



**HAL**  
open science

# Modélisation du récepteur aux chimiokines C-C de type 5 : caractérisation des états conformationnels et conception rationnelle de modulateurs de la dimérisation

Florian Koensgen

## ► To cite this version:

Florian Koensgen. Modélisation du récepteur aux chimiokines C-C de type 5 : caractérisation des états conformationnels et conception rationnelle de modulateurs de la dimérisation. Chemo-informatique. Université de Strasbourg, 2018. Français. NNT : 2018STRAF026 . tel-02072380

**HAL Id: tel-02072380**

**<https://theses.hal.science/tel-02072380>**

Submitted on 19 Mar 2019

**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.

**ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES**

**UMR 7200**

**THÈSE** présentée par :

**Florian KOENSGEN**

Soutenue le : **04 octobre 2018**

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

Discipline/ Spécialité : **Chimie informatique et théorique**

**Modélisation du récepteur aux chimiokines  
C-C de type 5 : caractérisation des états  
conformationnels et conception rationnelle  
de modulateurs de la dimérisation**

**THÈSE dirigée par :**

**Mme KELLENBERGER Esther**

Professeure, Université de Strasbourg

**Mr ROGNAN Didier**

Directeur de recherche, Université de Strasbourg

**RAPPORTEURS :**

**Mr BUREAU Ronan**

Professeur, Université de Caen

**Mr FLOQUET Nicolas**

Chargé de recherche, Université de Montpellier

---

**AUTRES MEMBRES DU JURY :**

**Mme BRELOT Anne**

Directeur de recherche, Institut Pasteur de Paris

Avant tout j'aimerais grandement remercier Didier et Esther de m'avoir accueilli dans leurs laboratoires, de leur confiance et pour m'avoir soutenu tout au long de cette thèse.

Je suis très reconnaissant et j'adresse mes remerciements au Professeur Ronan Bureau et au Docteur Nicolas Floquet qui ont accepté de juger mon travail en tant que rapporteur et également au Docteur Anne BreLOT qui me fait l'honneur de participer au jury de cette thèse.

Je tiens à remercier toutes les personnes qui m'ont permis d'arriver à ce moment et qui m'ont apporté leur aide durant mon cursus universitaire. Je tiens tout particulièrement à remercier le Docteur Anne Tromelin du CSGA de Dijon, pour m'avoir ouvert les portes du monde des RCPGs et donné envie de poursuivre ma formation de chercheur. Mais également l'ANRS et le Sidaction pour le financement des projets auquel j'ai participé, ainsi que les centres de calcul HPC de Strasbourg et IN2P3 de Lyon, pour l'accès à leurs ressources.

Bien entendu il serait difficile de ne pas remercier tous les membres, de l'équipe de chémogénomique structurale : Guillaume (S.O.S. médecin des ordinateurs toujours présent en cas de pépins) ; Franck (fana de film et grand débateur de l'univers Marvel) ; Priscila (parce que le quinoa en sachet c'est toujours mieux) ; tous les post-doctorants non-francophones qui ont eu le plaisir d'entendre mon accent anglais et m'ont aidé dans mes projets : Gosia (toujours présente pour les biscuits), Tim (comparse de cinéma à ses heures perdues) et Inna (parce que "tu verras c'est facile") ; Célien (également, grand consommateur de café) et bien sûr Joël, stagiaire Espagnol venu se perdre dans le froid alsacien. Je remercie de même, tous les autres membres de l'unité 7200 dont l'équipe de l'aquarium, Ludo, Patrick, Séverine, Stéphanie, Yann et tous les autres (pour tous ces midis de repas divers, kebab et raclettes) ; tous les autres doctorants de l'année 2015 : Laura, Laure, Lucie et Nicolas ; David, Sophie et Françoise (pour avoir répondu à mes requêtes diverses et variées de matériels et de bons de commande) ; et Bruno (pour l'accès à la partie réfrigérée de la chimiothèque).

Bien sûr, je remercie également tous mes amis, qui me supportent au quotidien depuis de multiples années, en commençant par Alex (on fera peut-être un jour cette randonnée dans les alpes), fournisseur officiel de divertissements et de vacances par procuration ; les membres de la ligue de donjon & dragon, Etienne, Odile, Greg, Roms, Sha, Matthieu, Thomas et Aurore ; ceux qui m'ont maintenu en forme physiquement : mes binômes de varappe, Adrien et Nicolas et Maxime pour le footing ; évidemment Vivien, Mika, Steve, Hervé et tous les

autres membres de la Movie Team ; toute la clique de Lyon; Romain pour les nombreuses soirées jeux de société et bien sur tous les camarades de master et notamment Benjy (« Si java bien c'est javamine »), Bruno, Nico (pour les weekends cochonnaille) et temps d'autre.

Enfin, je remercie tous les membres de ma famille, et en particulier ma mère, mon père et ma grand-mère qui m'ont toujours poussé à me dépasser et m'ont toujours soutenu au quotidien.



*“Intuition is the art, peculiar to the human mind, of working out the correct answer from data that is, in itself, incomplete or even, perhaps, misleading.”*

— Isaac Asimov, *Forward the Foundation*, 1993.

# Table des matières

**Table des matières** ..... IV

**Glossaire des Abréviations**..... VII

## **Chapitre 1 : Les récepteurs couplés aux protéines G et le récepteur aux chimiokines C-C de type 5... 1**

1.	Les récepteurs couplés aux protéines G.....	2
1.1	Une large famille de récepteurs.....	3
1.2	Une structure commune.....	6
2.	Relation entre la structure 3D des récepteurs et l'état d'activation.....	7
2.1	Historique et état des lieux des structures.....	7
2.2	Signatures structurales d'état d'activation.....	10
2.3	Domaine de liaison des ligands.....	12
2.4	Domaine de liaison des effecteurs.....	13
2.5	Influence des mutations d'acides aminés.....	14
3.	Simulation par dynamique moléculaire.....	14
4.	Récepteurs aux chimiokines C-C de type 5.....	15
5.	Conclusion.....	21
6.	Bibliographie de chapitre.....	22

## **Chapitre 2 : Distinction des récepteurs couplés aux protéines G et de leurs états d'activation par l'analyse de leurs interactions intramoléculaires transmembranaires. .... 26**

1.	Introduction.....	27
2.	Matériel et méthode.....	30
2.1	Matériel.....	30
2.1.1	Modèles de référence du domaine transmembranaire des récepteurs couplés aux protéines G.....	30
2.1.2	Structures de récepteurs couplés aux protéines G dans la PDB.....	30
2.1.3	Trajectoire de dynamique moléculaire de récepteurs couplés aux protéines G.....	33
2.2	Méthodes d'analyse et de classification.....	34
2.2.1	Interprétation des structures et création de représentations simplifiées.....	35
2.2.2	Comparaison des structures.....	40
2.2.3	Partitionnement de structures 3D.....	42
2.2.4	Création des empreintes d'interactions.....	43
3.	Résultats et discussion.....	44
3.1	Distinction des classes de récepteurs couplés aux protéines G et de leurs états d'activation	44

3.1.1	Description des structures expérimentales de la PDB .....	44
3.1.2	Partition des structures .....	46
3.1.3	Interactions prépondérantes.....	51
3.1.4	Conclusion .....	54
3.2	Distinctions des états conformationnels d'une trajectoire de dynamique moléculaire .....	55
3.2.1	Récepteur muscarinique de type 2 à l'acétylcholine (ACM2) .....	56
3.2.2	Récepteur adrénergique $\beta$ de type 2 (ADRB2) .....	59
3.2.3	Conclusion .....	65
4.	Conclusion .....	66
5.	Bibliographie de chapitre.....	67

### **Chapitre 3 : Analyse conformationnelle du récepteur CCR5 ..... 70**

1.	Introduction.....	71
2.	Matériel et méthodes.....	75
2.1	Matériel .....	75
2.1.1	Préparation des systèmes .....	75
2.1.2	Dynamiques moléculaires du récepteur CCR5 .....	78
2.2	Analyses des trajectoires de Dynamique moléculaire .....	81
2.2.1	Déviations par rapport aux coordonnées initiales.....	81
2.2.2	Fluctuations atomiques .....	81
2.2.3	Marqueurs structuraux de l'activation.....	81
2.2.4	Corrélations de mouvement .....	82
2.2.5	Caractérisation des mouvements.....	82
2.3	Identification de motifs d'interactions intramoléculaires.....	83
3.	Résultats .....	85
3.1	Dynamique moléculaire classique.....	85
3.1.1	Déviations par rapport aux coordonnées initiales.....	85
3.1.2	Paramètres d'accélération .....	87
3.2	Dynamique moléculaire accélérée .....	87
3.2.1	Déviations par rapport aux coordonnées initiales.....	87
3.2.2	Fluctuation atomique .....	89
3.2.3	Marqueurs structuraux de l'activation.....	91
3.2.4	Corrélations de mouvement .....	93
3.2.5	Mouvements du récepteur CCR5 .....	96
3.2.6	Identification de motifs d'interactions intramoléculaires.....	108
4.	Conclusion .....	120
5.	Bibliographie de Chapitre .....	122

<b>Chapitre 4 : Recherche de modulateurs de la dimérisation de CCR5 .....</b>	<b>126</b>
1. Introduction .....	127
2. Matériel et méthodes.....	130
2.1 Matériel .....	130
2.1.1 Chimiothèque .....	130
2.1.2 Structures tridimensionnelles des dimères I5 et I56.....	130
2.2 Sélection d'une cavité représentative.....	132
2.2.1 Caractérisation de l'interface .....	132
2.2.2 Détection de cavités .....	132
2.2.3 Sélection du site d'intérêt .....	133
2.3 Criblage virtuel .....	134
2.3.1 Création d'un pharmacophore .....	135
2.3.2 Criblage d'une chimiothèque .....	135
2.3.3 Sélection des touches d'intérêt.....	136
3. Résultats et discussion.....	137
3.1 Recherche d'un site droguable.....	138
3.1.1 Analyse des interfaces des dimères I5 et I56 .....	138
3.1.2 Analyse des cavités à l'interface des dimères .....	145
3.2 Criblage virtuel .....	153
3.2.1 Description du pharmacophore .....	154
3.2.2 Filtrage et tri de la chimiothèque .....	156
4. Conclusion .....	160
5. Bibliographie de chapitre.....	161
<b>Chapitre 5 : Conclusions et perspectives.....</b>	<b>164</b>
<b>Annexes .....</b>	<b>167</b>

# Glossaire des Abréviations

<b>3D :</b>	Tri-dimensionnel
<b>5THB1 :</b>	5-Hydroxytryptamine-1B
<b>5HTB2 :</b>	5-Hydroxytryptamine-2B
<b>AA2AR :</b>	Récepteur à l'adénosine de type 1
<b>AA2AR :</b>	Récepteur à l'adénosine de type 2
<b>ACM2 :</b>	Récepteur muscarinique de type 2 à l'acétylcholine
<b>ACP :</b>	Analyse en composante principale
<b>ADRB2 :</b>	Récepteur $\beta$ -adrénergique de type 2
<b>CALCR :</b>	Récepteur Calcitonine
<b>CD4 :</b>	Cluster de différenciation 4
<b>CHA :</b>	Classification hiérarchique ascendante
<b>CCR2 :</b>	Récepteur couplé aux chimiokines C-C de type 2
<b>CCR5 :</b>	Récepteur couplé aux chimiokines C-C de type 5
<b>CCR9 :</b>	Récepteur couplé aux chimiokines C-C de type 9
<b>cryo-EM :</b>	Cryo-microscopie électronique
<b>CXCR4 :</b>	Récepteur aux chimiokines CXC de type 4
<b>Droguabilité :</b>	Capacité d'un site de protéine à lier à un ligand droguable avec une haute affinité
<b>ECL :</b>	Boucle extracellulaire
<b>GaMD :</b>	Dynamique moléculaire accélérée par gaussienne
<b>GDP :</b>	Guanosine diphosphate
<b>Gp :</b>	Glycoprotéine
<b>GTP :</b>	Cofacteur guanosine triphosphate
<b>GLP1R :</b>	Récepteur glucagon like peptide 1
<b>HBA :</b>	Accepteur de liaison hydrogène
<b>HBA/HBD :</b>	Accepteur/donneur de liaison hydrogène
<b>HBD :</b>	Donneur de liaison hydrogène
<b>ICL :</b>	Boucle intracellulaire
<b>MVC :</b>	Maraviroc
<b>OPRK :</b>	Récepteur $\kappa$ -opiacé
<b>OPRM :</b>	Récepteur $\mu$ -opiacé

<b>OPSD :</b>	Récepteur rhodopsine
<b>PDB :</b>	Protein Data Bank
<b>PMF :</b>	Potentiel de force moyenne
<b>RCPG :</b>	Récepteur Couplé aux Protéines G
<b>RMN :</b>	Résonance magnétique nucléaire
<b>RMSD :</b>	Déviatiion de l'écart quadratique moyen
<b>RX :</b>	Rayons X
<b>Sida :</b>	Syndrome d'immunodéficience acquise
<b>TM :</b>	Transmembranaire
<b>VIH :</b>	Virus de l'immunodéficience humaine

# Chapitre 1

Les récepteurs couplés aux protéines G et le récepteur aux chimiokines C-C de type 5

Aux cours de mes trois années de thèse, je me suis intéressé à l'étude d'un récepteur membranaire, le récepteur aux chimiokines C-C de type 5 ou CCR5. Ce récepteur fait partie de la famille des récepteurs couplés aux protéines G ou RCPGs. Dans ce chapitre d'introduction je ferai une courte introduction sur les RCPGs, sur leurs caractéristiques structurales et fonctionnelles avant de finir sur les caractéristiques et informations générales sur le récepteur CCR5.

## 1. LES RÉCEPTEURS COUPLÉS AUX PROTÉINES G

---

Les récepteurs couplés aux protéines G sont des récepteurs membranaires. Leur fonction est de transmettre des signaux à travers la membrane cellulaire. Ils sont ainsi impliqués dans de nombreux mécanismes biologiques tels que la vision, l'odorat, le goût, la transmission neuronale, la régulation des émotions ou l'immunité.

La transmission du signal par les RCPGs à travers la membrane permet l'activation de différentes voies de signalisation (Figure 1- 1), généralement au travers du couplage d'une protéine G<sup>1</sup>. Une protéine G est formée par un hétérotrimère composé des sous-unités G $\alpha$ , G $\beta$  et G $\gamma$ . Au total, chez l'homme il existe 21 différentes sous-unités  $\alpha$ , 6 sous-unités  $\beta$  et 12 sous-unités  $\gamma$ . Au sein de l'hétérotrimère, la sous-unité  $\alpha$  est celle se liant au récepteur. Elle est représentée par quatre différentes familles selon leurs séquences en acides aminés et leurs fonctions : les protéines G $\alpha$ -s, G $\alpha$ -i/o, G $\alpha$ -q/11 et G $\alpha$ -12/13. La liaison de la sous-unité G $\alpha$  et du cofacteur guanosine diphosphate (GDP) induit la formation de l'hétérotrimère G $\alpha\beta\gamma$ . L'activation des RCPGs permet la liaison de la sous-unité G $\alpha$  de l'hétérotrimère au récepteur et promeut la dissociation du cofacteur GDP de celle-ci. Un fois la dissociation effectuée, le GDP est remplacé par le cofacteur guanosine triphosphate (GTP). Ce remplacement du GDP par du GTP entraîne la dissociation de la sous-unité G $\alpha$  des sous-unités G $\beta$  et G $\gamma$ , permettant la poursuite de la signalisation. Le cycle de signalisation se termine par le remplacement du GTP lié à la sous-unité G $\alpha$  par un GDP par un mécanisme d'hydrolyse et la reformation de l'hétérotrimère G $\alpha\beta\gamma$ .

Une seconde voie de signalisation peut être activée par les RCPGs. Elle implique le couplage et l'activation d'une arrestine<sup>2</sup>. Dans ce cas-ci, la partie C-terminale des RCPGs est



phosphorylée par des kinases permettant la liaison d'arrestine au récepteur et leur activation. Chez l'homme, il existe 4 types d'arrestine, les arrestine-1, 2, 3 et 4.

### Figure 1-1:

Figure n°1 de la revue :

Hilger, D., Masureel, M. & Kobilka, B. K. Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).

(<https://doi.org/10.1038/s41594-017-0011-7>)

Figure 1- 1: Schéma de la signalisation par les RCPGs<sup>3</sup>.

L'activation des RCPGs, résulte de leur stimulation extracellulaire par un ligand ou stimulus endogène (ion, lipide ou glycoprotéine par exemple), ou exogène (par exemple un médicament, un photon ou une molécule odorante). Elle est caractérisée au niveau moléculaire par des changements conformationnels. Les conformations activées sont caractérisées par la présence d'un ligand activateur (agoniste) et par leur aptitude à lier un effecteur dans le domaine intracellulaire. Les conformations inactivées sont, au contraire, inaptes à l'interaction avec une protéine G ou une arrestine.

#### 1.1 UNE LARGE FAMILLE DE RÉCEPTEURS

En fonction de leur séquence en acides aminés et de leurs rôles fonctionnels, les RCPGs peuvent être séparés dans différentes classes. L'une des classifications les plus fréquemment utilisées sépare les RCPGs en 6 familles<sup>4</sup>. Cette classification a été effectuée en comparant des empreintes spécifiques de la séquence en acides aminés des récepteurs. Les 6 familles sont : la famille A pour les récepteurs à la rhodopsine, B pour les récepteurs à la sécrétine, C pour

les récepteurs métabotropiques du glutamate et aux phéromones, D pour les récepteurs couplés aux phéromones fongiques, E pour les récepteurs à l'AMP cyclique et F pour les récepteurs *Frizzled/Smoothened*. Les RCPGs de classe A, B, C, F sont présents chez les mammifères contrairement aux RCPGs de classe D et E. Une classification plus récente, la classification GRAFS<sup>5</sup>, se focalise sur les séquences de RCPGs de mammifères et distingue 5 familles et diverses sous-familles (Figure 1- 2). Les cinq familles de la classification GRAFS, coïncident en grande partie avec la classification A-F. Les différentes classes sont : la classe des récepteurs métabotropiques glutamate (ou classe C), la classe des récepteurs de type rhodopsine (ou classe A), la classe des récepteurs de type adhésion, la classe des récepteurs *Frizzled/Taste 2* et la classe des récepteurs de type sécrétine.

Chez l'homme, les RCPGs forment l'une des plus grandes familles de récepteurs avec plus de 800 récepteurs fonctionnels recensés. Parmi ceux-ci, une grande partie correspond à la famille des récepteurs olfactifs. Ces récepteurs forment une sous-famille des récepteurs de classe A. Historiquement elle forme une classe d'étude spécifique au sein des RCPGs du fait de sa seule implication dans l'olfaction. Les récepteurs n'appartenant pas à la sous-famille des récepteurs olfactifs sont au nombre de 395 et se répartissent dans toutes les familles citées plus haut. Chez l'homme, la classe des récepteurs rhodopsine est la plus conséquente avec 289 récepteurs répartis dans 11 sous-familles. Les autres classes sont composées de 15 récepteurs de type sécrétine, de 33 récepteurs de type adhésion, de 22 récepteurs de type glutamate et de 36 récepteurs *Frizzled/Taste 2*.

De par leur nombre et les différentes fonctions qu'ils régulent, les RCPGs constituent une famille importante de cibles thérapeutiques. Approximativement 34% des médicaments actuellement approuvés par la FDA<sup>6</sup> ciblent effectivement les RCPGs. Parmi ces médicaments on peut citer des traitements contre le diabète ciblant le récepteur *glucagon-like peptide 1* (classe sécrétine), contre les nausées et vomissements ciblant les récepteurs neurokinine-1 (classe rhodopsine) ou contre l'infection par le virus de l'immunodéficience humaine (VIH) ciblant le récepteur aux chimiokines CC de type 5 (classe rhodopsine).

## Récepteurs de type Rhodopsine

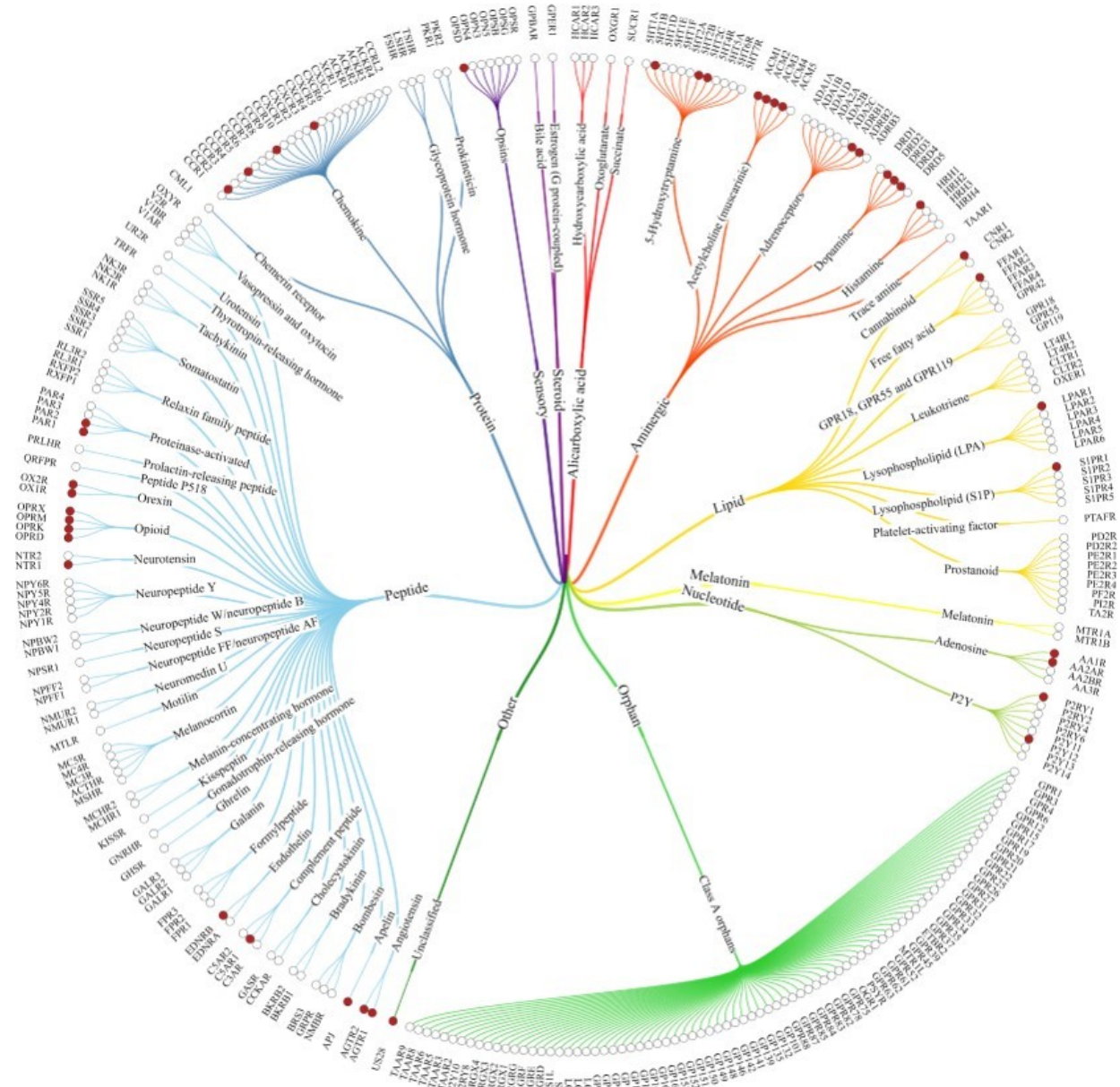


Figure 1- 2 : Classification des RCPGs. La classe des récepteurs de type rhodopsine est représentée sur cette page, les autres classes sont présentées sur la page suivante. Les familles sont divisées en sous-familles dont les membres sont indiqués par leur nom de gène. Les gènes marqués en rouge codent pour une protéine dont la structure 3D a été résolue par une méthode expérimentale. Source : <http://gpcrdb.org/structure/statistics>.



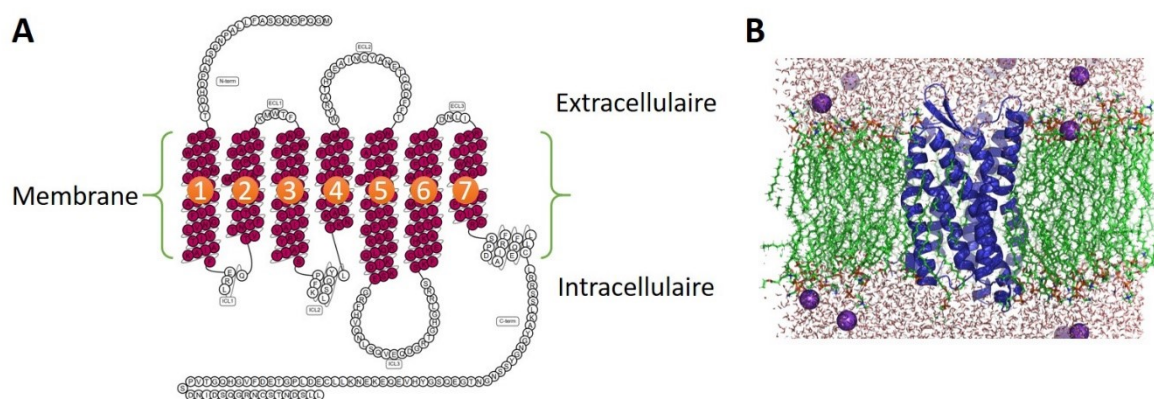


Figure 1- 3 : (A) Structure primaire et secondaire illustrée par un tracé de type " serpent" pour le récepteur à la rhodopsine (UNIPROT OPSD\_HUMAN). Source : [http://gpcrdb.org/proteine/OPSD\\_human/](http://gpcrdb.org/proteine/OPSD_human/). (B) Structure tertiaire du récepteur CCR5 modélisée dans une bicouche lipidique (vert) hydratée.

Le domaine intracellulaire des RCPGs est constitué de l'hélice 8, de la partie C-terminale et des boucles intracellulaires (ICL) 1, 2 et 3. La taille de la partie C-terminale varie d'une séquence courte de deux résidus pour le récepteur à la dopamine D2 à une séquence longue de 372 résidus pour le récepteur métabotrope glutamate 5. Parmi les trois boucles intracellulaires, ICL3 est celle dont la longueur est la plus variable, allant de quelques résidus à plus de 200 résidus pour les récepteurs muscariniques.

Le domaine transmembranaire est constitué d'un regroupement antiparallèle de 7 hélices  $\alpha$  qui composent le cœur du récepteur. Chaque hélice transmembranaire est dénommée par l'abréviation TM-X, X allant de 1 à 7. Chaque hélice est composée d'environ 25 résidus. Les structures en hélice peuvent contenir des cassures ou des irrégularités qui sont liées à la présence de résidus proline ou glycine. On peut également noter la présence de ponts disulfure entre résidus cystéine, qui participent à l'architecture générale des RCPGs. Ceux-ci sont variables dans les différentes familles de RCPG à l'exception du pont disulfure extrêmement conservé entre le TM3 et ECL2.

## 2. RELATION ENTRE LA STRUCTURE 3D DES RÉCEPTEURS ET L'ÉTAT D'ACTIVATION

### 2.1 HISTORIQUE ET ÉTAT DES LIEUX DES STRUCTURES

La première structure de RCPG a été résolue par diffraction aux rayons X en l'an 2000. Cette structure du récepteur de la rhodopsine bovine a une résolution de 2,8 Å et caractérise le



récepteur lié au 11-cis-rétinal dans un état inactivé (code pdb 1F88<sup>7</sup>). Le récepteur de la rhodopsine est resté pendant longtemps le seul récepteur cristallisé. Il a fallu attendre 2007 et 2008 pour obtenir la structure d'un second et troisième récepteur. La structure du second récepteur caractérise le récepteur adrénergique  $\beta$  de type 2 humain dans un état inactivé lié au ligand agoniste inverse carazolol (code pdb 2RH1<sup>8</sup>, résolution de 2,4 Å). La structure du troisième récepteur cristallisé caractérise le récepteur adénosine AA2AR dans un état inactivé lié au ligand antagoniste ZM241385 (code pdb 3EML<sup>9</sup>, résolution de 2,6 Å). Ces trois récepteurs font partis de la classe A. Le premier récepteur appartenant à une autre classe a été cristallisé en 2013. Il s'agit du récepteur « *Smoothened* » de la classe des récepteurs *Frizzled*, dans un état inactif lié au ligand antagoniste LY2940680 (code pdb 4JKV<sup>10</sup> résolution de 2.5 Å). A l'heure actuelle 52 récepteurs orthologues ont été décrits dans 270 structures cristallographiques. Parmi ces récepteurs, la grande majorité représente la classe des récepteurs de type rhodopsine, avec 45 récepteurs (240 structures). Les autres classes représentées sont la classe des récepteurs de type sécrétine, avec 4 récepteurs (13 structures), la classe des récepteurs de type glutamate avec deux récepteurs (6 structures) et la classe des récepteurs de type *Frizzled*/Goût 2, avec 1 récepteurs (11 structures) (Annexe 1).

La cristallisation des RCPGs est rendue difficile par leur flexibilité naturelle et la présence obligatoire d'un environnement mimant une membrane biologique. Un grand nombre de RCPGs ont pu être cristallisés grâce à l'introduction d'une protéine de fusion et/ou de mutations qui stabilisent une conformation inactive du récepteur. La cristallisation de RCPGs dans un état activé ou pré-activé a impliqué la formation d'un complexe entre le récepteur et une protéine mimant une protéine G, un anticorps, ou une protéine G mutée ayant une moindre flexibilité<sup>11</sup>. La première structure de ce type est celle du récepteur adrénergique  $\beta$  de type 2 lié à un anticorps de chameau, publiée en 2011 (code pdb 3POG<sup>12</sup>, résolution de 3.5 Å). La même année, le premier complexe de RCPG avec un hétérotrimère de protéines G a été cristallisé. Il s'agit du récepteur adrénergique  $\beta$  de type 2 lié à l'hétérotrimère de protéine Gs et à un fragment d'anticorps de chameau (code pdb 3SN6<sup>13</sup>, résolution de 3,2 Å). A l'heure actuelle, seules dix structures de RCPG ont été résolues, la plupart par cryomicroscopie électronique, avec un hétérotrimère de protéines G. En plus de celle du récepteur adrénergique, il s'agit des structures du récepteur  $\mu$  opioïde (une structure),  $\beta$ 2-adrenergique (une structure), GLP1R (deux structures), Calcitonine (une structure), Adénosine A2 (une

structure), Adénosine A1 (une structure) et 5HT-1B (une structure). Ces structures représentent des protéines G de type  $\alpha$ -s et  $\alpha$ -i. Sur les 52 récepteurs caractérisés par les structures de la PDB, 18 ont au moins une structure liée à une protéine G ou une arrestine (Tableau 1- 1).

Tableau 1- 1 : Structures de RCPG caractérisées dans un état activé. Les codes pdb en vert correspondent aux récepteurs liés à un hétérotrimère de protéine G

Récepteur	Entrées PDB de l'état actif	Effecteur
<b>5-Hydroxytryptamine-1B (5THB1)</b>	6G79	Protéine Gi
<b>5-Hydroxytryptamine-2B (5HTB2)</b>	5TUD	Anticorps Anti-5-HT2B Fab
<b>Adénosine A1 (AA1AR)</b>	6D9H	Protéine Gi
<b>Adénosine A2 (AA2AR)</b>	5G53, 6GDG	Protéine G $\alpha$ -s
<b>Calcitonine (CALCR)</b>	5UZ7	Protéine G-s
<b>Glucagon like peptide 1 (GLP1R)</b>	5VAI, 6B3J	Protéine G-s
<b>Muscarinique (ACM2)</b> 2	4MQS, 4MQT	Anticorps
<b>Rhodopsine (OPSD)</b>	2X72, 3DQB, 3PQR, 4A4M, 4BEY, 4J4Q, 4PXF, 4X1H, 4ZWJ, 5DGY, 5EN0, 5W0P, 5WKT, 6CMO	Protéine Gt, Protéine Gt-alpha3, Protéine G-i, fragment d'arrestine, S-arrestine
<b>US28</b>	4XT1	Protéine mimétique
<b><math>\beta</math>2-adrenergique (ADRB2)</b>	3P0G, 3SN6, 4LDO, 4LDL, 4LDE, 4QKX	Protéine mimétique, protéine G-s, Anticorps
<b><math>\kappa</math>-opiacé (OPRK)</b>	6B73	Anticorps
<b><math>\mu</math>-opiacé (OPRM)</b>	5C1M, 6DDF, 6DDE	Anticorps, Protéine G-i

Plusieurs sites web, comme notamment GPCRdb<sup>14</sup> (<http://gpcrdb.org/>) ou GPCR-exp (<https://zhanglab.cmb.med.umich.edu/GPCR-EXP/>) traitent des RCPGs et répertorient leurs différentes structures. En plus de la liste des diverses structures de RCPGs résolues, on peut trouver sur le site GPCRdb de nombreuses informations d'annotations. Cette liste et ces informations sont mises à jour régulièrement. L'un des principaux avantages de GPCRdb est son archivage des séquences des structures à l'aide de la numérotation Ballesteros-Weinstein<sup>15</sup>. Cette annotation, permet de comparer entre eux différents RCPGs sans avoir à aligner leur séquence. Elle caractérise les résidus du domaine transmembranaire par deux chiffres séparés par un point. Le premier chiffre caractérise le numéro de l'hélice auquel appartient le résidu et va de 1 à 7. Le second, caractérise la place du résidu au sein de son hélice et varie selon la taille de celle-ci entre 20 et 70, il augmente dans le sens de la séquence. Dans chaque hélice, le résidu X.50 est un résidu hautement conservé dans tous les RCPGs

d'une même classe. En plus d'archiver les structures de RCPG, le site GPCRdb fournit des outils pour générer des arbres phylogénétiques, pour modéliser la structure tridimensionnelle de RCPG, pour caractériser des acides aminés mutants (naturel et artificiel) ou des résidus impliqués dans la liaison des protéines G ou des ligands.

## 2.2 SIGNATURES STRUCTURALES D'ÉTAT D'ACTIVATION

Le lien entre la liaison d'un ligand, les changements conformationnels au sein du récepteur et l'activation des effecteurs de signalisation est aujourd'hui le sujet de nombreuses études. Avec de plus en plus de structures disponibles et surtout de structures résolues avec un effecteur de signalisation, des éléments structuraux communs aux changements conformationnels ont pu être identifiés au sein du domaine transmembranaire des RCPGs. Ces caractéristiques structurales de l'activation des RCPGs permettent d'identifier les états activés et inactivés des récepteurs. Du fait de la prépondérance de structures de classe A, ces caractéristiques ont été largement identifiées pour cette classe mais, avec l'émergence de structures des autres classes de récepteurs, certaines ont pu être généralisées.

Dans les classes A et B, l'activation des RCPGs est marquée par un mouvement concerté des hélices 5, 6 et 7 du domaine transmembranaire. Par rapport à l'état inactivé, l'hélice 6 est déplacée de 6 à 16 Å vers l'extérieur du récepteur, l'hélice 5 subit une rotation induisant son léger déplacement vers l'extérieur du récepteur, et l'hélice 7 est légèrement déplacée vers l'intérieur du récepteur (Figure 1- 4). En conséquence, la cavité intracellulaire du récepteur est élargie, permettant la liaison d'une protéine G ou d'une arrestine<sup>16</sup>. Ce mécanisme impliquant un mouvement de grande amplitude du TM6 a été caractérisé à partir de structures de complexes entre un GPCR et une sous-unité  $G\alpha$ -s. Le mouvement du TM6 est de moindre importance dans le cas d'un complexe impliquant une sous-unité  $G\alpha$ -i/o<sup>17</sup>. Conjointement, le positionnement de l'hélice 5 de la sous-unité  $G\alpha$ -i/o dans la cavité intracellulaire du RCPG est différent de celui observé pour la sous-unité  $G\alpha$ -s. L'interface entre le RCPG et la protéine G est ainsi plus réduite dans le complexe avec la sous-unité  $G\alpha$ -i/o. Ce résultat nouveau complique la caractérisation de l'état actif des RCPGs mais laisse entrevoir une explication à la sélectivité du couplage aux différentes protéines G.



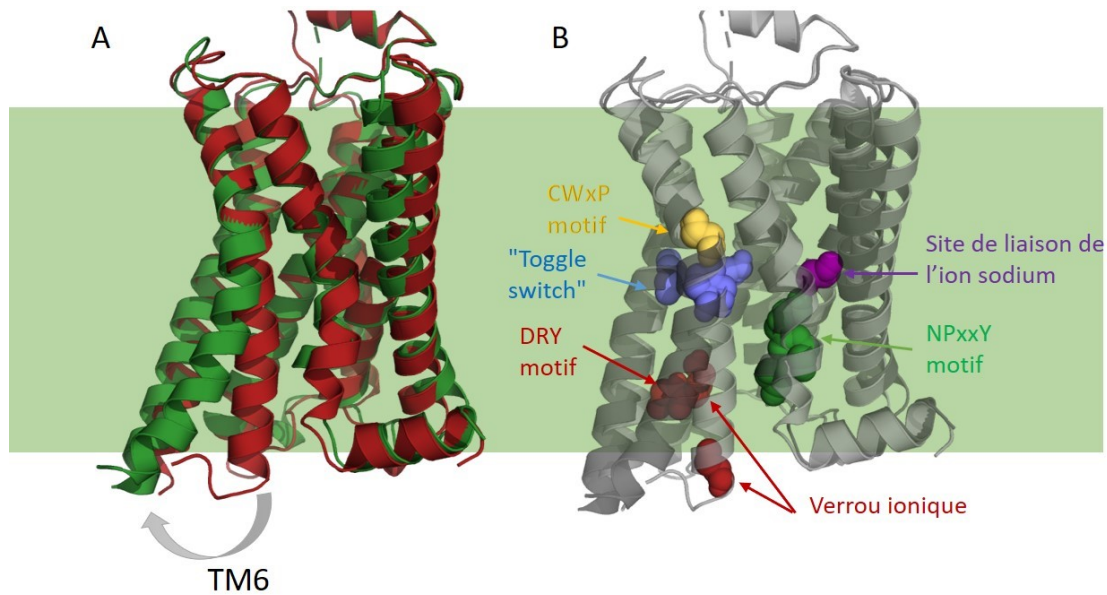


Figure 1- 4 : (A) Mouvement de l'hélice 6 d'un RCPG lors de l'activation. Les états activé et inactif sont représentés en vert (code pdb 3sn6) et rouge (code pdb 2rh1) respectivement. (B) Emplacement des Signatures structurales d'états d'activation des RCPGs.

Dans le cas de RCPGs de classe A liés à une sous-unité  $G\alpha_s$ , les mouvements concertés d'hélices transmembranaires sont associés à plusieurs liaisons inter-hélices appelées verrous moléculaires<sup>16</sup>. Les deux principaux verrous sont le *verrou ionique* présent dans les structures inactives et le *verrou aqueux*, présent dans les structures actives. Le verrou ionique est une combinaison de liaisons hydrogène et d'une liaison ionique. Le pont salin est établi entre le résidu Arg3.50 du motif conservé D(E)RY de l'hélice 3 et l'un des résidus 6.30 ou 6.34 de l'hélice 6<sup>18</sup>. Le verrou aqueux quant à lui, implique le motif conservé NPxxY de l'hélice 7. Une liaison hydrogène est établie entre les résidus 7.53 et 5.58 par l'intermédiaire d'une molécule d'eau. Dans certains cas de figure, le résidu 5.58 est également impliqué dans une liaison hydrogène avec le résidu du motif D(E)RY 3.50<sup>16</sup>.

D'autres caractéristiques structurales, autre que les verrous ioniques et aqueux, sont présentes dans certains récepteurs mais ne peuvent pas être étendues à l'ensemble des RCPGs. Par exemple, dans certaines structures inactives, un ion sodium est présent entre les hélices 2, 3 et 7. Coordiné par le résidu 2.50, cet ion joue le rôle d'un modulateur allostérique<sup>19</sup>. La comparaison des structures actives et inactives de certains RCPGs révèle aussi la perturbation d'un réseau de contacts hydrophobes impliquant les résidus L3.43, F6.40 et F/Y6.44 après la liaison d'un agoniste<sup>18</sup>.

### 2.3 DOMAINE DE LIAISON DES LIGANDS

La liaison du ligand au RCPG constitue le point de départ de la cascade de signalisation. La compréhension de leur mode de reconnaissance est essentielle pour appréhender l'étude des mécanismes d'activation. Les structures expérimentales de RCPGs sont riches d'information sur l'interaction ligand/récepteur. Sur les 270 structures de RCPG de la PDB, seules quatre montrent des récepteurs cristallisés sans aucun ligand (5XSZ, 5GLI, 4LR6, 4GPO). Au total, les structures décrivent 364 complexes protéines-ligand. Les ligands peuvent être classés en trois catégories fonctionnelles :

- les ligands agonistes, qui induisent le couplage à la protéine G.
- les ligands antagonistes, qui empêchent la fixation d'un agoniste sans modifier l'activité basale du récepteur.
- les ligands agonistes inverses, qui induisent le découplage de la protéine G et donc inhibent la signalisation constitutive induite par le récepteur.

Ces ligands se lient majoritairement à la partie extracellulaire ou à la cavité transmembranaire des RCPGs. Cette cavité qui a une grande variabilité de séquence lie une grande variété de ligands différents, allant de la petite molécule hétérocyclique, aux peptides pour les récepteurs à la neurotensine et jusqu'aux petites protéines pour les récepteurs aux chimiokines. Au sein d'une même famille de récepteurs cette variabilité est plus restreinte et fait qu'un même ligand peut avoir un effet sur plusieurs récepteurs.

Deux types de site de liaison coexistent : le site de liaison orthostérique et les sites de liaison allostériques. Le site orthostérique est le site principal de liaison des ligands agonistes. Il est situé au sein du domaine transmembranaire et est généralement entouré par les hélices 3, 6 et 7<sup>20</sup>. Les sites de liaisons allostériques sont des sites de liaisons secondaires et différents du site orthostérique (Figure 1- 5). Ils ont été observés dans la partie extracellulaire (ex : ACM2<sup>21</sup>), mais aussi intracellulaire (ex : CCR2<sup>22</sup>, CCR9<sup>23</sup>), permettent de moduler l'activation des récepteurs et sont spécifiques d'un RCPG <sup>24</sup>.

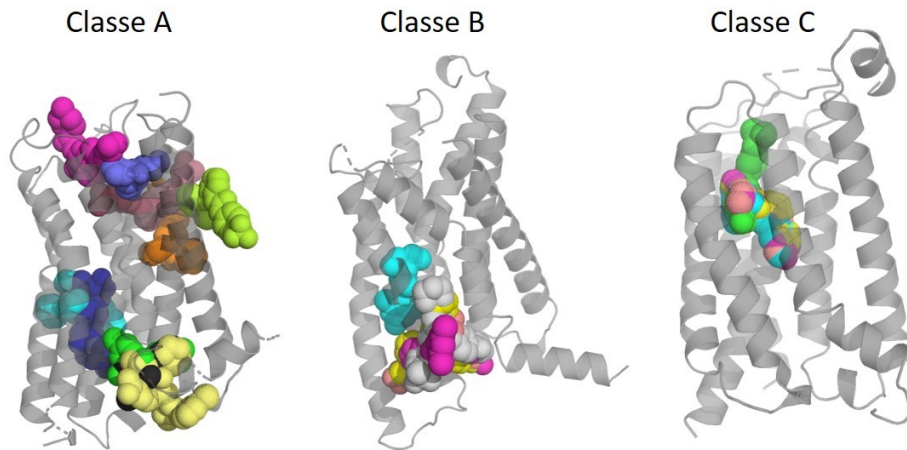


Figure 1- 5 : Emplacements de différents ligands allostériques pour les classes de récepteur A, B et C.

## 2.4 DOMAINE DE LIAISON DES EFFECTEURS

Comparativement au nombre de ligands capables de se lier au domaine extracellulaire, le nombre de protéines G et d'arrestines capables de se lier au domaine intracellulaire est beaucoup plus restreint avec au total une vingtaine de protéines G $\alpha$  et 4 arrestines. La résolution de la structure de différentes protéines G et arrestines associées à un RCPG a permis d'identifier les parties du récepteur qui sont importantes pour la liaison des effecteurs, à savoir ICL2 et ICL3, et les parties intracellulaires des TM3, 4 et 6<sup>25,26</sup>.

Par ailleurs, plusieurs motifs d'acides aminés dans les protéines G sont associés à la reconnaissance de régions distinctes des RCPGs<sup>27</sup>. Les résidus des RCPGs associés à ces interactions peuvent être récupérés pour chaque RCPG par le biais du site GPCRdb (Figure 1-6).

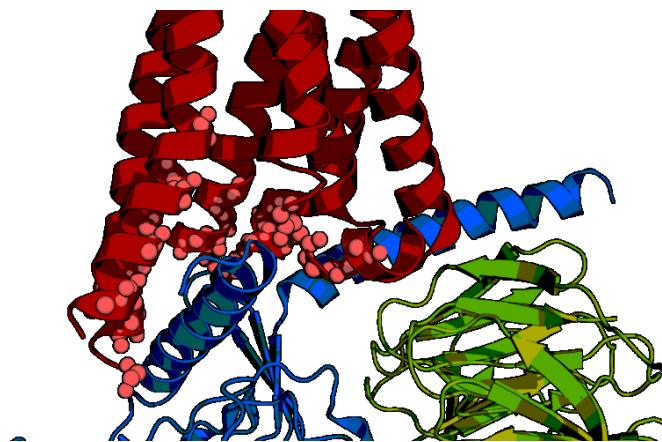


Figure 1- 6 : Structure du récepteur AA2AR (rouge) en complexe avec un hétérotrimère de protéine G (G $\alpha$  en bleu, G $\beta$  en vert) Les résidus prédits en interaction avec la protéine G $\alpha$  à partir du complexe ADRB2-G $\alpha$  sont indiqués par leurs atomes N,CA,C, et O (code pdb 3sn6). L'hélice 5 C-terminale de la sous-unité G $\alpha$ -s est enfouie dans la cavité du récepteur.

## 2.5 INFLUENCE DES MUTATIONS D'ACIDES AMINÉS

L'influence des mutations sur la signalisation par les RCPGs est bien documentée. Les mutations peuvent survenir naturellement ou sont introduites par mutagenèse dirigée pour moduler l'activité du récepteur<sup>28</sup>. Les mutations peuvent perturber la liaison des ligands endogènes et exogènes, altérer la liaison des effecteurs de signalisation<sup>29</sup> ou encore modifier la dynamique du récepteur. Certaines mutations changent le niveau d'activité basal du récepteur, on parle alors de mutation induisant un état actif ou inactif constitutif<sup>30</sup>.

## 3. SIMULATION PAR DYNAMIQUE MOLÉCULAIRE

---

La simulation par dynamique moléculaire joue un rôle de plus en plus important dans l'étude de la fonction des RCPGs. L'augmentation du nombre de récepteurs de structures connues fournit des points de départ pour la simulation de récepteurs différents. Les développements technologiques, notamment le calcul sur processeur graphique et les nouvelles méthodes de calcul dites par dynamique augmentée, ont permis de diminuer le temps nécessaire pour obtenir des simulations d'évènements biologiques longs tels que la liaison de ligand ou le changement d'état du récepteur (inactivation).

Ainsi, la simulation des RCPGs peut être utilisée pour étudier de nombreux aspects tels que la liaison des ligands aux sites orthostériques ou allostériques. Par exemple, l'utilisation de simulation par métadynamique du récepteur aux chimiokines CXCR3 a permis de définir deux cavités distinctes au sein de la poche extracellulaire pour la liaison d'un antagoniste et d'un agoniste inverse<sup>31</sup>. Dans un autre exemple, la simulation de type gros grain a prédit le mode de liaison de ligand peptidique pour les récepteurs neurotensine-1 et CXCR4<sup>32</sup>. De manière similaire, la simulation des RCPGs permet aussi d'étudier les modes et sites de liaison des effecteurs de signalisation. Par exemple la simulation de la rhodopsine bovine liée à l'arrestine-1 a révélé l'importance de l'extrémité C-terminale du récepteur pour la liaison de l'arrestine<sup>33</sup>.

La dynamique moléculaire peut aussi être utilisée sur les RCPGs afin d'observer les changements conformationnels lors de l'activation ou de la désactivation. Pour deux récepteurs, des transitions entre des états actifs et inactifs ont été prédites par le calcul. Plusieurs microsecondes de simulation du récepteur  $\beta$ 2-adrénérique, obtenues par

dynamique moléculaire classique à l'aide du superordinateur ANTON, ont révélé les transitions d'un état activé vers un état intermédiaire puis inactivé<sup>34</sup>. Plusieurs centaines de nanosecondes de simulation du récepteur muscarinique de type 2, obtenue par dynamique moléculaire accélérée, ont montré la transition du récepteur de l'état pseudoactivé vers un état inactivé et inversement<sup>35</sup>.

## 4. RÉCEPTEURS AUX CHIMIOKINES C-C DE TYPE 5

---

Le récepteur aux chimiokines C-C de type 5, fait partie de la famille de la classe A des récepteurs aux chimiokines. La famille de récepteurs aux chimiokines est composée de 23 récepteurs différents, nommée en fonction du type de chimiokines qu'ils lient. Les chimiokines sont séparées en classes, C, CC, CXC, CX3C en fonction du nombre et de l'emplacement de cystéines dans leur séquence. Le groupe des chimiokines C-C que lie le récepteur CCR5 est composé de 28 membres qui sont impliqués dans la migration des monocytes.

Le récepteur CCR5 est principalement situé à la surface des leucocytes, c'est à dire les cellules T, monocytes et macrophages et est par conséquent impliqué dans le système immunitaire notamment auprès des centres d'inflammation. On retrouve également ce récepteur à la surface de cellules du système nerveux central<sup>36</sup>. Des études récentes montrent l'implication de CCR5 dans certains cancers<sup>37,38</sup>, mais ce récepteur est principalement reconnu car il est l'un des corécepteurs du VIH-1 avec le récepteur CXCR4<sup>39</sup>.

CCR5 est constitué d'une séquence de 352 acides aminés. Son extrémité N-terminale a une longueur de 21 résidus tandis que son extrémité C-terminale a une longueur de 38 résidus. La boucle ECL2 est composée de 17 résidus et caractérisée par une sous-structure en feuillet  $\beta$  antiparallèle tandis que ICL3 est d'une taille réduite avec seulement 3 résidus. Le récepteur contient deux ponts disulfures, entre les résidus cystéine 20-269 et 101-178. Dans la cellule, il subit des modifications post-traductionnelles, qui se traduisent par la présence de résidus modifiés : 4 sulfotyrosines en position 3, 10, 14 et 15 (impliquées dans la liaison des chimiokines), O-glycosylation des résidus 6, 7, 16 et 17 (importante pour la liaison des chimiokines)<sup>40</sup>, palmitoylation des résidus cystéine 321, 323 et 324 (importante pour l'expression du récepteur à la membrane cellulaire), et phosphorylation des résidus serine 336, 337, 342 et 349 (favorisée par la liaison des chimiokines C-C)<sup>41</sup> (Figure 1- 7).

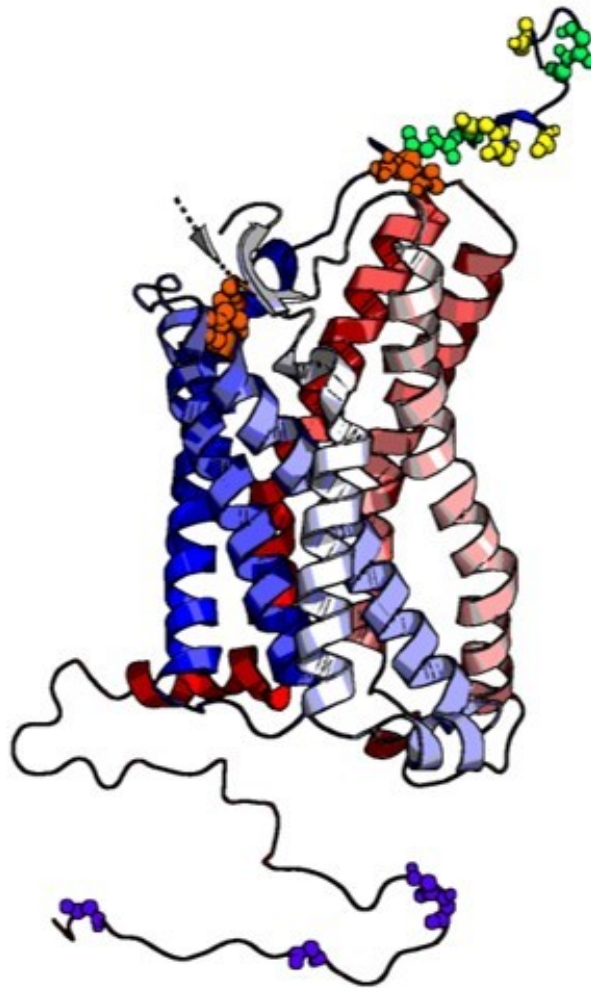


Figure 1- 7 : Structure du récepteur CCR5. Le gradient de couleur du ruban montrant la chaîne principale va du bleu pour le résidu 1 au rouge pour le résidu 352. Les sphères jaunes correspondent aux sulfotyrosines, les sphères vertes aux résidus O-glycosylés, les sphères orange aux ponts disulfures, les sphères cyan aux résidus palmitoylés et les sphères violettes aux résidus phosphorylés. Seuls les atomes de la chaîne principale sont représentés

De par son implication dans l'infection par le VIH, CCR5 a été abondamment étudié. L'impact de modification dans la séquence sur la fonction a été caractérisé pour de nombreux variants du récepteur. Par exemple, le variant  $\Delta 32$  du récepteur, majoritairement présent au sein de la population caucasienne et identifiée en 1996<sup>42</sup>, protège naturellement contre l'infection par le VIH. Ce variant naturel code pour un récepteur tronqué non fonctionnel qui n'est plus adressé à la membrane cellulaire et reste séquestré à l'intérieur de la cellule, empêchant ainsi l'entrée virale. De nombreuses mutations ont été artificiellement introduites pour cartographier les sites de liaison au ligand, par mesure directe ou par compétition de la liaison,





domaine extracellulaire apporte la plus grande variabilité entre les deux structures. On pourra remarquer dans ces structures inactivées l'absence d'ion sodium entre les hélices 2, 3 et 7 et l'absence de l'interaction ionique du motif conservé DRY.

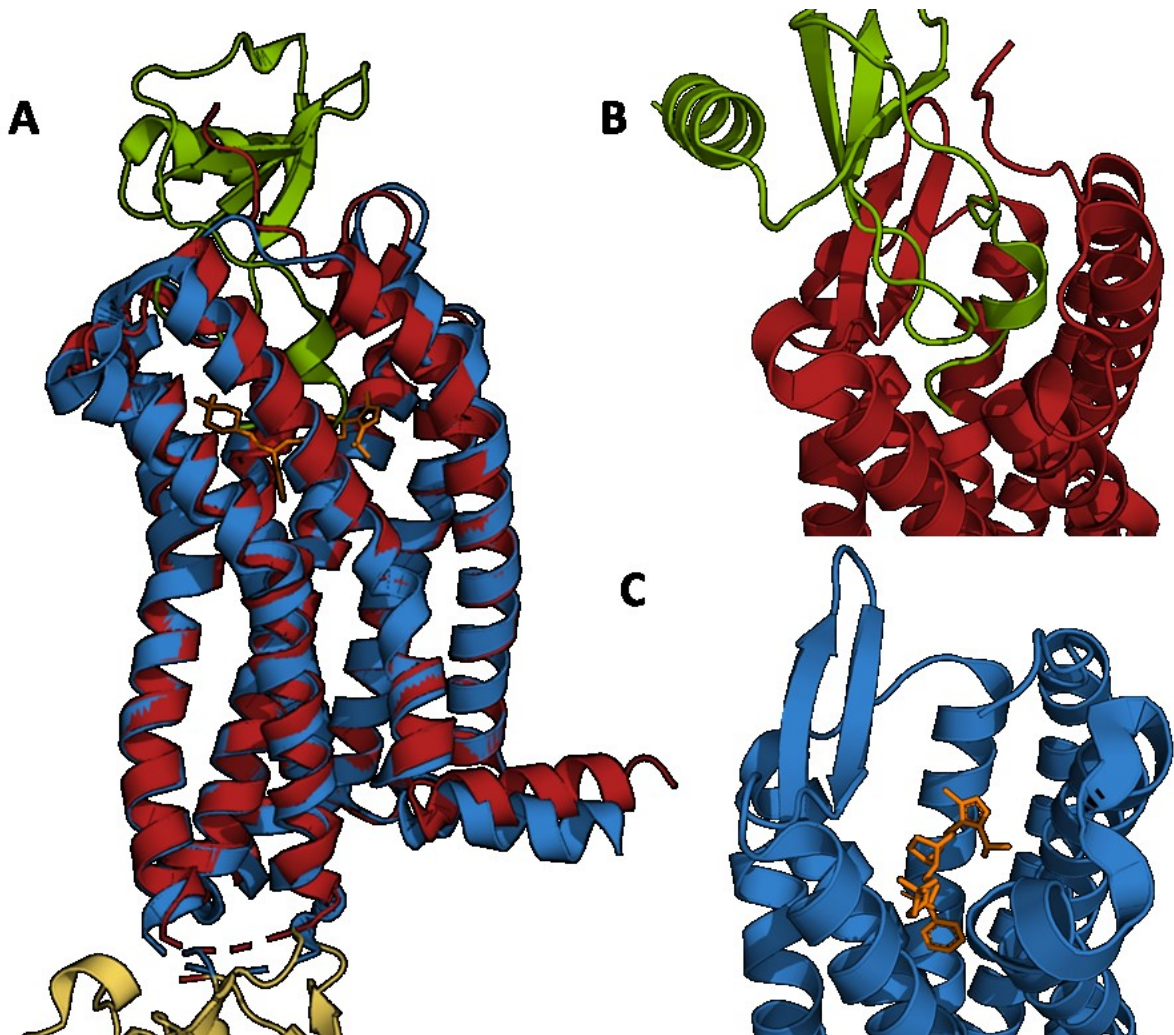


Figure 1-9 : Structures cristallographiques de CCR5. (A) Superposition du récepteur dans les entrées PDB 5UIW (CCR5 en rouge, [5P7]CCL5 en vert, rubredoxine en jaune) et 4MBS (CCR5 en bleu, MVC en orange, rubredoxine en jaune). (B) Zoom sur la cavité transmembranaire du complexe CCR5-[5P7]CCL5. (C) Zoom sur la cavité transmembranaire du complexe CCR5-MVC.

Les deux ligands cristallisés dans CCR5 interagissent avec la cavité transmembranaire. Malgré que les ligands inactivent tous les deux le récepteur, ils ont des modes de liaison différents, avec notamment le ligand MVC se logeant plus profondément dans la cavité. Seuls deux résidus (Tyr108 (3.32) et Glu283 (7.39)) sont communs aux deux modes d'interaction (Figure 1- 10).





récepteur entre la forme libre et liée à l'agoniste inverse<sup>46</sup>. Ces changements se traduisent par une plus grande flexibilité de ECL2 dans la forme libre et un déplacement du TM5 et de ICL3.

Une autre étude utilisant la dynamique moléculaire classique et des calculs d'énergie libre a permis de définir un modèle de liaison entre CCR5 et la boucle V3 de la protéine virale gp120 en accord avec des données biologiques<sup>47</sup>(Figure 1- 11). La modélisation du complexe entre CCR5 et V3 a notamment fourni une base moléculaire pour expliquer la reconnaissance du complexe CCR5/MVC par la gp120 d'un virus primaire résistant à l'antiviral<sup>48</sup>.

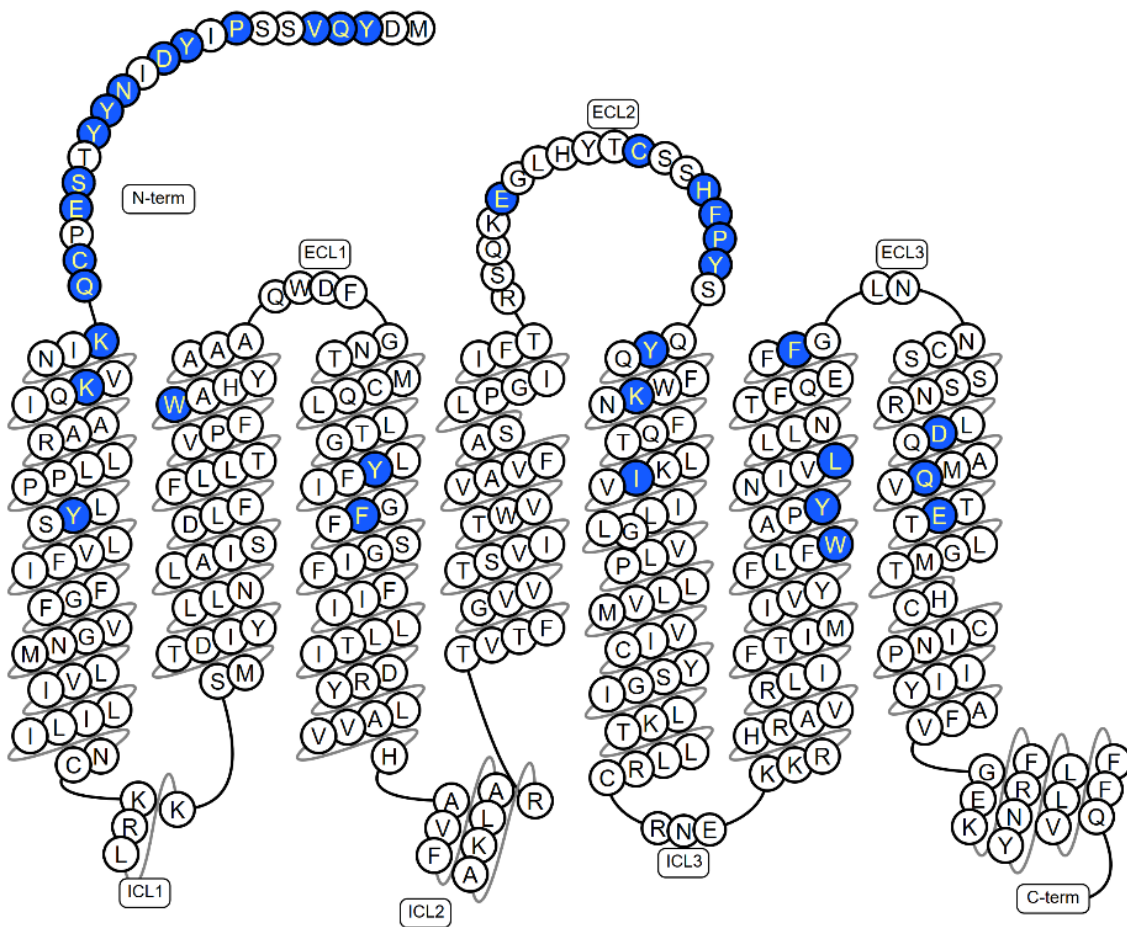


Figure 1- 11 : Résidus du CCR5 impliqués dans la liaison avec la boucle V3 de la protéine virale gp120.

## 5. CONCLUSION

---

Parmi les RCPGs, la famille des récepteurs aux chimiokines est une famille de grand intérêt thérapeutique du fait de leur implication dans le contrôle de l'activation, la migration, la différenciation et la survie des leucocytes. Ils constituent des cibles potentielles pour le traitement de maladies inflammatoires ou le cancer. Dans cette famille, CCR5 occupe une place particulière du fait de son implication dans l'infection par le VIH.

Au total, quatre membres de la famille des récepteurs aux chimiokines ont une structure résolue par cristallographie : CCR5, le récepteur aux chimiokines CC de type 2 (CCR2), le récepteur aux chimiokines CC de type 9 (CCR9) et le récepteur aux chimiokines CXC de type 4 (CXCR4). CCR2 est un homologue proche de CCR5 avec une conservation de séquence de 70%. CXCR4 est le corécepteur minoritaire du VIH. Ces structures révèlent les particularités de la famille : la structure de l'hélice 2 au niveau de la signature TXP, la longueur du TM5 et la structure en épingle à cheveux de ECL2 ou encore le pont disulfure en la partie N-terminale et ECL3. Par ailleurs, ces récepteurs ont été cristallisés avec des ligands variés : molécules de petite taille, endogènes (ex : acide oléique dans 3ODU) ou synthétiques (ex : Peptide it1 dans 3ODU) et occupant des sites distincts, mais aussi des chimiokines naturelle ou synthétiques, présentant des modes de liaison différents au niveau du site orthostérique.

Malgré l'abondance de l'information structurale, la relation entre structure et fonction n'est pas encore clairement comprise. En effet, le modèle binaire activé/inactivé est trop simpliste pour expliquer la fonction d'un récepteur et en particulier sa capacité à engager différentes voies de signalisation. Les questions actuelles portent sur la diversité conformationnelle de RCPGs et les mécanismes qui gouvernent l'équilibre entre les différents états. Cette question complexe nécessite d'approfondir nos connaissances sur l'oligomérisation des RCPGs, mais aussi sur l'impact de l'environnement du récepteur (ex : composition de la membrane) et de modifications post-traductionnelles. La modélisation et la simulation par dynamique moléculaires sont des méthodes de choix pour explorer les aspects structuraux de ces questions.

## 6. BIBLIOGRAPHIE DE CHAPITRE

---

1. Duc, N. M., Kim, H. R. & Chung, K. Y. Structural mechanism of G protein activation by G protein-coupled receptor. *Eur. J. Pharmacol.* **763**, 214–222 (2015).
2. Park, J. Y., Lee, S. Y., Kim, H. R., Seo, M.-D. & Chung, K. Y. Structural mechanism of GPCR-arrestin interaction: recent breakthroughs. *Arch. Pharm. Res.* **39**, 293–301 (2016).
3. Hilger, D., Masureel, M. & Kobilka, B. K. Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).
4. Attwood, T. K. & Findlay, J. B. C. Fingerprinting G-protein-coupled receptors. *Protein Eng. Des. Sel.* **7**, 195–203 (1994).
5. Fredriksson, R. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
6. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
7. Palczewski, K. *et al.* Crystal Structure of Rhodopsin: A G Protein–Coupled Receptor. **289**, 8 (2000).
8. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein-Coupled Receptor. *Science* **318**, 1258–1265 (2007).
9. Jaakola, V.-P. *et al.* The 2.6 Å Crystal Structure of a Human A2A Adenosine Receptor bound to ZM241385. **322**, 1211–1217 (2008).
10. Wang, C. *et al.* Structure of the human smoothed receptor bound to an antitumour agent. *Nature* **497**, 338–343 (2013).
11. Carpenter, B. & Tate, C. G. Engineering a minimal G protein to facilitate crystallisation of G protein-coupled receptors in their active conformation. *Protein Eng. Des. Sel.* **29**, 583–594 (2016).
12. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the  $\beta$ 2 adrenoceptor. *Nature* **469**, 175–180 (2011).
13. Rasmussen, S. G. F. *et al.* Crystal structure of the  $\beta$ 2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).

14. Pándy-Szekeres, G. *et al.* GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).
15. Ballesteros, J. A. & Weinstein, H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. in *Methods in Neurosciences* (ed. Sealfon, S. C.) **25**, 366–428 (Academic Press, 1995).
16. Manglik, A. & Kruse, A. C. Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry (Mosc.)* **56**, 5628–5634 (2017).
17. Capper, M. J. & Wacker, D. How the ubiquitous GPCR receptor family selectively activates signalling pathways. *Nature* **558**, 529–530 (2018).
18. Tehan, B. G., Bortolato, A., Blaney, F. E., Weir, M. P. & Mason, J. S. Unifying Family A GPCR Theories of Activation. *Pharmacol. Ther.* **143**, 51–60 (2014).
19. White, K. L. *et al.* Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling. *Structure* **26**, 259–269.e5 (2018).
20. Lee, S.-M., Booe, J. M. & Pioszak, A. A. Structural insights into ligand recognition and selectivity for classes A, B, and C GPCRs. *Eur. J. Pharmacol.* **763**, 196–205 (2015).
21. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
22. Zheng, Y. *et al.* Structure of CC Chemokine Receptor 2 with Orthosteric and Allosteric Antagonists. *Nature* **540**, 458–461 (2016).
23. Oswald, C. *et al.* Crystal structure of the human CC chemokine receptor type 9 (CCR9) in complex with vercirnon. *Nature* **540**, 462 (2016).
24. Thal, D. M., Glukhova, A., Sexton, P. M. & Christopoulos, A. Structural insights into G-protein-coupled receptor allostery. *Nature* **559**, 45–53 (2018).
25. Venkatakrishnan, A. J. *et al.* Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).
26. Carpenter, B. & Tate, C. G. Active state structures of G protein-coupled receptors highlight the similarities and differences in the G protein and arrestin coupling interfaces. *Curr. Opin. Struct. Biol.* **45**, 124–132 (2017).
27. Flock, T. *et al.* Selectivity determinants of GPCR–G-protein binding. *Nature* **545**, 317–322 (2017).

28. Fukami, M., Suzuki, E., Igarashi, M., Miyado, M. & Ogata, T. Gain-of-function mutations in G-protein-coupled receptor genes associated with human endocrine disorders. *Clin. Endocrinol. (Oxf.)* **88**, 351–359 (2018).
29. Stoy, H. & Gurevich, V. V. How genetic errors in GPCRs affect their function: Possible therapeutic strategies. *Genes Dis.* **2**, 108–132 (2015).
30. Tsukamoto, H. & Farrens, D. L. A Constitutively Activating Mutation Alters the Dynamics and Energetics of a Key Conformational Change in a Ligand-free G Protein-coupled Receptor. *J. Biol. Chem.* **288**, 28207–28216 (2013).
31. Milanos, L. *et al.* Identification of Two Distinct Sites for Antagonist and Biased Agonist Binding to the Human Chemokine Receptor CXCR3. *Angew. Chem. Int. Ed.* **55**, 15277–15281 (2016).
32. Delort, B. *et al.* Coarse-Grained Prediction of Peptide Binding to G-Protein Coupled Receptors. *J. Chem. Inf. Model.* **57**, 562–571 (2017).
33. Latorraca, N. R. *et al.* Molecular mechanism of GPCR-mediated arrestin activation. *Nature* **557**, 452–456 (2018).
34. Dror, R. O. *et al.* Activation mechanism of the  $\beta$ 2-adrenergic receptor. *Proc. Natl. Acad. Sci.* **108**, 18684–18689 (2011).
35. Miao, Y. & McCammon, J. A. Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci.* **113**, 12162–12167 (2016).
36. Sorce, S., Myburgh, R. & Krause, K.-H. The chemokine receptor CCR5 in the central nervous system. *Prog. Neurobiol.* **93**, 297–311 (2011).
37. Singh, S. K. *et al.* CCR5/CCL5 axis interaction promotes migratory and invasiveness of pancreatic cancer cells. *Sci. Rep.* **8**, (2018).
38. Jiao, X. *et al.* CCR5 Governs DNA Damage Repair and Breast Cancer Stem Cell Expansion. *Cancer Res.* **78**, 1657–1671 (2018).
39. Deng, H. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661–666 (1996).
40. Blanpain, C. *et al.* Palmitoylation of CCR5 Is Critical for Receptor Trafficking and Efficient Activation of Intracellular Signaling Pathways. *J. Biol. Chem.* **276**, 23795–23804 (2001).
41. Oppermann, M., Mack, M., Proudfoot, A. E. I. & Olbrich, H. Differential Effects of CC Chemokines on CC Chemokine Receptor 5 (CCR5) Phosphorylation and Identification of

- Phosphorylation Sites on the CCR5 Carboxyl Terminus. *J. Biol. Chem.* **274**, 8875–8885 (1999).
42. Samson, M. *et al.* Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
  43. Garcia-Perez, J. *et al.* Allosteric Model of Maraviroc Binding to CC Chemokine Receptor 5 (CCR5). *J. Biol. Chem.* **286**, 33409–33421 (2011).
  44. Tan, Q. *et al.* Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science* **341**, 1387–1390 (2013).
  45. Zheng, Y. *et al.* Structure of CC Chemokine Receptor 5 with a Potent Chemokine Antagonist Reveals Mechanisms of Chemokine Recognition and Molecular Mimicry by HIV. *Immunity* **46**, 1005-1017.e5 (2017).
  46. Salmas, R. E., Yurtsever, M. & Durdagi, S. Investigation of Inhibition Mechanism of Chemokine Receptor CCR5 by Micro-second Molecular Dynamics Simulations. *Sci. Rep.* **5**, (2015).
  47. Tamamis, P. & Floudas, C. A. Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS ONE* **9**, e95767 (2014).
  48. Garcia-Perez, J. *et al.* A single-residue change in the HIV-1 V3 loop associated with maraviroc resistance impairs CCR5 binding affinity while increasing replicative capacity. *Retrovirology* **12**, (2015).

## Chapitre 2

Distinction des récepteurs couplés aux protéines G et de leurs états d'activation par l'analyse de leurs interactions intramoléculaires transmembranaires.



## 1. INTRODUCTION

---

Les RCPGs constituent une classe pharmacologique de protéines membranaires impliqués dans la transduction du signal de part et d'autre de la membrane. Ils répondent à un stimulus par le couplage ou découplage à des effecteurs cellulaires, notamment les protéines G et la  $\beta$ -arrestine. Tous les RCPGs ont une architecture moléculaire commune, constituée de sept hélices transmembranaires (TM), notées TM1 à TM7. Le stimulus, s'il est de nature moléculaire, se lie soit à la partie extracellulaire soit à la partie transmembranaire. L'état activé du récepteur correspond à la forme liée à l'effecteur, alors que l'état inactivé est au contraire celui inapte à lier l'effecteur.

Les 398 RCPGs non olfactifs assurent le contrôle de fonctions biologiques variées comme la transmission de signaux lors du cycle visuel (rhodopsine (OPSD)), dans le contrôle du rythme cardiaque (récepteur muscarinique de type 2 à l'acétylcholine (ACM2)), dans la relaxation des tissus musculaires (récepteur adrénergique beta de type 2 (ADRB2)), dans le contrôle de la douleur (récepteur adénosine A2 (AA2AR)) ou de l'entrée de virus dans une cellule (récepteur aux chimiokines au motif C-C de type 5 (CCR5)), etc.

Ces récepteurs diffèrent par leur séquence en acides aminés. L'analyse phylogénétique des RCPGs<sup>1</sup> propose la définition de cinq classes, en adéquation avec leur pharmacologie. Ces cinq classes sont nommées par une lettre ou le nom d'un récepteur emblématique :

- Classe A ou classe des récepteurs rhodopsine
- Classe B1 ou classe des récepteurs sécrétine
- Classe B2 ou classe des récepteurs à adhésion
- Classe C ou classe des récepteurs métabotropique glutamate
- Classe F ou classe des récepteurs Frizzled/Smoothened

Plusieurs approches ont été utilisées pour déterminer des classes à partir des séquences de RCPG et toutes s'accordent sur la définition des cinq classes. Les travaux de J.S. SURGAND en thèse au laboratoire, ont montré qu'un sous-ensemble de 30 résidus, choisis comme étant ceux délimitant les cavités transmembranaires des RCPGs de classes A, est suffisant pour obtenir une partition pharmacologique correcte des 398 RCPGs non olfactif humains. Ainsi la particularité d'un RCPG ou d'une classe de RCPG ne réside pas uniquement dans les boucles

et parties extramembranaires mais aussi dans le domaine à sept hélices transmembranaires qui est commun à tous les RCPGs. L'existence de motifs de séquences associés à une famille ou un domaine fonctionnel a été largement démontrée par l'analyse bio-informatique des séquences de protéines. Dans le cas des RCPGs, des signatures ont également été observées dans les sept domaines transmembranaires des différentes classes<sup>2</sup>. Par exemple, pour les récepteurs de classe A, le motif NPxxY ou 'water lock' situé dans le domaine transmembranaire 7, est associé à l'ouverture d'une barrière hydrophobe. Le motif DRY dans le domaine transmembranaire 3 de la plupart des RCPGs de classes A a été proposé comme responsable du verrou ionique ('ionic lock') maintenant le récepteur à l'état inactivé.

Les structures 3D des RCPGs ont fourni une explication moléculaire à certaines signatures de séquence<sup>3,4</sup>. Ainsi, l'état inactivé des RCPGs est caractérisé par : la présence du 'ionic lock', associé au motif DRY et reliant le TM 3 au TM 6 par un pont salin entre les résidus Arg(3.50) et Glu(6.30), induisant ainsi une proximité accrue de ces TM ; et par la présence d'un ion sodium coordonné par les TM 2, 3 et 7. L'état activé des RCPGs est caractérisé par une rotation des TM 3, 5, 6 et 7 et une translation du TM6 pour former une cavité dans le domaine intracellulaire. Ces rotations/translations perturbent l'existence du verrou ionique et permettent également la formation du 'water lock' associé au motif NPxxY qui relie les résidus Tyr(7.53) et Tyr(5.58) par une molécule d'eau. Ces mouvements, motifs de résidus conservés et interactions sont actuellement admis comme étant caractéristiques des états d'activation pour la grande majorité des RCPGs. Malgré cela, il reste difficile de proposer une vision unique de l'activation des RCPGs, principalement du fait de la variabilité de séquence de ces récepteurs. Ainsi, même dans la classe A, il existe des variations à ces caractéristiques d'activation. Par exemple dans le cas des récepteurs  $\mu$ -opiacé et rhodopsine le « *water lock* » est étendu avec une interaction entre les résidus Tyr(5.58) et Arg(3.50). Pour le récepteur CCR5 l'ion sodium est absent et le résidu en position 6.30 est une arginine et non un acide glutamique. De ce fait, le verrou ionique est absent dans ce récepteur.

S'il existe clairement un lien entre séquence, structure et fonction, la description de ce lien pour les RCPGs n'est pas encore complète<sup>5</sup>. Par ailleurs, il n'a été essentiellement établi que dans un sens : de la séquence vers la structure. Avec l'accès à un nombre toujours croissant de RCPGs, il est désormais possible de faire la démarche inverse, à savoir l'étude comparative

des structures 3D pour expliquer les conservations dans la séquence<sup>6-8</sup>. Nous proposons ici d'analyser les structures à partir de leurs interactions intramoléculaires pour déterminer :

- la signature moléculaire d'un RCPG ou d'une classe de RCPGs
- la signature moléculaire des états d'activation / d'inactivation d'un RCPG ou d'une classe de RCPGs.

Ces signatures devraient nous révéler les acides aminés déterminant pour la stabilisation de la structure à sept domaines transmembranaires et pour la stabilisation des états conformationnels du domaine transmembranaire.

Nous avons tout d'abord analysé l'ensemble des structures cristallographique de RCPGs dans la PDB. Nous avons ensuite appliqué notre approche à des trajectoires de dynamique moléculaire simulant la transition entre deux états conformationnels distincts d'un RCPG.

## 2. MATÉRIEL ET MÉTHODE

---

Dans cette section, je décrirai la méthode développée pour l'analyse des interactions intramoléculaires du domaine transmembranaire et la classification des RCPGs. Je décrirai également les jeux de données et leur préparation.

### 2.1 MATÉRIEL

#### 2.1.1 Modèles de référence du domaine transmembranaire des récepteurs couplés aux protéines G

Des modèles 3D des RCPGs humains non olfactifs sont disponibles au laboratoire<sup>9,10</sup>. Ces modèles délimitent le domaine transmembranaire en suivant le système de numérotation des résidus proposé par Ballesteros-Weinstein<sup>11</sup> d'après les bornes des hélices transmembranaires dans la structure PDB 1F88 du récepteur de la rhodopsine (OPSD) bovine.

#### 2.1.2 Structures de récepteurs couplés aux protéines G dans la PDB

La construction de la base de structure 3D a été effectuée en plusieurs étapes : (1) le recensement des différentes structures dans la PDB, (2) la caractérisation de leur état d'activation, (3) la classification des différentes structures selon leurs caractéristiques structurales.

##### 2.1.2.1 Sélection et téléchargement

Les structures de RCPGs obtenues par diffraction des rayons X (RX), cryo-microscopie électronique (cryo-EM) ou par résonance magnétique nucléaire (RMN) ont été téléchargées depuis le site de la Protein DataBank (PDB) (accès le 27/06/17 et 14/08/17). ([www.rcsb.org](http://www.rcsb.org)). L'identification des RCPGs dans la PDB a été effectuée en intersectant les informations de plusieurs bases de données libres de droit.

La base de données UniprotKB<sup>12</sup> ([www.uniprot.com/database](http://www.uniprot.com/database)), a été utilisée pour associer les différents identifiants Uniprot des RCPGs à leurs structures PDB. L'accès à cette base a été faite le 4/8/17, en croisant les références de la UniprotKB et de la GPCRdb<sup>13</sup> pour un total de 140 501 entrées. Les champs recherchés sont : Entry (Uniprot ID), Entry name (Nom\_Espèce de gène), Organism, Entrée PDB (référence croisée PDB). Seules les entrées Uniprot possédant des structures PDB et pour lesquels nous avons un modèle de référence de son domaine

transmembranaire ont été sélectionnées pour un total de 110 références Uniprot correspondant à 418 entrées PDB.

La base de données SIFTS ([www.ebi.ac.uk/pdbe/docs/sifts/quick.html](http://www.ebi.ac.uk/pdbe/docs/sifts/quick.html), accès le 31/7/17, 372 900 entrées) a été utilisée pour filtrer les 418 entrées PDB sélectionnées de la UniprotKB. Les champs recherchés sont : entrées PDB, nom de chaîne, identifiant Uniprot associé et numéros de résidus de début et de fin de chaîne. Ces informations sont utilisées pour écarter les entrées PDB n'ayant pas entre 200 et 600 résidus, réduisant le total de structures d'entrées PDB à 258 (Annexe 2).

Pour finir, toutes les entrées restantes ont été comparées avec celles de la GPCRdb<sup>13</sup> (accès le 22/09/17). Les entrées PDB différent ont été évaluées manuellement et ajoutées/supprimées pour obtenir in fine 215 structures de RCPG.

Pour chacune de ces structures les informations suivantes ont été référencées : l'identifiant Uniprot, le gène associé, l'identifiant PDB, le nom de chaîne étudiée dans le fichier PDB, le nombre de résidus de la chaîne étudiée, le nom de toutes les chaînes composant le fichier PDB, la méthode de détermination de la structure, la résolution, l'espèce associée (Annexe 3).

#### 2.1.2.2 Annotation 1 : états d'activation

En plus des diverses informations extraites des bases de données, chaque structure 3D répertoriée a été associée à un état d'activation. Cette association a été effectuée en analysant plusieurs éléments : la structures 3D, la publication associée et les informations de la GPCRdb. En détail, nous avons distingué les états activé, préactivé et inactivé selon les points suivants de priorité décroissante :

- 1) L'état est clairement caractérisé dans la publication associée à la structure,
- 2) L'état est proche d'une autre structure dont l'état est caractérisé dans la publication associée,
- 3) Une arrestine, une protéine G, un fragment ou un anticorps est présent dans la structure 3D, l'état actif sera donc associé sauf désaccord dans la publication associée,
- 4) Une mutation constitutive activant ou inactivant le récepteur est décrite dans la publication associée,

- 5) Un ligand co-cristallisé est présent dans la structure. L'état activé est associé aux ligands caractérisés comme agonistes tandis que l'état inactivé est associé aux ligands antagonistes,
- 6) L'état préactivé est associé aux ligands agonistes partiels,
- 7) Aucune information n'étant disponible l'état est considéré comme inactivé.
  - Pour le récepteur OPSD, les états métarhodopsines 1/2/3 et opsine sont considérés comme activés et les états rhodopsine, photorhodopsine, bathorhodopsine, BL(BSI)rhodopsine et lumirhodopsine sont considérés comme inactivés.

### 2.1.2.3 Annotation 2 : Classes de récepteurs couplés aux protéines G

Nous avons considéré cinq niveaux de classification :

Niveau 1 - Les familles de RCPGs : A, répartie en 7 sous classes (sensorielle, nucléotide, lipide, chimiokine, protéine, aminergique et autre), B, C1 (Métabotropique glutamate/phéromone) et F

Niveau 2 - 28 sous familles de RCPGs (Opsines, cannabinoïdes, chimiokine, adrénorecepteurs, etc.)

Niveau 3 - 46 RCPGs différents (OPSD, ADRB2, CCR5, AA2AR, etc.). Cette classification est basée sur le nom de gène, qui est commun aux orthologues.

Niveau 4 - L'identifiant Uniprot, qui caractérise le gène et l'espèce associée à la structure

Niveau 5 - L'état d'activation d'un RCPG, les états activé et préactivé étant regroupés dans une même catégorie.

### 2.1.2.4 Préparation des structures PDB

Toutes les structures téléchargées de la PDB ont été préparées selon la même méthode. La première étape de la préparation consiste à préparer les structures au moyen du logiciel Protoss<sup>14,15</sup> (version 2.0.3) en utilisant les paramètres standards. Dans une seconde étape sont retirés tous les éléments non constitutifs ou non analysés du récepteur : les molécules d'eau, de cholestérol, de lipides mais aussi les ligands, les ions, les protéines (protéines G, anticorps), les protéines de fusion. Dans le cas d'une structure oligomérique, seule la première chaîne est gardée.

### 2.1.3 Trajectoire de dynamique moléculaire de récepteurs couplés aux protéines G

Nous avons choisi dans la littérature deux exemples de trajectoire de dynamique moléculaire simulant un changement d'état d'activation. Ils s'agit de la transition entre les états activé/pseudoactivé et inactivé de ACM2<sup>16,17</sup> et de ADRB2<sup>18,19</sup>. Ces trajectoires nous ont été gracieusement transmises par les Docteurs J. Andrew McCAMMON et Yongling MIAO, de l'Université de Californie de San Diego, pour la dynamique du récepteur ACM2, et par les Docteurs David SHAW et Ron DROR, de l'entreprise D. E. Shaw & Co., pour la dynamique du récepteur ADRB2.

#### 2.1.3.1 Récepteur muscarinique de type 2 à l'acétylcholine

L'analyse de la simulation de ACM2 a été publiée en 2016<sup>20</sup>. La simulation d'une durée de 409,8 ns a été générée en utilisant l'algorithme de dynamique moléculaire accélérée par gaussienne (GaMD)<sup>21</sup>. Dans l'article, elle est référencée comme la simulation numéro 10. Son point de départ est la structure RX du récepteur M2 liée à l'agoniste plein IXO et à l'anticorps Nb9-8 (identifiant PDB 4MQS<sup>17</sup>). L'anticorps Nb9-8 a été retiré de la structure avant simulation. De plus, la protéine de fusion a été supprimée et la boucle ICL3 non modélisée. La structure initiale comporte deux ponts disulfures, i.e., C96(3.25)-C176(ECL2) et C413(6.61)-C416(7.29), et a été protonée en utilisant CHARMM à pH neutre. Les acides aminés acides sont ainsi tous chargés négativement, à l'exception du résidu D69(2.50) qui a été modélisé sous forme neutre. La protéine a été insérée dans une bicouche lipidique de palmitoyl-oleoyl-phosphatidyl-choline (POPC), dans une boîte d'eau contenant du NaCl à une concentration de 0.15 M. Le système final consiste en une boîte de  $88 \times 88 \times 86 \text{ \AA}^3$  contenant 136 lipides, 28 Na<sup>+</sup>, 39 Cl<sup>-</sup>, et 10010 molécules d'eau, pour un total de  $\sim 55300$  atomes. Les champs de forces CHARMM36 et GAAMP ont été utilisés pour la protéine et les lipides d'une part, et pour la molécule IXO d'autre part. Le système a été minimisé et équilibré en utilisant le logiciel NAMD2.10 avant d'avoir été simulé en GaMD en utilisant la version GPU de AMBER14.

#### 2.1.3.2 Récepteur $\beta$ adrénergique de type 2

L'analyse de la simulation de ADRB2 a été publiée en 2011<sup>22</sup>. Cette dynamique, d'une durée de  $\sim 11 \mu\text{s}$  a été générée grâce au superordinateur ANTON<sup>23</sup>. Dans l'article, elle est référencée comme la dynamique numéro 11, son point de départ étant la structure RX du récepteur ADRB2 lié au ligand BI-167107 et à l'anticorps Nb80 (identifiant PDB 3POG<sup>19</sup>). L'anticorps Nb80 a été retiré de la structure avant simulation. De plus, la protéine de fusion a été retirée et la

boucle ICL3 non modélisée. La structure initiale a été protonée à un pH égal à 7.0, à l'aide du logiciel Maestro. Les acides aminés sont ainsi tous chargés négativement, à l'exception de résidus D130(3.49) et E122(3.41) qui ont été modélisés sous forme neutre. La protéine a été insérée dans une bicouche lipidique de palmitoyl-oleoyl-phosphatidyl-choline (POPC) dans une boîte d'eau contenant des ions NaCl à une concentration de 0.15 M. Le système final consiste en une boîte de  $80 \times 80 \times 85 \text{ \AA}^3$  avec 140 lipides, 10000 molécules d'eau, 30  $\text{Cl}^-$ , et 30  $\text{Na}^+$ . Les champs de forces CHARMM27, CMAP, TIP3P et CHARMM modifié (Chemistry at Harvard Molecular Mechanics) ont été utilisés pour la protéine, l'eau et les lipides. Le champ de forces utilisé pour BI-167107 a été décrit dans d'autres publications<sup>24,25</sup>.

La dynamique du récepteur ADRB2 comportant au total 55556 structures, seules les sections expérimentant une transition d'états ont été étudiées, c'est-à-dire la section allant de 0 à 720 ns et la section allant de 5580 à 6300 ns. La première transition va d'un état activé vers un état intermédiaire, caractérisé par l'écart quadratique moyen (RMSD) des coordonnées atomiques de la chaîne principale du motif NPxxY (322-327) par rapport à la structure inactivée du récepteur (identifiant PDB 2RH1<sup>18</sup>). La seconde transition va d'un état intermédiaire vers un état inactivé, caractérisé par la diminution de la distance entre les carbones alpha des résidus R131(3.50) et L272(6.34).

## 2.2 MÉTHODES D'ANALYSE ET DE CLASSIFICATION

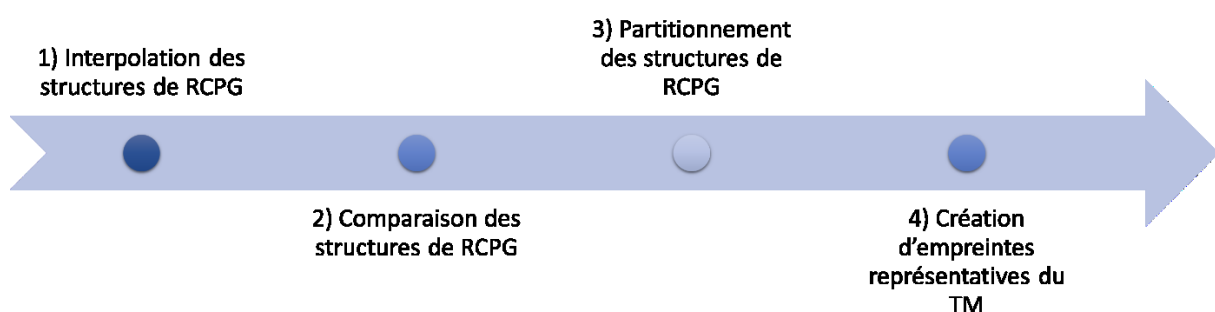


Figure 2- 1 : Protocole général d'analyse des structures en 3D de RCPG.

La méthode que nous avons développée est constituée de quatre étapes (Figure 2- 1). La première étape prend comme entrée les structures 3D protonées préparées au format PDB ou Mol2. Lors de cette étape, ces structures sont traitées afin de les valider, de détecter le



domaine transmembranaire, de calculer les interactions intramoléculaires et de convertir ces interactions sous forme de triplets de pseudo atomes ou de matrices d'interactions.

La seconde étape du protocole prend en entrée les représentations simplifiées obtenues lors de l'étape 1. Elle consiste à effectuer une comparaison par paires des structures de RCPG pour obtenir une matrice carrée regroupant les scores de similitude. Cette comparaison nécessite un alignement des représentations simplifiées dans le cas de triplets de pseudo-atomes. La comparaison des matrices d'interactions est indépendante de la position du récepteur dans le repère cartésien.

La troisième étape du protocole utilise la matrice de score de similitude calculée lors de l'étape 2 pour effectuer un partitionnement des structures de RCPGs.

Pour finir, l'étape 4, prend en compte les partitions de structures de RCPGs ainsi que les matrices d'interactions pour créer des empreintes d'interactions caractéristiques consensuelles.

## 2.2.1 Interprétation des structures et création de représentations simplifiées

### 2.2.1.1 Évaluation des structures et sélection du domaine transmembranaire

Cette partie est effectuée par les modules d'initiation des scripts python tMimiR, dans le cas de structures 3D (RX/RMN), et iMimiR, dans le cas de trajectoires de dynamique moléculaire. Elle vise à extraire une matrice de correspondance entre les sept hélices transmembranaires et leurs résidus de début et de fin.

Pour une structure expérimentale (RX/RMN), le module d'initiation nécessite plusieurs éléments : une structure 3D de RCPG au format Mol2, un fichier contenant les informations sur la structure (Tableau 2 - 1), et une base de données de modèles de référence de RCPG<sup>9</sup>. Chaque modèle de référence décrit au format Mol2 la zone transmembranaire annotée au format Ballesteros<sup>11</sup>.

Tableau 2 - 1 : Exemple de fichier contenant les informations pour trois structures de récepteurs différents.

Identifiant Uniprot	GPCR de référence	Identifiant PDB	Nom de chaîne	Nombre de résidus	Chaînes du fichier PDB	Méthode expérimentale	Résolution (Å) de la structure	Espèce	Etat d'activation	Ligand	Classe: Niveau 1	Classe: Niveau 2	Classe: Niveau 3	Classe: Niveau 4	Classe: Niveau 5
P35462	DRD3	3pbl	A	302	A/B	X-RAY	2.89	HUMAN	I	ANT	10	28	22	30	38
G1SGD4	G1SGD4	5vai	R	398	R	ELECTRON	4.1	RABBIT	A	Null	4	7	25	1	1
P25103	NK1R	2ks9	A	363	A/B	SOLUTION	NULL	HUMAN	pA	Null	9	18	32	18	25

Le script exécute ensuite plusieurs tâches :

- Pour chaque structure du fichier d'information, il convertit la séquence de fichier PDB et celle du modèle de RCPG de référence au format de Pearson (FASTA) <sup>26</sup>.
- Il vérifie, par recherche d'expression régulière entre les séquences FASTA, que les protéines de fusion (rubredoxin, endolysin, lysozyme, glutathione et maltose) sont bien absentes de la structure.
- Il effectue une comparaison des deux séquences par analyse d'expression régulière avec erreur pour identifier les positions de départ et de fin des hélices transmembranaires. Pour que deux séquences soient considérées comme correctement alignées, un maximum de cinq erreurs est autorisé. Ces cinq erreurs sont composées d'un maximum de cinq délétions/mutations, les délétions ne devant intervenir qu'en début ou fin d'hélice.
- Il récapitule les informations obtenues sur les hélices dans une matrice de correspondance. Pour chaque TM (1 à 7), les numéros de l'emplacement et du nom du résidu de début et de fin sont récupérés.

Dans le cas des structures de dynamique moléculaire, le module d'initiation ne nécessite qu'un élément : un fichier de correspondance entre les résidus dans le fichier de structure et leur numéro Ballesteros pour extraire une matrice de correspondance.

### 2.2.1.2 Calcul des interactions intramoléculaires du domaine transmembranaire

Cette partie est effectuée par le module de production des scripts python tMimiR (structures RX/RMN), et iMimiR (trajectoires de dynamiques moléculaires) et vise à calculer les interactions polaires intramoléculaires. Les scripts utilisent le logiciel IChem<sup>27</sup> (version 5.2.9) et notamment son module INTS<sup>28</sup> a été utilisé pour détecter les interactions non covalentes présentes dans les structures.

Dans le cas de structures expérimentales, les paramètres par défaut ont été utilisés, à l'exception de la distance maximale de détection des interactions apolaires, qui a été mise à 0 Å, et de la distance minimale de détection des liaisons hydrogène qui a été mise à 2,2 Å. Les interactions hydrophobes ne sont pas calculées. Le seuil de détection des liaisons hydrogène a été abaissé de 0.3 Å par rapport à la publication originale afin de ne pas omettre

d'interactions artificiellement contraintes lors de la construction du modèle dans la densité électronique.

Dans le cas de trajectoires de dynamique moléculaire, les paramètres de base ont été utilisés, à l'exception de la distance maximale de détection des interactions apolaires, qui a été définie à 0 Å pour les mêmes raisons que précédemment.

Les interactions détectées dans ce travail par le module INTS du logiciel IChem sont donc : les liaisons hydrogène, les interactions ioniques et les interactions aromatique  $\pi$ - $\pi$  (Figure 2- 2).

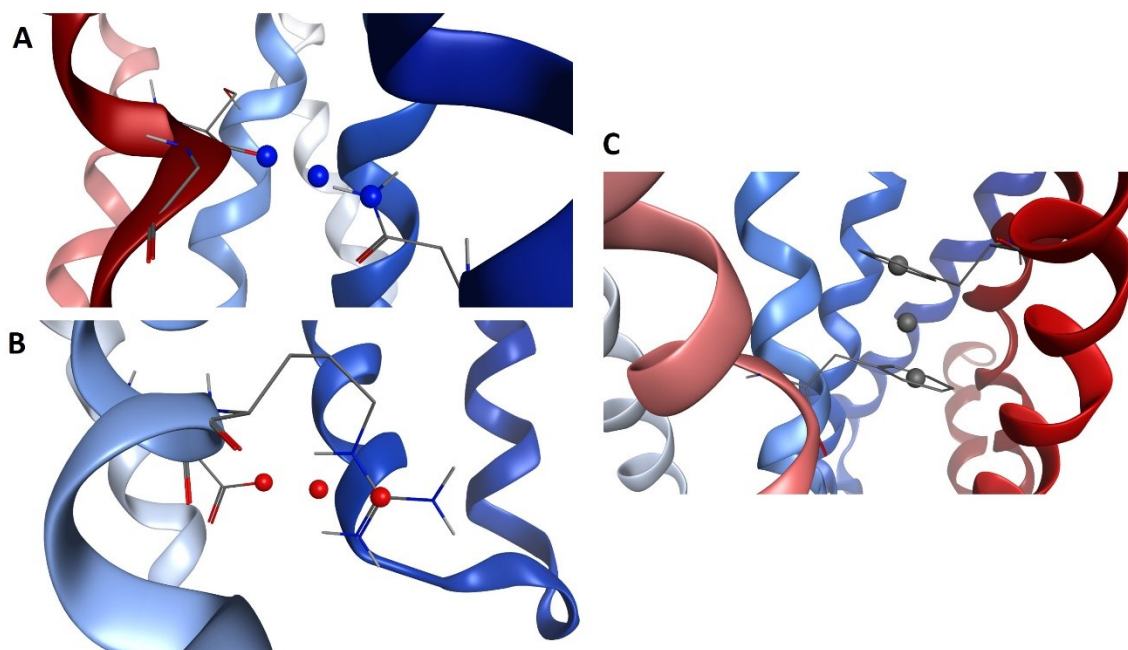


Figure 2- 2 : Types d'interactions détectées par le logiciel IChem-INTS pour la structure PDB 5jqh. De gauche à droite : une liaison hydrogène, une interaction ionique et une interaction aromatique. Les triplets de pseudo-atomes bleus (N), rouges (OD1) et gris (CZ) représentent dans le repère cartésien les interactions correspondantes. Deux pseudo atomes sont centrés sur les atomes impliqués dans l'interaction tandis que le troisième est centré au barycentre géométrique de la paire d'atomes en interaction.

Les interactions sont représentées sous forme de triplets de pseudo-atomes : un centré sur l'atome du premier résidu impliqué dans l'interaction, un autre centré sur l'atome du second résidu impliqué dans l'interaction et un troisième au centré au barycentre géométrique des atomes impliqué dans l'interaction. Ces atomes sont typés selon le type d'interaction que le triplet représente, atome CZ pour aromatique, N pour liaison hydrogène et OD1 pour liaison ionique.

Pour chaque structure analysée, le module INTS nous fournit donc un graphe d'interactions sous la forme d'un fichier Mol2 composé des triplets de toutes les interactions qu'il a détectées et un fichier texte listant ces interactions avec les informations sur le type d'interaction, les résidus et atomes concernés, les caractéristiques de distances et d'angle et un numéro d'identification de l'interaction.

### 2.2.1.3 Création de représentations simplifiées des structures

Cette partie est effectuée par le module de production des scripts python tMimiR, dans le cas de structures 3D de RX/RMN, et iMimiR, dans le cas de structures 3D de dynamiques moléculaires.

Dans les deux scripts d'analyse des structures de RCPG, les données prises en entrée sont d'une part le graphe d'interactions au format Mol2, le fichier texte de sortie du module INTS et d'autre par la matrice de correspondance. Les scripts mettent en relation les informations contenues dans le fichier texte du module INTS et la matrice de correspondance pour effectuer une sélection des interactions présentes dans le graphe d'interactions. Lors de cette sélection il a été choisi de séparer les liaisons hydrogène en trois types différents : inter-hélices, intra-hélices impliquant une chaîne latérale et intra-hélices impliquant uniquement la chaîne principale (interaction  $CO_{(i)}-NH_{(i+3/4)}$ ) abrégée  $(i,i+4)$  (Figure 2- 3).

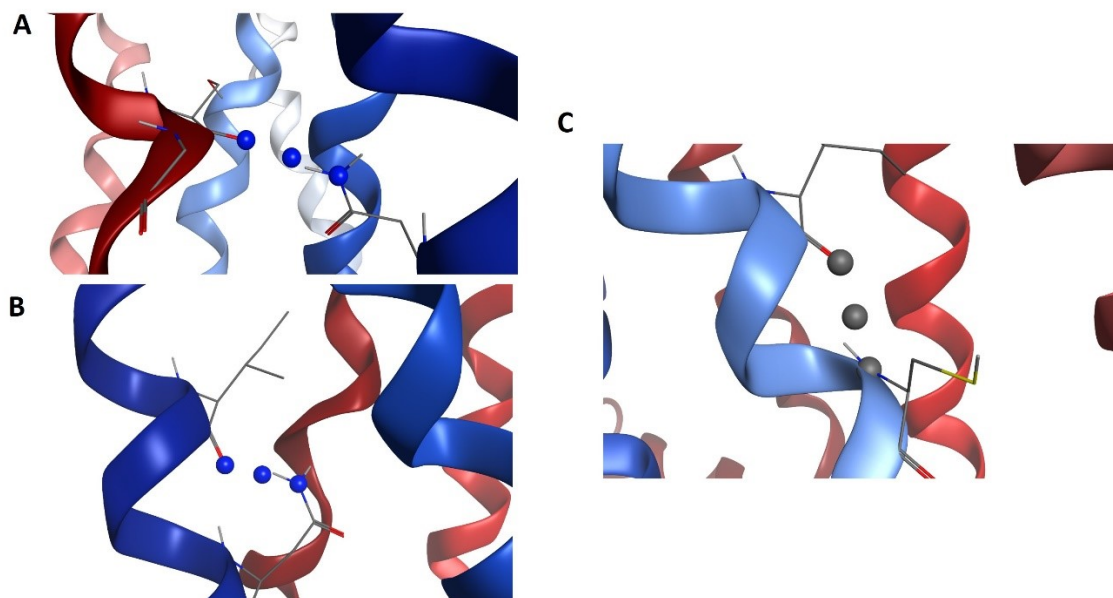


Figure 2- 3 : Distinction des liaisons hydrogène pour la structure PDB 5jqh. De gauche à droite les liaisons hydrogène inter-hélice, intra-hélice impliquant une chaîne latérale et intra-hélice impliquant uniquement la chaîne principale. Les triplets de pseudo-atomes bleu (N) et gris (CA) représentent dans le repère cartésien les interactions correspondantes, deux pseudo atomes sont centrés sur les atomes impliqués dans l'interaction tandis que le troisième est centré au barycentre géométrique de la paire d'atomes en interaction.

Les règles d'attribution des interactions à l'un des cinq types d'interaction sont les suivantes :

- Une interaction caractérisée par des pseudo-atomes de type CZ est aromatique.
- Une interaction caractérisée par des pseudo-atomes de type OD1 est ionique.
- Une interaction caractérisée par des pseudo-atomes de type N, ayant des numéros de résidu n'appartenant pas à la même hélice est une liaison hydrogène inter hélice.
- Une interaction caractérisée par des pseudo atomes de type N, ayant des numéros de résidu appartenant à la même hélice et pour laquelle les noms des deux atomes en interaction ne sont pas (N) et (O) est une liaison hydrogène intra-hélice impliquant une chaîne latérale.
- Une interaction caractérisée par des pseudo-atomes de type N, ayant des numéros de résidu appartenant à la même hélice et pour laquelle les noms des deux atomes en interactions sont (N) et (O) est une liaison hydrogène intra-hélice impliquant uniquement la chaîne principale (i,i+4). Le type atomique des pseudo-atomes caractérisant ces interactions ont été changé de (N) à (CA) pour faciliter leur différenciation.

Quatre combinaisons d'interactions ont été considérées pour former des graphes d'interactions : (1) les liaisons hydrogène inter-hélice, (2) les liaisons hydrogène sauf (i,i+4), (3) toutes les liaisons polaires sauf les liaisons hydrogène (i,i+4) et (4) toutes les liaisons polaires. Cette représentation sous forme de graphes d'interactions est la première forme de représentation simplifiée que les scripts produisent pour une structure.

La seconde représentation simplifiée est générée lors de l'étape d'attribution des interactions à l'un des cinq types d'interaction. À chaque évaluation d'une interaction prise en compte, elle est ajoutée à une matrice carré dissymétrique, qui est au final sauvegardée au format Numpy (format de sauvegarde binaire). Cette matrice répartit les interactions dans le triangle inférieur ou supérieure de la matrice selon leur nature. La répartition est schématisée dans la Figure 2- 4. Les axes horizontal et vertical de la matrice représentent les résidus composant les hélices, les entêtes de la matrice permettant de retrouver le numéro Ballesteros attribué au résidu. Les liaisons hydrogène sont ajoutées au triangle inférieur de la matrice, les interactions ioniques et aromatiques au triangle supérieur de la matrice. La diagonale correspond à une interaction d'un résidu avec lui-même quel que soit la nature de l'interaction.

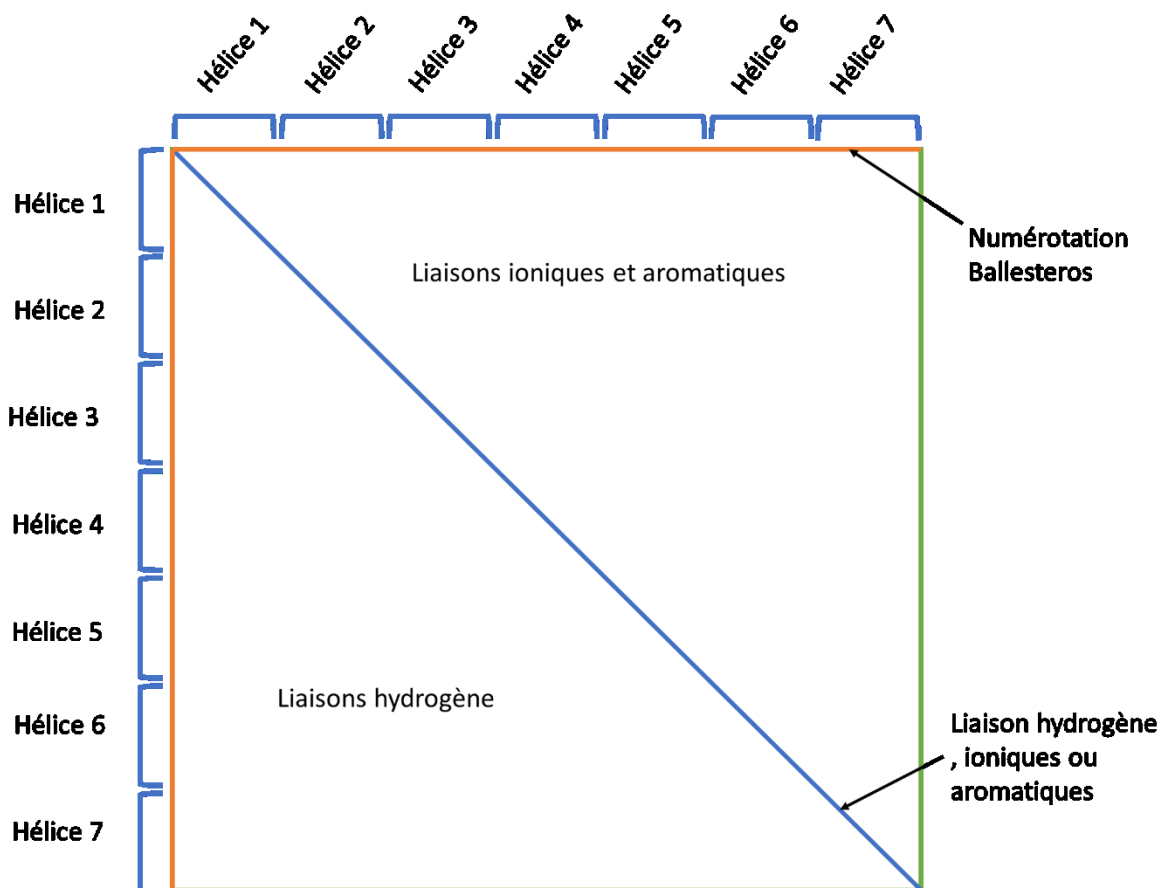


Figure 2- 4 : Représentation schématique d'une matrice regroupant toutes les interactions polaires d'une structure de RCPG.

## 2.2.2 Comparaison des structures.

La seconde étape du protocole vise à comparer les structures de RCPGs. Cette comparaison est effectuée à partir des représentations simplifiées des interactions intramoléculaires. Nous avons utilisé deux méthodes de comparaison, l'une avec alignement des représentations simplifiées, et l'autre sans alignement. Elles aboutissent toutes deux à une matrice contenant les scores de similitude de chaque couple possible de structures de RCPG.

### 2.2.2.1 Méthode avec alignement de graphes d'interaction

La méthode de comparaison avec alignement utilise en entrée les représentations simplifiées par graphes d'interaction des structures de RCPG. Pour comparer deux graphes, le script fait appel au logiciel IChem et plus particulièrement au module GRIM<sup>28,29</sup>. Il utilise l'algorithme modifié de Bron-Kerbosch<sup>30</sup> de détection de sous-graphe maximum commun. Les paramètres standard sont utilisés pour le module GRIM et permettent d'obtenir en sortie, la taille du sous-graphe maximum commun (clique), c'est à dire le nombre de pseudo atomes alignés.

Cet alignement est ensuite évalué, au moyen d'un score de similitude par indice de Tanimoto présenté dans l'Équation 1.

$$\text{Similarité}_{A \cdot B} = \frac{\sum \text{pseudoatomes}_{A \cap B}}{\sum \text{pseudoatomes}_A + \sum \text{pseudoatomes}_B - \sum \text{pseudoatomes}_{A \cap B}}$$

Équation 1 : Similitude entre les graphes d'interactions des structures A et B évalué par indice de Tanimoto.

### 2.2.2.2 Méthode sans alignement par comparaison de matrice d'interactions

La méthode de comparaison sans alignement utilise en entrée les représentations simplifiées par matrice d'interactions des structures de RCPGs. La comparaison entre deux matrices, est quantifiée selon l'Équation 2.

$$\text{Similarité}_{A \cdot B} = \frac{\sum \text{interactions}_{A \cap B}}{\sum \text{interactions}_A + \sum \text{interactions}_B - \sum \text{interactions}_{A \cap B}}$$

Équation 2 : Similitude entre les matrices d'interactions des structures A et B évalué par indice de Tanimoto

Il faut noter que le nombre d'interactions représentées par  $\sum \text{interactions}$  impose un poids de 1 pour chaque type de liaison (hydrogène, ionique, aromatique). Ainsi, si plusieurs interactions de même nature sont observées pour une même paire de résidus, elles seront comptées comme étant une seule et même interaction. Un exemple de ce genre de situation est présenté dans la Figure 2- 5.

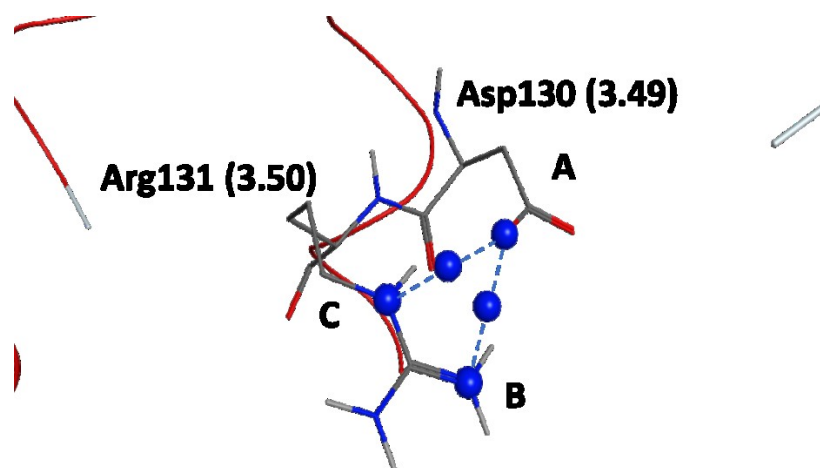


Figure 2- 5 : Représentation d'une double interaction ionique (A-B, A-C) calculée par le logiciel IChem-INTS pour deux mêmes résidus (Structures PDB 5jqh).

### 2.2.3 Partitionnement de structures 3D

La troisième étape du protocole vise à effectuer un partitionnement des structures de RCPG, il est effectué à partir des matrices de similitude issues de la comparaison des structures. Le partitionnement est effectué par un script python, dans le cas de structures expérimentales et prend en entrée le fichier d'information sur les structures à analyser (Tableau 2 - 1, p35), et un fichier texte de la matrice de similitude. Dans le cas des structures issues de dynamiques, moléculaires le partitionnement est effectué par un script R qui prend en entrée la matrice de similitude transformée et une liste de noms des structures étudiées.

#### 2.2.3.1 Classification hiérarchique ascendante à simple lien

Le script python dédié au partitionnement par classification hiérarchique ascendante (CHA) effectue plusieurs tâches. La première consiste à extraire du fichier d'information les classes attribuées à chaque structure PDB et à générer un partitionnement de référence pour chaque niveau de classe. La seconde tâche consiste à générer différents partitionnements à partir de la matrice de similitude. Ces partitionnements sont générés en considérant un seuil de coupure pour le score de similitude (qui varie de 0 à 1). Les différentes partitions qui en résulte agrègent les structures ayant un score de similitude supérieur au seuil de coupure. La troisième étape consiste à évaluer les différents partitionnements par le biais d'un score. Ce score caractérise la bonne prédiction d'un partitionnement issu de l'analyse des structures par rapport aux classes de référence. Il correspond à une proportion de structures bien classées. Les règles suivantes sont utilisées pour générer ce score.

- La classe d'une structure est celle qui est prédominante dans la partition.
- Une structure PDB dont la classe prédite est la même que celle de référence est correctement prédite.
- Une structure PDB qui est seul membre de sa classe est correctement prédite si le partitionnement la définit comme singulet.
- Une structure PDB dont la classe prédite est différente de celle de référence ou qui est incorrectement définit comme singulet est mal prédite.

Cette méthode de partitionnement a été implémentée dans le logiciel Cytoscape<sup>31</sup> (version 3.5.1). Dans le cas présent, ce logiciel n'a été utilisé que pour obtenir une visualisation graphique des partitions obtenues par le partitionnement par CHA.



### 2.2.3.2 Classification hiérarchique ascendante par la méthode de Ward

Le partitionnement utilisant la classification hiérarchique ascendante par la méthode de Ward a été réalisé à l'aide du logiciel R-cran<sup>32</sup>, à partir de la matrice de dissimilitude, obtenue par transformation de la matrice de similitude (1-similitude). La matrice de dissimilitude transformée en matrice de distance par le module "dist" est utilisée par le module "hclust" pour effectuer un partitionnement par la méthode de "ward.D2"<sup>33,34</sup>. La sortie de cette fonction peut être visualisée sous forme de dendrogramme et permet d'obtenir des partitions de structures si une coupure par nombre de partition est effectuée avec le module "cutree". L'interprétation du dendrogramme conduit à plusieurs partitions possibles. Le meilleur seuil de coupure est choisi par la comparaison des moyennes des similitudes intra-partition et de la similitude inter-partition. Un score calculé en soustrayant la moyenne de similitude intra-partition à la moyenne de similitude inter-partition et divisé par l'écart-type de la moyenne de similitude inter-partition, est calculé pour chaque partition. La coupure générant le plus grand nombre de partitions ayant un score supérieur à 1 est sélectionnée.

### 2.2.4 Création des empreintes d'interactions

La quatrième étape du protocole vise à obtenir une visualisation graphique des interactions. Cette visualisation est effectuée à partir des matrices d'interactions. Un script python a été écrit pour générer une matrice consensus des interactions présentes dans les structures considérées, puis générer la visualisation graphique.

Deux cas de figure ont été envisagés :

- Les structures considérées sont toutes issues du même RCPGs. Dans ce cas, la matrice consensus est calculée en effectuant une moyenne simple des matrices d'interactions.
- Les structures considérées sont de RCPGs différents. Dans ce cas, la matrice consensus est calculée en effectuant une moyenne de moyenne. Une première moyenne simple est calculée sur les matrices d'interactions de chaque RCPGs, puis une seconde moyenne simple est calculée sur les matrices moyennes des différents RCPGs.

Dans les deux cas, la représentation graphique d'une matrice d'interactions consensus représente la fréquence des interactions dans les structures considérées (Figure 2- 6).

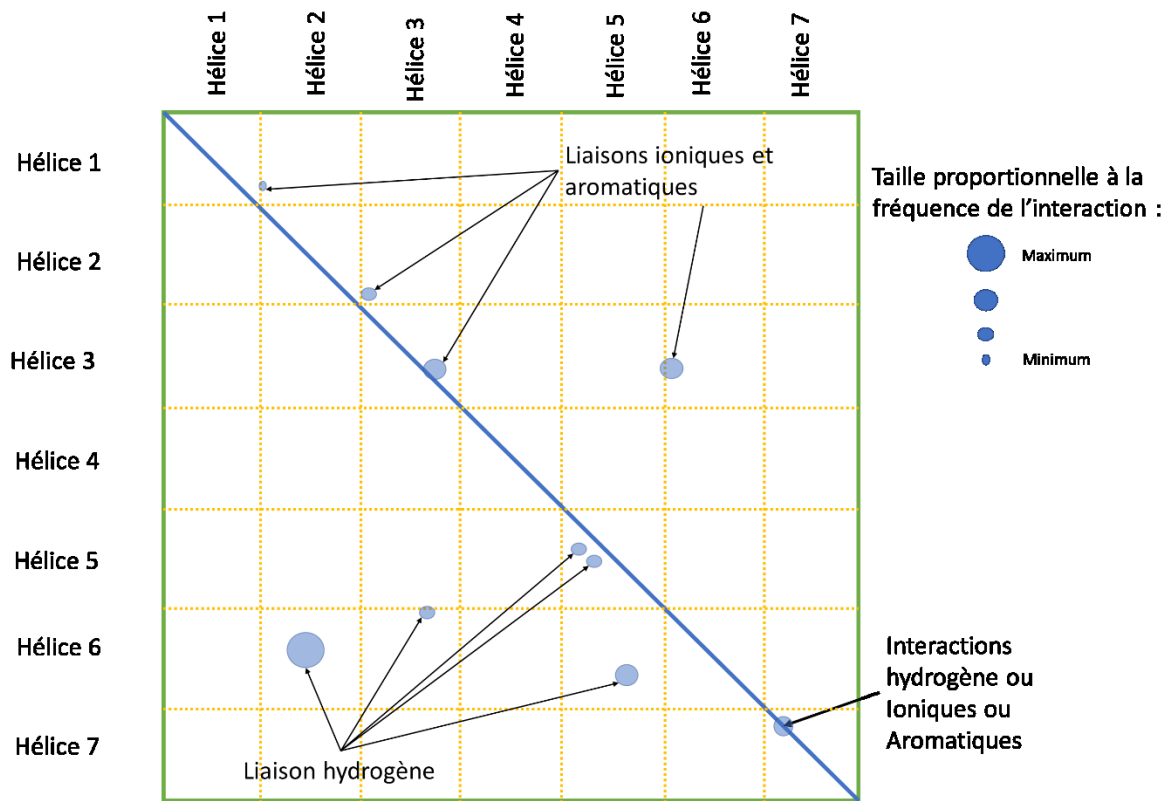


Figure 2- 6 : Représentation graphique de la matrice d'interaction consensus.

### 3. RÉSULTATS ET DISCUSSION

Dans cette partie, je décrirai les résultats obtenus à partir de l'analyse, par la méthode que j'ai développée, d'une part des structures de RCPG présent dans la PDB et d'autre part des structures de RCPG issues de dynamique moléculaire.

#### 3.1 DISTINCTION DES CLASSES DE RÉCEPTEURS COUPLÉS AU PROTÉINES G ET DE LEURS ÉTATS D'ACTIVATION

##### 3.1.1 Description des structures expérimentales de la PDB

L'espace structural des RCPGs est représenté au quatre août 2017 par 215 structures 3D dans la PDB. Ces structures représentent au total 47 récepteurs différents ou 50 si l'on considère les orthologues. Les récepteurs cristallisés sont issus à 66% de l'espèce humaine. Les 44% restant regroupent six espèces différentes (souris, rat, bœuf, lapin, sèche et dinde). Sur les 215 structures de la PDB, 13 ont été écartées car leurs séquences étaient trop éloignées de la

séquence native, avec notamment plus de cinq substitutions ou délétions consécutives au sein d'une hélice. Les 202 structures restantes ont une résolution allant de 1,72 Å à 7,7 Å et sont réparties selon leurs états d'activation soit : 59 activés, 126 inactivés et 17 pré-activés. 188 structures appartiennent à la classe A des RCPGs, sept à la classe F, six à la classe C et dix à la classe B1. Globalement quatre RCPGs représentent 45 % des structures PDB étudiées, les récepteurs ADRB1, ADRB2, AA2AR et OPSD. Seize récepteurs ne sont présents qu'en un exemplaire et 11 en deux exemplaires. Seuls dix récepteurs ont au moins une structure activée et une structure inactivée.

L'information associée à chaque structure expérimentale permet de les classer dans cinq différentes classes (Tableau 2 - 2). Ces classes représentent chacune un niveau croissant d'information sur la structure allant du niveau 1 qui correspond à la classe d'appartenance dans une classification phylogénétique, au niveau 5 qui correspond à l'état d'activation du récepteur associé.

Tableau 2 - 2 : Classification des structures de récepteurs couplés aux protéines G. Pour chaque niveau de classification un exemple d'annotation des différentes classes est fourni, le nombre total de classes est donné entre parenthèses.

Classification				
1 : Famille	2: Sous-famille	3: Orthologues	4 : Récepteur	5 : Etats d'activation ou fonctionnel par récepteur
Classe F	(1) Frizzled	(1) SMO	(1) Q99835	(2)
Classe C	(1) Métabotrope au glutamate	(2) GRM1, GRM5	(2) P41594, Q13255	(2)
Classe B1	(3) Glucagon, Calcitonin, etc.	(4) CALCR, CRFR1, etc.	(4) P43220 ,P47871 ,etc.	(5)
Classe A Autre	(2) US28, G1SGD4	(2) G1SGD4, U28	(2) P69332,G1SGD4	(2)
Classe A Sensorielles	(1) Opsines	(1) OPSD	(3) P02699	(4)
Classe A Nucleotide	(2) Adénosine, P2Y	(4) AA1R, AA2AR, etc.	(4) P29274, Q9H244 ,etc.	(6)
Classe A Lipide	(4) Cannabinoïde, acide gras libre, etc.	(4) FFAR1, S1PR1, etc.	(4) P21554, O14842 ,etc.	(5)
Classe A Protéine	(1) Chimiokine	(5) CCR2,CCR5, etc.	(5) P51681, P61073 ,etc.	(5)
Classe A Peptide	(8) Opioïde, Orexin, etc.	(13) OPRD, OPRM, etc.	(15) P41143, P30556 , etc.	(17)
Classe A Aminergique	(5) Adrenergique, Muscarinique, etc.	(10) ACM1, ACM2, etc.	(10) P07550 , P07700 ,etc.	(13)

Les structures expérimentales de RCPGs présentent un nombre variable de liaisons intramoléculaires dans le domaine transmembranaire (Tableau 2 - 3). Les liaisons les plus fréquemment observées sont les liaisons hydrogène intra-hélices ( $i, i+4$ ) tandis que les liaisons les plus rares sont les liaisons ioniques et aromatiques.

Tableau 2 - 3 : Tableau récapitulatif des différents jeux d'interactions intramoléculaires définis par le processus iMiR ainsi que le nombre minimum et maximum d'interactions présentes dans ses sets.

Nombre d'interactions Type de Liaison	Moyenne	Minimum	Maximum
Liaison hydrogène inter-hélice	9.72	2	26
Liaison hydrogène intra-hélice impliquant une chaîne latérale	16.9	6	27
Liaison hydrogène intra hélice impliquant la chaîne principale (i,i+4)	160.3	115	177
Ionique et aromatique	1.11	0	6

En moyenne, un RCPG présente dix liaisons hydrogène inter-hélices. Ce nombre varie d'un RCPG à l'autre avec un minimum de deux pour la structure 2LNL<sup>35</sup> du récepteur CXCR1 dans un état inactivé à 26 pour la structure 4XES<sup>36</sup> du récepteur de rat NTR1 dans un état pré-activé. Nous avons également pu noter que pour un même récepteur, dans le même état d'activation, le nombre de liaisons peut également varier. Par exemple pour les structures inactivées du récepteur ADRB2, le nombre d'interactions inter-hélice varie de six à 12.

### 3.1.2 Partition des structures

Pour évaluer la similitude entre structures et effectuer un partitionnement de celles-ci, nous avons considéré quatre combinaisons d'interactions (Tableau 2 - 4). La composition des graphes d'interactions issue de ces combinaisons est représentée dans la Figure 2- 1. Pour les quatre types de graphes, les interactions hydrophobes non pas été prises en compte, d'une part à cause de leur grand nombre qui résulte de l'architecture à sept hélices des RCPG, et d'autre part à cause de la sensibilité au résidu composant l'environnement utilisé pour leur génération.

Tableau 2 - 4 : Interactions considérées dans les représentations réduites des récepteurs couplés au protéines G.

		Liaison hydrogène Inter-hélice	Liaison hydrogène intra-hélices impliquant la chaîne latérale	Liaison ionique	Liaison aromatique	Liaison hydrogène intra-hélice impliquant uniquement la chaîne principale
Jeu 1	Liaison hydrogène inter hélice	X				
Jeu 2	Liaison hydrogène sauf (i,i+4)	X	X			
Jeu 3	Interactions polaires sauf (i,i+4)	X	X	X	X	
Jeu 4	Interaction polaire	X	X	X	X	X

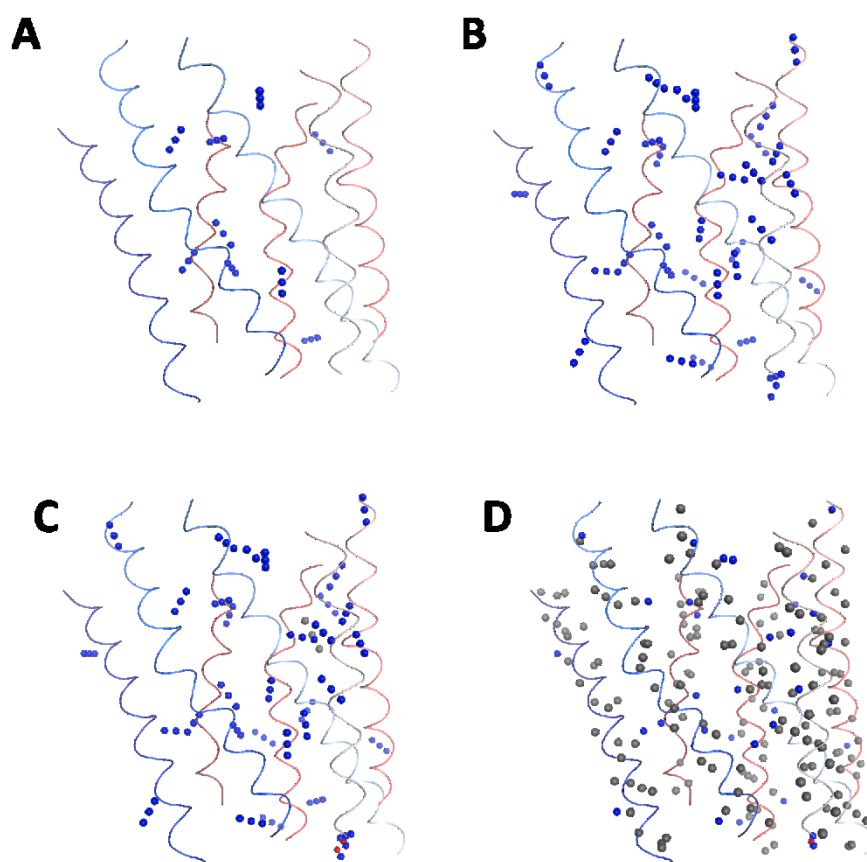


Figure 2- 7 : Représentation simplifiée de la structure 5jqh. (A) représentation par graphe de liaisons hydrogène inter-hélices, (B) représentation par graphe de liaisons hydrogène sauf (i,i+4), (C) représentation par graphe d'interactions polaires sauf les liaisons hydrogène (i,i+4), (D) représentation par graphe d'interactions polaires. (Les représentations des hélices en fil ne font pas partis des graphes.)

Le nombre moyen de point par graphe varie de 28, si l'on ne tient compte que des liaisons hydrogène inter-hélices, à 188 si l'on tient compte de toutes les interactions polaires au sein du domaine transmembranaire.

Les 202 structures de RCPG ont été groupées en fonction de la similitude entre leurs graphes d'interaction. Les proportions de structures bien classées, pour les partitions utilisant chacun des quatre jeux d'interactions par rapport au différentes classes définis dans le Tableau 2 - 1 sont répertoriées dans le Tableau 2 - 5.

Tableau 2 - 5 : Partition des structures expérimentales de RCPG de la PDB.

Classification	Jeu d'interactions	95 <sup>ème</sup> centile des scores de similarité	Structures bien classées (%)	Seuil de similarité	Similarité intra-partition		Similarité inter -partition		Score moyen de comparaison de moyenne
					Moyenne	Déviati on standard	Moyenne	Déviati on standard	
Famille	Jeu 1	0,19	44	0.300	0.501	0.126	0.090	0.031	13.245
	Jeu 2	0,13	63	0.130	0.351	0.129	0.039	0.013	23.606
	Jeu 3	0,14	63	0.140	0.361	0.104	0.038	0.013	24.831
	Jeu 4	0,14	63	0.140	0.388	0.163	0.050	0.016	20.616
Sous-famille	Jeu 1	0,19	57	0.300	0.501	0.126	0.090	0.031	13.245
	Jeu 2	0,13	83	0.130	0.351	0.129	0.039	0.013	23.606
	Jeu 3	0,14	83	0.140	0.361	0.104	0.038	0.013	24.831
	Jeu 4	0,14	80	0.130	0.368	0.158	0.049	0.016	20.165
Orthologue	Jeu 1	0,19	47	0.330	0.532	0.157	0.093	0.034	12.839
	Jeu 2	0,13	70	0.130	0.351	0.129	0.039	0.013	23.606
	Jeu 3	0,14	70	0.140	0.361	0.104	0.038	0.013	24.831
	Jeu 4	0,14	64	0.130	0.368	0.158	0.049	0.016	20.165
Récepteur	Jeu 1	0,19	57	0.330	0.532	0.157	0.093	0.034	12.839
	Jeu 2	0,13	77	0.130	0.351	0.129	0.039	0.013	23.606
	Jeu 3	0,14	78	0.140	0.361	0.104	0.038	0.013	24.831
	Jeu 4	0,14	70	0.130	0.368	0.158	0.049	0.016	20.165
Etats d'activation	Jeu 1	0,19	63	0.330	0.532	0.157	0.093	0.034	12.839
	Jeu 2	0,13	74	0.210	0.430	0.127	0.042	0.018	21.475
	Jeu 3	0,14	75	0.180	0.383	0.126	0.040	0.017	20.824
	Jeu 4	0,14	77	0.160	0.421	0.162	0.050	0.017	21.923

La partition des structures calculée à partir de la similitude des graphes discrimine efficacement les sous-familles de RCPG, et ce, quel que soit le jeu d'interactions, à l'exception des liaisons hydrogène inter-hélices.

Ainsi sur les quatre jeux d'interactions, représentant les RCPGs selon un niveau différent d'information, seules les liaisons hydrogène inter hélice (jeu 1) ne permettent pas d'obtenir un partitionnement correct des RCPGs. Les trois autres jeux d'interactions ont, quant à eux, des statistiques de classification similaires, réussissant à partitionner en moyenne plus de 70% des structures de RCPG. Parmi ces jeux d'interactions, les liaisons hydrogène inter/intra hélices (jeu 2) et les interactions polaires sauf les liaisons hydrogène (i,i+4) (jeu 3), permettent d'obtenir les meilleures classifications, à l'exception de la classification des états d'activation, qui est plus efficace en utilisant toutes les interactions polaires (jeu 4). Néanmoins, cette meilleure classification ne permet qu'un gain de 2% pour une multiplication du nombre d'interactions nécessaire à la classification par plus de 6, passant ainsi de ~30 à ~200

interactions. Cette observation indique ainsi que l'information apportés par l'architecture des hélices est soit non nécessaire ou déjà présente dans les autres jeux d'interactions.

En observant les statistiques non plus au regard des jeux d'interactions mais des différentes classifications utilisées (Tableau 2 - 5), on peut remarquer que trois classes sont mieux prédites : les sous-familles, les récepteurs et les états d'activation par récepteurs. Ce résultat suggère qu'il n'y a pas assez d'interactions transmembranaires intramoléculaires des classes A à F des RCPGs. Pour autant, le nombre de points communs est suffisant afin de bien classer les orthologues. Par ailleurs, la prédiction de récepteurs est meilleure que celle des orthologues, ce qui semble indiquer que deux récepteurs de deux espèces différentes sont autant dissimilaires du point de vue des interactions intramoléculaires transmembranaires que deux récepteurs différents d'une même espèce.

Si, pour chaque classification et jeu d'interactions, la valeur moyenne de similitude des classes varie significativement de la valeur moyenne inter classe, on peut remarquer que le jeu 1 obtient à chaque fois la plus haute valeur moyenne de similitude mais aussi le plus bas taux de structures bien classées. En considérant seulement les jeux d'interactions 2, 3 et 4, on peut remarquer que les seuils de coupure des taux maximum de structures bien classés sont les mêmes pour les classifications par famille, orthologue et récepteur. Au regard des statistiques, les interactions polaires ne prenant pas en compte les liaison hydrogène ( $i, i+4$ ) permettent la meilleure prédiction des RCPGs et de leurs états d'activation.

La visualisation graphique des partitions permet d'identifier les RCPGs et structures posant problème. Parmi les partitions par récepteur (Figure 2- 8) et par leur état d'activation de récepteur (Figure 2- 9), obtenues en utilisant le jeu 3, on peut se rendre compte que les erreurs générées par la méthode sont issues de la non-différentiation des structures de certaines sous-familles de récepteurs. En effet, les structures posant problème sont issues en grande majorité de quatre sous-familles de la classe A et regroupent des récepteurs homologues. Les récepteurs à l'adénosine (AA1AR/AA2AR), muscariniques (ACM1/ACM2/ACM3/ACM4) et opiacés (OPRD/OPRK/OPRM/OPRX) contribuent de manière limitée à l'erreur générée par la méthode en agréant des récepteurs de même sous-famille ayant un nombre de structures limitées. Les récepteurs adrénérgiques apportent la plus grande contribution négative en agréant des structures du récepteur ADRB1 de dinde aux structures de ADRB2 humain. Dans les deux cas de figure, ces erreurs impliquent qu'il n'y a pas assez de différence dans les



interactions intramoléculaires transmembranaires des structures de ces récepteurs pour permettre de les différencier. Ceci peut néanmoins être nuancé pour les récepteurs à l'adénosine, muscariniques et opiacés à cause du faible nombre de structures disponibles.

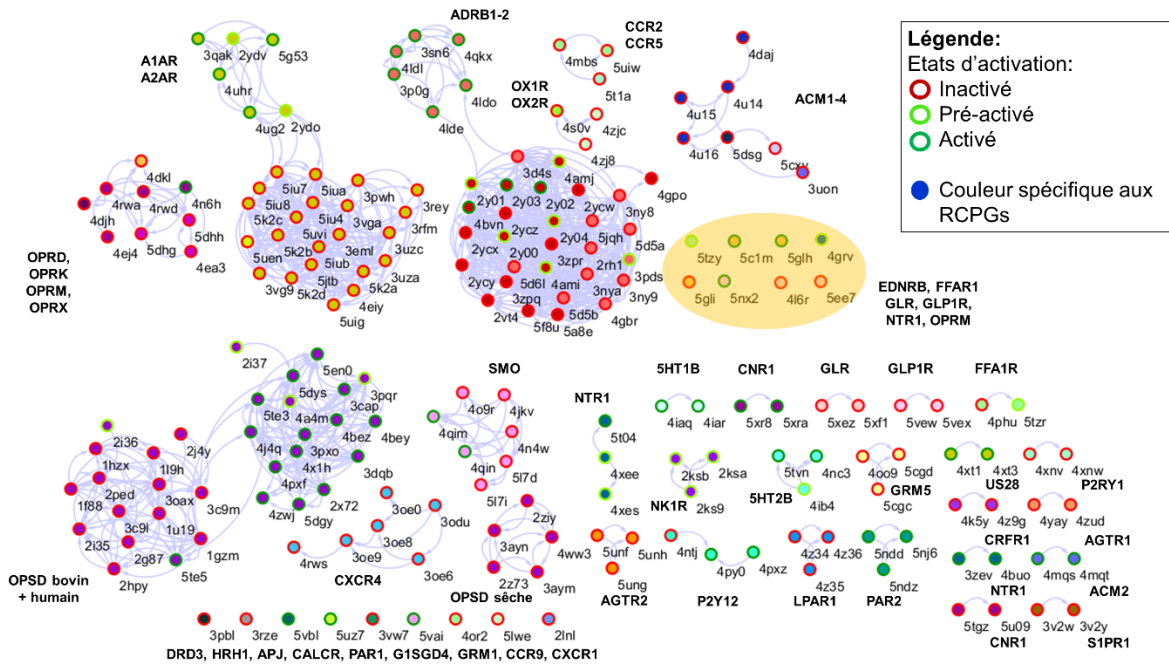


Figure 2- 8 : Partition des structures expérimentales de RCPG, obtenue en utilisant la similitude entre graphes issus du jeu 3 d'interactions (seuil de coupe de 0,14). Les structures sur fond jaune correspondent aux structures "faux" singulet.

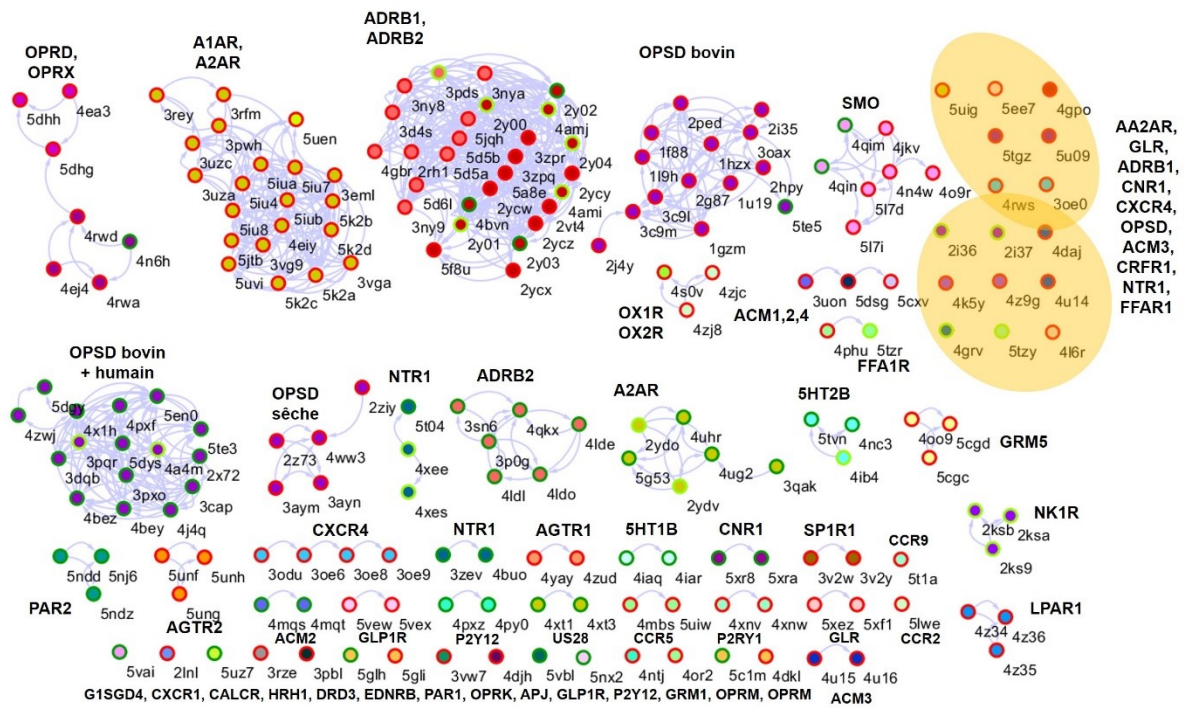


Figure 2- 9 : Partition des structures expérimentales de RCPG, obtenue en utilisant la similitude entre graphes issus du jeu 3. (Seuil de coupe de 0,18). Les structures sur fond jaune correspondent aux structures "faux" singulet.



En se focalisant sur les partitions impliquant les trois récepteurs les plus représentés, autant en termes de structures que d'états d'activation, c'est-à-dire les récepteurs OPSD, AA2AR et ADRB2 (Figure 2- 10), on peut voir que le partitionnement effectué par la méthode permet de différencier les structures caractérisant les états activé et inactivé. De plus dans le cas des récepteurs ADRB2 et ADRB1, on peut remarquer que toutes les structures du récepteur ADRB1 sont uniquement comprises dans la partition étiquetée comme inactivée.

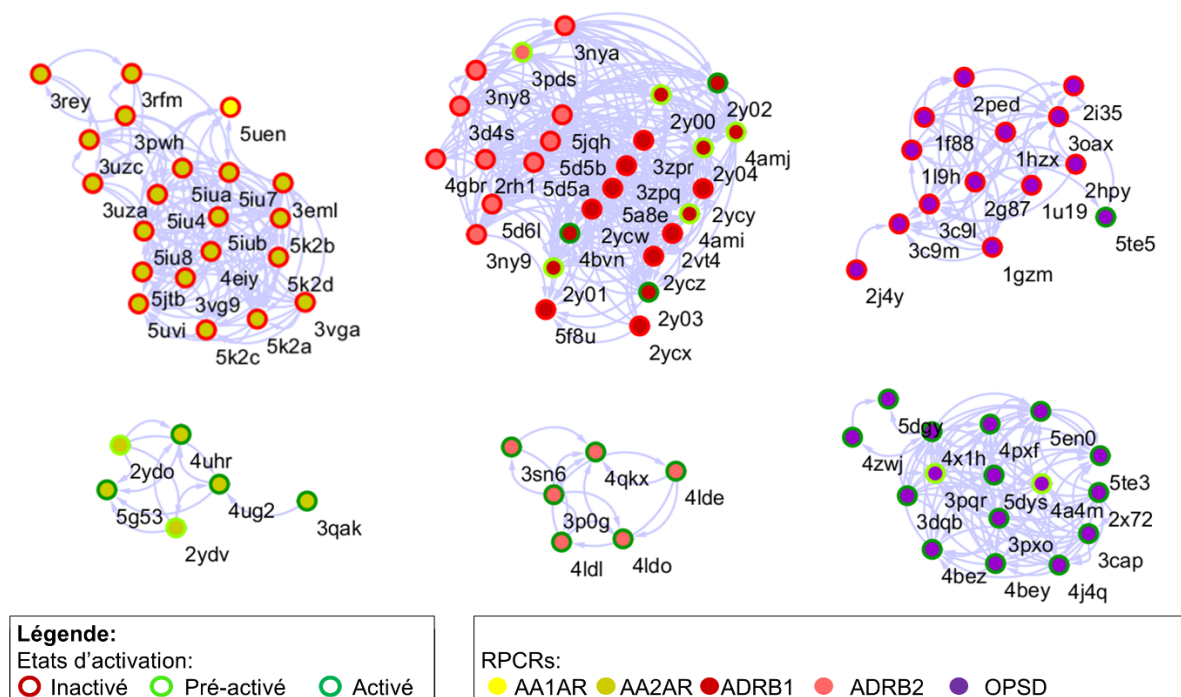


Figure 2- 10 : Zoom sur les récepteurs partitionnés de récepteurs AA1AR, AA2AR, ADRB1, ADRB2, et OPSD du partitionnement des structures expérimentales de RCPG, obtenue en utilisant la similitude entre graphes issus du jeu 3. (Seuil de coupure de 0,18).

### 3.1.3 Interactions prépondérantes

L'analyse des empreintes associées aux structures PDB permet de faire ressortir les interactions caractérisant les RCPGs. En sélectionnant les structures associées à un récepteur, on peut obtenir la fréquence des différentes interactions détectées. De même, on peut rechercher les interactions caractéristiques d'un état d'activation, ou encore celles qui seraient communes à tous les RCPGs.

Si on analyse l'ensemble des interactions présentes dans tous les RCPGs de la PDB, on peut remarquer qu'il y a peu de liaisons ioniques ou aromatiques et qu'aucune n'est commune à plusieurs RCPGs excepté le pont salin impliquant le motif DRY, présent dans les structures de 18 récepteurs. L'empreinte d'interaction des RCPGs révèle que des liaisons hydrogène intra-

hélice impliquant des résidus  $(i,i+4)$  s'établissent tout au long des sept hélices transmembranaire (cercle jaune sur la diagonale de la Figure 2- 11). Ces interactions sont moins fréquentes au centre des hélices 5 et 6. Il faut noter que dans cette représentation, les interactions entre résidus de la chaîne principale  $CO_{(i)}-NH_{(i+3/4)}$  ne sont pas représentées. Ces interactions correspondent donc aux interactions intra-hélices impliquant une chaîne latérale. L'empreinte montre également des liaisons hydrogène entre n'importe quelle paire d'hélices consécutive, avec une fréquence plus importante pour celles entre TM2 et TM3 et entre l'extrémité C-terminale de TM1 et TM7 (cercle bleu sur les contre-diagonales de la Figure 2- 11). Enfin, les liaisons hydrogène relient également des hélices non consécutives (cercle orange de la Figure 2- 11). Les liaisons les plus fréquemment observées impliquent les TM2 et TM4, TM2 et TM7 et TM3 et TM7. En proportion moindre, les liaisons hydrogène relient les TM3 et TM5, TM3 et TM6 et TM2 et TM6.

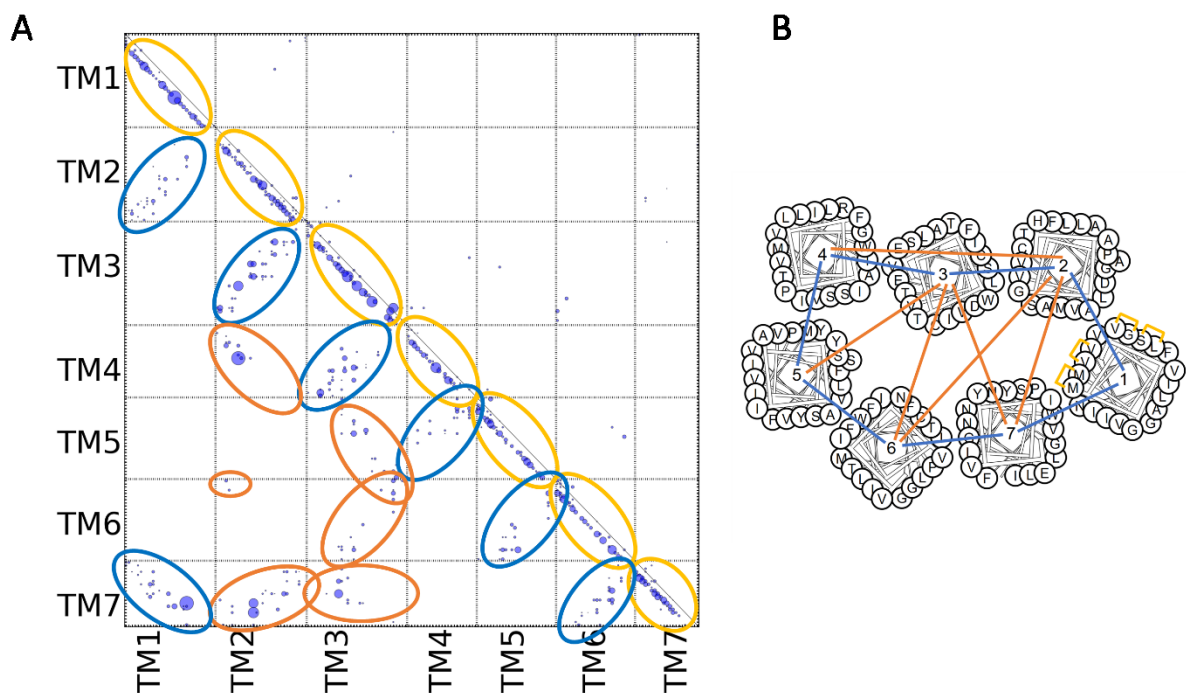


Figure 2- 11 : Empreinte cumulée des 50 RCPGs orthologues de la PDB. Les lignes bleues représentent des interactions hydrogène inter-TM entre TM consécutifs. Les lignes oranges représentent des liaisons hydrogène inter-TM entre TM non consécutifs. Les lignes jaunes représentent des interactions hydrogènes intra-TM. A) Représentation matricielle : Partie triangulaire supérieure : liaisons aromatiques et ioniques. Partie triangulaire inférieure : liaisons hydrogène sauf  $(i,i+4)$ . Diagonale : liaisons hydrogène, ioniques et aromatiques. B) Vue extracellulaire d'un RCPG schématisant les interactions entre TM.

Seules six interactions sur un total de 518 sont présentes dans plus de 40% des récepteurs dont trois dans plus de 65% des récepteurs (Figure 2- 12). Il s'agit des interactions entre résidus 1.46-1.50 ; 1.50-7.46, 2.45-4.50 pour celles présentes dans plus de 65% et 2.50-7.49,

3.35-3.39, 3.34-3.47 pour celles présent dans plus de 40%. Ces interactions impliquent des acides aminés conservés dans les classes A, B, C et F, notamment les résidus N(1.50), D(2.50), W(4.50), conservés respectivement à 98, 90, 97% dans la classe A.

Nous nous sommes ensuite intéressés aux états inactivés et activés des RCPGs et avons observé des couples d'états pour les récepteurs comportant le plus de structures des deux états, c'est-à-dire OPSD et AA2AR. OPSD est décrit par 16 structures activées et 13 inactivées. AA2AR est décrit par 6 structures activées et 20 inactivées.

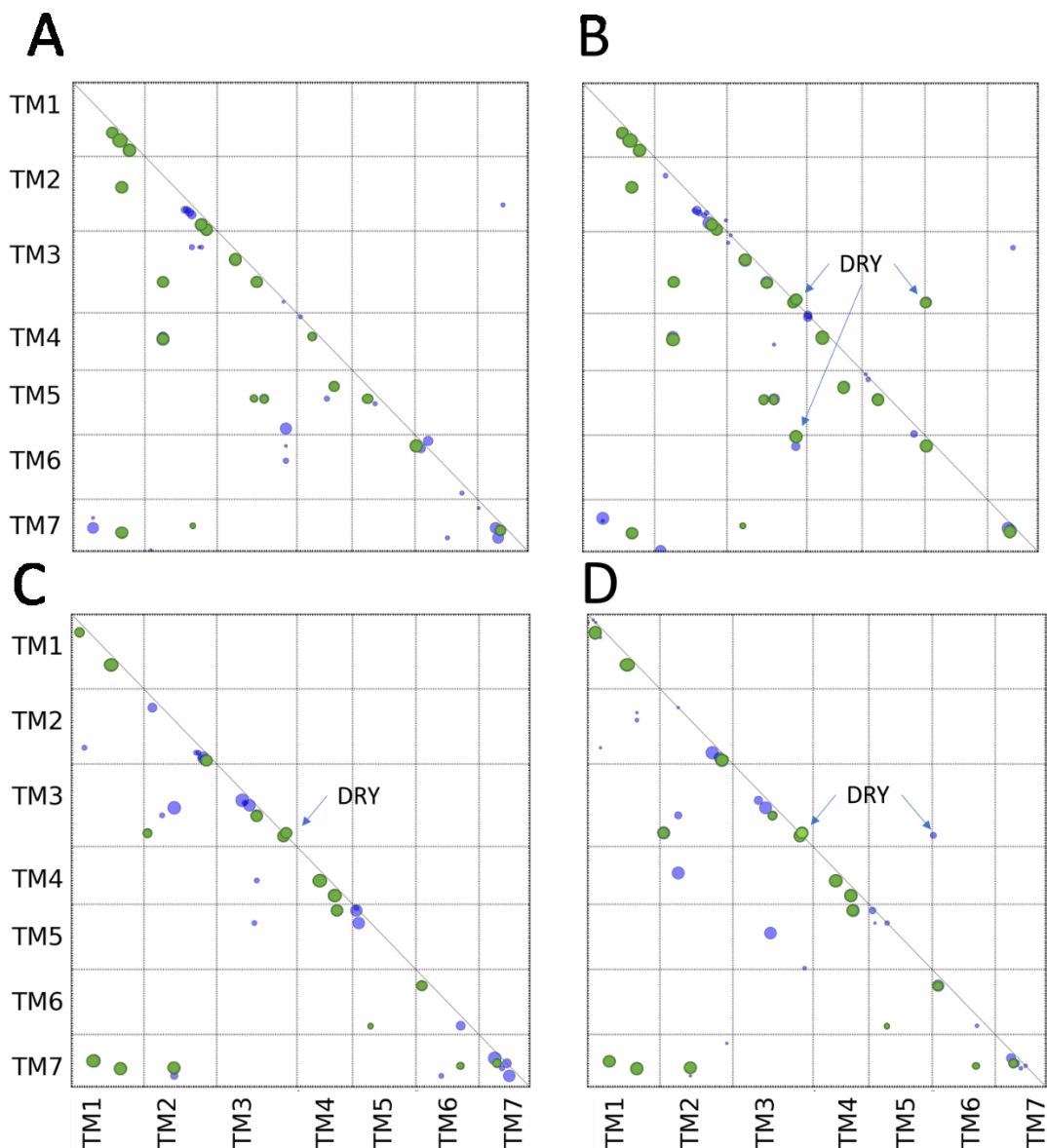


Figure 2- 12 : Empreinte des interactions des structures cristallographiques des récepteurs OPSD et AA2AR. A) Empreinte des structures du récepteur OPSD activé. B) Empreinte des structures du récepteur OPSD inactivé. C) Empreinte des structures du récepteur AA2AR activé. D) Empreinte des structures du récepteur AA2AR inactivé. Les points verts correspondent aux interactions communes aux deux états d'activation d'un récepteur. Les points d'interactions du motif "DRY" sont indiqués par des flèches.

En observant les empreintes pour les deux états fonctionnels du récepteur OPSD (Figure 2- 12 A, B), on peut remarquer 19 liaisons hydrogène communes aux deux états et présentes dans la majorité des structures. Dans le cas du récepteur AA2AR ce sont 16 liaisons hydrogène qui sont communes aux deux états (Figure 2- 12 C, D). Quatre de ces interactions sont présentes dans les deux récepteurs. Seule une liaison ionique entre deux mêmes résidus différents de manière significative selon les états dans OPSD et AA2AR. Cette interaction qui implique les résidus 3.50 et 6.30 est issue du motif DRY, spécifique de l'état d'activation. Elle n'est présente dans aucune structure active des récepteurs et dans respectivement 92 et 25% des structures inactivées des récepteurs OPSD et AA2AR.

#### 3.1.4 Conclusion

L'objectif de la méthode présentée dans ce chapitre est de pouvoir identifier, à partir des interactions intramoléculaires transmembranaires des structures de RCPG, des signatures moléculaires caractéristiques d'une classe, d'un récepteur ou d'un état d'activation de celui-ci.

L'analyse des interactions intramoléculaires transmembranaires des structures de RCPG de la PDB nous révèle qu'il est effectivement possible d'obtenir une signature des interactions spécifiques à un récepteur. Pour caractériser les sous-familles de récepteurs et les récepteurs eux-mêmes, il suffit de détecter leurs liaisons hydrogène inter-hélices et intra-hélices impliquant une chaîne latérale. En utilisant ce jeu d'interactions il est possible de regrouper 80% des structures de la PDB de manière similaire à une classification par comparaison de séquence qui reste à l'heure actuelle le meilleur moyen pour caractériser les familles de RCPG. Ce regroupement par liaisons intramoléculaires transmembranaires nous permet d'observer différentes associations entre récepteurs, avec notamment la différenciation des récepteurs orthologues (structures inactivées du récepteur OPSD bovin et OPSD sèche) et l'association de récepteurs homologues (structures inactivées du récepteurs ADRB2 humain avec les structures du récepteur ADRB1 de la dinde).

En caractérisant les liaisons hydrogène inter-hélices et intra-hélices impliquant une chaîne latérale, les liaisons ioniques et les liaisons aromatiques, il est possible de regrouper ensemble 75% des structures selon leurs états d'activation. Ces interactions permettent, pour un récepteur donné, de séparer ses structures activées de ses structures inactivées. La caractérisation de l'état d'activation d'un RCPG et le regroupement de ses structures est

usuellement effectuée par l'analyse de motif structuraux d'activation et de désactivation. Dans le cas de la méthode d'analyse par interactions transmembranaires, le regroupement des structures ayant le même état d'activation est effectué sans a priori sur la structure et connaissance de motif structuraux.

La création d'empreintes caractéristiques des interactions présentes dans les partitions de structures générées par la méthode permet d'identifier les architectures connues de RCPGs ainsi que les interactions spécifiques au récepteur et leurs états d'activation.

On retrouve dans ces empreintes l'architecture à sept hélices des RCPGs. L'analyse des empreintes permet d'identifier les interactions utilisées par la méthode pour partitionner les structures par sous-famille, récepteur et états d'activation. Ainsi, on peut remarquer que l'identification de partition associée à un RCPG ou à un état d'activation de RCPG est effectuée par des motifs spécifiques d'interactions. Dans ces motifs, le nombre d'interactions communes à toutes les sous-familles et aux récepteurs est extrêmement limité et semble être cantonné en partie aux interactions impliquant des résidus hautement conservés dans leur séquence. Pour les récepteurs de la classe A d'état fonctionnel inactif, une seule interaction est commune et correspond à la zone bien connue des résidus DRY des récepteurs.

On pourra noter que les spécificités structurales d'une hélice peuvent être encodées uniquement par les liaisons hydrogène impliquant une chaîne latérale, et ne nécessite pas de considérer les liaisons hydrogène  $(i, i+4)$ .

### 3.2 DISTINCTIONS DES ÉTATS CONFORMATIONNELS D'UNE TRAJECTOIRE DE DYNAMIQUE MOLÉCULAIRE

Nous voulons maintenant voir si les interactions intramoléculaires constituent un descripteur adapté à l'identification d'états conformationnels issus d'une simulation par dynamique moléculaire. Pour cela, nous avons sélectionné deux trajectoires publiées qui décrivent la transition d'un état activé ou pseudo activé à un état inactivé.

Pour des questions de temps de calcul, les similitudes entre interactions ont été évaluées par calcul matriciel sans alignement. Dans le cas de comparaison de structures d'un seul récepteur, la méthode utilisant des graphes à peu d'intérêt car il est peu probable que les

interactions entre deux paires différentes d'acides aminés occupent la même position relative aux sept hélices transmembranaires.

### 3.2.1 Récepteur muscarinique de type 2 à l'acétylcholine (ACM2)

La trajectoire de dynamique moléculaire du récepteur ACM2 montre trois états d'activation, deux intermédiaires (I1 et I2) et un état inactivé, distingués en fonction des distances entre résidus R121(3.50)–T386(6.34) et Y177(ECL2)–N410(6.58). La distance entre carbone alpha des résidus R121(3.50) et T386(6.34) est de 10.0 Å pour I1, de 12.0 Å pour l'état I2 et de ~6–7 Å pour l'état inactivé (Figure 2- 13).

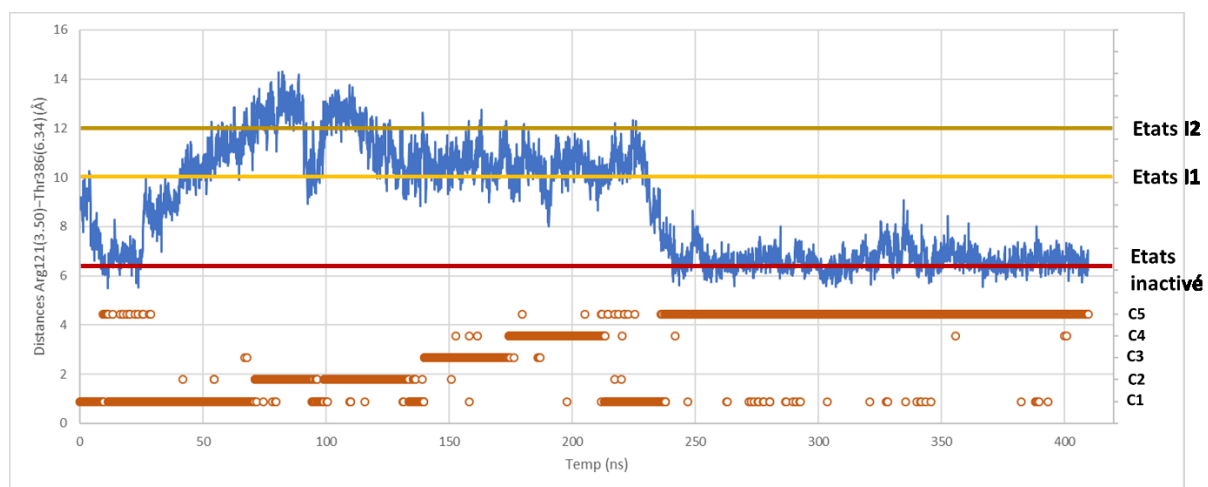


Figure 2- 13 : États d'activation de ACM2. En bleu est tracée la distance entre les carbones alpha des résidus Arg121(3.50) et Thr386(6.34) au long de la trajectoire. Les valeurs correspondant aux états inactivés, I1 et I2 sont indiquées par les lignes horizontales rouge, jaune et brune. La partition de la trajectoire d'après la similitude des interactions intra-TM est montrée en bas du graphique (C1-C5).

L'analyse de la dynamique a été effectuée en considérant le jeu d'interactions 3. La partition d'après la similitude d'interactions intramoléculaires dans le domaine transmembranaire découpe la trajectoire en cinq ensembles de structures (Figure 2- 13). Ces ensembles sont notés de C1 à C5 dans leur ordre d'apparition dans la dynamique. L'homogénéité des partitions est évaluée par l'analyse de la distribution des similitudes intra-partition en comparaison à celle obtenue pour les similitudes inter-partition (Figure 2- 14). Une partition est considérée comme homogène si la valeur de 1<sup>er</sup> quartile de la distribution de sa similitude intra-partition est supérieure à la valeur du 3<sup>ème</sup> quartile de la distribution de la similitude inter-partition. Dans le cas présent, deux des partitions, C2 et C3, ont une distribution se séparant significativement de la distribution inter-partitions permettant ainsi de les considérer comme caractéristiques d'un état stable du récepteur. On peut noter que la distribution de la partition C4 ne se sépare pas de la distribution inter-partition tandis que les

distributions des partitions C1 et C5 sont caractérisées par une valeur de 2<sup>ème</sup> quartile supérieure à la valeur du 3<sup>ème</sup> quartile de la distribution de la similitude inter-partition.

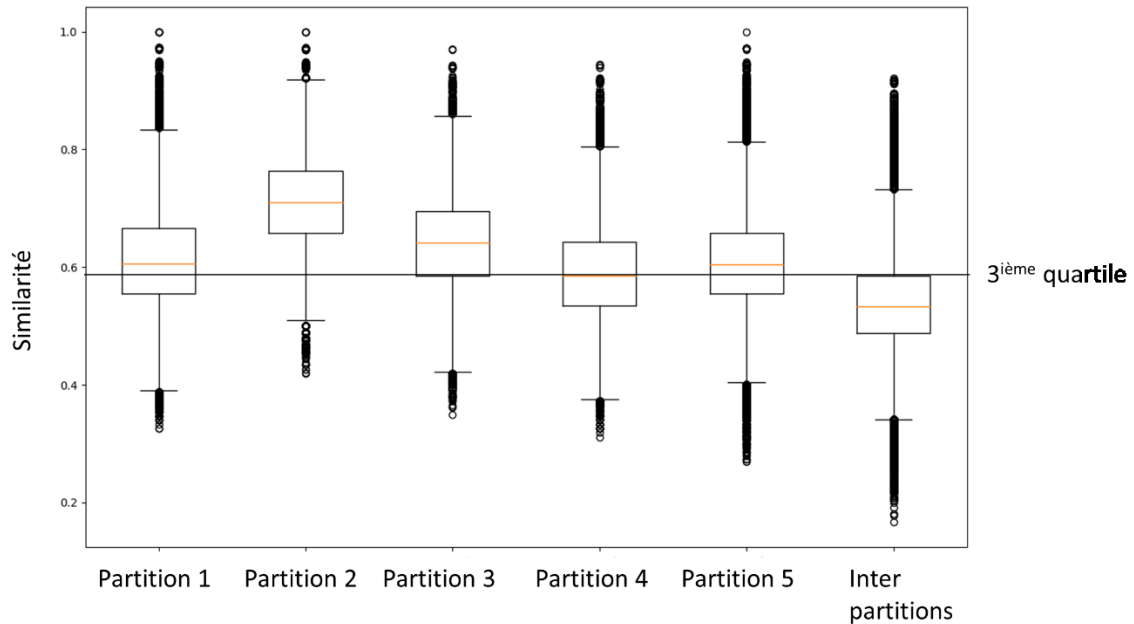


Figure 2- 14 : Représentation en boîte à moustache des similitudes inter-partitions et intra-partitions pour le partitionnement des structures du récepteur ACM2 selon le jeu d'interactions polaires n'impliquant pas les liaisons hydrogène ( $i, i+4$ ).

Dans le cas des partitions C2 et C3, on peut remarquer que les structures prises en compte dans ces partitions (Figure 2- 13) ne correspondent que partiellement aux états d'activation I1 et I2 définis par des caractéristiques de distance. Les structures de la partition C2 correspondent aux structures de l'état I2 mais aussi à des structures transitoires entre les états I2 et I1. Tandis que les structures de la partition C3 correspondent seulement à une fraction des structures de l'état I1. On peut remarquer que la partition C5, qui n'est pas considérée comme homogène, est la partition correspondant le mieux à un état d'activation, ici l'état inactivé.

Pour générer les représentations de ces partitions, nous avons choisi de considérer les partitions homogènes C2, C3 correspondant aux états I2, I1 mais aussi la partition C5 correspondant le mieux à l'état inactivé (Figure 2- 15).



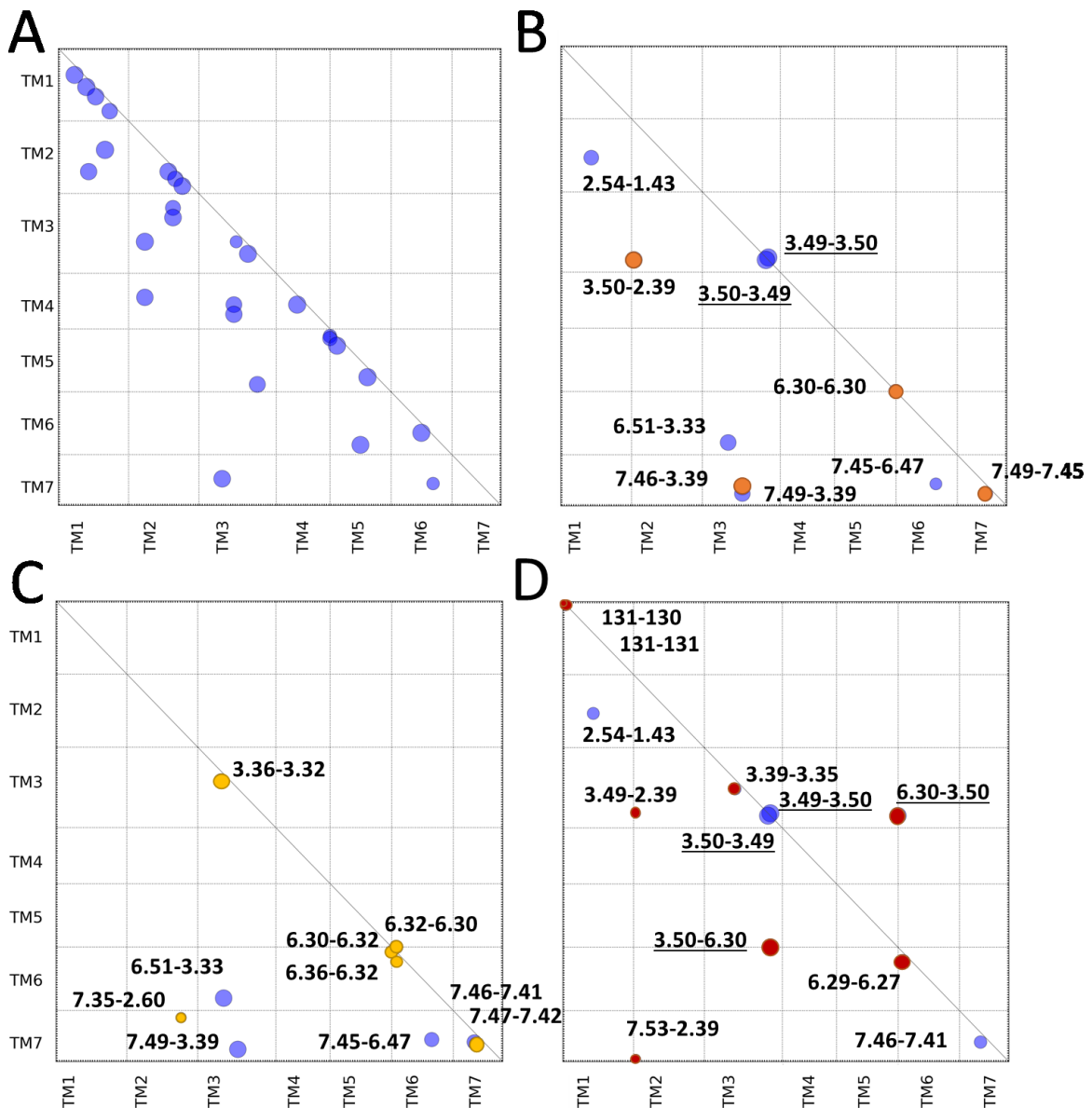


Figure 2- 15 : Représentation graphique des interactions intra-TM de ACM2. (A) Interactions communes à C2, C3 et C5. (B) Interactions présentes dans C2 et absentes dans C3 ou/et C5. (C) Interactions présentes dans C3 et absentes dans C2 ou/et C5. (D) Interactions présentes dans C5 et absentes dans C2 ou/et C3. Les couleurs orange, jaune et rouge montrent les interactions spécifiques de C2, C3 et C5, respectivement. Les interactions correspondant au motif DRY sont soulignées.

L'analyse des empreintes d'interactions des trois états définis pour ces trois partitions montre que 28 interactions observées tout au long de la trajectoire sont caractéristiques de ACM2, 24 autres interactions dépendent des différents états. Une partie de ces interactions est spécifique à chaque état fonctionnel tandis que certaines sont présentes dans deux partitions. On peut remarquer, parmi les interactions spécifiques, des interactions caractéristiques du motif DRY du récepteur dans le cas de l'état inactivé (Résidus 3.49-3.51). Une partie de ces interactions sont aussi présentes dans l'état I2. Pour les états I2 et I1, une série d'interactions spécifiques entre TM6-TM3, entre TM7-TM3 et entre TM7-TM2 sont probablement liées au



déplacement de l'hélice 6 lors de l'activation du récepteur. Ce déplacement semble aussi induire des modifications dans la partie intra cellulaire de l'hélice 6. L'ensemble des interactions moléculaires spécifiques aux différents états d'activation est résumé dans le Tableau 2 - 6.

Tableau 2 - 6 : Interactions spécifiques aux partitions C2, C3 et C5, respectivement associées aux états d'activations I2, I1 et Inactif.

Type de liaison	Résidu A	Résidu B	Proportion	État d'activation
Liaison Hydrogène	SER(7.46)	SER(3.39)	0,91	I2
Liaison Hydrogène	ARG(3.50)	ASN(2.39)	0,90	I2
Liaison Hydrogène	ASN(7.49)	ASN(7.45)	0,68	I2
Liaison Hydrogène/ionique	GLU(6.30)	GLU(6.30)	0,65	I2
Liaison Hydrogène	SER(3.36)	ASP(3.32)	0,73	I1
Liaison Hydrogène	THR(7.47)	TYR(7.42)	0,72	I1
Interactions Ionique	LYS(6.32)	GLU(6.30)	0,58	I1
Liaison Hydrogène	LYS(6.32)	GLU(6.30)	0,53	I1
Liaison Hydrogène	THR(6.36)	LYS(6.32)	0,42	I1
Hydrogène	THR(7.35)	TYR(2.60)	0,36	I1
Hydrogène	GLU(6.30)	ARG(3.50)	0,90	Inactif
Interaction Ionique	GLU(6.30)	ARG(3.50)	0,88	Inactif
Liaison Hydrogène	THR(6.36)	LYS(6.32)	0,70	Inactif
Liaison Hydrogène/ionique	GLU(1.31)	GLU(1.31)	0,62	Inactif
Liaison Hydrogène	GLU(1.31)	THR(1.29)	0,55	Inactif
Liaison Hydrogène	SER(3.39)	VAL(3.35)	0,47	Inactif
Liaison Hydrogène	ARG(6.29)	PRO(6.27)	0,36	Inactif
Liaison Hydrogène	TYR(7.53)	ASN(2.39)	0,31	Inactif
Liaison Hydrogène	ASP(3.49)	ASN(2.39)	0,30	Inactif

### 3.2.2 Récepteur adrénergique $\beta$ de type 2 (ADRB2)

La trajectoire de dynamique moléculaire du récepteur ADRB2, montre trois états fonctionnels : un activé, un intermédiaire et un inactif. Ils sont distingués par plusieurs mesures de distance inter-hélices et le RMSD par rapport à un état de référence, calculé sur trois zones particulières du récepteur. La zone de liaison du ligand est caractérisée par l'éloignement du carbone alpha de la sérine 207(5.47) à l'hélice 7 et la distance entre ce même résidu et l'agoniste. La zone de jonction est caractérisée par le RMSD aux structures activées et inactivées des résidus Ile121(3.40) et Phe282(6.44). Enfin, la zone de liaison de la protéine G distingue les états par la distance entre les carbones alpha des résidus Arg131(3.50) - Leu272(6.34) et le RMSD du motif NPxxY (3.22-3.27).

Dans mon analyse du récepteur, je reprendrai comme caractéristique des états fonctionnels ceux de la zone de liaison de la protéine G. Une première partie de la trajectoire simule la transition de l'état activé vers l'état intermédiaire, la seconde partie simulant la transition de l'état intermédiaire vers l'état inactif (Figure 2- 16).

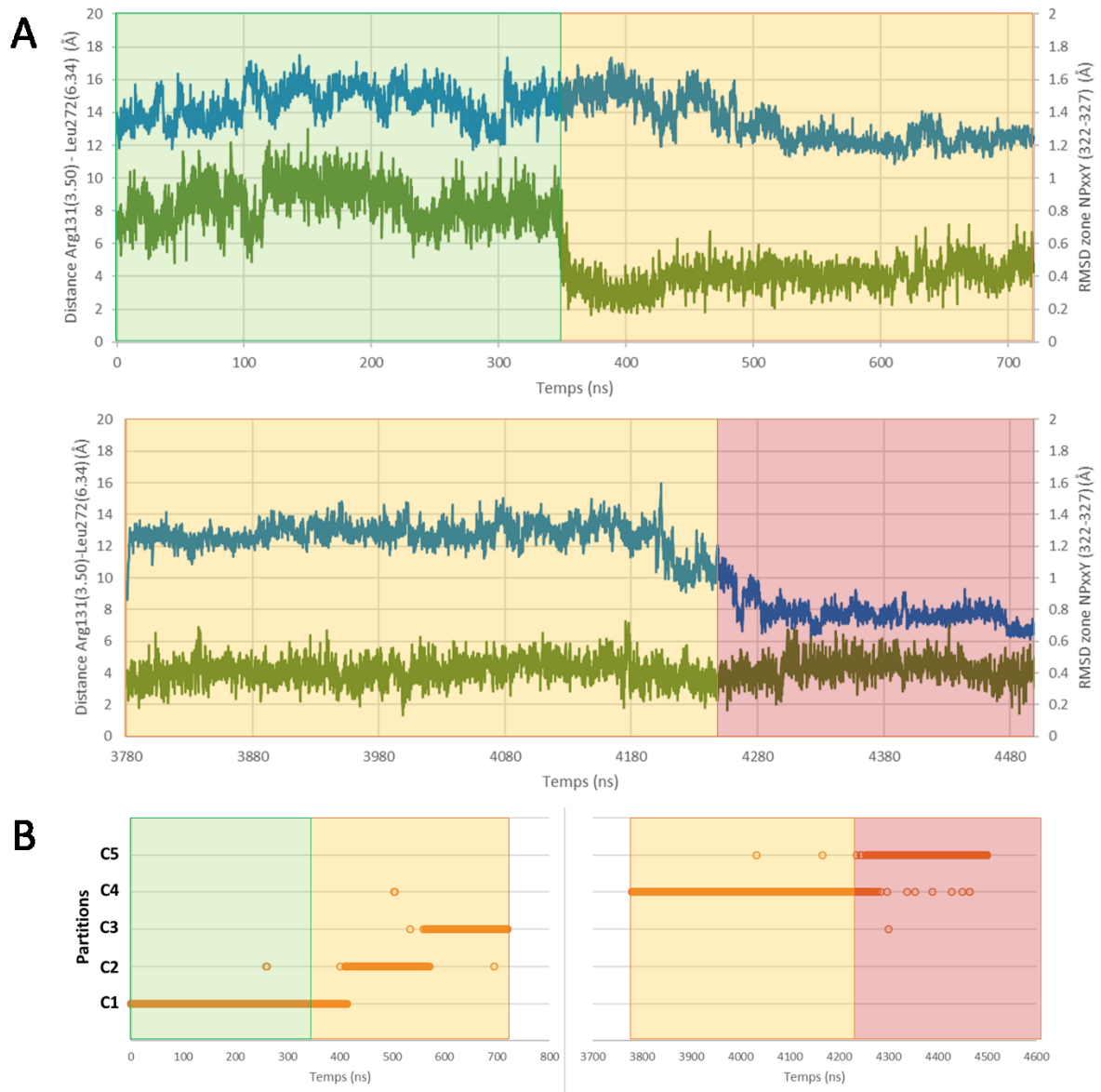


Figure 2- 16 : (A) Proportion de la trajectoire de ADRB2 dans les états activés (vert), intermédiaires (jaune) et inactifs (rouge). En bleu est tracée la distance entre les résidus Arg131(3.50)-Leu272(4.34) et en vert le RMSD du motif NPxxY (résidus 322 à 327) par rapport à la structure inactive 2RH1. (B) Partitions de la trajectoire d'après la similitude des interactions intra-TM (C1-C5).

L'analyse de la trajectoire de dynamique moléculaire a été effectuée en considérant le jeu d'interactions 3. La partition d'après la similitude d'interactions intramoléculeaires dans le domaine transmembranaire découpe la trajectoire en cinq ensembles de structures (Figure 2- 16). Ces ensembles sont notés de C1 à C5 dans leur ordre d'apparition dans la trajectoire.

L'homogénéité des partitions est évaluée par l'analyse de la distribution des similitudes intra-partition en comparaison à la similitude inter-partition (Figure 2- 17). Dans le cas présent trois de ces partitions, C1, C4 et C5, ont une valeur de 1<sup>ier</sup> quartile de leurs distribution supérieure à la valeur du 3<sup>ième</sup> quartile de la distribution inter-partitions et sont donc considérées comme caractéristique d'un état stable du récepteur.

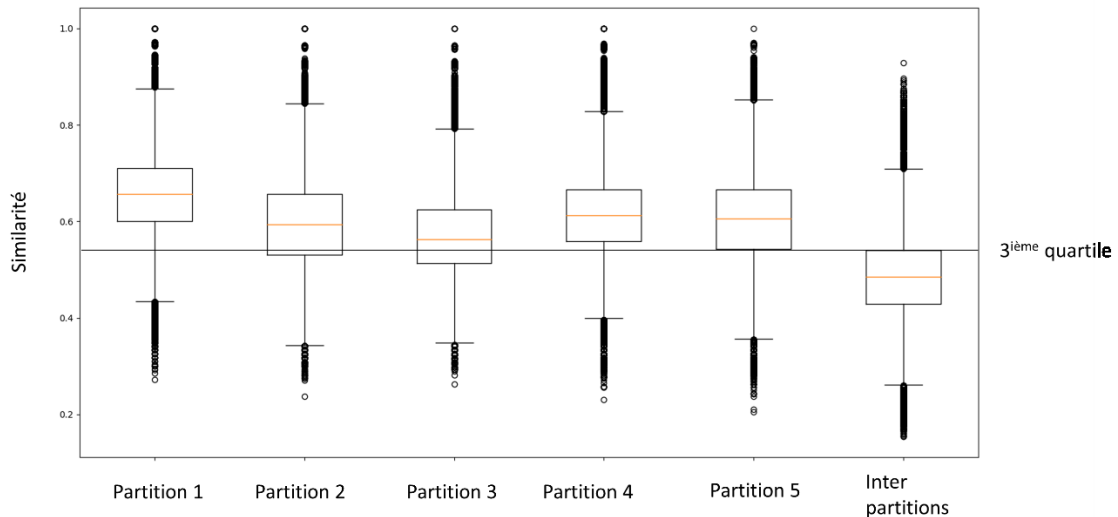


Figure 2- 17 : Représentation en boîte à moustache des similitudes inter partitions et intra partitions pour le partitionnement des structures du récepteur ADRB2 selon le jeu d'interactions polaires n'impliquant pas les liaisons hydrogène ( $i, i+4$ ).

Pour ces partitions (C1, C4, C5), on peut remarquer qu'elles correspondent aux zones de la trajectoire associées respectivement aux états activés, intermédiaires et inactivés, définis par la distance et le RMSD caractéristique. Comme dans la trajectoire du récepteur ACM2, ces partitions prennent en compte des ensembles de structures différents de ceux publiés par les auteurs, à l'exception de la partition correspondant à l'état inactivé.

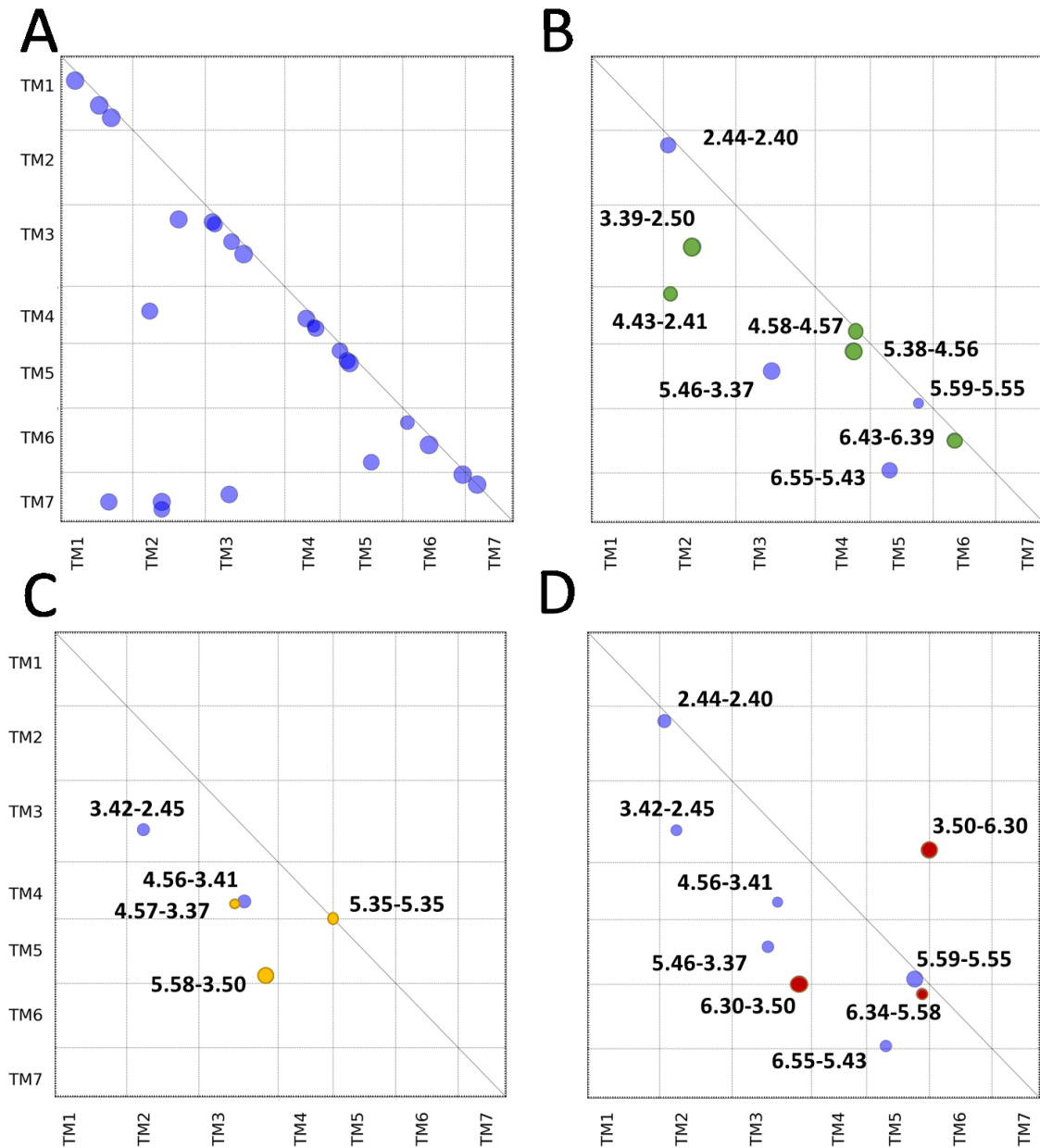


Figure 2- 18 : Représentations graphiques des interactions intra-TM de ADRB2. (A) Interactions communes à C1, C4 et C5. (B) Interactions présentes dans C1 et absentes dans C4 ou/et C5. (C) Interactions présentes dans C4 et absentes dans C1 ou/et C5. (D) Interactions présentes dans C5 et absentes dans C1 ou/et C4. Les couleurs vert, jaune et rouge montrent les interactions spécifiques de C1, C4 et C5, respectivement. Les interactions correspondant au motif DRY sont soulignées.

Pour ces trois partitions homogènes, l'analyse des empreintes d'interactions (Figure 2- 18) montre que 24 interactions sont observées tout au long de la trajectoire et sont caractéristiques de ADRB2 ; 17 interactions dépendent des états des partitions. Une partie de ces interactions est spécifique à chaque état d'activation (Tableau 2 - 7) tandis que certaines sont présentées dans deux partitions. On peut remarquer, parmi les interactions spécifiques, la présence d'interactions caractéristiques de la zone DRY du récepteur dans le cas de l'état inactivé.

Tableau 2 - 7 : Interactions spécifiques aux partitions C1, C4 et C5, respectivement associées aux états d'activation activé, intermédiaire et inactivé.

Type de liaison	Résidu A	Résidu B	Proportion	État fonctionnel
Hydrogène	SER(3.39)	ASP(2.50)	0,97	Activé
Hydrogène	TYR(5.38)	THR(4.56)	0,90	Activé
Hydrogène	THR(6.43)	ILE(6.39)	0,72	Activé
Hydrogène	PHE(4.58)	SER(4.57)	0,64	Activé
Hydrogène	ARG(4.43)	TYR(2.41)	0,60	Activé
Hydrogène	THR(4.56)	GLH(3.41)	0,53	Intermédiaire
Hydrogène	SER(4.57)	THR(3.37)	0,30	Intermédiaire
Hydrogène/ionique	ASN(5.36)	ASN(5.36)	0,30	Intermédiaire
Hydrogène	GLU(6.30)	ARG(3.50)	0,86	Inactivé
Ionique	GLU(6.30)	ARG(3.50)	0,84	Inactivé
Hydrogène	LEU(6.34)	TYR(5.58)	0,41	Inactivé

Parmi les interactions communes à chacune des partitions respectives des récepteurs ACM2 et ADRB2, on peut observer 9 interactions communes aux deux récepteurs (Figure 2- 19) : 1.50-1.46, 3.55-3.51, 4.53-4.49, 5.39-5.36, 5.42-5.38, 3.28-2.57, 4.50-2.45, 7.42-3.32, 6.52-5.48. Les récepteurs ACM2 et ADRB2 font partie de la classe A et plus particulièrement de la sous-famille des récepteurs aminergiques. On peut supposer que ces interactions contiennent les interactions caractéristiques de cette sous-famille de récepteurs d'autant que parmi les 18 résidus impliqués dans ces interactions, 13 sont conservés à plus de 60% (1.46 64%, 1.50 100%, 2.45 79%, 3.28 79%, 3.32 100%, 3.51 100%, 3.55 71%, 4.50 100%, 4.53 86%, 5.38 64%, 5.48 79%, 6.52 64%, 7.43 100%) dans les récepteurs muscarinique et adrénergiques. On notera la présence de deux interactions, 1.50-1.46 et 4.50-2.45, observables dans plus de 65% des structures expérimentales de la PDB (3.1.3, p51-53, Figure 2- 11).

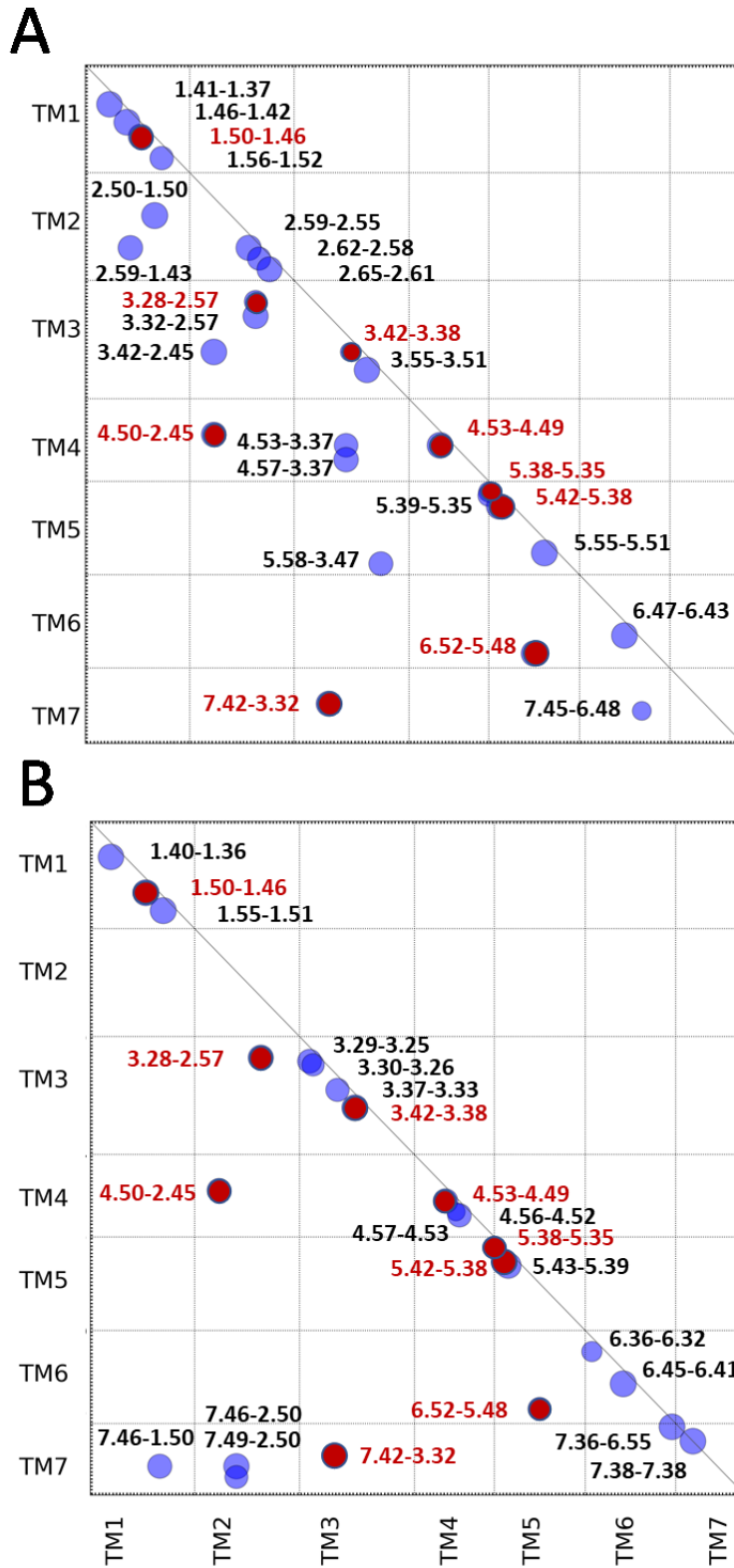


Figure 2- 19: Représentation des interactions présentes dans plus de 30% de la trajectoire de dynamique des récepteurs ACM2 (A) et ADRB2 (B). Les points rouges correspondent aux interactions communes entre les deux représentations.

### 3.2.3 Conclusion

L'objectif de la méthode dans le cas de l'analyse de dynamiques moléculaires est de pouvoir identifier des états conformationnels stables dans les simulations et de caractériser leurs interactions caractéristiques.

L'utilisation des liaisons hydrogène inter-hélices et intra-hélices impliquant une chaîne latérale, des liaisons ioniques et des liaisons aromatiques permet effectivement de caractériser des ensembles de structures partageant les mêmes interactions. Dans les deux dynamiques étudiées, les partitions homogènes générées correspondent en grande partie aux états d'activation définis par des marqueurs moléculaires d'activation. La caractérisation de ces partitions est effectuée sans connaissance sur les caractéristiques structurales d'activation du récepteur.

La génération par la méthode de partitionnement multiple pour l'état d'activation intermédiaire, indique que la méthode est sensible aux transitions conformationnelles. Pour autant ces partitions font partie des moins homogènes générées et sont donc considérées comme moins stables dans les dynamiques.

L'extraction des empreintes d'interactions associées aux différentes partitions homogènes permet d'identifier des interactions spécifiques aux RCPG et leurs états fonctionnels. De manière similaire à l'étude des structures cristallographiques, un motif d'interactions spécifiques semble apparaître pour chacun des deux récepteurs et des interactions spécifiques sont identifiables pour leurs états conformationnels stables. Le grand nombre d'interactions communes aux différentes partitions contraste avec le faible nombre d'interactions spécifiques. Dans les empreintes des partitions correspondant aux états inactivés des deux RCPGs, des interactions caractéristiques de la zone DRY sont à chaque fois présentes.

## 4. CONCLUSION

---

Ce chapitre décrit une méthode permettant d'analyser les structures de RCPGs, en extrayant leurs interactions intramoléculaires transmembranaires, et de comparer leurs similitudes. L'utilisation de cette méthode sur des ensembles de structures de RCPGs permet de partitionner celles-ci selon leur nature et leurs état conformationnel.

Plus précisément, l'utilisation de motifs de liaisons intramoléculaires ioniques, aromatiques et hydrogène impliquant une chaîne latérale, spécifique de la zone transmembranaire, permet de discriminer les structures représentant un même récepteur (ou récepteurs homologues) des autres récepteurs. L'utilisation de ces mêmes interactions permet aussi de discriminer les structures caractérisant les états fonctionnels activés et inactivés pour chaque récepteur. De manière similaire, l'utilisation de ces motifs, permet de discriminer les différents états conformationnels stables d'une trajectoire de dynamique moléculaire. Ces états conformationnels peuvent être mis en relation avec des états fonctionnels du récepteur.

L'identification d'ensembles de structures similaires du point de vue conformationnel et fonctionnel par les motifs d'interactions permet d'identifier les interactions communes ou spécifiques des différents ensembles. Ainsi, les ensembles de structures d'un même récepteur ou d'une même famille de récepteurs permettent d'identifier les interactions qui leurs sont caractéristiques. De manière similaire, les ensembles caractéristiques d'un état fonctionnel ou conformationnel permettent d'identifier les interactions caractéristiques des états d'activation.

En comparaison de méthodes de partitionnement usuellement utilisées, telle que le partitionnement par covariance de RMSD, la méthode présentée ici ne requière pas de connaissance préalable sur les résidus importants. Néanmoins, l'utilisation d'interactions pour effectuer le partitionnement rend la méthode sensible à l'état d'ionisation et de protonation de la structure et est caractérisée par un temps et un espace disque de calcul élevé.



## 5. BIBLIOGRAPHIE DE CHAPITRE

---

1. Fredriksson, R. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
2. Ballesteros, J. A. *et al.* Activation of the  $\beta_2$ -Adrenergic Receptor Involves Disruption of an Ionic Lock between the Cytoplasmic Ends of Transmembrane Segments 3 and 6. *J. Biol. Chem.* **276**, 29171–29177 (2001).
3. Trzaskowski, B. *et al.* Action of molecular switches in GPCRs-theoretical and experimental studies. *Curr. Med. Chem.* **19**, 1090–1109 (2012).
4. Manglik, A. & Kruse, A. C. Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry (Mosc.)* **56**, 5628–5634 (2017).
5. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem. Rev.* **117**, 139–155 (2017).
6. Dong, S. S., Goddard, W. A. & Abrol, R. Conformational and Thermodynamic Landscape of GPCR Activation from Theory and Computation. *Biophys. J.* **110**, 2618–2629 (2016).
7. Cvicsek, V., Goddard, W. A. & Abrol, R. Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLOS Comput. Biol.* **12**, e1004805 (2016).
8. Sanchez-Reyes, O. B. *et al.* G Protein-Coupled Receptors Contain Two Conserved Packing Clusters. *Biophys. J.* **112**, 2315–2326 (2017).
9. de Graaf, C. & Rognan, D. Customizing G Protein-coupled receptor models for structure-based virtual screening. *Curr. Pharm. Des.* **15**, 4026–4048 (2009).
10. Bissantz, C., Logean, A. & Rognan, D. High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional Model Building, and Receptor Library Screening. *J. Chem. Inf. Comput. Sci.* **44**, 1162–1176 (2004).
11. Ballesteros, J. A. & Weinstein, H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. in *Methods in Neurosciences* (ed. Sealfon, S. C.) **25**, 366–428 (Academic Press, 1995).

12. Pundir, S., Martin, M. J. & O'Donovan, C. UniProt Protein Knowledgebase. in *Protein Bioinformatics* (eds. Wu, C. H., Arighi, C. N. & Ross, K. E.) **1558**, 41–55 (Springer New York, 2017).
13. Pándy-Szekeres, G. *et al.* GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).
14. Bietz, S., Urbaczek, S., Schulz, B. & Rarey, M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J. Cheminformatics* **6**, (2014).
15. Lippert, T. & Rarey, M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *J. Cheminformatics* **1**, 13 (2009).
16. Haga, K. *et al.* Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**, 547–551 (2012).
17. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
18. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein-Coupled Receptor. *Science* **318**, 1258–1265 (2007).
19. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the  $\beta 2$  adrenoceptor. *Nature* **469**, 175–180 (2011).
20. Miao, Y. & McCammon, J. A. Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci.* **113**, 12162–12167 (2016).
21. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **11**, 3584–3595 (2015).
22. Dror, R. O. *et al.* Activation mechanism of the  $\beta 2$ -adrenergic receptor. *Proc. Natl. Acad. Sci.* **108**, 18684–18689 (2011).
23. Vilardaga, J.-P., Bünemann, M., Krasel, C., Castro, M. & Lohse, M. J. Measurement of the millisecond activation switch of G protein-coupled receptors in living cells. *Nat. Biotechnol.* **21**, 807–812 (2003).
24. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist- $\beta 2$  adrenoceptor complex. *Nature* **469**, 236–240 (2011).

25. Dror, R. O. *et al.* Identification of two distinct inactive conformations of the 2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci.* **106**, 4689–4694 (2009).
26. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85**, 2444–2448 (1988).
27. Da Silva, F., Desaphy, J. & Rognan, D. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein-Ligand Interactions. *ChemMedChem* **13**, 507–510 (2018).
28. Desaphy, J., Raimbaud, E., Ducrot, P. & Rognan, D. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *J. Chem. Inf. Model.* **53**, 623–637 (2013).
29. Slynko, I., Da Silva, F., Bret, G. & Rognan, D. Docking pose selection by interaction pattern graph similarity: application to the D3R grand challenge 2015. *J. Comput. Aided Mol. Des.* **30**, 669–683 (2016).
30. Bron, C. & Kerboscht, J. Finding All Cliques of an Undirected Graph [H]. *Commun. ACM* **16**, 575–577 (1973).
31. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
32. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2017).
33. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
34. Murtagh, F. & Legendre, P. Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? *J. Classif.* **31**, 274–295 (2014).
35. Park, S. *et al.* Structure of human CXCR1 in phospholipid bilayers. *Nature* **491**, 779–783 (2012).
36. Krumm, B. E., White, J. F., Shah, P. & Grisshammer, R. Structure of active-like neurotensin receptor. *Nat. Commun.* **6**, 7895 (2015).

# Chapitre 3

## Analyse conformationnelle du récepteur CCR5

## 1. INTRODUCTION

Le récepteur CCR5 est, avec le récepteur CXCR4, le principal corécepteur du VIH à la surface des cellules hôtes. L'entrée virale<sup>1</sup> (Figure 3 - 1) est initiée par la liaison de l'enveloppe du virus avec le récepteur CD4.

L'enveloppe du VIH est constituée de deux sous unités, la glycoprotéine 41 (gp41), insérée dans la membrane du virus, et la glycoprotéine 120 (gp120), liée de manière non covalente à gp41. La liaison de l'enveloppe du VIH au récepteur CD4 s'effectue par le biais de gp120. Cette liaison induit un changement conformationnel de la structure de gp120 qui permet la liaison de sa boucle v3 au corécepteur. Celui-ci induit un second changement conformationnel qui permet l'insertion de la gp41 dans la membrane cellulaire et par la suite la fusion entre les membranes virale et cellulaire.

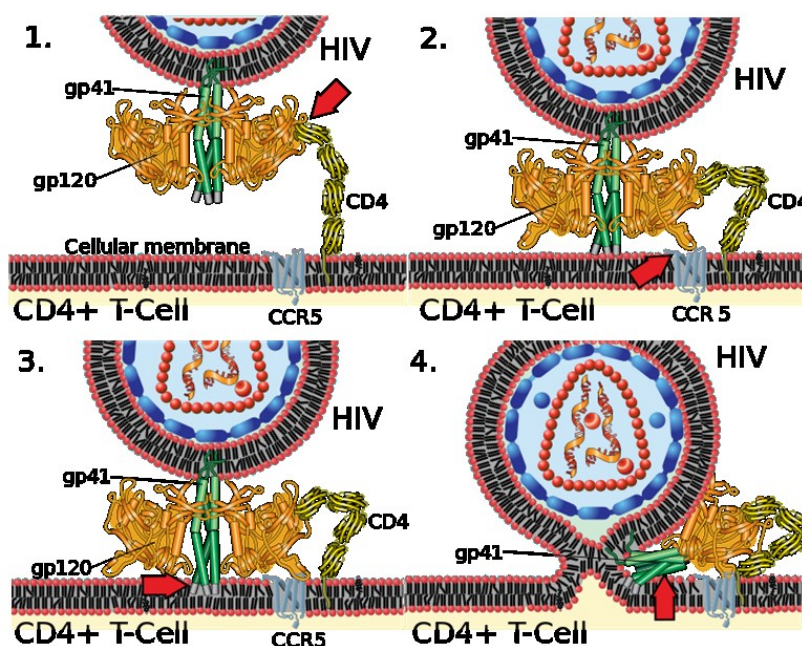


Figure 3 - 1 : Schéma de l'entrée du VIH dans une cellule. (1) Interaction initiale entre gp120 et CD4. (2) Changement conformationnel de gp120 permettant l'interaction avec CCR5. (3) Insertion de gp41 dans la membrane cellulaire. (4) Changements conformationnel de gp41 conduisant à la fusion des membranes cellulaire et virale. © Mikes Jones / Wikimedia Commons / [CC-BY-SA-3.0](https://commons.wikimedia.org/wiki/File:CC-BY-SA-3.0)

A l'heure actuelle, il n'existe pas de traitement permettant de guérir du syndrome d'immunodéficience acquise (SIDA) et d'éradiquer le VIH. Néanmoins, il existe des médicaments qui limitent l'infection virale. Le maraviroc (MVC) (Figure 3 - 2) stoppe la progression du virus<sup>2</sup> en ciblant le récepteur CCR5. Il le stabilise dans une conformation ne permettant pas la liaison de gp120, empêchant ainsi l'entrée du VIH dans les cellules.

Malheureusement des études *in vivo* et le suivi de patients traités au MVC ont montré l'émergence de résistance<sup>3</sup>. Cette résistance du virus se traduit par des modifications de gp120 permettant la liaison de CCR5, en présence de MVC<sup>4</sup>.

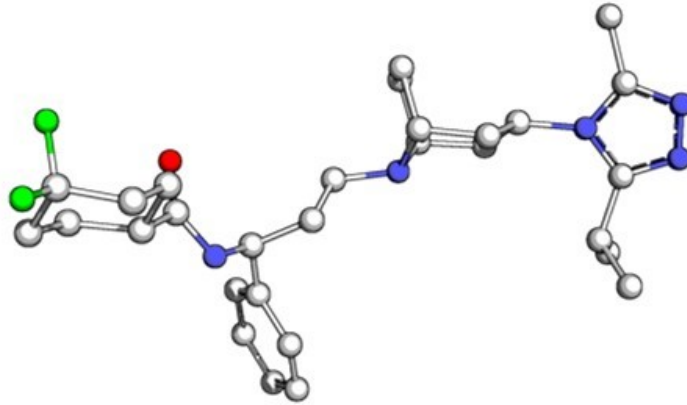


Figure 3 - 2 : Structure tridimensionnelle de maraviroc, telle que liée à CCR5 (identifiant PDB 4MBS). (Blanc : atomes de carbone, Bleu : atomes d'azote, Rouge : atome d'oxygène, Vert : fluor).

Le travail présenté dans ce chapitre s'intègre dans un projet de recherche collaboratif avec le Docteur Bernard LAGANE, membre du centre de pathophysiologie de l'hôpital universitaire Purpan de Toulouse. Ce projet a pour objectif de comprendre le lien entre la physiopathologie du VIH et la diversité conformationnelle du récepteur CCR5. Il vise à apporter des éléments de réponse à plusieurs questions :

- Quel est le rôle du tropisme cellulaire dans l'infection par le VIH ? Pourquoi le VIH a-t-il des préférences cellulaires, indépendamment de CCR5 qui est toujours exprimé ?
- Quel est le lien entre les conformations du récepteur et la possibilité de lier les effecteurs cellulaires ? Pourquoi le type des cellules qui expriment le récepteur CCR5 influe-t-il le mode de signalisation ?
- Quels sont les mécanismes de résistance du VIH à l'antiviral MVC ? Notamment comment le virus se lie à un récepteur occupé par MVC ?

Au sein de ce projet, mon objectif a été d'explorer l'éventail conformationnel disponible pour le récepteur CCR5 dans son état monomérique et de caractériser la dynamique et les spécificités structurales des principaux conformères du récepteur.

Tout comme les autres RCPGs de classe A, le récepteur CCR5 adapte son architecture structurelle en fonction des divers stimuli lui étant appliqués. Actuellement, seuls des

changements induits par l'inactivation du récepteur par un ligand sont caractérisés par deux structures résolues expérimentalement. Ces structures de l'état inactivé présentent le récepteur lié à un agoniste inverse (MVC)<sup>5</sup> ou à un antagoniste ([5P7]CCL5)<sup>6</sup>. Même s'il n'existe pas de structure résolue pour l'état activé de CCR5, des résidus essentiels à son activation par la liaison d'agoniste naturel ont été identifiés pour le ligand chimiokine CCL3<sup>7</sup>. Il a été montré que les voies de signalisation induite par l'activation par les chimiokines utilisent des protéines G $\alpha$ -i, G $\alpha$ -o et G $\alpha$ -12<sup>8</sup>.

On peut noter que CCR5 à un niveau basal d'activité. Ce niveau d'activité basal peut être modulé positivement ou négativement par des mutations spécifiques d'acides aminés. Ainsi, les mutations des résidus G286(7.42)<sup>9</sup> et T82(2.56)<sup>10</sup> respectivement en phénylalanine et proline rendent le récepteur constitutivement actif tandis que les mutations des résidus R126(3.50)<sup>11</sup> et P84(2.58)<sup>12</sup> respectivement en asparagine et alanine rendent le récepteur constitutivement inactif. De manière plus spécifique, les mutants des résidus T82(2.56) et P84(2.58) ciblent le motif TxP, conservé dans tous les récepteurs aux chimiokines. Le mutant T82P augmente spécifiquement la signalisation du récepteur CCR5 tandis que le mutant P84A diminue la liaison des chimiokines.

Pour obtenir un panel représentatif des différentes conformations, trois types de systèmes ont été sélectionnés pour caractériser le récepteur : un système natif libre, deux systèmes mutés et deux systèmes natifs liés à un ligand. Le système de référence est le système natif libre. Les deux systèmes mutés caractérisent le récepteur dans un état constitutivement actif (référéncé comme système T82P) et constitutivement inactif (référéncé comme P84A). Les deux systèmes liés caractérisent le récepteur lié à un ligand agoniste, la chimiokine de type C-C numéro 3 (référéncé comme CCL3) et un ligand agoniste inverse, l'antiviral MVC (référéncé comme MVC). Le choix d'étudier ces ligands a été motivé par leurs études poussées par notre laboratoire et nos collaborateurs. Le mutant constitutivement actif a été sélectionné car spécifique au récepteur CCR5 et le mutant constitutivement inactif car n'impactant pas le motif de résidus conservés DRY.

Pour proposer de nouvelles architectures du récepteur basé sur ces systèmes, nous nous sommes tournés vers la dynamique moléculaire. Les méthodes et technique de modélisation moléculaire ainsi que la capacité à modéliser des biomolécules ont été présentées dans de nombreux ouvrages et publications<sup>13,14</sup>. Un des facteurs limitant de l'utilisation de la

dynamique moléculaire pour modéliser des biomolécules telles que les RCPGs est le temps nécessaire pour observer les mouvements collectifs d'activation du domaine transmembranaire. Ces mouvements sont de l'ordre de la  $\mu$ s voir de la ms et prennent du temps à être simulé, en considérant une production de dynamiques classique de 100 ns par mois<sup>15,16</sup>. Le développement des méthodes et techniques de calcul tel que l'échantillonnage par réplica par méthode de Monte-Carlo, par chaîne de Markov, par métadynamique ou par hyperdynamique, ont permis d'accéder plus rapidement à de tels évènements. Notamment, l'une de ces techniques, la dynamique moléculaire accélérée<sup>17</sup>, a été utilisée avec succès sur plusieurs RCPG pour simuler un changement d'état<sup>18-20</sup>. Cette technique agit sur le profil énergétique de la dynamique moléculaire<sup>21</sup> en augmentant artificiellement différents niveaux énergétiques et permettant ainsi d'accéder aux états énergétiques de systèmes plus élevés (Figure 3 - 3), diminuant ainsi les temps de calcul nécessaire d'un facteur 4.

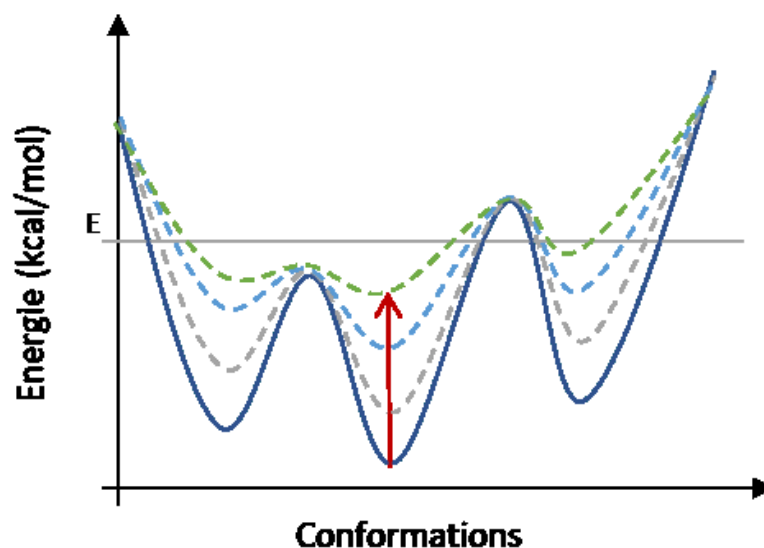


Figure 3 - 3 : Principe de l'accélération de dynamique moléculaire. Schéma du profil conformationnel en fonction de son énergie. La courbe pleine représente le profil énergétique de base et les courbes en pointiller les profils énergétiques, après accélération et représentée par la flèche rouge.



## 2. MATÉRIEL ET MÉTHODES

---

### 2.1 MATÉRIEL

#### 2.1.1 Préparation des systèmes

Cinq systèmes ont été modélisés : CCR5, CCR5-T82P, CCR5-P84A, CCR5/MVC et CCR5/CCL3 (Tableau 3 - 1). Le récepteur a été modélisé à partir de la structure cristallographique de l'entrée PDB 4MBS<sup>5</sup>. Cette structure comporte le récepteur lié à MVC. Un ion zinc, des molécules d'eau et de cholestérol sont également présentes. Le récepteur est cristallisé sous une forme tronquée et mutée, il contient les coordonnées des résidus 19 à 223 et 227 à 313 de la séquence du récepteur (Figure 3 - 4). Les résidus mutés sont : la tyrosine 58, l'asparagine 163, l'acide aspartique 233, et l'acide glutamique 303 qui sont respectivement des cystéine, glycine, alanine et lysine dans la séquence native (Uniprot : P51681, CCR5\_human). La protéine de fusion rubredoxine, utilisée pour faciliter la cristallisation, est insérée entre les résidus 223 et 227 à la place de la troisième boucle intracellulaire (ICL3).

Tableau 3 - 1 : Description des systèmes simulés.

Système	Ligand	Mutant
CCR5 libre	-	-
CCR5 ligand agoniste inverse	MVC	-
CCR5 ligand agoniste	CCL3	-
CCR5 mutant actif	-	T82P
CCR5 mutant inactif	-	P84A

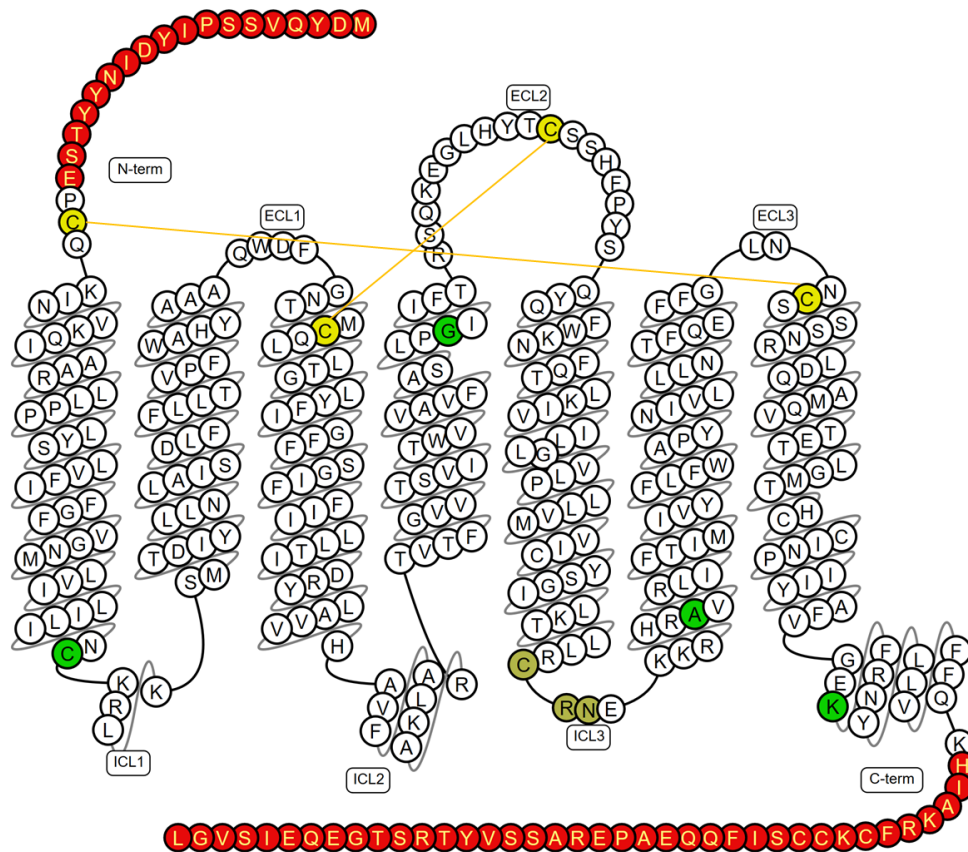


Figure 3 - 4 : Séquence de CCR5 dans l'entrée PDB 4MBS. En rouge, les résidus absents ou non résolus, en vert les résidus mutés, en brun les résidus remplacés par la rubredoxine et en jaune les deux ponts disulfures.

Le premier modèle de CCR5 a été préparé au laboratoire par le Professeur E. KELLENBERGER<sup>4,22</sup>. Dans ce modèle, l'ion et les lipides de la structure ont été retirés, les mutants modifiés en leurs résidus naturels et la rubredoxine supprimée et remplacée par les résidus manquants de la boucle ICL3. La partie N-terminale a été ajoutée en considérant la structure du peptide S7-Y15 en complexe avec la protéine virale gp120 et CD4 (identifiant PDB : 2RLL<sup>23</sup>). Les positions 10, 14 et 15 ont été modélisées par une sulfotyrosine. Le modèle décrit les résidus 7 à 313.

La partie N-terminale a été ajoutée par le Docteur Tim KNEHANS, post-doctorant au laboratoire (non publié). Tim KNEHANS a également construit le complexe CCR5-CCL3 à l'aide de Modeller version 9.14<sup>24</sup> en considérant la structure du récepteur CXCR4 en complexe avec la protéine virale vMIP-II (identifiant PDB : 4RWS<sup>25</sup>) et la structure de la chimiokine CCL3 (identifiant PDB : 5D65<sup>26</sup>). Les résidus 1-6, manquants dans le premier modèle du récepteur, ont été construits ab initio dans une conformation étendue. En pratique, la séquence de CCR5

(résidus 1 à 313) a été alignée sur le récepteur CXCR4, tandis que la chimiokine CCL3 (résidus 1 à 69) a été alignée sur la protéine virale à l'aide de la commande *align\_2d()*. Le modèle a été construit à partir de la ligne de commande *automodel()*. Durant la modélisation, des distances entre résidus ont été contraintes (Tableau 3 - 2) et des structures secondaires en hélice  $\alpha$  conservées pour les résidus 80 à 87 de CCR5 et 56 à 69 de CCL3. La structure obtenue a un score DOPE égal à 10039 kcal/mol, ne contient pas de collision atomique et est en accord avec le mode d'interaction prédit pas des expériences de liaison avec des protéines mutées (en particulier, Y187(5.31), S180, N258(6.58), Q280(7.36) et E283(7.39) dans CCR5 sont impliqués dans la liaison à CCL3<sup>7</sup>.

Tableau 3 - 2 : Contraintes de distance imposées durant la modélisation du complexe entre CCR5 et CCL3.

Atome : Numéro de résidu de CCL3	Atome : Numéro de résidu de CCR5	Distance (Å)	Déviations standard (Å)
N : 1	CD : 352	3.0	0.1
CG : 5	NZ : 260	3.0	0.1
OE1 : 33	OH : 253	3.0	0.1
CZ : 45	OH : 79	5.0	0.1
CZ : 47	OH : 83	5.0	0.1
CZ : 17	OH : 84	5.0	0.1
CA : 64	CA : 70	5.0	0.1

Une fois la structure modélisée, les quatre tyrosines 3, 10, 14 et 15 ont été mutées en sulfotyrosine et les deux paires de cystéines 20-269 et 101-178 pontées par deux ponts disulfures à l'aide du logiciel MOE 2017<sup>27</sup>. Les hydrogènes des résidus histidine 132 et 231 sont en position delta et les hydrogènes des résidus histidine 88, 175, 181 et 289 en position epsilon.

Les quatre autres systèmes ont été générés à partir de cette structure (Tableau 3 - 1). La pose du ligand MVC est calquée sur la structure cristallographique d'origine (4MBS). Les mutants T82P et P84A ont été obtenus à l'aide du logiciel MOE version 2017.

Dans les cinq systèmes, le récepteur libre ou lié a été inséré dans une bicouche lipidique au moyen du logiciel CHARMM-Gui<sup>28</sup>. Pour chaque couche de la membrane lipidique, un ratio de 2:2:1 de lipides palmitoyloleoylphosphatidylethanolamine (POPE), palmitoyloleoylphosphatidylcholine (POPC) et cholestérol (CHL) a été utilisé pour une boîte de

100 Å selon les axes X et Y. Une épaisseur de 20 est utilisée pour les couches d'eau dans lesquelles sont insérés des ions KCl à une concentration de 0.15 mol/L. Pour le complexe CCR5/CCL3 cette épaisseur a été augmentée à 25 Å. Les tailles des boîtes et les compositions des systèmes sont résumées dans le Tableau 3 - 3.

Tableau 3 - 3 : Composition des systèmes simulés.

Systèmes	Nombre de molécules					Nombre d'atomes	Dimension de la Boite (ÅxÅxÅ)
	POPC	POPE	CHL	Eau	Ions (K/Cl)		
<b>CCR5 libre</b>	124	124	62	30911	90/97	134812	102x102x139
<b>CCR5 ligand agoniste inverse</b>	124	124	62	30912	89/97	134893	102x102x139
<b>CCR5 ligand agoniste</b>	122	122	61	33499	100/105	143057	101x101x149
<b>CCR5 Mutant actif</b>	122	122	61	30413	89/96	132724	101x101x139
<b>CCR5 Mutant inactif</b>	122	122	61	30441	87/96	132796	101x101x139

Les cinq jeux de coordonnées ont ensuite été préparés à l'aide du module tleap du logiciel AMBER 14<sup>29</sup>. Différents champs de forces ont été utilisés pour paramétrer les systèmes : TIP3P pour l'eau, FF14SB et GAFF pour les protéines et ligands, lipid14 pour les lipides POPE et POPC et lipid11 pour le cholestérol.

### 2.1.2 Dynamiques moléculaires du récepteur CCR5

Les cinq systèmes ont tous été soumis au même protocole de dynamique moléculaire : une étape d'équilibrage suivie d'une étape de production par dynamique moléculaire classique, puis par une étape de production par dynamique moléculaire accélérée. Le programme utilisé est AMBER16 pour GPU (commande pmed.cuda), avec les paramètres par défaut sauf iwrap=1 et autres modifications mentionnées explicitement dans le texte ci-dessous.

#### 2.1.2.1 Équilibrage

Les structures de départ sont soumises à une étape de minimisation, dix étapes de montée en température et pression, et trois étapes de simulation sans contrainte. Pour chaque nouvelle étape, seules les coordonnées sont récupérées, de nouvelles vitesses étant générées.

La minimisation est effectuée sans contrainte, à volume constant, pour un maximum de 15000 cycles, avec un pas de calcul de 0.0001 ps. Au total 10000 cycles de minimisation par la

méthode de la plus forte pente sont suivis par 5000 cycles de minimisation par la méthode des gradients conjugués. Une limite de 10 Å pour le calcul des interactions non covalentes est utilisée, ainsi qu'une fréquence de sauvegarde des structures de 1/200. Les étapes de montée en température et pression ont été effectuées avec contraintes. Les différents paramètres et contraintes sont résumés dans le Tableau 3 - 4. Le thermostat de Langevin est utilisé pour le contrôle de la température en utilisant une graine aléatoire et une fréquence de collision ( $\gamma_{In}$ ) de 1.0 ps<sup>-1</sup>. Le baromètre de Berendsen est utilisé pour le contrôle de la pression avec un temps de relaxation de la pression de 2.0 ps.

Tableau 3 - 4 : Paramètres et contraintes pour les étapes de montée en température et pression de l'équilibrage.

Étape	Contrainte sur l'eau (kcal/mol <sup>2</sup> )	Contrainte sur les ions (kcal/mol <sup>2</sup> )	Contrainte sur les lipides (kcal/mol <sup>2</sup> )	Contrainte sur les atomes de protéines / ligands (kcal/mol <sup>2</sup> )	Nombre de cycles	Température (Kelvin)	Contrôle de la pression	Tolérance des coordonnées de l'algorithme Shake (Å)
1	2.5	10	2.5	10	25000	0 -> 100	X	
2	2.5	10	2.5	10	25000	100 -> 200	X	
3	2.5	10	2.5	10	25000	200 -> 300	X	
4	2.5		2.5	5	25000	300	O	
5	1.0		1.0	2.5	25000	300	O	
6	0.5		0.5	1.0	25000	300	O	0.0000001
7	0.1		0.1	1.0	25000	300	O	0.0000001
8				1.0	25000	300	O	0.0000001
9				0.5	25000	300	O	0.0000001
10					25000	300	O	0.0000001

À l'issue de l'étape 10, le système est simulé sans contraintes pendant 6ns pour lui permettre de se stabiliser. Les paramètres sont ceux de l'étape 10 à l'exception du pas de calcul mis à 0.002 ps, du nombre de cycle de mis à 1000000, d'une limite de 10 Å pour le calcul des interactions non covalentes et d'une valeur de 5 Å pour le paramètre *skinnb*.

### 2.1.2.2 Production par dynamique moléculaire classique

Pour chaque système cinq trajectoires de 60 ns sont effectuées, à l'exception du système CCR5 CCL3 pour lequel les trajectoires sont un peu plus courtes (50ns). Une limite de 10 Å pour le calcul des interactions non liantes est utilisée. Le thermostat de Langevin est utilisé pour une température de 300 K en utilisant une graine aléatoire et une fréquence de collision

(gamma\_In) de 1.0 ps<sup>-1</sup>. Le baromètre de Berendsen est utilisé pour le contrôle de la pression avec un temps de relaxation de la pression de 2.0 ps. Le pas de calcul est mis à 0.002 ps et la tolérance des coordonnées calculées par l'algorithme Shake est de 0.0000001 Å.

Pour la première étape de production, seules les coordonnées de l'étape finale de l'équilibrage sont récupérées et de nouvelles vitesses initiales sont utilisées. Pour les étapes suivantes, les coordonnées finales et la vitesse des étapes précédentes sont utilisées. Le pas de sauvegarde des structures est défini à 20 ps.

### 2.1.2.3 Production par dynamique moléculaire accélérée

Pour chaque système du récepteur CCR5, quatre trajectoires de 300 ns sont effectuées. Une limite de 10 Å pour le calcul des interactions non liantes est utilisée. Le thermostat de Langevin est utilisé pour une température 300 K en utilisant une graine aléatoire et une fréquence de collision (gamma\_In) de 1.0 ps<sup>-1</sup>. Le baromètre de Berendsen est utilisé pour le contrôle de la pression avec un temps de relaxation de la pression de 2.0 ps. Le pas de calcul entre chaque cycle est mis à 0.002 ps et la tolérance des coordonnées calculées par l'algorithme Shake est de 0.0000001 Å.

Pour la première étape de production, les coordonnées de l'étape finale de l'équilibrage sont utilisées et de nouvelles vitesses initiales sont utilisées. Pour les étapes suivantes, les coordonnées finales et la vitesse des étapes précédentes sont utilisées. Le pas de sauvegarde des structures est défini à 20 ps.

Un boost est appliqué à l'énergie potentielle totale et à l'énergie des dièdres en définissant le paramètre *iamd* à 3. Les paramètres de boost sont calculés selon les formules suivantes<sup>18,30</sup>:

$$E_{total} = V_{total\_avg} + 0.2 * N_{atoms}, \alpha_{total} = 0.2 * N_{atoms}$$

$$E_{dihed} = V_{dihed\_avg} + \varphi * V_{dihed\_avg}, \alpha_{dihed} = \frac{(\varphi * V_{dihed\_avg})}{5}$$

Avec  $N_{atoms}$  le nombre total d'atomes d'un système,  $V_{total\_avg}$  et  $V_{dihed\_avg}$  les valeurs moyennes de l'énergie potentielle totale et de celle des dièdres, calculées sur les trajectoires de dynamique classique des systèmes, et  $\varphi = 0.3$ <sup>19,31</sup>.

## 2.2 ANALYSES DES TRAJECTOIRES DE DYNAMIQUE MOLÉCULAIRE

Plusieurs analyses ont été effectuées afin de vérifier la stabilité du système, de caractériser différents marqueurs de l'activation des RCPGs et d'identifier des mouvements spécifiques aux systèmes étudiés. Les différentes analyses ont été effectuées au moyen du module *cpptraj*<sup>32</sup> du logiciel AMBER 16, du logiciel VMD<sup>33</sup> version 1.9.2 ou de la bibliothèque Bio3D<sup>34,35</sup> version 2.3 du logiciel R<sup>36</sup>. Ces trois outils prennent en entrée une trajectoire au format AMBER (ou Schrödinger) et des structures de protéines au format PDB.

### 2.2.1 Déviation par rapport aux coordonnées initiales

Le calcul de l'écart quadratique moyen (RMSD) des coordonnées atomiques au long de la dynamique a été effectué, avec l'option *rmsd* du module *cpptraj*. Les structures de la trajectoire ont été comparées à la structure de départ minimisée en considérant les atomes C, N, CA, O de la chaîne principale du domaine transmembranaire, tel que définis sur le site GPCRdb<sup>37</sup> et correspondant aux résidus : 22 à 58, 63 à 92, 97 à 132, 141 à 167, 186 à 224, 228 à 265 et 268 à 300 de CCR5.

### 2.2.2 Fluctuations atomiques

Le calcul de la fluctuation atomique est effectué avec l'option *rmsd perres* du module de *cpptraj*. La commande effectue un premier calcul de RMSD par rapport à la structure de référence, pour le meilleur ajustement de la chaîne principale (CA, C, N, O) des résidus du TM insérés dans la membrane (31 à 57, 64 à 88, 99 à 129, 143 à 164, 190 à 219, 235 à 256 et 277 à 300). Ce calcul est utilisé pour superposer les différentes structures. Un second calcul de RMSD est ensuite effectué, cette fois-ci sans alignement sur les coordonnées de la chaîne principale (CA, C, N, O) du récepteur.

### 2.2.3 Marqueurs structuraux de l'activation

Pour étudier l'état d'activation du récepteur, deux marqueurs des mouvements d'activation sont utilisés. Ces marqueurs sont la distance entre les hélices 3 et 6 et le RMSD de la zone « NPxxY » par rapport à une structure inactive. La distance entre les hélices 3 et 6 est mesurée en utilisant l'option *distance* du module *cpptraj*. Les atomes considérés sont les carbones  $\alpha$  des résidus acides aspartique 125 (3.50) et valine 234 (6.34). La déviation de la zone « NPxxY » (résidus 293 (7.49) à 297 (7.53)) est calculée au moyen du logiciel VMD avec l'extension d'analyse *RMSD Trajectory Tool*. La structure cristallographique 4MBS du récepteur est utilisée

comme référence et le RMSD est calculé, après alignement des atomes de la chaîne principale (C, CA et N) des résidus du motif NPxxY.

#### 2.2.4 Corrélations de mouvement

Les corrélations de mouvement sont recherchées en utilisant la bibliothèque Bio3d<sup>34</sup> du logiciel R. Pour cela, les structures sont alignées sur la première structure en considérant les atomes CA des résidus du TM insérés dans la membrane (31 à 57, 64 à 88, 99 à 129, 143 à 164, 190 à 219, 235 à 256 et 277 à 300), avec l'option en ligne de commande *fit.xyz*. Le calcul des coefficients de corrélation croisée est ensuite effectué pour chaque paire d'atome CA des 313 résidus de la protéine en utilisant la ligne de commande *dccm*. Cette ligne de commande permet d'obtenir une matrice de corrélation, utilisée pour générer une carte de corrélation des mouvements avec la ligne de commande *plot.dccm*.

#### 2.2.5 Caractérisation des mouvements

Trois analyses sont utilisées pour étudier les mouvements du récepteur dans les différents systèmes. Ces analyses sont : une analyse par composante principale (ACP) pour identifier les mouvements prépondérants durant la dynamique, la caractérisation des mouvements identifiés par ACP par des descripteurs structuraux et une analyse des structures secondaires.

##### 2.2.5.1 Analyse en composante principale

L'ACP identifie les mouvements prépondérants du récepteur au moyen du module *cpptraj*. Un premier calcul avec la ligne de commande *rmsd* est effectué pour aligner toutes les structures sur la première structure, en considérant les atomes de la chaîne principale (C, CA, N et O) des résidus du TM insérés dans la membrane (31 à 57, 64 à 88, 99 à 129, 143 à 164, 190 à 219, 235 à 256 et 277 à 300). Une structure moyenne est ensuite générée pour les atomes C, CA, N et O des résidus 22 à 313 en utilisant la commande *average*. La matrice de covariance est générée à partir de la commande *covar*. Cette matrice est utilisée par la commande *diagmatrix* pour générer les valeurs et les axes principaux de l'ACP. Les trois premiers axes de l'ACP sont sauvegardés dans un fichier ".nmd" grâce à l'option *nmwiz*. Ce fichier peut être lu par l'extension *Normal Mode Wizard* du logiciel VMD pour visualiser les 3 premiers axes.

##### 2.2.5.2 Analyse structurale

Les mouvements identifiés par l'analyse par ACP sont caractérisés en associant des mesures de distance et d'angle à un calcul d'enthalpie libre. Cette analyse est effectuée au moyen de



la commande *hist* et de l'option *free* du module *cpptraj*. L'enthalpie libre est calculée en considérant 200 intervalles, pour une température de 300 K avec la formule suivante :

$$G_i = -k_B T * \frac{N_i}{N_{Max}}$$

Avec  $k_B$ , la constante de Boltzmann,  $N_i$  la population d'un intervalle  $i$  et  $N_{max}$  la population de l'intervalle le plus peuplé. Les descripteurs structuraux consistent en des distances entre CA de résidus particuliers et des angles entre vecteurs. Le logiciel *cpptraj* a été utilisé pour calculer les distances avec la ligne de commande *distance*. La commande *vectormath* et l'option *dotangle* ont été utilisées pour calculer les angles entre vecteurs. Les différents vecteurs nécessaires au calcul d'angle ont été générés avec la commande *vector* entre les CA.

### 2.2.5.3 Structures secondaires

Les structures secondaires ont été déterminées avec la commande *secstruct* du module *cpptraj*. Cette commande utilise l'algorithme « Define Secondary Structure of Proteins »<sup>38</sup>.

## 2.3 IDENTIFICATION DE MOTIFS D'INTERACTIONS INTRAMOLÉCULAIRES

Les motifs d'interactions intramoléculeaires sont identifiés à l'aide de la méthode décrite au chapitre 1. En résumé, les lipides et les ions sont retirés des structures, les liaisons hydrogène sauf  $CO_{(i)}-NH_{(i+3/4)}$ , les interactions ioniques et les interactions aromatiques sont identifiées par le module *INTS* du logiciel *IChem*<sup>39</sup> (version 5.2.9) et stockées dans une matrice. Cette matrice carrée dissymétrique, sauvegardée au format *Numpy*, répartit les interactions dans le triangle inférieur ou supérieur selon leur nature (Figure 3 - 5). À partir de ces matrices, une représentation en 2D des interactions peut être générée pour un ensemble de structures. Cette représentation est effectuée en calculant la proportion de présence des différentes interactions au sein des structures.

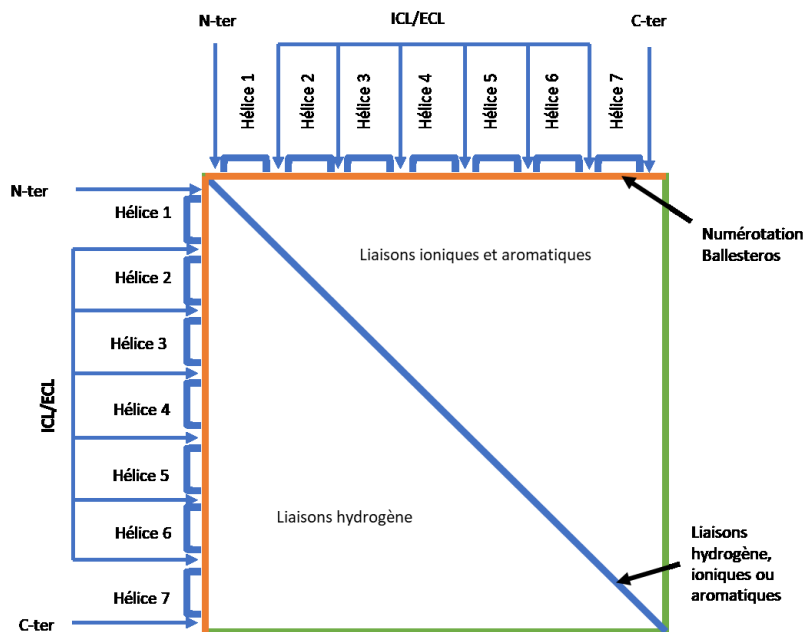


Figure 3 - 5 : Représentation schématique d'une matrice Numpy regroupant les interactions polaires, excluant les interactions  $CO_{(i)}-NH_{(i+3/4)}$  de la principale, d'une structure de récepteur couplé au protéines G.

La similitude entre deux structures est mesurée en considérant les interactions du domaine transmembranaire (22-58, 63-92, 97-132, 141-167, 186-224, 228-265 et 268-300) et utilise le calcul de similitude par indice de Tanimoto. Le partitionnement des structures par similitude de motifs d'interactions est effectué au moyen du logiciel R-cran par clustering hiérarchique ascendant en utilisant la méthode de "ward.D2"<sup>40,41</sup>. Le seuil de coupure a été ajusté pour maximiser le nombre de partitions ayant un score de similitude moyen supérieur à 1. Une partition est considérée comme homogène et donc contenant un motif pertinent, si la valeur de premier quartile de sa distribution des similitudes est supérieure à la valeur du 3<sup>ème</sup> quartile de la distribution des similitudes inter-partitions (Figure 3 - 6).

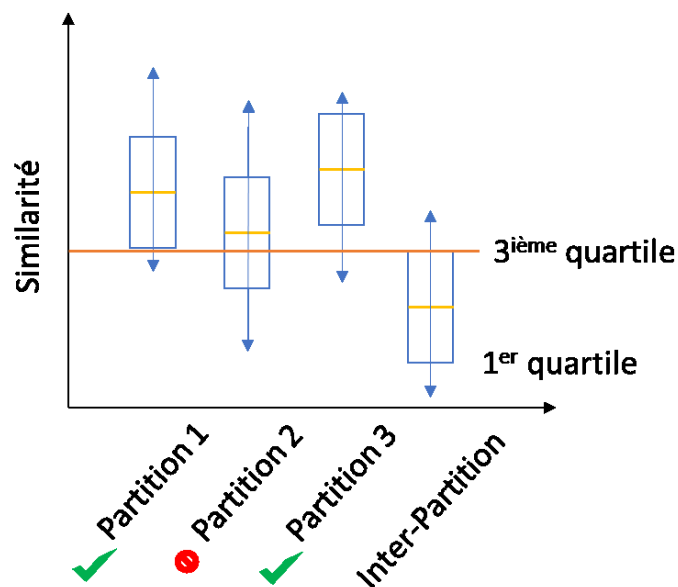


Figure 3 - 6 : Mode de selection des partitions pour l'identification de motifs d'interactions intra-TM.

## 3. RÉSULTATS

### 3.1 DYNAMIQUE MOLÉCULAIRE CLASSIQUE

#### 3.1.1 Déviation par rapport aux coordonnées initiales

Les trajectoires de dynamique moléculaire classique ont été réalisées afin de préparer les calculs par dynamique moléculaire accélérée. Dans les cinq systèmes étudiés, le RMSD est calculé après alignement du domaine transmembranaire à cause de la grande flexibilité du domaine N-terminal et permet de vérifier la stabilité du système. Dans toutes les trajectoires (Figure 3 - 7), on peut observer une rapide augmentation de la valeur de RMSD. Cette augmentation est normale et est due à l'étape d'équilibrage du système. Elle est suivie d'une stabilisation du RMSD dans tous les systèmes. Cependant CCR5/MVC et CCR5-T82P semblent subir assez rapidement des modifications du domaine transmembranaire dans certaines répétitions.

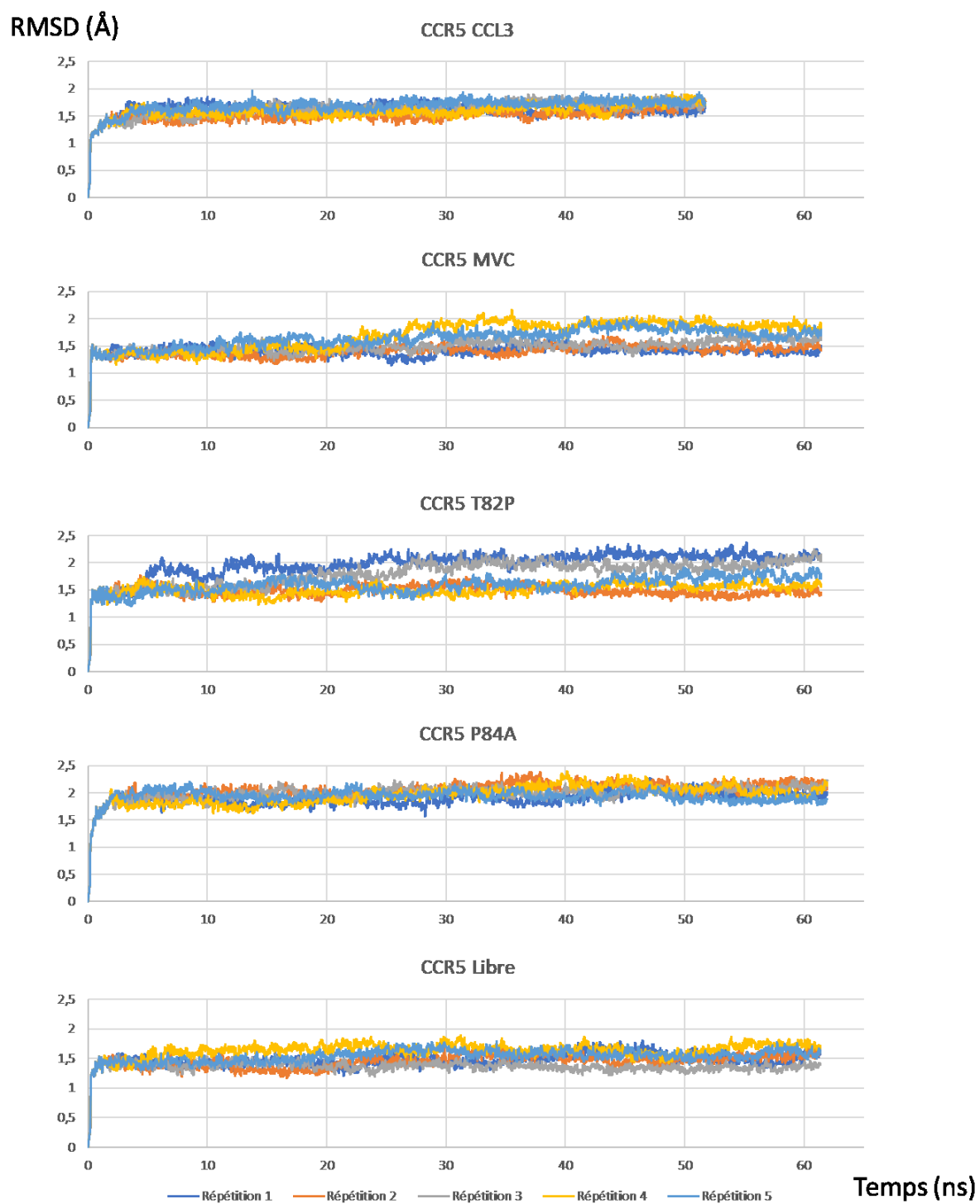


Figure 3 - 7 : RMSD calculés sur les coordonnées des atomes C,N, CA, O du domaine transmembranaire, par rapport à la structure minimisée de départ.

### 3.1.2 Paramètres d'accélération

Pour chaque système, les valeurs moyennes d'énergie potentielle et d'énergie des dièdres ont été calculées sur les phases de production des trajectoires et sont regroupées dans le tableau 5. Les différentes trajectoires de dynamique accélérée ont ensuite été lancées avec les paramètres théoriques calculés à partir des valeurs d'énergie (Tableau 3 - 1). Parmi tous les systèmes, les trajectoires 2, 3 et 4 du récepteur CCR5 T82P ont été initiées avec des valeurs modifiées du fait de boost d'énergie trop important induisant l'échec des simulations. Les valeurs modifiées ont été ajustées par essai/erreur.

Tableau 3 - 5 : Energie potentielle moyenne, énergie moyenne des dièdres (kcal/mol) et paramètres d'accélération .

Systèmes	$V_{dihed\_avg}$	$V_{total\_avg}$	$N_{atoms}$	$E_{dihed}$	$\alpha_{dihed}$	$E_{total}$	$\alpha_{total}$
CCR5 libre	17704.6093	-331053.8987	134812	23015.99209	1062.276558	-304091.4987	26962.4
CCR5/CCL3	18390.10416	-358833.1854	143057	23907.13541	1103.406249	-330221.7854	28611.4
CCR5/MVC	17718.5081	-330965.6172	134893	23034.06053	1063.110486	-303987.0172	26978.6
CCR5-P84A	17464.33669	-326237.9518	132798	22703.63769	1047.860201	-299678.3518	26559.6
CCR5-T82P Répétition 1	17483.36257	-325766.3302	132724	22728.37133	1049.001754	-299221.5302	26544.8
CCR5-T82P Répétitions 2, 3 et 4	17483.36257	-325766.3302	132724	22528.37133	1024.001754	-300221.5302	26444.8

## 3.2 DYNAMIQUE MOLÉCULAIRE ACCÉLÉRÉE

L'analyse des trajectoires de dynamique moléculaire accélérée est centrée sur la recherche des spécificités de chacun des systèmes. Les différentes analyses présentées dans ce paragraphe sont utilisées pour caractériser ces spécificités et leur pertinence.

### 3.2.1 Déviation par rapport aux coordonnées initiales

Pour les cinq systèmes on peut noter deux séries d'augmentation rapide du RMSD. La première augmentation est due à la phase d'équilibrage du système, comme pour les trajectoires de dynamiques classiques. La seconde phase quant à elle est due à l'utilisation du boost d'énergie de la dynamique accélérée. Pour ne pas prendre en compte cette seconde phase et intégrer des structures en transition, les 20 premières ns de chaque dynamique n'ont pas été prises en compte pour les analyses. Les cinq systèmes sont stables dans toutes les répétitions (Figure 3 - 8). On peut néanmoins remarquer que les répétitions 4 des systèmes libres et liés à MVC ont des profils différents. Les RMSD indiquent des changements dans le domaine transmembranaire. Dans le cas de la dynamique du système MVC, ce changement

apparaît rapidement et la structure résultante est stable. Dans le cas du système libre ce changement est initié à plusieurs reprises sans atteindre une conformation stable.

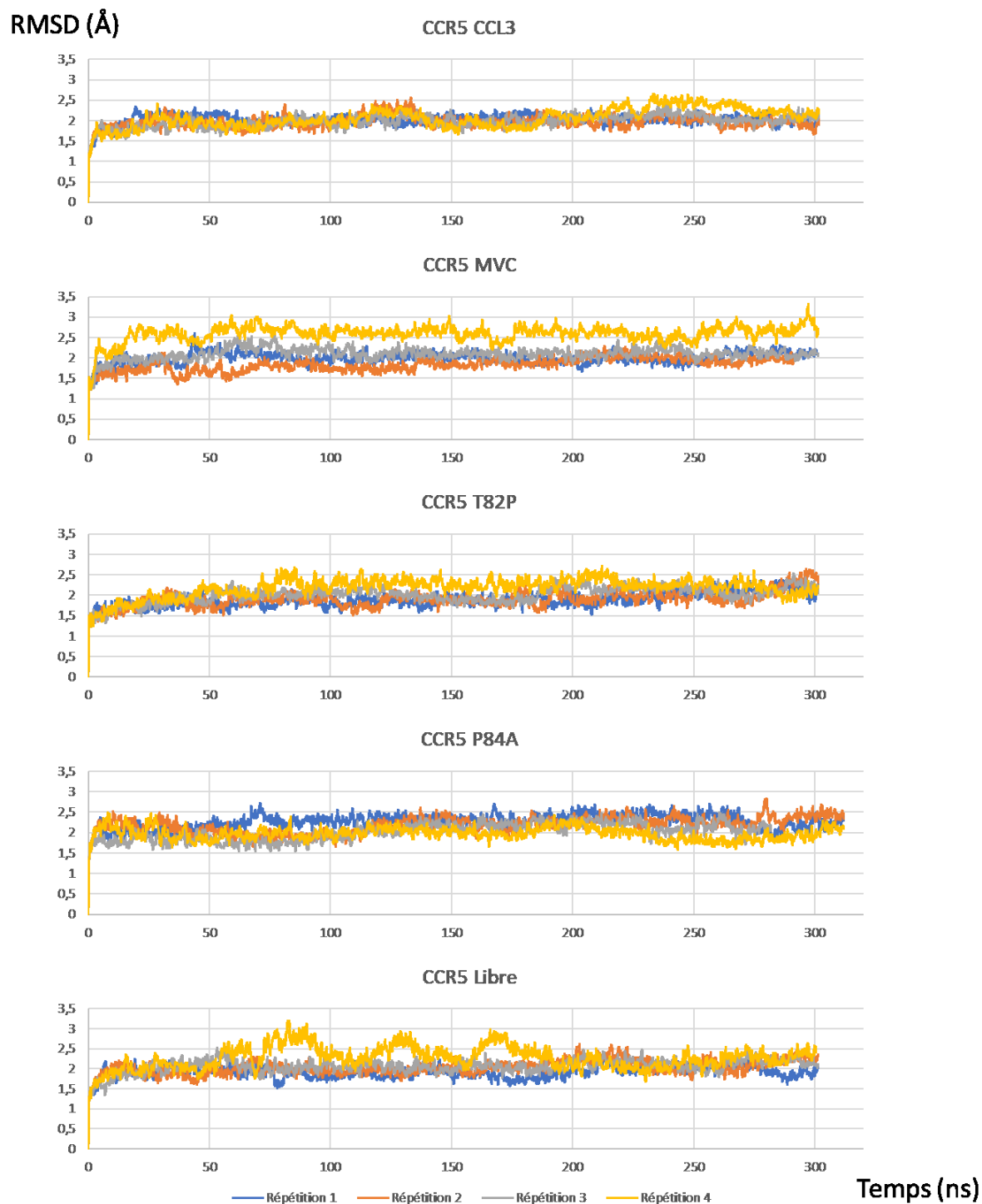


Figure 3 - 8 : RMSD des coordonnées des atomes de la chaîne principale du domaine transmembranaire, par rapport à la structure minimisée.

### 3.2.2 Fluctuation atomique

Les systèmes étant stables dans les différentes simulations, leur dynamique peut être étudiée. L'analyse de la fluctuation par résidu du récepteur tout au long de la dynamique permet de mettre en évidence les domaines les plus flexibles, qui sont la partie N-terminale, les boucles extracellulaires et intracellulaires et l'hélice 8. Cette variabilité conformationnelle est due au fait que ces domaines du récepteur ne sont pas insérés dans la membrane. Si ces domaines sont tous flexibles, certains subissent des fluctuations plus importantes que d'autres, notamment la partie N-terminale ainsi que l'hélice 8 et les boucles ECL2, ICL3 et ECL3 qui ont des RMSD supérieurs à 4 Å (Figure 3 - 9). À titre de comparaison les RMSD du domaine transmembranaire du récepteur est proche de 1 Å.

En observant les fluctuations atomiques, on peut observer des variations entre les systèmes (Figure 3 - 10) au niveau de l'hélice 5 et de bas de l'hélice 6 pour le domaine transmembranaire, et les boucles ICL2, ICL3, ECL3 ainsi que la partie N-terminale pour le domaine extracellulaire et intracellulaire. S'il est utile de noter ces différences de fluctuations, elles ne peuvent pas être considérées comme significatives d'après les moyennes et les écarts-types (Tableau 3 - 6).

Tableau 3 - 6 : Moyenne et écart type sur les quatre répétitions des fluctuations par résidus (Å).

	Libre		CCL3		MVC		T82P		P84A	
	Moyenne	Ecart type	Moyenne	Ecart type	Moyenne	Ecart type	Moyenne	Ecart type	Moyenne	Ecart type
N-ter	10.75	3.69	9.38	2.99	10.39	2.99	11.16	3.04	12.29	3.05
TM1	1.75	0.56	1.48	0.52	1.55	0.58	1.39	0.50	1.74	0.58
ICL1	1.71	0.59	1.52	0.54	1.61	0.72	1.62	0.65	2.72	0.91
TM2	0.97	0.35	1.08	0.37	0.80	0.30	0.83	0.31	0.94	0.33
ECL1	2.56	0.76	2.43	0.78	1.93	0.89	1.81	0.72	2.45	0.91
TM3	1.10	0.39	0.96	0.33	1.05	0.37	0.93	0.33	1.07	0.37
ICL2	3.29	1.34	2.25	0.84	2.25	0.87	2.53	0.91	2.40	0.84
TM4	1.16	0.45	1.26	0.42	1.13	0.44	1.06	0.39	1.18	0.44
ECL2	2.67	1.04	2.52	0.87	2.23	0.92	2.21	0.87	2.43	0.97
TM5	1.88	0.66	1.50	0.53	1.36	0.53	1.28	0.51	1.51	0.60
ICL3	4.88	1.71	5.03	1.52	4.71	1.60	5.05	1.61	3.58	1.30
TM6	1.97	0.77	1.96	0.65	1.57	0.57	1.93	0.69	2.07	0.74
ECL3	3.69	1.64	3.49	1.24	3.89	1.41	4.20	1.36	4.41	1.46
TM7	1.35	0.52	1.30	0.45	1.37	0.49	1.32	0.48	1.53	0.56
TM8	2.53	0.99	2.99	1.23	2.51	0.98	3.12	1.13	2.63	1.25

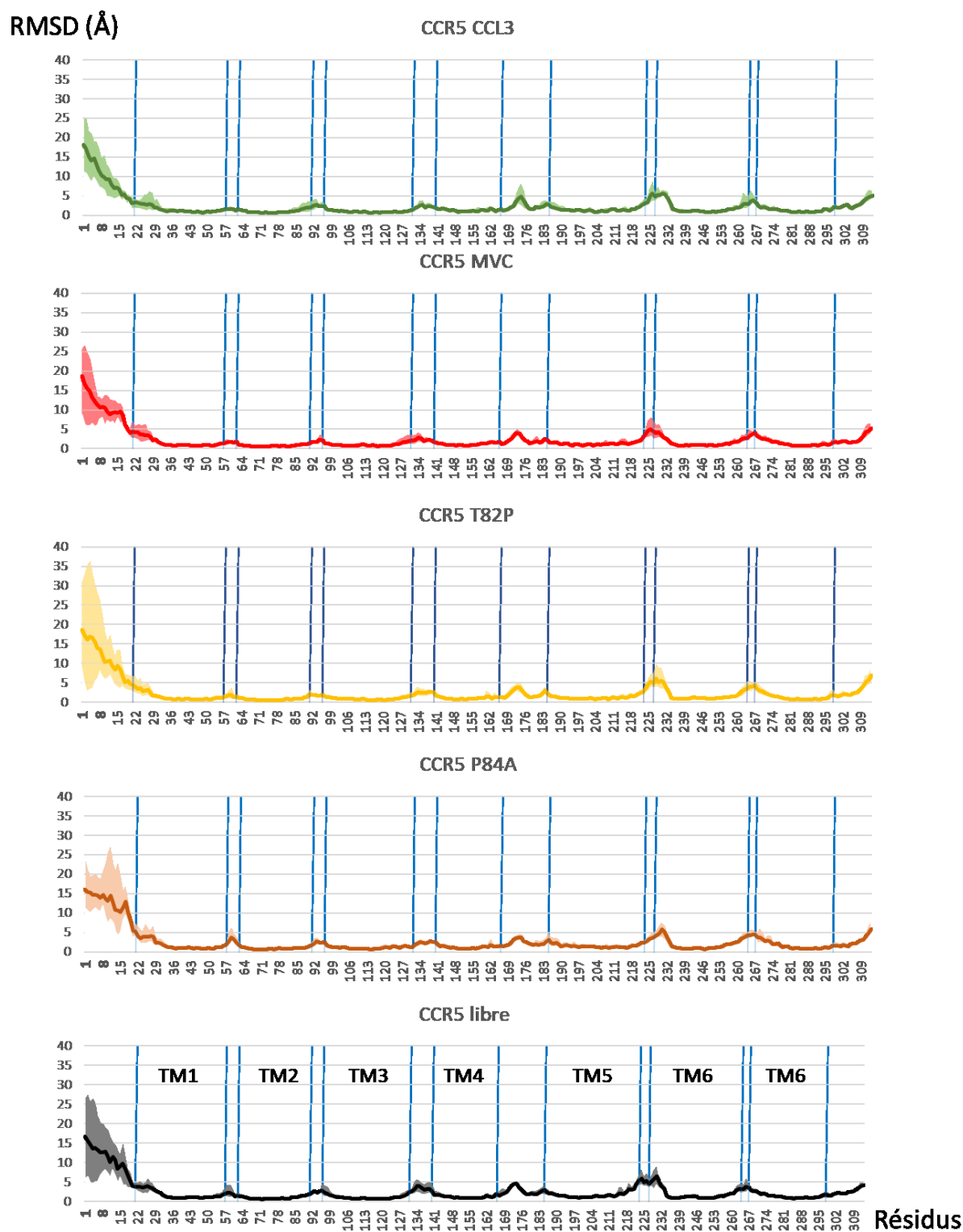


Figure 3 - 9 : Fluctuation moyennes (courbe), minimales et maximales (ombre).



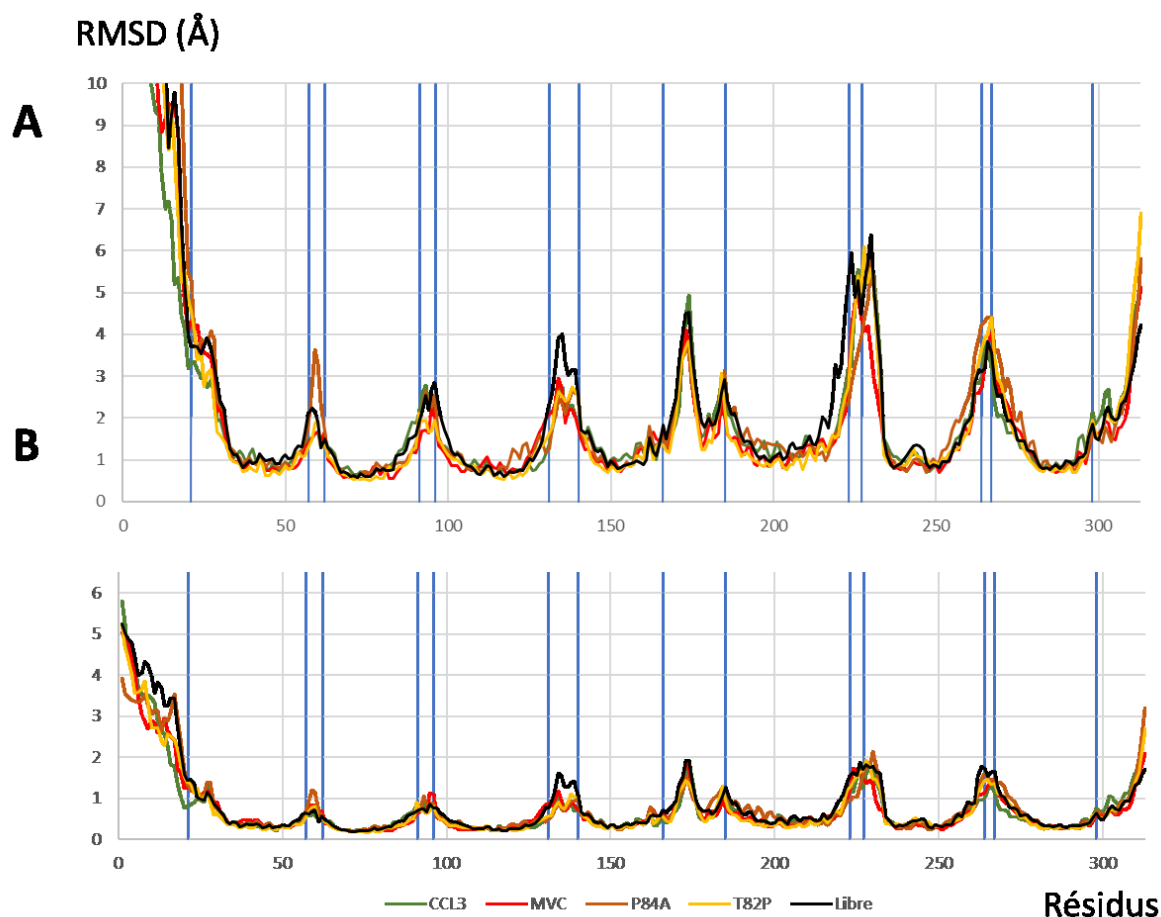


Figure 3 - 10 : Fluctuation atomique moyenne (A) et écart-type (B).

### 3.2.3 Marqueurs structuraux de l'activation

Les différents systèmes de CCR5 ont été construits dans le but théorique de simuler les prémices d'activation pour les systèmes CCL3 et T82P, la variabilité des états inactivés pour les systèmes MVC et P84A et de les comparer aux états explorés par le récepteur libre. Ces systèmes ayant tous été construits à partir d'une structure inactive, il est essentiel de vérifier par des marqueurs caractéristiques de l'activation, si un mouvement d'activation est observable. Cette évaluation est effectuée par la distance entre les hélices 3-6 et le RMSD du motif NPxxY par rapport à une structure inactive, et est présentée dans la Figure 3 - 11. On peut observer que le récepteur reste à l'état inactif dans tous les systèmes. Dans le cas des systèmes CCL3 et T82P on peut supposer qu'un changement d'état est initié du fait d'un échantillonnage légèrement plus important de structures ayant un RMSD du motif NPxxY supérieur à 0,3 Å. L'initiation d'un changement d'état est aussi visible dans le cas du système WT, cette fois-ci avec des distances échantillonnées entre hélice 3 et 6 supérieures à 8 Å.

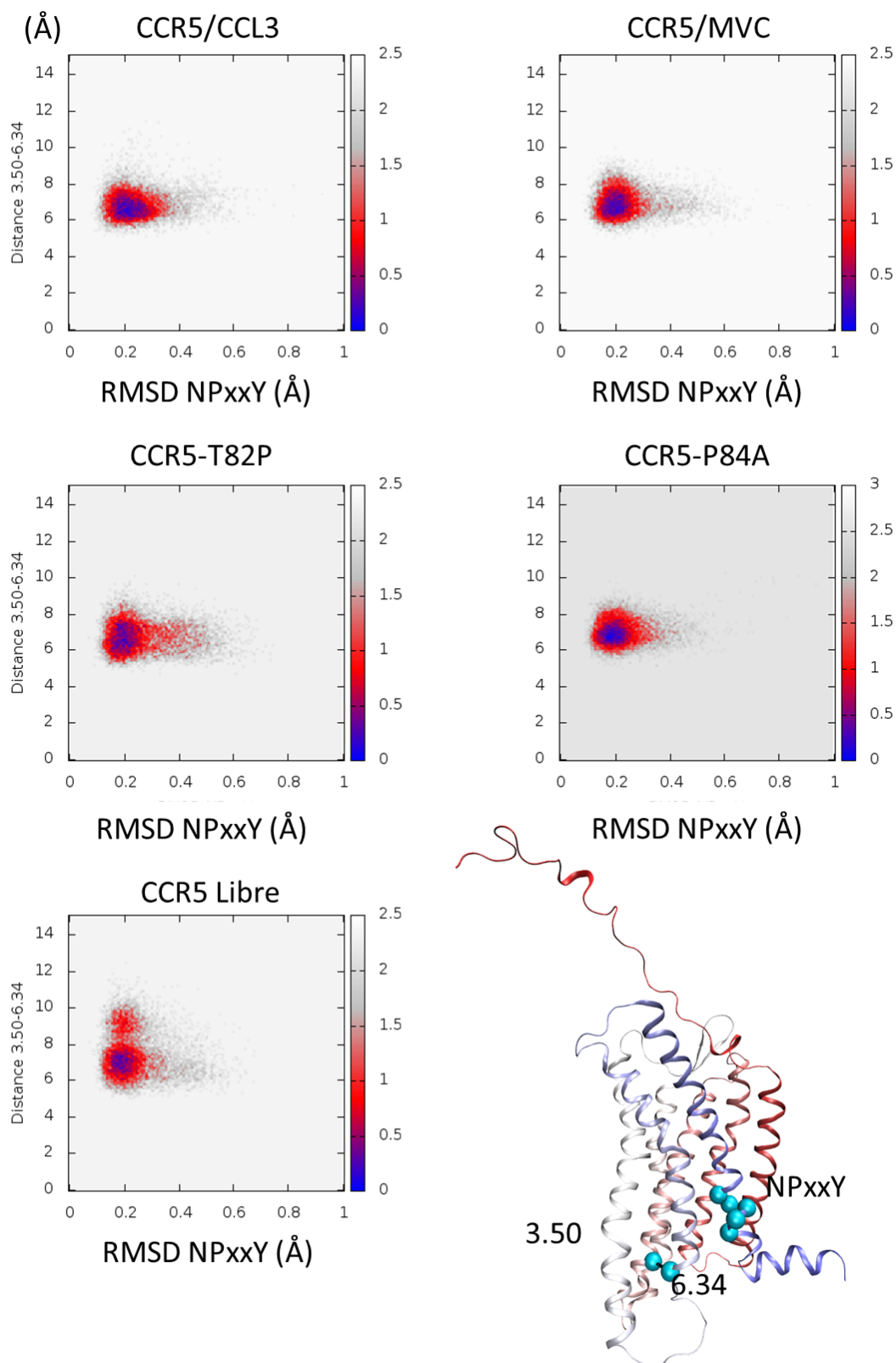


Figure 3 - 11 : Caractérisation de l'état d'activation du récepteur à partir des distances entre l'hélice 3 (atomes CA du résidu 3.50) et l'hélice 6 (CA du résidu 6.34) et du RMSD du motif NPxxY (atomes CA,C,N des résidus 7.49 à 7.53). L'image en bas à droite montre la position des CA des résidus 3.50 et 6.34 et du motif NPxxY dans CCR5, représenté par un ruban coloré par numéro de résidu.

### 3.2.4 Corrélations de mouvement

L'analyse de la corrélation croisée entre les résidus du récepteur a été effectuée en combinant les différentes répétitions des systèmes. Les cartes de corrélations montrent des similitudes entre les systèmes avec des corrélations entre les hélices adjacentes. Les cartes montrent également de nombreuses différences dans les corrélations/anti-corrélations faibles (inférieures à 0,25/ supérieures à -0,25) (Annexe 4). En se concentrant sur les corrélations/anti-corrélations fortes (supérieures à 0,5/ inférieures à -0,5) (Figure 3 - 12), les différences et similitudes entre systèmes sont plus visibles (Tableau 3 - 7). Globalement, les systèmes liés présentent plus de corrélations entre les domaines de CCR5 (9 couplages et 10 couplages), en comparaison aux systèmes libre (7 couplages pour chacun). On peut séparer ces couplages entre domaines en trois catégories : les couplages inter-TM, les couplages TM-boucle et les couplages boucle-boucle. Pour ces derniers, on peut remarquer que seul deux couplages de boucles consécutives sont observables pour les systèmes CCL3 et mutés.

Parmi les couplages inter-TM, certains impliquent des domaines proches dans l'espace du fait de l'architecture à sept hélices antiparallèles des RCPGs (i.e. TM1-TM2, TM2-TM3, TM3-TM4, TM4-TM5, TM5-TM6, TM6-TM7 et TM7-TM1). Étonnamment, aucun mouvement corrélé n'est observé avec le TM5, à l'exception d'un couplage atypique entre le TM1 et TM5 dans le système CCL3. De plus, on peut remarquer l'absence de couplage avec le TM 1 pour le système CCL3 et l'absence de couplage avec le TM2 dans les deux mutants du motif TxP. Les systèmes MVC et libre sont également caractérisés par deux couplages atypiques entre l'hélice 8 et respectivement le TM1 et le TM6.

Parmi les couplages TM-boucles, seul un mouvement de corrélation entre le TM3 et ECL2 est observé dans tous les systèmes. Cette corrélation est attendue à cause de la présence du pont disulfure entre les résidus C101(3.25) et C178. Étonnamment, aucun couplage entre la chaîne N-terminale et le TM7, porteur du pont disulfure entre les résidus C20 et C269(7.25), n'est observé. Seul un mouvement entre la chaîne N-terminale et le TM6 proche est observé dans le système MVC. On pourra remarquer que le système MVC effectue un autre couplage TM-boucle atypique entre le TM1 et ICL3 et que le système libre n'est caractérisé par aucun couplage de boucle à l'exception de TM3 et ECL2.

Globalement, les mouvements de corrélation et d'anti-corrélation de CCR5 sont caractéristiques de quatre types de systèmes : les systèmes liés à un agoniste (CCL3), lié à un agoniste inverse (MVC), libre et les systèmes mutés qui présentent les mêmes couplages sans différenciation de l'activité constitutive.

Tableau 3 - 7 : Corrélations/anti-corrélations observables entre différents domaines du récepteur selon les systèmes. Les cases rouges correspondent à des mouvements corrélés et les cases bleu à des mouvements anti-corrélés.

Domaines	Systèmes CCR5				
	CCL3	MVC	T82P	P84A	Libre
TM1-TM2					
TM1-TM7					
TM2-TM3					
TM2-TM4					
TM6-TM7					
TM1-TM5					
TM1-H8					
TM6-H8					
TM1-ECL1					
TM2-ECL2					
TM3-ECL2					
TM5-ICL2					
TM1-ICL3					
TM6-Nterminale					
ECL1-ECL2					
ICL2-ICL3					

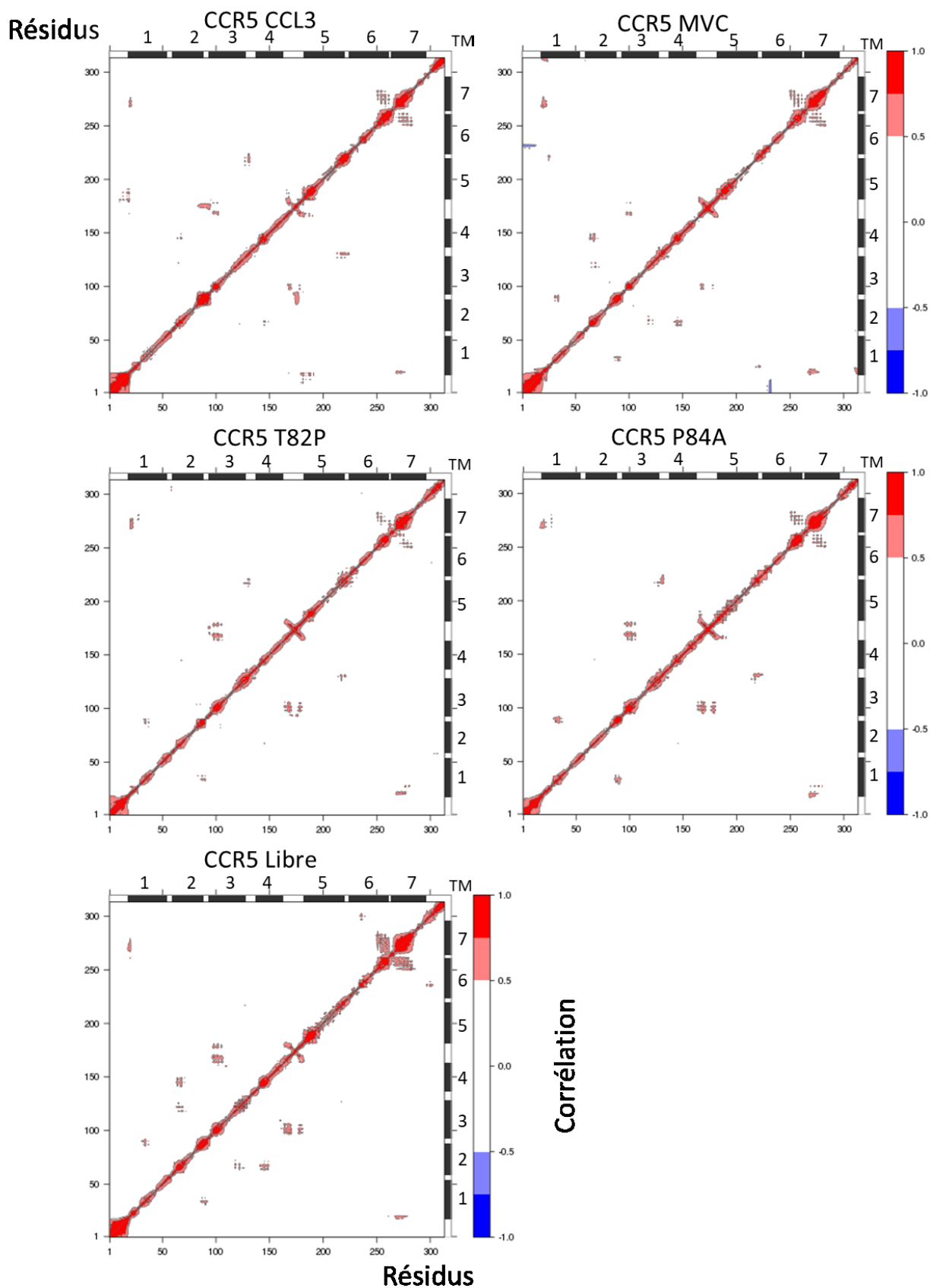


Figure 3 - 12 : Cartes de corrélation des mouvements dans CCR5. Les barres noires annotées de 1 à 7 définissent les hélices transmembranaires du récepteur. Les valeurs de corrélations des mouvements des résidus vont de 1 (couleur rouge), pour des mouvements corrélés, à -1 (couleur bleue) pour des mouvements anti-corrélés.

### 3.2.5 Mouvements du récepteur CCR5

Les différentes analyses effectuées jusqu'ici montrent des domaines de forte flexibilité et dont les mouvements sont corrélés. Pour caractériser les mouvements trois analyses sont utilisées. L'analyse par composante principale permet d'identifier les mouvements prépondérants. La seconde analyse utilise des caractéristiques physiques telle que des distances ou des angles entre atomes. La dernière analyse est utilisée pour vérifier les structures secondaires du récepteur pour la partie N-terminale - hélice 1 et la boucle ECL3. Toutes les analyses des mouvements ont été effectuées en combinant les différentes répétitions pour chacun des cinq systèmes.

#### 3.2.5.1 Identification des mouvements

L'analyse en composante principale permet de « déstructurer » les mouvements observables dans les trajectoires par l'analyse de la covariance entre les résidus. Elle permet de définir des vecteurs de mouvements et leur contribution à la dynamique. Cette analyse est biaisée par les mouvements de grande amplitude, c'est pourquoi la partie N-terminale du récepteur n'a pas été prise en compte. Pour chacun des cinq systèmes, les trois premières composantes ont été analysées.

Les trois premières composantes de l'ACP sont caractéristiques de moins de 40% de la simulation (Tableau 3 - 8). Seules celles du système CCR5-MVC sont caractéristiques de 50% des trajectoires. Les deux premières composantes apportant à chaque fois la plus grande contribution tandis que la troisième et les composantes suivantes varient entre 9 et 0 %.

Tableau 3 - 8 : Contribution aux trajectoires de dynamique des trois premières ACP des systèmes du récepteur CCR5.

Système	ACP 1	ACP 2	ACP 3	Somme des ACP 1-2-3
CCL3	0.18485286	0.13190461	0.06946036	0.38621783
MVC	0.23569649	0.18377817	0.08922015	0.5086948
T82P	0.20034173	0.10746379	0.08072553	0.38853105
P84A	0.15543268	0.13694536	0.08621017	0.37858822
Libre	0.1713075	0.10876221	0.07860862	0.35867833

Les descriptions spécifiques des trois premières composantes des systèmes sont décrites en annexe (Système CCL3 : Annexe 5, Système MVC : Annexe 6, Système T82P : Annexe 7, Système P84A : Annexe 8, Système WT : Annexe 9). Globalement, les différentes ACP des systèmes du récepteur CCR5 montrent très peu de mouvements au sein des TM à l'exception

de l'hélice 8, qui semble avoir des mouvements différents dans chacun des systèmes, et du haut des TM5 et TM6 dans le système libre. Au sein des régions flexibles, des mouvements de grande amplitude sont observés pour les domaines intracellulaires et extracellulaires, notamment pour les boucles ICL3, ECL2 et ECL3. Pour la boucle ICL3, les systèmes mutés et libre présentent des mouvements vers l'intérieur du récepteur, les systèmes CCL3 et libre des mouvements vers l'extérieur et le système MVC est caractérisé par un mouvement de cisaillement. Pour la boucle ECL2 les systèmes CCL3 et P84A présentent un mouvement de va et vient. Pour la boucle ECL3, les systèmes CCL3, T82P et P84A présentent tous des mouvements différents.

### 3.2.5.2 Caractérisation des mouvements

Si des mouvements sont identifiés par l'ACP des différentes trajectoires, il est nécessaire de vérifier leur nature dans les différents systèmes. Pour cela, les conformations associées aux différents mouvements peuvent être mises en évidence en associant des caractéristiques structurales et l'enthalpie libre calculée. Les mouvements étudiés ici sont ceux des boucles ECL2, ICL3 et ECL3 d'un côté et les mouvements de la partie N-terminale et de l'hélice 8 de l'autre. Les mouvements des boucles ont été caractérisés par des distances entre un de leurs résidus les plus fluctuants et des résidus dans la partie rigide du récepteur (Figure 3 - 13). Les mouvements de la partie N-terminale et de l'hélice 8 ont été caractérisés par des angles entre des vecteurs définis à partir des CA de trois résidus (Figure 3 - 13).

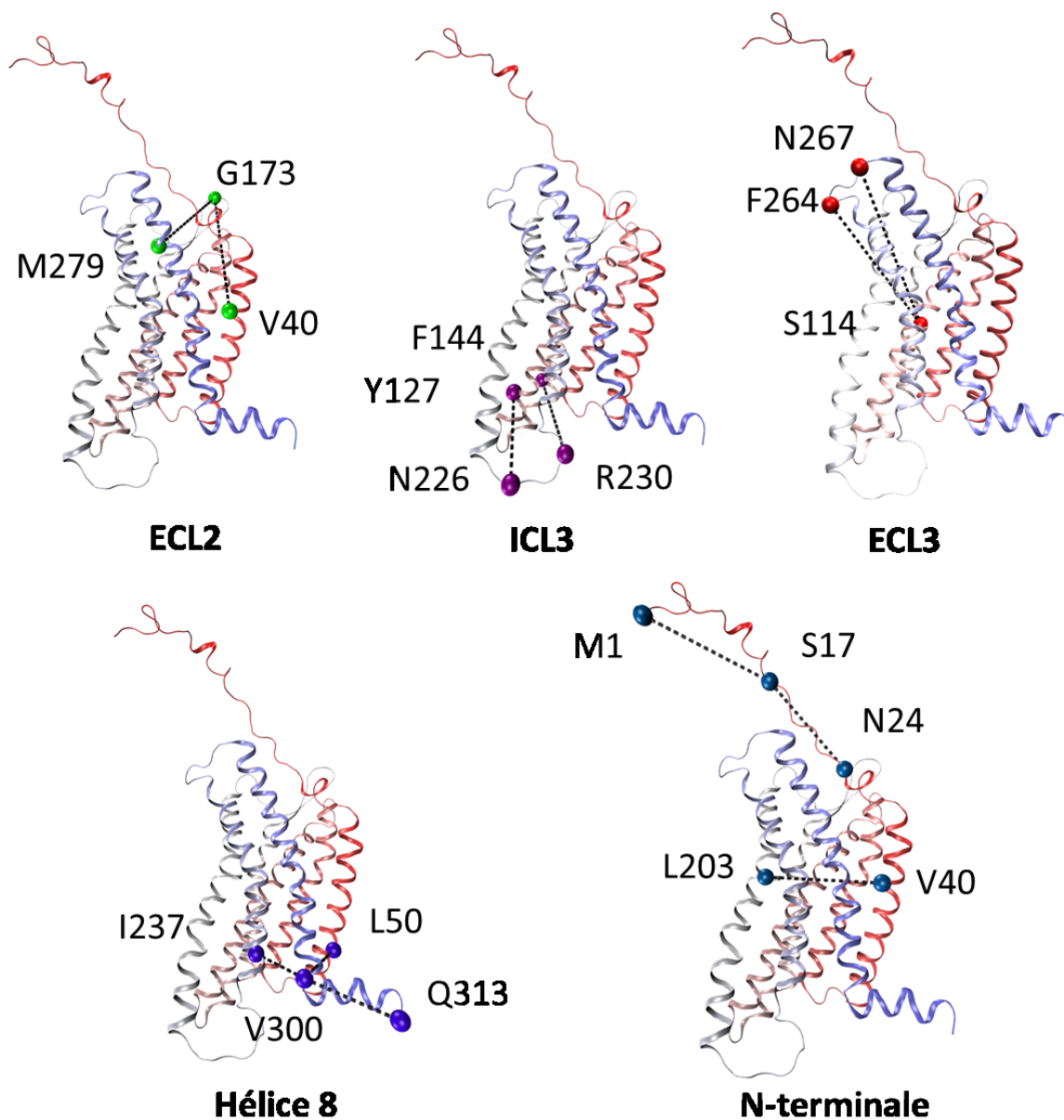


Figure 3 - 13 : Résidus utilisées pour définir les distances et les angles entre vecteurs caractéristiques.

Les mouvements ECL2 explorent un ensemble de conformations similaires dans tous les systèmes (Figure 3 - 14, p100). Selon les systèmes, des conformations particulières sont explorées de manière plus importante (zone bleue sur les images). Ces conformations préférentielles sont caractérisées par un minima d'enthalpie libre (Tableau 3 - 9). Seul le système MVC ne semble pas explorer de minima particulier.



Tableau 3 - 9 : Distances caractéristiques des minimas enthalpie libre des différentes boucles.

Boucle	Distance CA-CA	CCL3 (Å)	MVC (Å)	T82P (Å)	P84A (Å)	Libre (Å)
<b>ECL2</b>	M279-G173	22	16	16 et 18	17 et 22	20
	V40-G173	33	28	28 et 29	28 et 31	28
<b>ICL3</b>	F144-R230	25	16 et 24	25 et 27	25	29
	Y127-N226	17	16 et 17	15 et 20	13	19
<b>ECL3</b>	S114-F264	33	33	33	33	32
	S114-N267	38	37	33	35	37

Les mouvements de la boucle ICL3 explorent quant à eux un ensemble de conformations relativement étendues (Figure 3 - 15, p101). Les systèmes Libre/CCL3/T82P d'un côté et MVC/P84A de l'autre, explorent des conformations globalement similaires. Cependant, les différentes conformations préférentielles sont différentes dans les cinq systèmes (Tableau 3 - 9).

Les mouvements explorés par la boucle ECL3 et l'hélices 8 sont caractérisés par un ensemble restreint de conformations qui sont similaires dans les cinq systèmes (Figure 3 - 16 et Figure 3 - 17, p102-103). Les minimas pour la boucle (Tableau 3 - 9) et l'hélice (Tableau 3 - 10) sont proches dans les différents systèmes. On peut néanmoins noter une conformation spécifique de l'hélice 8 dans le système P84A, mais également pour ECL3 une tendance à explorer des mouvements similaires dans le cas des système mutés d'un côté et liés à un ligand de l'autre.

Tableau 3 - 10 : Angles caractéristiques des minima d'énergie libre de la partie N-terminale et de l'hélice 8.

	Vecteur 1 (CA-CA)	Vecteur 2 (CA-CA)	CCL3 (Degré)	MVC (Degré)	T82P (Degré)	P84A (Degré)	Libre (Degré)
<b>N-terminal</b>	S17-M1	S17-N24	33	113	81	49	40
	17-M1	V40-L203	127	26	75	143	85
<b>Hélice 8</b>	V300-I237	V300-L50	126	124	121	145	125
	V300-L50	V300-Q313	48	48	45	62	45

Pour finir, les mouvements explorés par l'extrémité N-terminale sont caractérisés par un ensemble varié de conformations (Figure 3 - 18, p104). Malgré la présence de minima (Tableau 3 - 10), ces conformations ne peuvent être considérées comme caractéristiques des différents systèmes du fait de la non reproductibilité des mouvements entre les répétitions d'un système.

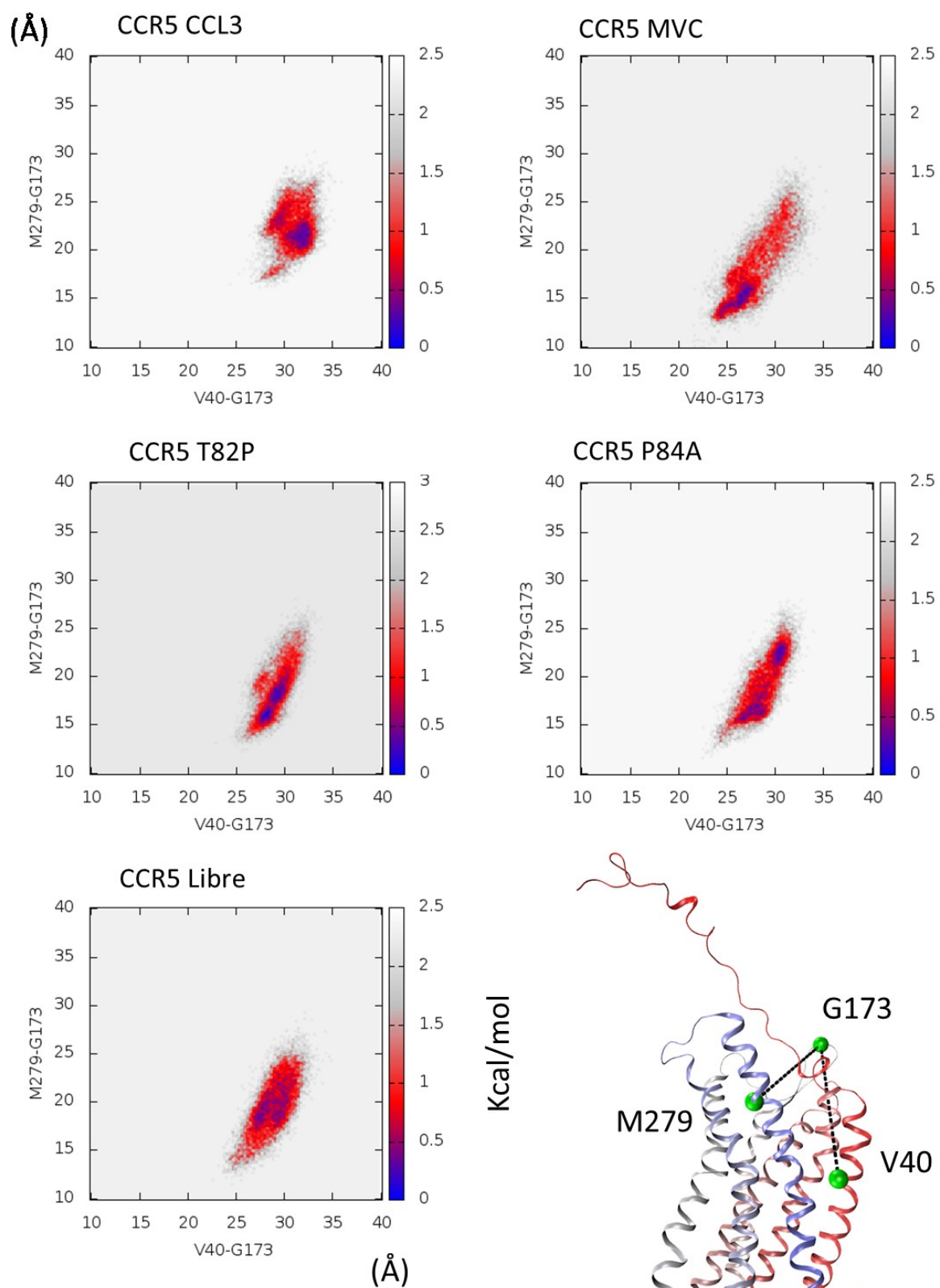


Figure 3 - 14 : Ensemble de conformations explorées par ECL2.

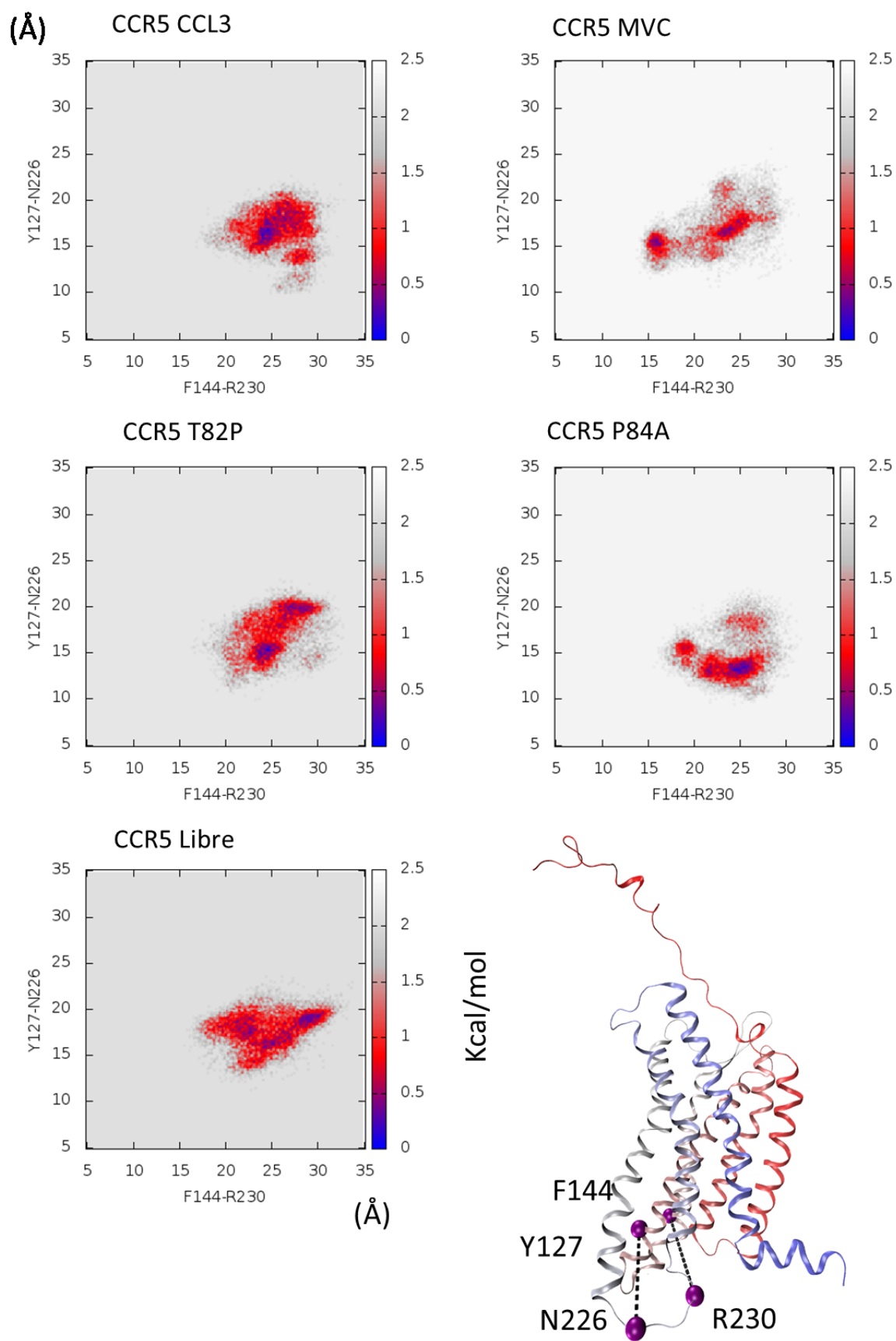


Figure 3 - 15 : Ensemble de conformations explorées par ICL3.

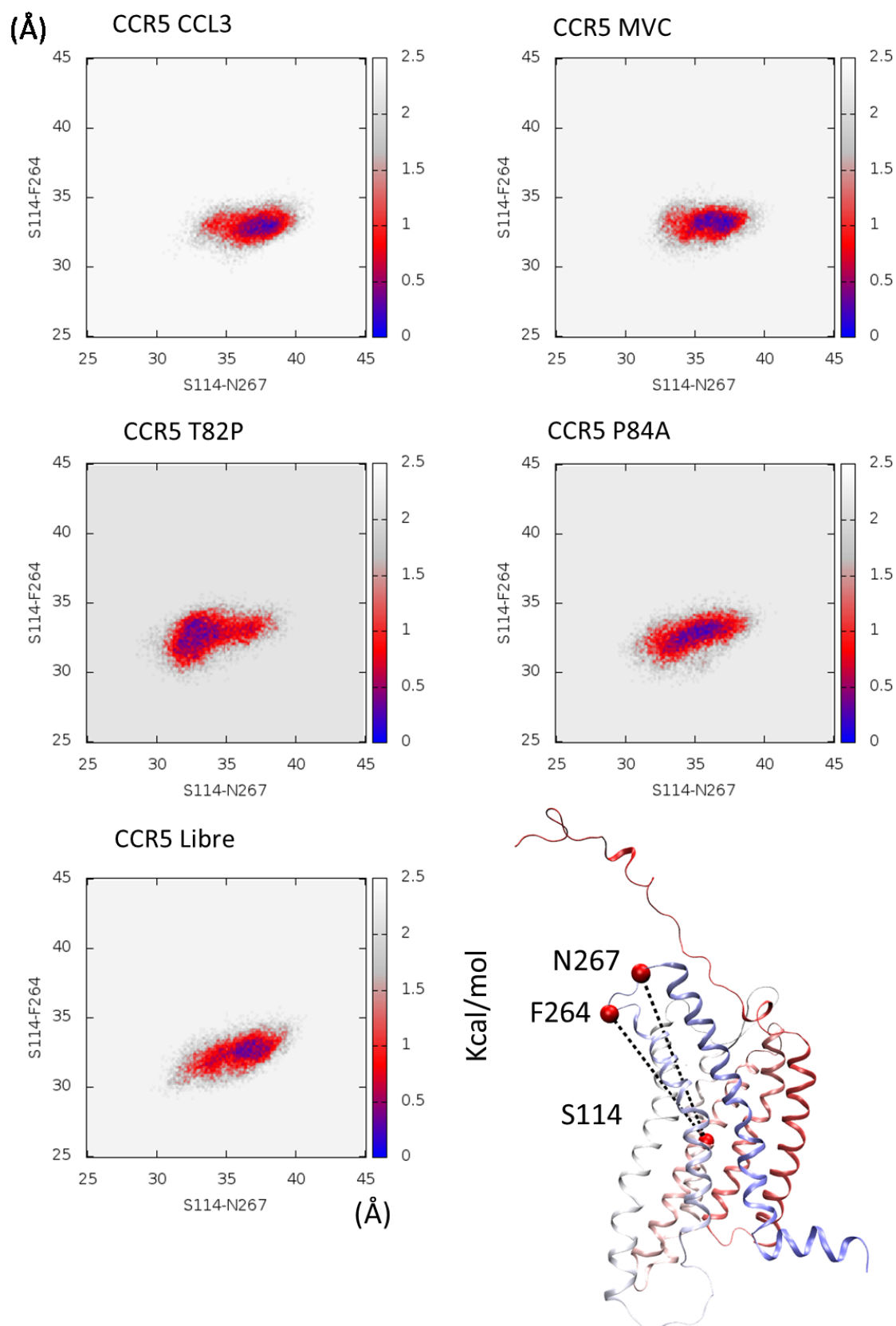


Figure 3 - 16 : Ensemble de conformations explorées par ECL3.

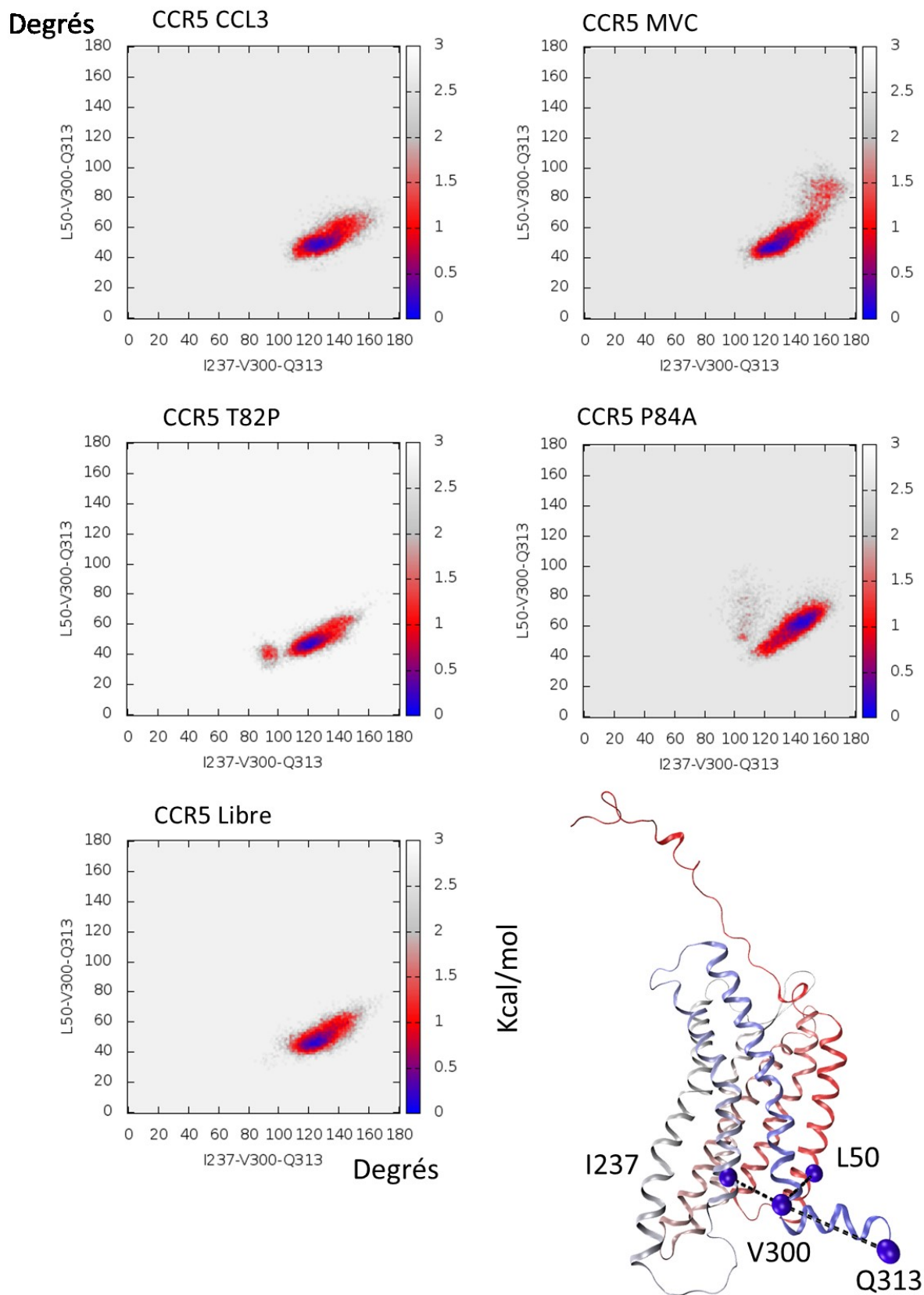


Figure 3 - 17 : Ensemble de conformations explorées par l'hélice 8.

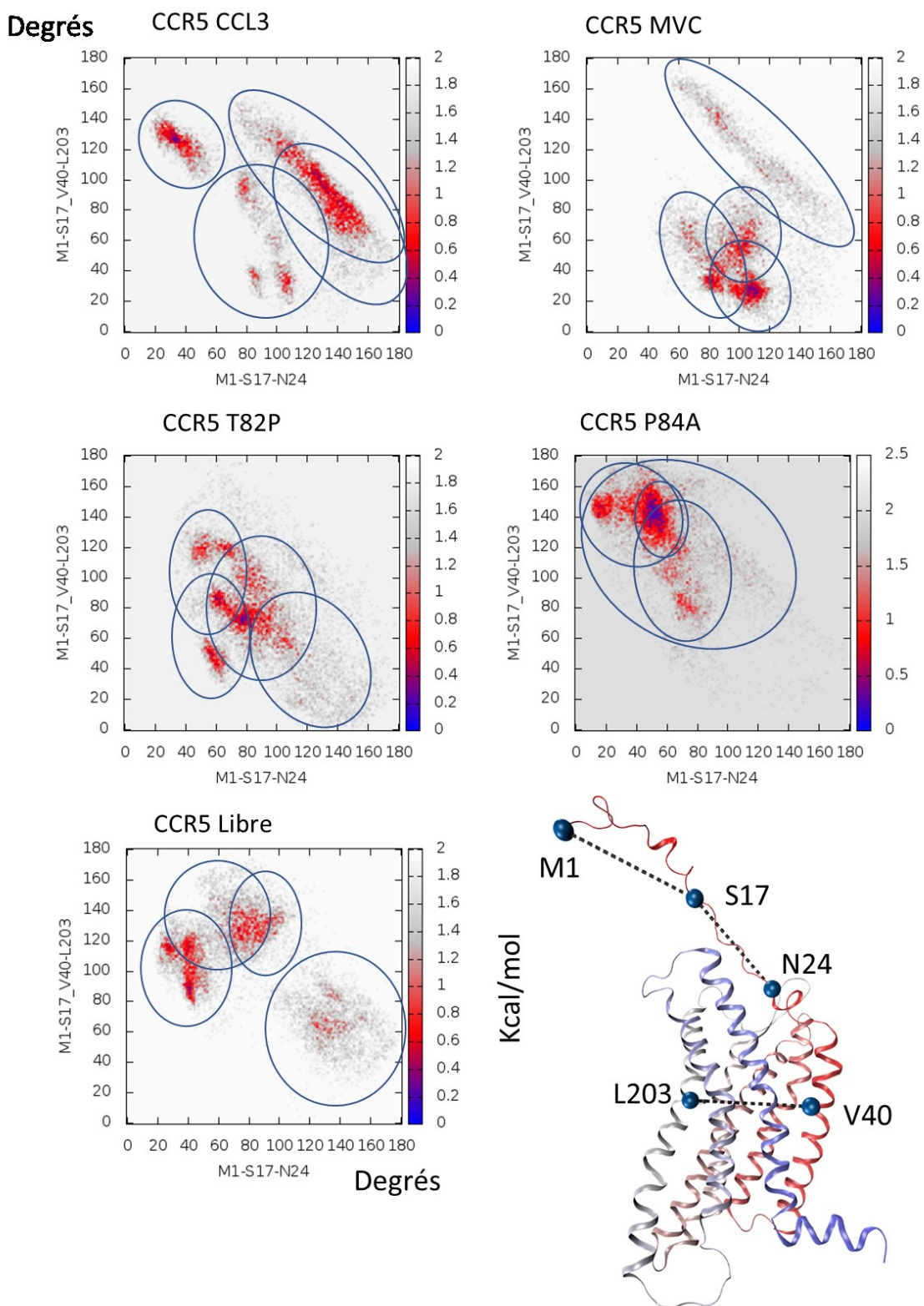


Figure 3 - 18 : Ensemble de conformations explorées par la partie N-terminale.



### 3.2.5.3 Caractérisation des structures secondaires

En visualisant les différentes trajectoires des systèmes, des changements caractéristiques de structure secondaire peuvent être identifiées pour la partie N-terminale et la boucle ICL3. Pour vérifier que ces changements sont statistiquement pertinents, une analyse de la structure secondaire a été effectuée en utilisant la méthode DSSP sur la totalité des répétitions.

Pour la partie N-terminale, la structure des résidus 1 à 50 du récepteur a été étudiée (Figure 3 - 19). Au sein de cette séquence, deux séries d'acides aminés adoptent des structures secondaires différentes selon les systèmes. La première série prend en compte les résidus 7 à 17. Ces résidus, sont caractérisés par une structure secondaire initiale du récepteur en hélice  $\alpha$ . Selon les systèmes étudiés, la proportion de présence de l'hélice  $\alpha$  est modifiée. Dans le cas des systèmes libre, CCL3 et T82P cette hélice est observable dans environ 90% des trajectoires. Tandis que pour les systèmes MVC et P84A, cette hélice n'est observable que dans environ 60% des trajectoires. La seconde série de résidus implique les résidus 29 à 32 du haut de l'hélice 1. Dans le cas du système CCL3, la structure secondaire en hélice  $\alpha$  de ces résidus est totalement absente en comparaison aux quatre autres systèmes où elle est observable dans 20 à 50% des trajectoires. Ce comportement spécifique du système CCL3 est un effet allostérique dû à l'interaction de la chaîne N-terminale de CCL3 avec les résidus Gln21 à Val25 de CCR5.

Pour la boucle ICL3 la structure des résidus 203 à 245 a été étudiée (Figure 3 - 20). Sur cette séquence, les résidus 225, 226 et 227, ont des structures secondaires différentes selon les systèmes. Dans le cas des systèmes libre, CCL3 et T82P, ces résidus sont caractérisés par une hélice  $\alpha$  dans respectivement 12, 8 et 17% des trajectoires. Dans le cas des systèmes MVC et P84A cette proportion diminue à 2%. On peut noter que les plus grandes proportions de sous-structures en hélice  $\alpha$  des résidus, pour les systèmes libre, CCL3 et T82P, sont majoritairement observables à la fin des trajectoires.

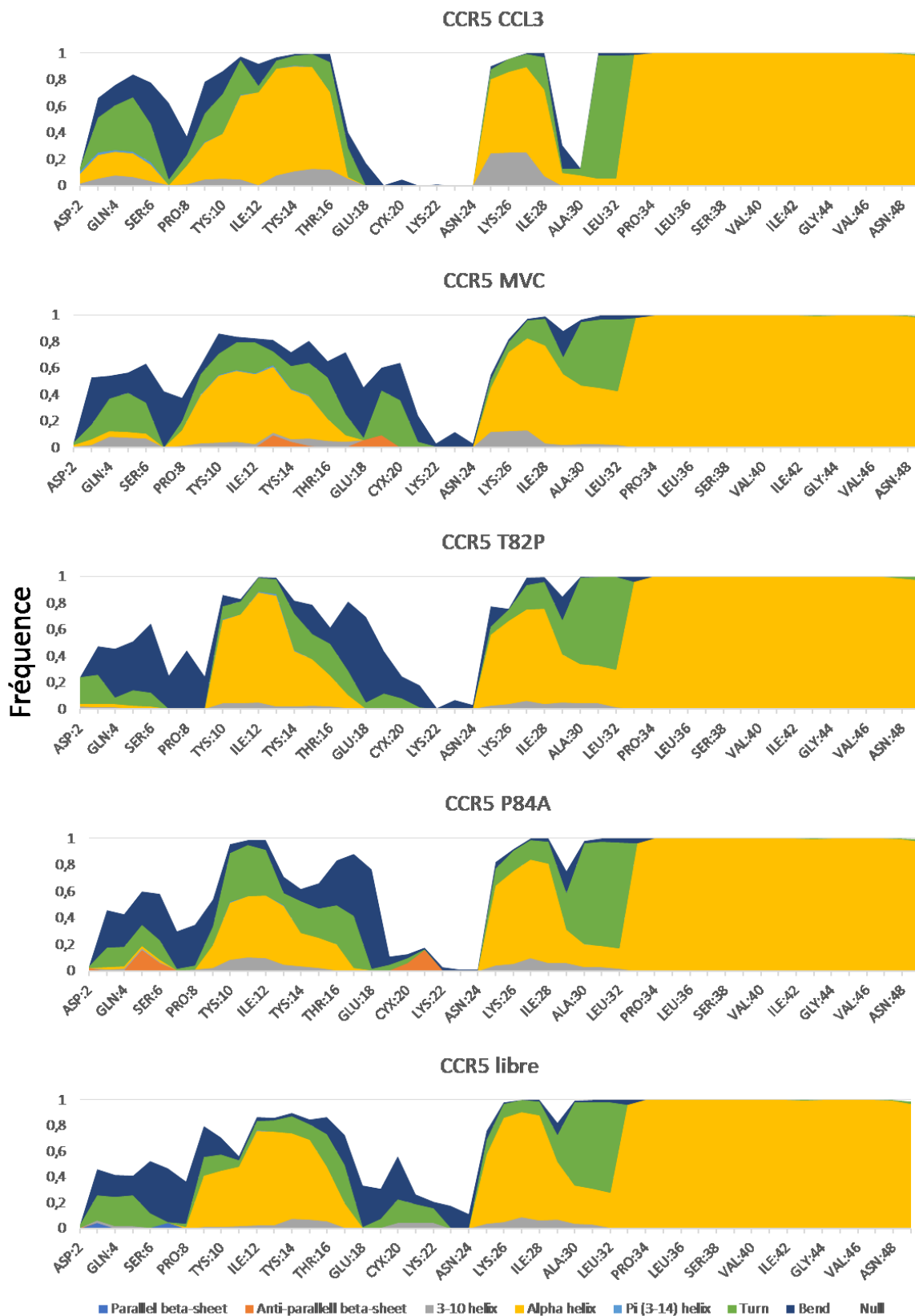


Figure 3 - 19 : Structure secondaire des résidus de la chaîne N-terminale et du début du TM1.



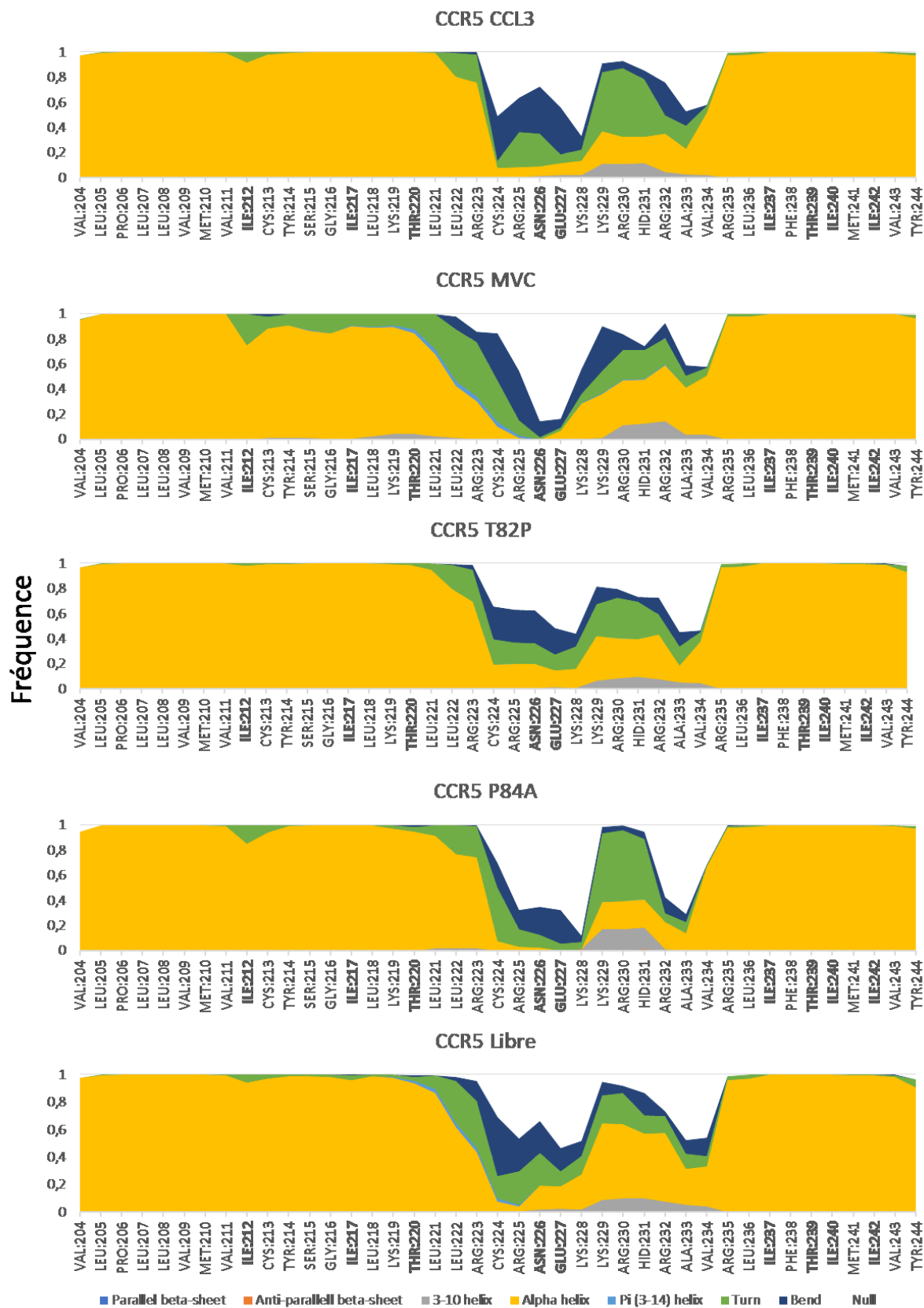


Figure 3 - 20 : Structure secondaire des résidus du TM 5, de la boucle ICL3 et du TM6.

### 3.2.6 Identification de motifs d'interactions intramoléculaires

Pour chaque système, des motifs d'interactions intramoléculaires ont été générés en considérant en même temps les quatre répétitions. Deux ensembles de motifs ont été générés, un premier pour lequel les trajectoires ont été considérées en totalité et un second pour lequel des partitions homogènes des trajectoires ont été considérées. Le second ensemble est utilisé pour identifier et caractériser des motifs d'interactions qui ne seraient pas visibles en utilisant la totalité de la dynamique. Dans les deux ensembles, seules les interactions du domaine transmembranaire (22 à 58, 63 à 92, 97 à 132, 141 à 167, 186 à 224, 228 à 265 et 268 à 300) sont considérées. La partie N-terminale et les boucles n'ont pas été prises en compte car apportant une trop grande variabilité et générant des partitions spécifiques aux répétitions dans le cas du second ensemble.

Les trajectoires ont été analysées en considérant un pas entre les structures de 200 ps. Pour chaque répétition 1400 structures ont été générées, amenant ainsi le nombre total de structures considérées à 5600 ou en équivalent de temps à 1,12  $\mu$ s. Ce pas a été choisi pour minimiser le temps de calcul tout en permettant d'obtenir le maximum d'information. Il a été sélectionné en considérant d'une part le pas de la dynamique du récepteur ADRB2, étudiée dans le chapitre 1, qui est de 180 ps, et en évaluant l'impact de l'échantillonnage sur les 120 premières ns de production de la répétition 1 du système CCR5 libre. Cette analyse a été effectuée en considérant des pas de 20 ps/40 ps/100 ps/200 ps/1 ns/2 ns. Sur la partie de trajectoire étudiée, le nombre d'interactions présentes dans plus de 10% des structures échantillonnées ne varie quasiment pas avec la diminution de la fréquence d'échantillonnage (Tableau 3 - 11). Néanmoins, une diminution conséquente du nombre d'interactions présentes dans moins de 10% de la trajectoire est observée pour les pas de 1 et 2 ns. L'échantillonnage correspondant aux pas de 20 ps, 100 ps et 200 ps produit des fréquences d'apparition de liaisons hydrogène qui sont identiques à 5% près (Tableau 3 - 12).

Tableau 3 - 11 : Nombre d'interactions de l'échantillonnage.

Pas entre les structures	Taux de présence de l'interaction											
	0 - 0.05	0.05 - 0.1	0.1 - 0.15	0.15 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1
20 ps	84	22	11	10	5	10	10	8	6	4	12	15
40 ps	85	26	10	9	7	9	10	8	6	4	12	15
100 ps	80	25	12	9	7	10	8	9	6	5	11	15
200 ps	79	23	13	9	9	10	8	8	6	5	11	15
1 ns	85	19	14	9	11	9	10	7	4	6	10	16
2 ns	63	15	15	7	11	10	9	8	6	3	10	16

Tableau 3 - 12 : Nombre d'interactions ayant une différence de leurs taux de conservation supérieur à 5% selon l'échantillonnage.

Pas entre les structures	20 ps	40 ps	100 ps	200 ps	1 ns	2 ns
20 ps						
40 ps	0					
100 ps	0	0				
200 ps	0	0	0			
1 ns	5	10	4	8		
2 ns	22	21	16	17	4	

### 3.2.6.1 Ensemble de motifs par dynamique

En considérant la totalité des trajectoires, le nombre d'interactions intramoléculaires est d'environ 260 (Tableau 3 - 13). Le nombre d'interactions est légèrement supérieur dans le récepteur libre, et au contraire inférieur dans le récepteur lié à CCL3 (Figure 3 - 21). Parmi ces interactions, 2% sont des interactions ioniques ou aromatiques et 7% des interactions s'effectuent entre un résidu et lui-même.

Tableau 3 - 13 : Nombre d'interactions par systèmes.

Systèmes	Nombre de structures	Nombre totale d'interactions	Proportion d'interactions ioniques/aromatiques	Proportion d'interactions intra-résidus
CCL3	5590	237	0.017	0.068
MVC	5600	260	0.019	0.069
T82P	5600	257	0.019	0.078
P84A	5600	263	0.027	0.065
Libre	5596	271	0.022	0.074

Selon les systèmes, la répartition des interactions selon leur taux de conservation est similaire. En moyenne, 23 interactions sont présentes dans 80% des trajectoires et 164 interactions sont présentes dans moins de 10% des trajectoires (Tableau 3 - 14 et Figure 3 - 21).

Tableau 3 - 14 : Taux de conservation des interactions.

Systèmes	Taux de présence des interactions											
	0 - 0.05	0.05 - 0.1	0.1 - 0.15	0.15 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1
CCL3	112	28	12	11	15	6	10	6	6	6	9	16
MVC	140	31	10	2	16	9	9	6	5	10	5	17
T82P	140	28	14	9	8	7	9	8	7	4	6	17
P84A	137	30	15	10	9	12	6	9	7	5	9	14
Libre	143	35	16	9	11	8	8	8	6	5	10	12

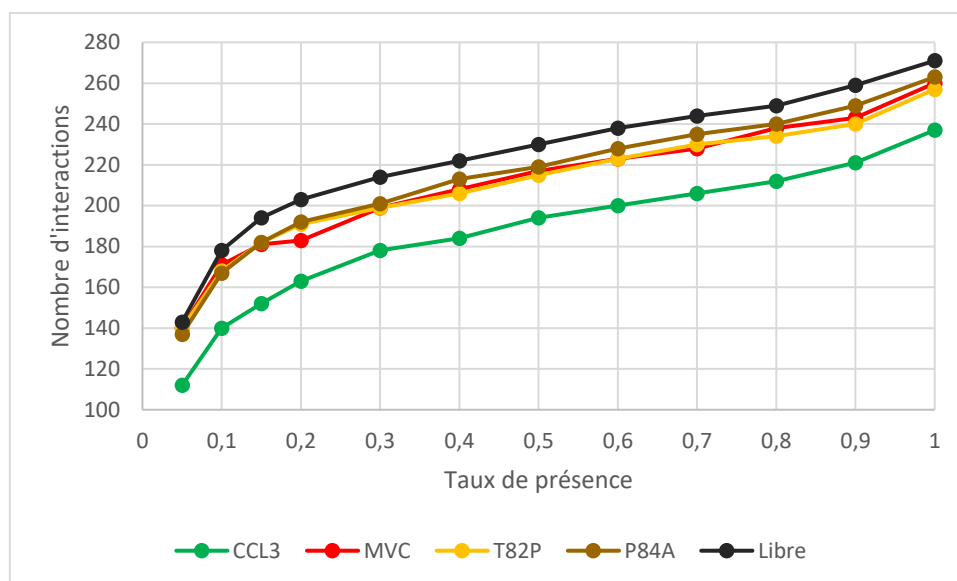


Figure 3 - 21 : Histogramme des valeurs cumulées du taux de conservation des interactions.

De par le nombre important d'interactions observées dans les systèmes, seules les interactions ayant un taux de conservation supérieure à 30% ont été étudiées. Cela correspond en moyenne dans les systèmes à 198 interactions (76% du total) présentes dans au moins 336 ns de simulations. Avec ce seuil, les motifs d'interactions des différents systèmes comportent en moyenne 60 interactions. Pour faciliter la lecture, les interactions présentes dans tous les cinq systèmes sont considérées comme communes au récepteur CCR5 et le motif de ces interactions est généré en effectuant une moyenne simple. Ce motif d'interaction commun et les motifs par système (interactions absentes d'au moins un système) sont présentés dans la Figure 3 - 22 (p112).

En observant les différents motifs, on peut remarquer que 48 interactions sont communes à tous les systèmes. Parmi ces interactions, 33 sont des liaisons hydrogène intra-TM, 13 des liaisons hydrogène inter-TM et seul deux interactions sont ioniques. L'une relie les résidus du motif DRY, D125(3.49) et R126(3.50) et l'autre les résidus E262(6.62)-K191(5.36). Les divers interactions inter-TM relient des hélices décrites comme couplées par l'analyse des corrélations de mouvement de CCR5 (TM2-TM4, TM2-TM3 et TM6-TM7). Mais elles relient également d'autre hélices sans couplage fort observé. Notamment, des interactions entre le TM5, seul TM caractérisé comme non couplé, et les TM3, 4 et 6 peuvent être identifiées.

Parmi les interactions spécifiques seules six sont uniquement présentes dans un système. En analysant leurs taux de présence dans les trajectoires sans prendre en compte la limite

minimale de conservation de 30% (Tableau 3 - 15), on peut remarquer que ces interactions ne sont pas strictement spécifiques mais surreprésentées. Les interactions des résidus E283(7.38), Q280(7.35) et Y251(6.51) spécifiques du système CCL3 et l'interaction entre les résidus A233(6.33) et R126(350) spécifiques du système MVC sont hautement intéressantes. Les résidus des interactions spécifiques du système CCL3(en interaction directe avec la chaîne N-terminal de CCL3 dans le modèle) ont été montré expérimentalement comme jouant un rôle important dans la liaison de la chimiokine<sup>7</sup>. L'interaction spécifique du système MVC, mime l'interaction caractéristique du DRY, sans néanmoins effectuer une interaction ionique. En plus de ces interactions surreprésentées dans un système, on peut remarquer la présence d'interactions sous-représentées dans un ou plusieurs systèmes (Tableau 3 - 16). Avec notamment, dans le cas du système T82P, deux interactions dans le motif TxP totalement absentes des trajectoires à cause de la mutation du résidu T82(2.56) ou une interaction ionique, entre les résidus 183(7.38) et 26(1.28), dans les systèmes CCL3 et MVC.

Tableau 3 - 15 : Taux de conservation des interactions identifiées comme spécifiques d'un système. Les cases colorées montrent le système ayant le taux de présence maximal.

Résidu 1	Résidu 2	Type d'interaction	Taux de conservation dans le système				
			CCL3	MVC	T82P	P84A	Libre
E283(7.38)	Q280(7.35)	Liaison Hydrogène	0,50	0	0,01	0,11	0,02
E283(7.38)	Y251(6.51)	Liaison Hydrogène	0,45	0,18	0,16	0,24	0,27
Q280(7.35)	Q280(7.35)	Liaison Hydrogène	0,35	0	0,02	0,05	0,02
Q102(3.26)	N98(3.22)	Liaison Hydrogène	0,34	0,23	0,21	0,27	0,18
A233(6.33)	R126(3.50)	Liaison Hydrogène	0,27	0,36	0,16	0,16	0,16
Y89(2.63)	K26(1.28)	Liaison Hydrogène	0,14	0,31	0,06	0,17	0,18

Tableau 3 - 16 : Taux de conservation des interactions identifiées comme absentes d'un système. Les cases colorées montrent le/les systèmes ayant le taux de présence minimal/minimaux.

Résidu 1	Résidu 2	Type d'interaction	Taux de conservation dans le système				
			CCL3	MVC	T82P	P84A	Libre
N48(1.50)	I44(1.46)	Liaison Hydrogène	0,28	0,38	0,77	0,57	0,57
T82(2.56)	F79(2.53)	Liaison Hydrogène	0,97	0,99	0	0,68	96
W86(2.60)	T82(2.56)	Liaison Hydrogène	0,46	0,84	0	0,32	0,59
T259(6.59)	T195(5.40)	Liaison Hydrogène	0,34	0,04	0,28	0,29	40
C290(7.46)	N48(1.50)	Liaison Hydrogène	0,23	0,29	0,54	0,45	0,45
D276(7.31)	K22(1.24)	Interaction ionique	0,44	0,53	0,19	0,32	0,18
E283(7.38)	K26(1.28)	Interaction ionique	0	0	0,35	0,30	0,30

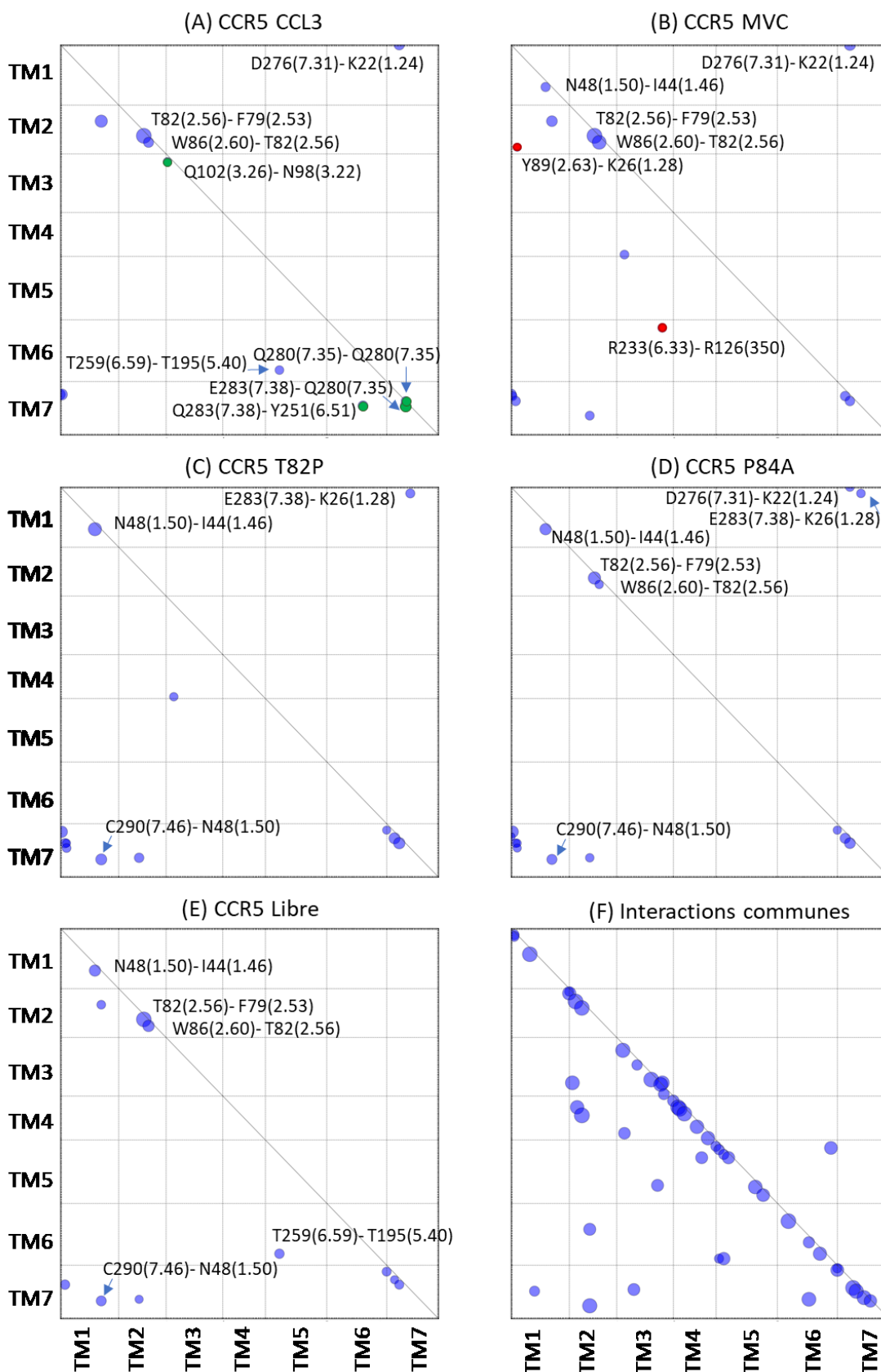


Figure 3 - 22 : Motifs d'interactions dans CCR5. (A-F) Interactions présentes dans au moins un système mais non commune à tous les systèmes pour le système (A) CCL3, (B) MVC, (C) T82P, (D) P84A, (E) Libre. (F) Interactions communes à tous les systèmes. La taille des points représentant l'interaction est proportionnelle à leur taux de conservation (minimum = 30%). Les interactions colorées représentent les interactions spécifiques dans le système.

### 3.2.6.2 Ensemble de motifs par dynamique partitionnée

Un certain nombre de partitions homogènes ont été générées à partir de la similitude des motifs d'interactions intramoléculaires du domaine transmembranaire pour chacun des systèmes. Le nombre total de partitions ainsi que le nombre de partitions considérées comme homogènes sont résumés dans le Tableau 3 - 17. La distribution des différentes partitions au long des trajectoires est représentée dans les Figure 3 - 23 à Figure 3 - 27 au côté des similitudes inter et intra-partitions.

Tableau 3 - 17 : Nombre de partitions. Une partition est considérée comme homogène lorsque la valeur du 1<sup>er</sup> quartile de la répartition des similitudes de celle-ci est supérieure à la valeur du 3<sup>ème</sup> quartile de la répartition des similitude inter-partitions.

	CCL3	MVC	T82P	P84A	Libre
Nombre total de partitions	10	7	11	9	9
Partitions considérées comme homogènes	5	4	4	4	3

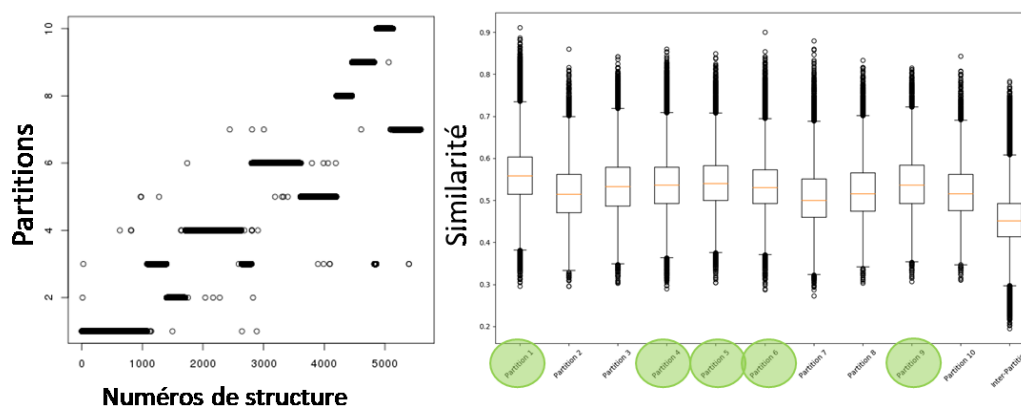


Figure 3 - 23 : Partitions du système CCL3. Les cercles verts correspondent aux partitions considérées homogènes.

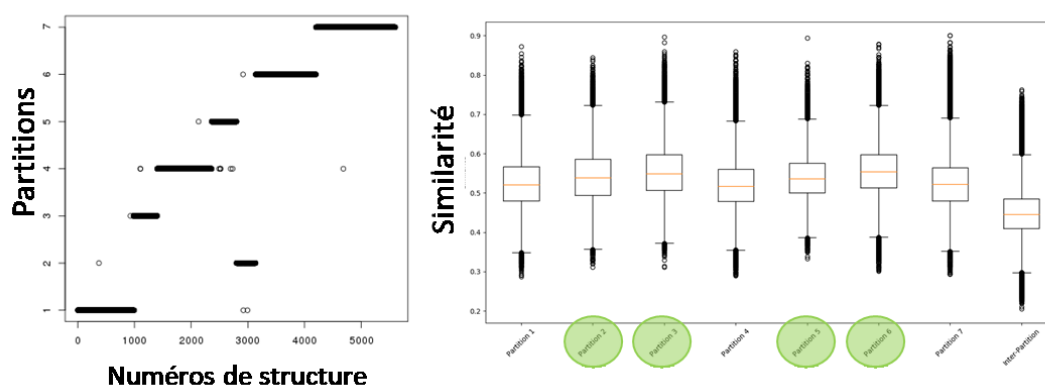


Figure 3 - 24 : Partitions du système MVC. Les cercles verts correspondent aux partitions considérées homogènes.

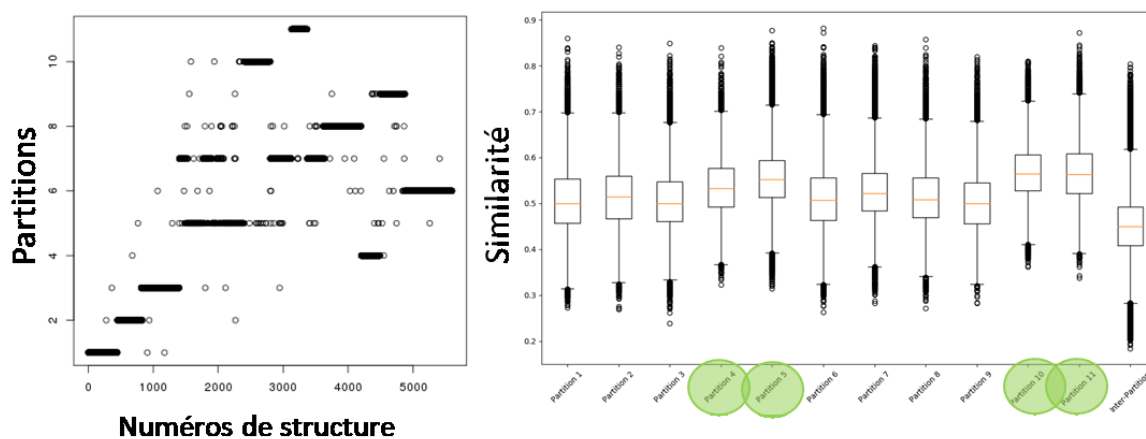


Figure 3 - 25 : Partitions du système T82P. Les cercles verts correspondent aux partitions considérées homogènes.

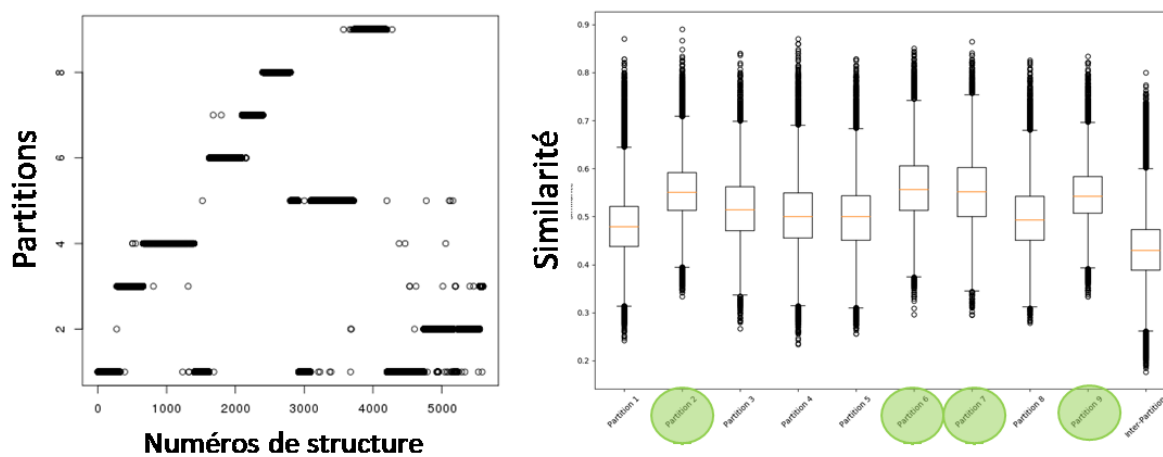


Figure 3 - 26 : Partitions du système P84A. Les cercles verts correspondent aux partitions considérées homogènes.

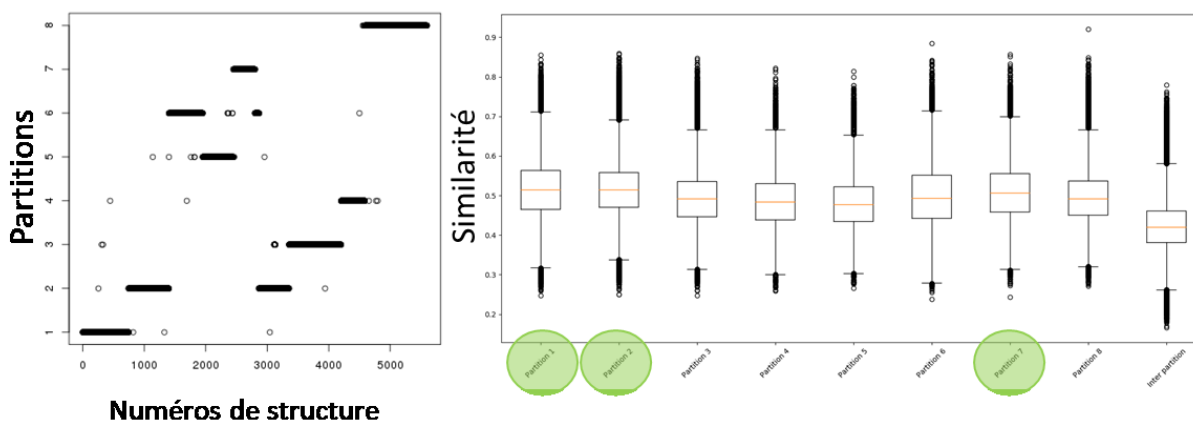


Figure 3 - 27 : Partitions du système Libre. Les cercles verts correspondent aux partitions considérées homogènes.



En considérant seulement les différentes partitions homogènes, le nombre d'interactions intramoléculaires est relativement similaire dans les différents systèmes avec environ 232 interactions. Parmi ces interactions, 2% sont des interactions ioniques ou aromatique et 7% des interactions s'effectuent entre un résidu et lui-même. En comparaison à l'étude précédente, seul 30% à 67% des trajectoires ont été étudiées (Tableau 3 - 18).

Tableau 3 - 18 : Statistique des structures et du nombre d'interactions pour les partitions homogènes des différents systèmes.

Systèmes	Proportion de dynamique considéré	Nombre de structures	Nombre total d'interactions	Proportion d'interactions ioniques/aromatiques	Proportion d'interaction intra-résidus
CCL3	0.67	3765	223	0.018	0.067
MVC	0.43	2407	225	0.018	0.076
T82P	0.30	1680	234	0.021	0.068
P84A	0.40	2250	238	0.025	0.071
Libre	0.40	2216	242	0.021	0.066

La répartition des interactions selon leur taux de présence est similaire dans les cinq systèmes. En moyenne, 24 interactions sont présentes dans 80% des partitions et 139 interactions sont présentes dans moins de 10% des partitions. (Tableau 3 - 19 et Figure 3 - 28). On peut remarquer que le nombre d'interactions est légèrement supérieur dans le système libre et au contraire légèrement inférieur dans les systèmes liés à un ligand.

Tableau 3 - 19 : Taux de conservation des interactions observables dans les motifs stables.

Systèmes	Taux de conservation des interactions											
	0 - 0.05	0.05 - 0.1	0.1 - 0.15	0.15 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1
CCL3	100	25	14	11	13	8	7	4	8	8	9	16
MVC	104	23	19	5	8	10	6	12	6	8	6	18
T82P	126	18	13	8	8	9	6	7	5	8	7	19
P84A	122	21	13	7	12	12	9	5	7	8	6	16
Libre	128	26	11	6	10	13	6	8	3	6	8	17

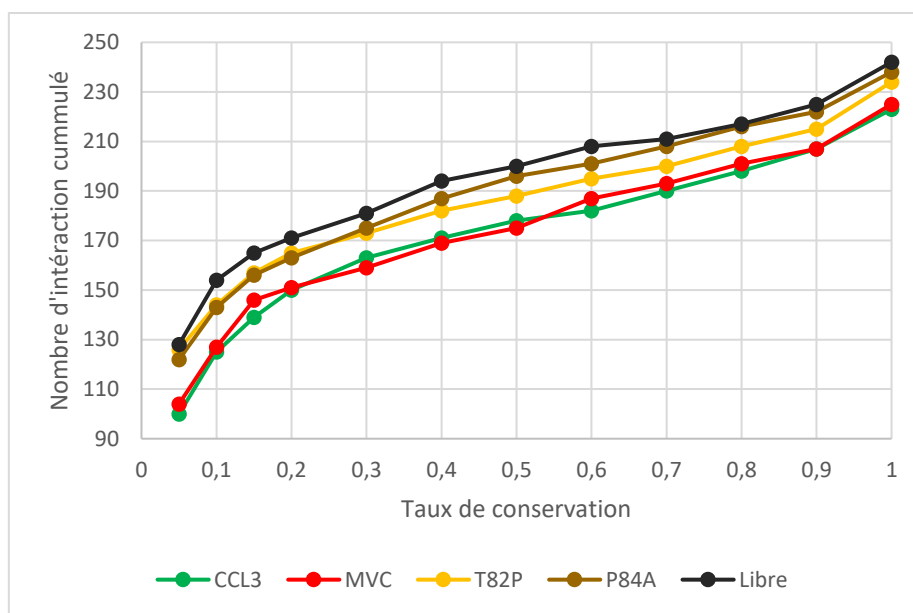


Figure 3 - 28 : Histogramme des valeurs cumulées du taux de conservation des interactions.

De par le nombre important d'interactions observés dans les systèmes, seules les interactions ayant un taux de présence supérieur à 30% ont été étudiées. Cela correspond en moyenne à 170 interactions (73% du total) présentes dans un équivalent en temps de moins de 100 à 225 ns de simulations selon les systèmes. Avec ce seuil, les motifs d'interactions des différents systèmes comportent en moyenne 62 interactions. Le motif commun et les motifs spécifiques ont été générés comme décrit précédemment. En observant ceux-ci, on peut remarquer que 45 interactions sont communes aux différents systèmes (Figure 3 - 29). En comparaison du motif commun généré pour l'ensemble des trajectoires, seulement trois interactions ne sont plus considérées comme communes. Ces interactions sont une liaison hydrogène entre les résidus E262(6.62) et Q188(5.33) et deux liaisons hydrogène intra-TM entre les résidus H132(3.56)-Y127(3.51) et K26(1.28)-N24(1.26). Ces liaisons se retrouvent parmi les motifs d'interactions spécifiques. Globalement, on retrouve les motifs décrits précédemment avec cette fois-ci 11 interactions surreprésentées et également des interactions sous-représentées (Tableau 3 - 20). Parmi les interactions surreprésentées on retrouve, pour le système CCL3, les résidus E283(7.38), Q280(7.35) et Y251(6.51) impliqués dans la liaison de CCL3 (Figure 3 - 30). On pourra noter, pour ce système, que l'interaction entre les résidus Q102(3.26) et N98(3.22) n'est plus considérée comme surreprésentée. Dans tous les motifs d'interactions spécifiques, de nouvelles interactions apparaissent ou disparaissent du fait de la sélection d'un ensemble

réduit de structures. Certaines nouvelles interactions sont surreprésentées. Avec, pour le système CCL3 la liaison hydrogène entre les résidus E283(7.38) et Y251(6.51) ; pour le système libre la liaison hydrogène entre les résidus S272(7.27) et Q261(6.61) ; et pour le système MVC, la liaison hydrogène entre les résidus K197(5.42) et P162(4.60) et une interaction hydrogène et ionique entre les résidus D125(3.49) et R230(6.30). Le fait d'observer une interaction ionique entre un résidu du motif DRY et un résidu du TM6 est surprenant car CCR5 est décrit comme n'ayant pas de verrou ionique. Néanmoins, cette interaction ionique couplée à la liaison hydrogène entre les résidus D125(3.49) et R230(6.30) (Figure 3 - 31) reproduit parfaitement le verrou ionique habituellement caractérisé dans les structures expérimentales inactivés des RCPGs.

Tableau 3 - 20 : Taux de conservation des interactions identifiées comme spécifiques d'un système. Les cases colorées montrent le système ayant le taux de présence maximal. La position des résidus impliqués selon la nomenclature Ballesteros est indiquée entre parenthèses.

Résidu 1	Résidu 2	Type d'interaction	Taux de présence dans le système				
			CCL3	MVC	T82P	P84A	Libre
E283(7.38)	Y251(6.51)	Liaison Hydrogène	0,40	0,20	0,18	0,23	0,24
E283(7.38)	Q280(7.35)	Liaison Hydrogène	0,59	0	0	0,08	0,03
Q280(7.35)	Q280(7.35)	Liaison Hydrogène	0,42	0	0,01	0,04	0,03
R235(6.35)	K229(6.29)	Liaison Hydrogène	0,31	0	0,02	0,02	0,02
Y89(2.63)	K26(1.28)	Liaison Hydrogène	0,18	0,52	0,04	0,08	0,27
A92(2.66)	K26(1.28)	Liaison Hydrogène	0,28	0,30	0,04	0,07	0,18
D125(3.49)	R230(6.30)	Liaison Hydrogène	0	0,52	0	0	0
D125(3.49)	R230(6.30)	Ionique	0	0,53	0	0	0
T167(4.65)	I164(4.62)	Liaison Hydrogène	0,03	0,35	0,08	0,02	0,011
K197(5.42)	P162(4.60)	Liaison Hydrogène	0,12	0,34	0,20	0,03	0,05
S272(7.27)	Q261(6.61)	Liaison Hydrogène	0,23	0,16	0,02	0,11	0,37

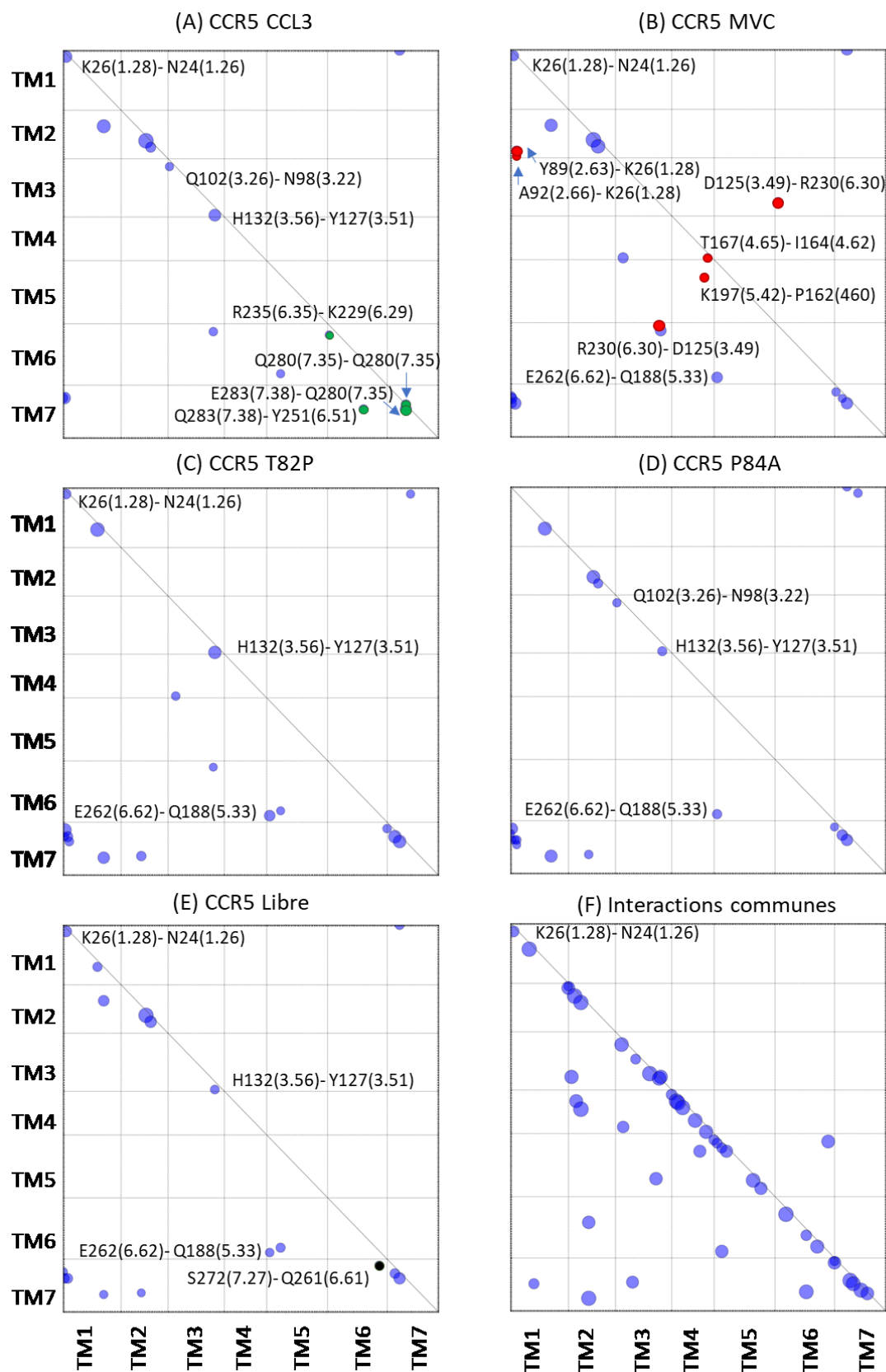


Figure 3 - 29 : Motifs d'interactions dans CCR5. (A-F) Interactions présentes dans au moins un système mais non commune à tous les systèmes pour le système (A) CCL3, (B) MVC, (C) T82P, (D) P84A, (E) Libre. (F) Interactions communes à tous les systèmes. La taille des points représentant l'interaction est proportionnelle à leurs taux de conservation (minimum = 30%). Les interactions colorées représentent les interactions spécifiques d'un système.

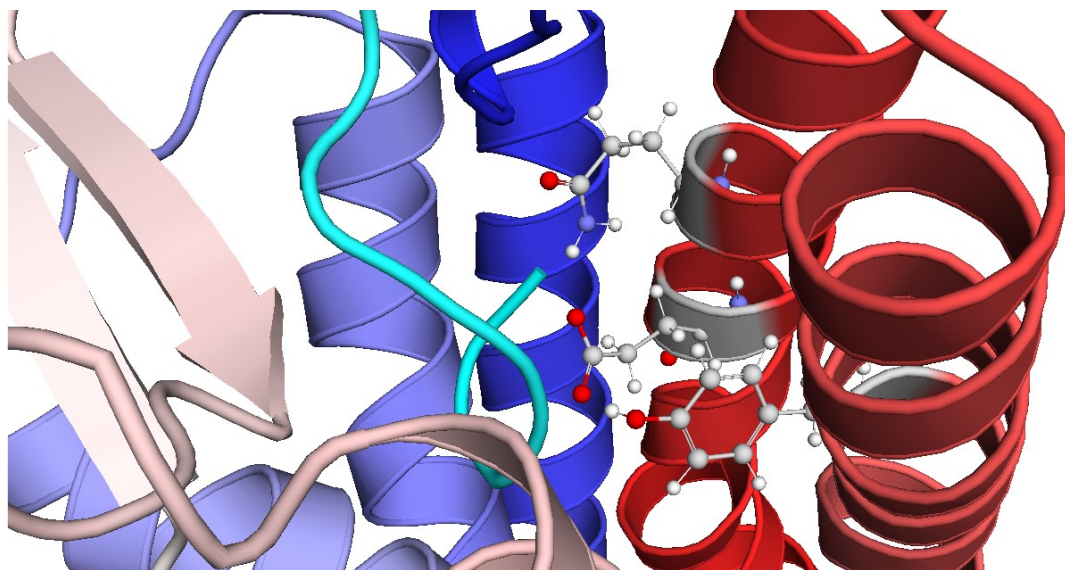


Figure 3 - 30 : Structure du système CCR5 CCL3, zoom sur les résidus Y251(6.51), Q280(7.35) et E283(7.38).

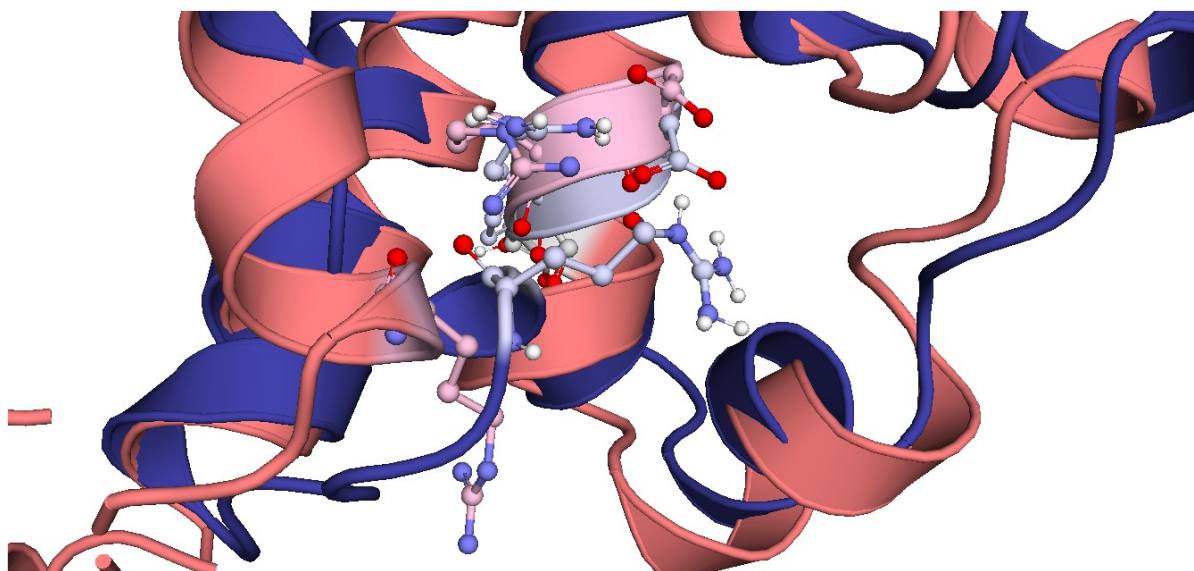


Figure 3 - 31 : Zoom sur le motif DRY, en bleu une structure d'une des partitions du système MVC et en rouge la structure cristallographique 4MBS.

## 4. CONCLUSION

---

L'objectif des travaux présentés dans ce chapitre vise à faire l'inventaire des différentes conformations du récepteur CCR5 sous sa forme monomérique dans le but d'étudier par la suite la liaison de la glycoprotéine virale gp120 au domaine extracellulaire de CCR5 ou la liaison d'effecteurs de signalisation au domaine intracellulaire de CCR5.

Pour pouvoir étudier les conformations et la dynamique de CCR5, les techniques de modélisation par dynamique moléculaire ont été utilisées et plus particulièrement, la dynamique moléculaire accélérée. Cette technique permet d'augmenter l'échantillonnage de conformations effectuées par les simulations et ainsi diminuer le temps de calcul nécessaire pour étudier le récepteur.

Cinq systèmes du récepteur CCR5 ont été simulés chacun par quatre répétitions de trajectoires moléculaires de 300 ns pour un totale de 1,2  $\mu$ s par système. Ces systèmes sont le récepteur libre, le récepteur lié à la chimiokine agoniste CCL3, un récepteur mutant T82P fonctionnellement constitutivement activé, le récepteur lié à l'agoniste inverse MVC et un récepteur mutant P84A s fonctionnellement constitutivement inactivé.

Tous les systèmes étant générés à partir d'une structure inactivée il était essentiel d'évaluer l'état d'activation du récepteur. Malgré l'utilisation de dynamique moléculaire accélérée aucun des différents systèmes, et notamment les CCR5/CCL3 et T82P, ne subissent de mouvements de grande amplitude caractéristique de l'activation des RCPG. Sur les cinq systèmes, seul le système CCR5 libre initie un changement conformationnel induisant une augmentation de la distance entre les hélices 3 et 6. Malheureusement, ce changement ne s'effectue que dans une des quatre répétitions et n'est pas stable. L'absence de changement conformationnel est également caractérisé par la faible fluctuation du domaine transmembranaire durant les simulations. Néanmoins, en prenant en compte les différences de mode de liaison des protéines  $G\alpha$ -s et  $G\alpha$ -i/o<sup>42</sup> et le fait que le récepteur CCR5 est montré comme pouvant recruter protéines  $G\alpha$ -i/o<sup>8</sup>, on ne peut pas exclure que les critères utilisés pour caractériser l'activation ne sont pas adaptés aux cas étudiés. D'autant que, l'analyse des trajectoires a permis d'identifier des caractéristiques spécifiques des différents systèmes.

L'étude des mouvements du récepteur dans les différentes trajectoires permet de caractériser des fluctuations spécifiques de certaines boucles du récepteur. Notamment, la boucle ECL2 effectue un mouvement de va et vient de l'intérieur vers l'extérieur de la protéine. Selon le type de stimulation du récepteur, ce mouvement est favorisé dans une position spécifique. La boucle ICL3 a également des préférences conformationnelles selon le type de stimulation du récepteur. Notamment, dans le cas des systèmes libre, CCL3 et T82P, c'est-à-dire activé et basal, une structure secondaire en hélice  $\alpha$  est observable de manière statistiquement significative. De façon similaire à la boucle ICL3, une structure secondaire en hélice  $\alpha$  est grandement favorisée dans la partie N-terminale par les systèmes libre, CCL3 et T82P et défavorisée dans les systèmes MVC et P84A.

L'analyse des motifs d'interactions intramoléculaires des différents systèmes permet également de définir des caractéristiques spécifiques. Dans la mesure où la variation du domaine transmembranaire est faible entre les systèmes, le nombre d'interactions caractéristiques est limité. Néanmoins les interactions observées sont caractéristiques d'états d'activation, que l'on considère la totalité des trajectoires ou seulement des partitions homogènes. Dans le cas du système MVC, l'utilisation de partitions de trajectoires permet d'observer une interaction mimant celle du motif DRY de l'hélice 3 avec l'hélice 6, interaction qui est invisible dans la structure cristallographique. Tandis que dans le cas du système CCL3 les interactions permettent d'identifier des résidus importants pour la liaison de la chimiokine (K229(6.29), Y251(6.51), Q280(7.35) et E283(7.38))

En résumé, les fluctuations spécifiques des boucles ECL2 et ICL3 qui ont déjà été rapportées comme importantes pour l'activation<sup>43,44</sup>, la présence d'interactions impliquant le motif DRY et l'identification de résidus nécessaires à la liaison de la chimiokine<sup>7</sup> nous permettent d'être confiant dans nos modèles. Il semble donc nécessaire de poursuivre les différentes trajectoires pour pouvoir observer des changements conformationnels d'amplitude dans les différents systèmes et ainsi obtenir de nouvelles conformations à partir de structure représentative de ces mouvements. Mais également de prendre en compte les nouveaux développements sur le mode de liaison des sous-unités  $G\alpha$ -i/o par rapport aux  $G\alpha$ -s.



## 5. BIBLIOGRAPHIE DE CHAPITRE

---

1. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harb. Perspect. Med.* **2**, a006866–a006866 (2012).
2. Dorr, P. *et al.* Maraviroc (UK-427,857), a Potent, Orally Bioavailable, and Selective Small-Molecule Inhibitor of Chemokine Receptor CCR5 with Broad-Spectrum Anti-Human Immunodeficiency Virus Type 1 Activity. *Antimicrob. Agents Chemother.* **49**, 4721–4732 (2005).
3. Tilton, J. C. *et al.* A Maraviroc-Resistant HIV-1 with Narrow Cross-Resistance to Other CCR5 Antagonists Depends on both N-Terminal and Extracellular Loop Domains of Drug-Bound CCR5. *J. Virol.* **84**, 10863–10876 (2010).
4. Garcia-Perez, J. *et al.* A single-residue change in the HIV-1 V3 loop associated with maraviroc resistance impairs CCR5 binding affinity while increasing replicative capacity. *Retrovirology* **12**, (2015).
5. Tan, Q. *et al.* Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science* **341**, 1387–1390 (2013).
6. Zheng, Y. *et al.* Structure of CC Chemokine Receptor 5 with a Potent Chemokine Antagonist Reveals Mechanisms of Chemokine Recognition and Molecular Mimicry by HIV. *Immunity* **46**, 1005-1017.e5 (2017).
7. Garcia-Perez, J. *et al.* Allosteric Model of Maraviroc Binding to CC Chemokine Receptor 5 (CCR5). *J. Biol. Chem.* **286**, 33409–33421 (2011).
8. Corbisier, J., Galès, C., Huszagh, A., Parmentier, M. & Springael, J.-Y. Biased Signaling at Chemokine Receptors. *J. Biol. Chem.* **290**, 9542–9554 (2015).
9. Steen, A. *et al.* Biased and Constitutive Signaling in the CC-chemokine Receptor CCR5 by Manipulating the Interface between Transmembrane Helices 6 and 7. *J. Biol. Chem.* **288**, 12511–12521 (2013).
10. Arias, D. A., Navenot, J.-M., Zhang, W., Broach, J. & Peiper, S. C. Constitutive Activation of CCR5 and CCR2 Induced by Conformational Changes in the Conserved T X P Motif in Transmembrane Helix 2. *J. Biol. Chem.* **278**, 36513–36521 (2003).
11. Lagane, B. Mutation of the DRY Motif Reveals Different Structural Requirements for the CC Chemokine Receptor 5-Mediated Signaling and Receptor Endocytosis. *Mol. Pharmacol.* **67**, 1966–1976 (2005).



12. Govaerts, C. *et al.* The T X P Motif in the Second Transmembrane Helix of CCR5: A STRUCTURAL DETERMINANT OF CHEMOKINE-INDUCED ACTIVATION. *J. Biol. Chem.* **276**, 13217–13225 (2001).
13. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 7 (2002).
14. Adcock, S. A. & McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **106**, 1589–1615 (2006).
15. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
16. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
17. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120**, 11919–11929 (2004).
18. Kastner, K. W. & Izaguirre, J. A. Accelerated molecular dynamics simulations of the octopamine receptor using GPUs: discovery of an alternate agonist-binding position: AMD Simulations of a GPCR Using GPUs. *Proteins Struct. Funct. Bioinforma.* **84**, 1480–1489 (2016).
19. Miao, Y., Nichols, S. E., Gasper, P. M., Metzger, V. T. & McCammon, J. A. Activation and dynamic network of the M2 muscarinic receptor. *Proc. Natl. Acad. Sci.* **110**, 10982–10987 (2013).
20. Gedeon, P. C., Thomas, J. R. & Madura, J. D. Accelerated Molecular Dynamics and Protein Conformational Change: A Theoretical and Practical Guide Using a Membrane Embedded Model Neurotransmitter Transporter. in *Molecular Modeling of Proteins* (ed. Kukol, A.) **1215**, 253–287 (Springer New York, 2015).
21. Markwick, P. R. L. & McCammon, J. A. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys. Chem. Chem. Phys.* **13**, 20053 (2011).
22. Jin, J. *et al.* CCR5 adopts three homodimeric conformations that control cell surface delivery. *Sci. Signal.* **11**, eaal2869 (2018).
23. Huang, C. -c. *et al.* Structures of the CCR5 N Terminus and of a Tyrosine-Sulfated Antibody with HIV-1 gp120 and CD4. *Science* **317**, 1930–1934 (2007).

24. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
25. Qin, L. *et al.* Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117–1122 (2015).
26. Liang, W. G. *et al.* Structural basis for oligomerization and glycosaminoglycan binding of CCL5 and CCL3. *Proc. Natl. Acad. Sci.* **113**, 5000–5005 (2016).
27. *Molecular Operating Environment (MOE)*, 2013.08. (Chemical Computing Group ULC, 2018).
28. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
29. Case, D. A. *et al.* *AMBER 2018*. (University of California).
30. Miao, Y., Nichols, S. E. & McCammon, J. A. Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. *Phys. Chem. Chem. Phys.* **16**, 6398–6406 (2014).
31. Kastner, K. W. *et al.* Characterization of the *Anopheles gambiae* octopamine receptor and discovery of potential agonists and antagonists using a combined computational-experimental approach. *Malar. J.* **13**, 434 (2014).
32. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
33. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
34. Skjærven, L., Yao, X.-Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, (2014).
35. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
36. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2017).
37. Pándy-Szekeres, G. *et al.* GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).
38. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

39. Da Silva, F., Desaphy, J. & Rognan, D. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein-Ligand Interactions. *ChemMedChem* **13**, 507–510 (2018).
40. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
41. Murtagh, F. & Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif.* **31**, 274–295 (2014).
42. Capper, M. J. & Wacker, D. How the ubiquitous GPCR receptor family selectively activates signalling pathways. *Nature* **558**, 529–530 (2018).
43. Venkatakrisnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
44. Salmas, R. E., Yurtsever, M. & Durdagi, S. Investigation of Inhibition Mechanism of Chemokine Receptor CCR5 by Micro-second Molecular Dynamics Simulations. *Sci. Rep.* **5**, (2015).

# Chapitre 4

Recherche de modulateurs de la dimérisation  
de CCR5

## 1. INTRODUCTION

---

L'oligomérisation est un mécanisme par lequel plusieurs récepteurs forment un complexe par le biais d'interactions non covalentes. Plusieurs degrés d'oligomérisation sont possibles. On parle de dimérisation dans le cas d'une complexation de deux chaînes peptidiques. Si la complexation s'effectue entre des récepteurs de même nature, le complexe est qualifié d'homo-oligomère. Dans le cas de récepteurs de nature différente, le complexe est qualifié d'hétéro-oligomère.

À leur découverte, les RCPGs ont été envisagés sous forme de monomères. Néanmoins depuis les deux dernières décennies, un nombre croissant d'études suggèrent l'existence d'oligomères. Ainsi, des récepteurs de type rhodopsine (classe A), des récepteurs métabotropique au glutamate (classe C)<sup>1</sup> et des récepteurs à la sécrétine (classe B)<sup>2</sup> ont été caractérisés comme formant des hétérodimères ou homodimères. Dans le cas du récepteur à l'acide  $\gamma$ -aminobutyrique<sup>3</sup> et du récepteur *Taste T1*<sup>4</sup>, tous deux appartenant à la classe C, il a été prouvé que la dimérisation est essentielle pour la fonction. Pour les autres classes et notamment pour la classe A, le rôle fonctionnel de la dimérisation est encore débattu.

Pour les RCPGs de classe A, plusieurs formes dimériques ont été observées dans des structures cristallographiques. Deux organisations dimériques peuvent être notamment distinguées<sup>5</sup>. La première implique les TM 1 et TM 2 ainsi que l'hélice 8, et a été observée pour les récepteurs ADRB1<sup>6</sup>, rhodopsine<sup>7</sup>,  $\kappa$ -opiacé<sup>8</sup> et  $\mu$ -opiacé<sup>9</sup>. La seconde implique TM 5 et a été observée pour les récepteurs *smoothened*<sup>10</sup>, ADRB2<sup>11</sup>,  $\mu$ -opioïde et CXCR4<sup>12</sup>. Cette interface autour de TM 5 varie selon le récepteur. Elle implique aussi TM4 dans le cas de CXCR4 et TM6 dans le cas du récepteur  $\mu$ -opioïde.

À ce jour, les structures cristallographiques de CCR5 montrent le récepteur sous forme monomérique. Néanmoins des données expérimentales suggèrent que CCR5 peut s'associer avec CCR2<sup>13</sup> et CXCR4<sup>14</sup>, et également former des homodimères<sup>15</sup>. Du point de vue fonctionnel, l'oligomérisation est susceptible d'influencer l'activité du récepteur<sup>16</sup> et d'inhiber l'entrée du VIH<sup>17</sup>.

Le travail présenté dans ce chapitre s'intègre dans un projet de recherche collaboratif avec le Docteur Anne BRELOT, membre de l'institut Pasteur de Paris. Ce projet a pour objectif

principal d'étudier la dimérisation de CCR5 et son influence sur l'infection par le VIH. Il vise notamment à répondre à plusieurs interrogations :

- Quelles sont les interfaces de dimérisation du récepteur ?
- Quel est l'impact de la perturbation de ces interfaces sur la dimérisation ?
- Quel est l'impact de la perturbation de la dimérisation sur la fonction de corécepteur du VIH ? sur l'activité de ligands antiviraux ?

Nous avons déjà prouvé expérimentalement l'existence d'homodimères de CCR5<sup>18</sup> et démontré que l'homodimérisation du récepteur est essentielle pour son expression à la surface de la membrane cellulaire<sup>18</sup>.

L'utilisation de techniques de mutagenèse dirigée et de pontage covalent nous a permis d'identifier des résidus essentiels à la dimérisation. Un travail de modélisation moléculaire a proposé deux modes de dimérisation en accord avec les résultats expérimentaux (Figure 4 - 1)<sup>18</sup>. Ces dimères, nommés I5 et I56, reproduisent les organisations observées dans les structures cristallographiques de dimères de CXCR4<sup>12</sup> pour le dimère I5 et du récepteur  $\mu$ -opioïde<sup>9</sup> pour le dimère I56.

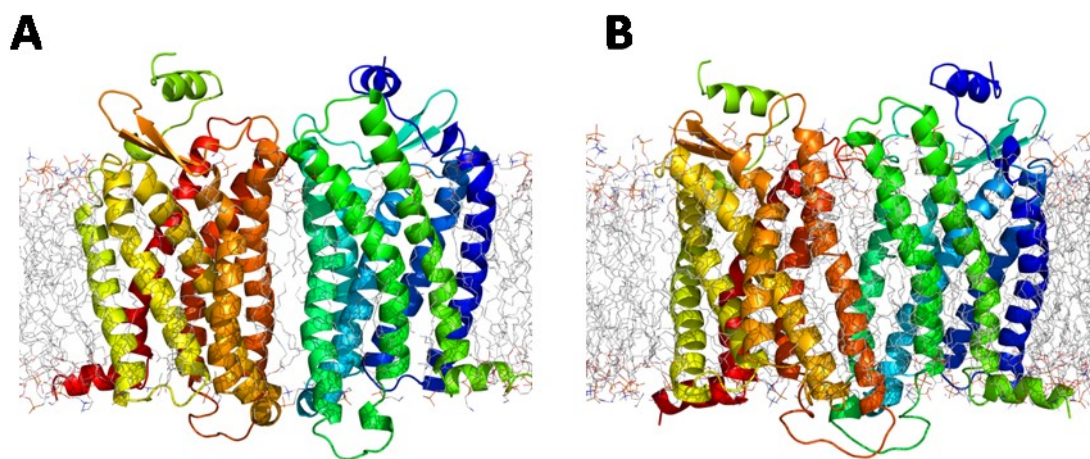


Figure 4 - 1 : Modèles de dimère de CCR5 inséré dans une membrane lipidique. Le dégradé de couleur du ruban représentant CCR5 correspond à la séquence de chaque chaîne. La chaîne A va du vert au rouge tandis que la chaîne B va du bleu au vert. (A) dimère I5. (B) dimère I56.

Outre ma participation aux travaux déjà publiés, qui ne seront pas présentés plus en détail ici, mon objectif dans cette étude a été de concevoir des modulateurs chimiques non covalents de la dimérisation de CCR5. Il a été montré par le Docteur Bernard LAGANE que différents variants de la protéine gp120 du VIH ciblent préférentiellement différentes formes

monomériques ou dimériques du récepteur. L'hypothèse émise dans ce projet est que des ligands interfaciaux permettront de moduler la capacité de CCR5 à former des complexes dimériques spécifiques, fournissant ainsi un outil supplémentaire pour étudier les implications fonctionnelles de la dimérisation, tout en étudiant un récepteur à l'état natif.

Pour atteindre cet objectif, un criblage virtuel par pharmacophore a été utilisé. Cette approche, basée sur la structure du récepteur, vise à générer un modèle en 3D, dit pharmacophore, caractérisant les propriétés physico-chimiques d'intérêt, ou points pharmacophoriques, de notre cible, ici les dimères I5 et I56. Ce modèle 3D, conçu à partir de cavités interfaciales ou proches de l'interface, est ensuite comparé à une chimiothèque pour identifier les molécules vérifiant le pharmacophore et susceptibles de stabiliser ou déstabiliser le dimère.

## 2. MATÉRIEL ET MÉTHODES

---

### 2.1 MATÉRIEL

#### 2.1.1 Chimiothèque

Pour effectuer le criblage virtuel, une chimiothèque de petites molécules disponibles dans le commerce a été utilisée. Cette chimiothèque, appelée Bioinfo<sup>19</sup> (<http://bioinfo-pharma.u-strasbg.fr/bioinfo/>), est compilée au sein du laboratoire. Elle consiste en un regroupement de catalogues de fournisseurs de chimiothèques, qui sont fusionnés et traités pour :

- Avoir un identifiant unique par molécule
- Avoir une structure 2D standardisée
- Supprimer les molécules redondantes
- Ne pas prendre en compte des molécules ayant des propriétés ou des groupements chimiques non désirés pour des molécules à visée thérapeutique

La version 16.1 de la Bioinfo, constituée de 5.498.747 molécules, a été utilisée pour les deux criblages. Cette version a été générée à partir de 23 bases de données commerciales totalisant 14.880.778 molécules.

La base Bioinfo n'a pas été considérée dans son intégralité, mais a fait l'objet d'un tri avec le logiciel Pipeline Pilote 9.5<sup>20</sup> pour sélectionner les molécules ayant une bonne solubilité aqueuse ( $\log S > -4,5$ ), et n'étant pas trop flexibles (nombre de liaison à rotation libre inférieur à 10).

Une structure en 3D de chaque molécule est générée avec le logiciel Corina version 3.4<sup>21-23</sup> et des options *wh* (ajout des hydrogènes) et *wb* (écriture de mauvais modèle), *stergen* (création de stéréoisomères), *msi* 4 (maximum de stéréoisomères), *msc* 2 (maximum de centres chiraux) et *preserve* (prise en compte des centres chiraux définis en 2D).

#### 2.1.2 Structures tridimensionnelles des dimères I5 et I56

Les dimères considérés, I5 et I56, sont tous les deux des modèles obtenus par assemblage des protomères en calquant des organisations de RCPGs cristallisées. Le dimère I5 a été construit pour reproduire l'organisation du dimère de CXCR4 (identifiant pdb 3OE0)<sup>12</sup>. Le dimère I56 a été construit pour reproduire celle du dimère de récepteur  $\mu$ -opioïde (identifiant pdb 4DKL)<sup>9</sup>.



La structure d'un protomère de CCR5 inclut les résidus 7 à 313 et sa préparation a été présentée dans le chapitre 3 (2.1.1, p75).

Les modèles ont été construits par la Professeure Esther KELLENBERGER. Leur stabilité a été testée par dynamique moléculaire. Les systèmes ont été préparés d'après le protocole décrit dans le chapitre 3 (2.1.2, p78), c'est-à-dire au moyen du site CHARMM-Gui<sup>24</sup> et de la version GPU du logiciel AMBER 14<sup>25</sup> par le biais de la commande *tleap*. Les étapes et leurs paramètres sont identiques à ceux décrits dans le chapitre 3 (2.1.2.1, p78), à l'exception de la hauteur de la couche d'eau qui est de 12 Å. La composition des différents systèmes est présentée dans le Tableau 4- 1.

Tableau 4- 1 : Composition des systèmes simulés par dynamique moléculaire.

Systèmes	Nombre de molécules					Nombre d'atomes	Dimension de la Boite (ÅxÅxÅ)
	POPC	POPE	CHL	Eau	Ions (K/Cl)		
<b>Dimère I5</b>	202	201	100	26724	75/95	143798	140x140x116
<b>Dimère I56</b>	180	180	90	24839	70/90	138133	135x135x102

Trois répétitions de dynamique moléculaire classique ont été générées pour chaque système. Ces répétitions ont une durée moyenne de 300 ns et ont été générées en suivant le protocole décrit dans le chapitre 3 (2.1.2.2, p79). Ce protocole consiste en une étape d'équilibrage commune aux répétitions et une étape de production. Les paramètres modifiés sont la durée de la phase de stabilisation lors de l'étape d'équilibrage, qui est de 5 ns et le pas de sauvegarde des structures lors de la phase de production qui est de 500 ps.

Au total, le système I5 est caractérisé par 1763 structures correspondant à 881,5 ns de simulation et le système I56 par 1675 structures correspondant à 837,5 ns de simulation.

Les structures du récepteur ont été extraites des trajectoires grâce au module *cpptraj*<sup>26</sup> du logiciel AMBER 14 en omettant les molécules d'eau, les ions et les lipides. Les coordonnées sont sauvegardées au format PDB. Les noms des résidus modifiés CYX, TYS, HIE et HID ont été transformés respectivement en CYS, TYR et HIS. Les fichiers PDB ont été convertis au format mol2 grâce au logiciel Sybyl v2.1.1<sup>27</sup>. Les résidus 1 à 307 forment la chaîne A et les résidus 308

à 614 forment la chaîne B. Pour chaque structure, ont été générés un fichier pour le dimère et un fichier séparé pour chacun des deux protomères.

## 2.2 SÉLECTION D'UNE CAVITÉ REPRÉSENTATIVE

La sélection d'une cavité représentative pour chaque dimère a été effectuée en suivant trois étapes. Ces étapes sont les mêmes pour les deux dimères étudiés mais les méthodes et paramètres utilisés peuvent varier.

La première étape caractérise l'interface de dimérisation. Elle identifie les résidus en interaction entre protomères. La seconde étape détecte les cavités interfaciales et à la surface des protomères. La troisième étape recherche une cavité droguable représentative de l'ensemble conformationnel analysé.

### 2.2.1 Caractérisation de l'interface

L'interface de dimérisation a été caractérisée par le biais du module IFP<sup>28</sup> du logiciel IChem v4.0<sup>29</sup> en utilisant les options par défaut. Les fichiers au format PDB des deux protomères d'une structure sont utilisés en entrée. En sortie sont énumérées toutes les interactions non covalentes établies entre les deux protomères. Pour chaque interaction est donné son type, les noms et numéros des résidus et atomes engagés dans l'interaction, et ses caractéristiques de distance et d'angle.

Ces fichiers ont servi au calcul de la fréquence d'apparition des interactions.

### 2.2.2 Détection de cavités

Les cavités ont été détectées dans les structures de dimères à l'aide d'un script python, développé par F. Da Silva. Ce script, appelé *analyseppi* correspond désormais au module *detectppi* du logiciel IChem<sup>30</sup>. Les fichiers au format PDB des structures de dimère sont utilisés en entrée. En sortie, des points de cavités sont donnés au format Mol2. Ces points sont représentés par des pseudo-atomes dont les propriétés pharmacophoriques sont complémentaires de ceux des atomes de protéines les plus proches. À chaque cavité sont associées une description comprenant sa taille, le nombre total de points de cavité, son volume, le nombre de point de cavités : hydrophobe, aromatique, accepteur de liaison hydrogène (HBA), donneur de liaison hydrogène (HBD), accepteur/donneur de liaisons hydrogène (HBA/HBD), chargés positivement, chargés négativement, *dummy* (pas d'atomes

de protéines a moins de 4,5 Å) ainsi qu'un score de droguabilité. Ces caractéristiques ainsi que le score de droguabilité ont été obtenus grâce au module Volsite<sup>31</sup> de IChem<sup>29</sup>. Un score positif suggère un site droguable.

Enfin, le script fait appel à un modèle mathématique de forêts aléatoires IChemPIC<sup>30</sup> (bioinforma.pharma.u-strasbg.fr/ICHEM PIC) pour prédire de manière binaire si l'interface a une pertinence biologique ou non.

### 2.2.3 Sélection du site d'intérêt

Pour chacun des deux dimères, le site d'intérêt a été choisi à partir des cavités de structures représentatives de l'interface en considérant des critères de tri successifs :

- Critère 1 : droguabilité
- Critère 2 : localisation de la cavité à l'interface du dimère
- Critère 3 : taille et polarité de la cavité
- Critère 4 : persistance de la cavité pendant la simulation par dynamique moléculaire
- Critère 5 : faible fluctuation du site pendant la simulation par dynamique moléculaire

#### 2.2.3.1 Structure représentative de l'interface de dimérisation

Pour chacun des deux dimères, une structure représentative a été choisie à l'aide du module *cpptraj* du logiciel AMBER 14.

Pour le dimère I5, les structures sont alignées à l'aide de la commande *RMSD* pour une superposition optimale des résidus de l'interface impliqués dans des interactions présentes dans plus de 75% de la simulation. La commande *cluster* a été ensuite utilisée pour obtenir les structures représentatives de cinq partitions basées sur le RMSD des résidus impliqués dans des interactions présentes dans plus de 1% de la dynamique.

Pour le dimère I56, les structures sont alignées à l'aide de la commande *RMSD* pour une superposition optimale des résidus de l'interface impliqués dans des interactions présentes dans plus de 1% de la simulation. La commande *cluster* a été ensuite utilisée pour obtenir les structures représentatives de cinq partitions basées sur le RMSD des atomes de la chaîne principale (C, CA, N, O) des résidus impliqués dans des interactions présentes dans plus de 50% de la simulation.

Enfin, pour chaque dimère, le choix final de la structure représentative est effectué en tenant compte de trois critères :

- Nombre important de structures dans la partition correspondante (> 15% des trajectoires)
- Temps d'apparition des structures dans la partition correspondante (>50ns)
- Interface de dimérisation prédite comme étant biologiquement pertinente

### 2.2.3.2 Sélection d'une cavité

Le tri des cavités de la structure représentative a été réalisé comme suit.

- Les cavités dont la droguabilité est négative ou nulle ont été écartées.
- Les cavités localisées au niveau du site de liaison de la protéine G ont été écartées.
- Les cavités trop grandes (taille supérieure à 1500 Å<sup>3</sup>), pas assez hydrophobes (nombre de points hydrophobe < 30) ou pas assez polaires (nombre de points polaires < 30) ont été écartées.
- Le calcul de persistance de la cavité est effectué au moyen du programme ROCs 3.2.0.4<sup>32</sup> (OpenEye) qui évalue la superposition des points de cavité de la structure représentative avec celles de toutes les structures issues de la simulation. Le score considéré est le tanimoto combo, obtenu sans alignement (option *scoreonly*) avec un seuil de 0,1.
- La fluctuation des résidus formant la cavité est estimée au moyen du module cpptraj du logiciel AMBER 14. La commande *RMSD* avec l'option *perres* est utilisée. Le RMSD est calculé sur les coordonnées des atomes de la chaîne principale (CA, C, N, O) après alignement sur la structure représentative pour la meilleure superposition des résidus à moins de 4,5 Å des points de la cavité.

## 2.3 CRIBLAGE VIRTUEL

Le criblage virtuel comporte trois étapes. Ces étapes sont les mêmes pour les deux dimères étudiés mais les méthodes et paramètres utilisés peuvent varier. La première étape consiste en l'élaboration d'un pharmacophore à partir du site d'intérêt. L'étape suivante évalue chaque

molécule de la chimiothèque par comparaison au pharmacophore. La dernière étape sélectionne une centaine de touches du criblage.

### 2.3.1 Création d'un pharmacophore

Pour chacun des deux dimères, le pharmacophore a été déterminé à partir du site d'intérêt de la structure représentative. Le module Volsite avec l'option --pharm du logiciel IChem permet de générer des points pharmacophoriques à partir de la cavité. Ces points pharmacophoriques sont des points hydrophobes, des donneurs de liaison hydrogène (HBD), des accepteurs de liaison hydrogène (HBA), des points aromatiques et des sphères d'exclusion. Le pharmacophore a nécessité un ajustement manuel à cause du grand nombre de points obtenus.

L'édition du pharmacophore a été effectuée avec la version graphique du logiciel Discovery Studio 4.5<sup>33</sup> comme suit :

- Les points hydrophobes proches ont été regroupés en un point unique de rayon augmenté.
- Lors de la présence simultanée de points hydrophobes et aromatiques dans une même région, les points hydrophobes-aromatiques ont été privilégiés.
- Les points HBA et HBD en interaction avec des atomes de la chaîne principale effectuant une interaction  $\text{CO}_{(i)}\text{-NH}_{(i+3/4)}$  ont été supprimés.
- Les points HBA et HBD pointant sur le même résidu ont été fusionnés.
- Les points en interaction avec un atome de soufre ont été supprimés.

### 2.3.2 Criblage d'une chimiothèque

Le criblage virtuel a été effectué avec la version en ligne de commande du logiciel Discovery Studio<sup>33</sup>. Cette version utilise le logiciel associé Pipeline Pilot 9.5<sup>20</sup>. La ligne de commande *catconf* a été utilisée pour effectuer un échantillonnage conformationnel des molécules. L'algorithme Fast a été utilisé pour générer un maximum de 255 conformations par molécule. Une fois l'échantillonnage effectué, la ligne de commande *citest* a été utilisée pour aligner de manière rigide les conformations aux pharmacophores et ensuite évaluer cet alignement par un score. Seul le meilleur alignement est considéré. La distance minimale entre point pharmacophoriques détectés par Discovery Studio est de 50 pm. Pour évaluer l'alignement, un certain nombre de points pharmacophoriques peuvent être omis. Pour le modèle I5, quatre

points ont pu être omis avec un minimum de points considéré de trois et un maximum de sept. Pour le modèle I56, le maximum de points pharmacophoriques peut être omis (valeur du paramètre n-1) avec un minimum de points considérés de trois et un maximum de huit.

### 2.3.3 Sélection des touches d'intérêt

La sélection d'un échantillon de molécules d'intérêt pour chaque dimère a été effectué sur des critères basés sur le pharmacophore et sur le châssis moléculaire des touches.

#### 2.3.3.1.1 Accord avec le pharmacophore

Cette première sélection a été effectuée avec le logiciel Pipeline Pilot 9.5 pour diminuer le nombre de touches du criblage virtuel.

Pour le dimère I5, les molécules considérées pour la suite de l'étude doivent se conformer aux critères suivants :

- Un score de *fitness* supérieur à 4
- Un point HBA obligatoire et un second point obligatoire parmi une sélection de trois points proches (HBA, HBD et aromatique).

Pour le dimère I56, les molécules considérées pour la suite de l'étude doivent se conformer aux critères suivants :

- Un score de *fitness* supérieur à 3
- Les points hydrophobes du pharmacophore doivent concorder avec des points pharmacophoriques de la molécule.
- Un rapport entre le nombre de points pharmacophoriques concordants d'une molécule et le nombre d'atome lourd de cette molécule supérieur à 0,4.
- Un rapport entre le nombre de points pharmacophoriques HBA et HBD concordants d'une molécule et le nombre d'hétéroatomes HBA et HBD de cette molécule supérieure à 0,225.
- Établir au moins une liaison hydrogène avec des résidus distants de moins de 4,5 Å des points de cavité. Ce dernier critère est vérifié au moyen du module IFP du logiciel IChem v4.0.

#### 2.3.3.1.2 Diversité chimique

Les molécules ayant passées le filtre pharmacophorique ont été groupées selon la similitude de leur châssis moléculaire. Ce partitionnement a été effectué avec le logiciel MedChem Studio 4.0<sup>34</sup> qui est actuellement intégré au logiciel Admet Predictor<sup>35</sup>.

Pour le dimère I5, le partitionnement a été effectué avec l'algorithme *topogenic fragments* et les options *weak match criteria*, *fuzzy atoms*, *forced closed rings*, *high homogeneity level* et *medium redundancy level*. Toutes les partitions générées n'ont pas été prises en compte. Ainsi les partitions ayant une prépondérance de score de criblage bas et ayant un squelette contenant un groupement sulfone ont été retirées manuellement. Deux molécules ont été sélectionnées par partition, celle ayant le meilleur score et celle ayant la meilleure solubilité. Ces molécules ont été évaluées manuellement en observant le mode de liaison potentiel. Les molécules non redondantes ayant un score de criblage supérieur à 4,6 ont également été rajoutées aux molécules sélectionnées.

Pour le dimère I56, le partitionnement a également été effectué avec l'algorithme *topogenic fragments* et les options *weak match criteria*, *fuzzy atoms*, *medium homogeneity level* et *medium redundancy level*. La redondance des partitions a été calculée avec l'option *calculate class redundancy* et les partitions ayant une redondance supérieure à 0,6 ont été regroupées. Des familles de classes ont été générées avec l'option *generate Class family*, en utilisant la méthode de fragment et en omettant les singulets, afin de regrouper les partitions similaires. Deux molécules ont été sélectionnées par partition, celle ayant le meilleur score et celle ayant le poids moléculaire le plus faible. Ces molécules ont été évaluées manuellement et les meilleures sélectionnées.

### 3. RÉSULTATS ET DISCUSSION

---

Dans cette partie, je décrirai la sélection d'un site droguable pour les dimères I5 et I56 de CCR5 ainsi que la recherche par criblage virtuel de chimiothèque par une approche pharmacophore de molécules potentiellement capables de se lier à ce site.

### 3.1 RECHERCHE D'UN SITE DROGUABLE

Afin de pouvoir générer les pharmacophores et d'effectuer les criblages virtuels, il a été nécessaire d'étudier l'interface de dimérisation et d'identifier une cavité d'intérêt de chaque dimère.

#### 3.1.1 Analyse des interfaces des dimères I5 et I56

Les dimères I5 et I56 de CCR5 ont été modélisés dans une bicouche lipidique et leurs mouvements ont été simulés dans une boîte d'eau par dynamique moléculaire<sup>18</sup>. Trois trajectoires de dynamique moléculaire d'environ 300 ns ont été générées pour chacun des dimères, délivrant respectivement 1763 et 1675 structures tridimensionnelles pour I5 et I56 pour l'analyse des interactions entre protomères.

L'interface de dimérisation modélisée est pseudo-symétrique dans I5 et I56. Elle implique majoritairement l'hélice 5 du récepteur dans les deux cas. Pour le dimère I5, l'interface implique majoritairement le haut du TM5 et dans une moindre mesure les TM3, TM4 et TM6 (Figure 4 - 2). Au total, 197 interactions entre résidus des deux protomères sont caractérisées. Les trois quarts de ces interactions sont hydrophobes (77% de contact hydrophobe). Parmi les interactions polaires, les liaisons hydrogène sont prédominantes (21% de liaisons hydrogène et 2% d'interactions autres -  $\pi$ /cation, aromatique, ou ionique). En ne considérant que les interactions ayant un taux de conservation supérieur à 1% ces proportions passent à 83% d'interaction hydrophobe, 16% de liaison hydrogène, 1% de liaison aromatique. Sur les 197 interactions caractérisées pour cette interface, seules 19 ont un taux de conservation supérieur à 50% (Tableau 4- 2).



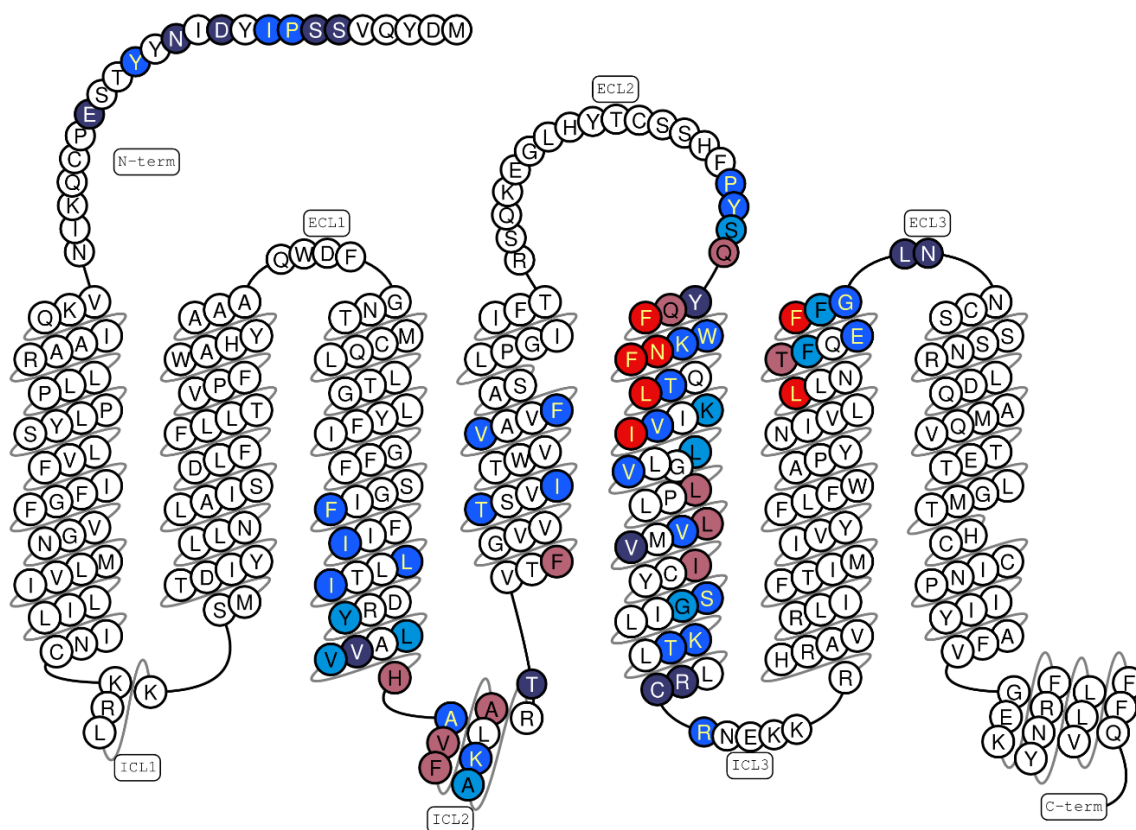


Figure 4 - 2 : Résidus impliqués dans l'interface du dimère I5. Le code couleur indique que le résidu est impliqué dans des interactions plus ou moins fréquentes. Bleu foncé : 0-1%, Bleu clair : 1-25%, Cyan 25-50%, Rose : 50-75%, Rouge 75-100%. Image réalisée à l'aide des outils de GPCRdb.

Tableau 4- 2 : Interactions ayant une fréquence de conservation d'au moins 50% dans les trajectoires de dynamique moléculaire du dimère I5.

Type d'interaction	Résidus de la chaîne A	Résidus de la chaîne B	Fréquence de conservation
Hydrophobe	L196(5.40)	F189(5.33)	0,98
Hydrophobe	L196(5.40)	F193(5.37)	0,98
Hydrophobe	F189(5.33)	F263(6.63)	0,97
Hydrophobe	L196(5.40)	L196(5.40)	0,95
Hydrophobe	L196(5.40)	N192(5.36)	0,88
Hydrophobe	I200(5.44)	L196(5.40)	0,87
Hydrophobe	N192(5.36)	F189(5.33)	0,87
Hydrophobe	I200(5.44)	I200(5.44)	0,86
Hydrophobe	L256(6.56)	F189(5.33)	0,85
Hydrophobe	F263(6.63)	Q186(5.30)	0,68
Hydrophobe	F189(5.33)	Q188(5.32)	0,67
Hydrophobe	F135	F135	0,63
Hydrophobe	F144(4.41)	I212(5.56)	0,61
Hydrophobe	F135	V134	0,60
Hydrophobe	F193(5.37)	L196(5.40)	0,57
Hydrophobe	T259(6.59)	F189(5.33)	0,55
Hydrophobe	F263(6.63)	F189(5.33)	0,53
Hydrophobe	L208(5.52)	L205(5.49)	0,53
Hydrophobe	A136	H132(3.56)	0,51

Certains des résidus impliqués dans les interactions conservées ont été décrits comme impactant la dimérisation<sup>18</sup> ou la fonction de CCR5<sup>36</sup>. Notamment, les résidus Q188(5.32), N192(5.36), L196(5.40), I200(5.44), L205(5.49), L256(6.56) renforcent la dimérisation de CCR5 par pontage covalent lorsqu'ils sont mutés en cystéine. Les résidus L196(5.40), I200(5.44), L205(5.49) compromettent l'export du récepteur à la membrane lorsqu'ils sont mutés en Lysine. La mutation du résidus F189 (5.33) en alanine compromet la liaison de la protéine gp120.

Les interfaces du dimère I5 sont prédites comme étant des interfaces biologiquement pertinentes uniquement dans 23% des structures. Elles sont caractérisées par une moyenne de  $57 \pm 10$  points d'interactions. Le jeu de données utilisé pour entraîner l'outil IChemPIC, comporte des interfaces caractérisées par une moyenne de 60–70 points d'interactions. Parmi ces interfaces, celle définies comme biologiquement pertinente ont une moyenne de  $86 \pm 40$  points d'interactions, tandis que celle définies comme non pertinente, ont une moyenne de  $50 \pm 30$  points. En comparaison, la taille de l'interface du dimère I5 est proche de celles des interfaces non pertinentes. En ne prenant en compte que les structures dont l'interface est prédite biologiquement pertinente, le nombre d'interactions interfaciales diminue à 157. Néanmoins, les interactions ayant un taux de conservation de 50% sont les mêmes, avant et après filtrage, avec des fréquences quasiment identiques (3% de variation en moyenne). La seule exception est l'interaction entre les résidus L208(5.52) et L205(5.49) qui passe d'un taux de conservation de 53% à 47%.

L'analyse des interactions entre les protomères du dimère I56 montre que son interface implique l'ensemble des TM5 et TM6 (Figure 4 - 3). Au total, 324 interactions sont caractérisées. Comme dans le dimère I5, les interactions hydrophobes sont majoritaires (77% de contacts hydrophobes) et les liaisons hydrogène représentent la plus grande partie des interactions polaires (20% de liaisons hydrogène et 3% d'interactions autres -  $\pi$ /cation, aromatique, ou ionique). En ne considérant que les interactions ayant un taux de conservation supérieur à 1% ces proportions passent à 84% d'interactions hydrophobes, 14% de liaisons hydrogène et 2% d'interactions autres. Au total, 50 interactions ont un taux de conservation supérieur à 50% (Tableau 4- 3). En comparant avec le dimère I5, il n'y a aucun recouvrement des interactions ayant un taux de conservation supérieure à 50%, malgré le fait qu'une partie des résidus impliqués soit les mêmes. Par exemple, le résidu L196(5.40) qui est impliqué dans

les deux dimères avec une fréquence de plus de 75%, interagit avec le résidu F189(5.33) dans le dimère I5, tandis qu'il interagit avec le résidu F260(6.60) dans le dimère I56.

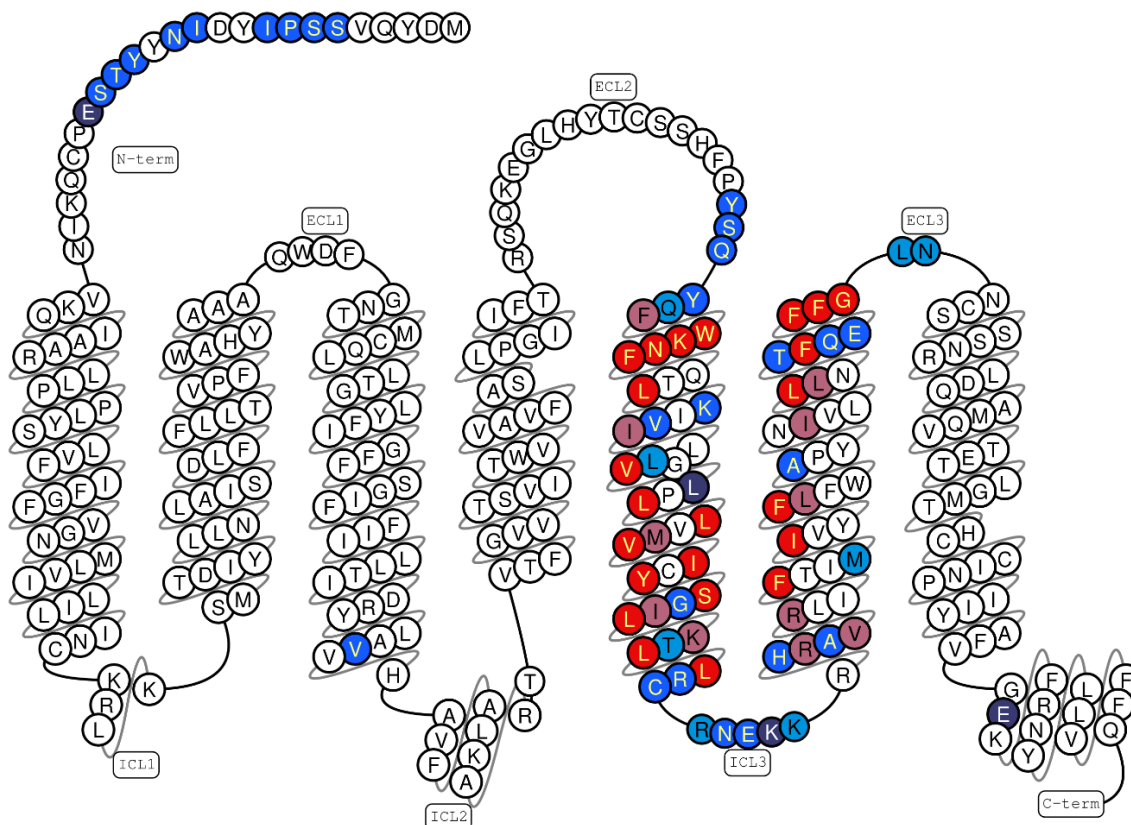


Figure 4 - 3 : Résidus impliqués dans l'interface du dimère I56. Le code couleur indique que le résidu est impliqué dans des interactions plus ou moins fréquentes. Bleu foncé : 0-1%, Bleu clair : 1-25%, Cyan 25-50%, Rose : 50-75%, Rouge 75-100%. Image réalisée à l'aide des outils de GPCRdb

Certains des résidus impliqués dans les interactions conservées de I56 ont été décrits comme impactant la dimérisation ou la fonction de CCR5. Notamment, les résidus N192(5.36), L196(5.40), I200(5.44), V204(5.48), L256(6.56), F260(6.60) renforcent la dimérisation de CCR5 par pontage covalent lorsqu'ils sont mutés en cystéine<sup>18</sup>. Le résidu I200(5.44) compromet l'export du récepteur à la membrane lorsqu'il est muté en Lysine, les mutations des résidus F189(5.33), W190(5.34) en alanine compromettent la liaison de la gp120 et de la chimiokine CCL3<sup>36</sup> et le résidu K191(5.35) est décrit par le site GPCRdb comme important pour la liaison de divers ligands du récepteur (TAK779, ancroviroc, aplaviroc, ...) <sup>37,38</sup>.

Tableau 4- 3 : Interactions, ayant une fréquence de conservation d'au moins 50% dans les trajectoires de dynamique moléculaire du dimère I56.

Type d'interaction	Résidus de la chaîne A	Résidus de la chaîne B	Fréquence de conservation
Hydrophobe	L207(5.51)	V211(5.55)	0,99
Hydrophobe	L218(5.62)	L221(5.65)	0,97
Hydrophobe	F238(6.38)	L218(5.62)	0,96
Hydrophobe	Y214(5.58)	L218(5.62)	0,96
Liaison hydrogène	N192(5.36)	F263(6.63)	0,95
Hydrophobe	F260(6.60)	L196(5.40)	0,94
Hydrophobe	L196(5.40)	F264(6.64)	0,94
Hydrophobe	S215(5.59)	F238(6.38)	0,93
Hydrophobe	F264(6.64)	L196(5.40)	0,92
Hydrophobe	F260(6.60)	L256(6.56)	0,92
Hydrophobe	F193(5.37)	F264(6.64)	0,92
Hydrophobe	N192(5.36)	F264(6.64)	0,91
Hydrophobe	L218(5.62)	L218(5.62)	0,91
Hydrophobe	V211(5.55)	Y214(5.58)	0,88
Hydrophobe	W190(5.34)	G265(6.65)	0,86
Hydrophobe	F264(6.64)	F193(5.37)	0,84
Hydrophobe	V204(5.48)	L207(5.51)	0,83
Hydrophobe	L221(5.65)	L221(5.65)	0,82
Hydrophobe	K191(5.35)	F263(6.63)	0,81
Hydrophobe	L208(5.52)	F245(6.45)	0,79
Hydrophobe	I212(5.56)	I242(6.42)	0,79
Hydrophobe	L207(5.51)	L207(5.51)	0,78
Hydrophobe	L196(5.40)	L256(6.56)	0,78
Hydrophobe	L256(6.56)	V204(5.48)	0,77
Hydrophobe	L221(5.65)	L222(5.66)	0,76
Hydrophobe	F260(6.60)	F260(6.60)	0,76
Hydrophobe	N192(5.36)	F263(6.63)	0,76
Hydrophobe	L218(5.62)	I217(5.61)	0,74
Hydrophobe	F245(6.45)	V211(5.55)	0,72
Hydrophobe	V211(5.55)	V211(5.55)	0,72
Hydrophobe	M210(5.54)	V211(5.55)	0,71
Hydrophobe	I200(5.44)	I253(6.53)	0,71
Hydrophobe	I253(6.53)	L208(5.52)	0,69
Hydrophobe	V234(6.34)	L222(5.66)	0,68
Hydrophobe	V211(5.55)	M210(5.54)	0,66
Hydrophobe	L222(5.66)	R232(6.32)	0,65
Hydrophobe	F263(6.63)	F263(6.63)	0,65
Hydrophobe	I217(5.61)	L221(5.65)	0,64
Hydrophobe	L196(5.40)	L257(6.57)	0,63
Liaison hydrogène	F189(5.33)	G265(6.65)	0,63
Hydrophobe	V211(5.55)	F238(6.38)	0,62
Hydrophobe	F263(6.63)	F260(6.60)	0,62
Hydrophobe	I212(5.56)	F238(6.38)	0,61
Liaison hydrogène	F263(6.63)	N192(5.36)	0,61
Hydrophobe	R232(6.32)	L222(5.66)	0,60
Hydrophobe	L208(5.52)	L207(5.51)	0,56
Hydrophobe	K219(5.63)	R235(6.35)	0,53
Hydrophobe	F264(6.64)	N192(5.36)	0,53
Hydrophobe	L208(5.52)	L246(6.46)	0,52
Hydrophobe	S215(5.59)	L218(5.62)	0,51

Au sein des trajectoires du dimère I56, les interfaces sont prédites comme étant biologiquement pertinentes dans 88% des structures. Le nombre moyen de points d'interactions de l'interface est de  $120 \pm 20$  points, dénotant une interface de très grande taille. En ne prenant en compte que les structures biologiquement pertinentes pour caractériser l'interface, le nombre d'interactions reste de 324. Si on considère celles ayant un taux de conservation de plus de 50%, la variation du taux de conservation est très faible ( $\pm 1\%$  en moyenne).

Pour chacun des dimères I5 et I56, nous avons recherché une structure représentative de l'interface. Les ensembles de structures ont d'abord été partitionnés par la similitude structurale de l'interface pour extraire cinq structures dont la meilleure a été choisie pour la recherche d'un site pour la conception d'un pharmacophore. Pour le dimère I5, la partition des trajectoires divise le jeu de structures en cinq ensembles de 569, 562, 459, 166 et 1 structures, respectivement (Figure 4 - 4). Les interfaces des structures de ces cinq ensembles sont très proches. Le RMSD calculé entre deux structures représentatives en considérant les atomes de la chaîne principale des résidus impliqués dans des interactions présentes dans plus de 50% des structures est compris entre 0,75 Å et 1,81 Å (moyenne de 1,3 Å). À titre de comparaison, le RMSD calculé sur des atomes de la chaîne principale de tous les résidus des TM (résidus 22 à 58, 63 à 92, 97 à 132, 141 à 167, 186 à 224, 228 à 265 et 268 à 300) est légèrement supérieur, de 1,12 Å à 1,95 Å (moyenne de 1,50 Å). Si les cinq structures représentatives sont similaires, elles diffèrent néanmoins suffisamment pour donner des prédictions contrastées en ce qui concerne la pertinence biologique de l'interface. En effet les interfaces des cinq structures sont caractérisées par des nombres de points d'interactions différents, respectivement 82, 51, 54, 48 et 44 (15, 11, 16, 16 et 14 avec un taux de conservation supérieur à 50%). Seules deux des structures présentent une interface évaluée comme biologiquement pertinente. Parmi elles, celle retenue présente un nombre de points d'interaction maximal (82 contre 54) et représente une majorité de structures produites après 50 ns de simulation.

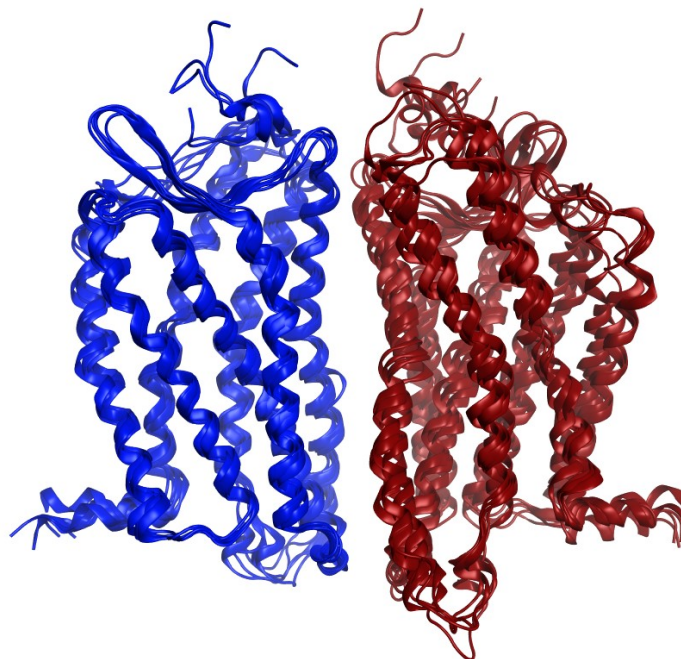


Figure 4 - 4 : Superposition des structures représentatives du dimère I5. La superposition effectuée est sur les C $\alpha$  des résidus impliqués dans des interactions présentes dans plus de 50% des structures de la trajectoire de dynamique moléculaire. La coloration est effectuée en fonction du numéro de résidus dans la chaîne du récepteur, partant du bleu et allant jusqu'au rouge. Le protomère de gauche correspond à la chaîne A tandis que le protomère de droite correspond à la chaîne B.

Pour le dimère I56, la partition des trajectoires divise le jeu de structures en cinq ensembles de 748, 441, 215, 186 et 80 structures, respectivement (Figure 4 - 5). Le RMSD calculé entre deux structures représentatives en considérant les atomes de la chaîne principale des résidus impliqués dans des interactions présentes dans plus de 50% des structures est compris entre 1,15 Å et 2,11 Å (moyenne de 1,62 Å). À titre de comparaison, le RMSD calculé sur des atomes de la chaîne principale de tous les résidus des TM est compris entre 1,51 Å et 2,41 Å (moyennes de 2,01 Å). À noter que ces valeurs sont sensiblement supérieures à celles observées pour I5, suggérant une plus grande flexibilité de l'interface. Cependant, les interfaces des cinq structures représentatives sont toutes évaluées comme biologiquement pertinentes. Le nombre de points d'interactions caractérisant ces structures sont respectivement de 106, 103, 146, 111 et 103. Au sein des cinq structures représentatives, on retrouve respectivement 40, 30, 42, 39 et 35 des 50 interactions ayant un taux de conservation supérieur à 50%. En conclusion, la structure représentative considérée pour ce dimère est la représentant une majorité de structure produite après 50 ns de simulation et représentant des structures observées issues de plusieurs répétitions (748 structures au total).



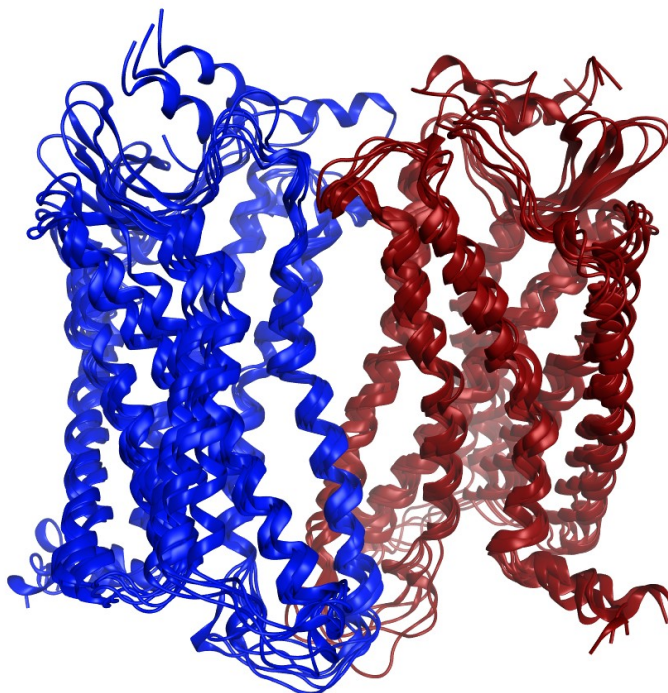


Figure 4 - 5 : Superposition des structures représentatives du dimère I56. Superposition effectuée sur les C $\alpha$  des résidus impliqués dans des interactions présentes dans plus de 50% des structures de la trajectoire de dynamique moléculaire. La coloration est effectuée en fonction du numéro de résidus dans la chaîne du récepteur, partant du bleu et allant jusqu'au rouge. Le protomère de gauche correspond à la chaîne A tandis que le protomère de droite correspond à la chaîne B.

### 3.1.2 Analyse des cavités à l'interface des dimères

La sélection d'un site d'intérêt pour le criblage virtuel a été effectuée pour chaque dimère en étudiant toutes les cavités présentes dans la structure représentative du dimère, directement à l'interface ou à proximité de celle-ci.

Les cavités générées à partir de la structure représentative du dimère I5 sont au nombre de 11 (Figure 4 - 6). Elles sont situées à l'interface des deux protomères, au niveau de la cavité transmembranaire et de la cavité intracellulaire. Leurs caractéristiques sont représentées dans le Tableau 4- 4. La cavité qui a été considérée comme pertinente est la numéro 3. Les cavités 7 et 23 ont été écartées à cause de leur faible indice de droguabilité ; Les cavités 6, 8 et 11 à cause de leur positionnement proche du site de liaison de la protéine G ou trop exposé au lipide de la membrane ; Les cavités 1 et 2 à cause de leur grand volume et de leur recouvrement avec la cavité transmembranaire ; les cavité 5 et 9 à cause de leur faible nombre de points polaires. Enfin la cavité 12 a été écartée car pas assez fréquente dans les structures

simulées (présente dans 25% des structures de la trajectoire pour un minimum de similitude de Tanimoto combo de 0,1).

Du fait de la taille de la cavité sélectionnée et de la présence de points trop exposés aux lipides, il a été décidé de ne considérer qu'une partie de celle-ci (Figure 4 - 7). Cette sélection manuelle, diminue le nombre de points de cavité de 356 à 181. Les points hydrophobes, aromatiques, HBA, HBD, HBA/HBD, de charge négative et de charge positive de la cavité réduite sont respectivement au nombre de 117, 29, 2, 19, 5 et 0. La cavité réduite est présente dans 75% des structures des trajectoires (similitude de Tanimoto combo minimum de 0,1).

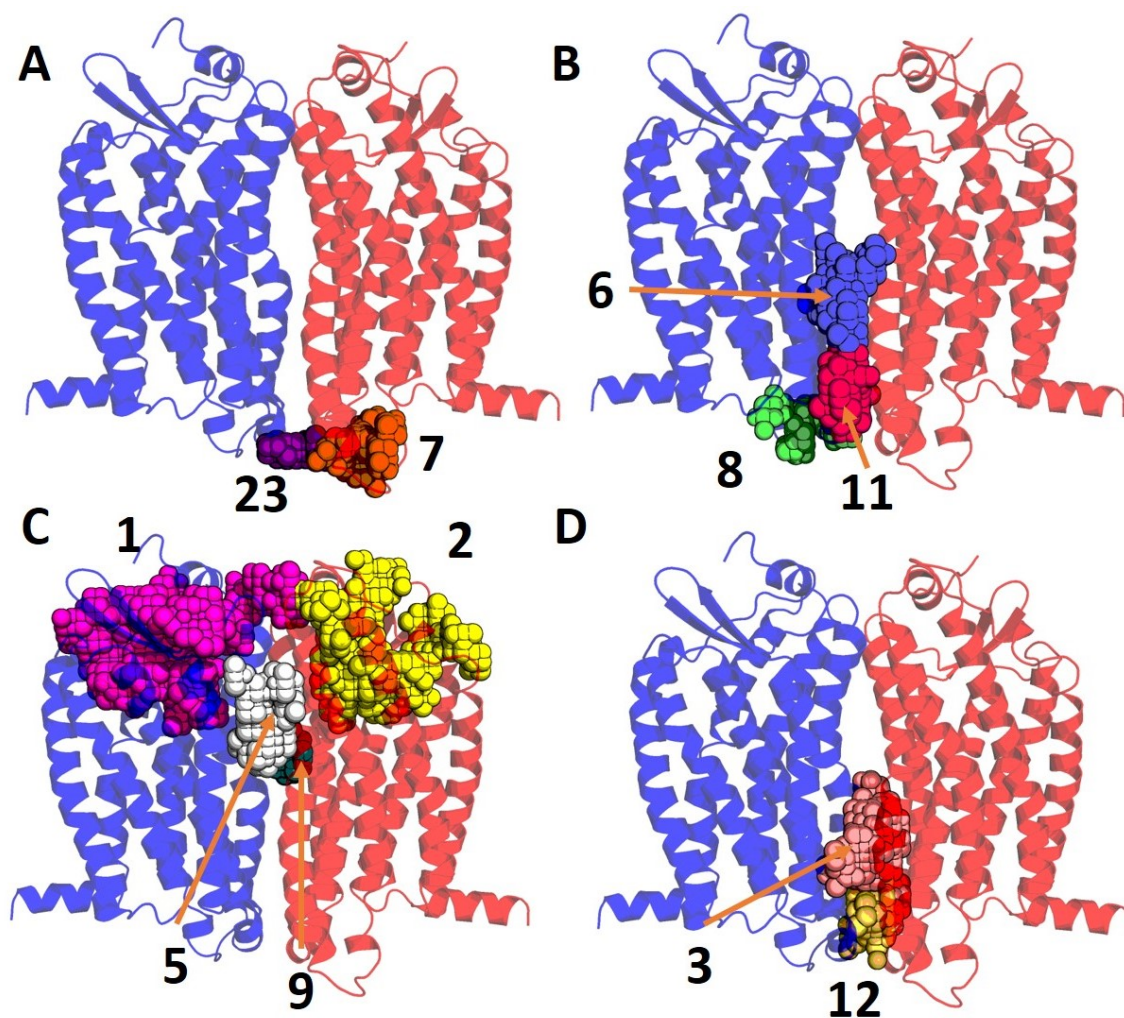


Figure 4 - 6 : Cavités de la structure représentative du dimère 15. A) Cavités écartées d'après le critère de drugabilité. B) Cavités écartées d'après le critère visuel de localisation de la cavité. C) Cavités écartées d'après le critère de taille et de polarité. D) Cavités restantes. Le protomère bleu correspond à la chaîne A (résidus 1-306) et le protomère rouge à la chaîne B (résidus 307-3014).



Tableau 4- 4 : Caractéristiques des cavités du dimère I5

Numéro de cavité	Nombre total de points	Volume (Å <sup>3</sup> )	Nombre de points hydrophobes	Nombre de points aromatiques	Nombre de points HBA, HBD, et HBA/HBD	Nombre de points de charge négative	Nombre de points de charge positive	Indice de droguabilité	Critère non conforme (cf 2.2.3)
1	783	2642.62	221	100	269	36	38	0.1350	3
2	678	2288.25	215	59	228	24	50	0.3599	3
3	356	1201.5	206	48	53	13	0	1.780	
5	169	570.375	80	49	6	3	0	1.540	3
6	159	536.625	125	9	11	0	0	2.088	2
7	141	475.875	41	0	52	26	13	-0.4270	1
8	113	381.375	48	5	25	16	5	0.1218	2
9	110	371.25	63	20	8	0	0	1.604	3
11	91	307.125	39	11	27	6	0	0.3378	2
12	82	276.75	39	23	20	0	0	1.279	4
23	36	121.5	20	2	11	1	0	-0.3877	1

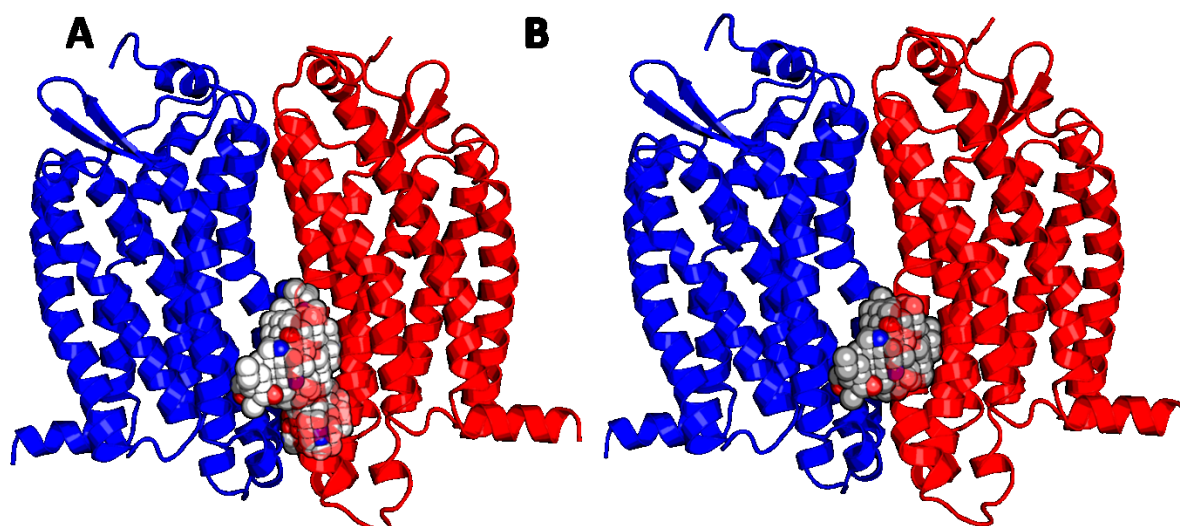


Figure 4 - 7 : (A) la cavité sélectionnée à partir de la structure représentative du dimère I5. (B) la cavité réduite. Les points de cavité blancs correspondent aux points hydrophobe et aromatique, les points de cavité bleus aux points HBD et les points de cavité rouges aux points HBA. Le protomère bleu correspond à la chaîne A (résidus 1-306) et le protomère rouge à la chaîne B (résidus 307-3014).

Le site englobant les résidus à moins de 4,5 Å des points de cavité est composée de 20 résidus de la chaîne A et 19 résidus de la chaîne B (Figure 4 - 8). La fluctuation des atomes de la chaîne principale de ces résidus par rapport à la structure de référence est comprise entre 0,82 Å et 2,17 Å (Figure 4 - 9, p149). Les résidus les plus fluctuants sont les résidus L208(5.52) et V209(5.59) de la chaîne A et F144(4.41), T148(4.45), I151(4.48) et I212(5.56) de la chaîne B. À l'exception du résidu I151(4.48), ces résidus sont présents au sein de la cavité dans les deux chaînes mais fluctuent de manière différente indiquant que les protomères ont des comportements différents l'un de l'autre. Les résidus Y68(2.42) et Y127(3.51) qui sont ciblés

par la suite pour le pharmacophore, ont une fluctuation de 0,94 Å et 0,85 Å pour la chaîne A et 1,25 Å et 1,19 Å pour la chaîne B.

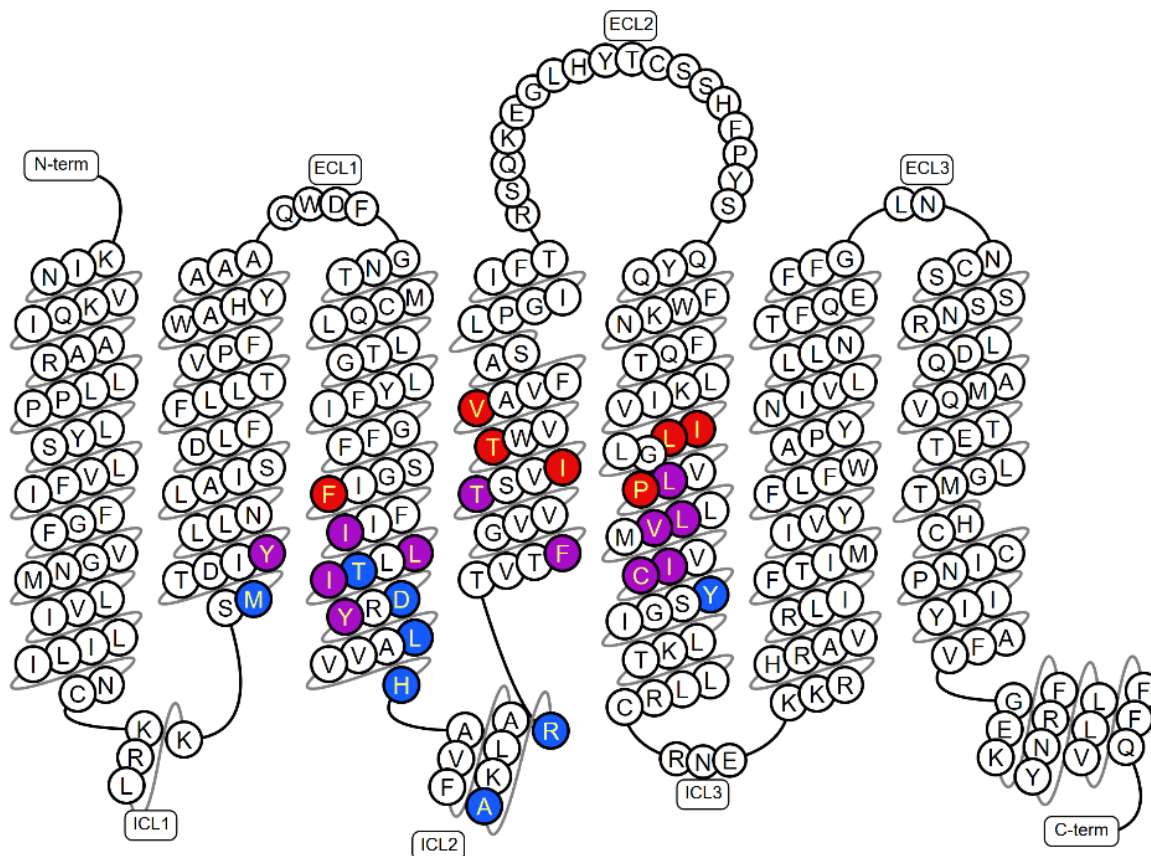


Figure 4 - 8 : Résidus à 4,5 Å des points de la cavité sélectionnée pour le dimère I5. Les résidus colorés en bleu sont issus de la chaîne A du dimère, les résidus en rouge, de la chaîne B et les résidus en violet des deux chaînes. Source : [http://gpcrdb.org/protein/ccr5\\_human/](http://gpcrdb.org/protein/ccr5_human/)

D'après les données de la littérature résumées par le site GPCRdb, la mutation des résidus Y68(2.42), F117(4.41), L121(3.45), T123(3.47), D125(3.49), Y127(3.51) et F144(4.41) n'impacte pas de manière significative la liaison de ligands "drug-like" à CCR5. Par contre les résidus H132(3.56), A139(34.56) et R140(34.57) peuvent être impliqués dans la liaison à la protéine Gas.

Parmi les 27 résidus de cette cavité, dix résidus sont conservés à plus de 50% pour la sous-famille des récepteurs aux chimiokines (Y68(2.42) 70%, D125(3.49) 87%, Y127(3.51) 91%, H132(3.56) 57%, R140(34.57) 57%, L205(5.49) 61%, P206(5.50) 91%, L208(5.52) 57%,

C213(5.57) 83% et Y214(5.58) 96%). Parmi ceux-ci seuls quatre sont conservés pour les récepteurs de classe A (D125(3.49) 64%, Y127(3.51) 66%, P206(5.50) 79% et Y214(5.58) 72%). Seuls Y68(2.42), Y127(3.51) et H132(3.56) exposent leur chaîne latérale à la surface de la cavité.

En conclusion, nous avons sélectionné pour le dimère I5 une cavité à l'interface entre les deux protomères au niveau de la région intracellulaire de la partie transmembranaire.

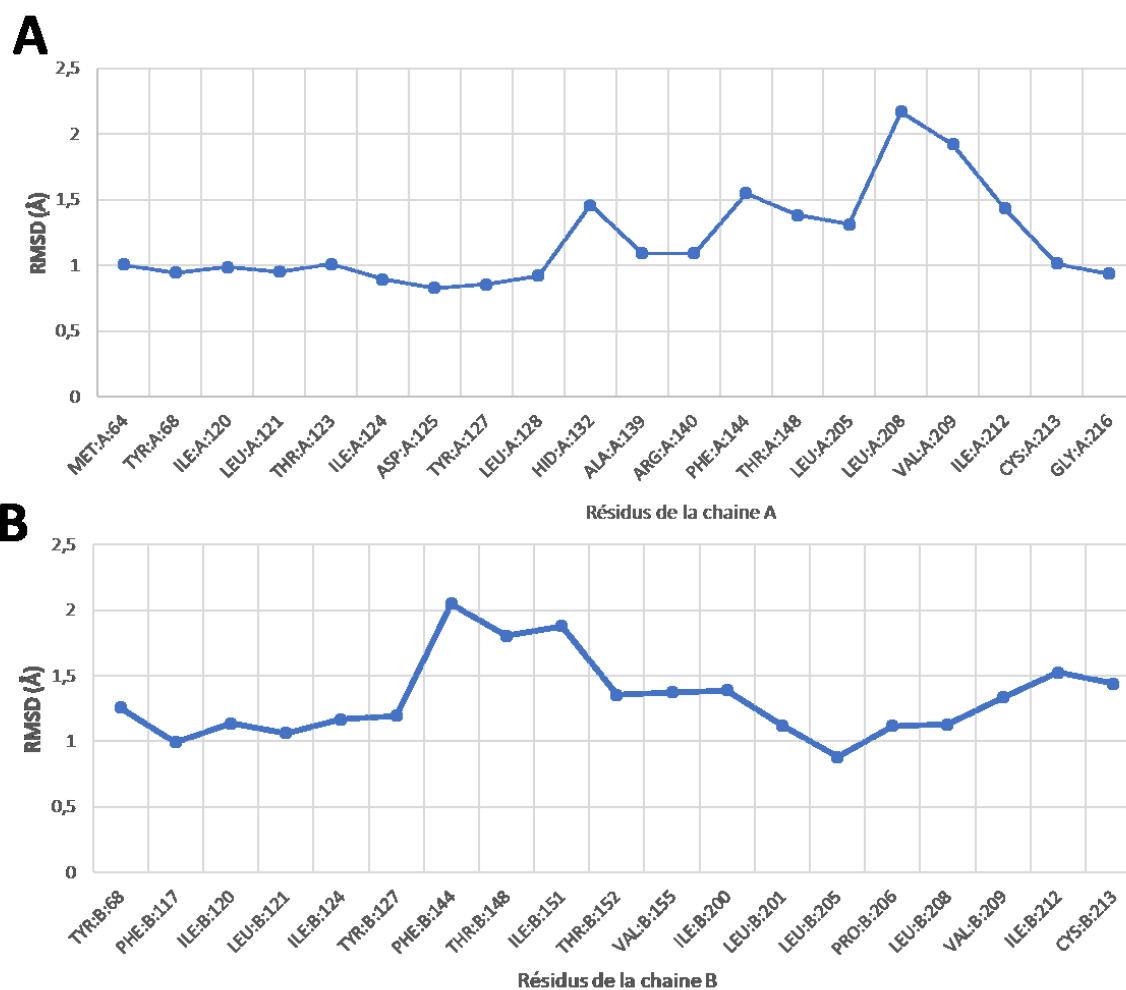


Figure 4 - 9 : RMSD par résidu par rapport à la structure représentative des atomes de la chaîne principale des résidus présents à moins de 4,5 Å des points de la cavité réduite du dimère I5. (A) résidus de la chaîne A, (B) résidus de la chaîne B.

En ce qui concerne le dimère I56, les cavités générées à partir de la structure représentative sont au nombre de 18 (Figure 4 - 10). Elles sont situées de part et d'autre de l'interface des deux protomères, au niveau de la cavité transmembranaire et de la cavité intracellulaire. Leurs caractéristiques sont représentées dans le Tableau 4- 5 (p151). La cavité qui a été identifiée

comme pertinente est la numéro 2. Les cavités 9, 19, 25, 27, 31 et 32 ont été écartées à cause faible droguabilité ; les cavités 3, 4, 5, et 6 à cause de leur positionnement dans la cavité intracellulaire ; la cavité 1 à cause de son grand volume ; les cavités 17, 18, 20 et 21 à cause de leur faible polarité ; et les cavités 11 et 24 à cause de leur faible hydrophobicité.

Du fait de la taille de la cavité sélectionnée et pour ne pas prendre en compte la cavité transmembranaire où se lie l'antiviral MVC, il a été décidé de n'en considérer qu'une partie (Figure 4 - 11). Cette sélection manuelle diminue le nombre de points de cavité de 414 à 133. Les points hydrophobes, aromatiques, HBA, HBD, HBA/HBD, de charge négative et de charge positive de la cavité réduite sont respectivement de 46, 3, 15, 33, 6, 6 et 13. La cavité réduite est présente dans une proportion importante de structure simulée, i.e. dans 93% des structures de la trajectoire (similitude minimum de Tanimoto combo de 0,1).

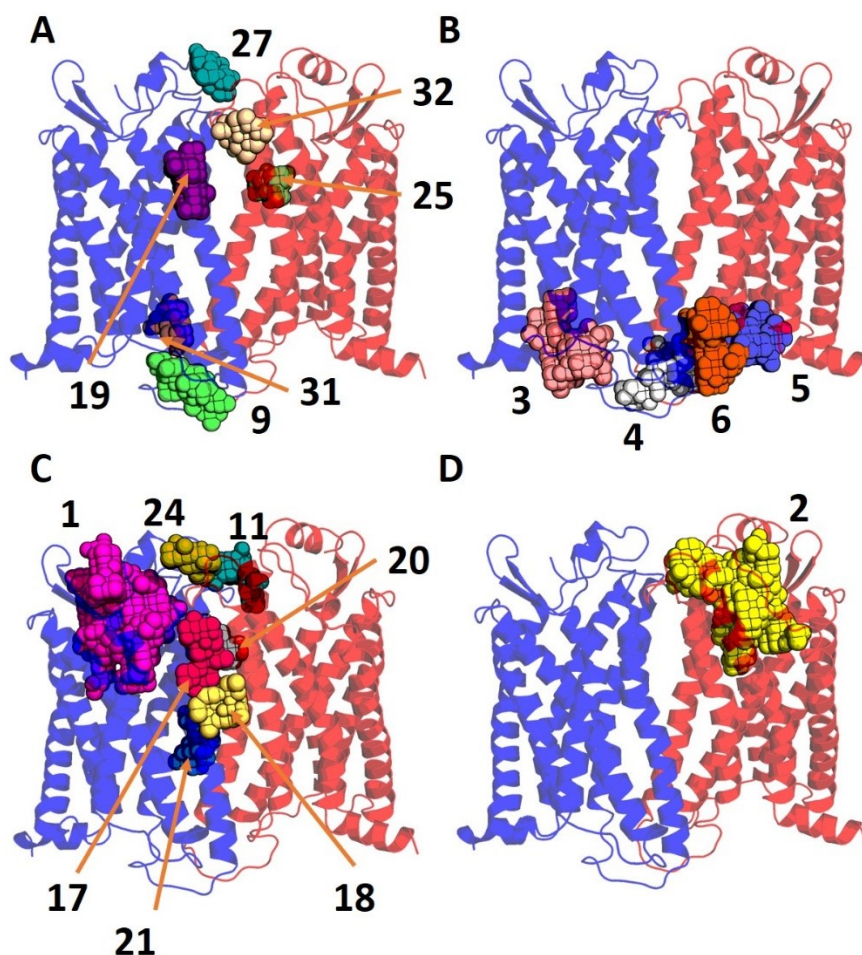


Figure 4 - 10 : Cavités de la structure représentative du dimère I56. A) Cavités écartées d'après le critère de droguabilité. B) Cavités écartées d'après le critère visuel de localisation de la cavité. C) Cavités écartées d'après le critère de taille et de polarité. D) Cavités restantes. Le protomère bleu correspond à la chaîne A (résidu 1-306) et le protomère rouge à la chaîne B (résidu 307-614).



Tableau 4- 5 : Caractéristiques des cavités du dimère I56

Numéro de cavité	Nombre total de points	Volume (Å <sup>3</sup> )	Nombre de points hydrophobe	Nombre de points aromatique	Nombre de points HBA, HBD, et HBA/HBD	Nombre de points de charge négative	Nombre de points de Charge Positive	Indice de ligandabilité	Critère non conforme (cf 2.2.3)
1	549	1852.88	197	91	152	16	22	0.6143	3
2	414	1397.25	116	39	157	24	31	1.109	
3	204	688.5	126	4	28	21	4	1.370	2
4	158	533.25	68	0	58	10	7	0.5629	2
5	146	492.75	65	6	18	14	14	0.3532	2
6	133	448.875	64	13	17	18	0	0.6255	2
9	78	263.25	23	0	22	14	1	-0.7600	1
11	74	249.75	14	15	35	0	2	0.0409	3
17	57	192.375	39	5	4	1	0	0.7223	3
18	57	192.375	43	2	2	0	0	0.6884	3
19	56	189	23	13	6	12	0	-0.6680	1
20	54	182.25	28	15	5	0	0	0.4510	3
21	54	182.25	41	5	4	0	0	0.8168	3
24	51	172.125	14	7	18	0	1	0.01183	3
25	48	162	26	4	11	4	0	-0.01385	1
27	46	155.25	9	17	8	0	4	-0.3003	1
31	38	128.25	17	5	6	8	0	-0.7681	1
32	37	124.875	13	0	18	0	0	-0.5045	1

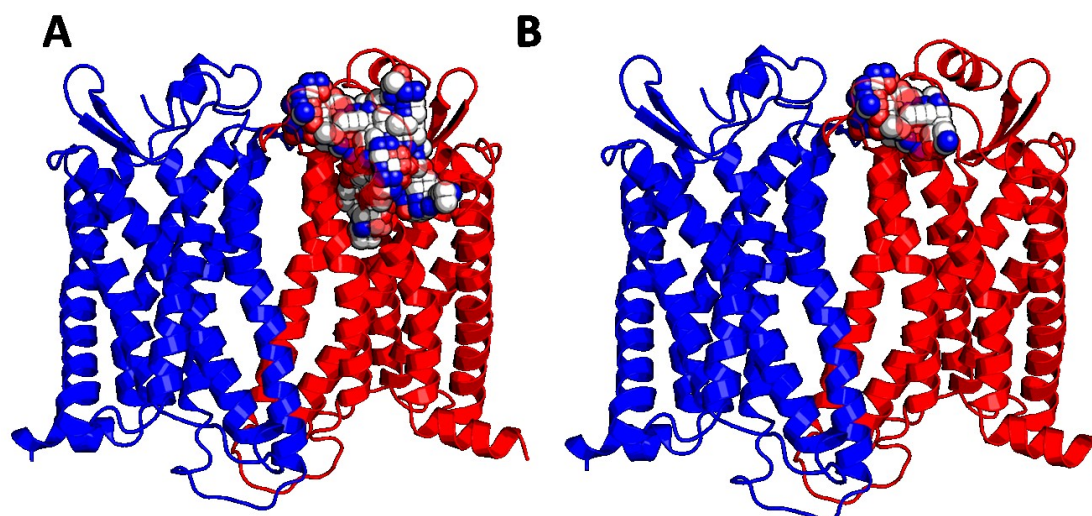


Figure 4 - 11 : (A) la cavité sélectionnée à partir de la structure représentative du dimère I5. (B) la cavité réduite. Les points de cavité blancs correspondent aux points hydrophobes et aromatiques, les points de cavité bleus aux points HBD et les points de cavité rouges aux points HBA. Le protomère bleu correspond à la chaîne A (résidus 1-306) et le protomère rouge à la chaîne B (résidus 307-3014).

Le site correspondant aux résidus distants de moins de 4,5 Å des points de cavités est composé d'un résidu de la chaîne A et de 67 résidus de la chaîne B (Figure 4 - 12). Le RMSD des atomes de la chaîne principale des résidus du site est compris entre 0,57 et 3,83 Å (Figure 4 - 13). Les résidus les plus fluctuants correspondent à des régions ne contenant pas ou peu d'éléments de structure secondaire, c'est-à-dire, l'extrémité N-terminale et le début du TM1 (résidus

14,15, 19-25), une partie de ECL2 (résidus 171-176) et de ECL3 (résidus 266-269) et le haut de TM 6 et de TM 7 (résidus 263 et 272). Les résidus C20, K22(1.24), Y187(5.31), N192(5.36), N258(6.58) et T259(6.59), qui exposent des groupements pharmacophoriques dans la cavité, ont une fluctuation de 2,76 Å, 2,29 Å, 1,10 Å, 0,85 Å, 1,12 Å, et 1,40 Å.

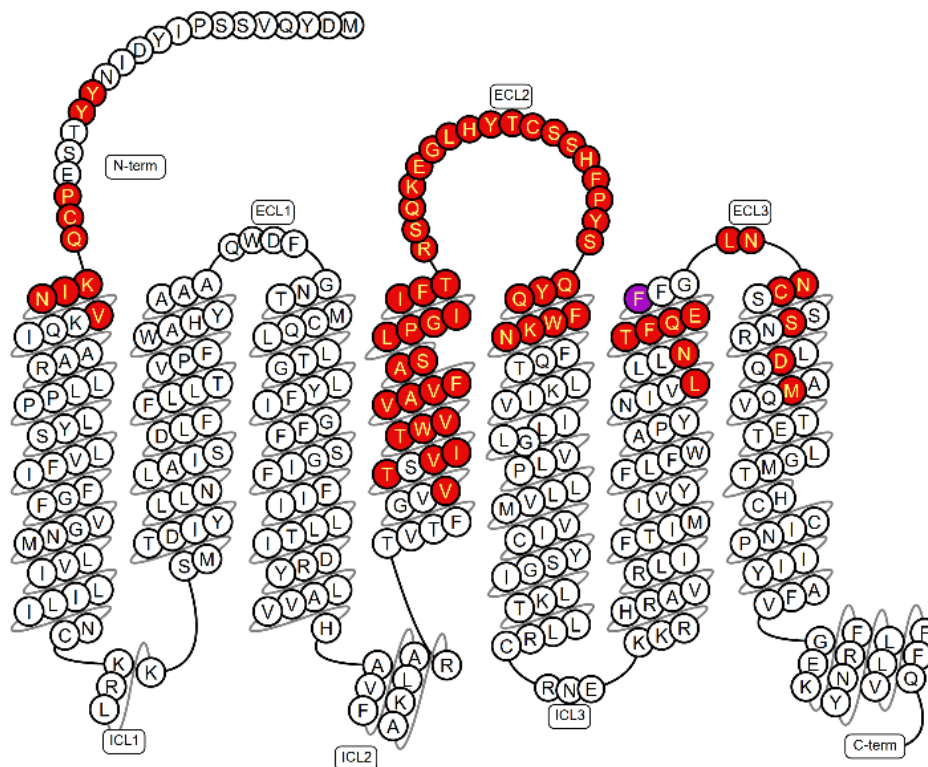


Figure 4 - 12 : Résidus à 4,5 Å des points de la cavité sélectionnée pour le dimère I56. Les résidus colorés en bleu sont issus de la chaîne A du dimère, les résidus en rouge, de la chaîne B et les résidus en violet des deux chaînes. Source : [http://gpcrdb.org/protein/ccr5\\_human/](http://gpcrdb.org/protein/ccr5_human/).

D'après les données de la littérature résumées par le site GPCRdb, la mutation des résidus G163(4.60), T177, C178, S180, S185, Y187(5.31), W190(5.34), K191(5.35) impacte de manière significative la liaison de ligand "drug-like". De plus, les résidus K22(1.24), G163(4.60), E172, S179, S180 sont impliqués dans la liaison de la chimiokine antagoniste [5P7]CCL5, le résidu L255 dans la liaison de MVC, et les résidus Y14, Y15, C20-K22, E172, C178, H181 -Y184, F264(6.64), L255(6.55) et D276(7.32) dans la liaison de la gp120<sup>39</sup>.

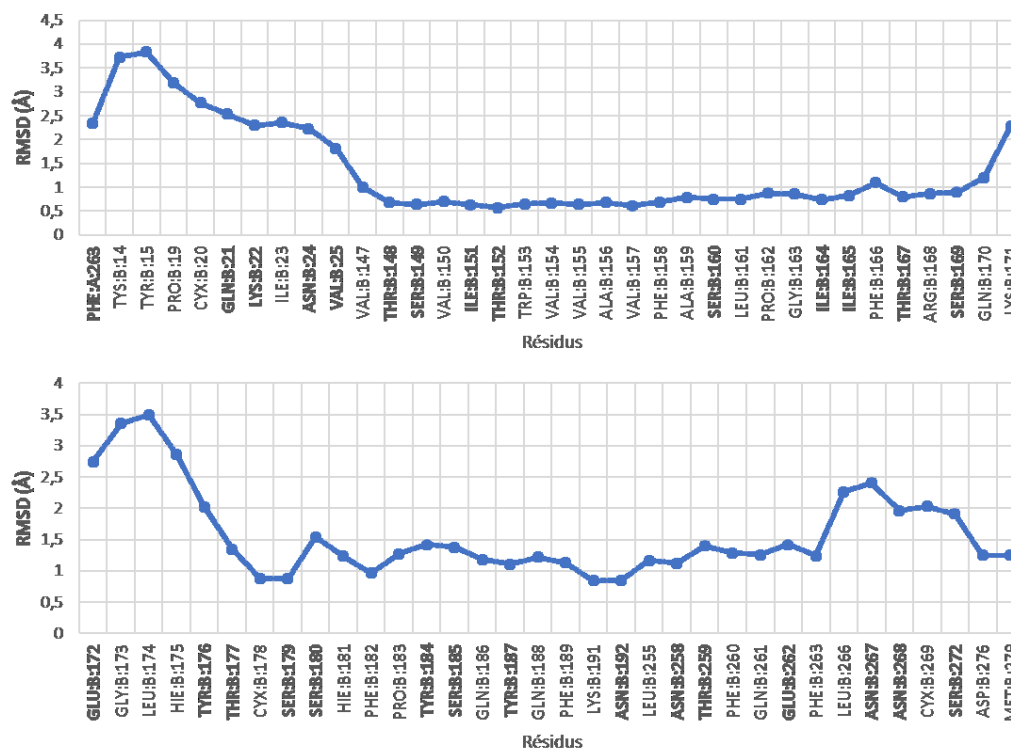


Figure 4 - 13 : RMSD par résidu par rapport à la structure représentative des atomes de la chaîne principale des résidus présents à moins de 4,5 Å des points de la cavité réduite du dimère 156.

Huit des résidus du site sont conservés à plus de 50% pour la sous-famille des récepteurs aux chimiokines : W153(4.50) 100%, A156(4.53) 61%, S160(4.57) 52%, P162(4.59) 91%, G166(4.63) 65%, W190(5.34) 65%, T259(6.59) 57% et C269(7.25) 96%. Parmi eux seuls deux sont conservés pour les récepteurs de classe A (W153(4.50) 96% et P162(4.59) 74%). Seuls T259(6.59) et C269(7.25) exposent leur chaîne latérale à la surface de la cavité.

### 3.2 CRIBLAGE VIRTUEL

Le criblage virtuel basé sur la structure est couramment utilisé afin d'identifier de nouvelles molécules actives. Dans le cas des RCPGs, le criblage virtuel peut se révéler compliqué. D'une part, il nécessite la structure de la cible étudiée ce qui, pour les RCPGs, réduit le champ d'étude possible à 52 récepteurs de structure 3D connues. D'autre part, la plasticité conformationnelle des RCPGs rend plus difficile la sélection d'un site d'intérêt. Malgré cela, de nombreux exemples dans la littérature montrent qu'il est possible d'identifier des ligands par criblage virtuel de chimiothèques en utilisant des structures cristallographiques de RCPG<sup>40,41</sup>, ou des modèles par homologie de RCPG. Notre équipe a été l'une des premières à utiliser une telle

approche pour identifier des agonistes non-peptidique du récepteur CCR5<sup>42</sup>. Depuis, la pertinence de l'utilisation de modèles par homologie ou des structures cristallographiques a été largement discutée. Une étude récente compare les performances en criblage virtuel de structures cristallographiques et de modèles pour 19 RCPGs des classes A et B. Elle révèle que, si la capacité prédictive des structures cristallographiques surpassent généralement celles des modèles, 15 des 38 modèles évalués conduisent des résultats comparables voire meilleurs que ceux obtenus en faisant appel aux structures cristallographiques<sup>43</sup>.

Dans le cas présent, nous avons choisi d'utiliser une méthode de criblage par pharmacophore. Notre choix a été motivé par l'absence de ligands ciblant l'interface dimérique de CCR5, excluant ainsi toutes les méthodes basées sur le ligand, mais également par la nécessité d'utiliser une méthode ayant un temps de calcul et de traitement réduit, rejetant ainsi les méthodes de docking. L'utilité des simulations de dynamique moléculaire pour la conception de pharmacophores est discutée depuis quelques années<sup>44</sup>. Le gain de tels pharmacophores a récemment été évalué sur le récepteur activé par les proliférateurs de peroxyosomes. Cette étude compare les résultats de criblage virtuelle par docking utilisant les pharmacophores issues de structures expérimentales à ceux utilisant les pharmacophores issue de dynamique moléculaire et semble indiquer une augmentation du taux de touche initiale des criblages pour ces derniers<sup>45</sup>. À notre connaissance aucun criblage virtuel de cavités interfaciales de RCPG n'a encore été répertorié dans la littérature.

### 3.2.1 Description du pharmacophore

Un pharmacophore a été conçu pour chaque dimère à partir du site d'intérêt décrit ci-dessus. Pour le dimère I5, les 181 points de la cavité réduite ont été transformés en un pharmacophore contenant 38 points hydrophobes, un point HBA/HBD, deux points HBA, 19 points HBD, deux points aromatiques et 44 sphères d'exclusion (Figure 4 - 14 (A)). Une sélection manuelle des points a conduit à un pharmacophore contenant huit points pharmacophoriques (Figure 4 - 14, (B)). Au final, trois résidus sont considérés comme importants dans ce pharmacophore car en interaction directe avec les molécules. Ces résidus sont la tyrosine Y68(2.42) qui est ciblée par un point pharmacophorique HBA, la tyrosine Y127(3.51) qui est ciblée par trois points pharmacophoriques, un de type HBA, un de type HBD et un de type aromatique, et la phénylalanine F117(3.41) qui est dans le voisinage d'un point hydrophobe/aromatique.



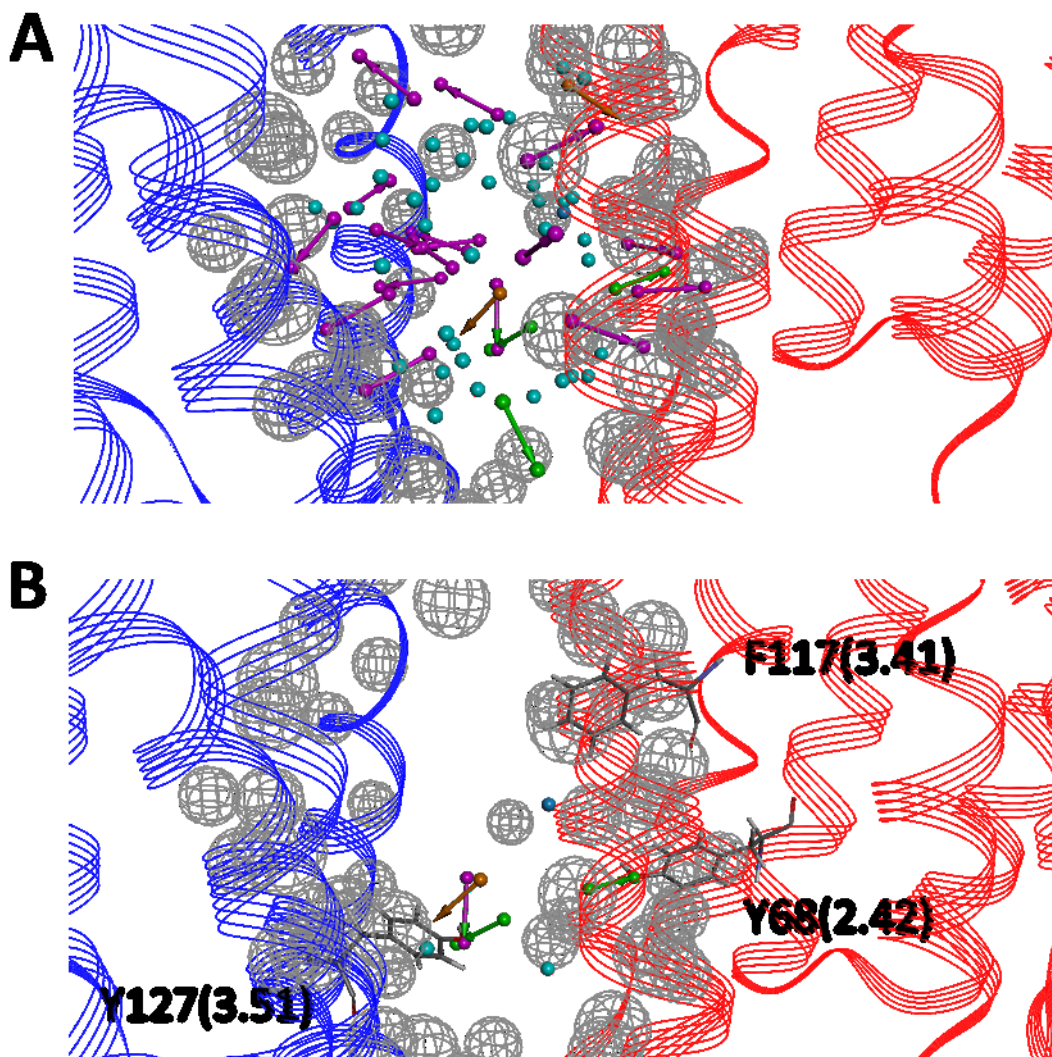


Figure 4 - 14 : (A) pharmacophore généré à partir de la cavité réduite du dimère I5. (B) pharmacophore réduit de la cavité réduite du dimère I5. Les points hydrophobes sont colorés en bleu, les points aromatiques en orange, les points hydrophobes-aromatiques en bleu foncé, les points HBD en mauve et les points HBA en vert. Le ruban bleu correspond à la chaîne A (résidus 1-306) et le ruban rouge à la chaîne B (résidus 307-3014).

Pour le dimère I56, le pharmacophore généré à partir de la cavité réduite de 133 points est constitué de 11 points hydrophobes, deux points HBA/HBD, huit points chargés, sept points HBA, 27 points HBD, un point aromatique et 49 points d'exclusion (Figure 4 - 15 (A)). Le pharmacophore réduit contient neuf points pharmacophoriques (Figure 4 - 15 (B)). Au final, six résidus sont considérés comme importants dans le pharmacophore car pouvant interagir avec les molécules. Ces résidus sont, la cystéine 20, cible d'un point pharmacophorique HBD, la lysine L22(1 .24) cible d'un point pharmacophorique HBA, la tyrosine Y187(5.31) cible d'un point pharmacophorique HBA et influençant un pharmacophore hydrophobe aromatique, l'acide aspartique N192(5.36), l'asparagine N258(6.58) cible d'un point pharmacophorique HBD et la thréonine T259(6.59) cible d'un point pharmacophorique HBD.

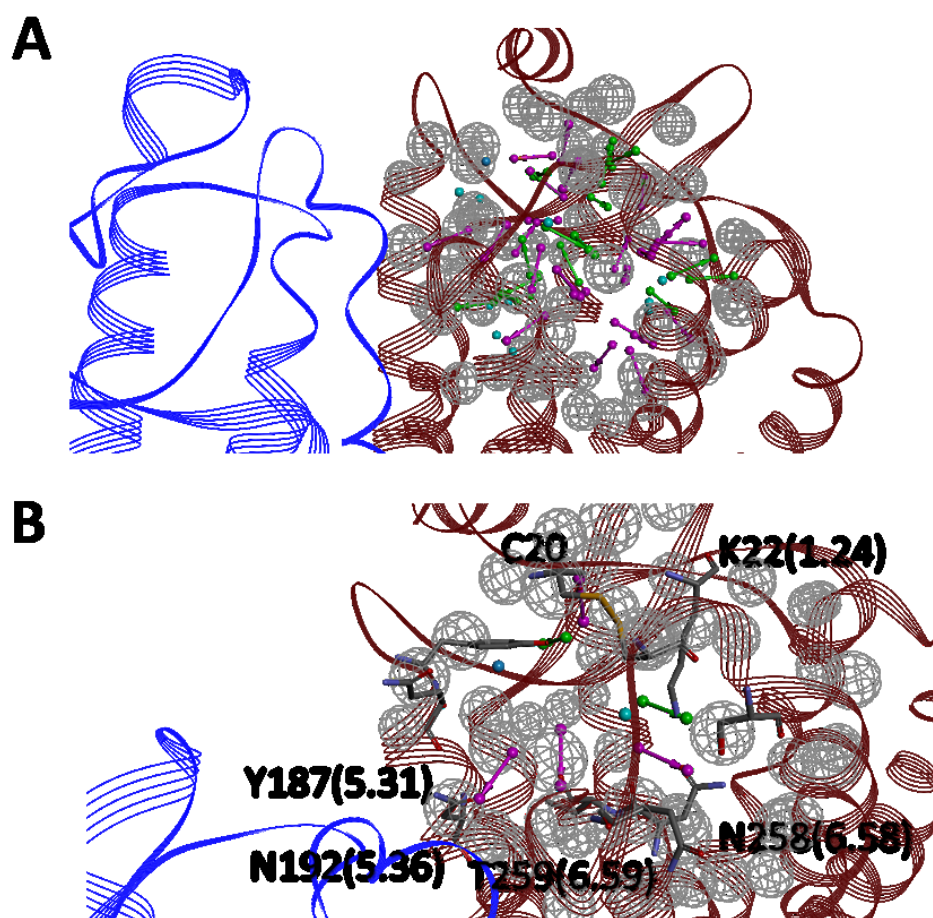


Figure 4 - 15 : (A) pharmacophore généré à partir de la cavité réduite du dimère I5. (B) pharmacophore réduit. Les points hydrophobes sont colorés en bleu, les points aromatiques en orange, les points hydrophobes-aromatiques en bleu foncé, les points HBD en mauve et les points HBA en vert. Le ruban bleu correspond à la chaîne A (résidus 1-306) et le ruban rouge à la chaîne B (résidus 307-3014).

### 3.2.2 Filtrage et tri de la chimiothèque

Le premier tri de la chimiothèque Bioinfo a été effectué en considérant la nature des points dans les deux pharmacophores réduits. Pour le dimère I5 sont requis au moins un cycle aromatique et au moins deux donneurs ou accepteurs de liaison hydrogène. Pour le dimère I56, la somme du nombre d'atome donneur et accepteur de liaison hydrogène doit être strictement supérieure à 3.

Le nombre de molécules à cribler, initialement de 5 498 747, est ainsi ramené à 1 941 768 dans le cas du dimère I5, et à 2 013 058 dans le cas du dimère I56. La génération des structures 3D accompagnées de la création de tous les stéréoisomères si un centre chiral n'est pas défini

augmente de moitié le nombre final de molécules à tester (3 013 533 pour le dimère I5 et 3 141 739 pour le dimère I56). Après génération de conformères et leur évaluation par comparaison avec le pharmacophore réduit, 2 946 439 des molécules obtiennent un score supérieur à zéro pour le dimère I5 et 1 647 867 pour le dimère I56.

La sélection des touches repose sur le score de criblage et sur l’empreinte pharmacophorique qui indique pour chaque point du pharmacophore s’il est retrouvé ou non après alignement du conformère de meilleur score pour une molécule donnée (*pharmprint* dans *catalyst*). Ainsi, pour le dimère I5, n’ont été retenues que les molécules susceptibles d’interagir avec les tyrosines 68 et 127. Pour le dimère I56, les molécules retenues doivent coïncider avec le point hydrophobe et le point hydrophobe/aromatique. Au total, 1 499 molécules ont été sélectionnées pour le dimère I5 et 3 055 molécules pour le dimère I56.

Les deux ensembles de molécules sélectionnées ont chacun été soumis à un partitionnement par similitude chimique, en considérant les châssis chimiques communs. Un total de 245 partitions sont obtenues pour le dimère I5, et de 299 pour le dimère I56. Pour chacune des partitions, l’inspection visuelle de l’alignement sur le pharmacophore de deux molécules a abouti à la sélection de 105 molécules pour le dimère I5 et 91 molécules pour le dimère I56. Un exemple de molécules placé sur le pharmacophore de chaque dimère est présenté en Figure 4 - 16.

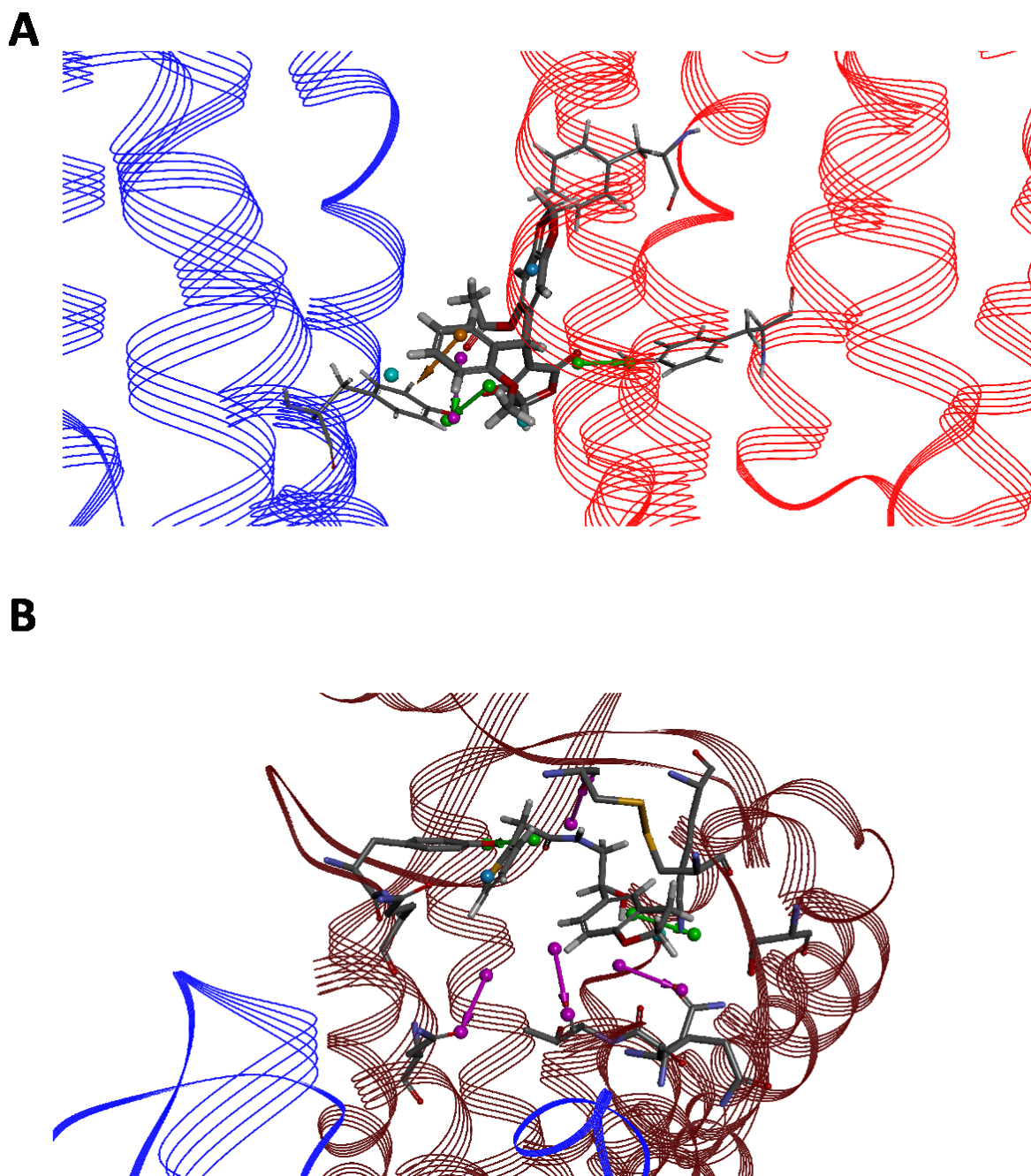


Figure 4 - 16 : Exemple de molécules sélectionnées lors du criblage virtuel du pharmacophore du dimère I5 (A) et I56 (B). Les points hydrophobes sont colorés en bleu, les points aromatiques en orange, les points hydrophobes-aromatiques en bleu foncé, les points HBD en mauve et les points HBA en vert. Le ruban bleu correspond à la chaîne A (résidus 1-306) et le ruban rouge à la chaîne B (résidus 307-3014).

Au total 90 des molécules sélectionnées ont pu être achetées pour le dimère I5 et 75 pour le dimère I56. Ces molécules sont actuellement testées expérimentalement par nos collaborateurs pour caractériser :

- Leur toxicité cellulaire
- Leur impact sur l'export de CCR5 à la membrane cellulaire
- Leur impact sur l'infection des cellules par des souches du VIH ciblant le récepteur CCR5 en présence ou non de l'antiviral MVC
- Leur impact sur l'infection des cellules par des souches du VIH ciblant le récepteur CXCR4 en présence de l'inhibiteur AMD3100
- Leur impact sur l'infection des cellules par le virus de la stomatite vésiculaire
- Leur impact sur la fusion du VIH avec des cellules T
- Leur impact sur la liaison de chimiokines
- Leur impact sur les niveaux de dimérisation

## 4. CONCLUSION

---

Les travaux présentés dans ce chapitre visent à proposer un ensemble de petites molécules pour moduler l'état conformationnel de CCR5 afin d'étudier l'influence de la dimérisation sur l'infection par le VIH.

Deux interfaces dimériques ont été étudiées et utilisées pour sélectionner les molécules. Ces interfaces, modélisées lors de travaux antérieurs, ont été validées expérimentalement. Elles impliquent toute les deux une organisation symétrique autour de l'hélice 5. Le premier dimère considéré, I5, implique également dans son interface les hélices 3 et 4 tandis que le second dimère, I56, implique également l'hélice 6.

Une chimiothèque de plus de 5 millions de molécules disponibles dans le commerce a été criblée par une approche de pharmacophore en 3D. Cette technique nécessite la conception d'un pharmacophore représentant les caractéristiques physico-chimiques que l'on veut retrouver dans les molécules d'intérêt. Pour chacun des dimères étudiés, le pharmacophore a été généré à partir d'une cavité dans l'interface (I5) ou à proximité directe de l'interface (I56). Le ciblage d'une telle cavité peut agir sur la dimérisation en augmentant celle-ci, par la création d'un pont entre les protomères ; en fixant l'architecture dimérique, en verrouillant la conformation du récepteur ; ou en inhibant la formation de dimère en jouant le rôle de perturbateur.

Le criblage virtuel a abouti à la sélection de 196 molécules, dont 165 molécules ont pu être achetées pour être testées in vitro.



## 5. BIBLIOGRAPHIE DE CHAPITRE

---

1. Kniazeff, J., Prézeau, L., Rondard, P., Pin, J.-P. & Goudet, C. Dimers and beyond: The functional puzzles of class C GPCRs. *Pharmacol. Ther.* **130**, 9–25 (2011).
2. Ng, S. Y. L., Lee, L. T. O. & Chow, B. K. C. Receptor oligomerization: from early evidence to current understanding in class B GPCRs. *Front. Endocrinol.* **3**, (2013).
3. Margeta-Mitrovic, M., Jan, Y. N. & Jan, L. Y. A Trafficking Checkpoint Controls GABAB Receptor Heterodimerization. *Neuron* **27**, 97–106 (2000).
4. Franco, R., Martínez-Pinilla, E., Lanciego, J. L. & Navarro, G. Basic Pharmacological and Structural Evidence for Class A G-Protein-Coupled Receptor Heteromerization. *Front. Pharmacol.* **7**, (2016).
5. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-Function of the G Protein–Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
6. Huang, J., Chen, S., Zhang, J. J. & Huang, X.-Y. Crystal structure of oligomeric  $\beta$ 1-adrenergic G protein–coupled receptors in ligand-free basal state. *Nat. Struct. Mol. Biol.* **20**, 419–425 (2013).
7. Salom, D. *et al.* Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl. Acad. Sci.* **103**, 16123–16128 (2006).
8. Wu, H. *et al.* Structure of the human kappa opioid receptor in complex with JD1c. *Nature* **485**, 327–332 (2012).
9. Manglik, A. *et al.* Crystal structure of the  $\mu$ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326 (2012).
10. Wang, C. *et al.* Structure of the human smooth muscle  $\alpha$ 1A adrenergic receptor bound to an antitumour agent. *Nature* **497**, 338–343 (2013).
11. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein-Coupled Receptor. *Science* **318**, 1258–1265 (2007).
12. Wu, B. *et al.* Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* **330**, 1066–1071 (2010).
13. Springael, J.-Y. Allosteric Modulation of Binding Properties between Units of Chemokine Receptor Homo- and Hetero-Oligomers. *Mol. Pharmacol.* **69**, 1652–1661 (2006).
14. Sohy, D. *et al.* Hetero-oligomerization of CCR2, CCR5, and CXCR4 and the Protean Effects of “Selective” Antagonists. *J. Biol. Chem.* **284**, 31270–31279 (2009).

15. Issafras, H. *et al.* Constitutive Agonist-independent CCR5 Oligomerization and Antibody-mediated Clustering Occurring at Physiological Levels of Receptors. *J. Biol. Chem.* **277**, 34666–34673 (2002).
16. El-Asmar, L. Evidence for Negative Binding Cooperativity within CCR5-CCR2b Heterodimers. *Mol. Pharmacol.* **67**, 460–469 (2004).
17. Martinez-Munoz, L. *et al.* CCR5/CD4/CXCR4 oligomerization prevents HIV-1 gp120IIIB binding to the cell surface. *Proc. Natl. Acad. Sci.* **111**, E1960–E1969 (2014).
18. Jin, J. *et al.* CCR5 adopts three homodimeric conformations that control cell surface delivery. *Sci. Signal.* **11**, eaal2869 (2018).
19. Rognan, D. un inventaire de molécules commercialement disponibles à des fins de criblage biologique. 4
20. Dassault Systèmes BIOVIA, Pipeline Pilot. (Dassault Systèmes).
21. Sadowski, J., Gasteiger, J. & Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Model.* **34**, 1000–1008 (1994).
22. Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discov. Today Technol.* **7**, e245–e253 (2010).
23. 3D Structure Generator CORINA Classic. (Molecular Networks GmbH).
24. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
25. Case, D. A. *et al.* AMBER 2018. (University of California).
26. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
27. Sybyl. (Certera U.S.A).
28. Marcou, G. & Rognan, D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J. Chem. Inf. Model.* **47**, 195–207 (2007).
29. Da Silva, F., Desaphy, J. & Rognan, D. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein-Ligand Interactions. *ChemMedChem* **13**, 507–510 (2018).
30. Da Silva, F., Desaphy, J., Bret, G. & Rognan, D. IChemPIC: A Random Forest Classifier of Biological and Crystallographic Protein–Protein Interfaces. *J. Chem. Inf. Model.* **55**, 2005–2014 (2015).



31. Desaphy, J., Azdimousa, K., Kellenberger, E. & Rognan, D. Comparison and Druggability Prediction of Protein–Ligand Binding Sites from Pharmacophore-Annotated Cavity Shapes. *J. Chem. Inf. Model.* **52**, 2287–2299 (2012).
32. Hawkins, P. C. D., Skillman, A. G. & Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **50**, 74–82 (2007).
33. Dassault Systèmes BIOVIA, *Discovery Studio*. (Dassault Systèmes).
34. *MedChem Studio*. (Simulations Plus, Inc.).
35. *Admet Predictor*. (Simulations Plus, Inc.).
36. Garcia-Perez, J. *et al.* Allosteric Model of Maraviroc Binding to CC Chemokine Receptor 5 (CCR5). *J. Biol. Chem.* **286**, 33409–33421 (2011).
37. Maeda, K. *et al.* Structural and Molecular Interactions of CCR5 Inhibitors with CCR5. *J. Biol. Chem.* **281**, 12688–12698 (2006).
38. Arimont, M. *et al.* Chemokine G Protein-Coupled Receptor (Gpcr) Mutation Data Set. (2016). doi:10.5281/zenodo.58160
39. Tamamis, P. & Floudas, C. A. Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS ONE* **9**, e95767 (2014).
40. Kolb, P. *et al.* Structure-based discovery of 2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci.* **106**, 6843–6848 (2009).
41. Rodríguez, D., Brea, J., Loza, M. I. & Carlsson, J. Structure-Based Discovery of Selective Serotonin 5-HT 1B Receptor Ligands. *Structure* **22**, 1140–1151 (2014).
42. Kellenberger, E. *et al.* Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. *J. Med. Chem.* **50**, 1294–1303 (2007).
43. Lim, V. J. Y., Du, W., Chen, Y. Z. & Fan, H. A benchmarking study on virtual ligand screening against homology models of human GPCRs. *Proteins Struct. Funct. Bioinforma.* (2018). doi:10.1002/prot.25533
44. Wieder, M., Perricone, U., Seidel, T., Boesch, S. & Langer, T. Comparing pharmacophore models derived from crystal structures and from molecular dynamics simulations. *Monatshefte Für Chem. - Chem. Mon.* **147**, 553–563 (2016).
45. Perricone, U. *et al.* A Molecular Dynamics-Shared Pharmacophore Approach to Boost Early-Enrichment Virtual Screening: A Case Study on Peroxisome Proliferator-Activated Receptor  $\alpha$ . *ChemMedChem* **12**, 1399–1407 (2017).

# Chapitre 5

## Conclusions et perspectives

Les RCPGs forment une famille de récepteurs fascinante et d'une complexité fonctionnelle incroyable malgré une architecture élémentaire commune. Les données récentes de la littérature suggèrent que la pléthore de fonctions contrôlées par les RCPGs est finement régulée, suggérant que le récepteur n'oscille pas simplement entre deux états, l'un activé et l'autre inactivé, mais expérimente une large gamme de conformations. Les architectures oligomériques participent à la multiplicité structurale des RCPGs, et peuvent à ce titre biaiser la signalisation.

Les travaux de cette thèse se sont focalisés sur l'étude de la plasticité structurale de CCR5 dans le contexte de l'infection par VIH, et ce afin de proposer une interprétation moléculaire des observations expérimentales faites par nos collaborateurs, les Docteurs Anne BRELOT et Bernard LAGANE. Dans ce but, CCR5 a été modélisé dans plusieurs contextes : libre/lié, natif/muté, monomérique/homodimérique. À partir de l'unique structure cristallographique disponible au début de ma thèse, j'ai obtenu un éventail varié de conformations.

Un résultat marquant de mon travail, et non décrit dans ce manuscrit, est l'identification par docking protéine-protéine d'un mode de dimérisation original, non documenté dans la littérature ni observé dans les structures cristallographiques de RCPGs. Ce mode de liaison a été validé expérimentalement par pontage covalent. Nos collaborateurs ont montré qu'il était largement favorisé par la liaison de MVC, et que le récepteur libre pouvait adopter deux autres modes de dimérisation, que j'ai également modélisé, en accord avec les données expérimentales, puis utilisé pour rechercher des modulateurs de dimérisation par criblage moléculaire par pharmacophore. Des molécules modifiant l'expression, la dimérisation et la fonction de corécepteur de CCR5 ont été identifiées et sont en cours de validation.

Les données de nos collaborateurs suggérant que CCR5 peut exister sous forme monomérique et oligomérique, nous avons choisi de travailler la forme la plus simple, le monomère, pour simuler les prémices des mécanismes d'activation ou de blocage du récepteur. Cinq systèmes ont été simulés par dynamique moléculaire : le récepteur dans la forme libre, lié à un ligand agoniste (la chimiokine CCL3), lié à un ligand agoniste inverse (l'antiviral MVC), constitutivement activé par mutation et constitutivement inactivé par mutation. Malgré un temps de production limité à ~300ns par répétition, les trajectoires montrent des comportements spécifiques, notamment des boucles ECL2, ICL3 et ECL3. Des différences de structure secondaire sont observées en particulier dans la partie N-terminale et ECL3. La

grande quantité de données générées pour cette étude et le manque d'information sur l'état activé de CCR5 nous ont amené, au court de l'analyse des trajectoires, à chercher un moyen d'identifier et de caractériser automatiquement les états d'activation. En partant du principe que des interactions intramoléculaires sont caractéristiques des états d'activation et de leurs marqueurs structuraux (comme le mouvement de TM6), nous avons développé une méthode pour identifier des motifs d'interactions intramoléculaires transmembranaires. Ainsi nous avons aisément repéré un petit nombre de résidus impliqués dans le maintien du récepteur dans un état non compétent pour la liaison d'une protéine G. Comme attendu, ces résidus impliquent le motif DRY, mais aussi les résidus Y89(2.63) et K26(1.28). A l'opposé, des résidus impliqués spécifiquement dans un motif caractéristique du complexe CCR5/CCL3, les résidus K229(6.29), Y251(6.51), Q280(7.35) et E283(7.38), ont également été identifiés par cette méthode.

Cette méthode d'analyse a également été validée sur des trajectoires de dynamiques moléculaires de référence. Elle a aussi été appliquée à l'ensemble des structures expérimentales de RCPGs de la PDB, révélant ainsi que les motifs d'interactions intramoléculaires transmembranaires sont caractéristiques du récepteur. L'état d'activation du récepteur est lui codé par une petite part du motif. Excepté le verrou ionique, les interactions associées aux états d'activation semblent dépendantes du récepteur considéré.

Le développement des travaux présentés dans cette thèse peut prendre plusieurs directions dans le futur. Naturellement, l'étude des dimères de CCR5 se poursuivra par l'analyse des différentes touches issues du criblage virtuel, une fois les tests biologiques finalisés. La méthode d'identification de motifs d'interactions intramoléculaires transmembranaires dans les RCPGs peut être développée pour prendre en compte les interactions avec les ions, les lipides ou médiées par une molécule d'eau. Par ailleurs les aspects graphiques peuvent compléter l'approche avec par exemple un affichage dynamique et interactif des interactions sur la structure d'un récepteur. Enfin, il conviendra de poursuivre les simulations du monomère de CCR5 pour conforter les tendances dégagées et aussi de revoir les critères d'analyse pour prendre en compte les différences potentielles dans la reconnaissance par CCR5 des diverses protéines G et notamment des sous-unités  $G\alpha$ -s et  $G\alpha$ -i/o.

# Annexes

Annexe 1 : Structures expérimentales de RCPG. Données issues du site internet GPCrdB. (<http://gpcrdb.org/structure/>)

Récepteur	Famille de récepteur	Classe	Espèce	Identifiant PDB	Résolution	Date
5-HT <sub>1B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">6G79</a>	3.8	20/06/2018
PAF receptor	Platelet-activating factor receptors	Class A (Rhodopsin)	Human	<a href="#">5ZKP</a>	2.8	20/06/2018
A <sub>1</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">6D9H</a>	3.6	20/06/2018
PAF receptor	Platelet-activating factor receptors	Class A (Rhodopsin)	Human	<a href="#">5ZKQ</a>	2.9	20/06/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Human	<a href="#">6CMO</a>	4.5	20/06/2018
US28	Unclassified	Class A (Rhodopsin)	strain AD169	<a href="#">5WB1</a>	3.5	13/06/2018
μ receptor	Opioid receptors	Class A (Rhodopsin)	Mouse	<a href="#">6DDF</a>	3.5	13/06/2018
US28	Unclassified	Class A (Rhodopsin)	strain AD169	<a href="#">5WB2</a>	3.5	13/06/2018
μ receptor	Opioid receptors	Class A (Rhodopsin)	Mouse	<a href="#">6DDE</a>	3.5	13/06/2018
C5a <sub>1</sub> receptor	Complement peptide receptors	Class A (Rhodopsin)	Human	<a href="#">6C1R</a>	2.2	30/05/2018
C5a <sub>1</sub> receptor	Complement peptide receptors	Class A (Rhodopsin)	Human	<a href="#">6C1Q</a>	2.9	30/05/2018
SMO	Frizzled	Class F (Frizzled)	African clawed frog	<a href="#">6D35</a>	3.9	23/05/2018
SMO	Frizzled	Class F (Frizzled)	African clawed frog	<a href="#">6D32</a>	3.8	23/05/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">6GDG</a>	4.1	16/05/2018
FFA1 receptor	Free fatty acid receptors	Class A (Rhodopsin)	Human	<a href="#">5KW2</a>	2.8	02/05/2018
Y <sub>1</sub> receptor	Neuropeptide Y receptors	Class A (Rhodopsin)	Human	<a href="#">5ZBH</a>	3	25/04/2018
Y <sub>1</sub> receptor	Neuropeptide Y receptors	Class A (Rhodopsin)	Human	<a href="#">5ZBQ</a>	2.7	25/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FK7</a>	2.6	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FK6</a>	2.4	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FK8</a>	2.9	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FKB</a>	3	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FKA</a>	2.7	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FKD</a>	2.5	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FK9</a>	2.6	04/04/2018
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">6FKC</a>	2.5	04/04/2018

D <sub>2</sub> receptor	Dopamine receptors	Class A (Rhodopsin)	Human	<a href="#">6CM4</a>	2.9	14/03/2018
mGlu <sub>5</sub> receptor	Metabotropic glutamate receptors	Class C (Glutamate)	Human	<a href="#">6FFI</a>	2.2	07/03/2018
mGlu <sub>5</sub> receptor	Metabotropic glutamate receptors	Class C (Glutamate)	Human	<a href="#">6FFH</a>	2.7	07/03/2018
GLP-1 receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">6B3J</a>	3.3	21/02/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5WF5</a>	2.6	21/02/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5WF6</a>	2.9	21/02/2018
5-HT <sub>2C</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">6BQH</a>	2.7	14/02/2018
5-HT <sub>2C</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">6BQG</a>	3	14/02/2018
5-HT <sub>1B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">5V54</a>	3.9	07/02/2018
κ receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">6B73</a>	3.1	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OLO</a>	3.1	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OLH</a>	2.6	17/01/2018
glucagon receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5YQZ</a>	3	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OLV</a>	2	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OM1</a>	2.1	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OLG</a>	1.9	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OLZ</a>	1.9	17/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5OM4</a>	2	17/01/2018
C5a <sub>1</sub> receptor	Complement peptide receptors	Class A (Rhodopsin)	Human	<a href="#">5O9H</a>	2.7	10/01/2018
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">6AQF</a>	2.5	10/01/2018
BLT <sub>1</sub> receptor	Leukotriene receptors	Class A (Rhodopsin)	Guinea pig	<a href="#">5X33</a>	3.7	03/01/2018
OX <sub>2</sub> receptor	Orexin receptors	Class A (Rhodopsin)	Human	<a href="#">5WS3</a>	2.3	13/12/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5VRA</a>	2.4	13/12/2017
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">5WKT</a>	3.2	13/12/2017
OX <sub>2</sub> receptor	Orexin receptors	Class A (Rhodopsin)	Human	<a href="#">5WQC</a>	2	29/11/2017
D <sub>4</sub> receptor	Dopamine receptors	Class A (Rhodopsin)	Human	<a href="#">5WIV</a>	2.1	18/10/2017
D <sub>4</sub> receptor	Dopamine receptors	Class A (Rhodopsin)	Human	<a href="#">5WIU</a>	2	18/10/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5NLX</a>	2.1	27/09/2017



A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5NM2</a>	2	27/09/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5NM4</a>	1.7	27/09/2017
ET <sub>B</sub> receptor	Endothelin receptors	Class A (Rhodopsin)	Human	<a href="#">5X93</a>	2.2	16/08/2017
LPA <sub>6</sub> receptor	Lysophospholipid (LPA) receptors	Class A (Rhodopsin)	Zebrafish	<a href="#">5XSZ</a>	3.2	16/08/2017
ET <sub>B</sub> receptor	Endothelin receptors	Class A (Rhodopsin)	Human	<a href="#">5XPR</a>	3.6	16/08/2017
β <sub>2</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">5X7D</a>	2.7	16/08/2017
Rhodopsin	Opsins	Class A (Rhodopsin)	Human	<a href="#">5WOP</a>	3	09/08/2017
5-HT <sub>2B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">5TUD</a>	3	26/07/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5N2R</a>	2.8	26/07/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5MZJ</a>	2	26/07/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5MZP</a>	2.1	26/07/2017
A <sub>1</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5N2S</a>	3.3	26/07/2017
CB <sub>1</sub> receptor	Cannabinoid receptors	Class A (Rhodopsin)	Human	<a href="#">5XRA</a>	2.8	12/07/2017
CB <sub>1</sub> receptor	Cannabinoid receptors	Class A (Rhodopsin)	Human	<a href="#">5XR8</a>	3	12/07/2017
CCR5	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">5UIW</a>	2.2	28/06/2017
GLP-1 receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5NX2</a>	3.7	14/06/2017
FFA1 receptor	Free fatty acid receptors	Class A (Rhodopsin)	Human	<a href="#">5TZY</a>	3.2	07/06/2017
FFA1 receptor	Free fatty acid receptors	Class A (Rhodopsin)	Human	<a href="#">5TZR</a>	2.2	07/06/2017
apelin receptor	Apelin receptors	Class A (Rhodopsin)	Human	<a href="#">5VBL</a>	2.6	31/05/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5JTB</a>	2.8	31/05/2017
GLP-1 receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5VEW</a>	2.7	24/05/2017
glucagon receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5XEZ</a>	3	24/05/2017
GLP-1 receptor	Glucagon receptor family	Class B1 (Secretin)	Rabbit	<a href="#">5VAI</a>	4.1	24/05/2017
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">5V56</a>	2.9	24/05/2017
glucagon receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5XF1</a>	3.2	24/05/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5UVI</a>	3.2	24/05/2017
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">5V57</a>	3	24/05/2017
GLP-1 receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5VEX</a>	3	17/05/2017

PAR2	Proteinase-activated receptors	Class A (Rhodopsin)	Human	<a href="#">5NDD</a>	2.8	03/05/2017
CT receptor	Calcitonin receptors	Class B1 (Secretin)	Human	<a href="#">5UZ7</a>	4.1	03/05/2017
PAR2	Proteinase-activated receptors	Class A (Rhodopsin)	Human	<a href="#">5NJ6</a>	4	03/05/2017
PAR2	Proteinase-activated receptors	Class A (Rhodopsin)	Human	<a href="#">5NDZ</a>	3.6	03/05/2017
AT <sub>2</sub> receptor	Angiotensin receptors	Class A (Rhodopsin)	Human	<a href="#">5UNG</a>	2.8	05/04/2017
AT <sub>2</sub> receptor	Angiotensin receptors	Class A (Rhodopsin)	Human	<a href="#">5UNF</a>	2.8	05/04/2017
AT <sub>2</sub> receptor	Angiotensin receptors	Class A (Rhodopsin)	Human	<a href="#">5UNH</a>	2.9	05/04/2017
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">5TE3</a>	2.7	15/03/2017
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">5TE5</a>	4	15/03/2017
A <sub>1</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5UEN</a>	3.2	01/03/2017
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5UIG</a>	3.5	08/02/2017
5-HT <sub>2B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">5TVN</a>	2.9	01/02/2017
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">5T04</a>	3.3	21/12/2016
CCR2	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">5T1A</a>	2.8	14/12/2016
CCR9	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">5LWE</a>	2.8	07/12/2016
CB <sub>1</sub> receptor	Cannabinoid receptors	Class A (Rhodopsin)	Human	<a href="#">5U09</a>	2.6	07/12/2016
CB <sub>1</sub> receptor	Cannabinoid receptors	Class A (Rhodopsin)	Human	<a href="#">5TGZ</a>	2.8	02/11/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5K2C</a>	1.9	21/09/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5K2B</a>	2.5	21/09/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5K2A</a>	2.5	21/09/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5K2D</a>	1.9	21/09/2016
ET <sub>B</sub> receptor	Endothelin receptors	Class A (Rhodopsin)	Human	<a href="#">5GLH</a>	2.8	07/09/2016
ET <sub>B</sub> receptor	Endothelin receptors	Class A (Rhodopsin)	Human	<a href="#">5GLI</a>	2.5	07/09/2016
β <sub>2</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">5D6L</a>	3.2	17/08/2016
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">5DYS</a>	2.3	10/08/2016
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">5EN0</a>	2.8	10/08/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5G53</a>	3.4	03/08/2016
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">5L7D</a>	3.2	20/07/2016

SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">5L7I</a>	3.3	20/07/2016
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">5JQH</a>	3.2	13/07/2016
CRF <sub>1</sub> receptor	Corticotropin-releasing factor receptors	Class B1 (Secretin)	Human	<a href="#">4Z9G</a>	3.2	29/06/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5IU8</a>	2	29/06/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5IUB</a>	2.1	29/06/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5IUA</a>	2.2	29/06/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5IU4</a>	1.7	29/06/2016
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">5IU7</a>	1.9	29/06/2016
glucagon receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">5EE7</a>	2.5	20/04/2016
Rhodopsin	Opsins	Class A (Rhodopsin)	Human	<a href="#">5DGY</a>	7.7	23/03/2016
M <sub>4</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Human	<a href="#">5DSG</a>	2.6	16/03/2016
M <sub>1</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Human	<a href="#">5CXV</a>	2.7	09/03/2016
OX <sub>1</sub> receptor	Orexin receptors	Class A (Rhodopsin)	Human	<a href="#">4ZJ8</a>	2.8	09/03/2016
OX <sub>1</sub> receptor	Orexin receptors	Class A (Rhodopsin)	Human	<a href="#">4ZJC</a>	2.8	09/03/2016
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">5D5A</a>	2.5	13/01/2016
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">5D5B</a>	3.8	13/01/2016
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">5F8U</a>	3.4	23/12/2015
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">4X1H</a>	2.3	04/11/2015
NOP receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">5DHG</a>	3	21/10/2015
NOP receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">5DHH</a>	3	21/10/2015
AT <sub>1</sub> receptor	Angiotensin receptors	Class A (Rhodopsin)	Human	<a href="#">4ZUD</a>	2.8	07/10/2015
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">5A8E</a>	2.4	30/09/2015
mGlu <sub>5</sub> receptor	Metabotropic glutamate receptors	Class C (Glutamate)	Human	<a href="#">5CGC</a>	3.1	12/08/2015
mGlu <sub>5</sub> receptor	Metabotropic glutamate receptors	Class C (Glutamate)	Human	<a href="#">5CGD</a>	2.6	12/08/2015
$\mu$ receptor	Opioid receptors	Class A (Rhodopsin)	Mouse	<a href="#">5C1M</a>	2.1	05/08/2015
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">4XES</a>	2.6	29/07/2015
Rhodopsin	Opsins	Class A (Rhodopsin)	Human	<a href="#">4ZWJ</a>	3.3	29/07/2015
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">4XEE</a>	2.9	29/07/2015

Rhodopsin	Opsins	Class A (Rhodopsin)	Japanese flying squid	<a href="#">4WW3</a>	2.8	17/06/2015
LPA <sub>1</sub> receptor	Lysophospholipid (LPA) receptors	Class A (Rhodopsin)	Human	<a href="#">4Z35</a>	2.9	03/06/2015
LPA <sub>1</sub> receptor	Lysophospholipid (LPA) receptors	Class A (Rhodopsin)	Human	<a href="#">4Z34</a>	3	03/06/2015
LPA <sub>1</sub> receptor	Lysophospholipid (LPA) receptors	Class A (Rhodopsin)	Human	<a href="#">4Z36</a>	2.9	03/06/2015
AT <sub>1</sub> receptor	Angiotensin receptors	Class A (Rhodopsin)	Human	<a href="#">4YAY</a>	2.9	22/04/2015
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">4UHR</a>	2.6	08/04/2015
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">4UG2</a>	2.6	08/04/2015
P2Y <sub>1</sub> receptor	P2Y receptors	Class A (Rhodopsin)	Human	<a href="#">4XNW</a>	2.7	01/04/2015
P2Y <sub>1</sub> receptor	P2Y receptors	Class A (Rhodopsin)	Human	<a href="#">4XNV</a>	2.2	01/04/2015
US28	Unclassified	Class A (Rhodopsin)	strain AD169	<a href="#">4XT1</a>	2.9	04/03/2015
US28	Unclassified	Class A (Rhodopsin)	strain AD169	<a href="#">4XT3</a>	3.8	04/03/2015
CXCR4	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">4RWS</a>	3.1	11/02/2015
δ receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">4RWD</a>	2.7	14/01/2015
δ receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">4RWA</a>	3.3	14/01/2015
OX <sub>2</sub> receptor	Orexin receptors	Class A (Rhodopsin)	Human	<a href="#">4S0V</a>	2.5	14/01/2015
M <sub>3</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Rat	<a href="#">4U15</a>	2.8	26/11/2014
M <sub>3</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Rat	<a href="#">4U14</a>	3.6	26/11/2014
M <sub>3</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Rat	<a href="#">4U16</a>	3.7	26/11/2014
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">4PXF</a>	2.8	17/09/2014
β <sub>2</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">4QKX</a>	3.3	23/07/2014
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">4QIM</a>	2.6	23/07/2014
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">4QIN</a>	2.6	23/07/2014
FFA <sub>1</sub> receptor	Free fatty acid receptors	Class A (Rhodopsin)	Human	<a href="#">4PHU</a>	2.3	16/07/2014
mGlu <sub>5</sub> receptor	Metabotropic glutamate receptors	Class C (Glutamate)	Human	<a href="#">4009</a>	2.6	02/07/2014
P2Y <sub>12</sub> receptor	P2Y receptors	Class A (Rhodopsin)	Human	<a href="#">4PY0</a>	3.1	30/04/2014
P2Y <sub>12</sub> receptor	P2Y receptors	Class A (Rhodopsin)	Human	<a href="#">4PXZ</a>	2.5	30/04/2014
β <sub>1</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">4BVN</a>	2.1	02/04/2014

P2Y <sub>12</sub> receptor	P2Y receptors	Class A (Rhodopsin)	Human	<a href="#">4NTJ</a>	2.6	26/03/2014
mGlu <sub>1</sub> receptor	Metabotropic glutamate receptors	Class C (Glutamate)	Human	<a href="#">4OR2</a>	2.8	19/03/2014
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">4O9R</a>	3.2	05/03/2014
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">4BUO</a>	2.8	29/01/2014
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">3ZEV</a>	3	29/01/2014
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">4BWB</a>	3.6	29/01/2014
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">4BVO</a>	3.1	29/01/2014
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">4N4W</a>	2.8	22/01/2014
δ receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">4N6H</a>	1.8	25/12/2013
5-HT <sub>2B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">4NC3</a>	2.8	18/12/2013
M <sub>2</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Human	<a href="#">4MQT</a>	3.7	27/11/2013
M <sub>2</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Human	<a href="#">4MQS</a>	3.5	27/11/2013
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">4J4Q</a>	2.7	30/10/2013
β <sub>2</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">4LDE</a>	2.8	25/09/2013
β <sub>2</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">4LDL</a>	3.1	25/09/2013
β <sub>2</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">4LDO</a>	3.2	25/09/2013
CCR5	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">4MBS</a>	2.7	11/09/2013
glucagon receptor	Glucagon receptor family	Class B1 (Secretin)	Human	<a href="#">4L6R</a>	3.3	24/07/2013
CRF <sub>1</sub> receptor	Corticotropin-releasing factor receptors	Class B1 (Secretin)	Human	<a href="#">4K5Y</a>	3	17/07/2013
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">4BEY</a>	2.9	08/05/2013
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">4BEZ</a>	3.3	24/04/2013
SMO	Frizzled	Class F (Frizzled)	Human	<a href="#">4JKV</a>	2.5	24/04/2013
β <sub>1</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">3ZPR</a>	2.7	03/04/2013
β <sub>1</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">3ZPQ</a>	2.8	03/04/2013
5-HT <sub>2B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">4IB4</a>	2.7	13/03/2013
5-HT <sub>1B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">4IAQ</a>	2.8	13/03/2013
5-HT <sub>1B</sub> receptor	5-Hydroxytryptamine receptors	Class A (Rhodopsin)	Human	<a href="#">4IAR</a>	2.7	13/03/2013
β <sub>1</sub> -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">4GPO</a>	3.5	27/02/2013

PAR1	Proteinase-activated receptors	Class A (Rhodopsin)	Human	<a href="#">3VW7</a>	2.2	12/12/2012
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">4GBR</a>	4	24/10/2012
NTS <sub>1</sub> receptor	Neurotensin receptors	Class A (Rhodopsin)	Rat	<a href="#">4GRV</a>	2.8	17/10/2012
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">4E1Y</a>	1.8	25/07/2012
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">4AMJ</a>	2.3	23/05/2012
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">4AMI</a>	3.2	23/05/2012
$\delta$ receptor	Opioid receptors	Class A (Rhodopsin)	Mouse	<a href="#">4EJ4</a>	3.4	16/05/2012
NOP receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">4EA3</a>	3	25/04/2012
$\kappa$ receptor	Opioid receptors	Class A (Rhodopsin)	Human	<a href="#">4DJH</a>	2.9	21/03/2012
$\mu$ receptor	Opioid receptors	Class A (Rhodopsin)	Mouse	<a href="#">4DKL</a>	2.8	21/03/2012
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3U2C</a>	3.3	21/03/2012
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3UZA</a>	3.3	21/03/2012
M <sub>3</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Rat	<a href="#">4DAJ</a>	3.4	22/02/2012
S1P <sub>1</sub> receptor	Lysophospholipid (S1P) receptors	Class A (Rhodopsin)	Human	<a href="#">3V2Y</a>	2.8	15/02/2012
S1P <sub>1</sub> receptor	Lysophospholipid (S1P) receptors	Class A (Rhodopsin)	Human	<a href="#">3V2W</a>	3.4	15/02/2012
M <sub>2</sub> receptor	Acetylcholine receptors (muscarinic)	Class A (Rhodopsin)	Human	<a href="#">3UON</a>	3	01/02/2012
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3VG9</a>	2.7	01/02/2012
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3VGA</a>	3.1	01/02/2012
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">4A4M</a>	3.3	25/01/2012
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3REY</a>	3.3	07/09/2011
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3PWH</a>	3.3	07/09/2011
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3RFM</a>	3.6	07/09/2011
Rhodopsin	Opsins	Class A (Rhodopsin)	Japanese flying squid	<a href="#">3AYM</a>	2.8	17/08/2011
Rhodopsin	Opsins	Class A (Rhodopsin)	Japanese flying squid	<a href="#">3AYN</a>	2.7	17/08/2011
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3SN6</a>	3.2	20/07/2011
H <sub>1</sub> receptor	Histamine receptors	Class A (Rhodopsin)	Human	<a href="#">3RZE</a>	3.1	15/06/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2YCY</a>	3.2	08/06/2011

$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2YCX</a>	3.3	01/06/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2YCW</a>	3	01/06/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2YCZ</a>	3.7	01/06/2011
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">2YDV</a>	2.6	18/05/2011
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">2YDO</a>	3	18/05/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2Y01</a>	2.6	30/03/2011
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2X72</a>	3	16/03/2011
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3PQR</a>	2.9	09/03/2011
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3QAK</a>	2.7	09/03/2011
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3PXO</a>	3	09/03/2011
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3OAX</a>	2.6	19/01/2011
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3POG</a>	3.5	19/01/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2Y00</a>	2.5	12/01/2011
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3PDS</a>	3.5	12/01/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2Y04</a>	3.1	12/01/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2Y03</a>	2.9	12/01/2011
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2Y02</a>	2.6	12/01/2011
D <sub>3</sub> receptor	Dopamine receptors	Class A (Rhodopsin)	Human	<a href="#">3PBL</a>	2.9	03/11/2010
CXCR4	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">3ODU</a>	2.5	27/10/2010
CXCR4	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">3OE9</a>	3.1	27/10/2010
CXCR4	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">3OE0</a>	2.9	27/10/2010
CXCR4	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">3OE8</a>	3.1	27/10/2010
CXCR4	Chemokine receptors	Class A (Rhodopsin)	Human	<a href="#">3OE6</a>	3.2	27/10/2010
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3NY8</a>	2.8	11/08/2010
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3NYA</a>	3.2	11/08/2010
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3NY9</a>	2.8	11/08/2010
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3KJ6</a>	3.4	16/02/2010
A <sub>2A</sub> receptor	Adenosine receptors	Class A (Rhodopsin)	Human	<a href="#">3EML</a>	2.6	14/10/2008



Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3DQB</a>	3.2	23/09/2008
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3C9M</a>	3.4	05/08/2008
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3C9L</a>	2.7	05/08/2008
$\beta_1$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Common turkey	<a href="#">2VT4</a>	2.7	24/06/2008
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">3CAP</a>	2.9	24/06/2008
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">3D4S</a>	2.8	17/06/2008
Rhodopsin	Opsins	Class A (Rhodopsin)	Japanese flying squid	<a href="#">2Z73</a>	2.5	13/05/2008
Rhodopsin	Opsins	Class A (Rhodopsin)	Japanese flying squid	<a href="#">2Z1Y</a>	3.7	06/05/2008
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">2R4R</a>	3.4	06/11/2007
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">2R4S</a>	3.4	06/11/2007
$\beta_2$ -adrenoceptor	Adrenoceptors	Class A (Rhodopsin)	Human	<a href="#">2RH1</a>	2.4	30/10/2007
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2PED</a>	3	30/10/2007
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2J4Y</a>	3.4	25/09/2007
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2I35</a>	3.8	17/10/2006
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2I37</a>	4.2	17/10/2006
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2I36</a>	4.1	17/10/2006
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2G87</a>	2.6	02/09/2006
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">2HPY</a>	2.8	22/08/2006
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">1U19</a>	2.2	12/10/2004
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">1GZM</a>	2.7	20/11/2003
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">1L9H</a>	2.6	15/05/2002
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">1HZX</a>	2.8	04/07/2001
Rhodopsin	Opsins	Class A (Rhodopsin)	Bovine	<a href="#">1F88</a>	2.8	04/08/2000



Annexe 2. Entrées PDB sélectionnées d'après la base de données SIFTS et n'ayant pas entre 200 et 600 résidus.

Identifiant Uniprot	Gène	Identifiant PDB	Chaîne	Nombre de résidus	Chaînes du fichiers PDB	Méthode	Résolution	Espèce
BOBLW3	LGR4	4li1	A	431	A/B	X-RAY	2.658	XENTR
BOBLW3	LGR4	4li2	A	431	A	X-RAY	3.19	XENTR
O14842	FFAR1	4phu	A	296	A	X-RAY	2.332	HUMAN
O43613	OX1R	4zj8	A	330	A	X-RAY	2.751	HUMAN
O43613	OX1R	4zjc	A	330	A	X-RAY	2.832	HUMAN
O43614	OX2R	4s0v	A	338	A	X-RAY	2.5	HUMAN
O60242	AGRB3	4dlo	A	370	A/B	X-RAY	2.3	HUMAN
O75473	LGR5	4bsr	A	521	A/B	X-RAY	3.2	HUMAN
O75473	LGR5	4bss	A	521	A/B/E/F	X-RAY	3.2	HUMAN
O75473	LGR5	4bst	A	521	A/B	X-RAY	4.3	HUMAN
O75473	LGR5	4bsu	A	521	A/B/E/F	X-RAY	3.2	HUMAN
O75473	LGR5	4kng	A	525	A/B	X-RAY	2.5	HUMAN
O75473	LGR5	4ufr	A	525	A/C	X-RAY	2.2	HUMAN
O75473	LGR5	4ufs	A	525	A	X-RAY	4.8	HUMAN
O75899	GABR2	4f11	A	424	A	X-RAY	2.38	HUMAN
O75899	GABR2	4f12	A	424	A	X-RAY	3.02	HUMAN
O88917	AGRL1	4dlq	A	377	A/B	X-RAY	1.85	RAT
P02699	OPSD	1f88	A	347	A/B	X-RAY	2.8	BOVIN
P02699	OPSD	1gzm	A	347	A/B	X-RAY	2.65	BOVIN
P02699	OPSD	1hzx	A	347	A/B	X-RAY	2.8	BOVIN
P02699	OPSD	1jfp	A	347	A	SOLUTION	NULL	BOVIN
P02699	OPSD	1l9h	A	347	A/B	X-RAY	2.6	BOVIN
P02699	OPSD	1ln6	A	347	A	SOLUTION	NULL	BOVIN
P02699	OPSD	1u19	A	347	A/B	X-RAY	2.2	BOVIN
P02699	OPSD	2g87	A	347	A/B	X-RAY	2.6	BOVIN
P02699	OPSD	2hpy	A	347	A/B	X-RAY	2.8	BOVIN
P02699	OPSD	2i35	A	347	A	X-RAY	3.8	BOVIN
P02699	OPSD	2i36	A	347	A/B/C	X-RAY	4.1	BOVIN
P02699	OPSD	2i37	A	347	A/B/C	X-RAY	4.15	BOVIN
P02699	OPSD	2j4y	A	347	A/B	X-RAY	3.4	BOVIN
P02699	OPSD	2ped	A	347	A/B	X-RAY	2.95	BOVIN
P02699	OPSD	2x72	A	347	A	X-RAY	3	BOVIN
P02699	OPSD	3c9l	A	347	A	X-RAY	2.646	BOVIN
P02699	OPSD	3c9m	A	347	A	X-RAY	3.4	BOVIN
P02699	OPSD	3cap	A	347	A/B	X-RAY	2.9	BOVIN
P02699	OPSD	3dqb	A	347	A	X-RAY	3.2	BOVIN
P02699	OPSD	3oax	A	347	A/B	X-RAY	2.6	BOVIN
P02699	OPSD	3pqr	A	347	A	X-RAY	2.85	BOVIN
P02699	OPSD	3pxo	A	347	A	X-RAY	3	BOVIN
P02699	OPSD	4a4m	A	347	A	X-RAY	3.3	BOVIN
P02699	OPSD	4bey	A	347	A	X-RAY	2.9	BOVIN
P02699	OPSD	4bez	A	347	A	X-RAY	3.3	BOVIN
P02699	OPSD	4j4q	A	347	A	X-RAY	2.65	BOVIN
P02699	OPSD	4pxf	A	347	A	X-RAY	2.75	BOVIN
P02699	OPSD	4x1h	A	347	A	X-RAY	2.29	BOVIN
P02699	OPSD	5dys	A	347	A	X-RAY	2.3	BOVIN
P02699	OPSD	5en0	A	347	A	X-RAY	2.81	BOVIN
P02699	OPSD	5te3	A	347	A	X-RAY	2.7	BOVIN
P02699	OPSD	5te5	A	347	A	X-RAY	4.01	BOVIN
P07550	ADRB2	2r4r	A	364	A	X-RAY	3.4	HUMAN

P07550	ADRB2	2r4s	A	341	A	X-RAY	3.4	HUMAN
P07550	ADRB2	2rh1	A	331	A	X-RAY	2.4	HUMAN
P07550	ADRB2	3d4s	A	314	A	X-RAY	2.8	HUMAN
P07550	ADRB2	3kj6	A	363	A	X-RAY	3.4	HUMAN
P07550	ADRB2	3ny8	A	314	A	X-RAY	2.84	HUMAN
P07550	ADRB2	3ny9	A	314	A	X-RAY	2.84	HUMAN
P07550	ADRB2	3nya	A	314	A	X-RAY	3.16	HUMAN
P07550	ADRB2	3p0g	A	331	A	X-RAY	3.5	HUMAN
P07550	ADRB2	3pds	A	317	A	X-RAY	3.5	HUMAN
P07550	ADRB2	4gbr	A	336	A	X-RAY	3.993	HUMAN
P07550	ADRB2	4lde	A	319	A	X-RAY	2.79	HUMAN
P07550	ADRB2	4ldl	A	319	A	X-RAY	3.1	HUMAN
P07550	ADRB2	4ldo	A	319	A	X-RAY	3.2	HUMAN
P07550	ADRB2	4qkx	A	319	A	X-RAY	3.3	HUMAN
P07550	ADRB2	5d5a	A	331	A	X-RAY	2.4826	HUMAN
P07550	ADRB2	5d5b	A	331	A	X-RAY	3.8	HUMAN
P07550	ADRB2	5d6l	A	323	A	X-RAY	3.2	HUMAN
P07550	ADRB2	5jqh	A	318	A/B	X-RAY	3.2	HUMAN
P07700	ADRB1	2vt4	A	337	A/B/C/D	X-RAY	2.7	MELGA
P07700	ADRB1	2y00	A	335	A/B	X-RAY	2.5	MELGA
P07700	ADRB1	2y01	A	335	A/B	X-RAY	2.6	MELGA
P07700	ADRB1	2y02	A	335	A/B	X-RAY	2.6	MELGA
P07700	ADRB1	2y03	A	335	A/B	X-RAY	2.85	MELGA
P07700	ADRB1	2y04	A	335	A/B	X-RAY	3.05	MELGA
P07700	ADRB1	2ycw	A	334	A/B	X-RAY	3	MELGA
P07700	ADRB1	2ycx	A	334	A/B	X-RAY	3.25	MELGA
P07700	ADRB1	2ycy	A	334	A/B	X-RAY	3.15	MELGA
P07700	ADRB1	2ycz	A	334	A/B	X-RAY	3.65	MELGA
P07700	ADRB1	3zpq	A	335	A/B	X-RAY	2.8	MELGA
P07700	ADRB1	3zpr	A	335	A/B	X-RAY	2.7	MELGA
P07700	ADRB1	4ami	A	335	A/B	X-RAY	3.2	MELGA
P07700	ADRB1	4amj	A	335	A/B	X-RAY	2.3	MELGA
P07700	ADRB1	4bvn	A	335	A	X-RAY	2.1	MELGA
P07700	ADRB1	4gpo	A	337	A/B	X-RAY	3.5	MELGA
P07700	ADRB1	5a8e	A	335	A	X-RAY	2.4	MELGA
P07700	ADRB1	5f8u	A	335	A/B	X-RAY	3.35	MELGA
P08100	OPSD	4zwj	A	345	A/B/C/D	X-RAY	3.302	HUMAN
P08100	OPSD	5dgy	A	345	A/B/C/D	X-RAY	7.7	HUMAN
P08172	ACM2	3uon	A	343	A	X-RAY	3	HUMAN
P08172	ACM2	4mqs	A	464	A	X-RAY	3.5	HUMAN
P08172	ACM2	4mqt	A	464	A	X-RAY	3.7	HUMAN
P08173	ACM4	5dsg	A	303	A/B	X-RAY	2.6	HUMAN
P08483	ACM3	4daj	A	357	A/B/C/D	X-RAY	3.4	RAT
P08483	ACM3	4u14	A	331	A	X-RAY	3.57	RAT
P08483	ACM3	4u15	A	283	A/B	X-RAY	2.8	RAT
P08483	ACM3	4u16	A	283	A/B	X-RAY	3.7	RAT
P11229	ACM1	5cxv	A	319	A	X-RAY	2.7	HUMAN
P16473	TSHR	2xwt	C	238	C	X-RAY	1.9	HUMAN
P16473	TSHR	3g04	C	238	C	X-RAY	2.55	HUMAN
P20789	NTR1	3zev	A	340	A/B	X-RAY	3	RAT
P20789	NTR1	4buo	A	340	A/B	X-RAY	2.75	RAT
P20789	NTR1	4bv0	A	340	A/B	X-RAY	3.1	RAT
P20789	NTR1	4bwb	A	340	A/B	X-RAY	3.57	RAT
P20789	NTR1	4grv	A	322	A	X-RAY	2.802	RAT
P20789	NTR1	4xee	A	353	A	X-RAY	2.9	RAT

P20789	NTR1	4xes	A	353	A	X-RAY	2.6	RAT
P20789	NTR1	5t04	A	325	A	X-RAY	3.3	RAT
P21554	CNR1	5tgz	A	289	A	X-RAY	2.8	HUMAN
P21554	CNR1	5u09	A	298	A	X-RAY	2.6	HUMAN
P23385	GRM1	1ewk	A	489	A/B	X-RAY	2.2	RAT
P23385	GRM1	1ewt	A	489	A/B	X-RAY	3.7	RAT
P23385	GRM1	1ewv	A	489	A/B	X-RAY	4	RAT
P23385	GRM1	1isr	A	489	A	X-RAY	4	RAT
P23385	GRM1	1iss	A	489	A/B	X-RAY	3.3	RAT
P23945	FSHR	1xwd	C	251	C/F	X-RAY	2.92	HUMAN
P23945	FSHR	4ay9	X	349	X/Y/Z	X-RAY	2.5	HUMAN
P23945	FSHR	4mqw	X	350	X/Y/Z	X-RAY	2.9	HUMAN
P24530	EDNRB	5glh	A	332	A	X-RAY	2.8	HUMAN
P24530	EDNRB	5gli	A	334	A	X-RAY	2.5	HUMAN
P25024	CXCR1	2lnl	A	308	A	SOLID-STATE	NULL	HUMAN
P25103	NK1R	2ks9	A	363	A	SOLUTION	NULL	HUMAN
P25103	NK1R	2ksa	A	363	A	SOLUTION	NULL	HUMAN
P25103	NK1R	2ksb	A	363	A	SOLUTION	NULL	HUMAN
P25116	PAR1	3vw7	A	305	A	X-RAY	2.2	HUMAN
P28222	5HT1B	4iaq	A	294	A	X-RAY	2.8	HUMAN
P28222	5HT1B	4iar	A	292	A	X-RAY	2.7	HUMAN
P29274	AA2AR	2ydo	A	316	A	X-RAY	3	HUMAN
P29274	AA2AR	2ydv	A	316	A	X-RAY	2.6	HUMAN
P29274	AA2AR	3eml	A	299	A	X-RAY	2.6	HUMAN
P29274	AA2AR	3pwh	A	316	A	X-RAY	3.296	HUMAN
P29274	AA2AR	3qak	A	299	A	X-RAY	2.71	HUMAN
P29274	AA2AR	3rey	A	316	A	X-RAY	3.309	HUMAN
P29274	AA2AR	3rfm	A	316	A	X-RAY	3.598	HUMAN
P29274	AA2AR	3uza	A	316	A	X-RAY	3.273	HUMAN
P29274	AA2AR	3uzc	A	316	A	X-RAY	3.341	HUMAN
P29274	AA2AR	3vg9	A	315	A	X-RAY	2.7	HUMAN
P29274	AA2AR	3vga	A	315	A	X-RAY	3.1	HUMAN
P29274	AA2AR	4eiy	A	307	A	X-RAY	1.8	HUMAN
P29274	AA2AR	4ug2	A	316	A/B	X-RAY	2.6	HUMAN
P29274	AA2AR	4uhr	A	316	A	X-RAY	2.6	HUMAN
P29274	AA2AR	5g53	A	307	A/B	X-RAY	3.4	HUMAN
P29274	AA2AR	5iu4	A	309	A	X-RAY	1.72	HUMAN
P29274	AA2AR	5iu7	A	306	A	X-RAY	1.9	HUMAN
P29274	AA2AR	5iu8	A	306	A	X-RAY	2.002	HUMAN
P29274	AA2AR	5iua	A	306	A	X-RAY	2.2	HUMAN
P29274	AA2AR	5iub	A	306	A	X-RAY	2.1	HUMAN
P29274	AA2AR	5k2a	A	307	A	X-RAY	2.5	HUMAN
P29274	AA2AR	5k2b	A	307	A	X-RAY	2.5	HUMAN
P29274	AA2AR	5k2c	A	307	A	X-RAY	1.9	HUMAN
P29274	AA2AR	5k2d	A	307	A	X-RAY	1.9	HUMAN
P29274	AA2AR	5uig	A	308	A	X-RAY	3.5	HUMAN
P30542	AA1R	5uen	A	294	A/B	X-RAY	3.2	HUMAN
P30556	AGTR1	4yay	A	304	A	X-RAY	2.9	HUMAN
P30556	AGTR1	4zud	A	300	A	X-RAY	2.8	HUMAN
P30988	CALCR	5uz7	R	465	R	ELECTRON	NULL	HUMAN
P31356	OPSD	2z73	A	447	A/B	X-RAY	2.5	TODPA
P31356	OPSD	2ziy	A	371	A	X-RAY	3.7	TODPA
P31356	OPSD	3aym	A	447	A/B	X-RAY	2.8	TODPA

P31356	OPSD	3ayn	A	447	A/B	X-RAY	2.7	TODPA
P31356	OPSD	4ww3	A	349	A/B	X-RAY	2.8	TODPA
P31422	GRM3	2e4u	A	550	A/B	X-RAY	2.35	RAT
P31422	GRM3	2e4v	A	550	A/B	X-RAY	2.4	RAT
P31422	GRM3	2e4w	A	550	A/B	X-RAY	2.4	RAT
P31422	GRM3	2e4x	A	550	A/B	X-RAY	2.75	RAT
P31422	GRM3	2e4y	A	550	A/B	X-RAY	3.4	RAT
P32300	OPRD	4ej4	A	298	A	X-RAY	3.4	MOUSE
P34998	CRFR1	4k5y	A	292	A/B/C	X-RAY	2.977	HUMAN
P34998	CRFR1	4z9g	A	294	A/B/C	X-RAY	3.183	HUMAN
P35367	HRH1	3rze	A	467	A	X-RAY	3.1	HUMAN
P35400	GRM7	2e4z	A	488	A	X-RAY	3.3	RAT
P35462	DRD3	3pbl	A	302	A/B	X-RAY	2.89	HUMAN
P41143	OPRD	4n6h	A	302	A	X-RAY	1.8	HUMAN
P41143	OPRD	4rwa	A	299	A/B	X-RAY	3.28	HUMAN
P41143	OPRD	4rwd	A	299	A/B	X-RAY	2.7	HUMAN
P41145	OPRK	4djh	A	313	A/B	X-RAY	2.9	HUMAN
P41146	OPRX	4ea3	A	296	A/B	X-RAY	3.013	HUMAN
P41146	OPRX	5dhg	A	296	A/B	X-RAY	3	HUMAN
P41180	CASR	5fbh	A	521	A/B	X-RAY	2.7	HUMAN
P41180	CASR	5fbk	A	521	A/B	X-RAY	2.1	HUMAN
P41180	CASR	5k5s	A	587	A/B	X-RAY	2.6	HUMAN
P41180	CASR	5k5t	A	587	A	X-RAY	3.1	HUMAN
P41594	GRM5	3lmk	A	487	A/B	X-RAY	2.44	HUMAN
P41594	GRM5	4oo9	A	265	A	X-RAY	2.6	HUMAN
P41594	GRM5	5cgc	A	265	A	X-RAY	3.101	HUMAN
P41594	GRM5	5cgd	A	265	A	X-RAY	2.603	HUMAN
P41595	5HT2B	4ib4	A	308	A	X-RAY	2.7	HUMAN
P41595	5HT2B	4nc3	A	308	A	X-RAY	2.8	HUMAN
P41595	5HT2B	5tvn	A	299	A	X-RAY	2.9	HUMAN
P41597	CCR2	5t1a	A	315	A	X-RAY	2.806	HUMAN
P42866	OPRM	4dkl	A	300	A	X-RAY	2.8	MOUSE
P42866	OPRM	5c1m	A	295	A	X-RAY	2.1	MOUSE
P43220	GLP1R	5vew	A	291	A/B	X-RAY	2.7	HUMAN
P43220	GLP1R	5vex	A	291	A/B	X-RAY	3	HUMAN
P47871	GLR	4l6r	A	311	A	X-RAY	3.3	HUMAN
P47871	GLR	5ee7	A	276	A	X-RAY	2.5	HUMAN
P47900	P2RY1	4xnv	A	365	A	X-RAY	2.2	HUMAN
P47900	P2RY1	4xnw	A	365	A/C	X-RAY	2.7	HUMAN
P50052	AGTR2	5unf	A	300	A/B	X-RAY	2.8	HUMAN
P50052	AGTR2	5ung	B	300	B	X-RAY	2.8	HUMAN
P50052	AGTR2	5unh	A	300	A/B	X-RAY	2.9	HUMAN
P51681	CCR5	4mbs	A	346	A/B	X-RAY	2.71	HUMAN
P51686	CCR9	5lwe	A	318	A/B	X-RAY	2.8	HUMAN
P55085	PAR2	5ndd	A	310	A	X-RAY	2.801	HUMAN
P55085	PAR2	5ndz	A	310	A	X-RAY	3.6	HUMAN
P55085	PAR2	5nj6	A	315	A	X-RAY	4	HUMAN
P61073	CXCR4	3odu	A	316	A/B	X-RAY	2.5	HUMAN
P61073	CXCR4	3oe0	A	316	A	X-RAY	2.9	HUMAN
P61073	CXCR4	3oe6	A	322	A	X-RAY	3.2	HUMAN
P61073	CXCR4	3oe8	A	316	A/B/C	X-RAY	3.1	HUMAN
P61073	CXCR4	3oe9	A	316	A/B	X-RAY	3.1	HUMAN
P61073	CXCR4	4rws	A	316	A	X-RAY	3.1	HUMAN
Q13255	GRM1	3ks9	A	490	A/B	X-RAY	1.9	HUMAN
Q13255	GRM1	4or2	A	279	A/B	X-RAY	2.8	HUMAN

Q14416	GRM2	4xaq	A	491	A/B	X-RAY	2.21	HUMAN
Q14416	GRM2	4xas	A	491	A/B	X-RAY	2.35	HUMAN
Q14416	GRM2	5cni	A	491	A/B	X-RAY	2.69	HUMAN
Q14416	GRM2	5cnj	A	491	A/B	X-RAY	2.65	HUMAN
Q14416	GRM2	5kzn	A	563	A	X-RAY	2.8	HUMAN
Q14416	GRM2	5kzq	A	563	A	X-RAY	2.8	HUMAN
Q14831	GRM7	3mq4	A	476	A	X-RAY	2.8	HUMAN
Q14831	GRM7	5c5c	A	476	A	X-RAY	1.862	HUMAN
Q14832	GRM3	3sm9	A	478	A	X-RAY	2.26	HUMAN
Q14832	GRM3	4xar	A	506	A	X-RAY	2.26	HUMAN
Q14832	GRM3	5cnk	A	505	A/B/C	X-RAY	3.15	HUMAN
Q14832	GRM3	5cnm	A	505	A	X-RAY	2.84	HUMAN
Q80TS3	AGRL3	4rmk	A	296	A	X-RAY	1.606	MOUSE
Q80TS3	AGRL3	4rml	A	296	A	X-RAY	1.601	MOUSE
Q80TS3	AGRL3	4yeb	A	296	A	X-RAY	3.19	MOUSE
Q80TS3	AGRL3	5ftt	C	371	C/D/G/H	X-RAY	3.4	MOUSE
Q80TS3	AGRL3	5ftu	C	371	C/D/G/H/ K/L	X-RAY	6.01	MOUSE
Q8K209	AGRG1	5kvm	A	354	A/B	X-RAY	2.449	MOUSE
Q92633	LPAR1	4z34	A	308	A	X-RAY	3	HUMAN
Q92633	LPAR1	4z35	A	308	A	X-RAY	2.9	HUMAN
Q92633	LPAR1	4z36	A	308	A	X-RAY	2.9	HUMAN
Q99835	SMO	4jkv	A	365	A/B	X-RAY	2.45	HUMAN
Q99835	SMO	4n4w	A	365	A	X-RAY	2.8	HUMAN
Q99835	SMO	4o9r	A	357	A	X-RAY	3.204	HUMAN
Q99835	SMO	4qim	A	357	A	X-RAY	2.61	HUMAN
Q99835	SMO	4qin	A	357	A	X-RAY	2.6	HUMAN
Q99835	SMO	5l7d	A	508	A/B	X-RAY	3.2	HUMAN
Q99835	SMO	5l7i	A	508	A/B	X-RAY	3.3	HUMAN
Q9BXB1	LGR4	4kt1	A	501	A	X-RAY	2.497	HUMAN
Q9BXB1	LGR4	4qxe	A	385	A	X-RAY	2.2	HUMAN
Q9BXB1	LGR4	4qxf	A	256	A/B	X-RAY	2.25	HUMAN
Q9H244	P2Y12	4pxz	A	339	A	X-RAY	2.5	HUMAN
Q9H244	P2Y12	4py0	A	339	A	X-RAY	3.1	HUMAN
Q9HAR2	AGRL3	5cmn	E	260	E/F/G/H	X-RAY	3.605	HUMAN
Q9UBS5	GABR1	4mqe	A	411	A	X-RAY	2.35	HUMAN
Q9UBS5	GABR1	4mqf	A	411	A	X-RAY	2.22	HUMAN
Q9UBS5	GABR1	4mr7	A	411	A	X-RAY	2.15	HUMAN
Q9UBS5	GABR1	4mr8	A	411	A	X-RAY	2.15	HUMAN
Q9UBS5	GABR1	4mr9	A	411	A	X-RAY	2.35	HUMAN
Q9UBS5	GABR1	4mrm	A	411	A	X-RAY	2.86	HUMAN
Q9UBS5	GABR1	4ms1	A	411	A	X-RAY	2.25	HUMAN
Q9UBS5	GABR1	4ms3	A	411	A	X-RAY	2.5	HUMAN
Q9UBS5	GABR1	4ms4	A	411	A	X-RAY	1.9	HUMAN
Q9UHX3	AGRE2	2bo2	A	235	A/B	X-RAY	2.6	HUMAN
Q9UHX3	AGRE2	2box	A	235	A	X-RAY	2.5	HUMAN

Annexe 3 : Entrées PDB d'après le site GPCRdb.

Identifiant Uniprot	Gène	Identifiant PDB	Chaîne	Nombre de résidus	Chaînes du fichiers PDB	Méthode	Résolution	Espèce
O14842	FFAR1	4phu	A	296	A	X-RAY	2.332	HUMAN
O43613	OX1R	4zj8	A	330	A	X-RAY	2.751	HUMAN
O43613	OX1R	4zjc	A	330	A	X-RAY	2.832	HUMAN
O43614	OX2R	4s0v	A	338	A	X-RAY	2.5	HUMAN
P02699	OPSD	1jfp	A	347	A	SOLUTION	NULL	BOVIN
P02699	OPSD	1ln6	A	347	A	SOLUTION	NULL	BOVIN
P02699	OPSD	4bey	A	347	A	X-RAY	2.9	BOVIN
P02699	OPSD	4bez	A	347	A	X-RAY	3.3	BOVIN
P02699	OPSD	4j4q	A	347	A	X-RAY	2.65	BOVIN
P02699	OPSD	4pxf	A	347	A	X-RAY	2.75	BOVIN
P02699	OPSD	4x1h	A	347	A	X-RAY	2.29	BOVIN
P02699	OPSD	5dys	A	347	A	X-RAY	2.3	BOVIN
P02699	OPSD	5en0	A	347	A	X-RAY	2.81	BOVIN
P02699	OPSD	5te3	A	347	A	X-RAY	2.7	BOVIN
P02699	OPSD	5te5	A	347	A	X-RAY	4.01	BOVIN
P07550	ADRB2	2r4r	A	364	A	X-RAY	3.4	HUMAN
P07550	ADRB2	2r4s	A	341	A	X-RAY	3.4	HUMAN
P07550	ADRB2	3kj6	A	363	A	X-RAY	3.4	HUMAN
P07700	ADRB1	2vt4	A	337	A/B/C/D	X-RAY	2.7	MELGA
P07700	ADRB1	2y00	A	335	A/B	X-RAY	2.5	MELGA
P07700	ADRB1	2y01	A	335	A/B	X-RAY	2.6	MELGA
P07700	ADRB1	2y02	A	335	A/B	X-RAY	2.6	MELGA
P07700	ADRB1	2y03	A	335	A/B	X-RAY	2.85	MELGA
P07700	ADRB1	2y04	A	335	A/B	X-RAY	3.05	MELGA
P07700	ADRB1	2ycw	A	334	A/B	X-RAY	3	MELGA
P07700	ADRB1	2ycx	A	334	A/B	X-RAY	3.25	MELGA
P07700	ADRB1	2ycy	A	334	A/B	X-RAY	3.15	MELGA
P07700	ADRB1	2ycz	A	334	A/B	X-RAY	3.65	MELGA
P07700	ADRB1	3zpq	A	335	A/B	X-RAY	2.8	MELGA
P07700	ADRB1	3zpr	A	335	A/B	X-RAY	2.7	MELGA
P07700	ADRB1	4ami	A	335	A/B	X-RAY	3.2	MELGA
P07700	ADRB1	4amj	A	335	A/B	X-RAY	2.3	MELGA
P07700	ADRB1	4bvn	A	335	A	X-RAY	2.1	MELGA
P07700	ADRB1	4gpo	A	337	A/B	X-RAY	3.5	MELGA
P07700	ADRB1	5a8e	A	335	A	X-RAY	2.4	MELGA
P07700	ADRB1	5f8u	A	335	A/B	X-RAY	3.35	MELGA
P08100	OPSD	4zwy	A	345	A/B/C/D	X-RAY	3.302	HUMAN
P08100	OPSD	5dgy	A	345	A/B/C/D	X-RAY	7.7	HUMAN
P08173	ACM4	5dsg	A	303	A/B	X-RAY	2.6	HUMAN
P08483	ACM3	4daj	A	357	A/B/C/D	X-RAY	3.4	RAT
P08483	ACM3	4u14	A	331	A	X-RAY	3.57	RAT
P08483	ACM3	4u15	A	283	A/B	X-RAY	2.8	RAT

P08483	ACM3	4u16	A	283	A/B	X-RAY	3.7	RAT
P11229	ACM1	5cxv	A	319	A	X-RAY	2.7	HUMAN
P20789	NTR1	3zev	A	340	A/B	X-RAY	3	RAT
P20789	NTR1	4buo	A	340	A/B	X-RAY	2.75	RAT
P20789	NTR1	4bv0	A	340	A/B	X-RAY	3.1	RAT
P20789	NTR1	4bwb	A	340	A/B	X-RAY	3.57	RAT
P20789	NTR1	5t04	A	325	A	X-RAY	3.3	RAT
P21554	CNR1	5tgz	A	289	A	X-RAY	2.8	HUMAN
P21554	CNR1	5u09	A	298	A	X-RAY	2.6	HUMAN
P24530	EDNRB	5glh	A	332	A	X-RAY	2.8	HUMAN
P24530	EDNRB	5gli	A	334	A	X-RAY	2.5	HUMAN
P25024	CXCR1	2lnl	A	308	A	SOLID-STATE	NULL	HUMAN
P25103	NK1R	2ks9	A	363	A	SOLUTION	NULL	HUMAN
P25103	NK1R	2ksa	A	363	A	SOLUTION	NULL	HUMAN
P25103	NK1R	2ksb	A	363	A	SOLUTION	NULL	HUMAN
P25116	PAR1	3vw7	A	305	A	X-RAY	2.2	HUMAN
P28222	5HT1B	4iaq	A	294	A	X-RAY	2.8	HUMAN
P28222	5HT1B	4iar	A	292	A	X-RAY	2.7	HUMAN
P29274	AA2AR	5iu8	A	306	A	X-RAY	2.002	HUMAN
P29274	AA2AR	5iua	A	306	A	X-RAY	2.2	HUMAN
P29274	AA2AR	5iub	A	306	A	X-RAY	2.1	HUMAN
P29274	AA2AR	5k2a	A	307	A	X-RAY	2.5	HUMAN
P29274	AA2AR	5k2b	A	307	A	X-RAY	2.5	HUMAN
P29274	AA2AR	5k2c	A	307	A	X-RAY	1.9	HUMAN
P29274	AA2AR	5k2d	A	307	A	X-RAY	1.9	HUMAN
P29274	AA2AR	5uig	A	308	A	X-RAY	3.5	HUMAN
P30542	AA1R	5uen	A	294	A/B	X-RAY	3.2	HUMAN
P30556	AGTR1	4yay	A	304	A	X-RAY	2.9	HUMAN
P30556	AGTR1	4zud	A	300	A	X-RAY	2.8	HUMAN
P30988	CALCR	5uz7	R	465	R	ELECTRON	NULL	HUMAN
P31356	OPSD	2z73	A	447	A/B	X-RAY	2.5	TODPA
P31356	OPSD	2ziy	A	371	A	X-RAY	3.7	TODPA
P31356	OPSD	3aym	A	447	A/B	X-RAY	2.8	TODPA
P31356	OPSD	3ayn	A	447	A/B	X-RAY	2.7	TODPA
P31356	OPSD	4ww3	A	349	A/B	X-RAY	2.8	TODPA
P32300	OPRD	4ej4	A	298	A	X-RAY	3.4	MOUSE
P34998	CRFR1	4k5y	A	292	A/B/C	X-RAY	2.977	HUMAN
P34998	CRFR1	4z9g	A	294	A/B/C	X-RAY	3.183	HUMAN
P35367	HRH1	3rze	A	467	A	X-RAY	3.1	HUMAN
P35462	DRD3	3pbl	A	302	A/B	X-RAY	2.89	HUMAN
P41143	OPRD	4n6h	A	302	A	X-RAY	1.8	HUMAN
P41143	OPRD	4rwa	A	299	A/B	X-RAY	3.28	HUMAN
P41143	OPRD	4rwd	A	299	A/B	X-RAY	2.7	HUMAN
P41145	OPRK	4djh	A	313	A/B	X-RAY	2.9	HUMAN
P41146	OPRX	4ea3	A	296	A/B	X-RAY	3.013	HUMAN



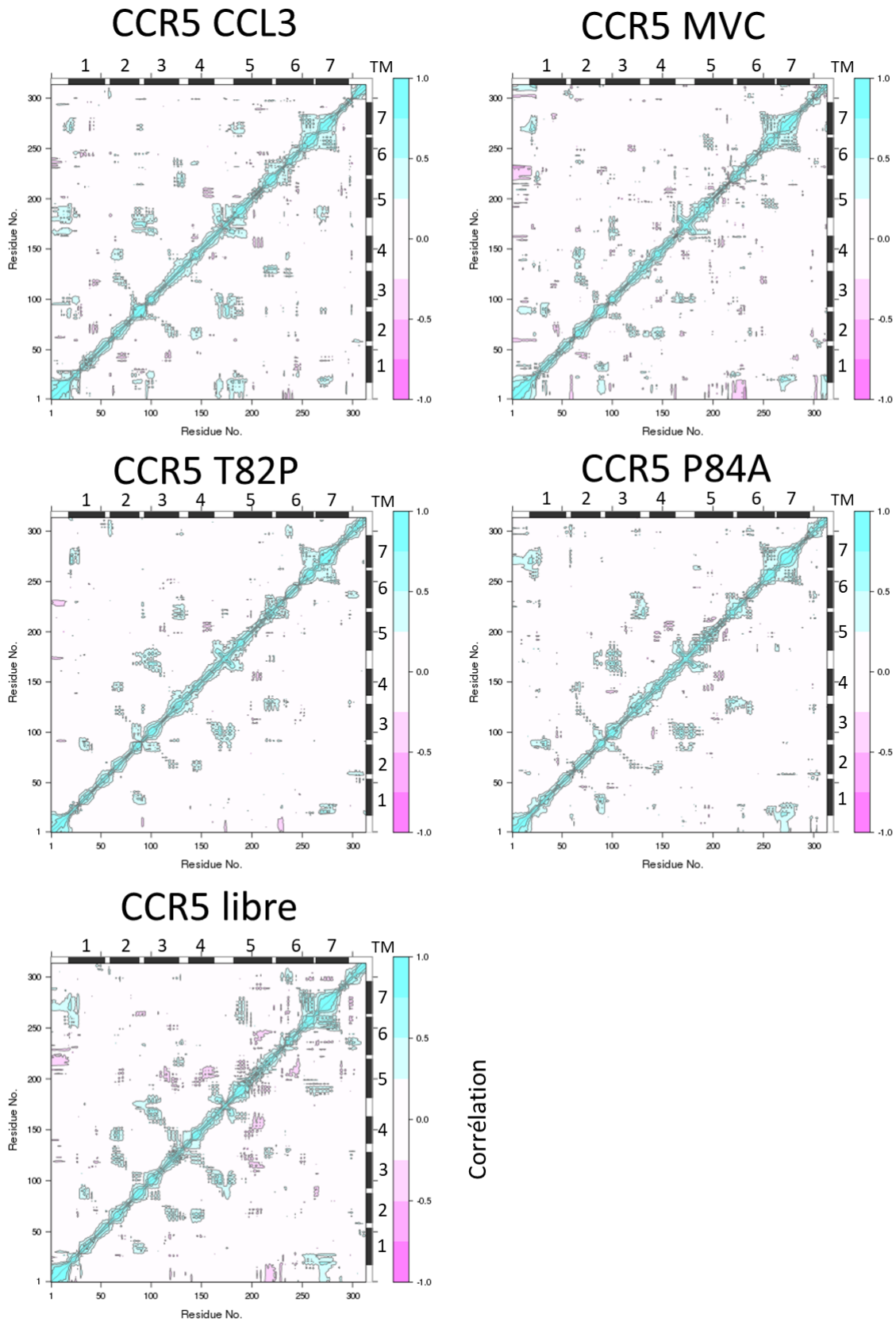
P41146	OPRX	5dhg	A	296	A/B	X-RAY	3	HUMAN
P41594	GRM5	4oo9	A	265	A	X-RAY	2.6	HUMAN
P41594	GRM5	5cgc	A	265	A	X-RAY	3.101	HUMAN
P41594	GRM5	5cgd	A	265	A	X-RAY	2.603	HUMAN
P41595	5HT2B	4ib4	A	308	A	X-RAY	2.7	HUMAN
P41595	5HT2B	4nc3	A	308	A	X-RAY	2.8	HUMAN
P41595	5HT2B	5tvn	A	299	A	X-RAY	2.9	HUMAN
P43220	GLP1R	5vew	A	291	A/B	X-RAY	2.7	HUMAN
P43220	GLP1R	5vex	A	291	A/B	X-RAY	3	HUMAN
P47871	GLR	4l6r	A	311	A	X-RAY	3.3	HUMAN
P47871	GLR	5ee7	A	276	A	X-RAY	2.5	HUMAN
P47900	P2RY1	4xnv	A	365	A	X-RAY	2.2	HUMAN
P47900	P2RY1	4xnw	A	365	A/C	X-RAY	2.7	HUMAN
P50052	AGTR2	5unf	A	300	A/B	X-RAY	2.8	HUMAN
P50052	AGTR2	5ung	B	300	B	X-RAY	2.8	HUMAN
P50052	AGTR2	5unh	A	300	A/B	X-RAY	2.9	HUMAN
P51681	CCR5	4mbs	A	346	A/B	X-RAY	2.71	HUMAN
P51686	CCR9	5lwe	A	318	A/B	X-RAY	2.8	HUMAN
P55085	PAR2	5ndd	A	310	A	X-RAY	2.801	HUMAN
P55085	PAR2	5ndz	A	310	A	X-RAY	3.6	HUMAN
P55085	PAR2	5nj6	A	315	A	X-RAY	4	HUMAN
P61073	CXCR4	3odu	A	316	A/B	X-RAY	2.5	HUMAN
P61073	CXCR4	3oe0	A	316	A	X-RAY	2.9	HUMAN
P61073	CXCR4	3oe6	A	322	A	X-RAY	3.2	HUMAN
P61073	CXCR4	3oe8	A	316	A/B/C	X-RAY	3.1	HUMAN
P61073	CXCR4	3oe9	A	316	A/B	X-RAY	3.1	HUMAN
P61073	CXCR4	4rws	A	316	A	X-RAY	3.1	HUMAN
Q13255	GRM1	4or2	A	279	A/B	X-RAY	2.8	HUMAN
Q14416	GRM2	5kzn	A	563	A	X-RAY	2.8	HUMAN
Q14416	GRM2	5kzq	A	563	A	X-RAY	2.8	HUMAN
Q92633	LPAR1	4z34	A	308	A	X-RAY	3	HUMAN
Q92633	LPAR1	4z35	A	308	A	X-RAY	2.9	HUMAN
Q92633	LPAR1	4z36	A	308	A	X-RAY	2.9	HUMAN
Q99835	SMO	4jkv	A	365	A/B	X-RAY	2.45	HUMAN
Q99835	SMO	4n4w	A	365	A	X-RAY	2.8	HUMAN
Q99835	SMO	4o9r	A	357	A	X-RAY	3.204	HUMAN
Q99835	SMO	4qim	A	357	A	X-RAY	2.61	HUMAN
Q99835	SMO	4qin	A	357	A	X-RAY	2.6	HUMAN
Q99835	SMO	5l7d	A	508	A/B	X-RAY	3.2	HUMAN
Q99835	SMO	5l7i	A	508	A/B	X-RAY	3.3	HUMAN
Q9H244	P2Y12	4pxz	A	339	A	X-RAY	2.5	HUMAN
Q9H244	P2Y12	4py0	A	339	A	X-RAY	3.1	HUMAN
Q9H244	P2Y12	4ntj	A	339	A	X-RAY	2.62	HUMAN
P25024	CXCR1	2lnl	A	308	A	SOLID-STATE	NULL	HUMAN
P25103	NK1R	2ks9	A	363	A	SOLUTION	NULL	HUMAN
P25103	NK1R	2ksa	A	363	A	SOLUTION	NULL	HUMAN



P25103	NK1R	2ksb	A	363	A	SOLUTION	NULL	HUMAN
P69332	US28	4xt1	A	362	A	X-RAY	2.89	HUMAN
P69332	US28	4xt3	A	362	A	X-RAY	2.89	HUMAN
P41146	OPRX	5dhh	A	296	A/B	X-RAY	3	HUMAN
P29274	AA2AR	5jtb	A	412	A	X-RAY	2.8	HUMAN
P43220	GLP1R	5nx2	A	422	A	X-RAY	3.7	HUMAN
O14842	FFAR1	5tzt	A	300	A	X-RAY	2.2	HUMAN
O14842	FFAR1	5tzy	A	300	A	X-RAY	3.22	HUMAN
P51681	CCR5	5uiw	A	352	A	X-RAY	2.2	HUMAN
P29274	AA2AR	5uvi	A	316	A	X-RAY	3.2	HUMAN
G1SGD4	G1SGD4	5vai	R	398	R	ELECTRON	4.1	RABBIT A
P35414	APJ	5vbl	B	407	B	X-RAY	2.6	HUMAN
P47871	GLR	5xez	A	406	A/B	X-RAY	3	HUMAN
P47871	GLR	5xf1	A	406	A/B	X-RAY	3	HUMAN
P21554	CNR1	5xr8	A	291	A	X-RAY	2.95	HUMAN
P21554	CNR1	5xra	A	291	A	X-RAY	2.8	HUMAN
P29274	AA2AR	3qak	A	299	A	X-RAY	2.71	HUMAN
P29274	AA2AR	4ug2	A	316	A/B	X-RAY	2.6	HUMAN
P29274	AA2AR	4uhr	A	316	A	X-RAY	2.6	HUMAN
P29274	AA2AR	5g53	A	307	A/B	X-RAY	3.4	HUMAN
P29274	AA2AR	5iu4	A	309	A	X-RAY	1.72	HUMAN
P29274	AA2AR	5iu7	A	306	A	X-RAY	1.9	HUMAN
P29274	AA2AR	3eml	A	299	A	X-RAY	2.6	HUMAN
P29274	AA2AR	3rey	A	316	A	X-RAY	3.309	HUMAN
P29274	AA2AR	3rfm	A	316	A	X-RAY	3.598	HUMAN
P29274	AA2AR	3uza	A	316	A	X-RAY	3.273	HUMAN
P29274	AA2AR	3uzc	A	316	A	X-RAY	3.341	HUMAN
P29274	AA2AR	3vg9	A	315	A	X-RAY	2.7	HUMAN
P29274	AA2AR	3vga	A	315	A	X-RAY	3.1	HUMAN
P29274	AA2AR	4eiy	A	307	A	X-RAY	1.8	HUMAN
P29274	AA2AR	3pwh	A	316	A	X-RAY	3.296	HUMAN
P29274	AA2AR	2ydo	A	316	A	X-RAY	3	HUMAN
P29274	AA2AR	2ydv	A	316	A	X-RAY	2.6	HUMAN
P08172	ACM2	4mq5	A	464	A	X-RAY	3.5	HUMAN
P08172	ACM2	4mqt	A	464	A	X-RAY	3.7	HUMAN
P08172	ACM2	3uon	A	343	A	X-RAY	3	HUMAN
P07550	ADRB2	3p0g	A	331	A	X-RAY	3.5	HUMAN
P07550	ADRB2	3sn6	R	335	R	X-RAY	3.2	HUMAN
P07550	ADRB2	4lde	A	319	A	X-RAY	2.79	HUMAN
P07550	ADRB2	4ldl	A	319	A	X-RAY	3.1	HUMAN
P07550	ADRB2	4ldo	A	319	A	X-RAY	3.2	HUMAN
P07550	ADRB2	4qkx	A	319	A	X-RAY	3.3	HUMAN
P07550	ADRB2	3nya	A	314	A	X-RAY	3.16	HUMAN
P07550	ADRB2	2rh1	A	331	A	X-RAY	2.4	HUMAN
P07550	ADRB2	3d4s	A	314	A	X-RAY	2.8	HUMAN
P07550	ADRB2	3ny8	A	314	A	X-RAY	2.84	HUMAN
P07550	ADRB2	3ny9	A	314	A	X-RAY	2.84	HUMAN
P07550	ADRB2	4gbr	A	336	A	X-RAY	3.993	HUMAN

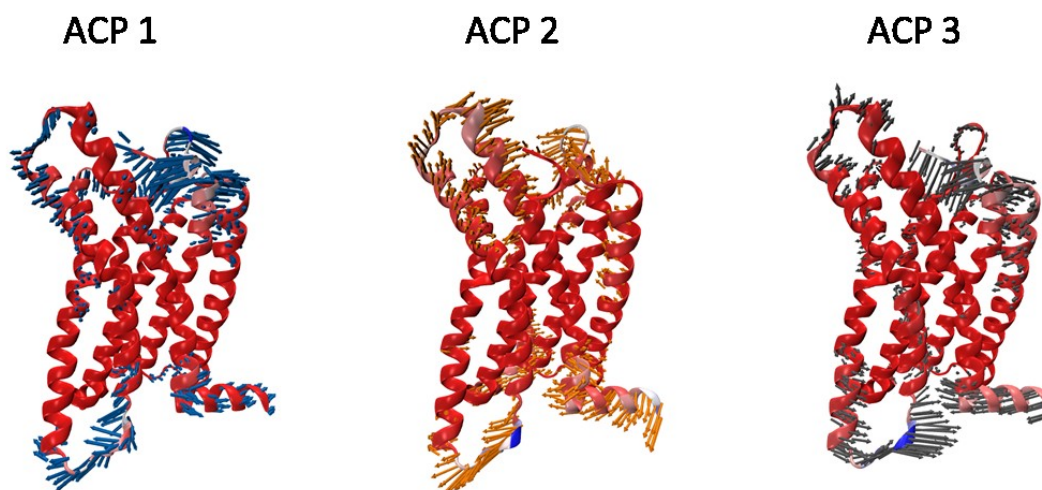
P07550	ADRB2	5d5a	A	331	A	X-RAY	2.4826	HUMAN
P07550	ADRB2	5d5b	A	331	A	X-RAY	3.8	HUMAN
P07550	ADRB2	5d6l	A	323	A	X-RAY	3.2	HUMAN
P07550	ADRB2	5jqh	A	318	A/B	X-RAY	3.2	HUMAN
P07550	ADRB2	3pds	A	317	A	X-RAY	3.5	HUMAN
P41597	CCR2	5t1a	A	315	A	X-RAY	2.806	HUMAN
P20789	NTR1	4grv	A	322	A	X-RAY	2.802	RAT
P20789	NTR1	4xee	A	353	A	X-RAY	2.9	RAT
P20789	NTR1	4xes	A	353	A	X-RAY	2.6	RAT
P42866	OPRM	5c1m	A	295	A	X-RAY	2.1	MOUSE
P42866	OPRM	4dkl	A	300	A	X-RAY	2.8	MOUSE
P02699	OPSD	2x72	A	347	A	X-RAY	3	BOVIN
P02699	OPSD	3dqb	A	347	A	X-RAY	3.2	BOVIN
P02699	OPSD	2g87	A	347	A/B	X-RAY	2.6	BOVIN
P02699	OPSD	3pxo	A	347	A	X-RAY	3	BOVIN
P02699	OPSD	1f88	A	347	A/B	X-RAY	2.8	BOVIN
P02699	OPSD	1gzm	A	347	A/B	X-RAY	2.65	BOVIN
P02699	OPSD	1hzx	A	347	A/B	X-RAY	2.8	BOVIN
P02699	OPSD	1l9h	A	347	A/B	X-RAY	2.6	BOVIN
P02699	OPSD	1u19	A	347	A/B	X-RAY	2.2	BOVIN
P02699	OPSD	2j4y	A	347	A/B	X-RAY	3.4	BOVIN
P02699	OPSD	2ped	A	347	A/B	X-RAY	2.95	BOVIN
P02699	OPSD	3c9l	A	347	A	X-RAY	2.646	BOVIN
P02699	OPSD	3c9m	A	347	A	X-RAY	3.4	BOVIN
P02699	OPSD	3oax	A	347	A/B	X-RAY	2.6	BOVIN
P02699	OPSD	3cap	A	347	A/B	X-RAY	2.9	BOVIN
P02699	OPSD	2hpy	A	347	A/B	X-RAY	2.8	BOVIN
P02699	OPSD	3pqr	A	347	A	X-RAY	2.85	BOVIN
P02699	OPSD	4a4m	A	347	A	X-RAY	3.3	BOVIN
P02699	OPSD	2i35	A	347	A	X-RAY	3.8	BOVIN
P02699	OPSD	2i36	A	347	A/B/C	X-RAY	4.1	BOVIN
P02699	OPSD	2i37	A	347	A/B/C	X-RAY	4.15	BOVIN
P21453	S1PR1	3v2w	A	312	A	X-RAY	3.35	HUMAN
P21453	S1PR1	3v2y	A	312	A	X-RAY	2.8	HUMAN

Annexe 4 : Carte de corrélations des mouvements dans CCR5. Les barres noires annotées de 1 à 7 définissent les hélices transmembranaires du récepteur. Les valeurs de corrélation des mouvements des résidus vont de 1 (couleur cyan), pour des mouvements corrélés, à -1 (couleur rose) pour des mouvements anti-corrélés.



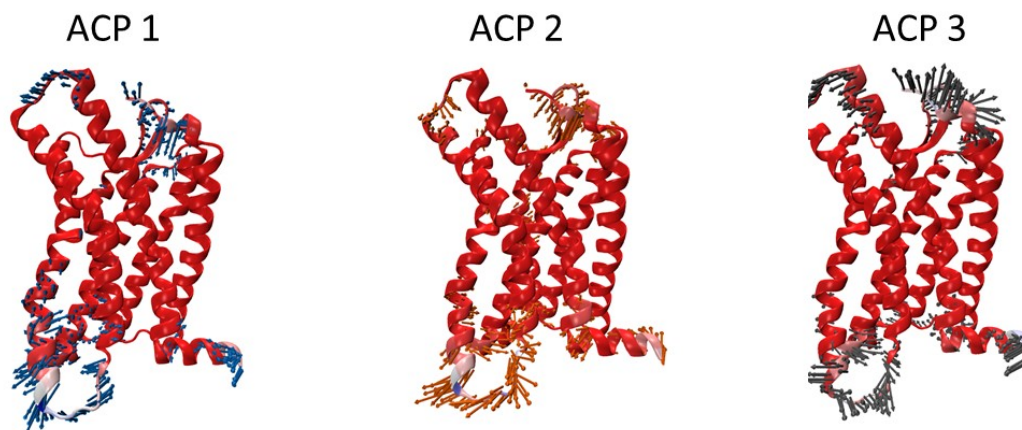
*Annexe 5 : Analyse et image des vecteurs des mouvements des trois premières composantes du système CCL3.*

Les trois premières composantes du système CCR5-CCL3 (ci-dessous) montrent des mouvements de grande amplitude de la boucle ICL3 principalement vers l'extérieur du récepteur. Elles permettent d'observer également des mouvements de va et vient de la boucle ECL2, un déplacement vers l'intérieur du récepteur de la boucle ECL3.



*Annexe 6 : Analyse et image des vecteurs des mouvements des trois premières composantes du système MVC.*

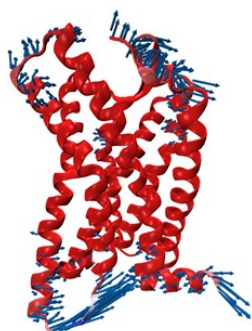
En comparaison avec le système CCL3, l'ACP du système CCR5-MVC (ci-dessous) montre peu de mouvements de grande amplitude. Les seuls mouvements conséquents sont un mouvement de va et vient du bout de l'hélice 8 et un mouvement de cisaillement de la boucle ICL3 caractérisé par des sens de mouvement différents entre la partie de la boucle attachée au TM 5 et la partie attachée au TM6.



*Annexe 7 : Analyse et image des vecteurs des mouvements des trois premières composantes du système T82P.*

Pour le système CCR5-T82P (ci-dessous), les mouvements caractérisés par l'ACP présentent un mouvement de grande amplitude de la boucle ICL3 vers l'hélice 8 et vers l'extérieur de la protéine. Un mouvement de la boucle ECL3 vers l'extérieur est également caractérisé par la seconde composante de l'ACP.

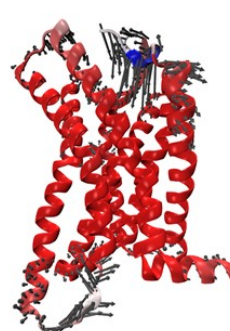
**ACP 1**



**ACP 2**



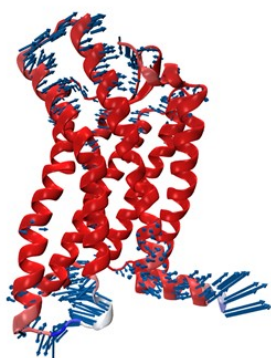
**ACP 3**



*Annexe 8 : Analyse et image des vecteurs des mouvements des trois premières composantes du système P84A.*

Les trois premières composantes du système CCR5-P84A (ci-dessous) montrent des mouvements de grande amplitude de ICL3 vers l'intérieur du récepteur et des légers mouvements de va et vient des boucles ECL2 et ECL3.

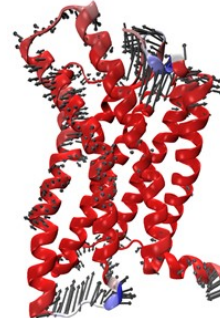
**ACP 1**



**ACP 2**

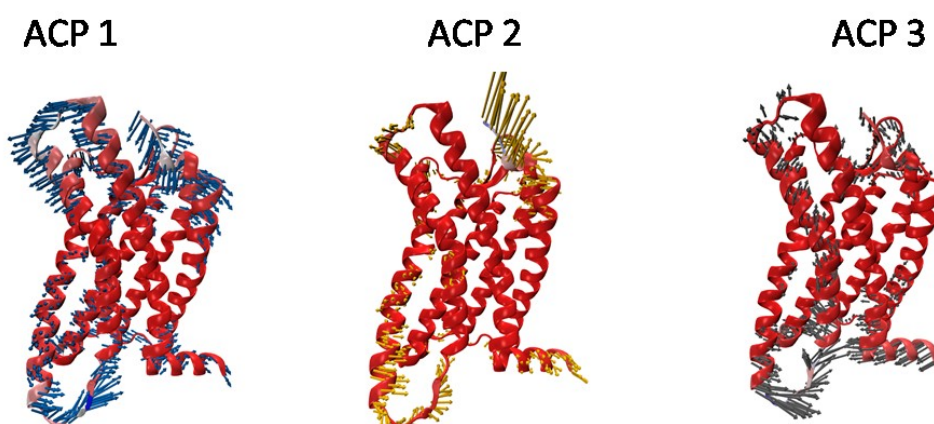


**ACP 3**



*Annexe 9 : Analyse et image des vecteurs des mouvements des trois premières composantes du système libre.*

Le système CCR5-WT est celui où le plus de mouvements au sein de la protéine peuvent être observés (ci-dessous). Des mouvements de grande amplitude de la boucle ICL3 vers l'intérieur et l'extérieur du récepteur sont caractérisés par les trois composantes mais également des mouvements de grande amplitude de la boucle ECL3 vers l'extérieur du récepteur. On pourra noter un léger mouvement de l'hélice 8 de l'extérieur vers l'intérieur et de légers mouvements de l'hélice 6 vers le domaine extracellulaire et du bas de l'hélice 5 vers le domaine intracellulaire.





## **Modélisation du récepteur aux chimiokines C-C de type 5 : caractérisation des états conformationnels et conception rationnelle de modulateurs de la dimérisation**

De nombreuses études des RCPGs révèlent que leur activation n'implique pas que deux états conformationnels, l'un activé et l'autre inactivé, mais une diversité plus importante de ces états, impliquant également des états intermédiaires. Nous avons utilisé la simulation par dynamique moléculaire et l'analyse des interactions intramoléculaires non-covalentes pour étudier la plasticité structurale du récepteur CCR5 sous sa forme monomérique et dimérique. En couplant notre analyse avec diverses données expérimentales, nous avons pu proposer trois architectures dimériques du récepteur et associer des mouvements et des interactions clés aux états conformationnels de CCR5 libre, lié à un agoniste, lié à un agoniste inverse et constitutivement activé ou inactivé par mutation d'acides aminés. Nous avons également développé une méthode d'identification des motifs d'interactions intramoléculaires transmembranaires, permettant de discriminer les états d'activations des RCPGs.

Many studies reveal that GPCR activation does not simply involve two conformational states, one activated and the other inactivated, but a variety of these states coupled to intermediate states. By using molecular dynamics simulations and analyzing non-covalent intramolecular interactions, we studied the structural plasticity of monomeric and dimeric CCR5. By coupling our analysis with various experimental data, we identified three dimeric organizations states of the receptor and associated key motions and intramolecular interactions to free CCR5, bound to an agonist, bound to an inverse agonist and constitutively activated or inactivated by mutated residues. We have also developed a method to identify intramolecular transmembrane interactions patterns, which allow the discrimination of GPCRs activation states.