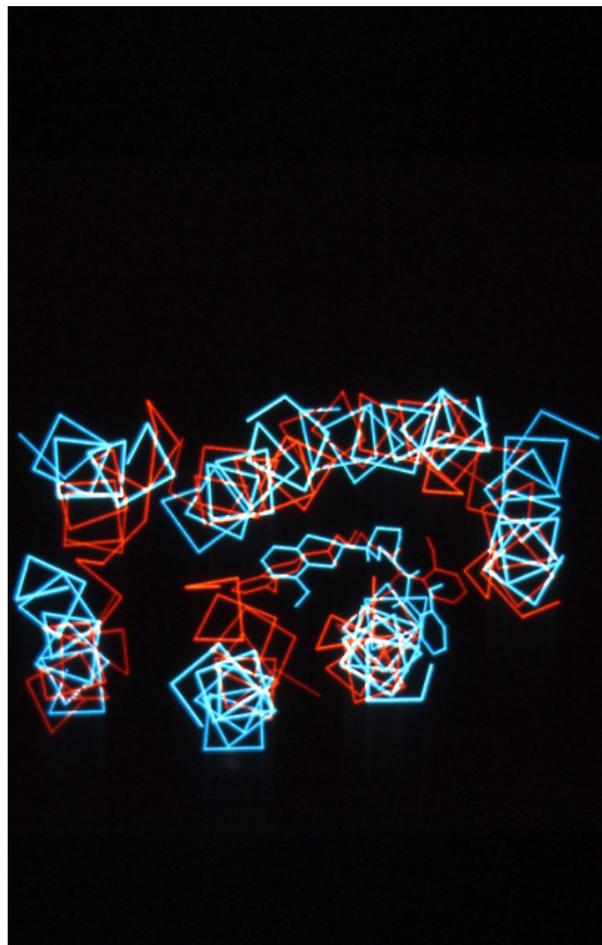


Photo 18. Complexe C41/récepteur.

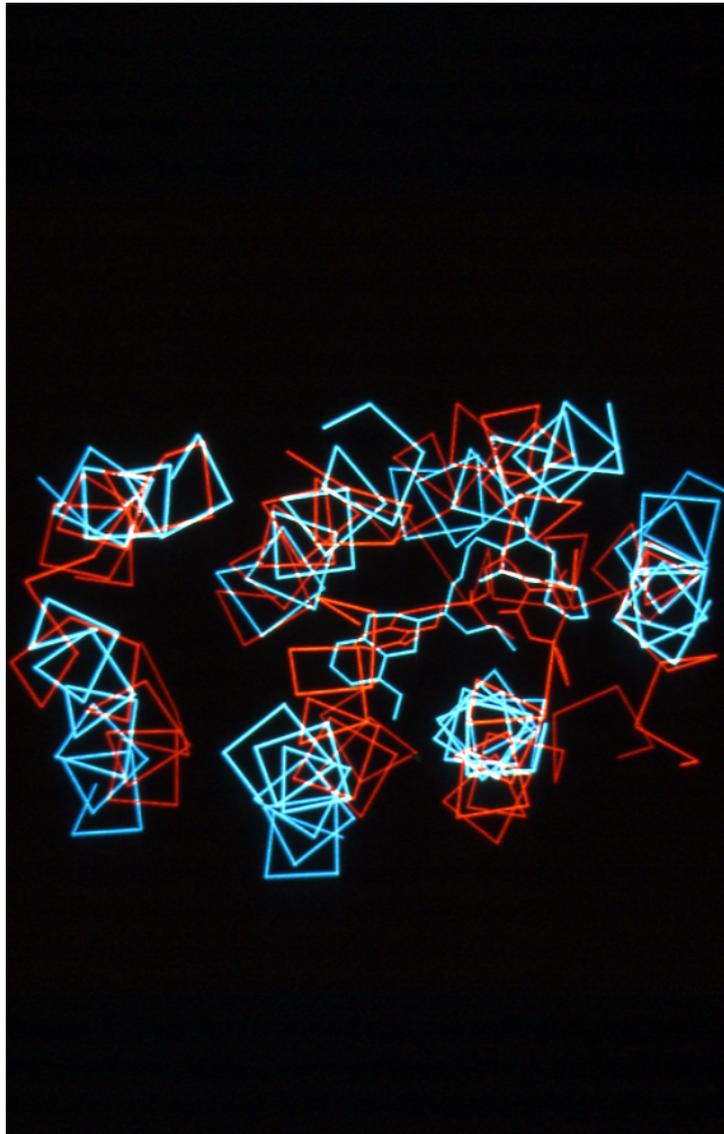


Photo 19. Complexe **C44**/récepteur.

Les résultats de cette visualisation graphique peuvent être résumés ainsi :

- durant les simulations de dynamique, la majorité des ligands effectuent une rotation pouvant aller jusqu'à 90° , afin de se positionner correctement au niveau du site actif.
- les chaînes latérales des ligands les plus actifs ont tendance à se placer entre l'hélice 7 et l'hélice 1. Les fonctions polaires comme par exemple les fonctions imide des chaînes latérales de ces ligands actifs établissent des interactions préférentielles avec les résidus de l'hélice 7. Ceci est vérifié pour tous les ligands ayant cette chaîne sauf pour le composé **C49** où cette partie mobile reste au centre du récepteur (Photo 16) et est dirigée vers la partie intra-cellulaire. Ce phénomène peut sans doute être relié à l'aspect rigide de ce composé favorisé par sa structure spiranique.

- la manipulation graphique utilisée pour positionner les ligands dans le site actif au début de l'analyse est importante. En effet, les composés n'effectuent pas de déplacement (longitudinale, latérale ou transversale) important dans l'ensemble de la cavité de la protéine comme on le voit sur ces photos.

- on observe en général des déformations plus importantes du récepteur lorsque le site actif est occupé par les ligands les plus actifs.

III. CONCLUSION ET PERSPECTIVES

Les études des interactions entre les ligands et les résidus des hélices transmembranaires du récepteur 5-HT_{1A} montrent que les ligands ont en général un comportement assez différent dans le site actif suivant leur affinité vis-à-vis de ce récepteur.

La majorité de ces ligands interagissent avec les résidus Asp82 de l'hélice 2, Ser393 (et Asn396 de l'hélice 7 seulement pour les résidus les plus actifs). Ceci est en bon accord avec les études de mutagenèse menées par Chanda *et al.*²⁸ qui ont révélées l'importance de ces trois résidus pour la reconnaissance des ligands 5-HT_{1A}.

Ces études d'interactions mettent aussi en évidence l'importance de deux autres acides aminés Ser123 et Tyr390 dans le processus de stabilisation des composés actifs au sein du site actif. L'hydrogène de la fonction alcool du résidu Ser123 de l'hélice 3, qui forme très souvent une liaison hydrogène avec l'atome d'oxygène du squelette chromane, semble avoir une importance non négligeable, dans la stabilité de cette famille de composés.

Les résultats des études de comparaison (RMS et RMSD) entre les différentes structures des complexes montrent l'influence des ligands sur la mobilité des hélices 5, 6 et 7 du récepteur. Les ligands les plus actifs (**C18**, **C49**, **C30**, **C41**, **C44**) semblent responsables d'une plus grande mobilité de ces hélices 5, 6 et 7. Cet aspect dynamique probablement relié à des modifications structurales des hélices TMH5, 6 et 7 peut être un facteur important dans les relations fonctionnelles du récepteur. Il peut être supposé qu'une mobilité plus importante des hélices 5, 6 et 7 associée à leur facilité de déformation puisse permettre la transmission du message aux protéines G via les résidus intra-cellulaires présents dans la boucle reliant les hélices 5 et 6 et dans la partie C-terminale.

Pour expliquer la différence d'activité pharmacologique des ligands étudiés ici, nous avancerons que les principales différences entre les ligands inactifs et actifs sont dues à :

- des interactions plus importantes de la partie commune des ligands (noyau chromane, fonctions méthoxy et amine protonée) avec les résidus du récepteur.

- des interactions plus spécifiques des fonctions polaires (par exemple fonction imide) de la chaîne latérale des composés actifs avec les résidus Tyr390 et Asn396 de l'hélice TMH7.

Suite à ces résultats, nous pouvons tenter de répondre aux questions que nous nous posions au début de ce chapitre.

La première question et la seconde question

"Les informations obtenues avec un ensemble homogène de ligands sont-elles similaires à celles obtenues avec des ensembles hétérogènes de composés? Sinon, met-on en évidence des interactions spécifiques à cette famille de composés?"

"Met-on en évidence une différence de comportement du complexe liée à l'affinité de ces ligands?"

Nous avons obtenu des résultats ayant beaucoup de points communs avec ceux obtenus dans des études antérieures (même interaction entre les ligands agonistes et le résidu Asp82 ; mobilité des hélices 5 et 6 plus importante avec la présence d'un agoniste fort dans le site actif) mais aussi démontré quelques spécificités de nos ligands comme le rôle de certaines fonctions, lié aux résidus Ser123, Ser393, Tyr390 et Asn396, dans la stabilité des composés dans le site actif.

La troisième question

"Peut-on par dynamique moléculaire (DM) obtenir des informations qualitatives et/ou quantitatives permettant de mieux prédire l'activité potentielle des ligands?"

Une information qualitative importante qui émerge de notre étude concerne la chaîne latérale des ligands. L'intérêt de la chaîne latérale possédant des fonctions polaires comme par exemple des fonctions imide est bien connu en SAR : son ajout sur un motif déjà actif en augmente en général l'activité. Nous pouvons, par nos résultats justifier et expliquer ce rôle. La chaîne latérale semble augmenter notablement l'affinité des ligands en se fixant principalement sur l'hélice 7 du récepteur. Lorsque la conformation des cycles spiraniques ne permet pas cette position de la chaîne latérale, l'affinité est moindre. Une étude plus exhaustive sur un ensemble de composés ayant des longueurs de chaîne et des fonctions polaires différentes permettrait de renforcer ces résultats.

Par contre il est certainement trop tôt pour conclure sur la question : "Si oui, ces informations peuvent-elles être introduites dans une approche QSAR afin de dépasser les limitations que nous avons trouvées précédemment?"

Le nombre limité de complexes ligand/récepteur que nous avons étudiés ne le permet pas actuellement. Néanmoins les données quantitatives telles que l'évaluation de la stabilité des complexes, les modifications de position et de mobilité des hélices transmembranaires sont des informations certainement utilisables dans une table QSAR.

Références Chapitre 4

-
- ¹ Sylte, I.; Edvardsen, O.; Dahl, S. G. *Protein Engineering* **1993**, 7, 691-700.
- ² Kuipers, W.; Van Wijngaarden, I.; Ijzerman, A. P. *Drug Design and Discovery* **1994**, 11, 231-249.
- ³ Zhang, D.; Weinstein, H. *J. Med. Chem.* **1993**, 36, 934-938.
- ⁴ Hibert, M. F.; Trumpp-Kallmeyer, S.; Bruinvels, A.; Hoflack, J. *Mol. Pharmacol.* **1992**, 40, 8-15.
- ⁵ Trumpp-Kallmeyer, S.; Hoflack, J.; Bruinvels, A.; Hibert, M. *J. Med. Chem.* **1992**, 35, 3448-3462.
- ⁶ Kobilka, B. K.; Frielle, T.; Collins, S.; Yang-Feng, T.; Kobilka, T. S.; Francke, U.; Lefkowitz, R. J.; Caron, M. G. *Nature* **1987**, 329, 75-79.
- ⁷ Bernstein, P. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, 112, 535-542.
- ⁸ Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. *J. Mol. Biol.* **1990**, 213, 899-929.
- ⁹ Baldwin, J. M.; Henderson, R.; Beckman, E.; Zemlin, F. *J. Mol. Biol.* **1988**, 202, 585-591.
- ¹⁰ Henderson, R.; Unwin, P.N.T. *Nature* **1975**, 257, 28-32.
- ¹¹ O'Dowd, B. F.; Hnatowich, M.; Regan, J. W.; Leader, W. M.; Caron, M. G.; Lefkowitz, R. J. *J. Biol. Chem.* **1988**, 263, 15985-15992.
- ¹² Higashijima, T. K.; Uzu, S.; Nakajima, T.; Ross, E. M. *J. Biol. Chem.* **1988**, 263, 6491-6494.
- ¹³ Lefkowitz, R. J.; Benovic, J. L.; Kobilka, B.; Caron, M. G. *Trends Pharmacol. Sci.* **1986**, 7, 444-448.
- ¹⁴ Dohlman, H. G.; Caron, M. G.; Lefkowitz, R. *J. Biochemistry* **1987**, 26, 2657-2664.
- ¹⁵ Haussdorff, W. P.; Hnantomich, M.; O'Dowd, B. F.; Caron, M. G.; Lefkowitz, R. J. *J. Biol. Chem.* **1990**, 265, 1388-1393.
- ¹⁶ Deisenhofer, J.; Michel, H. *Science* **1989**, 245, 1463-1467.

-
- ¹⁷ Deisenhofer, J.; Epp, O.; Miki, K.; Hubert, R.; Michel, M. *Nature* **1985**, 318, 618- 624.
- ¹⁸ Kühlbrandt, W.; Wang, D. N. *Nature* **1991**, 350, 130-134.
- ¹⁹ Dixon, R. A. F.; Sigal, I. S.; Rands, E.; Register, R. B.; Candelore, M. R.; Blake, A. D.; Strader, C. D. *Nature* **1987**, 326, 73-77.
- ²⁰ Dahl, S. G.; Edvardsen, O.; Sylte, I. *Proc. Natl. Acad. Sci. USA* **1991**, 88, 8111- 8115.
- ²¹ Hartig, P. R. *Trends Pharmacol. Sci.* **1989**, 10, 64-69.
- ²² Baldwin, J. M. *EMBO J.* **1992**, 13, 1693-1703.
- ²³ Tota, M. R.; Strader, C. D. *J. Biol. Chem.* **1990**, 265, 16891-16897.
- ²⁴ Rands, E.; Candelore, M. R.; Cheung, A. H.; Hill, W. S.; Strader, C. D.; Dixon, R. A. F. *J. Biol. Chem.* **1990**, 265, 10759-10764.
- ²⁵ George, S. T.; Ruoho, A. E.; Malbon, C. C. *J. Biol. Chem.* **1986**, 261, 16559-16564.
- ²⁶ Lefkowitz, R. J.; Kobilka, B. K.; Caron, M. G. *Biochem. Pharmacol.* **1989**, 38, 2941-2948.
- ²⁷ Ho, B. Y.; Karschin, A.; Brancheck, T.; Davidson, N.; Lester, H. A. *FEBS Lett.* **1992**, 312, 259-262.
- ²⁸ Chanda, P. K.; Minchin, M. C. W.; Davis, A. R.; Greenberg, L.; Reilly, Y.; McGregor, W. H.; Bhat, R.; Lubeck, M. D.; Mizutani, S.; Hung, P. P. *Mol. Pharmacol.* **1993**, 43, 516-520.
- ²⁹ Oxford Molecular Ltd, The Magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kingdom.
- ³⁰ Kyte, J.; Doolittle, R. *J. Mol. Biol.* **1982**, 157, 105-132.
- ³¹ Goldman, A.; Engelman, D.; Steitz, T. *Ann. Rev. Biophys. Chem.* **1986**, 15, 321- 353.
- ³² Van Gunsteren, W. F.; Berendsen, H. J. C. **1987**. Groningen Molecular Simulation (GROMOS) Library Manual. Biomos, Groningen.
- ³³ Stewart, J. J. P. *J. Comput.-Aided Mol. Design.* **1990**, 4, 1-105.
- ³⁴ Saudou, F.; Hen, R. *Med. Chem. Res.* **1994**, 4, 16-84.

-
- ³⁵ Van Gunsteren, W. F.; Berendsen, H. J. C. *J. Mol. Phys.* **1977**, 34, 1311-1327.
- ³⁶ Kjelsberg, M. A.; Cotecchia, S.; Ostrowski, J.; Caron, M. G.; Lefkowitz, R. J. J. *Biol. Chem.* **1992**, 267, 1430-1433.
- ³⁷ Luo X.; Zhang, D.; Weinstein, H. *Protein Engineering.* **1994**, 7, 1441-1448.

CONCLUSION GENERALE

Dans ce travail nous avons abordé par deux voies fondamentalement différentes l'étude par modélisation moléculaire de l'affinité de composés à visée sérotoninergique. La première se basait sur des approches QSAR, la deuxième était orientée sur l'étude des interactions ligand/récepteur par dynamique moléculaire.

Nous avons cherché les limites d'application des différentes techniques de QSAR. Des résultats obtenus, nous pouvons extraire les potentialités et limites suivantes. L'approche par les indices de similarité moléculaire conduit globalement à une validité statistique et un pouvoir prédictif meilleurs que l'approche par les descripteurs "classiques". Ces résultats sont cependant fortement liés à la composition de l'échantillon test utilisé et aux méthodes statistiques choisies. Il n'est pas possible a priori de définir la meilleure méthode qui devrait être utilisée pour étudier un nouvel échantillon. Pour notre part, afin d'obtenir un pouvoir prédictif optimum, nous avons testé sur nos échantillons les différentes méthodes et, si nécessaire, combiné l'ensemble de leurs résultats. Nous avons ainsi obtenu un écart moyen d'environ 1 unité de pIC_{50} sur l'ensemble des produits prédits. L'utilisation d'un échantillon important de composés à structure hétérogène permet d'appliquer le modèle à une grande gamme de produits mais ceci au détriment de la précision. A la fin de ce travail, nous sommes donc en possession de trois modèles quantitatifs de l'affinité sur le récepteur $5HT_{1A}$ directement utilisables avec une bonne précision (indoles, chromanes et thiochromanes) et un modèle généraliste (382 composés) moins précis.

L'approche par dynamique moléculaire des complexes ligand/récepteur a conduit par contre à des résultats principalement qualitatifs. Nous avons sur un échantillon de 8 ligands à squelette identique (Chromane), mais possédant une large gamme d'activité (4.70 à 9.70 unités de pIC_{50}), mis en évidence des différences de comportement des complexes ligand/récepteur lors d'études des interactions dans le site actif et lors d'études de l'influence des ligands sur la mobilité des hélices transmembranaires du récepteur. Ces résultats sont à rapprocher de certains obtenus antérieurement sur des études de mécaniques et dynamiques moléculaire effectuées sur des complexes

ligand/récepteur transmembranaire. Nous avons également noté des différences d'interaction spécifiques entre certains résidus (Asp82, Ser123, Tyr390, Ser393 et Asn396) et certaines fonctions des ligands ainsi que des interactions générées par le bras mobile qui expliquent l'augmentation d'affinité induite par la présence de ce bras. Ces résultats sont limités à un trop faible nombre de ligands pour pouvoir être introduits dans une étude QSAR mais certaines corrélations semblent potentiellement utilisables d'un point de vue statistique.

Ce travail a un prolongement évident dans l'approfondissement de l'étude de la dynamique des complexes ligand/récepteur et de l'introduction de ses résultats dans une approche QSAR. Cela devrait conduire à un meilleur pouvoir prédictif surtout dans le cas de structures proches comme des isomères de position ou des diastéréoisomères.