



HAL
open science

Mécanismes moléculaires de la réponse des récepteurs AT1 et AT2 de l'angiotensine II à des stimuli mécaniques : influence de la composition membranaire

Rym Ben Boubaker

► **To cite this version:**

Rym Ben Boubaker. Mécanismes moléculaires de la réponse des récepteurs AT1 et AT2 de l'angiotensine II à des stimuli mécaniques : influence de la composition membranaire. Médecine humaine et pathologie. Université d'Angers, 2022. Français. NNT : 2022ANGE0011 . tel-04055301

HAL Id: tel-04055301

<https://theses.hal.science/tel-04055301>

Submitted on 2 Apr 2023

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.



Thèse de doctorat Université d'Angers

Ecole doctorale Biologie Santé

Doctorat

Par

Madame Ben Boubaker Rym

Mécanismes moléculaires de la réponse des récepteurs AT1 et

AT2 de l'angiotensine II à des stimuli mécaniques :

Influence de la composition membranaire

Unité de recherche : MITOVASC UMR Inserm U1083 - CNRS 6015

Thèse présentée et soutenue à Angers le 31/03/2022

Composition du Jury

Président : Daniel HENRION

Rapporteurs : Hélène DEMENE

Nicolas FLOQUET

Examineurs : Guillaume LEBON

Thérèse MALLIAVIN

Bernard OFFMANN

Dir. de thèse : Marie CHABBERT

DR1, Université d'Angers

CRCN, Université de Montpellier

DR2, Université de Montpellier

CRCN, Université de Montpellier

DR2 CNRS /institut Pasteur

PR, Université de Nantes

CRHC, Université d'Angers

L'auteur du présent document vous autorise à le partager, reproduire, distribuer et communiquer selon les conditions suivantes :



- Vous devez le citer en l'attribuant de la manière indiquée par l'auteur (mais pas d'une manière qui suggérerait qu'il approuve votre utilisation de l'œuvre).
- Vous n'avez pas le droit d'utiliser ce document à des fins commerciales.
- Vous n'avez pas le droit de le modifier, de le transformer ou de l'adapter.

**Consulter la licence creative commons complète en français :
<http://creativecommons.org/licences/by-nc-nd/2.0/fr/>**

Ces conditions d'utilisation (attribution, pas d'utilisation commerciale, pas de modification) sont symbolisées par les icônes positionnées en pied de page.



REMERCIEMENTS

Tout d'abord, je souhaiterais exprimer ma gratitude à mon encadrante de thèse, le Dr Marie Chabbert, qui m'a accordé sa confiance et m'a accepté dans son équipe alors que j'étais néophyte dans le domaine. Je la remercie aussi pour ses multiples conseils et tout le savoir qu'elle m'a transmis durant ces 3 années. J'aimerais également lui dire à quel point j'ai apprécié ses qualités humaines et sa grande disponibilité. Elle est, pour moi, mon enseignante mais aussi une amie et parfois ... une mère.

Marie tu es sans doute une inspiration et tu fais partie des personnes qui ont marqué ma vie !

Mes remerciements vont également à M. le Dr Daniel Henrion, Directeur de MITOVASC, pour m'avoir accueilli durant 3 ans au sein de son laboratoire.

Je remercie également les membres du jury qui ont pris de leur temps pour examiner et juger ce travail.

Merci à l'école doctorale Biologie-Santé et à l'université d'Angers d'avoir financé mon parcours doctoral. Merci aussi à l'IDRIS pour l'allocation du temps de calcul ainsi que pour leur assistance technique rapide.

Je remercie chaleureusement Asma Tiss, ma chère amie et collègue, avec qui j'ai partagé des moments inoubliables au travail et en dehors du travail.

Je remercie les membres de MITOVASC, en particulier notre *GPCR Team* : Valentine Suteau, Maria Marengo et Fabienne Emeriau pour leur soutien moral ainsi que pour la bonne ambiance dans le bureau.

Je remercie mes amis éparpillés dans le monde entier mais toujours unis. En particulier je cite Zinouba, Shirshah, Eddy, Nourhen, Wala, Mourad, Ichrak, Arij, Ghita, Aymen, MariaCarmela, Pete et Ouyem pour tous les moments authentiques passés ensemble et leurs attentions et encouragements qui m'ont accompagnée tout au long de ces années.

Ces remerciements seraient incomplets si je n'en adressais pas à ma famille en Tunisie.

A mon frère, à tous les moments d'enfance passés avec toi, mon frère, ainsi qu'à ton soutien moral et tes encouragements tout le long de ce parcours.

A ma sœur, tu es toujours la flamme de force dans mon cœur... sans dire un mot tu as réussi à toujours toucher mon âme et mon esprit.

A mes parents qui ont forgé la personne que je suis et qui n'ont jamais cessé de croire en moi et mes capacités. Merci pour les appels journaliers pour s'assurer que j'étais vivante, les colis d'objets et de nourriture bienveillants. Papa, merci pour ta confiance et d'être mon inspiration pour voyager et voir le monde. Maman, merci pour ta générosité, ton amour inconditionnel et tes prières. Vos encouragements furent la torche de lumière dans tout mon parcours et dans ma vie. Je vous aime énormément.

Je dédie cette thèse à ma grand-mère.

Paix à son âme.

Table des matières

TABLE DES ILLUSTRATIONS	VII
LISTE DES TABLEAUX	IX
LISTE DES ANNEXES	IX
LISTE DES ABREVIATIONS	X
AVANT-PROPOS	1
INTRODUCTION	2
SYNTHESE BIBLIOGRAPHIQUE	4
1. Les récepteurs couplés aux protéines G	5
1.1. Présentation des récepteurs couplés aux protéines G	5
1.1.1. Structure des récepteurs couplés aux protéines G	5
1.2. Classification des récepteurs couplés aux protéines G	6
1.3. Les récepteurs de classe A	9
1.3.1. Evolution des récepteurs de classe A par multi-dimensional scaling	9
1.3.2. Structures cristallographiques des GPCR de classe A	9
1.3.3. La nomenclature Ballesteros-Weinstein	11
1.4. Mécanisme d'activation des récepteurs couplés aux protéines G	12
1.4.1. Structures actives et inactives	12
1.4.2. Les motifs d'activation des RCPGs	13
1.4.3. Structures actives non canoniques	14
1.5. Rôle de l'eau dans l'activation des récepteurs couplés aux protéines G	15
1.6. Rôle de l'ion sodium chez les récepteurs couplés aux protéines G	16
1.7. Effet de la composition lipidique dans l'activation des récepteurs couplés aux protéines G	17
1.7.1. Effet du cholestérol sur le fonctionnement des RCPG	17
1.7.2. Effet des lipides anioniques : exemple du phosphatidylglycérol (PG)	18
1.7.3. Lipides à action fluidifiante sur la membrane : les oméga 3	19
1.8. Activation des récepteurs couplés aux protéines G par le stress mécanique	20
1.8.1. Stress mécanique subi par les cellules	20
1.8.2. Réponse des RCPGs au stress mécanique	21
1.8.3. Mécanismes moléculaires d'activation des RCPG par le stress mécanique	22
a) Rôle de l'hélice 8 dans la mécanosensation des RCPG	22
b) Rôle des N-Glycanes dans la mécanosensation des RCPG	23
2. Le système rénine-angiotensine	24
2.1. Généralités	24
2.2. Evolution des récepteurs de l'angiotensine II	25
2.3. Rôle de l'ion sodium dans la régulation des récepteurs de l'angiotensine	26
2.4. Le récepteur AT1	27
2.4.1. Le récepteur AT1 et ses voies de signalisation	27
2.4.2. Les biais de signalisation d'AT1	29
2.4.3. Les structures cristallographiques résolues d'AT1	29
2.4.4. Mécanismes moléculaires des biais de signalisation d'AT1	30
2.5. Le récepteur AT2	31
2.5.1. L'énigme du récepteur AT2	31
2.5.2. Les structures cristallographiques d'AT2	32
2.6. Les récepteurs de l'angiotensine et le stress mécanique	33
2.6.1. Le stress mécanique dans les vaisseaux sanguins	33
2.6.2. AT1 et stress mécanique	34
2.6.3. Exemples de pathologies où AT1 et le stress mécanique sont impliqués	35
a) Exemple 1: l'athérosclérose	35
b) Exemple 2: Cancer et angiogenèse	36
3. Modélisation et dynamique moléculaires	37
3.1. La modélisation moléculaire	37
3.1.1. La modélisation par homologie	38
3.1.2. La modélisation en « Twilight zone »	39

a)	Recherche des patrons par les méthodes d'enfilage ou « threading »	39
b)	Choix des patrons en utilisant le MDS	39
3.1.3.	Les outils de modélisation moléculaire	40
3.2.	Les simulations dynamiques moléculaires classiques	41
3.2.1.	Définition et applications	41
3.2.2.	Principe et algorithmes.....	42
3.2.3.	Les champs de force	44
3.2.4.	Les conditions limites périodiques.....	44
3.2.5.	Les ensembles thermodynamiques	45
3.2.6.	Outils des simulations dynamiques	46
3.3.	Les simulations dynamiques moléculaires dirigées	46
3.3.1.	Principes.....	46
3.3.2.	Quaternion d'orientation.....	46
MATERIELS & METHODES		48
4.	Matériel	49
4.1.	Migration vers le supercalculateur Jean-Zay et optimisation du temps de calcul	49
5.	Méthodes	49
5.1.	Modélisation moléculaire des récepteurs	49
5.1.1.	Modélisation du récepteur AT1	49
5.1.2.	Modélisation du récepteur AT2	50
5.2.	Modélisation du stress mécanique	50
5.2.1.	Modélisation du stress mécanique en tension de surface.....	50
5.2.2.	Suivi des paramètres physiques et calcul de la tension de surface	50
5.3.	Les simulations de dynamique moléculaire	51
5.3.1.	Les simulations dynamiques classiques	51
5.3.2.	Les simulations dynamiques dirigées (SMD)	52
a)	Mise au point des simulations dynamiques dirigées par quaternion	52
b)	Etape des SMD	53
5.4.	Analyses des simulations dynamiques moléculaires	54
5.4.1.	Visualisation et analyses des trajectoires.....	54
5.4.2.	Analyse des changements conformationnels	54
5.4.3.	Analyse de la coordination de l'ion sodium	55
5.4.4.	Quantification des molécules d'eau	55
5.4.5.	Quantification des liaisons H	56
5.4.6.	Caractérisation des lipides	56
RESULTATS		57
6.	Effet du stress mécanique sur les paramètres physiques du système de simulation.....	58
6.1.	Effet du stress mécanique sur les dimensions de la boîte de simulation	58
6.2.	Calcul de la tension de surface et de la pression latérale du système	58
6.3.	Effet du stress mécanique sur les caractéristiques des lipides.....	59
6.4.	Distribution des lipides autours du récepteur	60
7.	Simulations de dynamique moléculaire classique des récepteurs AT1 et AT2 sous l'effet d'un stress mécanique.....	61
7.1.	Observations générales	61
7.2.	Activation du récepteur AT1	65
a)	Analyse détaillée.....	65
b)	Interactions de l'hélice 8 avec les DOPG	66
c)	Screening des conformations actives des RCPG.....	66
7.3.	Activation du récepteur AT2.....	67
7.3.1.	Activation du récepteur AT2 en présence de l'ion sodium dans une bicouche de POPC	67
a)	Changement conformationnel et verrouillage ionique.....	67
b)	Rôle de l'eau dans l'activation du récepteur	68
7.3.2.	Activation du récepteur AT2 en absence de l'ion sodium dans une bicouche de POPC.....	69
7.3.3.	Activations d'AT2 en l'absence de sodium dans une bicouche de DOPG	70
7.4.	Structures canoniques et non canoniques.....	70
7.5.	Effets de la tension sur la coordination sodium	71
8.	Analyse des transitions conformationnelles par SMD	73

8.1.	Récepteur AT1.....	73
8.1.1.	Observations générales	73
8.1.2.	Les liaisons H	74
8.1.3.	Validation par mutagenèse in silico	75
8.1.4.	Comparaison avec le récepteur AT1 d'anguille	75
8.2.	Récepteur AT2.....	76
8.3.	Mécanismes proposés	77
8.4.	Les transitions conformationnelles d'AT1 dans une bicouche de DOPG.....	79
8.5.	Les transitions conformationnelles d'AT1 sous l'effet de la tension	79
	DISCUSSION	81
	CONCLUSION & PERSPECTIVES	89
	BIBLIOGRAPHIE	91
	ANNEXE I	103
	ANNEXE II.....	110

Table des illustrations

Figure 1 : Structure des récepteurs couplés aux protéines G.	5
Figure 2 : Voies de signalisation des récepteurs couplés aux protéines G.	6
Figure 3 : Exemples de structures cristallographiques des différentes classes des RCPG.	8
Figure 4 : Evolution des RCPG humains de classe A non olfactifs.	9
Figure 5 : Evolution du nombre de structures cristallographiques de RCPG.	10
Figure 6 : Arbre évolutif des RCPG humains de classe A indiquant les structures résolues.	11
Figure 7 : Changements conformationnels des RCPG de classe A lors de l'activation.	12
Figure 8 : Les cinq motifs structuraux conservés des RCPG de classe A.	14
Figure 9 : Comparaison structurale de hNTR1 dans les états canonique C et non canonique NC.	15
Figure 10 : Canal d'eau du récepteur A2AR actif couplé à la protéine G α et au ligand NECA.	16
Figure 11 : Rôle de l'ion sodium chez les RCPG de classe A.	17
Figure 12 : Sites de fixation du cholestérol dans les RCPG de classe A.	18
Figure 13 : Présentation schématique des phosphatidycholines et les phosphatidylglycerols.	19
Figure 14 : Influence de la composition membranaire sur le nombre de contacts entre récepteurs.	20
Figure 15 : Stress mécanique subi par les cellules endothéliales et mécanosenseurs potentiels.	21
Figure 16 : Rôle de l'hélice 8 dans l'activation des RCPG par le stress mécanique.	23
Figure 17 : Rôle des N-glycanes dans la mécanosensation des RCPG.	23
Figure 18 : Schéma simplifié du système rénine-angiotensine	25
Figure 19 : Arbre phylogénique des récepteurs AT1 et AT2.	26
Figure 20 : Visualisation de l'entrée de l'ion sodium dans AT1 humain.	27
Figure 21 : Schéma simplifié des voies de signalisation médiées par le récepteur AT1.	28
Figure 22 : Structures cristallographiques d'AT1 présentes dans la Protein Data Bank.	30
Figure 23 : Représentation schématique du rôle de N3.35 dans l'activation biaisée d'AT1.	31
Figure 24 : Schéma simplifié des voies de signalisation médiées par le récepteur AT2.	32
Figure 25 : Structures cristallographiques d'AT2 présentes dans la Protein data Bank.	33
Figure 26 : Contraintes mécaniques associées au flux sanguin.	33
Figure 27 : Mécanisme autocrine d'activation du récepteur AT1 par l'effet d'étirement.	34
Figure 28 : Activation du récepteur AT1 par le stress mécanique.	35
Figure 29 : Présentation schématique des différentes zones de modélisation moléculaire des protéines.	38
Figure 30 : Analyse statistique des longueurs de boucle dans les RCPG humains (classe A).	39
Figure 31 : Informations sur l'évolution de PRF1.	40
Figure 32 : Comparaison des modèles de PRF1.	41
Figure 33 : Exemples d'application de la dynamique moléculaire	42
Figure 34 : Equations et représentation schématique des énergies liées et non liées.	44
Figure 35 : Conditions limites périodiques.	44
Figure 36 : Modélisation du stress mécanique par une tension de surface.	45
Figure 37 : Evolution du temps de calcul physique et alloué en fonction du nombre de nœuds.	49

Figure 38 : Suivi des paramètres physiques et calcul de la tension de surface.	51
Figure 39 : Etapes des simulations de dynamique moléculaire.	52
Figure 40 : Description 3D de l'ouverture de l'hélice TM6 sous forme de quaternion.	53
Figure 41 : Simulations dynamiques dirigées.	54
Figure 42 : Distances entre TM3 et TM6 et entre TM3 et TM7.	55
Figure 43 : Zones de quantification des molécules d'eau dans la cavité interne du récepteur AT1.	55
Figure 44 : Paramètres de caractérisation des lipides.	56
Figure 45 : Evolution des dimensions de la boîte de simulation en fonction de la tension.	58
Figure 46 : Calcul de la tension de surface et de la pression latérale du système en fonction de la tension.	59
Figure 47 : Comparaison des propriétés physicochimique des POPC et des DOPG.	60
Figure 48 : Distributions des lipides autour du récepteur AT1 pour une membrane POPC/DOPG (90:10).	61
Figure 49 : Heatmaps des distances entre TM3 et TM6 lors des simulations dynamiques classiques.	63
Figure 50 : Snapshots des différents états observés chez AT2 dans une bicouche de DOPG.	64
Figure 51 : Déstabilisation d'AT1 sans sodium à haute tension.	64
Figure 52 : Activation du récepteur AT1 par le double effet de l'étirement et des lipides anionique DOPG.	65
Figure 53 : Liaisons hydrogènes entre l'hélice 8 d'AT1 et les molécules DOPG.	66
Figure 54 : Comparaison de la structure active d'AT1 obtenue par MD et des structures cristallographiques d'AT1 et de NTSR1.	67
Figure 55 : Activation d'AT2 en présence de sodium dans une bicouche de POPC à la tension de 20 dyne/cm.	68
Figure 56: Devenir de l'ion sodium pendant l'activation transitoire du récepteur AT2.	69
Figure 57 : Activation d'AT2 en l'absence sodium dans une bicouche de POPC à la tension de 20 dyne/cm.	69
Figure 58 : Activations d'AT2 en absence de l'ion sodium dans une bicouche de DOPG.	70
Figure 59 : Bilan des activations observées lors des simulations dynamiques classique.	71
Figure 60 : Mode de liaison de l'ion sodium sous l'effet de la tension chez AT1 et AT2	72
Figure 61 : Énergie libre électrostatique de l'interaction de l'ion sodium avec les récepteurs AT1 et AT2.	72
Figure 62 : Transition conformationnelle d'AT1 par SMD dans une bicouche de POPC.	74
Figure 63 : Rôle des liaisons H dans la formation de la structure non canonique d'AT1.	74
Figure 64 : Transition conformationnelle d'AT1 D6.32A/K7.58A par SMD dans une bicouche de POPC.	75
Figure 65 : Transition conformationnelle d'AT1 d'anguille par SMD dans une bicouche de POPC.	76
Figure 66 : Transition conformationnelle d'AT2 par SMD dans une bicouche de POPC.	77
Figure 67 : Etapes mécanistiques d'activation du récepteur AT1 dans une bicouche de POPC.	78
Figure 68 : Etapes mécanistiques d'activation du récepteur AT2 dans une bicouche lipidique POPC.	78
Figure 69 : Transition conformationnelle d'AT1 par SMD dans une bicouche de DOPG.	79
Figure 70 : Transition conformationnelle d'AT1 dans une bicouche de POPC sous tension de 20 dyne/cm.	80
Figure 71: Effet des lipides anioniques sur AT1.	84
Figure 72 : Sites allostériques potentiels d'AT1.	85
Figure 73 : Comparaison du verrouillage ionique entre AT1 et AT2.	86
Figure 74 : Liaisons H des résidus 6.32 et 7.58 dans les simulations dynamiques dirigées.	87
Figure 75 : Logoplots des résidus 6.32 et 7.58 des récepteurs AT1 et AT2 terrestres et aquatiques.	87

Liste des tableaux

Tableau 1 : RCPG mécanosensibles. ¹	21
Tableau 2 : Les ligands d'AT1 et leurs fonctions	29
Tableau 3 : Paramètres des simulations dynamiques de bicouches lipidiques ¹	59
Tableau 4 : Simulations dynamiques classiques d'AT1 et AT2 ¹	62

Liste des annexes

Annexes I : Résultats supplémentaires

Figure S1 : RMSD et RMSF des simulations dynamiques classiques.	104
Figure S2 : Heatmaps des distances entre les hélices 3 et 7 (Distance 3.50-7.55) lors des simulations dynamiques classiques.	105
Figure S3 : Heatmaps des distances entre l'ion sodium et les résidus protéiques de son site lors des simulations dynamiques classiques.	106
Figure S4 : Tableau des moyennes et écarts type des liaisons H entre les résidus 6.32 et 7.58 lors des simulations dynamiques classiques (n=3).	107
Figure S5 : Heatmaps des liaisons H entre les résidus 6.32 et 7.58 lors des simulations dynamiques classiques	108
Figure S6 : Heatmaps des simulations dynamiques dirigées	109

Annexe II: Publications

Annexe II.1: Article

Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota.

Annexe II.2: Chapitres méthodologiques

Chapitre 1: Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom.

Chapitre 2: Homology modeling in the twilight zone: Improved accuracy by sequence space analysis.

Annexe II.3: Abstracts d'autres publications

Abstract 1: G Protein-coupled Receptors in Radioiodine-refractory Thyroid Cancer in the Era of Precision Medicine.

Abstract 2: Identification of dysregulated expression of G protein coupled receptors in endocrine tumors by bioinformatics analysis: potential drug targets?

Liste des abréviations

RCPG : Récepteur couplé aux protéines G
AT1 : récepteur de l'angiotensine type 1
AT2 : récepteur de l'angiotensine de type 2
 β 2AR : récepteur β 1 adrenergic receptor
NTSR1 : récepteur à la neurotensine 1
 μ OR : récepteur μ -opioïde
 A_{2A} AR : récepteur de l'adénosine
CCR5: chemokine receptor type 5
CXCR4: chemokine receptor type 4
OPRD : récepteur opoïde de type β
PAR1: Protease activated receptor
aMD: accelerated Molecular Dynamics
SMD: steered Molecular Dynamics
cMD: classical Molecular Dynamics
 C_{α} : carbones α
CA : carbones de système protéiques
DM : Dynamique Moléculaire
ECL : boucle extracellulaire
ICL : boucle intracellulaire
H₂O : molécule d'eau
IDRIS : Institut du développement et des ressources en informatique scientifique
NAMD : Nanoscale Molecular Dynamics
NPxxY : acides aminés de la terminaison de l'hélice 7
PDB : Protein Data Bank
RMN: Résonance magnétique nucléaire
RMSD: Root Mean Square Deviation (Å°)
RMSF: Root Mean Square Fluctuations (Å°)
TIP3P : modèle des molécules d'eau
TM : hélice transmembranaire
VMD: Visual Molecular Dynamics
G: Glycine (Gly)
P : Proline (Pro)
A : Alanine (Ala)
V : Valine (Val)
L : Leucine (Leu)
I : Isoleucine (Ile)
M : Methionine (Met)
C : Cysteine (Cys)
F : Phenylalanine (Phe)
Y : Tyrosine (Tyr)
W : Tryptophan (Trp)
H : Histidine (His)
K : Lysine (Lys)
R : Arginine (Arg)
Q : Glutamine (Gln)
N : Asparagine (Asn)
E : Glutamic Acid (Glu)
D : Aspartic Acid (Asp)
S: Serine (Ser)
T: Threonine (Thr)
SCD: Deuterium order parameter

DOPG: 1,2-Dioleoyl-sn-Glycero-3-[Phospho-rac-(1-glycerol)]
POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PC: phosphatidylcholine
PG: phosphatidylglycerol
Etat NC: état non canonique
Etat C: état canonique
MDS: Multidimensional scaling
CRAN: Comprehensive R Archive Network
bios2mds: from Biological Sequences to Multidimensional Scaling
cAMP: Cyclic adenosine monophosphate
ERK: extracellular-signal-regulated kinase
SRA: système rénine-angiotensine
ACE: enzyme de conversion de l'angiotensine

Avant-propos

Ce rapport présente les résultats obtenus durant ma thèse au sein du groupe "Evolution, Structure et Fonction des RCPG" sous la direction du Dr Marie Chabbert de l'équipe CARME du laboratoire MITOVASC à l'université d'Angers.

L'objectif principal de notre groupe est de mieux comprendre les mécanismes moléculaires conduisant à l'activation des RCPG et à la sélectivité fonctionnelle (activation d'une voie spécifique), avec un but à long terme dans la conception de médicaments. Notre approche est basée sur le couplage entre les simulations de dynamique moléculaire et les informations évolutives. L'analyse de l'évolution révèle des positions dont la mutation a été cruciale pour l'évolution et la divergence des familles de RCPG et peut donc avoir un rôle important dans la spécificité des récepteurs et leur mécanisme d'action.

Au sein de notre groupe, nous nous concentrons sur les mécanismes moléculaires conduisant à l'activation et à la signalisation biaisée des récepteurs AT1 et AT2 de l'angiotensine II en réponse à la liaison du ligand ou à des stimuli physiques. Plus précisément, cet objectif a été subdivisé en 2 projets visant à répondre à 2 questions principales:

Projet 1 : Quel est le rôle de l'ion sodium dans le mécanisme d'activation et la signalisation biaisée ?

Projet 2 : Quels sont les effets du stress mécanique (pression, tension, forces de cisaillement) et de la composition de la membrane sur l'activation des récepteurs ?

Le projet 2 est mon sujet de thèse principal. Il s'intéresse à élucider les mécanismes moléculaires de la réponse des récepteurs AT1 et AT2 de l'angiotensine II à des stimuli mécaniques et à étudier l'influence de la composition membranaire. Ce projet a été initié avec mon arrivée au laboratoire, il a demandé de nombreux développements méthodologiques et d'approches exploratoires de ma part. C'est donc lui que je présenterai en détail dans cette thèse, accompagné de résultats complémentaires en Annexe I.

Pendant mes 3 années de thèse, je me suis aussi impliquée dans le premier projet sur le rôle de l'ion sodium et l'évolution des récepteurs AT1 et AT2. Ce projet a fait l'objet d'une publication récente dont je suis co-premier auteur. Elle est intitulée «Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota» (*Plos Comput Biol*, 2022, doi: 10.1371/journal.pcbi.1009732). Je présenterai cet article, accompagné de quelques commentaires pour y détailler mon rôle en Annexe II.1.

Au cours de cette thèse, j'ai aussi participé à deux articles méthodologiques, publiés sous forme de chapitres dans *Methods in Molecular Biology*. Ils seront présentés en Annexe II.2. Finalement, j'ai collaboré en support pour l'analyse des données à plusieurs projets de mes collègues. Les abstracts seront présentés en Annexe II.3.

Certains des résultats des articles en Annexe sont insérés dans la partie bibliographique soit à titre didactique pour les chapitres méthodologiques, soit pour leur importance pour mon projet de thèse principal.

INTRODUCTION

Les récepteurs couplés aux protéines G (RCPG) sont des récepteurs transmembranaires qui agissent comme une boîte de communication transformant des messages extracellulaires en réponses cellulaires. Les RCPG jouent un rôle important dans des nombreuses fonctions physiologiques et pathologies du corps humain. Une meilleure compréhension de ces récepteurs aura un impact considérable sur la conception des médicaments. En effet, les chercheurs estiment qu'environ un tiers de tous les médicaments commercialisés agissent en se liant aux RCPG [1]. L'évolution technologique et la disponibilité récente de nombreuses structures cristallines de RCPG contribuent fortement à la conception de nouveaux ligands basée sur leur structure.

Un des systèmes les plus ciblés est le système rénine-angiotensine (SRA). Ce système joue un rôle essentiel dans la physiopathologie cardiovasculaire et rénale, en exerçant un large éventail de processus homéostatiques et modulateurs, notamment la régulation de l'équilibre hydrique et salin, la vasoconstriction, le remodelage cellulaire/tissulaire et le dysfonctionnement du système cardiovasculaire [2].

Au cœur de ce système se trouve l'angiotensine II (AngII), un peptide bioactif. Un grand nombre de ses effets s'exerce en se liant à deux récepteurs couplés aux protéines G : le récepteur de l'AngII de type 1 (récepteur AT1) et le récepteur de l'AngII de type 2 (récepteur AT2) [3,4]. AT1 est exprimé dans la plupart des organes et des tissus et est fortement impliqué dans l'hypertension, le dysfonctionnement endothélial et les maladies cardiovasculaires [5]. L'activation du récepteur AT1 médiée par l'AngII induit l'activation de diverses voies de signalisation cytoplasmique qui conduisent, entre autres, au remodelage des cellules musculaires lisses vasculaires (CMLV) et des cardiomyocytes, y compris l'hypertrophie et la migration [6].

AT2 est exprimée majoritairement dans la vie fœtale. Son expression diminue rapidement après la naissance, sauf dans des conditions pathologiques ou certains organes comme l'utérus ou le cerveau. Le récepteur AT2 est décrit comme ayant des effets opposés à AT1 [7]. Il confère une protection cardiaque et une diminution de la pression artérielle mais le mécanisme d'action de ce récepteur reste inconnu à ce jour [8].

Le mécanisme d'activation d'AT1 et AT2 par l'angiotensine II est beaucoup plus clair depuis la récente résolution des structures cristallographique d'AT1 (pdb ID : 6OS0) [9] et AT2 (pdb ID : 6JOD) [8] complexés avec l'AngII. Néanmoins de nombreuses questions restent en suspens. En effet, ces structures donnent une image statique des récepteurs et ne reflètent pas leur nature dynamique qui est fondamentale pour leurs fonctions *in vivo* (ref Latorraca et al., 2017)

Des études expérimentales montrent que certains RCPG vasculaires, dont AT1, sont mécanosensibles et activables par les forces mécaniques [10], soit par l'effet des forces de cisaillement générées par le flux sanguin [11] ou par l'effet de l'étirement membranaire [12], généré par l'environnement des vaisseaux. Cette activation peut être directe ou indirecte par l'intermédiaire de la membrane lipidique. En effet, des études montrent que la tension de surface membranaire et les propriétés physico-chimiques des lipides sont extrêmement sensibles aux forces mécaniques [13]. De plus, une composition lipidique particulière peut induire l'activation spontanée de certains GPCR, en particulier les lipides anioniques [14]. Parmi ces lipides, on note la famille des phosphatidylglycérols tel que les DOPG (1,2-Dioleoyl-sn-Glycero-3-[Phospho-rac-(1-glycerol)]) [15,16].

Pour le récepteur AT1, l'activation générée par le stress mécanique contribue à des voies de signalisation semblables à celles induites par l'angiotensine II [11], mais l'influence de la composition membranaire et le mécanisme moléculaire d'activation restent inconnus.

L'objectif de ce projet est d'approfondir la compréhension des mécanismes moléculaires conduisant à l'activation d'AT1 et AT2 en absence du ligand naturel, l'angiotensine II, en particulier par l'effet du stress mécanique. Il s'agira en particulier d'éclaircir le rôle de l'eau, de l'ion sodium et de la composition lipidique dans les changements conformationnels de ces récepteurs induits par un stress mécanique.

En utilisant une approche *in silico* de modélisation moléculaire et de simulations de dynamique moléculaire classique et dirigée, nous visons à répondre aux questions suivantes :

- Des facteurs mécaniques modifient-ils la conformation des récepteurs AT1 et à l'AT2, en absence de ligand, influençant ainsi l'activité basale des récepteurs et la réponse ultérieure à la liaison du ligand ?
- Certains types de lipides dans l'environnement des récepteurs peuvent-ils agir sur l'activation des récepteurs par stress mécanique ? Si oui, leur action est-elle due à une interaction spécifique avec le récepteur ou à une modification de la fluidité membranaire ?
- Quelles sont les étapes du mécanisme d'activation d'AT1 et d'AT2 affectées par les facteurs externes que sont le stress mécanique et les lipides de l'environnement ?
- Quelles sont les différences structurales entre AT1 et AT2 qui peuvent avoir un rôle dans la réponse différentielle de ces deux récepteurs aux facteurs externes ?

Les réponses à ces questions visent à comprendre la base moléculaire de l'activation des récepteurs AT1 et AT2 et le rôle supposé de divers facteurs (composition de la membrane et stress mécanique) dans ce mécanisme.

Dans ce rapport de thèse, nous allons tout d'abord commencer par une synthèse bibliographique présentant tout d'abord les récepteurs couplés aux protéines G, en insistant sur l'effet du stress mécanique et des lipides sur l'activité de ces récepteurs en absence du ligand naturel, et sur les récepteurs de l'angiotensine II, AT1 et AT2, puis les principes de la modélisation moléculaire et des simulations dynamiques classiques et dirigées. Enfin, nous présenterons le sujet principal de thèse, en décrivant en détail les méthodes utilisées et nos résultats, pour les discuter ainsi que les perspectives qu'ils ouvrent. Les travaux complémentaires que j'ai réalisés au cours de cette thèse seront présentés en Annexe.

SYNTHESE

BIBLIOGRAPHIQUE

1. Les récepteurs couplés aux protéines G

1.1. Présentation des récepteurs couplés aux protéines G

1.1.1. Structure des récepteurs couplés aux protéines G

Les récepteurs couplés aux protéines G (RCPG) constituent la plus grande famille de protéines membranaires du génome humain et une cible importante pour l'industrie pharmaceutique [1].

Le génome humain présente plus de 800 RCPG. Chacun d'entre eux est hautement spécifique pour un signal particulier [17]. Les RCPG sont constitués d'une seule chaîne polypeptidique intégrée à la membrane plasmique. Ils sont formés de 7 segments transmembranaires en hélice α reliés par des boucles à l'intérieur (ICL) et à l'extérieur (ECL) de la cellule. Ils sont activés par la fixation de ligands spécifiques, appelés agonistes. Pour la majorité des récepteurs, la cavité extracellulaire est le principal site de fixation des agonistes (site orthostérique) auquel peuvent contribuer les boucles extracellulaires [17]. Il peut aussi exister des sites de fixation dits allostériques, distincts du site orthostérique qui permettent à un ligand « allostérique » de moduler la fixation de l'agoniste et l'activité du récepteur [18]

Les ligands agonistes des RCPG sont chimiquement et physiquement diversifiés et peuvent inclure : des photons, des ions (H^+ , Zn^{++} , Ca^{++} , etc.), des substances odorantes; des substances gustatives, des vitamines (niacine, vitamine A, etc.), des protéines (chimiokines, hormones glycoprotéiques, etc.), des peptides, des amines (dopamine, sérotonine, etc.), des nucléotides, des lipides et un grand nombre de métabolites intermédiaires [18]. L'activation des RCPG se traduit par divers signaux médiés par les protéines G hétérotrimériques, les β -arrestines ou d'autres effecteurs (Figure 1) [18]. Ces voies de signalisation contrôlent de nombreux processus biologiques et pathologiques majeurs dans les systèmes neuronaux, cardiovasculaires, immunitaires et endocriniens [19]. En conséquence, les RCPG sont la cible thérapeutique de plus de 35 % des médicaments cliniquement approuvés. Environ 700 médicaments approuvés ciblent 134 RCPG [1].

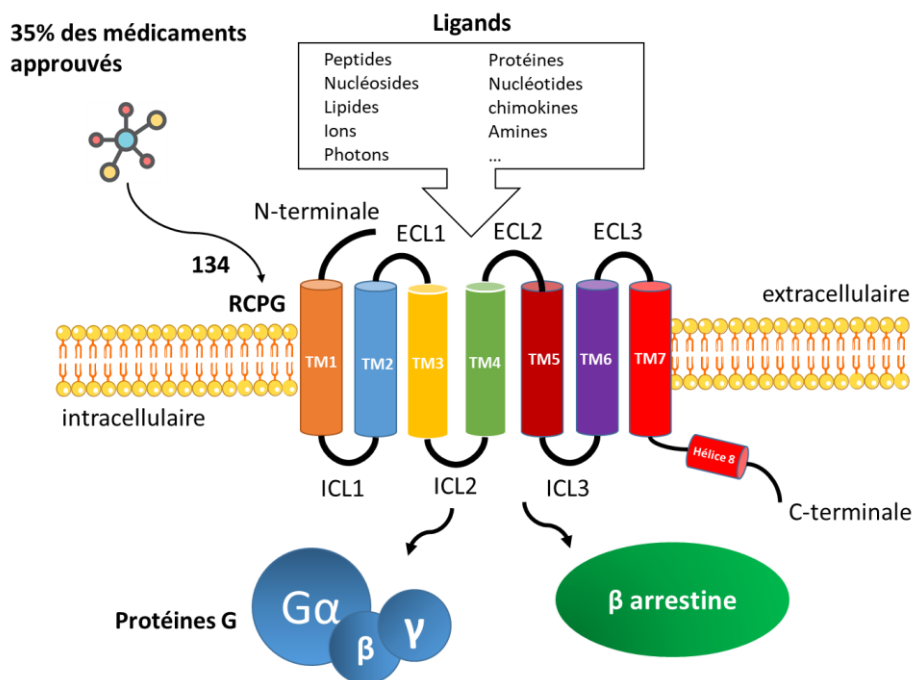


Figure 1 : Structure des récepteurs couplés aux protéines G.

Les RCPG constituent donc une grande famille de récepteurs qui se lient à une grande variété de ligands et sont impliqués dans diverses voies de signalisation. La liaison d'un agoniste sur son récepteur induit un réarrangement conformationnel et le passage d'un état inactif vers un état actif de celui-ci. Ce changement structural permet le couplage d'une protéine effectrice (protéine G ou β -arrestine) et l'initiation d'une cascade de voies de signalisation intracellulaire.

Les protéines G sont composées de 3 sous-unités $G\alpha$, $G\beta$ et $G\gamma$. Elles sont classées en quatre familles (G_i , G_s , $G_{12/13}$, et G_q) en fonction de leur sous-unité α qui induit des réponses intracellulaires différentes. Les protéines G_s et G_i régulent l'activité de l'adénylyl cyclase, tandis que G_q active la phospholipase $C\beta$ et que $G_{12/13}$ peut activer les familles de petites GTPases Rho. La famille G_q comprend quatre membres : G_q , G_{11} , G_{14} , et $G_{15/16}$ [20]. Certains récepteurs peuvent se lier à différentes protéines G. A l'état inactif, les protéines G sont hétérotrimériques et la sous-unité $G\alpha$ fixe une molécule de GDP. Lors de l'activation par la fixation au récepteur, la sous-unité $G\alpha$ libère le GDP et fixe du GTP. Cet échange est suivi par la dissociation du récepteur de $G\alpha$ d'un côté et du dimère $G\beta$ - $G\gamma$ de l'autre. $G\alpha$ reste actif jusqu'à l'hydrolyse du GTP en GDP qui permet la reformation de l'hétérotrimère inactif. A l'état actif, $G\alpha$ active les enzymes cibles, libérant des effecteurs secondaires tels que l'AMPc, les phosphates d'inositol et l'ion Ca^{2+} , induisant différentes réponses cellulaires [21].

Le récepteur actif peut aussi fixer les β -arrestines, en particulier après avoir été phosphorylé par des kinases spécifiques [21]. Les β -arrestines désensibilisent le récepteur en empêchant les protéines G de se coupler au récepteur activé et activent l'internalisation du récepteur par endocytose. De plus, les β -arrestines elles-mêmes déclenchent des voies de signalisation spécifiques, indépendantes des protéines G [21] (Figure 2).

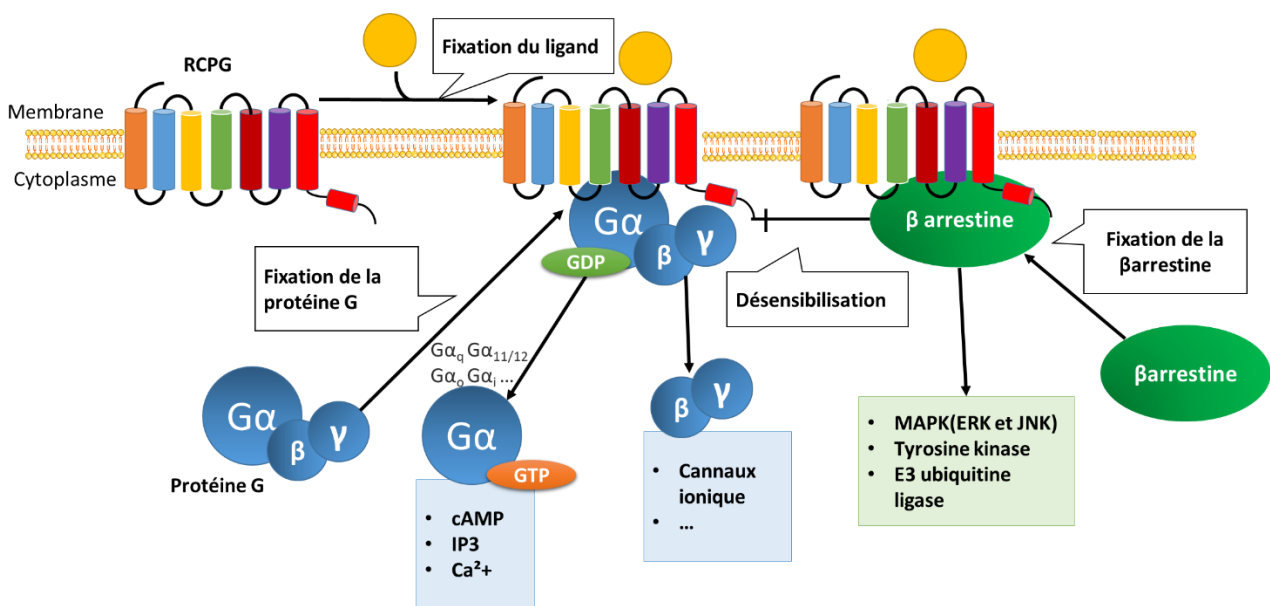


Figure 2 : Voies de signalisation des récepteurs couplés aux protéines G.

1.2. Classification des récepteurs couplés aux protéines G

En 2003, Fredriksson *et al.* ont identifié plus de 800 séquences de RCPG humains et réalisé une analyse phylogénétique de 342 séquences uniques de RCPG humains fonctionnels non olfactifs. Ils ont identifié 5

grandes familles, nommées Glutamate, Rhodopsine, Adhésion, Frizzled/taste2, et Sécrétine, formant ainsi le système de classification GRAFS (Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, Secretin) [22,23].

Famille de la rhodopsine (classe A) : C'est la plus grande famille de RCPG. Elle comprend environ 700 membres chez les humains, avec environ 400 récepteurs olfactifs et 300 récepteurs non olfactifs. Selon Fredriksson *et al.*, cette famille est en outre divisée 4 groupes principaux [22]:

Groupe α : Ce groupe comprend les récepteurs des prostaglandines, les récepteurs des amines, les opsines et le groupe des récepteurs MECA (Melanocortin, EDG, Cannabinoïd, Adenosine).

Groupe β : Ce groupe n'a pas de branche principale et comprend 36 récepteurs (Figure 3). Tous les ligands connus de ces récepteurs sont des peptides (par exemple, les récepteurs de l'hypocrétine (HCRTRs), les récepteurs du neuropeptide FF (NPFFs), les récepteurs de la tachykinine (TACR), etc.)

Groupe γ : Ce groupe a trois branches principales : le groupe des récepteurs SOG (somatostatine, opioïde, galanine), le groupe des récepteurs MCH (melanin-concentrating hormone), et le groupe des récepteurs des chimiokines.

Groupe δ : Ce groupe comprend trois branches principales : les récepteurs de la famille MAS, le groupe des récepteurs des hormones glycoprotéines et le groupe des récepteurs des purines. Les récepteurs olfactifs se brancheraient sur ce groupe.

Famille des récepteurs de la sécrétine (classe B1) : Elle comprend environ 60 membres et se caractérise par la présence d'un grand domaine N-terminal (entre 60 et 80 acides aminés), appelé ectodomaine, contenant des ponts disulfure conservés. L'ectodomaine est particulièrement important pour la liaison du ligand à ces récepteurs [22]. Les RCPG de la famille B ont une morphologie similaire à celle des récepteurs d'hormones glycoprotéiques (groupe δ de la classe A), mais ils ne partagent pas d'homologie de séquence. Leurs ligands comprennent des hormones de poids moléculaire élevé telles que le glucagon, la sécrétine, la calcitonine, etc. [23,24]

Famille des récepteurs d'Adhésion (classe B2) : appelé B2 en raison de leur similarité avec les récepteurs de la sécrétine pour la partie transmembranaire. Les membres de cette famille sont définis par leur structure particulière avec une grande région extracellulaire. Les extrémités N-terminales sont de longueur variable, d'environ 200 jusqu'à 2800 acides aminés, et sont souvent riches en sites de glycosylation et en résidus proline, formant ce qui a été décrit comme des tiges ressemblant à des mucines susceptibles de participer à l'adhésion cellulaire [22,23] .

Famille des récepteurs de glutamate (classe C) : La famille C recrute environ deux douzaines de RCPG tels que les récepteurs métabotropiques du glutamate (mGluR) et les récepteurs de détection du Ca^{2+} . Cette famille comprend également les récepteurs $\text{GABA}_{\text{B}1}$ (2 variants d'épissage) et $\text{GABA}_{\text{B}2}$ et certains récepteurs du goût TAS1. Comme la famille B, ces récepteurs possèdent de grands ectodomains responsables de la liaison du ligand [22,23].

Famille des récepteurs Frizzled/taste2 (classe F) : Cette famille comprend 2 groupes : les récepteurs Frizzled et les récepteurs de goût TAS2 [22]. Les récepteurs Frizzled sont définie par des caractéristiques structurales conservées, avec un domaine de liaison au ligand riche en cystéines. Les récepteurs Frizzled sont les récepteurs des protéines sécrétées Wnt (glycoprotéines), ainsi que d'autres ligands. Ils jouent un rôle essentiel dans le développement embryonnaire et, en particulier, dans la polarité et la segmentation cellulaire [22,25]

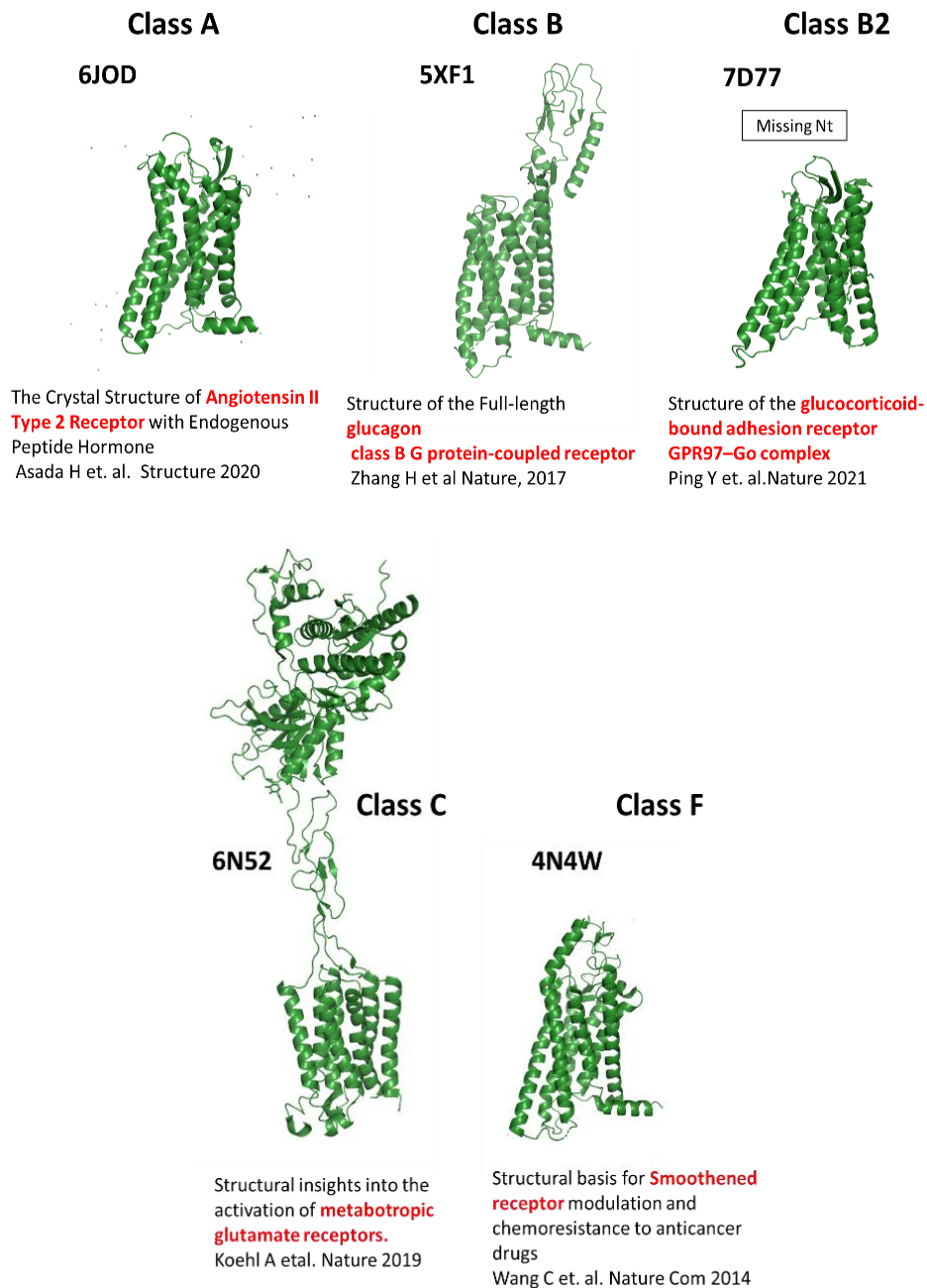


Figure 3 : Exemples de structures cristallographiques des différentes classes des RCPG.
Les structures sont issues de la Protein Data Bank. Les légendes correspondent à l'article correspondant.

1.3. Les récepteurs de classe A

1.3.1. Evolution des récepteurs de classe A par multi-dimensional scaling

La famille des récepteurs de classe A représente la famille la plus diversifiée avec environ 700 récepteurs olfactifs et non olfactifs. Ses membres sont caractérisés par des motifs de séquence conservés qui impliquent des caractéristiques structurales et des mécanismes d'activation communs [26].

Notre équipe a contribué à la classification des récepteurs de classe A non olfactifs en utilisant l'approche de multi-dimensional scaling [27,28]. Cette méthode permet d'avoir une vue tri-dimensionnelle de l'espace de séquence et de distinguer ainsi différents groupes de récepteurs. Notre équipe a mis en évidence quatre groupes de récepteurs dont un groupe central et 3 groupes périphériques, suggérant un mécanisme d'évolution radial :

Groupe central 0 : regroupant les récepteurs de peptides, les opsines et les récepteurs de la mélatonine

Groupe périphérique 1 : regroupant les récepteurs somatostatine/opiacés, chimiotactiques et purinergiques

Groupe périphérique 2 : regroupant les récepteurs d'amines et de l'adénosine

Groupe périphérique 3 : regroupant les récepteurs de la mélanocortine, les récepteurs des prostaglandines, les récepteurs Mas et les récepteurs des hormones glycoprotéiques [28].

Cette classification met en évidence le rôle des mutations des résidus proline des hélices transmembranaires 2 et 5 dans l'évolution des RCPG de classe A. Elle diffère de celle de Fredriksson *et al.* par la position des récepteurs purinergiques. Notre équipe a confirmé indépendamment le lien entre les récepteurs somatostatine/opiacés, les récepteurs chimiotactiques et les récepteurs purinergiques [29].

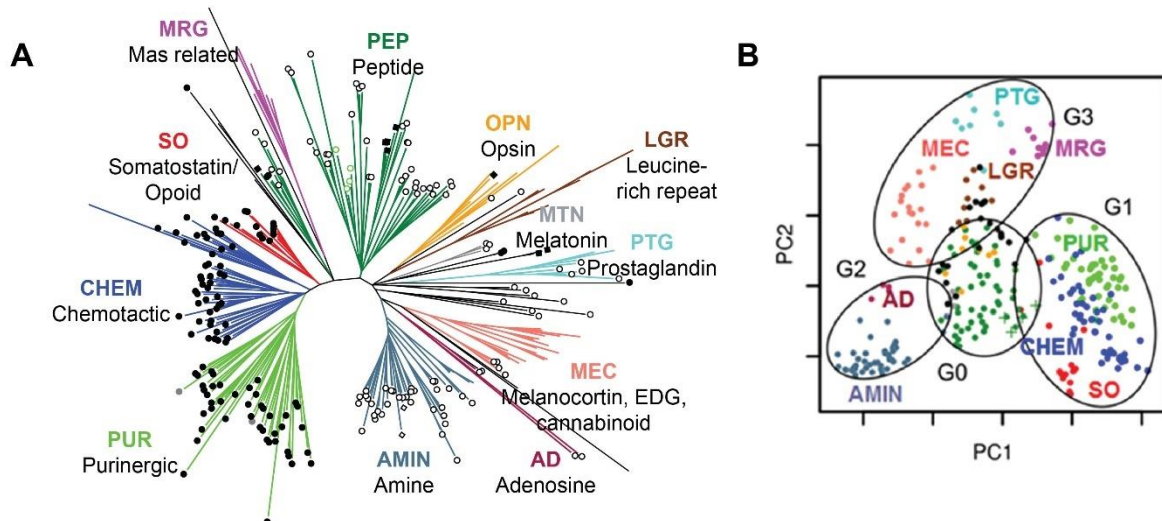


Figure 4 : Evolution des RCPG humains de classe A non olfactifs.

(A) Arbre évolutif obtenu par la méthode neighbour-joining; (B) Représentation par multi-dimensional scaling. Le code couleur indique les différentes sous-familles [28].

1.3.2. Structures cristallographiques des GPCR de classe A

Grâce à la cristallographie, depuis 2007, il y a eu une augmentation exponentielle du nombre de RCPG complexés à différents ligands (agonistes, antagonistes, agonistes inverses) dont la structure a été résolue (Figure 5a). Ces structures ont apporté de nombreuses informations sur le mécanisme moléculaire d'activation de ces récepteurs [30].

Plus récemment, le développement de la technologie de cryo-microscopie électronique (cryo-ME) a permis de déterminer certaines structures de RCPG couplés aux protéines G hétérotrimériques (Figure 5b), par exemple le récepteur 1 de la neurotensine (PDB ID : 6OS9). Ces progrès ont permis de mieux comprendre les mécanismes moléculaires de régulation de ces récepteurs.

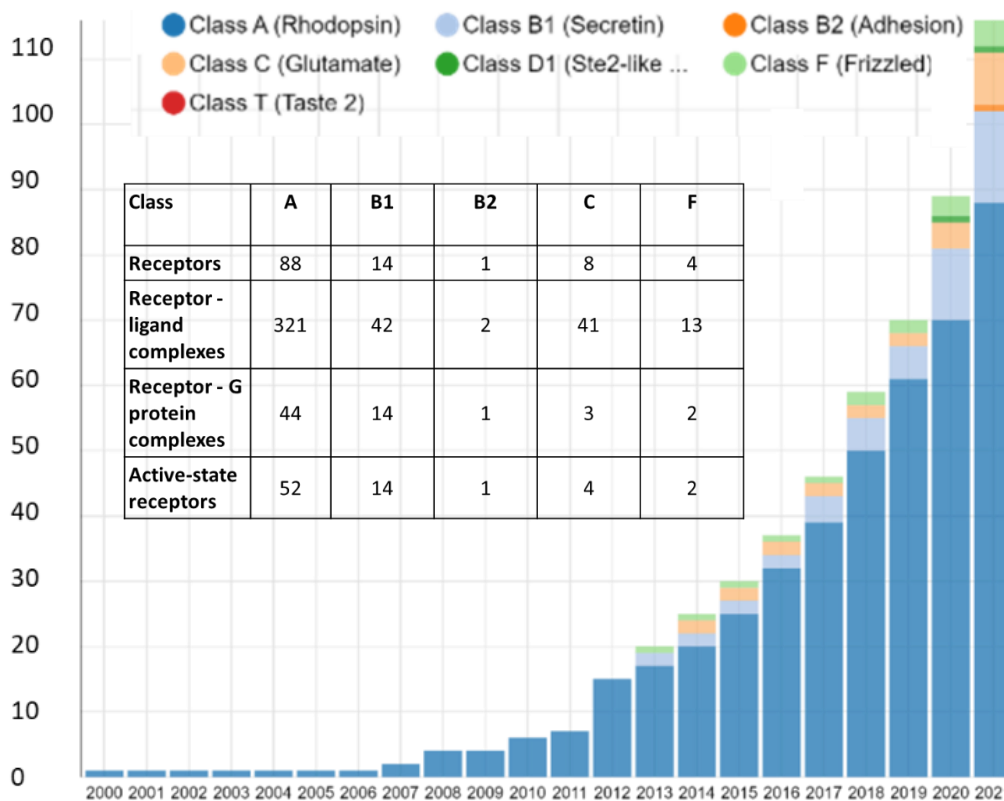


Figure 5 : Evolution du nombre de structures cristallographiques de RCPG.
La figure est issue de [31]. Le Tableau indique le nombre de structures en 2022.

Dans un recensement des structures cristallographiques des RCPG de classe A effectué en 2020, notre équipe a compté 57 RCPG dont 39 récepteurs dans la forme inactive, 2 récepteurs dans la forme active et 16 récepteurs dans différents états d'activation. La Figure ci-dessous (Figure 6) montre les différentes sous-familles des RCPG de classe A et les structures résolues dans chaque sous famille [30].

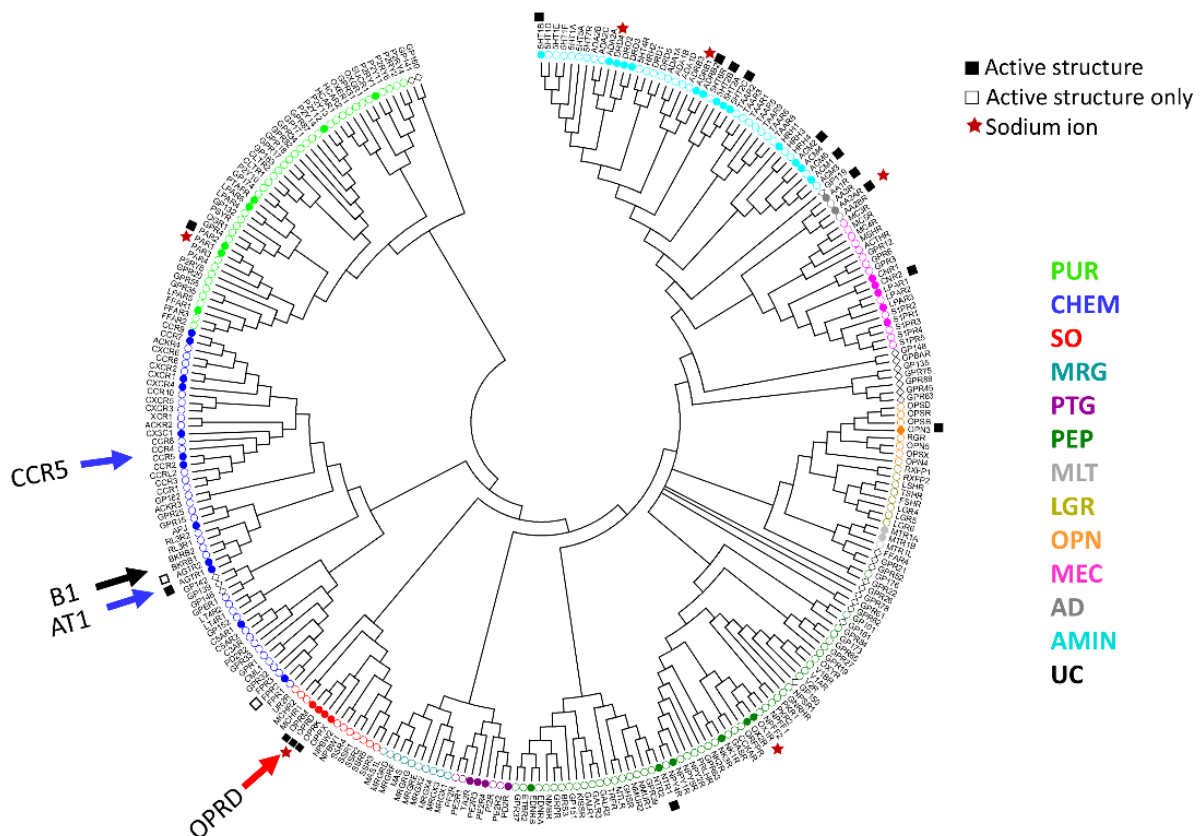


Figure 6 : Arbre évolutif des RCPG humains de classe A indiquant les structures résolues.

La couleur des cercles correspond aux sous-familles indiquées à gauche. Les cercles pleins indiquent au moins une structure résolue. Les carrés indiquent les structures actives [30].

1.3.3. La nomenclature Ballesteros-Weinstein

Afin de faciliter la comparaison des structures des récepteur de classe A, la méthode de numérotation de Ballesteros-Weinstein [32] est amplement utilisée. Cette méthode est basée sur la présence de résidus hautement conservés dans chacune des sept hélices transmembranaires (TM).

La numérotation se compose de deux nombres. Le premier désigne l'hélice n (1-7), le second la position du résidu par rapport au résidu le plus conservé de l'hélice n , défini comme le numéro $n.50$. Ci-dessous sont listés les résidus les plus conservés des récepteurs de classe A qui servent d'ancrage et leur nomination respective selon la méthode de Ballesteros-Weinstein [33]:

TM1 : Une asparagine (N1.50)

TM2 : Un acide aspartique (D2.50)

TM3 : L'arginine du motif E/DRY (R3.50)

TM4 : Un tryptophane (W4.50)

TM5 : Une proline (P5.50)

TM6 : La proline du motif CWxP (P6.50)

TM7 : La proline du motif NPxxY (P7.50)

Les autres résidus dans l'hélice seront numérotés selon leur position relative par rapport à la position $n.50$. Par exemple, le résidu 2.43 correspond à un résidu situé dans l'hélice 2 et positionné 7 résidus en amont du résidu le plus conservé dans cette hélice, D2.50.

1.4. Mécanisme d'activation des récepteurs couplés aux protéines G

L'explosion structurale des RCPG de classe A a permis de dévoiler leur similarité structurale très importante au niveau des hélices transmembranaires malgré de faibles identités de séquences (autour de 30% dans les différentes sous-familles et en-dessous de 20% entre les sous-familles). Ceci explique à la fois la conservation d'un mécanisme général et la diversité des ligands et des effecteurs.

1.4.1. Structures actives et inactives

La comparaison des structures actives et inactives des RCPG de classe A montre les changements majeurs de conformation du récepteur lors de l'activation qui se produisent du côté intracellulaire. A l'état actif, les hélices transmembranaires TM5 à TM7 se réarrangent pour permettre l'insertion de la protéine G (Figure 7). TM6 fait un mouvement vers l'extérieur ou « avant » et s'éloigne de près de 14 Å de TM3, alors que TM7 fait un mouvement vers l'intérieur ou « arrière » [34]. Dans certains cas, on observe aussi un redressement de TM5 et la formation d'une hélice α dans ICL2.

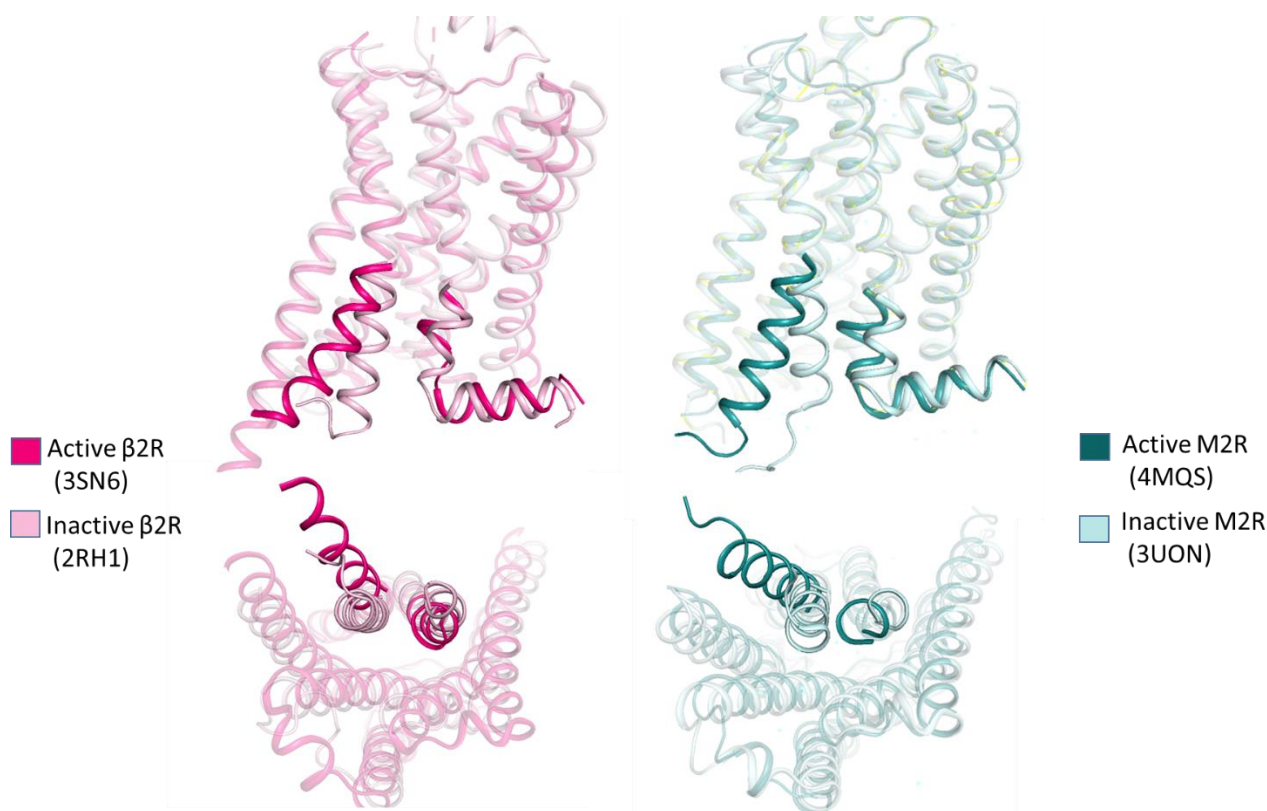


Figure 7 : Changements conformationnels des RCPG de classe A lors de l'activation.

Deux RCPG de classe A capturés par cristallographie dans leurs conformations inactive et active révèlent des changements de conformation communs lors de l'activation. La réorganisation de TM6 et TM7 est mise en évidence par une couleur plus foncée dans la conformation active. M2 est le récepteur muscarinique 2 de l'acétylcholine et β 2R est le récepteur beta 2 adrénergique.

1.4.2. Les motifs d'activation des RCPGs

Les GPCR de classe A partagent un mécanisme d'activation commun liant le site de fixation du ligand au site de couplage avec les protéines G. En analysant 234 structures de 45 sous-familles de RCPG de classe A, Zhou *et al.* ont détecté 34 paires de résidus impliqués avec un réarrangement conservé chez les récepteurs de classe A pendant l'activation [26]. Ces paires de résidus sont impliquées dans des motifs clés très conservés tels que CWxP, DRY, NPxxY, PIF et le site de fixation de l'ion Na⁺, reliant ainsi directement le site de fixation du ligand à la région de couplage avec les protéines G [26] (Figure 8). Ces motifs ont un rôle important dans le mécanisme d'activation des RCPG :

Le motif CWxP : appelé aussi « transmission switch/Trp Rotamer Toggler switch », consiste en la relocation de W6.48 vers P5.50 à l'état actif des récepteurs.

Le motif PIF : forme une interface entre TM5, TM6 et TM3, près du site de fixation du ligand dans le récepteur. Pendant l'activation, on observe un déplacement vers l'intérieur du résidu P5.50 couplé avec une rotamérisation du résidu I3.40, un déplacement important de la chaîne latérale de F6.44, et la rotation vers l'extérieur de l'hélice 6 du côté cytoplasmique.

Le site de fixation de l'ion sodium : certaines structures cristallographiques montrent la présence d'un ion sodium dans la partie transmembranaire du récepteur. L'ion sodium a un effet de régulateur allostérique négatif stabilisant l'état inactif du récepteur. Le site de fixation du sodium est bien conservé chez les RCPG de classe A et certaines mutations des résidus de ce site peuvent déclencher des activations constitutives. Le site se réorganise lors de l'activation et ne permet plus la fixation du sodium.

Le motif NPxxY : le motif NPxxY a été initialement proposé pour expliquer la présence de l'interaction entre Y7.53 dans TM7 et F7.60 dans l'hélice H8 dans la structure inactive de la rhodopsine. Une telle interaction peut stabiliser l'état inactif du récepteur [35]. Lors de l'activation Y7.53 forme de nouveaux contacts avec les résidus 3.43, 3.46 et R3.50.

Le DRY motif : appelé aussi verrouillage ionique ou « ionic lock », décrit les liaisons ioniques fortes entre R3.50 et D/E3.49 de TM3 et dans certains cas avec E6.30 de TM6, stabilisant l'état inactif des RCPG et empêchant l'interaction de R3.50 avec l'extrémité C-terminale de la sous-unité α de la protéine G. Dans certains RCPG à l'état actif, R3.50 forme des liaisons hydrogènes avec Y5.58 [35].

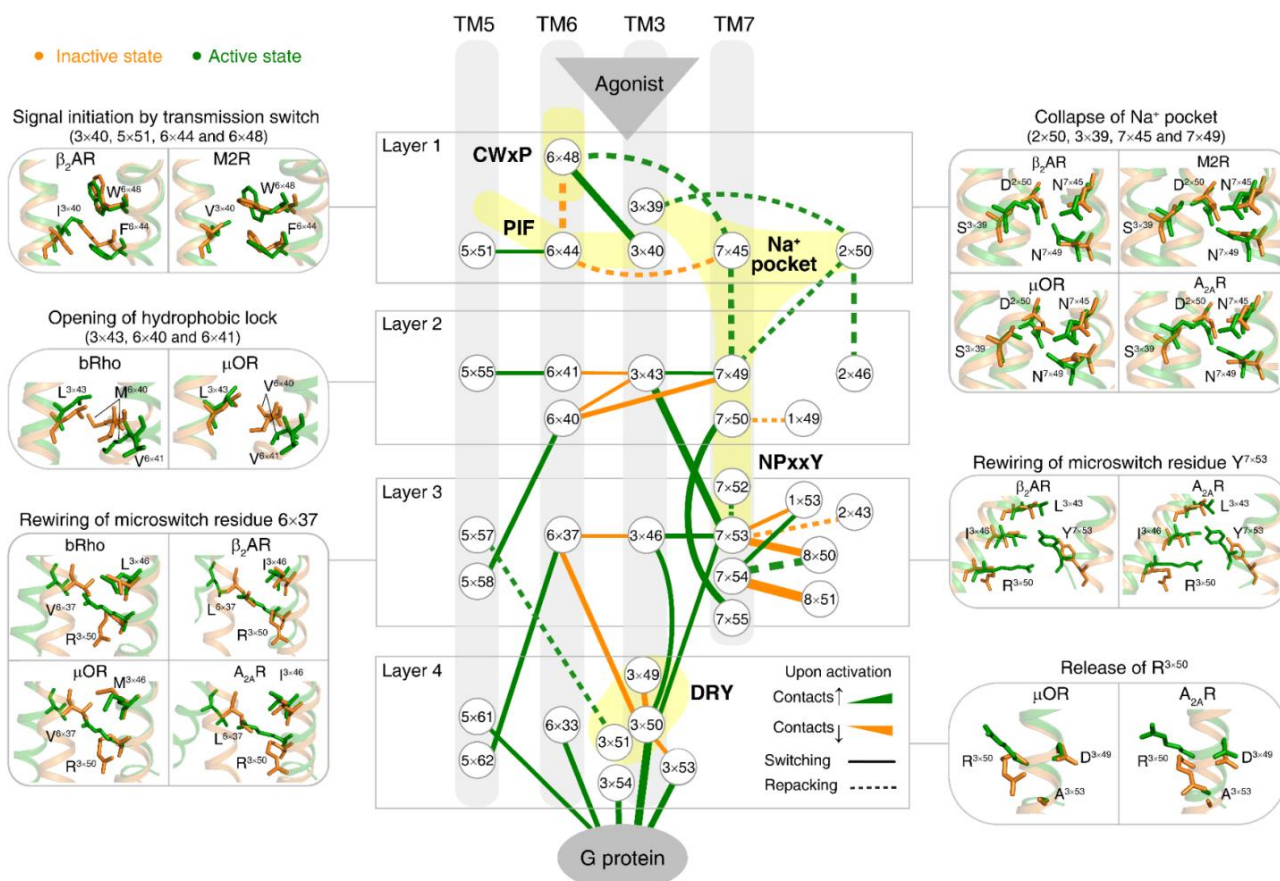


Figure 8 : Les cinq motifs structuraux conservés des RCPG de classe A.

Ces motifs relient le site de fixation du ligand au site de couplage avec les protéines G. Ils sont organisés en couches avec les motifs CWxP, PIF, site sodium, NPxxY, DRY positionnés de haut en bas [26].

1.4.3. Structures actives non canoniques

La plupart des structures cristallographiques actives présentent les mêmes caractéristiques d'activation (mouvement en avant de TM6 et mouvement en arrière de TM7). Mais parfois, il existe des exceptions à la norme où une structure active présente des caractéristiques non canoniques. C'est le cas pour le récepteur humain 1 de la neurotensine (hNTSR1). Kato *et al.* ont publié récemment la structure cristallographique active du NTSR1 couplé à la protéine G_i sous deux formes, canonique C et non canonique NC [36]. Les comparaisons structurales révèlent que hNTSR1 à l'état C adopte la conformation canonique d'un RCPG activé, mais que hNTSR1 à l'état NC présente certaines caractéristiques des conformations actives et inactives; les motifs TM5, TM6 et DRY (déplacement de R3.50 de 2.1Å) adoptent des conformations actives tandis que TM7 (y compris le motif NPXXY) est dans une conformation inactive (Figure 9d,e). Notons cependant que la structure est à une résolution de 3Å. Ces caractéristiques non canoniques ont été aussi observées pendant des simulations de dynamique moléculaire du récepteur β_2 adrenergique (β_2AR) comme un état intermédiaire d'activation [37].

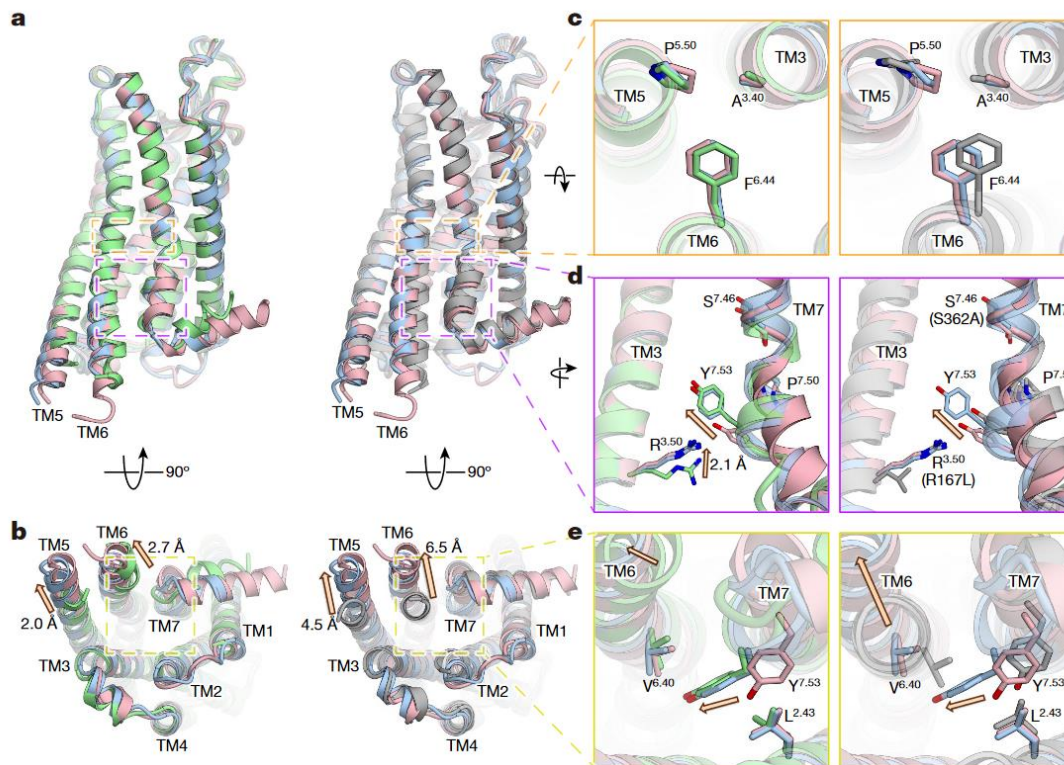


Figure 9 : Comparaison structurale de hNTSR1 dans les états canonique C et non canonique NC.

(a-e) Structures superposées de hNTSR1, à l'état C (bleu) et à l'état NC (rouge), et de rNTSR1, à l'état actif (vert) et à l'état inactif (gris). Vue latérale (a) et intracellulaire (b) de la structure globale, et vue agrandie des motifs PIF ou PAF (c), DRY (d) et NPXXY (e). Les hNTSR1 à l'état C et NC sont superposés au rNTSR1 à l'état actif (a-e, gauche) et au rNTSR1 à l'état inactif (a-e, droite). Les flèches marquent les différences entre les structures superposées [36].

1.5. Rôle de l'eau dans l'activation des récepteurs couplés aux protéines G

Les structures cristallographiques montrent la présence des molécules d'eau dans la partie transmembranaire des récepteurs [38]. Ces molécules contribuent à 2 réseaux polaires conservés chez les RCPG. Le premier réseau est présent dans les états actif et inactif. Le deuxième réseau se forme lors de l'activation. Ces réseaux impliquant des molécules d'eau jouent un rôle important dans l'activation et la stabilisation des RCPG. En effet, ces molécules d'eau stables permettent des interactions entre résidus protéiques polaires stabilisant ainsi certaines conformations [38]. Le nombre des molécules d'eau présenté dans les structures cristallographiques en particulier dans les structures actives, peut dépendre de la résolution cristallographique de la structure d'où l'importance des simulations dynamiques pour estimer et valider le nombre et le réseau des molécules d'eau plus correctement [39,40].

Les simulations de dynamique moléculaire montrent que, lors de l'activation de certains RCPG, un canal continu d'eau se forme dans la partie transmembranaire [41,42]. A l'état inactif, ce canal est bloqué par une couche d'acides aminés hydrophobes qui est située près du motif NPxxY. Cette couche forme un portail qui s'ouvre lors de l'activation, du fait de la rotamérisation de la tyrosine Y7.53 (Figure 10) [42].

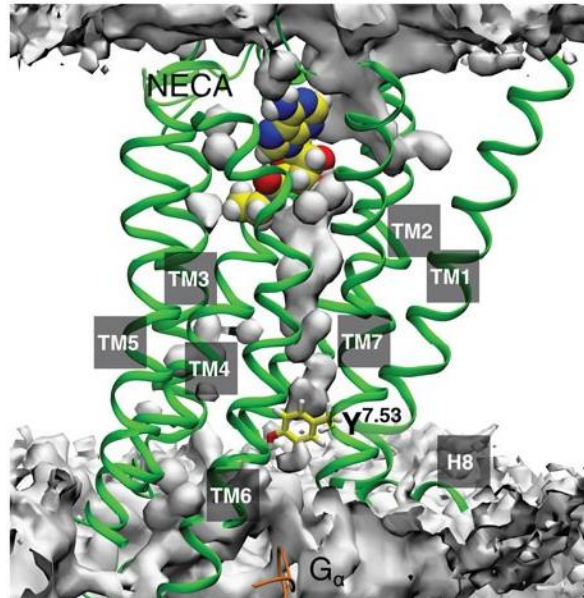


Figure 10 : Canal d'eau du récepteur A2AR actif couplé à la protéine G α et au ligand NECA.
Le canal est visualisé par la densité moyenne de l'eau. La tyrosine Y7.53 est représentée en bâtonnets jaunes [42].

1.6. Rôle de l'ion sodium chez les récepteurs couplés aux protéines G

L'un des motifs les plus conservés chez les RCPG de classe A est le site de fixation de l'ion sodium. La présence d'un ion sodium a été mise en évidence dans les structures cristallographiques à l'état inactif de plusieurs récepteurs appartenant à différentes branches des RCPG de classe A (le récepteur de l'adénosine A2A, le récepteur PAR1, le récepteur adrénergique β 1AR, le récepteur opiacé δ -OR, et le récepteur dopaminergique D4). Ces récepteurs partagent les mêmes positions de fixation de l'ion sodium [43]. Dans tous les cas étudiés, l'ion sodium interagit avec le résidu D2.50 chargé négativement, en présence d'un maximum de 10 molécules d'eau participant à la coordination du sodium avec les résidus de son site. Le site de fixation du sodium inclut des résidus des hélices TM1, TM2, TM3, TM6, TM7. Les résidus formant le site sodium sont très conservés (Figure 11a). Parmi eux, l'acide aspartique D2.50 est toujours impliqué dans l'interaction avec l'ion sodium. Les autres résidus interagissant directement avec l'ion changent selon la nature du récepteur (par exemple 7.45, 7.46, 3.35, 3.39, 7.49, etc.) [36].

L'ion sodium a un rôle essentiel dans de nombreuses fonctions physiologiques. Il peut avoir des rôles directs et indirects dans le fonctionnement des RCPG. Le premier rôle du sodium sur les RCPG qui a été mis en évidence est un effet allostérique négatif [44]. Cet effet a d'abord été mis en évidence chez les récepteurs opiacés [45]. L'étude montre que le sodium module négativement la fixation de l'agoniste sans affecter l'affinité des antagonistes [45]. Par la suite, cet effet a été observé pour plus de 15 RCPG différents, y compris les récepteurs α -adrénergiques, dopaminergiques, sérotoninergiques, neurotensines et autres [44].

En plus de l'effet allostérique négatif de l'ion sodium, la mutation des résidus du site de fixation du sodium, en particulier D2.50, peut inhiber la signalisation par les protéines G en présence d'un agoniste (cas du récepteur AT1 de rat [46] ou du récepteur A2A de l'adénosine [47]) ou induire une signalisation biaisée (cas du récepteur opiacé δ [48,49]) (Figure 11a). L'analyse des structures cristallographiques et les simulations dynamiques des RCPG montrent qu'à l'état actif, le site de fixation du sodium est déstabilisé par les mouvements drastiques des hélices. La réorganisation des hélices est incompatible avec la fixation de l'ion sodium dans son site. Cela

suggère que la relocation du sodium est nécessaire pour l'activation du récepteur [44]. Parmi les facteurs qui peuvent influencer le rôle de l'ion sodium sur l'activation des RCPG, se trouvent le pH et le potentiel de la membrane. En effet, la protonation de D2.50 à un faible pH, dans un environnement acide, peut faciliter la sortie de l'ion sodium et activer certains RCPG de classe A [50,51]. Le sodium étant chargé positivement, il devrait aussi être sensible au potentiel électrostatique membranaire, ce qui permettrait à ce potentiel d'agir via le sodium sur l'activité des RCPG [50].

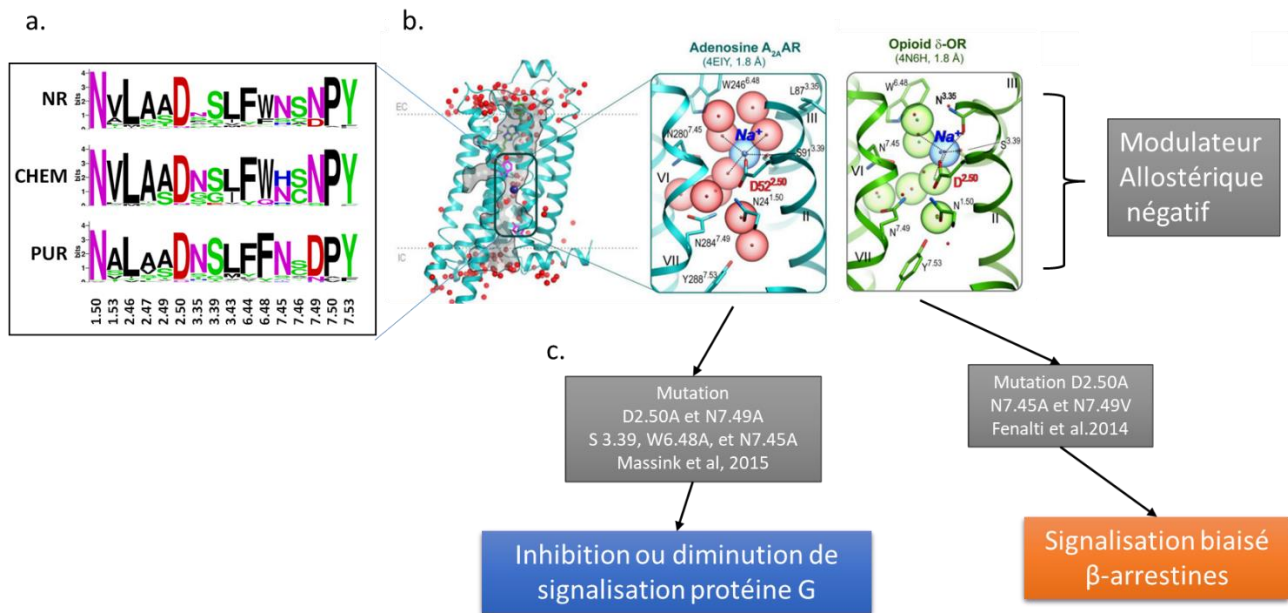


Figure 11 : Rôle de l'ion sodium chez les RCPG de classe A.

(a) Logo plots du site sodium pour les RCPG de classe A non-olfactifs (NR), les récepteurs chimiotactiques (CHEM) et les récepteurs purinergiques (PUR) [52]; (b) Réseau Na⁺/eau du site sodium des RCPG de classe A; (c) Rôle du sodium et effet des mutations du site de fixation du sodium. Figure adaptée de [44].

1.7. Effet de la composition lipidique dans l'activation des récepteurs couplés aux protéines G

La membrane cellulaire est formée d'une bicouche lipidique dans laquelle des protéines transmembranaires peuvent s'insérer. Les protéines membranaires se sont adaptées pour tenir compte des propriétés physico-chimiques de la bicouche lipidique. Par conséquent, tout changement dans la composition de la bicouche lipidique peut affecter la structure et la fonction des protéines transmembranaires et en particulier les RCPG [53]. Ces effets peuvent être attribués soit à une modification des propriétés de la bicouche (épaisseur, courbure et tension superficielle), soit aux interactions de lipides spécifiques avec les protéines [53]. Les lipides liés peuvent influencer également les interactions ternaires entre les RCPG et les protéines G [54]. Parmi les lipides les plus documentés pour avoir un effet sur les RCPG, se trouvent le cholestérol, les omega-3 et les phospholipides chargés négativement.

1.7.1. Effet du cholestérol sur le fonctionnement des RCPG

Le cholestérol peut affecter les RCPG de deux manières. Il peut soit se lier directement au récepteur et ainsi moduler de manière allostérique l'affinité des ligands, l'efficacité des agonistes et l'activité spontanée du

récepteur. Il peut également affecter indirectement les RCPG en modifiant la fluidité et l'organisation de la membrane, ce qui a une incidence sur la signalisation des RCPG [55].

Le cholestérol a été trouvé co-cristallisé avec de nombreux RCPG de classe A, ce qui suggère une possible liaison spécifique. En 2021, Jakubik *et al.* ont comptabilisé 44 structures cristallographiques de récepteurs complexés à une molécule de cholestérol, appartenant à 18 RCPG de classe A différents (Figure 12) [55]. En fonction de l'emplacement du site de fixation du cholestérol, ce dernier peut affecter la fixation du ligand, la fonction du récepteur ou l'oligomérisation du récepteur [55]. Par exemple le groupement hydroxyle du cholestérol peut avoir une interaction directe avec le motif DRY stabilisant ainsi l'état actif du récepteur [14]. La présence du cholestérol peut aussi moduler la dimérisation de certains récepteurs, soit directement quand le cholestérol est une partie intégrante de l'interface protomère-protomère (par exemple dans le cas du récepteur α_2c -adrenergique 6KUW), soit indirectement quand le cholestérol affecte l'organisation de la membrane plasmique qui, à son tour, affecte le processus d'oligomérisation en altérant la diffusion des récepteurs [55].

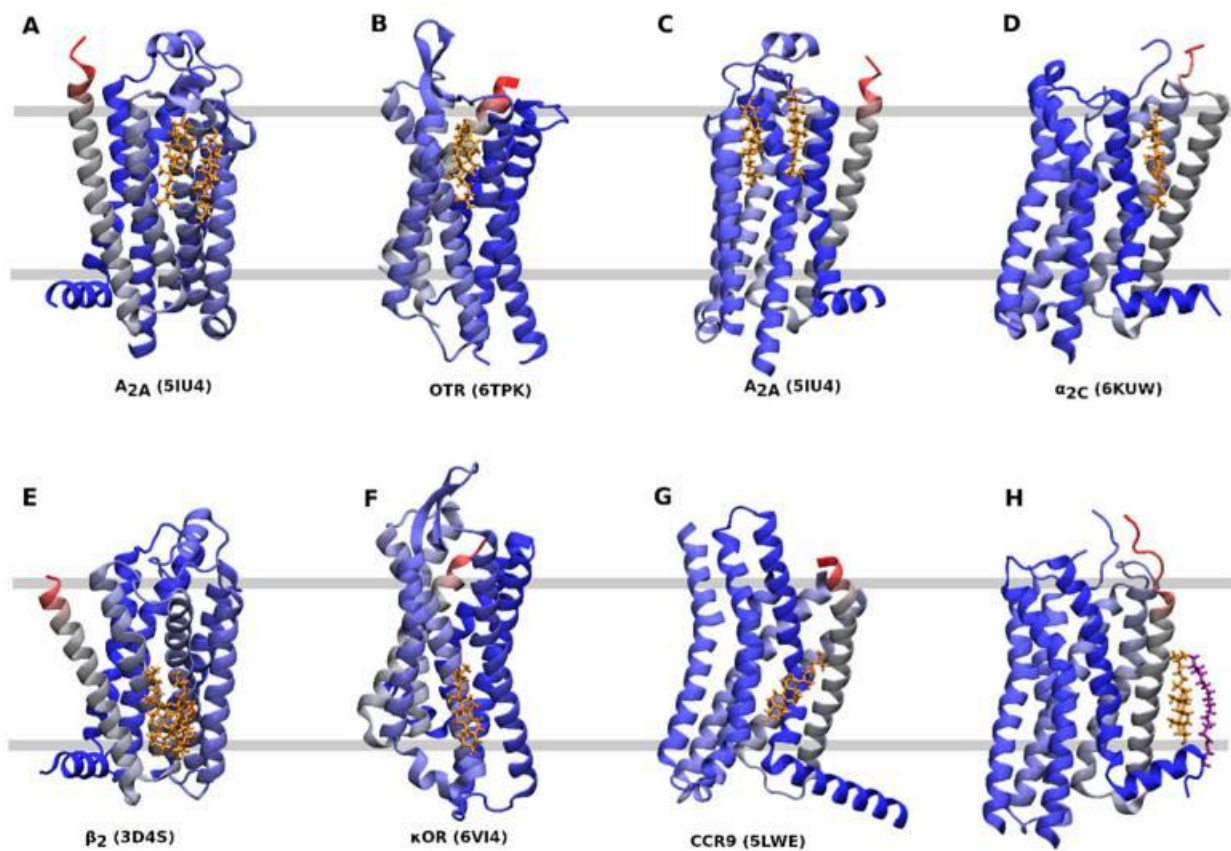


Figure 12 : Sites de fixation du cholestérol dans les RCPG de classe A.
Le cholestérol est visualisé en bâtonnets jaunes. Figure issue de [55].

1.7.2. Effet des lipides anioniques : exemple du phosphatidylglycérol (PG)

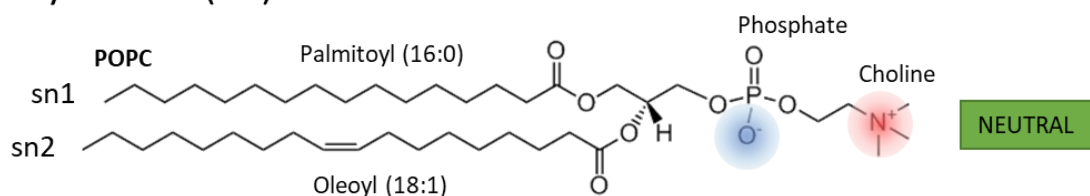
Des études antérieures montrent que les RCPG peuvent être affectés par les lipides anioniques tels que les DOPG (1,2-Dioleoyl-*sn*-Glycero-3-[Phospho-*rac*-(1-glycérol)]) ou les PIP (phosphatidylinositol phosphate) [56]. Ici nous détaillerons les données sur les DOPG que nous avons utilisés au cours de la thèse. Les DOPG appartiennent à la famille des phosphatidyl-glycérols. Ces lipides diffèrent des POPC par la longueur d'une des chaînes lipidiques et par la charge électrostatique de la tête des phospholipides. En effet, dans la famille des phosphatidylcholines, tels que le POPC (1-palmitoyl-2-oleoyl-glycero-3-phosphocholine), la charge positive du

groupement choline neutralise la charge négative du groupement phosphate, alors que, dans le cas des DOPG, les lipides sont chargés négativement par le groupement phosphate (Figure 13).

L'action de ces lipides sur les RCPG a été étudiée expérimentalement. Par exemple, dans le cas des récepteurs β_2 adrénergiques, les DOPG présentent 3.6% de la composition membranaire mais 16% des lipides liés au récepteur, facilitant ainsi son activation [53]. Dans le cas du récepteur 1 de la neurotensine (NTSR1), les DOPG augmentent le couplage de la protéine G au récepteur [15].

Des études bioinformatiques de simulations dynamiques révèlent que, dans le cas des récepteurs β_2 adrénergiques, les DOPG stabilisent la conformation active plus longtemps que les phosphatidcholines [57] et que, dans le cas des récepteurs de l'adénosine dans une bicouche DOPG, des préactivations du récepteur en absence de son ligand et des activations reproductibles en présence du ligand ont été observées [16]. Les lipides anioniques DOPG facilitent ou maintiennent donc l'activation de certains RCPG. Le but de notre étude est de chercher si les lipides anioniques, en particulier les DOPG, peuvent intervenir dans le mécanisme d'activation des récepteurs mécanosensibles à l'étirement de la membrane.

Phosphatidylcholine (PC)



Phosphatidylglycerol (PG)

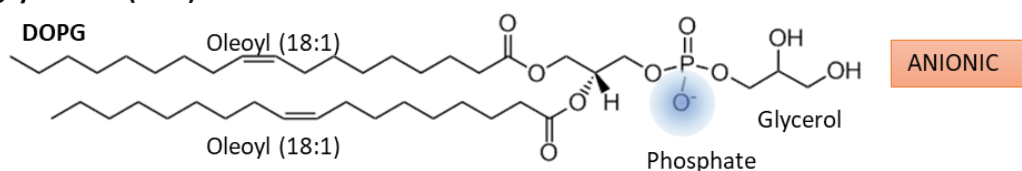


Figure 13 : Présentation schématique des phosphatidycholines et les phosphatidylglycérols.

1.7.3. Lipides à action fluidifiante sur la membrane : les oméga 3

Les acides gras oméga-3 polyinsaturés, tels que le DHA (acide docosahexaénoïque) et l'EPA (acide eicosapentaénoïque) ont généralement des effets bénéfiques sur la santé humaine. Ils sont considérés comme des composants membranaires majeurs des cellules du cerveau et ils sont fortement impliqués dans les fonctions cérébrales [58]. L'oméga-3 module des propriétés clés des membranes cellulaires, comme la fluidité, et affecte ainsi le comportement des protéines transmembranaires comme les récepteurs couplés aux protéines G. Guixà-González *et al.* ont montré que les niveaux de DHA dans la membrane module la cinétique de l'oligomérisation des RCPG de classe A, en particulier les oligomères des récepteurs A2A de l'adénosine et D2 de la dopamine [59]. En effet, les résultats montrent que des niveaux plus élevés de DHA accélèrent l'agrégation des protéines, grâce à un nombre de contacts protéine-protéine plus élevé, et soulignent le rôle de la cinétique dans la modulation de l'oligomérisation des récepteurs A2A et D2 (Figure 14) [59]. D'autres études montrent que le récepteur GPR120 (GPCR de classe F) est un senseur fonctionnel d'oméga-3 qui exerce une puissante action de sensibilisation à l'insuline et des effets antidiabétiques *in vivo* en réprimant l'inflammation tissulaire induite par les macrophages [60].

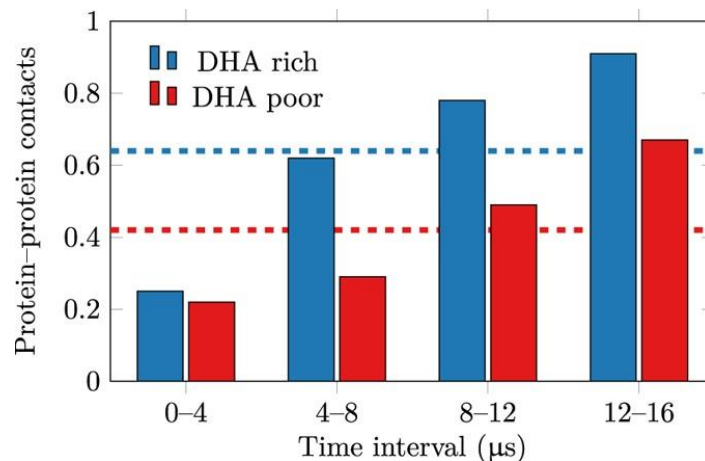


Figure 14 : Influence de la composition membranaire sur le nombre de contacts entre récepteurs.
L'évolution du nombre moyen de contacts protéine-protéine par protomère a été mesurée sur un ensemble des simulations des récepteurs A2A et D2 [59].

1.8. Activation des récepteurs couplés aux protéines G par le stress mécanique

La capacité de sentir et de s'adapter à l'environnement est essentielle à la survie de tous les organismes. Au cœur de ces mécanismes de survie, se trouve la mécanotransduction. La mécanotransduction désigne le phénomène biologique par lequel les contraintes mécaniques appliquées aux cellules sont traduites en signaux biochimiques qui provoquent des réponses adaptatives [61]. La mécanotransduction est impliquée dans de nombreuses réponses biologiques essentielles, de la vasodilatation à l'audition, l'équilibre, la sensation de la position des articulations, la contraction musculaire et le toucher [61].

Plusieurs avancées récentes dans ce domaine ont suscité un grand intérêt, qui s'est traduit par l'attribution du prix Nobel de médecine 2021 à David Julius et Ardem Patapoutian pour leur découverte de transducteurs thermiques et mécaniques [62].

1.8.1. Stress mécanique subi par les cellules

Une cellule peut subir trois types de conditions externes entraînant un stress mécanique sur la membrane plasmique (Figure 15) : (1) les forces de cisaillement [13] qui résultent du contact d'une cellule avec une surface mobile (par exemple à l'interface cellule endothéliale - sang); (2) la pression hydrostatique lorsqu'une cellule est placée dans un environnement basse ou haute pression causant ainsi un effet d'étirement ou de contraction de la membrane ; et (3) la pression osmotique lorsqu'une cellule est placée dans un environnement aqueux hypotonique ou hypertonique. Ces conditions externes conduisent à une modification de la tension de surface de la membrane, conduisant à des tensions superficielles positives ou négatives [63].

Les cellules endothéliales subissent un stress mécanique permanent, résultant de la pression artérielle et du flux sanguin. Pour répondre à ces signaux mécaniques, les cellules endothéliales utilisent plusieurs voies de signalisation activées par des protéines membranaires mécanosensibles. Parmi les protéines mécanosensibles, on note les RCPG, les canaux ioniques, les intégrines et quelques enzymes membranaires tel que les tyrosines kinases [12] (Figure 15).

Stress mécanique

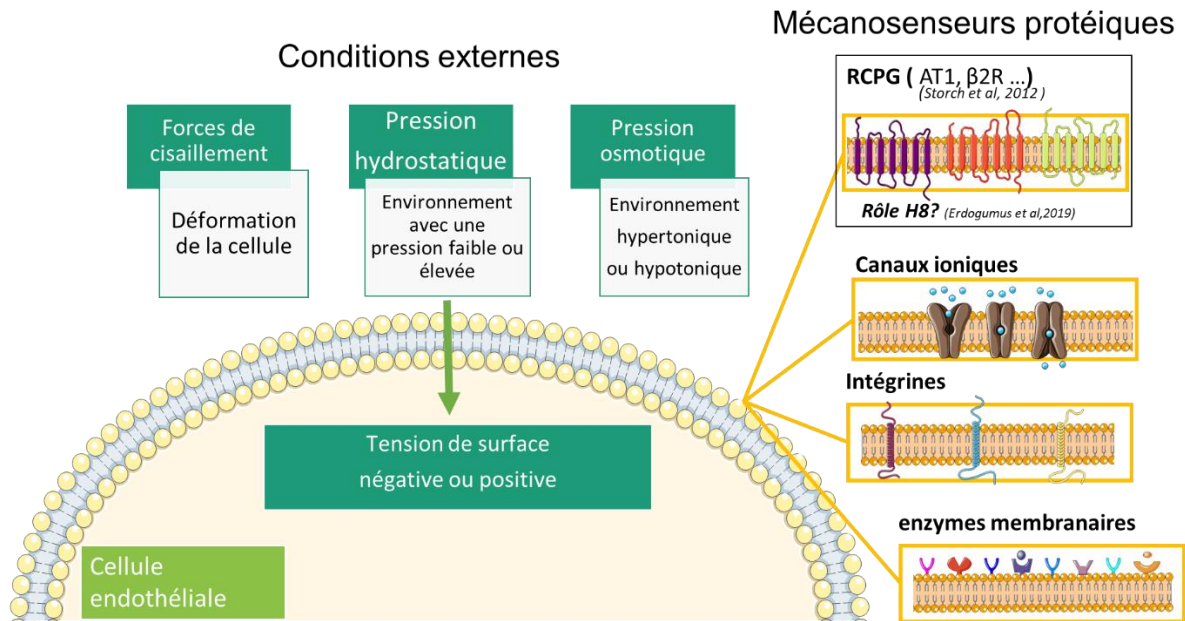


Figure 15 : Stress mécanique subi par les cellules endothéliales et mécanosenseurs potentiels.

1.8.2. Réponse des RCPGs au stress mécanique

Les premières études sur l'activation des RCPG par un stress mécanique semblaient montrer qu'ils étaient activés par la libération paracrine ou autocrine de peptides, de lipides ou de protéines [63,64]. Plus récemment, il a été démontré que des RCPG peuvent avoir un rôle de mécanosenseur [12,65] et être activés directement par l'effet du stress mécanique, en absence de ligands.

Une liste de RCPG mécanosenseurs est présentée dans le tableau 1 :

Tableau 1 : RCPG mécanosensibles.¹

RCPG	Stress mécanique	Fonction possible	Références
AT1R	Forces de cisaillement, étirement direct de la membrane, environnement hypotonique	Mécanosenseur médiant la vasoconstriction myogénique et l'hypertrophie cardiaque	[11,66,67]
B2R	Forces de cisaillement, environnement hypotonique, agent fluidifiant de la membrane	Senseur des forces de cisaillement dans les cellules endothéliales, augmentation de l'activité de B2R dans les cellules endothéliales	[66]
beta2-adrénergique	Forces de cisaillement + bactéries pathogènes	Senseur des forces de traction induites par la bactérie dans le flux sanguin et appliquées sur des chaînes de glycanes de l'extrémité N terminale du récepteur	[64]
D5R	Forces de cisaillement	Senseur des forces de cisaillement dans les cellules endothéliales	[69]
ET1AR	Environnement hypotonique	?	[70]
FPR1	Forces de cisaillement	Rétraction des pseudopodes dans les neutrophiles	[71]
H1R	Environnement hypotonique/étirement direct de la membrane	Senseur des forces de cisaillement dans les cellules endothéliales	[72]

M5R	Environnement hypotonique/étirement direct de la membrane	?	[70]
PTH1R	Forces de cisaillement	Croissance des cellules osseuses	[73]
V1AR	Environnement hypotonique	?	[70]
GPR68	Forces de cisaillement	Senseur des forces de cisaillement dans les cellules endothéliales	[73,74]
APJ/apelin	Forces de cisaillement	Autorégulation de l'expression d'APJ induite par la contrainte de cisaillement	[76]
SIP1/sphingosine phosphate	Rigidité de la matrice extracellulaire	Rôle mécanotransducteur pour la réparation/régénération tissulaire intrinsèque	[77]

¹ Tableau adapté de [12] et [64]

1.8.3. Mécanismes moléculaires d'activation des RCPG par le stress mécanique

Même si le rôle mécanosenseur des RCPG est bien établi, les mécanismes moléculaires induisant l'activation mécanique des RCPG restent inconnus. Au cours de ces dernières années, quelques études ont cherché à dévoiler le mécanisme d'activation possible des RCPG par les forces mécaniques.

a) Rôle de l'hélice 8 dans la mécanosensation des RCPG

En 2019, Erdogumus *et al.* ont montré l'implication de l'hélice 8 C-terminale dans la mécanosensation du récepteur de l'histamine 1 (H1R) et suggéré la présence d'une conformation distincte d'activation par le stress mécanique [72]. En utilisant des outils pharmacologiques, ils ont mis en évidence la présence d'une hélice 8 intacte comme une condition préalable essentielle à la mécanosensation des RCPG. Les RCPG dépourvus d'une hélice 8 fonctionnelle en raison d'une altération de l'intégrité structurelle de l'hélice 8 ou d'une troncature C-terminale ne sont pas mécanosensibles (Figure 16) et, inversement, le transfert de l'hélice 8 d'un récepteur mécanosensible à un autre qui ne l'est pas confère la mécanosensibilité.

En particulier, ces auteurs ont vérifié que l'hélice 8 est également essentielle à la mécanosensibilité du récepteur AT1 par la perturbation de l'intégrité structurelle de l'hélice 8 due à des mutations d'acides aminés sélectionnés (F309P et F309P-F313P). En réalisant une imagerie calcique, ils ont observé que les augmentations calciques induites mécaniquement étaient significativement réduites chez les récepteurs AT1 mutés. Ils ont ainsi conclu que l'hélice H8 d'AT1 exprimée dans le système vasculaire est un motif structurel essentiel pour détecter les forces mécaniques et les convertir en réponses vasculaires, contribuant ainsi à l'autorégulation du tonus myogénique [72].

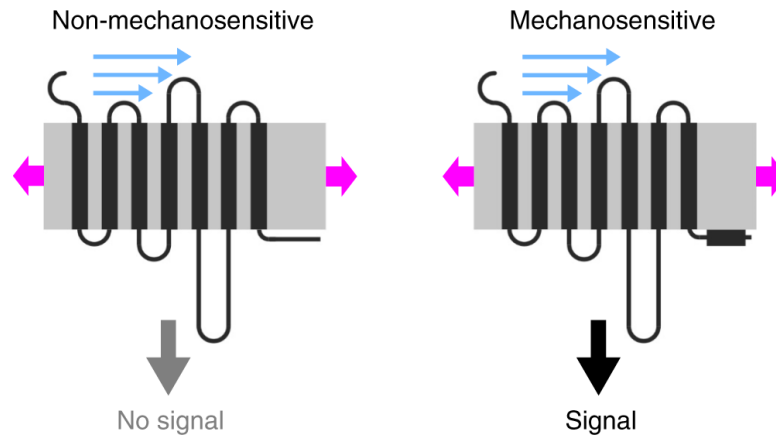


Figure 16 : Rôle de l'hélice 8 dans l'activation des RCPG par le stress mécanique.
Figure issue de [72].

b) Rôle des N-Glycanes dans la mécanosensation des RCPG.

En 2020, Marullo *et al.* ont rapporté un mécanisme d'activation mécanique du récepteur β 2-adrénergique, β 2AR, par un pathogène bactérien. Cette activation implique des forces de traction induites par la bactérie dans le flux sanguin et appliquées sur des chaînes de glycanes de l'extrémité N-terminale du récepteur [64] (Figure 17).

Plus généralement, la plupart (80 %) des RCPG expriment au moins une chaîne N-glycane dans leur extrémité N-terminale [64]. Ces chaînes glycaniques pourraient également être impliquées dans des interactions avec des lectines de surface cellulaire et activées allostériquement dans le contexte de la signalisation mécanique intercellulaire. Cependant, il est actuellement difficile d'évaluer combien de RCPG pourraient effectivement être activés par une traction appliquée sur leurs N-glycanes. Par exemple, les billes enduites de lectine n'ont pas réussi à activer AT1, ce qui indique que chaque récepteur doit être testé individuellement pour son aptitude à être activé directement par des forces mécaniques appliquées sur l'extrémité N-terminale du récepteur [64].

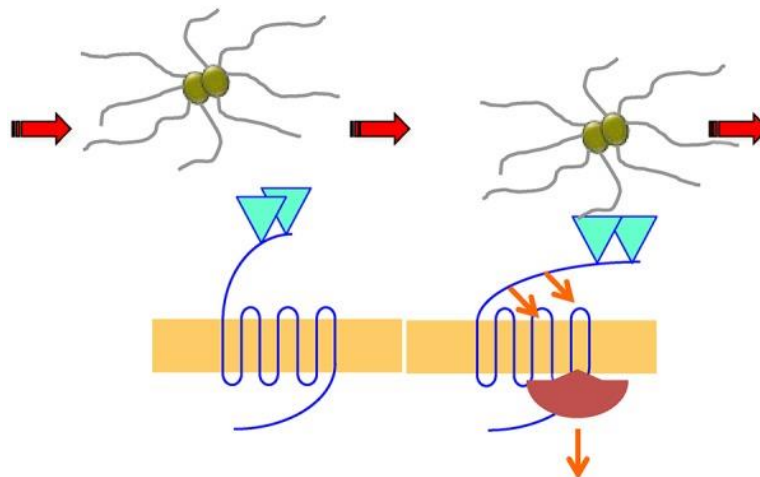


Figure 17 : Rôle des N-glycanes dans la mécanosensation des RCPG.

Certains RCPG peuvent être activés par les forces de traction agissant directement sur les N-glycanes de l'extrémité N-terminale du récepteur. Figure issue de [64]

2. Le système rénine-angiotensine

2.1. Généralités

Le système rénine-angiotensine (SRA) joue un rôle important dans le contrôle de la pression artérielle et dans l'homéostasie vasculaire. Il est impliqué dans la physiopathologie de l'hypertension et dans les altérations structurelles du système vasculaire [78]. Au cœur de ces mécanismes physiopathologiques se trouve l'angiotensineII.

L'angiotensine II est un octapeptide produit par l'hydrolyse de l'angiotensinogène hépatique en l'angiotensine I (AngI), par l'action de la rénine, puis par l'hydrolyse de celle-ci en angiotensine II (AngII), par l'action de l'enzyme de conversion de l'angiotensine ACE. Elle agit principalement par l'intermédiaire de deux RCPG de classe A : les récepteurs de l'AngII de type 1 (AT1) et de type 2 (AT2) (Figure 18). La liaison de l'AngII au récepteur AT1 induit une vasoconstriction qui entraîne une augmentation de la pression artérielle. L'activation d'AT1 contribue également de manière indépendante à la pathologie de maladies chroniques en favorisant la croissance et la prolifération vasculaires, ainsi que le dysfonctionnement endothélial. Ces conséquences négatives de l'angiotensine II sont partiellement contrecarrées par la stimulation du récepteur AT2, qui a des effets favorables sur la croissance des tissus et les processus de réparation [79].

L'angiotensine 1-7 (Ang1-7) est aussi un peptide actif dans le système rénine-angiotensine. Ce peptide est un agoniste du récepteur MAS. Il est produit par la conversion de l'angiotensine 1 en angiotensine 1-9 par l'action d'ACE2, puis la conversion de l'angiotensine 1-9 en angiotensine 1-7 par l'action d'ACE. Semblable à AT2, l'activation du récepteur MAS par l'angiotensine 1-7 induit des voies de signalisation généralement protectrices qui contrebalancent l'effet du récepteur AT1. L'axe ACE2-angiotensine-(1-7)-Mas est désormais sérieusement considéré comme une cible potentielle pour le développement de nouveaux médicaments cardiovasculaires [4,79]

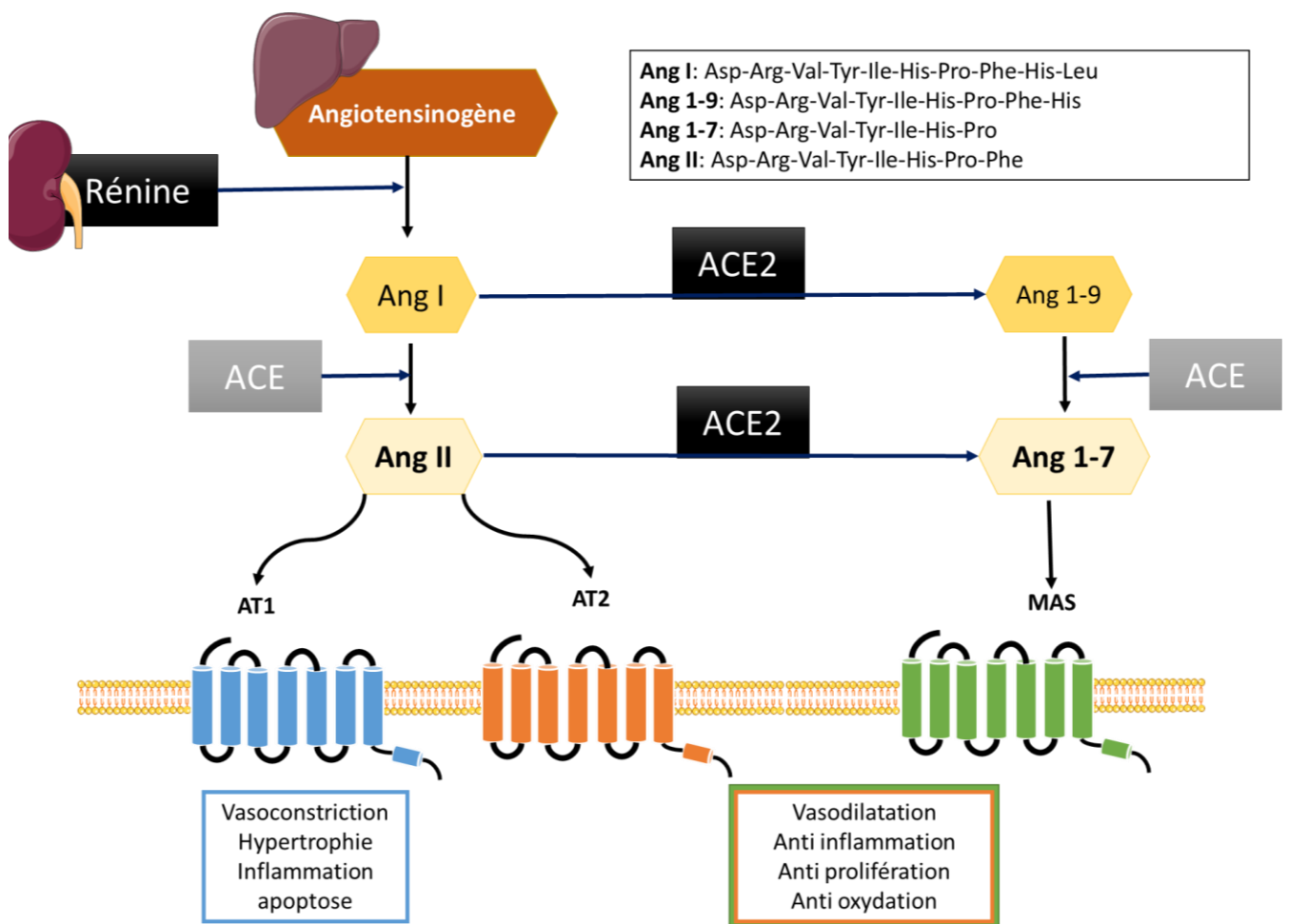


Figure 18 : Schéma simplifié du système rénine-angiotensine

L'angiotensinogène d'origine hépatique produit le décapeptide angiotensine I par l'action de la rénine produite par le rein. L'angiotensine I à son tour va être clivée en angiotensine II par l'action de l'enzyme de conversion de l'angiotensine (ACE). L'angiotensine II est le ligand naturel des récepteurs AT1 et AT2. La fixation de l'angiotensine II sur le récepteur AT1 induit l'activation de plusieurs voies de signalisation induisant la vasoconstriction et l'augmentation de la pression artérielle. Les effets d'AT1 sont contrebalancés par l'action des récepteurs AT2 et MAS, induisant une vasodilatation et une baisse de la pression artérielle. ACE2 est une carboxypeptidase qui élimine un acide aminé de l'angiotensine I produisant l'angiotensine 1-9. Cette dernière peut être convertie par ACE en angiotensine 1-7, un ligand du récepteur MAS. ACE2 peut également convertir l'angiotensine II en angiotensine 1-7.

2.2. Evolution des récepteurs de l'angiotensine II

En utilisant des méthodes phylogénétiques, notre équipe a mis en évidence l'évolution miroir des récepteurs AT1 et AT2 en deux groupes séparés par l'unique récepteur de l'AngII de la lamproie (*Lampetra Fluviatilis*, une espèce ancienne de poisson sans mâchoire). Chacun des deux récepteurs est aussi réparti en deux sous-groupes qui séparent les récepteurs des poissons des récepteurs terrestres. De façon intéressante, des mutations, survenues lors de la transition à la vie terrestre, sont situées au niveau du site de fixation du sodium [81]. En effet pour les deux récepteurs, le passage à la vie terrestre s'accompagne de la mutation d'une sérine en une alanine en position 2.49. Dans le cas d'AT1, une mutation d'une sérine à une asparagine au niveau de la position 7.46 a eu lieu chez les amniotes (les espèces chez lesquelles l'embryon puis le fœtus sont protégés par un sac amniotique) [81]. Parallèlement, une mutation V3.36M est apparue dans le cas d'AT2 chez les amniotes [81].

Ces mutations du site sodium semblent modifier la balance entre spécificité et promiscuité des deux récepteurs, ce qui pourrait avoir contribué à la séparation des deux axes du système rénine-angiotensine chez les amniotes (Figure 19) [81].

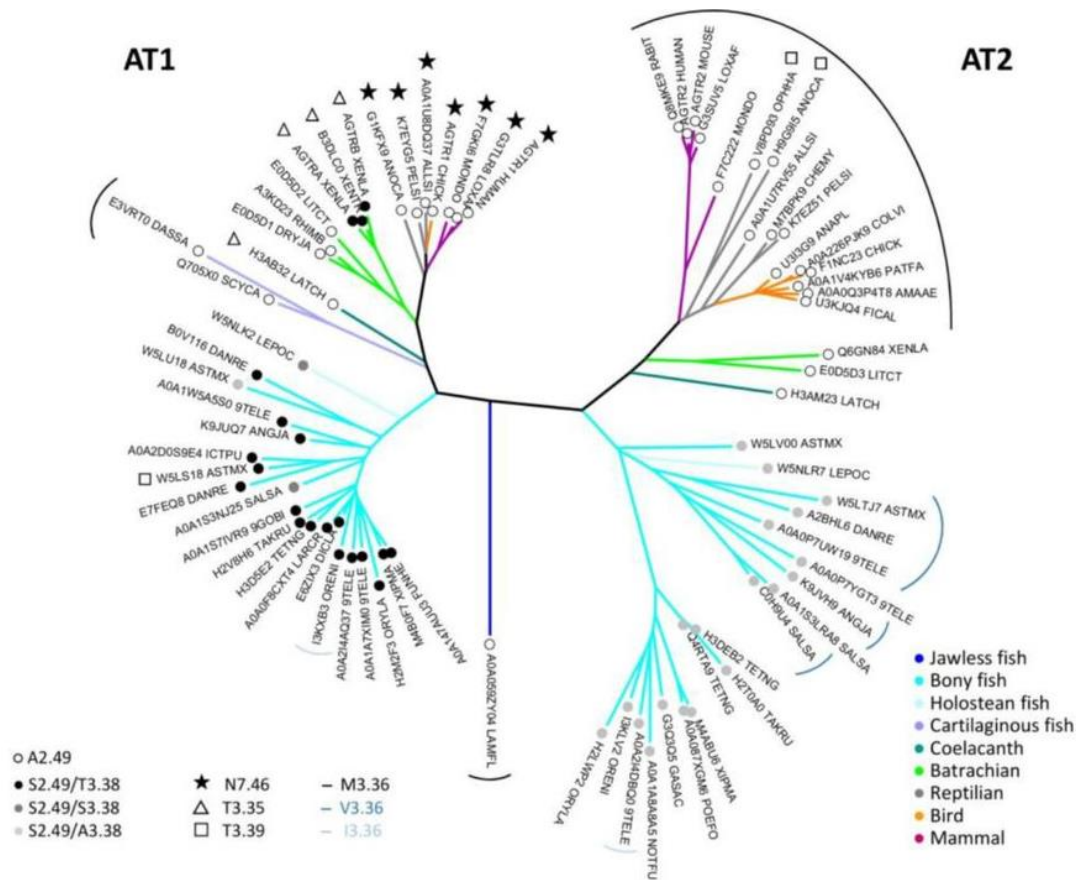


Figure 19 : Arbre phylogénique des récepteurs AT1 et AT2.
Figure issue de [81]

2.3. Rôle de l'ion sodium dans la régulation des récepteurs de l'angiotensine

La mutation de certains résidus du site de fixation de l'ion sodium a un effet sur l'activation du récepteur AT1. Ainsi, la substitution de l'Asn111 (N3.35) par Gly a été signalée comme provoquant une activation constitutive partielle du récepteur [82]. De même, des mutations au niveau de l'Asn295 (N7.46) dans TM7 produisent une faible activation constitutive [83]. Un effet allostérique négatif de l'ion sodium a été clairement observé avec le mutant AT1 N7.46S dont l'activité constitutive dépend de la concentration de sodium [84].

Malgré la conservation du site de fixation du sodium des récepteurs AT1 et AT2, la possibilité de fixer l'ion sodium dans le domaine transmembranaire de ces récepteurs reste controversée. Des études antérieures suggèrent que le sodium ne peut pas se fixer au récepteur AT1 humain en se basant sur des études expérimentales. Zheng *et al.* ont en effet montré que le sodium ne présente pas un effet allostérique négatif et une modification de la signalisation du récepteur [85].

Dans le cadre de l'étude évolutive d'AT1 et AT2, nous avons mis en évidence, par les approches de simulations dynamiques et de calculs électrostatiques, que l'ion sodium peut se fixer sur AT1 humain et que cette fixation stabilise la structure du récepteur [81]. La fixation du sodium a été observée dans des simulations accélérées.

Un clash entre deux ions sodium et un changement des rotamères des résidus 6.51 et 3.32 a permis l'entrée du sodium à partir du milieu extracellulaire vers le domaine transmembranaire d'AT1, en franchissant des barrières énergétiques. Après avoir gagné son site, le sodium interagit préférentiellement avec les chaînes latérales de D2.50 et de N7.46 (Figure 20).

Les analyses électrostatiques montrent que le pKa du résidu D2.50 des récepteurs AT1 et AT2 est semblable à celui du récepteur opiacé δ . Cela renforce l'hypothèse de la fixation du sodium pour les récepteurs AT1 et AT2 [81].

Chez AT1, la cinétique de la fixation de l'ion sodium est lente à cause de présence d'un résidu chargé positivement, K5.42, au niveau du site orthostérique. Ce résidu qui est indispensable à la fixation de l'AngII est conservé chez tous les récepteurs AT1 et AT2. Il induit un potentiel électrostatique positif qui est très défavorable à l'entrée et à la sortie du sodium dans le récepteur inactif. Cela pourrait expliquer l'absence d'effet sodium dans les expériences in vitro pour des raisons cinétiques.

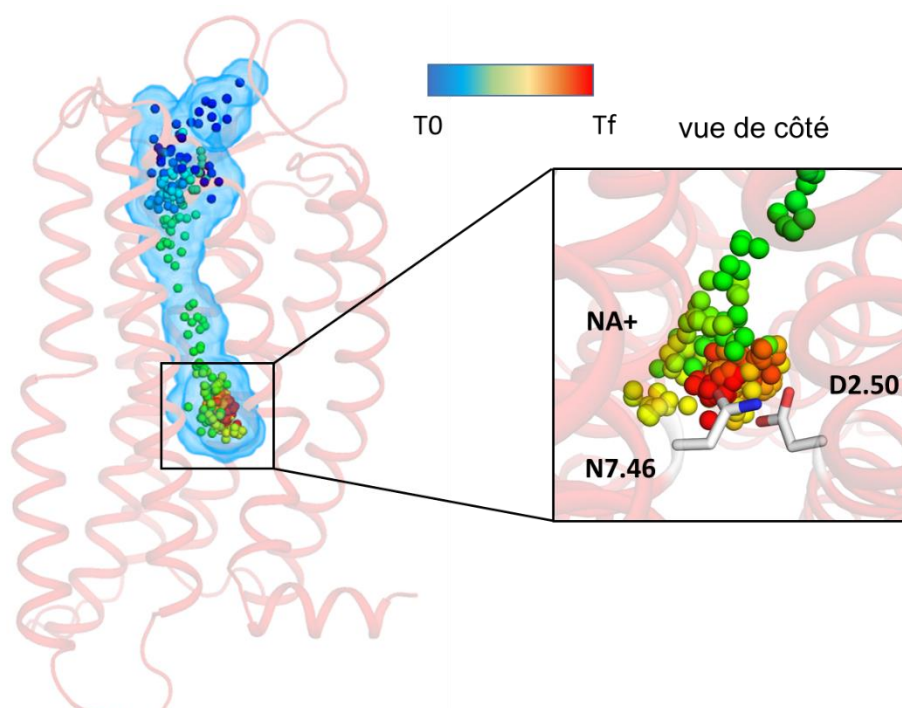


Figure 20 : Visualisation de l'entrée de l'ion sodium dans AT1 humain.

Les snapshots sont distants de 8 ps. L'évolution temporelle de la position du sodium est indiquée par les couleurs de l'arc en ciel du bleu vers le rouge. Le nuage bleu présente la densité volumétrique de l'ion sodium durant sa trajectoire. Le cadre représente le site de fixation du sodium vu de côté.

2.4. Le récepteur AT1

2.4.1. Le récepteur AT1 et ses voies de signalisation

Le récepteur AT1 est un récepteur exprimé dans tous les organes, en particulier les cellules du système cardiovasculaire telles que les cellules endothéliales, les cellules musculaires lisses, les cardiomyocytes. Chez l'humain, AT1 est codé par le gène *AGT1R* du chromosome 3. L'expression du gène donne lieu à un récepteur de 359 acides aminés dont la partie extracellulaire est stabilisée par deux ponts disulfure [80]. AT1 peut activer plusieurs voies de signalisation médiées par les protéines G ou les β -arrestines.

Le récepteur AT1 peut se coupler à différentes protéines G : Gq, Gi/o et G12/13. L'activation de Gq induit l'hydrolyse du phosphatidylinositol bisphosphate par la phospholipase C β spécifique des phosphoinositides, ce qui entraîne à terme l'augmentation de la concentration de Ca²⁺ intracellulaire. Dans les cellules musculaires lisses vasculaires (CMLV), l'augmentation de la concentration calcique va induire la phosphorylation de la chaîne légère de la myosine (CLM) et la contraction musculaire. AT1 peut également se coupler aux protéines Gi/o, ce qui entraîne l'inhibition de l'adénylyl cyclase et de la protéine kinase A, et aux protéines G12/13, ce qui entraîne l'activation de la Rho-kinase. Dans tous les cas, ces voies de signalisation conduisent à la phosphorylation de CLM et la contraction des CMLV [86]. L'activation d'AT1 induit aussi l'activation de diverses protéines kinases, de sous-unités de la nicotinamide adénine dinucléotide phosphate (NADPH) oxydase, de la transactivation des récepteurs de facteurs de croissance, ou des interactions directes avec des protéines interagissant avec le récepteur AT1, telles que la Janus kinase 2 (JAK2), la phospholipase C (PLC) γ 1, la protéine associée au récepteur AT1 (ATRAP), la protéine associée au récepteur de l'angiotensine II de type 1 (ARAP1) et la protéine semblable au facteur d'échange de nucléotides de la guanine (GEF) [6].

Suite à l'activation d'AT1 par l'AngII, les mécanismes médiés par les β -arrestines peuvent être considérés comme une deuxième vague d'événements de transduction du signal aboutissant à l'internalisation du récepteur [86] ou d'autres voies de signalisation. Les événements de signalisation médiés par la β -arrestine comprennent l'activation des tyrosines kinases Src, de la kinase régulée par le signal extracellulaire 1/2 (ERK 1/2), de la kinase c-Jun N-terminale (JNK), des cascades de protéines kinases activées par des agents mitogènes (MAPK), des protéines kinases activées par des agents mitogènes Akt et p38 [86]. Ces événements sont favorisés par certains ligands (ligands biaisés vers la voie des β -arrestines) (Figure 21).

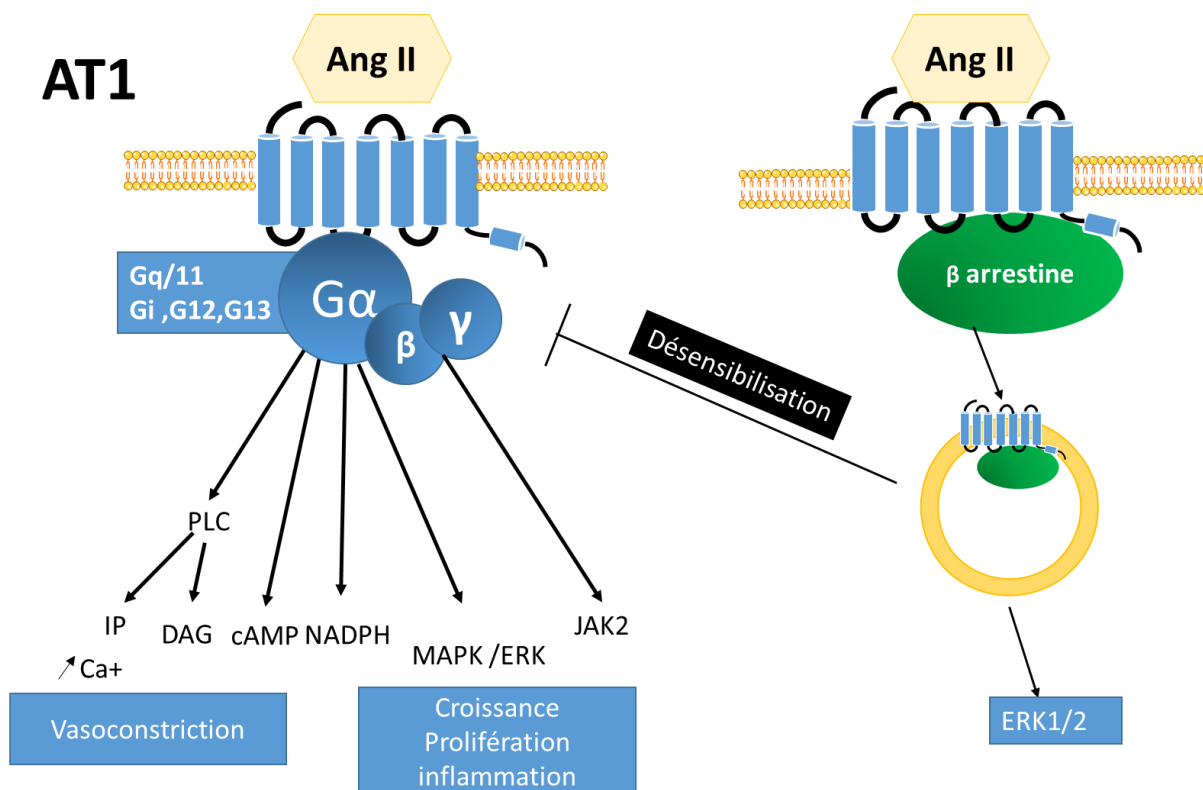


Figure 21 : Schéma simplifié des voies de signalisation médiées par le récepteur AT1.

2.4.2. Les biais de signalisation d'AT1

Les ligands des récepteurs membranaires étaient à l'origine classés en agonistes et antagonistes (ou plus récemment en agonistes inverses). Cependant, plusieurs ligands sont capables d'initier sélectivement une ou plusieurs voies de transduction du signal. Ce phénomène a été appelé "agonisme biaisé" ou "sélectivité fonctionnelle" [86].

Il existe des ligands synthétiques dérivés de l'angiotensine II qui peuvent jouer le rôle d'agoniste biaisé, d'antagoniste ou d'agoniste partiel. Les agonistes biaisés favorisent une voie de signalisation spécifique. Cette propriété peut être très intéressante à des fins thérapeutiques, pour favoriser les voies de signalisation bénéfiques et réduire les voies de signalisation délétères provoquant des effets secondaires négatifs. Le tableau 2 ci-dessous résume les ligands d'AT1 et leur fonctions [84,86].

Tableau 2 : Les ligands d'AT1 et leurs fonctions

Ligand AT1	Rôle	Fonction et voies de signalisation
Angiotensine II	Agoniste naturel	Active les voies de signalisation médiées par les protéines G et les β -arrestines
[Sar1, Ile4, Ile8] Angiotensine II (SII-AngII)	Agoniste biaisé β -arrestines	Favorise la phosphorylation d'AT1, l'internalisation et l'activation d'ERK, mais pas le renouvellement du phosphatidylinositol
Losartan, Olmesartan, Candesartan, Eprosartan, Valsartan, Azilsartan et Telmisartan	Antagonistes	Bloqueurs d'AT1, utilisés pour prévenir ou traiter certaines maladies cardiovasculaires tel que l'hypertension, l'insuffisance cardiaque, l'infarctus du myocarde, l'accident vasculaire cérébral, etc.
TRV023, TRV026 TRV027	Agonistes biaisés β -arrestines	Favorisent les voies de signalisation des β -arrestines, inhibant l'hypertension médiée par l'AngII et augmentant la contractilité cardiaque
TRV055, TRV056	Agonistes biaisés Gq	Présentent un couplage allostérique à Gq > 10 fois supérieur à celui de l'AngII et sont également plus efficaces pour stimuler la signalisation cellulaire médiée par Gq (par exemple, la génération d'IP1), mais ils présentent un couplage allostérique à la β -arrestine similaire à celui de l'AngII.

2.4.3. Les structures cristallographiques résolues d'AT1

Durant les 5 dernières années, six structures cristallographiques du récepteur AT1 ont été résolues par cristallographie aux rayons X, dont deux à l'état inactif et quatre à l'état actif. La conformation globale du récepteur AT1 est très similaire à celle des récepteurs des chimiokines et des opioïdes, avec une structure en « kink » de TM2 [89]. Deux ponts disulfure, semblables à ceux observés dans les récepteurs des chimiokines CXCR4 et CCR5, contribuent à façonner le côté extracellulaire d'AT1. Le pont Cys18-Cys274 relie l'extrémité N-

terminale et ECL3, tandis que le pont Cys101-Cys180 relie TM3 et ECL2 [89]. L'état inactif du récepteur est représenté par les structures 4ZUD et 4YAY d'AT1 en interaction avec l'agoniste inverse olmesartan ou l'antagoniste sélectif ZD7155. L'état inactif est caractérisé par une disposition en arrière de TM6 et en avant de TM7. Des liaisons hydrogène intramoléculaires sont observées entre N3.35 et N7.46, stabilisant la structure inactive en absence de l'ion sodium. La perte de ces liaisons-H pourrait expliquer l'activation constitutive d'AT1 lors des mutations de N3.35 et N7.46 [83].

En 2020, 3 structures cristallographiques d'AT1 à l'état actif, stabilisées par des nanocorps synthétiques, ont été résolues. Les récepteurs sont liés à 3 ligands différents : l'angiotensine II, qui est l'agoniste naturel, et 2 analogues biaisés vers la voie des β -arrestines, TRV026 et TRV023 (codes PDB : 6OS0, 6OS1 et 6OS2) [9]. La superposition des structures inactive (4YAY) et active d'AT1 (6OS0) (Figure 22) montre un mouvement remarquable vers l'avant de l'hélice TM6, accompagné par un mouvement vers l'arrière de l'hélice TM7 ainsi que la stabilisation de la boucle intracellulaire ICL2 sous forme d'une hélice α à l'état actif.

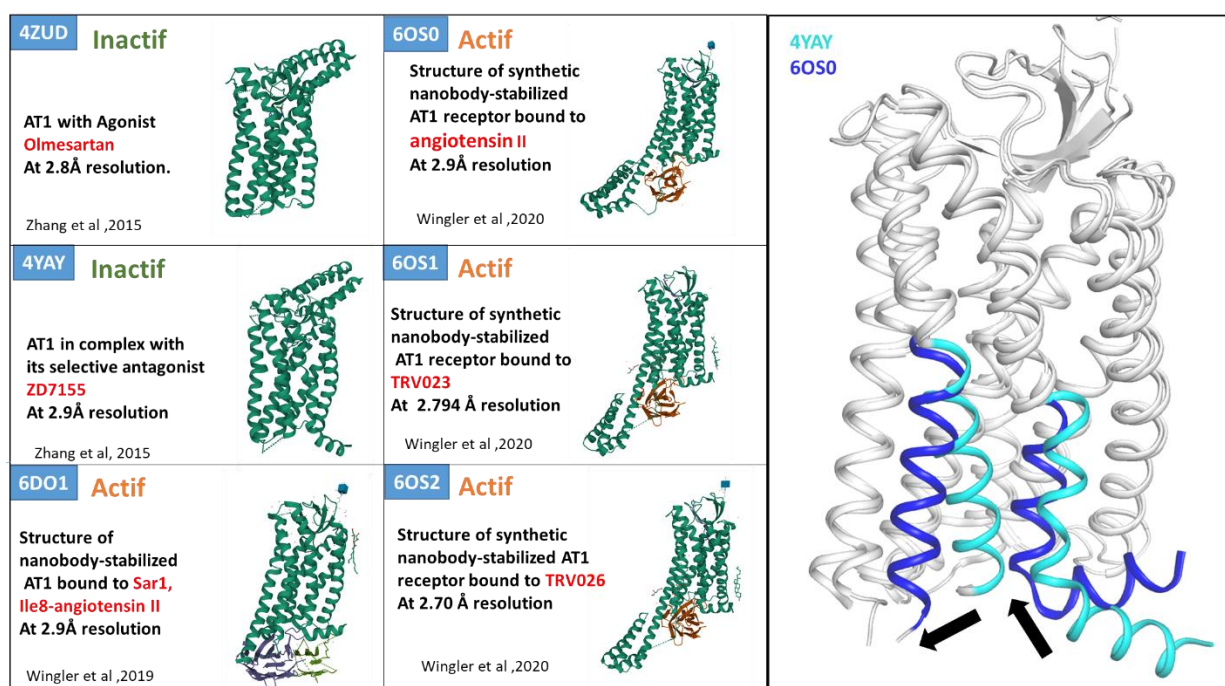


Figure 22 : Structures cristallographiques d'AT1 présentes dans la Protein Data Bank.
Le panneau de droite montre la superposition d'une structure inactive (4YAY) et d'une structure active (6OS0).

2.4.4. Mécanismes moléculaires des biais de signalisation d'AT1

La comparaison des conformations actives induites par l'AngII et les ligands biaisés met en évidence des différences très importantes au niveau du site de la fixation de l'ion sodium. En particulier, N3.35 est orienté vers l'extérieur dans le complexe avec l'AngII et vers l'intérieur dans les complexes avec les ligands biaisés. Cela suggère que N3.35 peut jouer un rôle de commutateur pour favoriser des conformations capables d'activer sélectivement les β -arrestine sans activer les voies de signalisation médiées par les protéines G [9].

De manière intéressante, en réalisant des simulations dynamiques extensives, Suomivuori et al. ont aussi mis en évidence que la position N3.35 peut jouer un rôle de commutateur entre les voies des protéines G et des β -arrestines. En particulier, l'orientation de N3.35 vers l'intérieur conduirait, via un réseau de liaisons-H, à des orientations alternatives de R3.50 (motif DRY) et de Y7.53 (motif NPxxY) qui empêcheraient la fixation des protéines G mais non des β -arrestines. Une orientation vers l'extérieur de N3.35 permettrait la fixation aussi bien des protéines G que des β -arrestines, favorisant ainsi une signalisation « neutre » (Figure 23) [9,89].

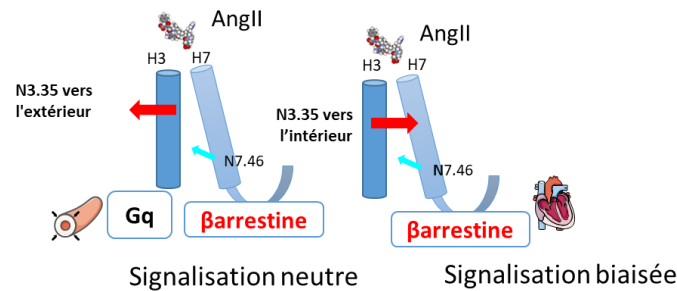


Figure 23 : Représentation schématique du rôle de N3.35 dans l'activation biaisée d'AT1.

2.5. Le récepteur AT2

2.5.1. L'énigme du récepteur AT2

Le récepteur AT2 est codé par le gène *AGTR2* situé sur le chromosome X [91]. Ce récepteur a une masse moléculaire de 41 000 Da [91]. AT2 est exprimé principalement au cours de la vie fœtale, son expression diminue rapidement après la naissance, sauf dans des conditions pathologiques ou certains tissus ou organes [7]. L'utérus exprime le récepteur AT2 en abondance même chez l'adulte, ce qui suggère un rôle dans la reproduction des mammifères [92]. Il est aussi décrit comme ayant des effets opposés au récepteur AT1, participant à des mécanismes de cardioprotection et à une diminution de la pression artérielle [7,91]. De plus, le récepteur AT2 est une nouvelle cible pour le développement de nouveaux agents de traitement de l'hypertension [7] et de la douleur neuropathique [93]. La séquence d'AT2 présente 32 % d'homologie avec le récepteur AT1. Une des différences majeures se situe au niveau du site de fixation du sodium qui est un modulateur allostérique négatif des RCPG de la famille A [44]. Le résidu 7.46 est une sérine pour AT2 et une asparagine pour AT1, un résidu dont la mutation chez AT1 peut induire l'activation constitutive. La présence d'une sérine en position 7.46 pourrait participer à l'activation constitutive d'AT2, observée en l'absence de stimulation par l'AngII et induisant l'apoptose [94].

Les mécanismes de transduction du signal pour le récepteur AT2 ne sont pas encore définis. Bien que la structure moléculaire du récepteur AT2 ressemble à celle de la superfamille des récepteurs couplés aux protéines G, la transduction du signal par couplage direct aux protéines G reste incertaine [90,93].

Les effets inhibiteurs de la croissance cellulaire du récepteur AT2 sont au moins partiellement médiés par l'activation de la protéine tyrosine phosphatase (PTPase), ce qui entraîne l'inactivation de la kinase régulée par le signal extracellulaire (ERK) [96]. Diverses phosphatases, telles que PP2A, MKP-1 et SHP-1, sont activées par la stimulation du récepteur AT2, via un mécanisme auquel participent les protéines G, ce qui entraîne l'inactivation de ERK, l'ouverture des canaux K⁺ et l'inhibition du courant Ca²⁺. De plus, l'activation des récepteurs AT2 produit du NO induisant la vasodilatation vasculaire [96]. Des études suggèrent que le récepteur AT2 active Gi dans la modulation des canaux K. Dans l'ensemble, les preuves actuelles suggèrent que la troisième boucle intracellulaire du récepteur AT2 est liée à une voie de signalisation impliquant Gi et l'inactivation de ERK [95] (Figure 24). Cependant, aucune activation *directe* des protéines G par AT2 n'a pu être mise en évidence dans des conditions physiologiques [90,93].

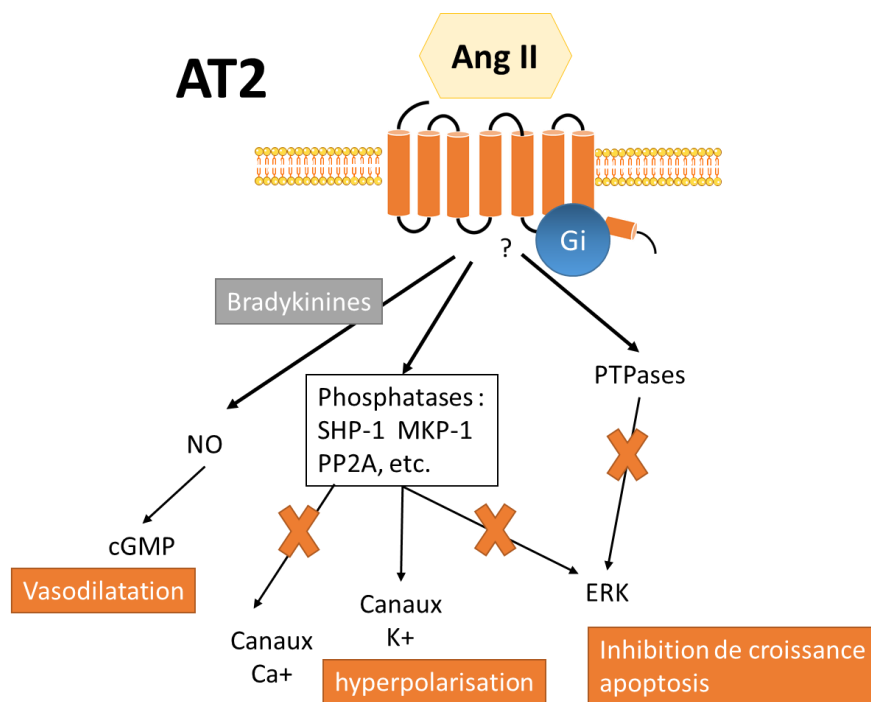


Figure 24 : Schéma simplifié des voies de signalisation médiées par le récepteur AT2.
Figure inspiré de [92]

2.5.2. Les structures cristallographiques d'AT2

A partir de 2017, six structures du récepteur AT2 ont été résolues par cristallographie aux rayons X. Ces structures présentent la position en avant de l'hélice TM6 et la position en arrière de l'hélice TM7, caractéristiques de l'activation, ce qui suggère que ces structures correspondent à l'état actif du récepteur (Figure 25).

Dans la plupart des structures de RCPG, l'hélice H8 est parallèle à la membrane et pointe vers l'extérieur de TM7. Cependant, dans les structures d'AT2 lié à des inhibiteurs sélectifs de la liaison de l'AngII à AT2 (codes PDB : 5UNH, 5UNF, 5UNG, 5XJM), l'hélice 8 adopte une conformation inhabituelle en se retournant pour interagir avec les extrémités intracellulaires de TM3, TM5 et TM6. Cette position non canonique de l'hélice 8 suggère que, malgré la conformation de type actif de TM6 et TM7, l'hélice 8 peut bloquer la liaison des protéines G et des β -arrestines [97]. Cependant, la structure 6JOD résolue en 2020, dans laquelle AT2 est couplé à son ligand naturel, l'angiotensine II, révèle que, dans ce cas, l'hélice 8 est positionnée dans la conformation canonique des RCPG, ce qui permettrait au récepteur de lier les protéines G et les β -arrestines [98]. Cette structure dévoile aussi les interactions entre l'angiotensine II et AT2 qui jouent un rôle important dans l'activation du récepteur. En particulier, l'interaction entre la phénylalanine 8 de l'AngII et la méthionine 128 (M3.36) du récepteur peut contribuer au changement structural d'AT2, conduisant à la normalisation de la position de l'hélice 8 [98].

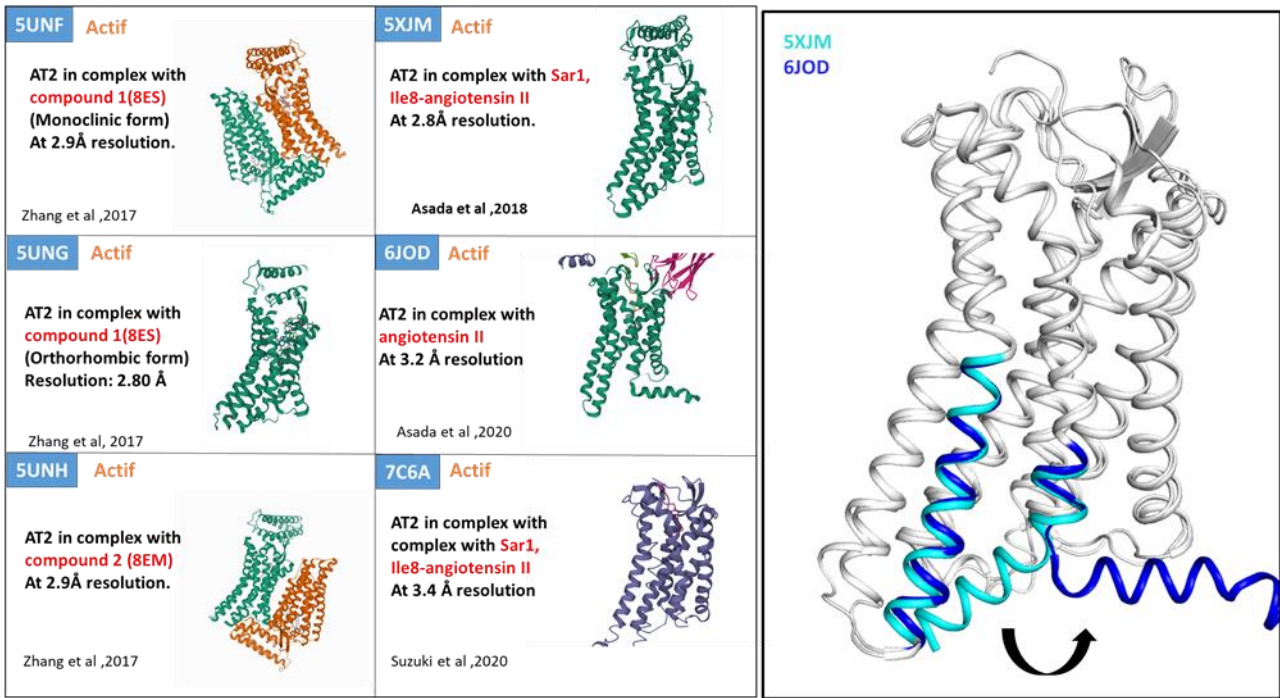


Figure 25 : Structures cristallographiques d'AT2 présentes dans la Protein data Bank.
Le panneau de droite montre la superposition de deux structures actives se différenciant par la position de l'hélice 8.

2.6. Les récepteurs de l'angiotensine et le stress mécanique

2.6.1. Le stress mécanique dans les vaisseaux sanguins

Les vaisseaux sont soumis en permanence à des contraintes mécaniques associées au flux sanguin. La pression induit une vasoconstriction myogénique (tonus myogénique, TM) et les forces de cisaillement induisent une vasodilatation dépendante de l'endothélium (dilatation dépendante du flux, FMD) (Figure 26). L'équilibre entre TM et FMD permet un contrôle précis des débits sanguins locaux qui sont perturbés dans de nombreuses pathologies telles que l'hypertension, l'insuffisance cardiaque et le diabète [99]. Pour pouvoir réajuster cet équilibre, il est nécessaire de mieux comprendre les mécanismes impliqués dans la mécano-sensation et les réponses cellulaires à des stimuli physiques.

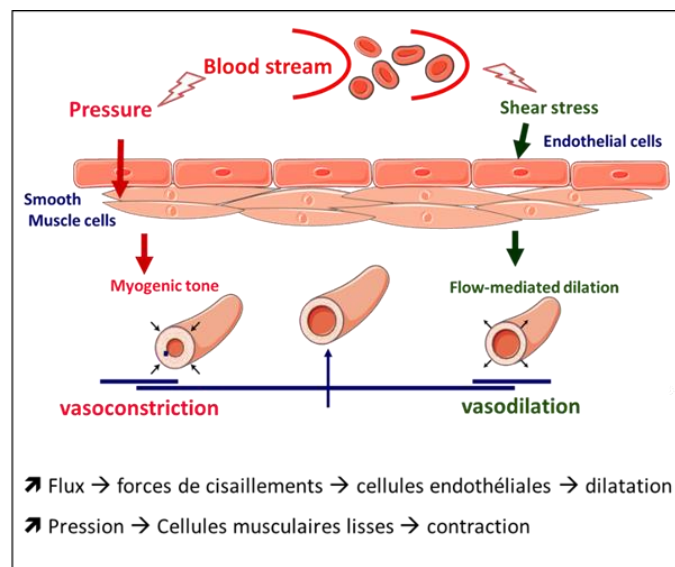


Figure 26 : Contraintes mécaniques associées au flux sanguin.

2.6.2. AT1 et stress mécanique

La notion d'activation du récepteur AT1 par l'effet de l'étirement membranaire a été introduite pour la première fois par Sadoshima *et al.* en 1993 [65]. Ces auteurs ont montré que l'étirement mécanique provoque la libération d'Ang II des myocytes cardiaques et que l'Ang II agit comme un médiateur initial de la réponse hypertrophique induite par l'étirement (Figure 27). Ces données suggèrent un mécanisme autocrine pour l'activation d'AT1 par le stress mécanique et, par conséquent, la croissance des cellules musculaires cardiaques induite par l'étirement [65].

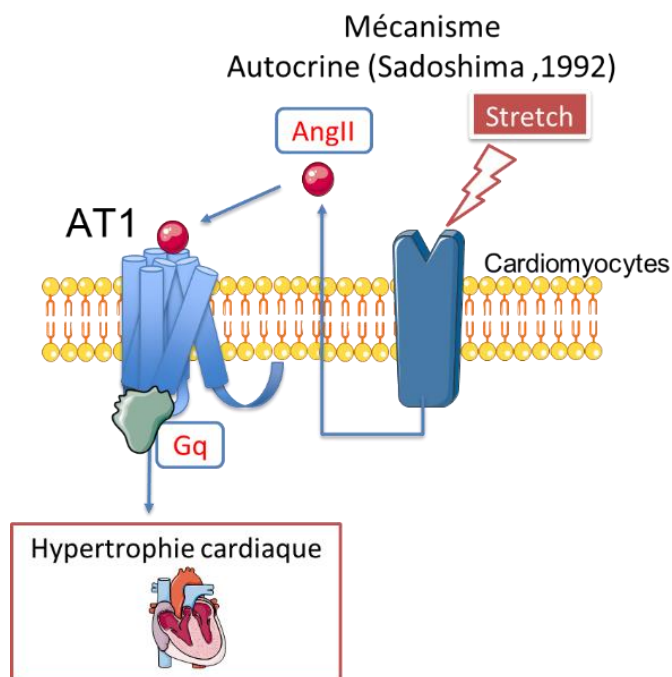


Figure 27 : Mécanisme autocrine d'activation du récepteur AT1 par l'effet d'étirement.

La libération autocrine d'AngII est responsable de l'hypertrophie des cardiomyocytes induite par l'étirement [65].

Plus récemment, d'autres études montrent un effet plus direct et indépendant de l'angiotensine II pour l'activation d'AT1 induite par l'étirement membranaire dans l'hypertrophie cardiaque (Figure 28). En utilisant des anticorps ciblant l'AngII ou l'utilisation des souris déficientes en angiotensinogène [10], Zhou *et al* ont éliminé la possibilité d'un mécanisme autocrine ou paracrine d'activation d'AT1 par le stress mécanique, suggérant ainsi un rôle mécanosenseur d'AT1. L'effet de l'étirement sur AT1 active la voie Gq induisant une hypertrophie cardiaque, en activant les voies de signalisation ERK. De plus, cet effet peut être inhibé en bloquant le récepteur AT1 avec le candésartan [10,98].

Wang *et al.* ont montré que le stress mécanique peut aussi activer la voie des β -arrestines induisant ainsi l'augmentation de la contractilité cardiaque mais le mécanisme d'activation est plus complexe que celui induit par les agonistes biaisés vers la β -arrestine 2, TRV023 and TRV026. En effet, l'étirement de la membrane favorise uniquement le couplage de la protéine Gi au récepteur AT1 qui est nécessaire pour le recrutement de la β -arrestine 2 et l'activation des voies de signalisation dépendantes de la β -arrestine 2 [67].

L'étirement mécanique induit aussi l'activation des canaux TRPC par une stimulation mécanique d'AT1 via des voies de signalisation dépendantes de phospholipases, induisant la dépolarisation de la membrane en augmentant la concentration du calcium intracellulaire, provoquant ainsi une vasoconstriction [101].

Une autre étude expérimentale a démontré que les forces de cisaillement de 15 dyne/cm² peuvent activer le

récepteur AT1 en induisant des voies de signalisation dépendante ou indépendantes de la protéine Gq [11]. Les mécanismes moléculaires induisant la réponse d'AT1 à des stimuli mécaniques ne sont pas très bien compris: les facteurs physiques peuvent agir directement sur l'équilibre conformationnel d'AT1, en appliquant des forces externes, ou indirectement en altérant les propriétés physiques des lipides et agissant sur la fluidité membranaire.

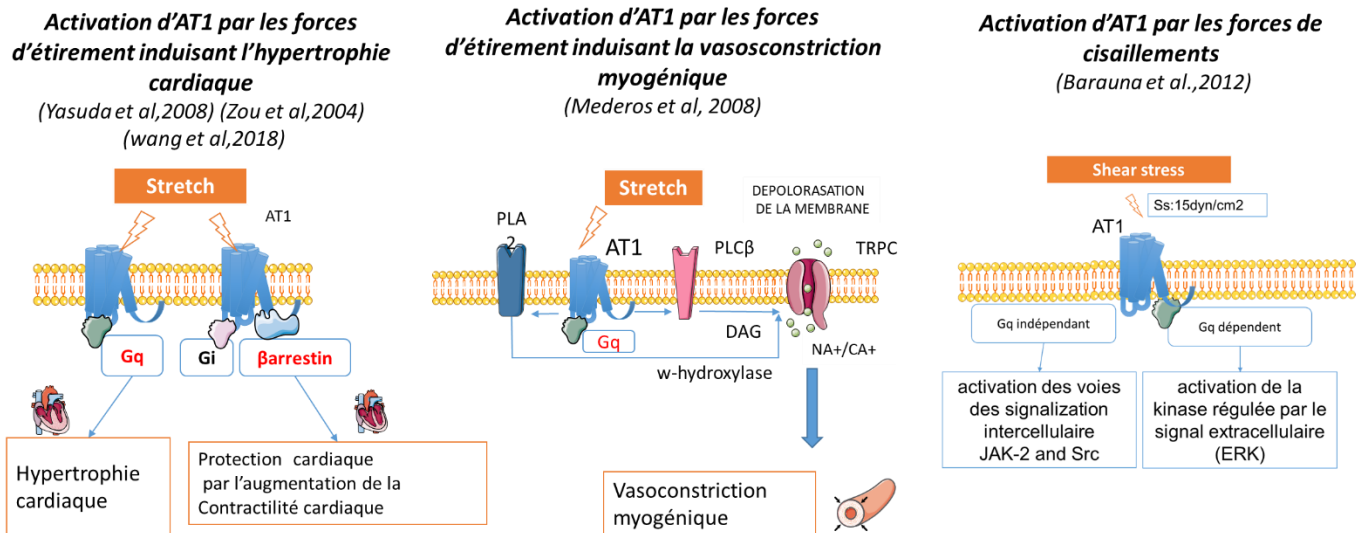


Figure 28 : Activation du récepteur AT1 par le stress mécanique.
Les forces d'étirement et de cisaillement peuvent activer AT1.

2.6.3. Exemples de pathologies où AT1 et le stress mécanique sont impliqués

Le système rénine-angiotensine est l'un des systèmes de régulation les plus importants des fonctions cardiovasculaires, via les récepteurs de l'angiotensine II, AT1 et AT2. Un déséquilibre de ce système contribue au développement du dysfonctionnement endothélial et de nombreuses pathologies vasculaires. En effet, le récepteur AT1 est fortement impliqué dans la pathologie de l'hypertension et constitue la cible thérapeutique de certains médicaments hypotenseurs qui sont des antagonistes des récepteurs de l'angiotensine II (famille des sartans). Avantageusement, des études montrent que les antagonistes d'AT1 peuvent aussi bloquer l'activation spontanée d'AT1 par le stress mécanique [10].

Plus récemment, le rôle du stress mécanique et la composition lipidique membranaire ont été mis en évidence dans d'autres pathologies telles que l'athérosclérose et la croissance tumorale.

a) Exemple 1: l'athérosclérose

Dans le cas de l'athérosclérose, AT1 et AT2 sont surexprimés [100,101]. Plusieurs études montrent la participation du récepteur AT1 à la pathogénie de l'athérosclérose par son action pro-inflammatoire et par l'augmentation de la pression artérielle et le stress oxydatif [104]. AT2 semble avoir des effets opposés à AT1 dans le remodelage vasculaire, la prévention de l'athérosclérose et la baisse de la pression [101,103].

Plusieurs études suggèrent que les traitements visant à inhiber l'activation des récepteurs AT1 sont une option thérapeutique prometteuse dans la prévention de l'athérosclérose et dans le retardement de sa progression avec des molécules déjà disponibles ou de nouveaux ligands sélectifs [106]. Il y a donc un intérêt thérapeutique

certain à approfondir les mécanismes moléculaires impliqués dans l'activation/inhibition des récepteurs de l'angiotensine dans le cas de l'athérosclérose.

Le développement de l'athérosclérose est sensible aux changements dans les caractéristiques biochimiques et physiques du microenvironnement cellulaire dont la modification de la composition membranaire [107] et le stress mécanique [108]. En effet, de faibles forces de cisaillement (shear stress) induites par le flux sanguin constituent un environnement mécanique qui favorise le dysfonctionnement vasculaire et l'athérosclérose. À l'inverse, une force de cisaillement physiologiquement élevée est protectrice [106,107]

Le récepteur AT1 fait partie des RCPG dont l'expression [110] et l'activation [10] sont affectés par un stress mécanique mais le mécanisme moléculaire induisant l'activation reste inconnu en particulier dans le cas d'athérosclérose où l'effet du shear stress varie selon son intensité (fort/faible) et sa nature (laminaire/oscillatoire). Dans tous ces cas, le stress mécanique induit une modification de la tension de surface de la membrane [13]. Cette dernière est sensible au profil lipidique de la membrane. L'hypercholestérolémie est un des facteurs de risque du développement de l'athérosclérose. Des études montrent une corrélation entre l'hypercholestérolémie et la surexpression du récepteur AT1 et de l'angiotensine II [100,109]. Le cholestérol régule la fluidité membranaire vers un état plus dense, ce qui diminue la diffusion latérale des bloqueurs d'AT1 tel que le Candesartan [112]. D'autres études suggèrent la présence d'un site de fixation du cholestérol chez les RCPG modulant ainsi leur activité [113].

b) Exemple 2: Cancer et angiogénèse

Dans le contexte tumoral, l'équilibre du système rénine-angiotensine est aussi altéré. En effet, dans de nombreux types de cancer et à des stades élevés, le récepteur AT1 est surexprimé et joue un rôle important dans la vascularisation et la croissance tumorales et dans la formation de métastases [46]. Le récepteur AT2 est surexprimé dans quelques types de cancer mais reste peu étudié [114]. AT2 est décrit comme ayant des effets d'inhibition de la prolifération cellulaire [91]. Des récentes études suggèrent de cibler le système rénine-angiotensine pour le traitement complémentaire de certains cancers, en bloquant les voies de signalisation d'AT1 [112-113], avec des molécules déjà disponibles ou de nouveaux ligands sélectifs. Ici aussi, il y a donc un intérêt thérapeutique certain à approfondir les mécanismes d'action des récepteurs AT1 et AT2 dans un environnement tumoral.

Les tissus cancéreux, notamment les tumeurs, présentent des changements importants dans les caractéristiques biochimiques et physiques du microenvironnement cellulaire. Le profil et l'hémostase lipidique sont altérés, ce qui conduit à la modification de la composition membranaire et de la tension de surface de la membrane [116]. Cette dernière est aussi sensible au stress mécanique généré par les forces cisaillement du flux sanguin et les forces d'étirement et de compression. Les forces mécaniques dans le microenvironnement de la tumeur jouent un rôle important dans sa croissance et sa vascularisation anormale [117]. Pour répondre aux signaux mécaniques, les cellules endothéliales tumorales utilisent plusieurs voies de signalisation activées par des protéines membranaires mécanosensibles dont certains RCPG [117]. Le récepteur AT1 fait partie des RCPG qui peuvent être activés par un stress mécanique [66]. D'autre part, dans plusieurs types de cancer (sein, col de l'utérus, carcinome et lymphome), une augmentation des lipides anioniques tels que les phosphatidylglycerols (PG) est observée avec des taux de variation corrélés à l'évolution du cancer [114,116-117]. Les PG peuvent faciliter ou maintenir l'activation de certains RCPG [57].

Etudier l'effet du stress mécanique et de la composition lipidique membranaire sur les réponses moléculaires des récepteurs AT1 et AT2 permettra, dans un premier temps, de comprendre comment les fonctions de ces récepteurs sont altérées dans un environnement pathologique et dans un deuxième temps de cibler éventuellement des conformations spécifiques de cet environnement dans un but thérapeutique.

3. Modélisation et dynamique moléculaires

3.1. La modélisation moléculaire

La modélisation moléculaire permet de simuler et reconstruire en 3 dimensions la structure d'une molécule en utilisant divers outils bioinformatiques.

Dans cette partie, nous allons présenter nos travaux publiés dans *Methods in Molecular Biology* sous la forme de 2 chapitres. Ces deux chapitres décrivent les différentes étapes et les astuces pour réussir la modélisation moléculaire, tout en comparant les techniques de modélisation à l'aide du logiciel Modeller [120] avec les programmes de modélisation automatiques disponibles en ligne. Nous allons étudier deux types de modélisation : la modélisation par homologie [30] et la modélisation par threading (Ben Boubaker et al., 2022, *Methods Mol Biol*, accepted) (Annexe II.2).

Les deux types de modélisation de la séquence « cible » dépendent de la présence d'un patron ou « template » qui est généralement une structure cristallographique d'une protéine homologue, c'est-à-dire partageant un ancêtre commun, ce qui se traduit par une similarité structurale, séquentielle, et éventuellement fonctionnelle. Les deux types de modélisation diffèrent par la facilité avec laquelle le patron est trouvé. Plus le pourcentage de similarité et le nombre d'acides aminés alignés sont élevés, plus le patron et le modèle qui en résulte sont fiables. En fonction de ces paramètres, on distingue 3 zones de modélisation (Figure 29):

La zone de confort ou « Safe zone » qui concerne principalement la modélisation dite par homologie : lorsque l'identité de séquence entre la cible et le patron est supérieur à 25-30% pour une longueur alignée d'au moins 100 acides aminés, le modèle est fiable car l'identité de séquence est synonyme d'origine commune.

La zone crépusculaire ou « Twilight zone » : Lorsque l'identité de séquence entre la cible et le template descend en dessous de 25-30%, la modélisation par homologie devient progressivement de plus en plus difficile mais il est possible grâce à l'exploitation des méthodes d'enfilage ou « threading » de trouver les patrons compatibles. La connaissance d'informations sur l'évolution des séquences aide pour choisir et valider un patron qui doit avoir une origine commune avec la cible.

La zone noire ou « Dark zone » : En dessous d'un certain seuil d'identité de l'ordre de 10%, la modélisation devient pratiquement impossible en l'absence de patrons homologues. Afin de modéliser une molécule dans la zone noire, il existe la possibilité de déduire la structure 3D à partir des principes physiques et physico-chimiques (méthode *ab initio*) mais cette méthode est limitée aux petites protéines et ne réussit pas aussi bien que les approches basées sur des modèles lorsque ceux-ci sont disponibles.

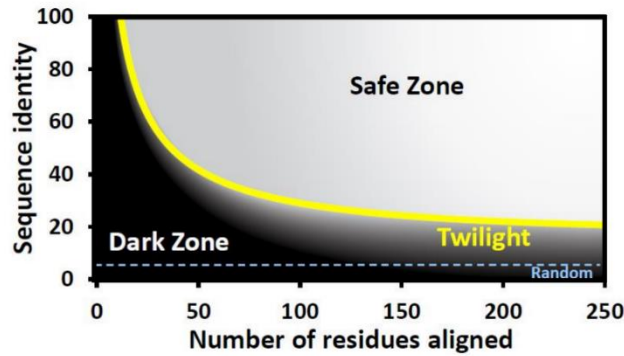


Figure 29 : Présentation schématique des différentes zones de modélisation moléculaire des protéines.
Les zones se différencient par le pourcentage d'identité de séquence et le nombre des résidus alignés.

3.1.1. La modélisation par homologie

Malgré l'explosion structurale des RCPG, 80 % des RCPG non olfactifs n'ont toujours pas de structures résolues. Dans certains cas, la modélisation des RCPGs est nécessaire pour étudier les mécanismes moléculaires associés à ces récepteurs. Notre premier chapitre de modélisation décrit étape par étape la modélisation par homologie d'un récepteur de classe A, le récepteur de la bradykinine B1 à l'état inactif, complexé avec l'ion sodium.

Comme première étape, nous cherchons les patrons des récepteurs les plus proches phylogénétiquement pour modéliser la structure souhaitée. Cette étape peut être effectuée de manière directe en faisant un blast de la séquence cible à partir d'UniProtKB [121] dans la Protein Data Bank (PDB) [122]. Dans notre cas, à partir des résultats obtenus, nous avons choisi trois récepteurs ayant des structures résolues : CCR5 résolu à l'état inactif, AT1 résolu à l'état actif et inactif, et OPRD résolu à l'état inactif avec un ion sodium et des molécules d'eau au niveau du site allostérique. Ce choix se base non seulement sur le pourcentage d'homologie élevé mais aussi sur les caractéristiques souhaitées de la structure à modéliser : présence de l'ion sodium dans son site de fixation ainsi que des molécules d'eau de coordination, état inactif du récepteur, présence des boucles intra/extracellulaire et leur structure (présence d'hélice α ou de feuillet β etc) et l'orientation de l'hélice 8. Ces caractéristiques nécessitent d'avoir de multiples patrons pour modéliser le récepteur B1.

Généralement, la modélisation des boucles est l'étape la plus difficile, dû à leur variabilité structurale et séquentielle ou leur absence dans les patrons homologues. Modeller propose deux fonctions (loopmodel et dope_loopmodel) pour générer/affiner automatiquement les boucles mais le résultat n'est pas toujours fiable. Actuellement, de nombreuses structures de RCPG sont disponibles. Il est hautement préférable de rechercher des boucles équivalentes de même longueur résolues et de les utiliser comme patrons. Dans ce but, j'ai contribué à l'inspection de toutes les structures cristallographiques des RCPG, afin d'identifier toutes les boucles résolues d'une longueur donnée et ainsi les patrons potentiels pour modéliser les boucles du récepteur cible. Un script R a été créé pour l'analyse des longueurs de boucle dans les RCPG humains. A partir d'un fichier d'alignement des RCPG humains, la longueur relative des boucles est mesurée comme le nombre de résidus entre les résidus d'ancrage $n.50$ présents dans chaque hélice n , puis le nombre des RCPG pour chaque longueur de boucle est calculé et visualisé sous forme d'histogramme avec des barres bleues indiquant les boucles résolues (Figure 30).

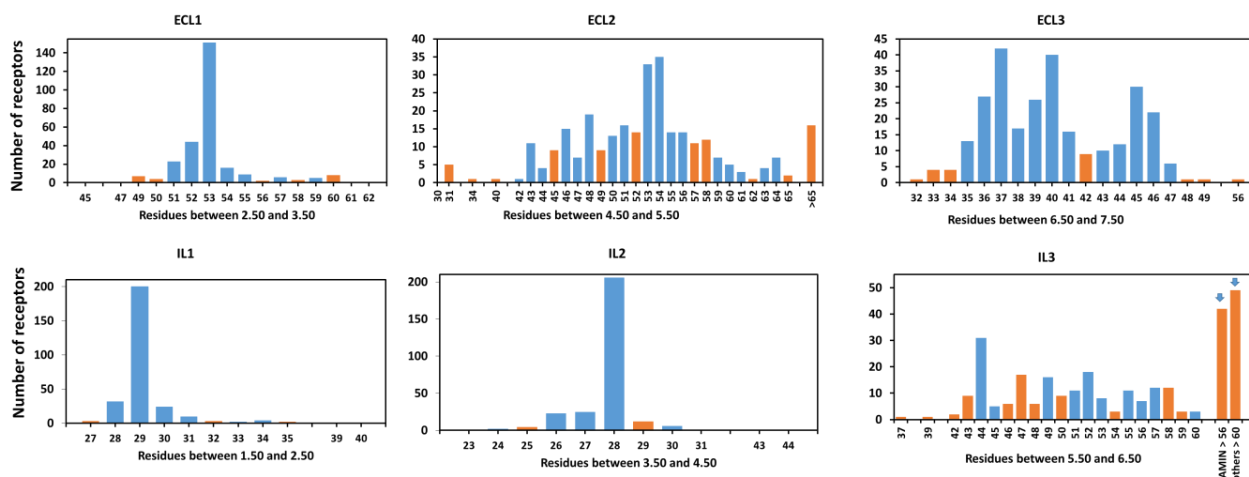


Figure 30 : Analyse statistique des longueurs de boucle dans les RCPG humains (classe A).

Les barres bleues indiquent que des boucles résolues de la longueur indiquée sont présentes dans les structures de RCPG disponibles. Les barres orange indiquent l'absence de boucles résolues. Pour ICL3, les flèches pour les longues boucles correspondent à une longueur de 87 résidus pour le récepteur PEP OX1R (6TP3) et à une longueur de 78 résidus pour le récepteur AMIN ADRB2 (6MXT).

3.1.2. La modélisation en « Twilight zone »

a) Recherche des patrons par les méthodes d'enfilage ou « threading »

Pour modéliser dans la « twilight zone », nous avons choisi l'exemple de la phéromone de la salamandre PRF1 découverte en 1999 mais dont les informations liées à son évolution sont parcellaires. PRF1 présente une faible identité de séquence (<20%) avec les cytokines de la famille IL6, ce qui ne permet de conclure que ces protéines sont homologues. Comme première étape, nous avons cherché les patrons potentiels pour cette séquence. En utilisant un blast classique, aucun résultat n'est apparu indiquant l'absence d'homologues proches ayant des structures résolues. Ainsi, nous avons eu recours à des méthodes de recherche plus élaborées, appelées les méthodes par enfilage ou « threading ». Ces méthodes fonctionnent en utilisant des alignements de famille de protéines et des méthodes statistiques comme les modèles de Markov cachés. Parmi les méthodes de threading testées, on note PSI-BLAST, HMMER et HHpred. Ces recherches conduisent aux patrons potentiels CNTF et LIF, deux cytokines de la famille de l'IL-6.

b) Choix des patrons en utilisant le MDS

La modélisation dans la « Twilight zone » requiert la connaissance d'informations phylogéniques sur la séquence à modéliser pour valider un patron suggéré par les méthodes de threading.

La nouvelle version du package R Bios2mds[27] nous a permis de visualiser l'espace de séquence de la phéromone PRF1 de la salamandre et de détecter des séquences intermédiaires entre CNTF et PRF1, ce qui indique l'origine commune de CNTF et de PRF1. En se basant sur le principe de transitivité de l'homologie, si les protéines A et B sont homologues à une protéine C commune, alors elles sont homologues même si l'identité des séquences A et B est inférieure à la valeur seuil. Les séquences identifiées comme intermédiaires dans l'espace de séquence montrent une identité de séquence supérieure à 20% à la fois pour PRF1 et pour CNTF humain. On peut déduire alors que PRF1 et CNTF sont homologues même si elles présentent une identité de

séquence entre elles inférieure à 20%. La transitivité de l'homologie permet ainsi la modélisation de la protéine PRF1 à partir de la structure cristallographique de CNTF en utilisant le logiciel Modeller (Figure 31).

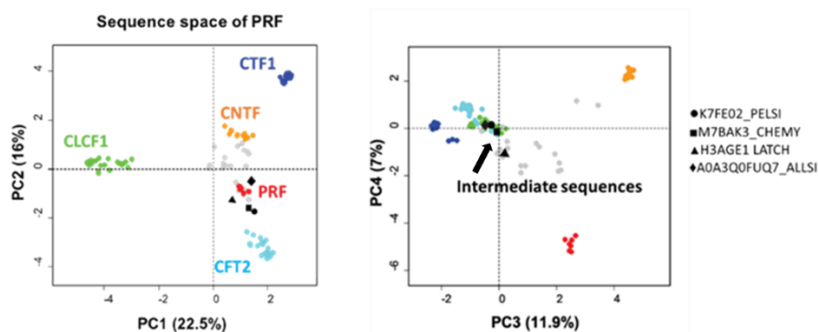


Figure 31 : Informations sur l'évolution de PRF1.

Espace de séquence des homologues de PRF1 avec projection des séquences sur les première et deuxième dimensions (à gauche) et sur les troisième et quatrième dimensions (à droite) de l'analyse MDS.

3.1.3. Les outils de modélisation moléculaire

Il existe plusieurs outils en ligne gratuits et conviviaux pour réaliser des modèles protéiques de façon automatique soit par homologie directe, soit par les méthodes de threading. Cependant, pour avoir plus de contrôle sur le modèle choisi, il vaut mieux utiliser le logiciel Modeller [120] qui construit des modèles moléculaires de la requête à partir d'un ou plusieurs patrons. L'utilisateur peut ainsi personnaliser son modèle afin de répondre aux exigences structurales et fonctionnelles, en rajoutant des contraintes (par exemple angles dièdres ou distances définis par l'utilisateur) ou combiner des fragments de structures patrons pour réaliser un modèle hybride.

Dans le deuxième chapitre de modélisation (annexe II.2) nous avons comparé les modèles de PRF1 réalisés avec des outils automatiques (Robetta avec l'option RosettaCM, Phyre2 et I-TASSER) basés sur des méthodes de threading, avec le modèle personnalisé obtenu avec Modeller.

Aucune différence significative n'a été observée pour les quatre hélices, mais les boucles entre les hélices AB sont modélisées différemment. Parmi les modèles automatisés, seul le modèle Phyre2 a prédit une position de Pro-61 similaire à celle du modèle personnalisé. En revanche, seul RosettaCM a prédit un élément hélicoïdal à l'extrémité C-terminale de la boucle AB. Cependant, l'orientation exposée de Phe-70, similaire à l'orientation de Trp-64 dans CNTF, a été observée seulement dans le modèle personnalisé (Figure 32).

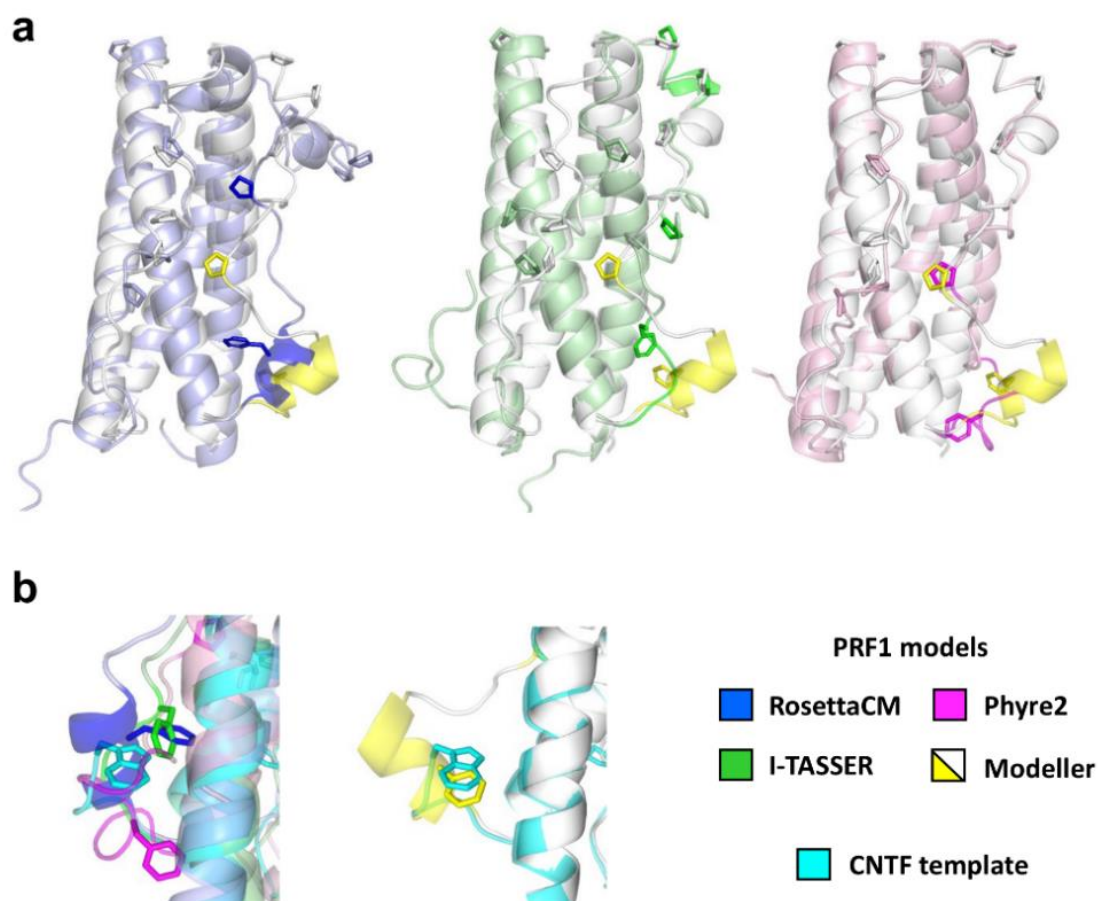


Figure 32 : Comparaison des modèles de PRF1.

Dans (a), le modèle Modeller personnalisé (blanc et jaune) est superposé aux meilleurs modèles obtenus avec RosettaCM (bleu), ITASSER (vert) et Phyre2 (rose). Pro-61 et Phe-70 sont indiquées par des bâtonnets. Le ruban jaune dans le modèle Modeller indique les contraintes hélicoïdales ajoutées par l'utilisateur. Dans (b), le zoom sur la partie C-terminale de la boucle AB révèle que la superposition des cycles phényles du Trp-64 du CNTF (cyan) et de la Phe-70 de PRF1 (jaune) dans le modèle Modeller personnalisé n'est pas observée dans les trois autres modèles.

3.2. Les simulations dynamiques moléculaires classiques

3.2.1. Définition et applications

Les simulations de dynamique moléculaire (MD) sont une méthode computationnelle qui utilise les lois de Newton pour évaluer les mouvements des atomes dans des systèmes complexes comme une macromolécule dans un solvant [123]. Ces simulations peuvent décrire une grande variété de processus biomoléculaires importants, y compris les changements de conformation d'une protéine, la liaison des ligands et l'effet des perturbations moléculaires telles qu'une mutation, la phosphorylation, la protonation. Ces simulations peuvent également prédire comment les biomolécules réagiront aux changements environnementaux tels que l'effet de différentes compositions lipidiques ou du stress mécanique, ainsi que la fixation et la dissociation des ions (Figure 33).

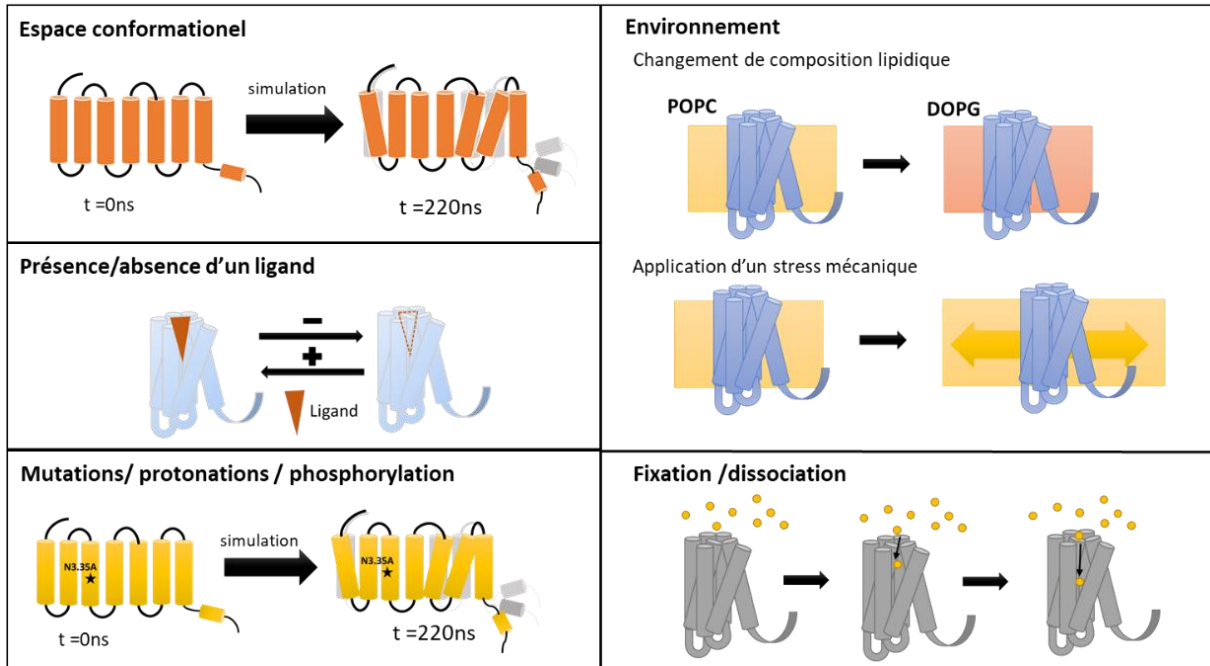


Figure 33 : Exemples d'application de la dynamique moléculaire

3.2.2. Principe et algorithmes

Les simulations dynamiques moléculaires consistent à prédire comment chaque atome d'une protéine ou d'un autre système moléculaire se déplacera dans le temps en appliquant les lois de la mécanique classique newtonienne. En effet selon la deuxième loi de Newton, chaque un atome i de coordonnées r_i ayant une masse m_i subit l'accélération a_i reliée à la force F_i .

$$F_i = m_i a_i \quad (1)$$

Etant donné que

$$a_i = \frac{dv_i(t)}{dt} \quad (2)$$

Et

$$v_i = \frac{dr_i(t)}{dt} \quad (3)$$

la combinaison de ces équations donne

$$F_i = m_i a_i = m_i \frac{dv_i(t)}{dt} = m_i \frac{d^2r_i(t)}{dt^2} \quad (4)$$

L'accélération a dépend de la dérivée de l'énergie potentielle par rapport à la position, r ,

$$a = -\frac{1}{m} \frac{dE(r_1, \dots, r_n)}{dr_i} \quad (5)$$

Par conséquent, pour calculer une trajectoire, il suffit de connaître les positions initiales des atomes, la distribution initiale des vitesses et l'accélération, qui est déterminée par le gradient de la fonction d'énergie potentielle. Les équations du mouvement sont déterministes, c'est-à-dire que les positions et les vitesses au temps zéro déterminent les positions et les vitesses à tous les autres temps, t . Les positions initiales peuvent

être obtenues à partir de structures expérimentales, telles que la structure cristallographique de la protéine ou la structure en solution déterminée par spectroscopie RMN [124].

L'énergie potentielle est une fonction des positions atomiques de tous les atomes du système. En raison de la nature complexe de cette fonction, il n'existe pas de solution analytique aux équations du mouvement ; elles doivent être résolues numériquement.

Les algorithmes d'intégration permettent ainsi de faire évoluer le système d'un petit pas de temps Δt pendant lequel les forces sont considérées comme constantes. Cela permettra de modifier les positions, recalculer les forces et les vitesses et de répéter le processus.

De nombreux algorithmes numériques ont été développés pour intégrer les équations du mouvement : Algorithme de Leap-frog, Vitesse de Verlet, Algorithme de Beeman [124].

Tous les algorithmes d'intégration supposent que les positions, les vitesses et les accélérations peuvent être approximées par une expansion en série de Taylor et le déplacement d'un atome pendant l'intervalle de temps Δt est ainsi donné par l'équation :

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2} \mathbf{a}(t)\Delta t^2 + \frac{1}{6} \frac{d\mathbf{a}(t)}{dt} \Delta t^3 + \sigma(\Delta t^4) \quad (6)$$

Afin de minimiser les erreurs, l'algorithme de Verlet prend en compte le déplacement inverse :

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2} \mathbf{a}(t)\Delta t^2 - \frac{1}{6} \frac{d\mathbf{a}(t)}{dt} \Delta t^3 + \sigma(\Delta t^4) \quad (7)$$

En additionnant ces deux équations, on obtient

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 + \sigma(\Delta t^4) \quad (8)$$

L'algorithme de Verlet utilise les positions et les accélérations au temps t et les positions du temps $(t - \Delta t)$ pour calculer les nouvelles positions au temps $(t + \Delta t)$. Une fois que les positions et les vitesses de chaque atome sont connues, l'état du système peut être prédit à tout moment dans le futur ou dans le passé.

En pratique, l'algorithme le plus utilisé, y compris dans NAMD, est l'algorithme dit « Velocity Verlet ». Il utilise une approche similaire mais incorpore la vitesse. Le calcul s'effectue en 3 étapes :

- Calcul de la position à $t + \Delta t$:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2} \mathbf{a}(t)\Delta t^2 \quad (9)$$

- Calcul de l'accélération à $t + \Delta t$ à partir du potentiel à la position $\mathbf{r}(t + \Delta t)$:
- Calcul de la vitesse à $t + \Delta t$:

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{a}(t) + \mathbf{a}(t + \Delta t)}{2} \Delta t \quad (10)$$

Cette procédure est répétée à des intervalles de temps discrets jusqu'à l'identification de toute la trajectoire. L'énergie totale du système doit être conservée durant toute la simulation.

Les simulations de dynamique moléculaire peuvent prendre du temps et coûter cher en termes de temps de calcul. Cependant, avec le développement d'ordinateurs de plus en plus performants, les simulations des protéines peuvent aller jusqu'à la milliseconde [124].

3.2.3. Les champs de force

Un champ de force est une expression mathématique décrivant la dépendance de l'énergie d'un système par rapport aux coordonnées de ses particules [125]. Il consiste en une forme analytique de l'énergie potentielle interatomique, $U(r_1, r_2, \dots, r_n)$, et en un ensemble de paramètres entrant dans cette forme. Les paramètres sont généralement obtenus soit par des calculs de mécanique quantique, soit par ajustement à des données expérimentales. Il existe de nombreux champs de force disponibles dans la littérature (CHARMM, AMBER, GROMOS...etc), ayant différents degrés de complexité, et orientés pour traiter différents types de systèmes [125].

L'expression typique d'un champ de force $U(R)$, décrit dans la figure 34, regroupe l'énergie entre atomes liés comportant l'énergie d'étirement des liaisons, de flexion des angles et de torsion des angles dièdres et impropres, et l'énergie d'atomes non liés résultant des interactions électrostatiques et de Van der Waals.

$$U(R) = \sum_{\text{liaison}} k_r (r - r_{eq})^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dièdre}} k_\phi (1 + \cos[n_\phi - \gamma]) + \sum_{\text{impropres}} k_\omega (\omega - \omega_{eq})^2 + \sum_{i < j}^{\text{atomes}} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^2 - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] + \sum_{i < j}^{\text{atomes}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij} \epsilon} \quad (11)$$

Energie potentielle = Energie d'atomes liés + Energie d'atomes non liés

Figure 34 : Equations et représentation schématique des énergies liées et non liées.

L'énergie potentielle regroupe les énergies liées (comportant les énergies de liaison entre 2 atomes, les énergies de flexion entre 3 atomes et les énergies de torsion entre 4 atomes) et les énergies non-liées (comportant les énergies électrostatiques et de van der Waals)

3.2.4. Les conditions limites périodiques

Les conditions limites périodiques regroupent un ensemble de conditions aux limites utilisées afin de simuler un système et faire en sorte que le système ressemble à un système infini. La boîte de simulation est donc entourée de multiples répliques d'elle-même, comme le montre la figure 35. Seuls les atomes à l'intérieur de la cellule principale sont considérés explicitement, mais dès qu'un des atomes quitte la cellule, une particule image entre par le côté opposé pour le remplacer [125].

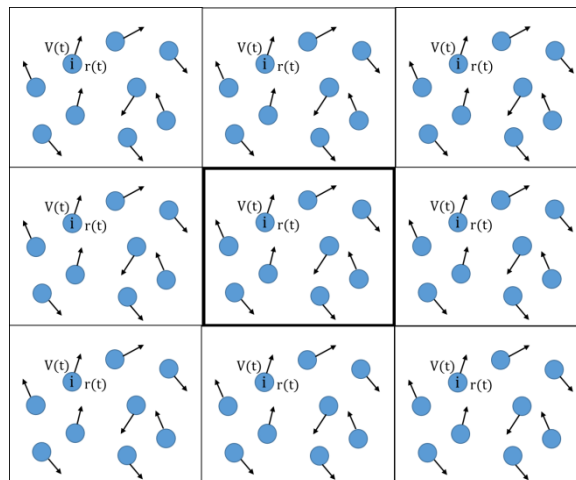


Figure 35 : Conditions limites périodiques.

Illustration en 2D de la boîte de simulation mise en évidence au centre et entourée d'images périodiques d'elle-même

3.2.5. Les ensembles thermodynamiques

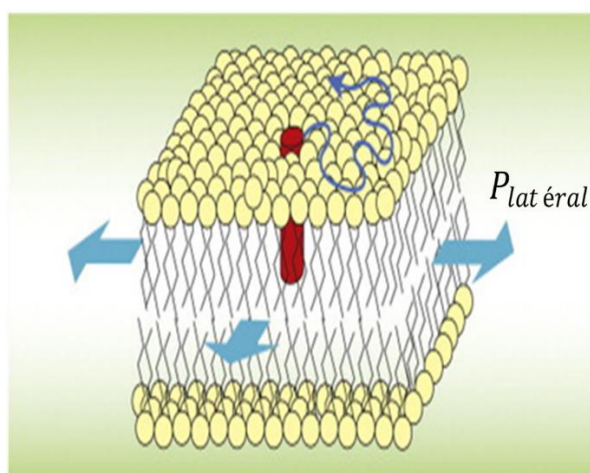
Les ensembles thermodynamiques regroupent un ensemble de grandeurs thermodynamiques dont le nombre de particules N , la température T , la pression P , le volume V , ou l'Énergie totale E . Ainsi, les ensembles thermodynamiques fondamentaux les plus utilisées sont l'ensemble microcanonique NVE, l'ensemble canonique NVT et l'ensemble isotherme isobare NPT. Ces ensembles sont utilisés en dynamique moléculaire afin de maintenir la température et la pression du système constantes, dans le but de mimer les conditions expérimentales.

Pour les membranes, les propriétés thermodynamiques de la pression du système peuvent être différentes entre la direction latérale et la direction normale. Pour contrôler la pression dans ces directions, l'ensemble NPyT doit être utilisé [126].

Deux ensembles thermodynamiques, **NPT** et **NPyT**, ont été utilisés dans ce travail.

L'ensemble NPT maintient constants le nombre de particules N , la pression P et la température T . Cet ensemble est utilisé fréquemment pour décrire un grand nombre de systèmes, y compris l'environnement des cellules biologiques. Il est important pour obtenir des états d'équilibre dans les simulations biomoléculaires. Il est aussi important pour mimer des conditions expérimentales car les réactions chimiques sont généralement conduites sous des conditions de pression constante (exemple d'un tube sous une pression atmosphérique).

L'ensemble NPyT permet de simuler un stress mécanique subi par la membrane. Il maintient constant le nombre de particules N , la pression normale P , la tension superficielle latérale γ des membranes et la température T . Il est utilisé pour caractériser les propriétés physicochimiques des lipides [127] ou évaluer le comportement membranaire sous un stress mécanique, par exemple dans le cas de simulations d'étirement et de compression membranaire suite à un stress mécanique [63]. Cet ensemble permet de contrôler la direction de la pression latérale selon les axes x et y couplée à une tension de surface γ de la membrane (Figure 36). Nous utilisons la méthode développée par K. Schulten (Urbana) qui a ajouté une tension de surface γ à l'équation du mouvement de Newton (simulations en conditions NPyT) dans le logiciel NAMD [128]. Des valeurs de γ positives correspondent à une pression latérale négative mimant un effet d'étirement, alors que des valeurs γ négatives correspondent à une pression latérale positive mimant un effet de compression.



$$\gamma = \int_{-h/2}^{h/2} dz [P_{zz} - 1/2(P_{xx} + P_{yy})] \quad (12)$$

$$P_{lat\ éral} = \frac{P_{xx}(z) + P_{yy}(z)}{2} - P_{zz} \quad (13)$$

Figure 36 : Modélisation du stress mécanique par une tension de surface.

3.2.6. Outils des simulations dynamiques

Pour réaliser des études de dynamique moléculaire, il existe de nombreux logiciels fiables pour effectuer les simulations (par exemple NAMD, Amber, Gromacs, etc.) et pour visualiser et analyser leurs résultats (VMD, Chimera, Pymol etc.) [129]–[132]

3.3. Les simulations dynamiques moléculaires dirigées

3.3.1. Principes

Les simulations dynamiques dirigées ou « steered molecular dynamics » (SMD) sont des outils de simulations hors équilibre permettant d'explorer un processus ou un espace conformationnel biologique rare ou inaccessible par des simulations dynamiques classiques. Cela peut permettre d'étudier la relation structure-fonction d'un complexe macromoléculaire impliquant des interactions protéine-ligand ou protéine-protéine. Cette méthode peut être aussi utilisée comme précurseur pour le calcul de l'énergie libre de transformations conformationnelles [133]

Le principe des SMD est d'appliquer une force externe uniquement à un atome ou groupe d'atomes que l'on veut déplacer afin d'observer le comportement de la protéine dans ces conditions [134]. Pour exemple, si l'on veut déplacer un groupe d'atomes par SMD, le centre de masse de ceux-ci sera lié de manière harmonique avec une constante de force k à un atome fictif contraint de se déplacer avec une vitesse v dans la direction \vec{n} . La méthode SMD entraîne donc l'application à un instant t d'un potentiel au système suivant l'équation suivante :

$$U(r_1^{\rightarrow}, r_2^{\rightarrow} \dots, t) = \frac{1}{2} k [vt - (\vec{R}(t) - \vec{R}_0) \cdot \vec{n}]^2 \quad (14)$$

où $\vec{R}(t)$ est le centre de masse à l'instant t des atomes à déplacer et \vec{R}_0 est le centre de masse initial. Le vecteur \vec{n} est normalisé par NAMD avant d'être utilisé.

3.3.2. Quaternion d'orientation

Dans le cas des protéines transmembranaires (TM), les orientations et les directions des hélices TM sont très importantes pour décrire les changements conformationnels de la protéine. Par exemple l'orientation de l'hélice TM6 des RCPG détermine l'état d'activation des récepteurs. De ce fait, les orientations des hélices TM se sont avérées être des variables collectives efficaces décrivant les changements de conformation des protéines membranaires. L'orientation des hélices peut être décrite en utilisant le formalisme du quaternion Q . Le quaternion est une notation mathématique introduite par le mathématicien William Rowan Hamilton en 1843. Le quaternion décrit les orientations et les rotations spatiales des éléments en 3 dimensions :

$$Q = (\cos \frac{\theta}{2}, \sin \frac{\theta}{2} \vec{u}) \quad (15)$$

avec θ et \vec{u} (le vecteur unitaire) décrivant l'angle et l'axe de rotation. Le formalisme du quaternion peut être utilisé comme variable collective efficace pour introduire un changement rotationnel dans un domaine protéique. Moradi *et al.* ont introduit cette variable dans les simulations dynamiques dirigées de NAMD afin

d'étudier les transitions conformationnelles d'un transporteur membranaire MsbA (membre de la famille des exportateurs de cassettes de liaison à l'ATP) [135]. Ce transporteur ayant un large espace conformationnel fonctionnel, il est difficile et coûteux d'étudier toutes les transitions possibles par simulations dynamiques classiques. Ils ont alors conçu plusieurs variables collectives basées sur le formalisme du "quaternion" comme méthode efficace pour induire des transformations rotationnelles dans un système moléculaire. En réalisant 200 simulations extensives, ils ont pu estimer les énergies libres associées à l'espace conformationnel et déterminer les chemins favorisés énergétiquement à partir du travail fourni pour chaque transition dans les simulations hors équilibre.

Matériels & Méthodes

4. Matériel

Les étapes d'équilibrage et d'analyse des trajectoires ont été réalisées localement sur une station de travail DELL avec le système d'exploitation Ubuntu 16.04.4 LTS version 64bits (processeur Intel® Core™ i7-7700 CPU Quadro K620/PCIe/SSE2).

La phase de production des trajectoires qui nécessite une capacité de calcul très importante a été réalisée sur le serveur national de l'IDRIS (Institut du développement et des ressources en informatique scientifique) en utilisant les serveurs Ada (en 2018-2019) puis Jean-Zay (de 2019 à 2022).

Le logiciel utilisé pour les simulations dynamiques est NAMD version 2.9 sur Ada et version 2.13 sur Jean-Zay.

4.1. Migration vers le supercalculateur Jean-Zay et optimisation du temps de calcul

Au cours du 2^{ème} semestre de 2019, le nouveau supercalculateur Jean-Zay a été installé à l'IDRIS. Nos simulations ont migré vers ce supercalculateur, tout en utilisant la nouvelle version de NAMD2.13. De nouveaux scripts de lancement de travaux ont été créés. L'architecture de ce nouveau supercalculateur étant différente du précédent, nous avons dû réaliser des tests pour définir les conditions pour un usage optimal du temps de calcul alloué (5 noeuds avec 40 tâches). Pour une simulation donnée, cette configuration entraîne un faible gain de temps physique (4%) mais un gain marqué dans le temps alloué (25%) par rapport au serveur Ada (Figure 37).

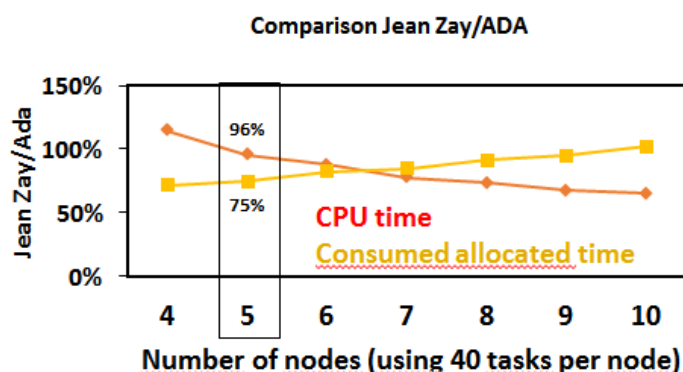


Figure 37 : Evolution du temps de calcul physique et alloué en fonction du nombre de nœuds.

5. Méthodes

5.1. Modélisation moléculaire des récepteurs

La modélisation moléculaire des récepteurs étudiés a été réalisée par homologie à l'aide du logiciel de modélisation moléculaire MODELLER [136].

5.1.1. Modélisation du récepteur AT1

AT1 humain avec l'ion sodium :

La structure inactive d'AT1 humain lié à l'ion sodium a été modélisée de la position 17 à 334 à partir des structures cristallographiques d'AT1 sans sodium 4YAY [89] et 4ZUD [85] et un fragment de la structure du

récepteur opiacé δ , OPRD (4N6H) [49], correspondant au site de fixation du sodium (TM2, TM3, TM6 et TM7), et comprenant l'ion sodium et les molécules d'eau de coordination.

L'hélice H8 C-terminale a été modélisée par homologie avec la structure 4N6H d'OPRD car dans les structures 4YAY et 4ZUD, H8 est soit manquante (4ZUD : résidus 289-301 manquants), soit inclinée par rapport au plan de la membrane (4YAY). Le récepteur OPRD a été choisi car il est phylogénétiquement proche d'AT1.

AT1 humain sans l'ion sodium :

Le modèle d'AT1 sans sodium a été obtenu en remplaçant l'ion sodium du modèle précédent par une molécule d'eau.

AT1 d'anguille : Le modèle d'AT1 d'anguille (*Anguilla japonica*) avec l'ion sodium lié a été construit par homologie à partir de la structure d'AT1 humain.

5.1.2. Modélisation du récepteur AT2

Pour AT2, seule la structure de la forme active a été résolue. Nous avons donc réalisé 2 modèles du récepteur AT2 de la position 35 à 350 à l'état inactif, lié ou non à l'ion sodium, à partir de la structure active d'AT2 5UNG [97] et du modèle d'AT1 humain inactif. Les moitiés intracellulaires de TM5, TM6 et TM7 et la partie C-terminale H8 de la structure 5UNG ont été tronquées et remplacées par les parties équivalentes du modèle d'AT1 humain, y compris l'ion sodium et les molécules d'eau. Les deux modèles diffèrent par la longueur de TM6.

Dans le modèle A, TM6 a la même longueur que dans AT1. Ce modèle a été utilisé dans les simulations dynamiques dirigées, vu sa stabilité dans des conditions hors équilibre.

Dans le modèle B, des contraintes α -hélicoïdales ont été ajoutées à la procédure de modélisation pour s'assurer que la longueur de TM6 était identique à la structure 5UNG. Ce modèle a été utilisé dans les simulations dynamiques classiques.

5.2. Modélisation du stress mécanique

5.2.1. Modélisation du stress mécanique en tension de surface

Trois types de conditions externes peuvent entraîner une contrainte mécanique sur une bicouche lipidique : les forces de cisaillement, la pression hydrostatique et la pression osmotique. Ces conditions externes conduisent à une modification de la tension de surface de la membrane [93].

La tension de surface γ d'une membrane correspond à une pression latérale qui pourra être modélisée en dynamique moléculaire en réalisant des simulations en conditions NPT avec NAMD (méthode décrite dans la synthèse bibliographique au paragraphe 3.2.5). On maintient la pression normale à la bicouche (P_{zz}) à 1 atm et la tension de surface constante aux valeurs désirées.

5.2.2. Suivi des paramètres physiques et calcul de la tension de surface

Après les générations des trajectoires sous une tension de surface γ , il est indispensable de suivre les paramètres physiques suivants (Figure 38):

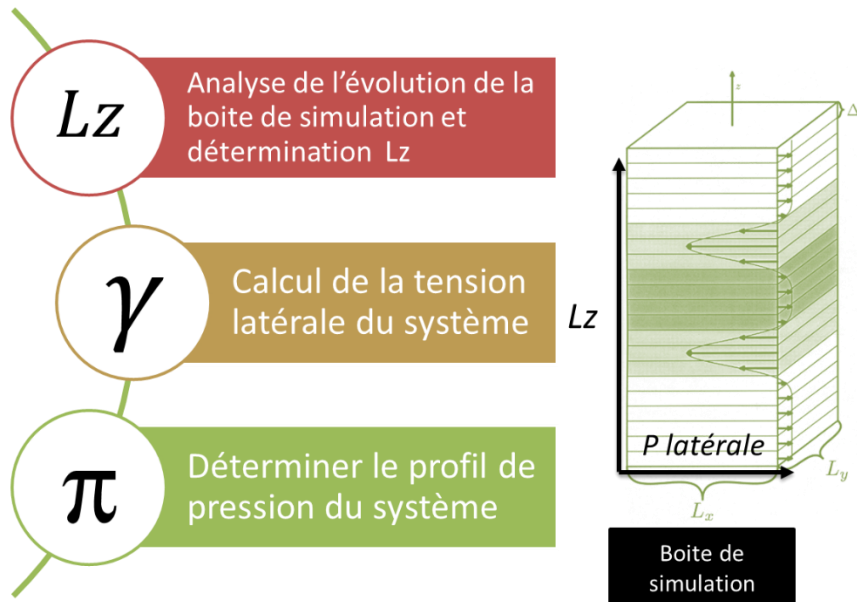
- l'évolution des dimensions de la boîte de simulation au cours de la trajectoire et la détermination de la hauteur de la boîte L_z (dimension de la boîte sur l'axe z , perpendiculaire au plan de la membrane)
- la valeur de la tension latérale afin de vérifier si la tension émise est égale à la tension reçue par le système

- le profil de la pression latérale locale à l'intérieur de la membrane dans la direction normale à la surface de la membrane.

Pour calculer le profil de pression latérale, la bicouche lipidique est divisée en 30 tranches. Le profil de pression latérale $n(z)$ est la différence entre le tenseur de pression normale et le tenseur de pression latérale.

Paramètres physiques

$$\gamma = Lz[P_{zz} - \frac{1}{2}(P_{xx} + P_{yy})] \quad (12)$$



$$Lateral\ pressure = [P_{zz} - 1/2(P_{xx} + P_{yy})] \quad (13)$$

Figure 38 : Suivi des paramètres physiques et calcul de la tension de surface.

5.3. Les simulations de dynamique moléculaire

5.3.1. Les simulations dynamiques classiques

A partir des modèles des récepteurs AT1 et AT2 avec et sans sodium, nous construisons le système qui contient plus de 60000 atomes à l'aide de la plateforme CHARMM-GUI [137]. Elle permet d'orienter le récepteur, de l'insérer dans une bicouche lipidique homogène de 120 molécules de POPC ou DOPG (60 molécules par feuillet), de solvater et de neutraliser électrostatiquement le système en ajoutant de l'eau de type TIP3 et du KCl à une concentration de 150 mM. Le champ de force CHARMM36 est utilisé pour paramétrer le système. Le système passe par une étape de minimisation d'énergie suivie par une étape d'équilibration d'une durée d'une nanoseconde au cours de laquelle il y a un relâchement progressif des contraintes tout en portant le système à une température de 310K. Les trajectoires sont générées en utilisant NAMD (versions 2.9 et 2.13) à des valeurs de tension différentes en utilisant le superordinateur IDRIS. La température est maintenue constante par la méthode de Langevin. La pression est maintenue constante par la méthode dite de « Nosé-Hoover Langevin piston pressure control ». Il s'agit d'une méthode de Nosé-Hoover modifiée dans laquelle la dynamique de Langevin est utilisée pour contrôler le barostat. Sauf indication contraires, les trajectoires sont d'une durée de 220 ns (Figure 39). Les trajectoires ayant pour but de décrire les propriétés des lipides ont été réalisées une fois par condition. Les trajectoires avec les récepteurs ont été répliquées trois fois par condition.

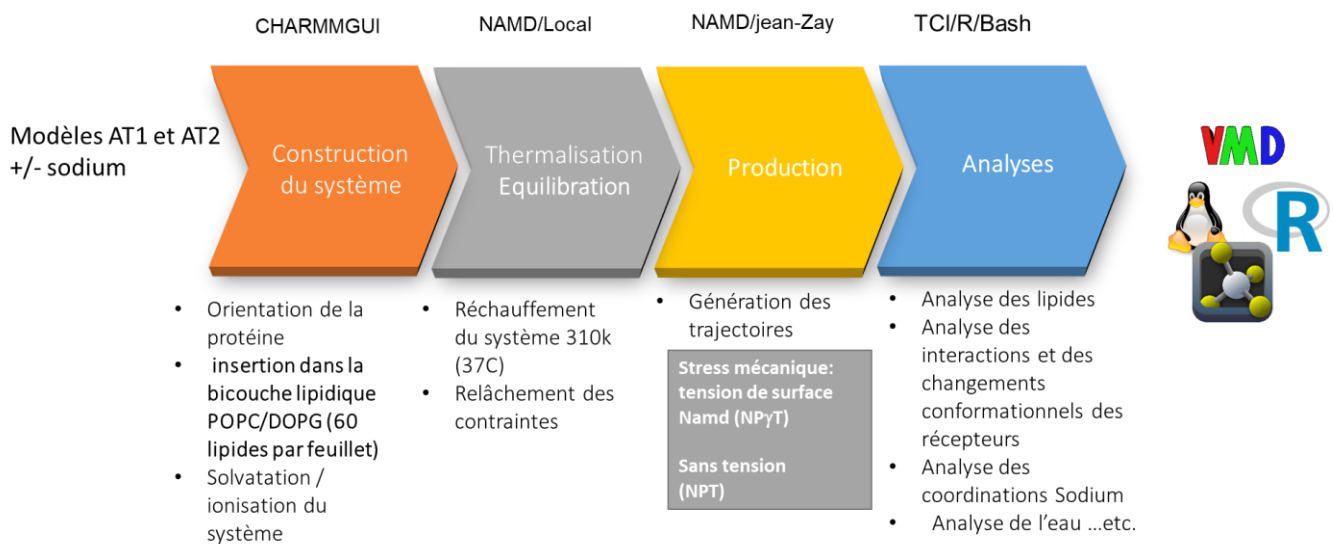


Figure 39 : Etapes des simulations de dynamique moléculaire.

5.3.2. Les simulations dynamiques dirigées (SMD)

Les simulations dynamiques dirigées permettent d'explorer un processus ou un espace conformationnel biologique rare ou inaccessible par des simulations dynamiques classiques. Le principe des SMD est d'appliquer une force externe à un atome ou groupe d'atomes afin de faciliter les transitions conformationnelles et ainsi d'observer le comportement de la protéine dans différentes conditions.

a) Mise au point des simulations dynamiques dirigées par quaternion

Pour mieux comparer les différentes conditions (nature des lipides, présence de ligands, et stress mécanique) et dévoiler le mécanisme d'activation d'AT1 et AT2, nous avons initié la mise en place des simulations de dynamique moléculaire dirigée ou « Steered Molecular dynamics » (SMD). Dans cette méthode, une force appliquée à une variable collective permet de franchir la barrière de potentiel entre deux conformations et obtenir un changement conformationnel.

La principale caractéristique de l'activation d'un RCPG est le mouvement vers l'extérieur de TM6 qui ouvre la cavité intracellulaire pour faciliter le couplage avec la protéine G.

Dans notre cas, nous avons choisi d'appliquer une force de traction externe aux résidus N-terminaux de l'hélice 6 pour induire leur déplacement vers une structure active (mouvement vers l'extérieur). Pour utiliser cette méthode, il faut bien décrire le mouvement de l'hélice 6 en 3 dimensions.

Dans ce but, nous avons développé un modèle de quaternion pour l'ouverture de TM6 en s'inspirant des travaux de Moradi *et al.* [133], [135]. Dans ce modèle, la rotation d'un vecteur (axe de l'hélice N-terminale de TM6) est décrite par l'axe de rotation \hat{u} (u_1, u_2, u_3), obtenu à partir du produit vectoriel des axes \hat{i} (TM6 inactive) et \hat{j} (TM6 active) avant et après la rotation, et par l'angle de rotation θ autour de l'axe \hat{u} . Ces paramètres sont condensés en un quaternion Q (Figure 40):

$$Q = (\cos\theta/2, \sin\theta/2 \hat{u}) = (\cos \theta/2, \sin \theta/2 u_1, \sin \theta/2 u_2, \sin \theta/2 u_3) \quad (16)$$

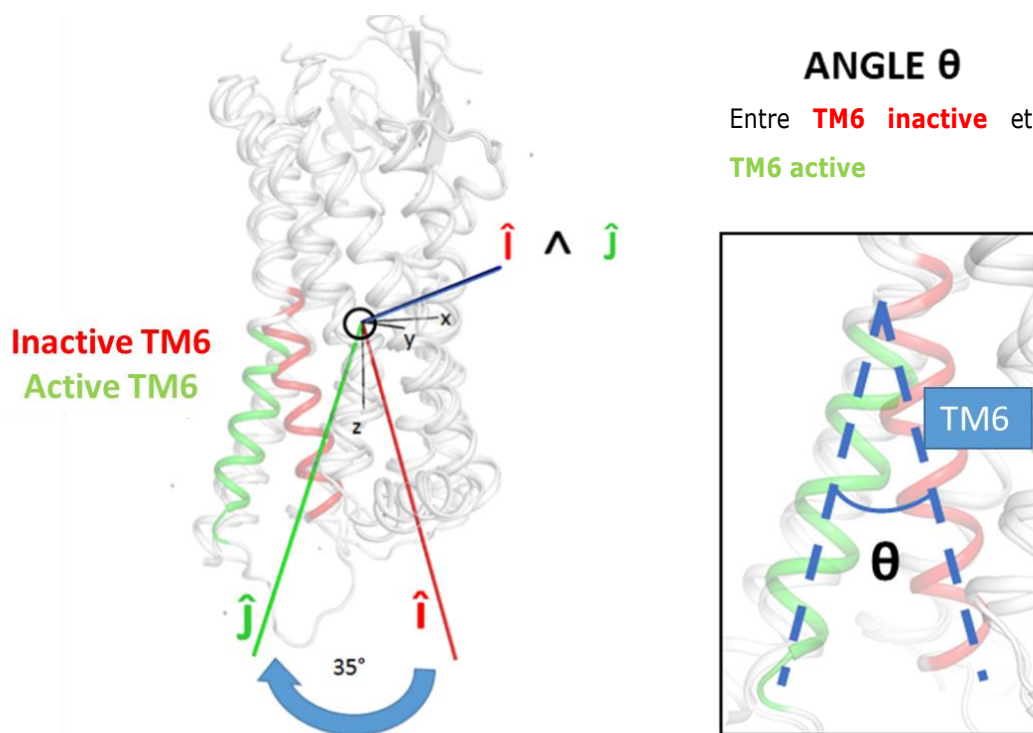


Figure 40 : Description 3D de l'ouverture de l'hélice TM6 sous forme de quaternion.

Le quaternion représentant le mouvement de la partie N-terminale de l'hélice 6 est défini par l'axe de rotation \hat{u} obtenu à partir du produit vectoriel des axes \hat{i} et \hat{j} avant et après la rotation et par l'angle de rotation θ autour de l'axe de rotation.

b) Etape des SMD

Les étapes des simulations dirigées sont les suivantes (Figure 41a) :

- 1- A partir d'une simulation dynamique classique (cMD) d'AT1 ou AT2 réalisée sous diverses conditions (POPC/DOPG), en présence de l'ion sodium lié au récepteur, une structure initiale équilibrée est choisie pour servir comme base de départ pour les SMD.
- 2- Comme deuxième étape, il faut choisir les paramètres indispensables pour le fonctionnement des SMD tels que les coordonnées des résidus de TM6 qui vont subir la force de traction, les valeurs initiales et finales du quaternion, la force appliquée ainsi que le nombre d'étapes au cours desquels TM6 est soumise à la force K. Dans notre cas, la force K de traction est appliquée pendant 20 ns de SMD suivies par 10 ns de simulations classiques sans application de force pour rééquilibrer le système. Au cours de cette deuxième étape, le stress mécanique (tension de surface 20 dyne/cm) peut être rajouté sur l'ensemble de la simulation (SMD puis cMD) (Figure 41b).
- 3- Les trajectoires sont générées en utilisant le logiciel NAMD avec le supercalculateur Jean Zay.
- 4- L'analyse des trajectoires se fait essentiellement en calculant l'évolution de l'angle θ au cours du temps pour juger l'état d'activation du récepteur ainsi que l'évolution des distances TM3-TM6 et TM3-TM7 pour décrire les mouvements respectifs des hélices TM6 et TM7.

Nos tests préliminaires montrent que, au-delà de la force de 1000 kcal/mol/rad², nous observons une ouverture complète de l'hélice 6 ($\theta > 24^\circ$) vers l'état actif avec des conditions optimales à 5000 kcal/mol/rad² (Figure 41b).

Utiliser cette méthode dans des conditions optimales et stables nous permettra d'étudier les étapes mécanistiques d'activation d'AT1 et AT2 et observer les mouvements de TM7.

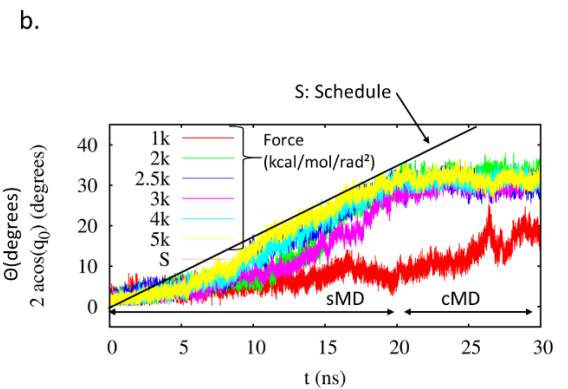
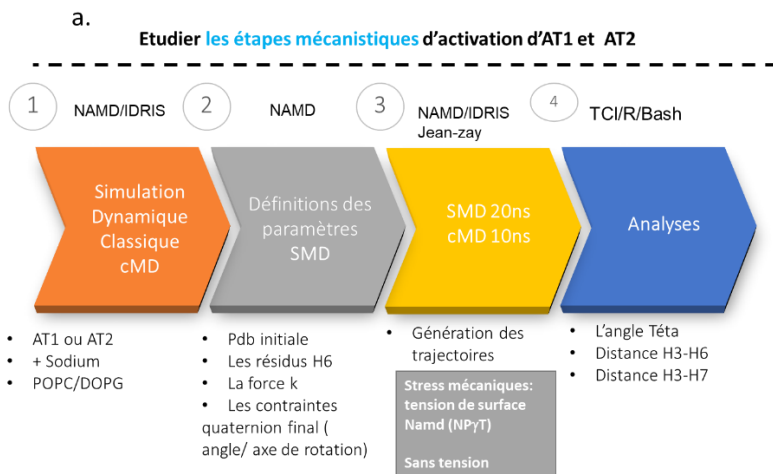


Figure 41 : Simulations dynamiques dirigées.

(a) Les différentes étapes des simulations dynamiques dirigées, (b) Exemple représentatif de l'évolution de l'angle θ au cours de temps en fonction de la force de traction appliquée.

5.4. Analyses des simulations dynamiques moléculaires

5.4.1. Visualisation et analyses des trajectoires

Des scripts R basés sur bio3D [138] et des scripts tcl/tk utilisant les fonctionnalités du programme VMD [132] ont été développés pour analyser systématiquement dans toutes les trajectoires :

- Les RMSD (root mean square deviations) par rapport à la structure initiale
- Les RMSF (root mean square fluctuations)
- Les distances du sodium aux atomes d'oxygène présents au niveau du site sodium
- Certains angles dièdres.

D'autres scripts R ont été développés pour afficher certains de ces résultats en forme de heatmaps par trajectoire en utilisant le package R ComplexHeatmaps [139].

La visualisation des trajectoires est faite avec les logiciels VMD et Pymol [140].

5.4.2. Analyse des changements conformationnels

Afin de mieux décrire les mouvements de TM6 et TM7, nous avons mesuré les distances TM3-TM6, en prenant comme référence les distances entre les atomes CA des résidus 3.50 et 6.34, et TM3-TM7, en prenant comme référence les distances entre les atomes CA des résidus 3.50 et 7.55. Le mouvement en avant de l'hélice 6 est traduit par une augmentation de la distance entre TM3-TM6. De la même façon, le mouvement en arrière ou en avant de l'hélice 7 est traduit par, respectivement, la diminution ou l'augmentation de la distance entre TM3 et TM7 (Figure 42). Ces distances sont ensuite représentées sous forme de heatmaps ou sous forme de cartes 2D, générés dans les deux cas par des scripts R.

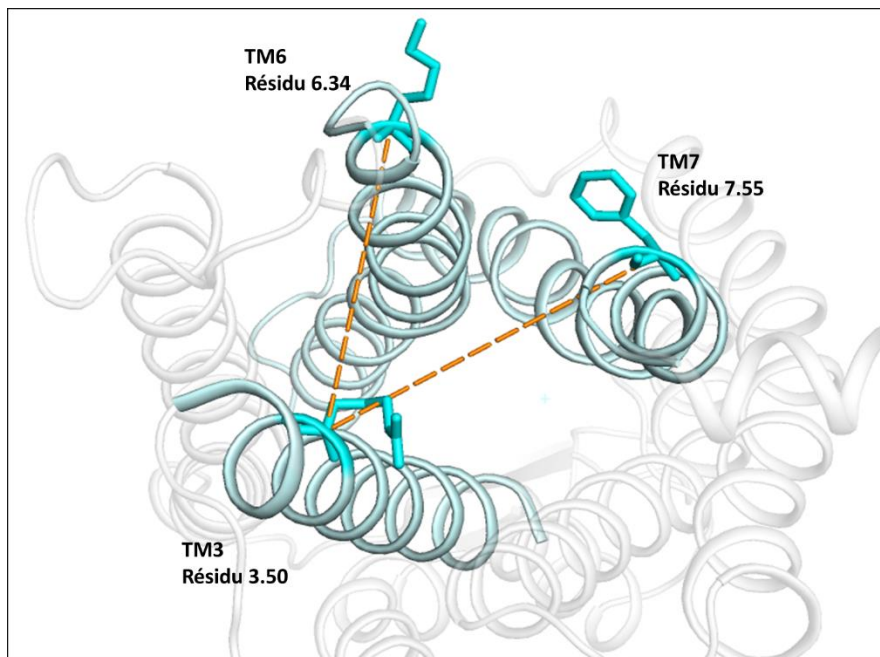


Figure 42 : Distances entre TM3 et TM6 et entre TM3 et TM7.

Les distances sont mesurées entre les atomes CA des résidus 3.50 et 6.34 et des résidus 3.50 et 7.55.

5.4.3. Analyse de la coordination de l'ion sodium

Il s'agit de mesurer les distances entre l'ion sodium et d'autres atomes d'intérêt tels que l'oxygène des molécules d'eau et des résidus protéiques du site sodium. Nous avons défini la première couche de coordination du sodium comme formée par les atomes d'oxygène protéiques et aqueux situés à une distance inférieure ou égale à, respectivement, 2.8 et 3.0 Å.

5.4.4. Quantification des molécules d'eau

Afin d'estimer les modifications de la cavité interne des récepteurs dans différentes conditions, nous avons quantifié le nombre des molécules d'eau dans cette cavité.

Nous avons étudié le nombre total des molécules d'eau distantes de 5 Å des résidus 3.32, 7.39, 6.44, 7.45, L2.43 et L3.46 couvrant le domaine transmembranaire interne du récepteur.

Pour faire une analyse plus fine, nous avons subdivisé ce domaine en 3 parties définissant 3 zones. Nous avons quantifié le nombre des molécules d'eau à 5 Å des résidus 3.32 et 7.39 pour la zone 1, des résidus 6.44 et 7.45 pour la zone 2 et des résidus 2.43 et 3.46 pour zone 3 (Figure 43).

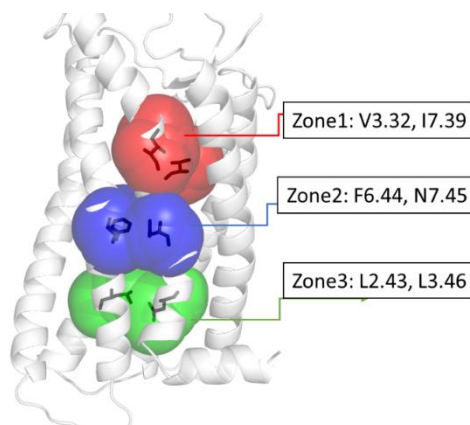


Figure 43 : Zones de quantification des molécules d'eau dans la cavité interne du récepteur AT1.

5.4.5. Quantification des liaisons H

Les liaisons hydrogène ont été mesurés avec la fonctionnalité *Hydrogen Bonds* de VMD, en utilisant un angle de 30° et une distance de 3.5 Å comme limites et en prenant comme sélections les chaînes latérales des résidus protéiques d'intérêt. Les analyses sont faites avec un script tcl regroupant les liaisons d'intérêt à mesurer.

5.4.6. Caractérisation des lipides

Afin de caractériser les bicouches lipidiques et surveiller les transitions de phase, 3 paramètres importants ont été mesurés en utilisant le MEMBRANEPLUGIN de VMD [141] (Figure 44):

- Aire par lipide (**A**) : en appliquant la méthode du diagramme de Veronoi qui se base sur la projection du centre de masse des lipides sur une surface plane
- Epaisseur de la bicouche lipidique (**E**): mesure de la distance entre les 2 têtes des phospholipides des bicouches
- Paramètre d'ordre de la chaîne lipidique (**Scd**) : un paramètre typiquement dérivé des expériences de RMN qui reflète la mobilité orientationnelle de chaque liaison C-H le long des queues aliphatiques des lipides et donc la fluidité de la membrane.

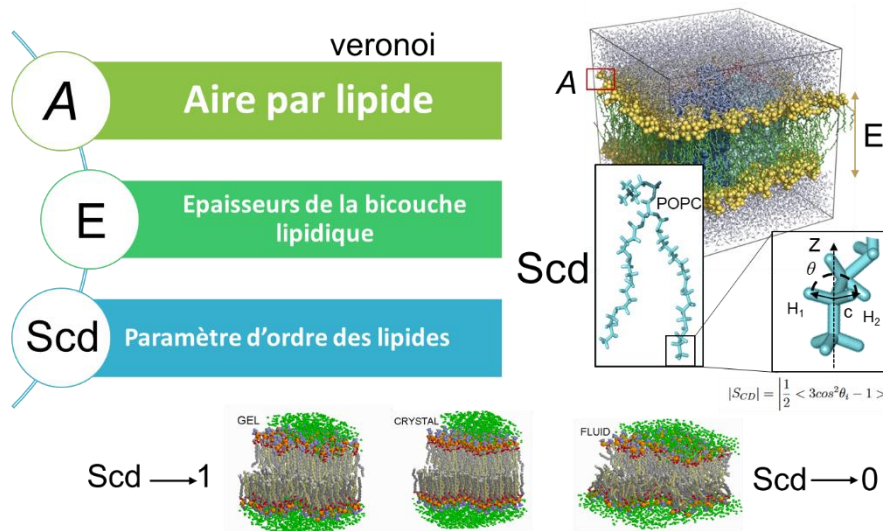


Figure 44 : Paramètres de caractérisation des lipides.

Résultats

6. Effet du stress mécanique sur les paramètres physiques du système de simulation

6.1. Effet du stress mécanique sur les dimensions de la boîte de simulation

Afin de vérifier si la taille des boîtes de simulations est adaptée aux simulations dynamiques sous l'ensemble NPT, nous avons examiné les dimensions des boîtes de simulations sous différentes contraintes.

Cette étude a été réalisée sur un système composé d'une bicouche lipidique hydratée avec insertion ou non du récepteur AT1. Les contraintes mécaniques appliquées induisent un changement radical des dimensions de la boîte de simulation, que ce soit en présence ou en absence du récepteur. En appliquant des tensions positives, on observe un effet d'étirement du système alors qu'en appliquant des tensions négatives, nous observons un effet de compression qui est plus visible à la tension -100 dyne/cm (Figure 45a). Afin de quantifier cet effet, nous avons mesuré les dimensions x , y et z de la boîte de simulation en présence du récepteur (POPC + récepteur) et en absence du récepteur (POPC seul) sous les différentes tensions. Nous remarquons une diminution de la dimension Z de 60% en allant de $\gamma = -100$ dyne/cm vers $\gamma = 50$ dyne/cm (Figure 45b). Néanmoins, le volume de la boîte change beaucoup moins ($< 1\%$) car les variations de z sont compensées par les variations en x et y .

Lorsque nous comparons les dimensions des boîtes de simulation en présence et en absence de récepteur, nous notons que l'effet de la tension est moins marqué en présence du récepteur. Ceci semble indiquer que la tension a un effet moins marqué sur le récepteur que sur les lipides, en lien avec la différence de caractéristiques physico-chimiques des protéines et des lipides.

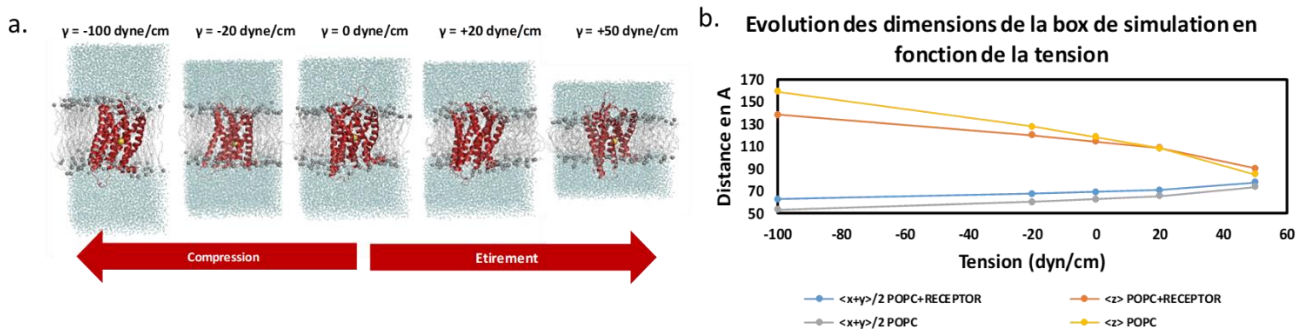


Figure 45 : Evolution des dimensions de la boîte de simulation en fonction de la tension.

Le système est formé d'une bicouche de POPC hydratée avec insertion ou non du récepteur AT1 lié au sodium.

(a) Boîtes de simulations dynamiques lors de l'application de tension variant de -100 à 50 dyne/cm. (b) Courbes d'évolution des dimensions $(x + y/2)$ et z en fonction de la tension.

6.2. Calcul de la tension de surface et de la pression latérale du système

En déterminant L_z (dimension de la boîte sur l'axe Z) ainsi que la pression suivant les différentes coordonnées du système, en particulier P_{xx} , P_{yy} et P_{zz} , nous avons recalculé la tension de surface ainsi que la pression

latérale du système selon les équations décrites dans la Figure 46a. La bicouche lipidique est divisée en 30 tranches. Le profil de pression latérale est calculé dans chaque tranche.

Nous avons bien validé que la tension mesurée est égale à la tension initiale (Figure 46a). Par contre, la pression latérale totale calculée du système ne reflète pas la pression microscopique du système. En effet, le profil de pression fluctue au long de l'axe Z. Cette fluctuation peut être due à la variation de la densité locale ou la variation de la composition de la membrane : Au niveau des molécules d'eau, la pression latérale est nulle alors qu'au contact avec les têtes des phospholipides ou la protéine, le système est plus dense et la pression fluctue à des valeurs positives ou négatives dans différentes positions de la bicouche phospholipidique (Figure 46b). Ce profil est aussi sensible aux valeurs de tension. Plus la tension est élevée, plus le profil de pression est faible.

a. Calcul de la tension du surface et la pression latérale

$$\gamma = \int_{-h/2}^{h/2} dz [P_{zz} - 1/2(P_{xx} + P_{yy})]$$

$$\gamma = Lz [P_{zz} - 1/2(p_{xx} + p_{yy})]$$

↓

La pression latérale = $[P_{zz} - 1/2(P_{xx} + P_{yy})]$

Tension initiale (dyne/cm)	Lz(Å)	tension mesurée γ (dyne/cm)	Pression latérale(bar)
γ_0	115	0	0
γ_{20}	107	20	-18
γ_{50}	89	49	-55

b. Profil de la pression latérale

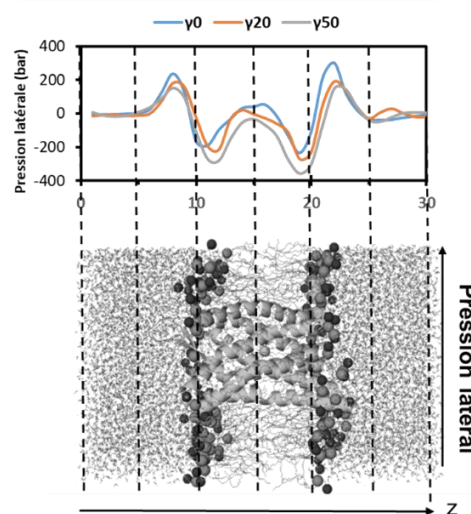


Figure 46 : Calcul de la tension de surface et de la pression latérale du système en fonction de la tension.

Le système est formé du récepteur AT1 lié au sodium et inséré dans une membrane de POPC hydratée.

6.3. Effet du stress mécanique sur les caractéristiques des lipides

Afin de caractériser les bicouches et surveiller les transitions de phase, nous avons réalisé des simulations dynamiques des bicouches lipidiques (POPC/DOPG) en absence du récepteur et avec une épaisseur d'eau de 45Å et nous avons appliqué des tensions de surface négatives et positives pour mimer des effets de la compression et de la dilatation selon le tableau suivant :

Tableau 3 : Paramètres des simulations dynamiques de bicouches lipidiques¹

Lipide	Tension (dyne/cm)				
	-20	Sans tension (NPT)	0 (NPγT)	20	50
POPC	-20	Sans tension (NPT)	0 (NPγT)	20	50
DOPG	-20	Sans tension (NPT)	0 (NPγT)	20	50

¹ Durée des simulations: 220 ns ; 1 réplica.

Trois paramètres importants ont été mesurés en utilisant le MEMBRANEPLUGIN [141] de VMD : l'aire par lipide, l'épaisseur de la membrane et le paramètre d'ordre S_{CD} .

Nous avons constaté que les contraintes mécaniques affectent les propriétés physicochimiques de la bicouche lipidique quelle que soit la nature des lipides : l'étirement conduit à une augmentation de l'aire par lipide (Figure 47a) et une diminution de l'épaisseur de la membrane (Figure 47b) et des paramètres d'ordre des chaînes lipidiques aliphatiques SN1 et SN2 (Figure 47c). Ces résultats indiquent que l'étirement conduit à un état plus fluide des lipides, ce qui facilite les mouvements moléculaires et les interactions protéine-protéine ou lipide-protéine. Nous avons aussi remarqué que l'utilisation des ensembles NPT ou NPγT avec $\gamma=0$ dyn/cm pour représenter des conditions stables sans stress mécanique donne les mêmes résultats au niveau des propriétés physicochimiques des lipides.

En comparant les deux lipides, les effets semblent plus prononcés pour les DOPG que pour les POPC (Figure 47). En effet, l'aire par lipide est supérieure pour les DOPG par rapport aux POPC avec une différence qui varie de 6 Å ($\gamma = -20$) à 12 Å ($\gamma = 50$). L'épaisseur des bicouches lipidique est plus faible chez les DOPG par rapport aux POPC avec une différence de 2 Å. Les paramètres d'ordre des chaînes lipidiques aliphatiques SN1 et SN2 sont plus faible pour les DOPG que pour les POPC.

Par conséquent, la sensibilité des DOPG à l'étirement peut faciliter les mouvements moléculaires et les interactions lipide-protéine par rapport aux POPC.

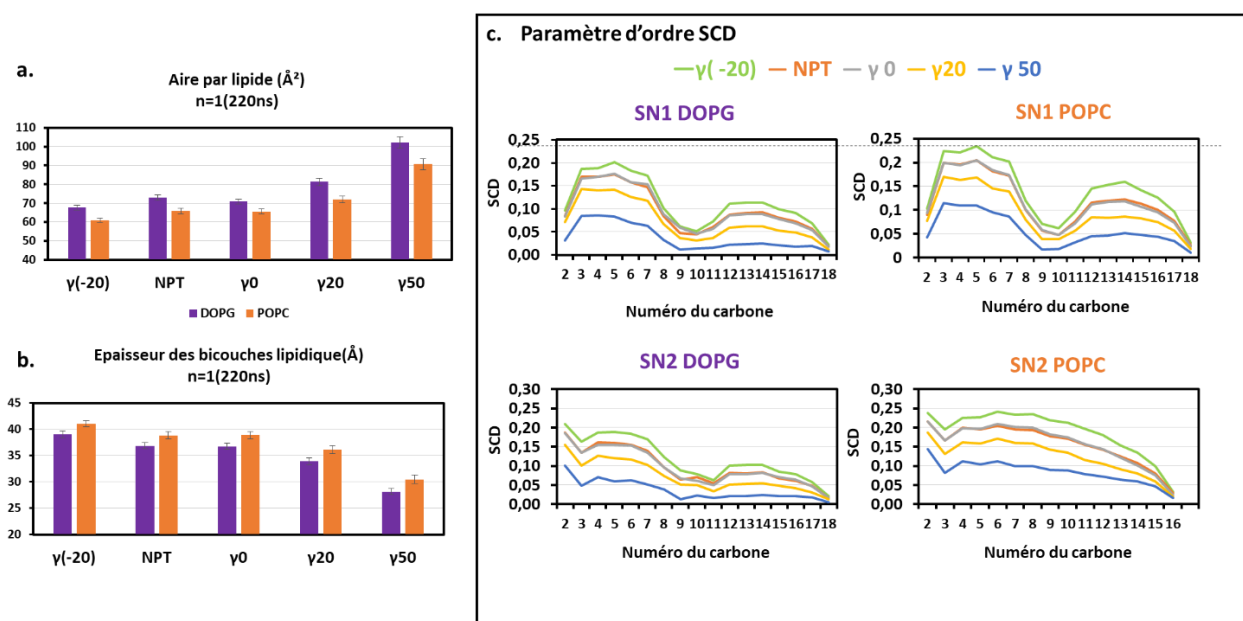


Figure 47 : Comparaison des propriétés physicochimique des POPC et des DOPG.

(a) Aire par lipide ; (b) Epaisseur des lipides ; (c) Paramètres d'ordre des chaînes SN1 et SN2 des DOPG et POPC.

6.4. Distribution des lipides autour du récepteur

Etant donné que les DOPG présentent 3.6% de la composition membranaire mais 16% des lipides liés au récepteur $\beta 2$ adrénergique facilitant ainsi son activation [53], nous avons analysé la distribution des lipides d'une mélange POPC/DOPG autour du récepteur AT1. Afin de mimer des conditions plus réelles de la membrane, nous avons réalisé 3 répliquas du système d'AT1 lié à l'ion sodium dans une bicouche lipidique hétérogène POPC/DOPG avec le ratio 90:10.

En considérant l'ion sodium comme le centre du récepteur, nous avons calculé par la suite la distribution radiale des lipides jusqu'à 30 Å de l'ion sodium en utilisant le plugin RADIAL PAIR DISTRIBUTION FUNCTION de VMD [142].

Nos simulations montrent que 60 à 80% des DOPG enveloppent le récepteur AT1 pratiquement dès le début de la simulation et maintiennent cette distribution tout le long des trajectoires (Figure 48a). Pour appuyer ces observations, la fonction de distribution radiale de chaque type de lipides autour du récepteur AT1 (n=3) confirme que la première couche de solvatation autour de la protéine est principalement peuplée par des phospholipides DOPG (Figure 48b). Ceci suggère que même à faible concentration de DOPG et dans des conditions plus physiologiques, les DOPG migrent à proximité du récepteur ce qui augmente les interactions intermoléculaires en particulier avec les résidus chargés positivement.

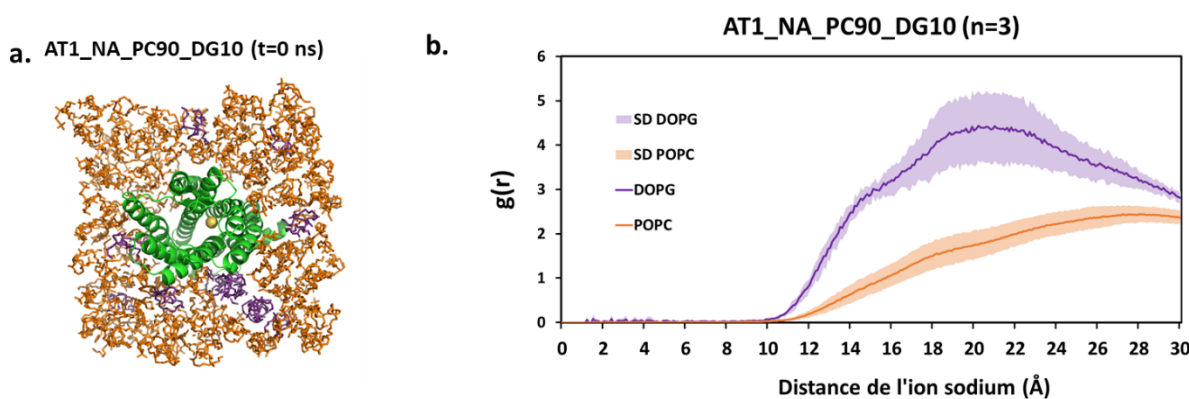


Figure 48 : Distributions des lipides autour du récepteur AT1 pour une membrane POPC/DOPG (90:10).
(a) Snapshot du système à t0. (b) Fonction de la distribution radiale $g(r)$ des lipides autour de la protéine (n=3).

7. Simulations de dynamique moléculaire classique des récepteurs AT1 et AT2 sous l'effet d'un stress mécanique

7.1. Observations générales

Dans cette partie, nous avons étudié l'effet du stress mécanique induit par l'application de différentes tensions de surface sur les récepteurs AT1 et AT2 insérés dans des bicouches homogènes de POPC et DOPG. Afin de déterminer l'effet du stress mécanique sur le récepteur AT1, nous avons réalisé des simulations dynamiques en triplet en présence et en absence de l'ion sodium avec des tensions positives différentes (étirement). Nous avons aussi testé d'une tension négative (compression) dans le cas du récepteur AT1 lié au sodium dans une bicouche de POPC

Les simulations ont été étudiées à la fois par des analyses quantitatives, telles que le RMSD et RMSF (Figure S1) et la mesure des distances TM3-TM6 (Figure 50), et qualitatives avec inspection visuelle des trajectoires. Nous avons utilisé le terme « activation » pour définir un changement de conformation caractérisé par l'ouverture de l'hélice 6 qui est détectable par l'augmentation des distances TM3-TM6.

Lors de l'analyse des distances, nous avons noté que les distances entre les mêmes positions 3.50 et 6.34 ne correspondent pas exactement aux mêmes changements conformationnels. Pour AT1, un début d'ouverture de TM6 correspond à une distance entre 10 et 11 Å alors que, pour AT2, elle se situe autour de 8-9 Å. Ceci pourrait

être dû au fait que l'hélice 6 d'AT1 est basée directement sur une structure cristallographique inactive alors que celle d'AT2 a été modélisée.

Nous avons observé trois types de changement conformationnels :

a. Préactivation : Pour AT2, nous avons observé de fréquentes « préactivations » transitoires ou continues qui se traduisent par une légère ouverture de TM6 dans différentes conditions. Nous en donnons quelques exemples Figure 51.

b. Déstabilisation : Pour AT1, en l'absence de sodium lié et à haute tension (50 dyne/cm), nous avons observé des instabilités de la structure protéique qui se traduisent par des modifications de la cavité interne du récepteur (Figure 51a). Pour quantifier ces modifications, nous avons étudié le nombre total des molécules d'eau distantes de 5 Å des résidus couvrant le domaine transmembranaire interne d'AT1. Pour une analyse plus fine, nous avons subdivisé ce domaine en 3 parties définissant 3 zones de la cavité interne du système. Ainsi, pour AT1 dans une bicouche de POPC, le nombre des molécules d'eau ne varie pas de manière significative en présence de sodium, lors de l'application d'une tension. Cependant, en l'absence de sodium, il y a une augmentation très significative du nombre de molécules d'eau à une tension de 50 dyne/cm (30 ± 7 molécules d'eau au total, à comparer à 21 ± 3 en l'absence de tension). Cette augmentation résulte d'une entrée de l'eau dans les zones 1 et 2. Elle traduit un élargissement de la cavité interne du récepteur (Figure 51b). Pour AT2, dans une des simulations à haute tension, nous avons aussi observé un état instable avec déformation des hélices et mouvement en arrière de TM6 (Figure 50).

c. Activation : Nous avons aussi observé des activations transitoires ou stables que nous allons analyser en détail. Elles sont indiquées sur le Tableau 4 ainsi que sur les heatmaps de la distance TM3-TM6 (Figure 49).

Concernant ces activations nous pouvons faire quelques remarques :

- (1) Elles sont peu fréquentes, mais plus faciles à observer pour AT2 qu'AT1
- (2) Dans le cas d'AT1, elles sont observables uniquement dans une bicouche lipidique anionique de DOPG couplé à un effet de tension
- (3) AT1 est très stable dans une bicouche de POPC en particulier en présence de l'ion sodium qui stabilise la structure même à haute tension (50 dyne/cm).

Tableau 4 : Simulations dynamiques classiques d'AT1 et AT2¹

	conditions		Etat du récepteur observé	
	sodium	Tension (dyne/cm)	POPC	DOPG
AT1	x	-20	Non	-
	x	Sans tension	Non	Non
	x	20	Non	Non
	x	50	Non	1 activation
AT1	-	Sans tension	Non	Non
	-	20	Non	1 activation
	-	50	instabilité	instabilité
AT2	x	Sans tension	Non	Non
	x	20	1 activation	Non
	x	50	Non	structures instables
AT2	-	Sans tension	Non	1 activation
	-	20	1 activation	1 activation
	-	50	Non	1 activation

¹ Durée : 220 ns ; 3 réplicas.

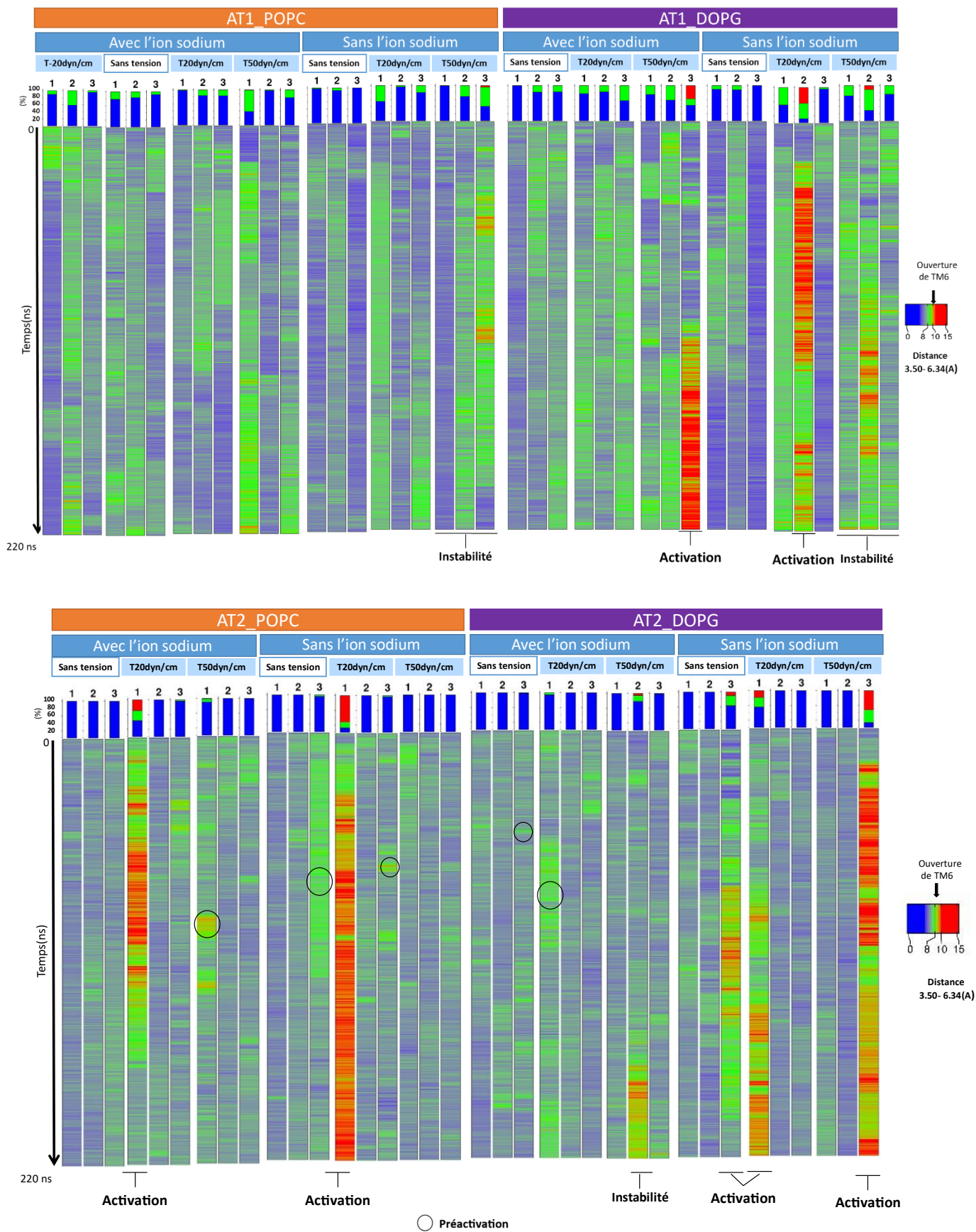


Figure 49 : Heatmaps des distances entre TM3 et TM6 lors des simulations dynamiques classiques.
 Les distances sont mesurées entre les atomes CA des résidus 3.50 et 6.34.

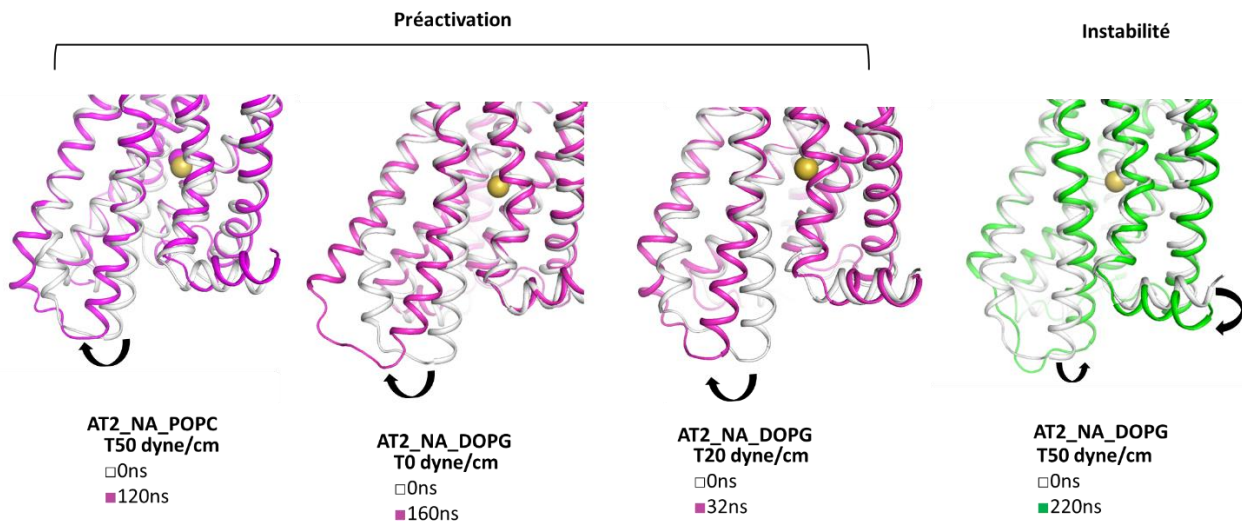
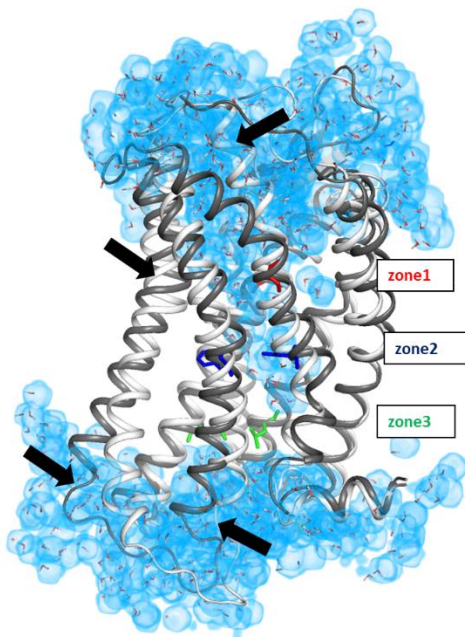


Figure 50 : Snapshots des différents états observés chez AT2 dans une bicouche de DOPG.
 Les états préactifs observés à $\gamma=0$ et $\gamma=20$ dyne/cm sont représentés en magenta. L'état instable observé à $\gamma = 50$ dyne/cm est représenté en vert.

a. AT1 sans l'ion sodium cas d'étirement dans une bicouche POPC ($\gamma=50$)



b.

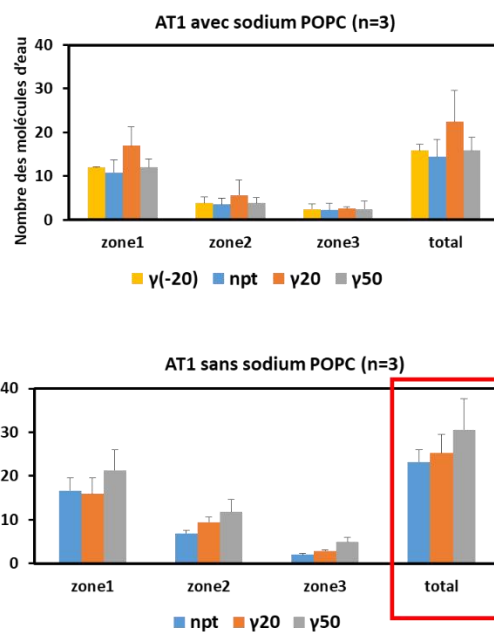


Figure 51 : Déstabilisation d'AT1 sans sodium à haute tension.

(a) densité volumique d'eau du récepteur AT1 lors d'un étirement en l'absence de l'ion sodium (50dyn/cm). La structure à $t=0$ ns est en blanc et la structure à $t=57.3$ ns est en gris, les flèches montrent les zones déstabilisées ; (b) Evolution du nombre des molécules d'eau transmembranaire d'AT1 dans une membrane POPC en fonction de la tension, en présence ou absence de sodium.

7.2. Activation du récepteur AT1

Les lipides anioniques semblent faciliter l'activation de certains RCPG [16]. Dans cette étude, nous avons testé l'effet de ces lipides sur le récepteur AT1 en présence et en absence des effets mécaniques. En présence des forces d'étirement, nous avons observé **2 activations du récepteur AT1 uniquement dans une bicouche de DOPG** (Figure 52):

- Une activation du récepteur AT1 en présence de l'ion sodium à une tension élevée de 50 dyne/cm
- Une activation du récepteur AT1 en absence de l'ion sodium à une tension modérée de 20 dyne/cm.

Les deux activations partagent le même mécanisme d'activation mais l'activation à la tension élevée 50 dyne/cm présente quelques déformations dues à l'instabilité des conditions. Ci-dessous, je détaille l'activation d'AT1 en absence de l'ion sodium à la tension 20 dyne/cm.

a) Analyse détaillée

La transition est initiée par l'étirement de l'hélice 8 et l'avancement de l'hélice 7 dès les premières ns (augmentation du RMSD de 1 à 5 Å dans les 20 premières nanosecondes) (Figure 52a). Cela a permis l'entrée de l'eau dans le domaine transmembranaire (augmentation de nombre des molécules d'eau de 10 à 30 molécules) (Figure 52a,b), provoquant l'ouverture progressive de l'hélice 6 (la distance TM3-TM6 augmente de 6 à 12 Å) jusqu'à l'activation (Figure 52). L'hélice 7 garde la position en avant, ce qui suggère soit un état intermédiaire d'activation, soit une conformation spécifique d'activation.

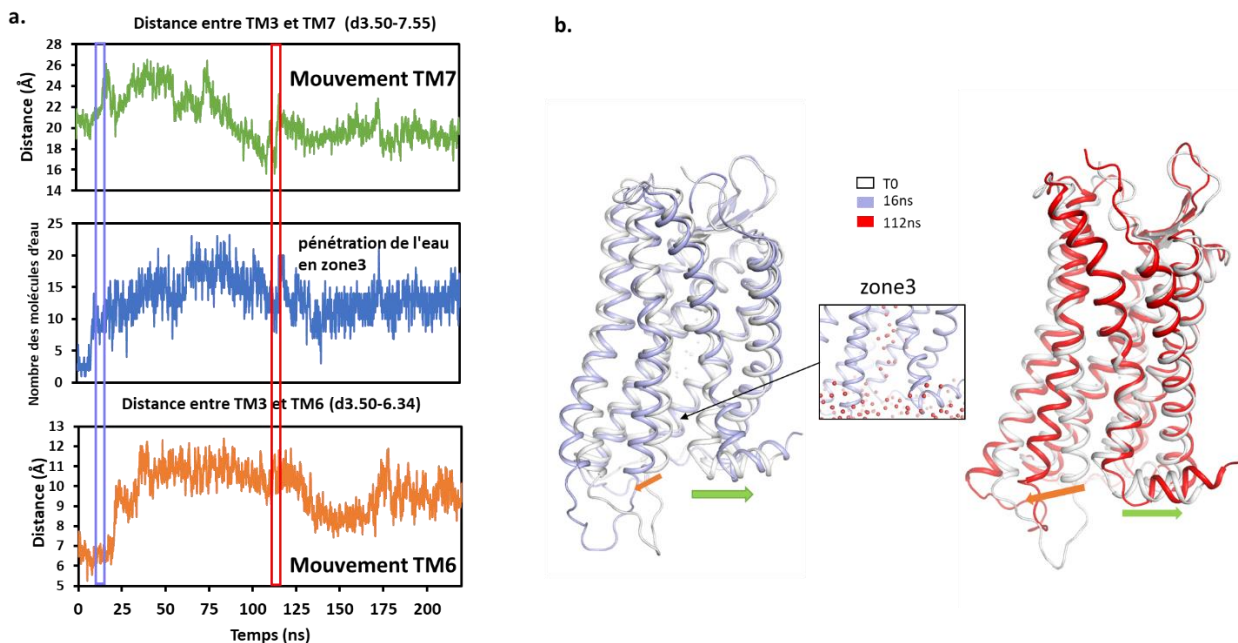


Figure 52 : Activation du récepteur AT1 par le double effet de l'étirement et des lipides anionique DOPG.

(a) Distance entre TM3 et TM7, nombre des molécules d'eau dans le domaine transmembranaire, et distance entre TM3 et TM6. En (b), la structure initiale d'AT1 (blanche) est superposée à la structure obtenue après 16 ns (bleue) et 112 ns (rouge). L'insert montre la pénétration des molécules d'eau lors de l'activation.

b) Interactions de l'hélice 8 avec les DOPG

Le mécanisme d'activation observé en présence à la fois des DOPG et des forces d'étirement débute toujours par l'étirement de l'hélice 8. Nous avons alors cherché la cause de ces mouvements en calculant les interactions entre les résidus de l'hélice 8 et les DOPG. En effet, dans les deux activations observées, les DOPG interagissent avec les résidus cationiques de l'hélice 8 en particulier la lysine K7.61 (Figure 53ab) par le biais de liaisons H orchestrant ainsi les mouvements des hélices 7 et 8. Ces interactions sont relativement de longue durée (environ 50% de la trajectoire pour K294/7.61), capables de perturber le mouvement de l'hélice 8 (Figure 53b,c)

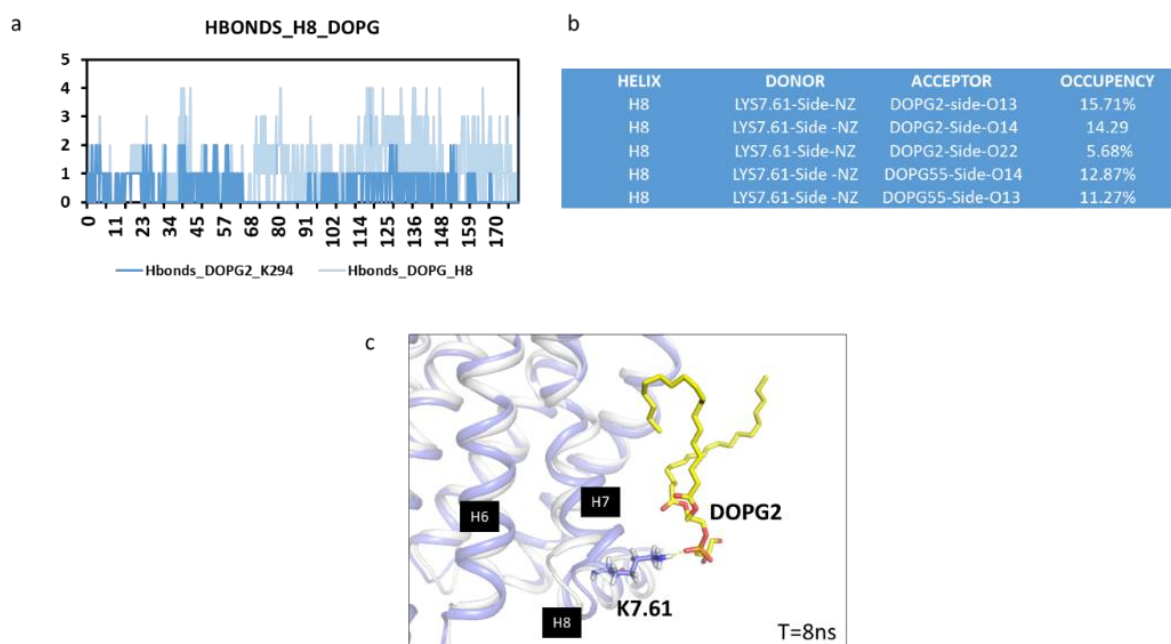


Figure 53 : Liaisons hydrogènes entre l'hélice 8 d'AT1 et les molécules DOPG.

(a) Evolution des H-Bonds entre H8 et les DOPG au cours du temps et spécifiquement la liaison entre DOPG2 et la lysine K7.61 ; (b) Ensemble des liaisons hydrogènes entre la lysine K7.61 et les DOPG et le pourcentage de chaque liaison ; (c) Snapshot de l'interaction entre la lysine K7.61 chargée positivement et la tête du phospholipide DOPG2.

c) Screening des conformations actives des RCPG

La superposition de la structure active obtenue par étirement avec la structure cristallographique active d'AT1 couplé à l'angiotensine II (PDB 6OS0) [143] montre une similarité dans le mouvement en avant de l'hélice 6 mais une différence au niveau du mouvement de l'hélice 7 (Figure 54a). En effet, l'activation par l'angiotensine II est caractérisée par un recul de l'hélice 7 alors que, dans notre cas d'activation par étirement, l'hélice 7 reste dans son état initial (inactive) ou avance par l'étirement de l'hélice 8.

Afin de chercher s'il existe des structures similaires chez les RCPG, nous avons comparé la structure résultante avec toutes les structures cristallographiques de RCPG actives présentes dans Protein Data Bank. Cette exploration 3D a révélé que la structure active d'AT1 obtenue par étirement est semblable à la structure "non canonique" du récepteur 1 de la neurotensine (NTSR1) complexé avec la protéine Gi1 (PDB 6OSA) [36] (Figure 54b).

Dans cet état non canonique, le récepteur NTSR1 présente des caractéristiques de conformation active et inactive, ce qui suggère que la structure pourrait représenter une forme intermédiaire d'activation [6].

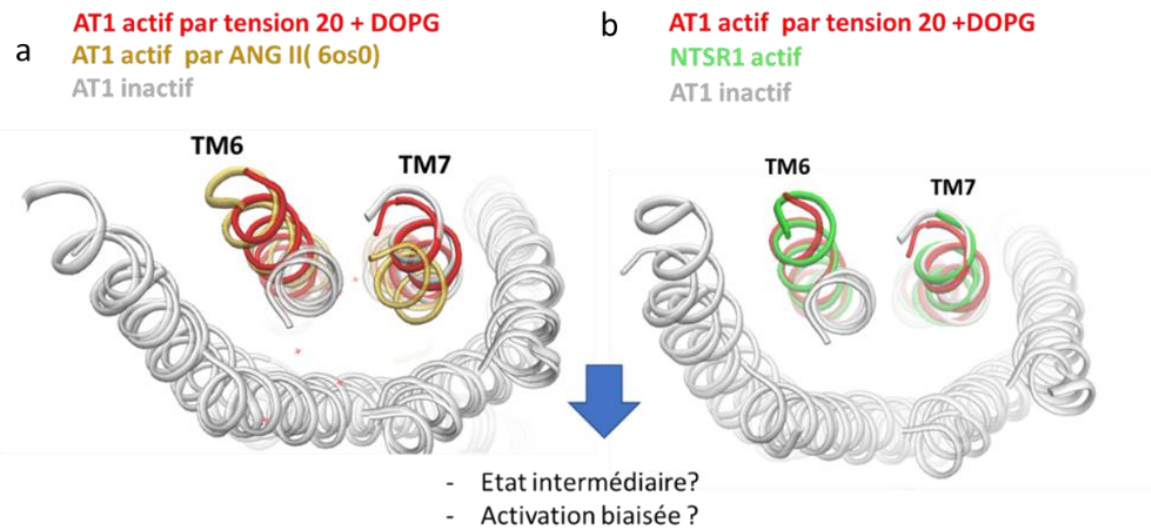


Figure 54 : Comparaison de la structure active d'AT1 obtenue par MD et des structures cristallographiques d'AT1 et de NTSR1.

(a) Snapshot de la superposition de la structure inactive d'AT1 (blanc) avec la structure obtenue par l'application de la tension de surface dans une membrane DOPG et la structure cristallographique active d'AT1 couplé à l'AngII (jaune) ; (b) Snapshot de la superposition de la structure inactive d'AT1 (blanc) avec la structure obtenue par l'application de la tension de surface dans une membrane DOPG et la structure cristallographique non canonique du récepteur NTSR1 (vert).

7.3. Activation du récepteur AT2

A notre connaissance, l'effet du stress mécanique sur le récepteur AT2 n'a pas été étudié expérimentalement ou *in silico*. Donc nous avons soumis ce récepteur en présence et en absence de l'ion sodium à des simulations dynamiques aux mêmes conditions de tensions positives à 20 et 50 dyne/cm. Nous avons pu observer plusieurs activations dans différentes conditions :

- Une activation transitoire du récepteur AT2 en présence de l'ion sodium à une tension de 20 dyne/cm dans une bicouche lipidique POPC
- Une activation continue du récepteur AT2 en absence de l'ion sodium à une tension de 20 dyne/cm dans une bicouche lipidique POPC
- Trois activations continues du récepteur AT2 en absence de l'ion sodium dans une bicouche lipidique de DOPG. L'une a été observée sans tension appliquée, la deuxième à une tension modérée de 20 dyne/cm et la troisième à une tension élevée de 50 dyne/cm.

7.3.1. Activation du récepteur AT2 en présence de l'ion sodium dans une bicouche de POPC

a) Changement conformationnel et verrouillage ionique

Dans l'une des simulations d'AT2 lié au sodium dans une membrane de POPC, en appliquant une tension de 20 dyne/cm, nous avons constaté une activation transitoire du récepteur AT2 durant l'intervalle [40ns, 130ns] au cours duquel le récepteur montre une activation maximale aux alentours de 100ns. Nous observons un mouvement vers l'avant de l'hélice 6 (Figure 55a) qui est quantifié par l'augmentation de la distance entre les hélices 3 et 6 de 6 à 11 Å (Figure 55b). Parallèlement, nous observons aussi un mouvement vers l'arrière de l'hélice 7 quantifié par la diminution de la distance entre les hélices 3 et 7 de 20 à 15 Å.

Nous avons aussi étudié le verrouillage ionique dans l'état actif et inactif du récepteur AT2. D3.49 et R3.50 dans l'hélice 3 sont des résidus très conservés des RCPG. Il a été montré dans certains RCPG que R3.50 peut

interagir avec D7.49 et/ou le résidu 6.30 pour stabiliser l'état inactif, c'est ce qu'on appelle verrouillage ionique (ionic lock) [144]. Nous avons quantifié les liaisons hydrogènes impliquant R3.50 dans l'état actif et inactif. Hors activation, R3.50 interagit essentiellement avec D3.49, alors que, pendant l'activation, la rotamérisation de 3.50 favorise l'interaction avec Y5.58 de l'hélice 5, stabilisant ainsi l'état actif (Figure 55c).

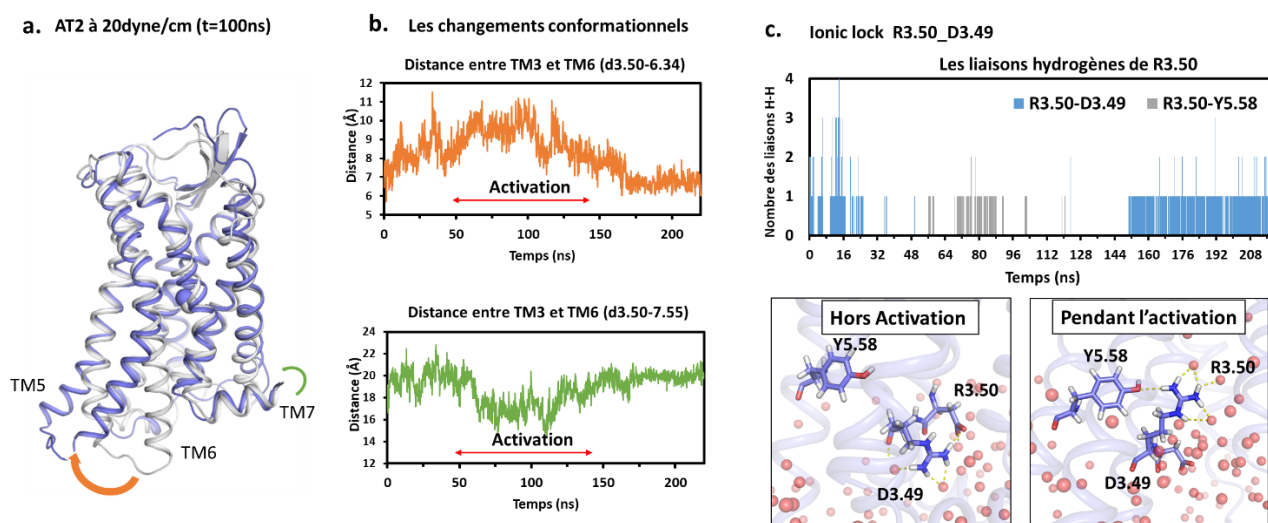


Figure 55 : Activation d'AT2 en présence de sodium dans une bicouche de POPC à la tension de 20 dyne/cm.
 (a) Screenshot de l'activation d'AT2 à la tension de 20 dyne/cm à t=0ns (blanc) et t=100ns (bleue) ; (b) Changements conformationnels mesurés par la distance entre les hélices 3 et 6 (d3.50-6.34) et entre les hélices 3 et 7 (d3.50-7.53) ; (c) Verrouillage ionique lors de l'activation et de l'inactivation du récepteur.

b) Rôle de l'eau dans l'activation du récepteur

Nous nous sommes ensuite intéressés au rôle de l'eau dans l'activation d'AT2 et au changement de coordination de l'ion sodium. Nous avons quantifié le nombre des molécules d'eau à une distance inférieure à 3 Å de D2.50, ce qui correspond à la cavité interne intracellulaire du récepteur. Deux pics d'augmentation significative des molécules d'eau sont observés (Figure 56a). Le premier pic correspond aux 10 premières ns de la simulation où commence le mouvement de l'hélice 6. Le sodium interagit avec D2.50 (Figure 56).

Le deuxième pic coïncide avec l'activation maximale du récepteur où nous avons une augmentation significative des molécules d'eau à l'intérieur du récepteur permettant la formation d'un canal d'eau liant la partie intracellulaire vers la partie extracellulaire et stabilisant l'activation pour quelques dizaines de nanosecondes (Figure 56a). En effet, l'analyse détaillée de l'activation transitoire montre que la rotation simultanée de V6.40 et R3.50 permet l'entrée d'eau au voisinage de D2.50. L'eau permet un effet d'écran entre l'ion sodium et D2.50, ce qui conduit à une réorganisation des interactions impliquant D2.50 et l'ion sodium. Celui-ci se déplace vers le groupe hydroxyle de S3.39 (Figure 56). Cela explique le changement du mode de coordination observé chez AT2 dans son état actif (Figure 56b).

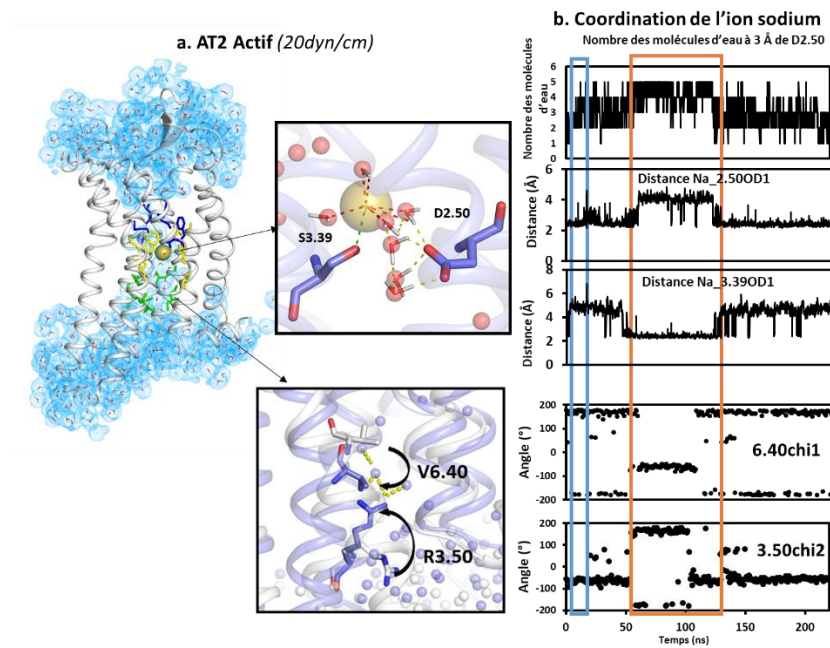


Figure 56: Devenir de l'ion sodium pendant l'activation transitoire du récepteur AT2.

(a) Dans AT2, la rotamérisation de V6.40 et R3.50 permet l'entrée d'eau et une réorganisation de la coordination du sodium avec S3.39; (b) De haut au bas: évolution du nombre des molécules d'eau à 3 Å de D2.50, évolution de la coordination de l'ion sodium avec D2.50 OD1 et S3.39 OD1, évolution des angles chi1 de V6.40 et chi2 de R3.50.

7.3.2. Activation du récepteur AT2 en absence de l'ion sodium dans une bicouche de POPC

A la même tension de 20 dyne/cm, une activation du récepteur AT2 est observée en absence de l'ion sodium. Contrairement à l'activation observée en présence de l'ion sodium, cette activation est continue et caractérisée par un mouvement de l'hélice 7 qui fluctue entre états actif et inactif. Il est quantifié par la distance entre TM3 et TM7 qui varie entre 17 Å et 24 Å. L'ouverture de l'hélice 6 n'est pas maximale (la distance TM3-TM6 augmente de 6 à 11 Å maximum) mais elle reste stable tout le long de la simulation, possiblement par l'effet des forces d'étirement et l'entrée des molécules d'eau (Figure 57).

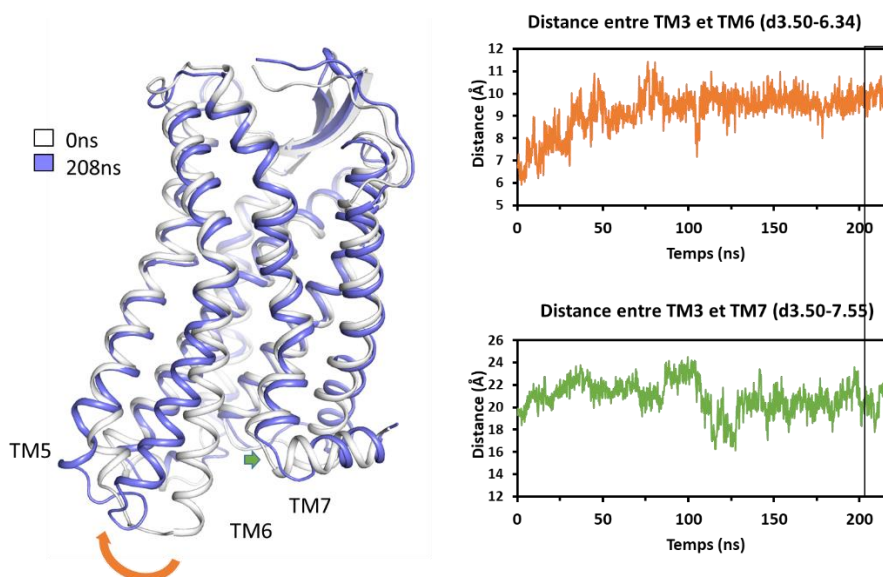


Figure 57 : Activation d'AT2 en l'absence sodium dans une bicouche de POPC à la tension de 20 dyne/cm.

7.3.3. Activations d'AT2 en l'absence de sodium dans une bicouche de DOPG

En absence de l'ion sodium, dans une bicouche lipidique DOPG, nous avons observé 3 activations du récepteur AT2 en absence et en présence de la tension de surface.

Dans des conditions de pression constante (NPT), nous avons observé une activation transitoire avec un maximum d'ouverture de TM6 vers les 110ns, couplé à un léger recul de TM7 dans le même intervalle temporaire (Figure 58a).

Dans des conditions de tension de surface de 20 et 50 dyne/cm, les activations d'AT2 sont maintenues tout le long de la trajectoire avec une ouverture plus importante de TM6 (distance entre TM6 et TM3 maximale de 12Å).

Au cours de ces activations, les mouvements de l'hélice 7 varient d'une simulation à une autre. En effet à 20 dyne/cm, nous observons un mouvement en avant et latéral de TM7 (la distance entre TM7 et TM3 augmente de 19 à 28Å). Au contraire, à 50 dyne/cm, nous observons un recul de TM7 (la distance entre TM7 et TM3 diminue de 20 à 16Å). Dans les 2 cas d'activation, une rotamérisation de R3.50 est aussi observée, facilitant ainsi l'entrée des molécules d'eau et affaiblissant le verrouillage ionique 3.50-3.49 (Figure 58bc).

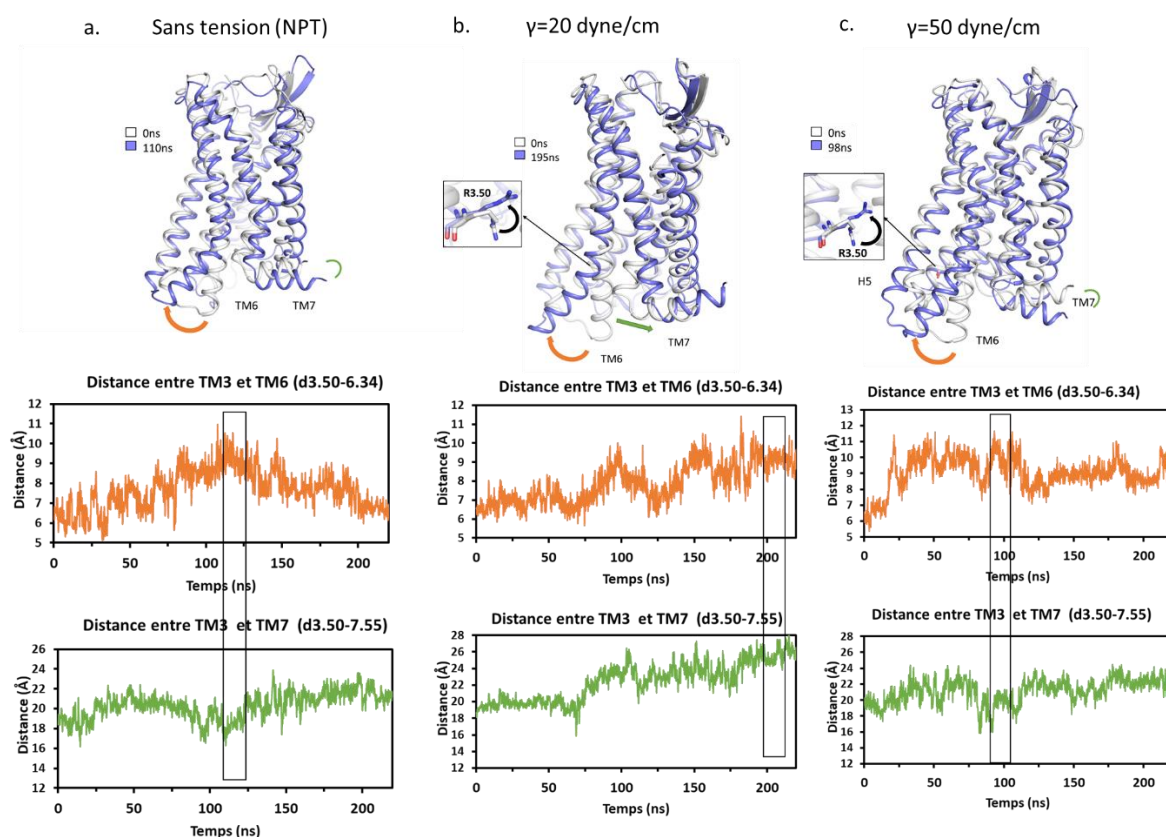


Figure 58 : Activations d'AT2 en absence de l'ion sodium dans une bicouche de DOPG.

(a) Changements conformationnels observés lors de l'activation d'AT2 en l'absence de tension; (b) Changements conformationnels observés lors de l'activation d'AT2 à la tension de 20 dyne/cm; (c) Changements conformationnels observés lors de l'activation d'AT2 à la tension de 50 dyne/cm. Le cadre indique la rotamérisation du résidu R3.50.

7.4. Structures canoniques et non canoniques

De manière globale, nous avons observé 2 activations pour AT1 seulement en conditions extrêmes (100% DOPG) avec tension de surface, alors que, pour AT2, l'activation est plus fréquente dans des multiples conditions (Figure 59). En analysant les mouvements de TM7 (Figure 59 & S2), pour AT1, seules des activations

non canoniques ont été observés avec TM7 soit en position inactive (heatmap TM3-TM7 bleue), soit en mouvement en avant (heatmap TM3-TM7 vert). Ceci peut être expliqué non seulement par la fluidité membranaire et les interactions fréquentes des hélices 7 et 8 avec les molécules DOPG mais aussi par la limitation du temps de calcul (220ns). Chez AT2, l'hélice 7 est plus lâche et flexible, dans la même trajectoire d'activation, nous pouvons observer l'état actif canonique (heatmap TM3-TM7 rouge) et non canonique (heatmap TM3-TM7 rouge) (Figure 59).

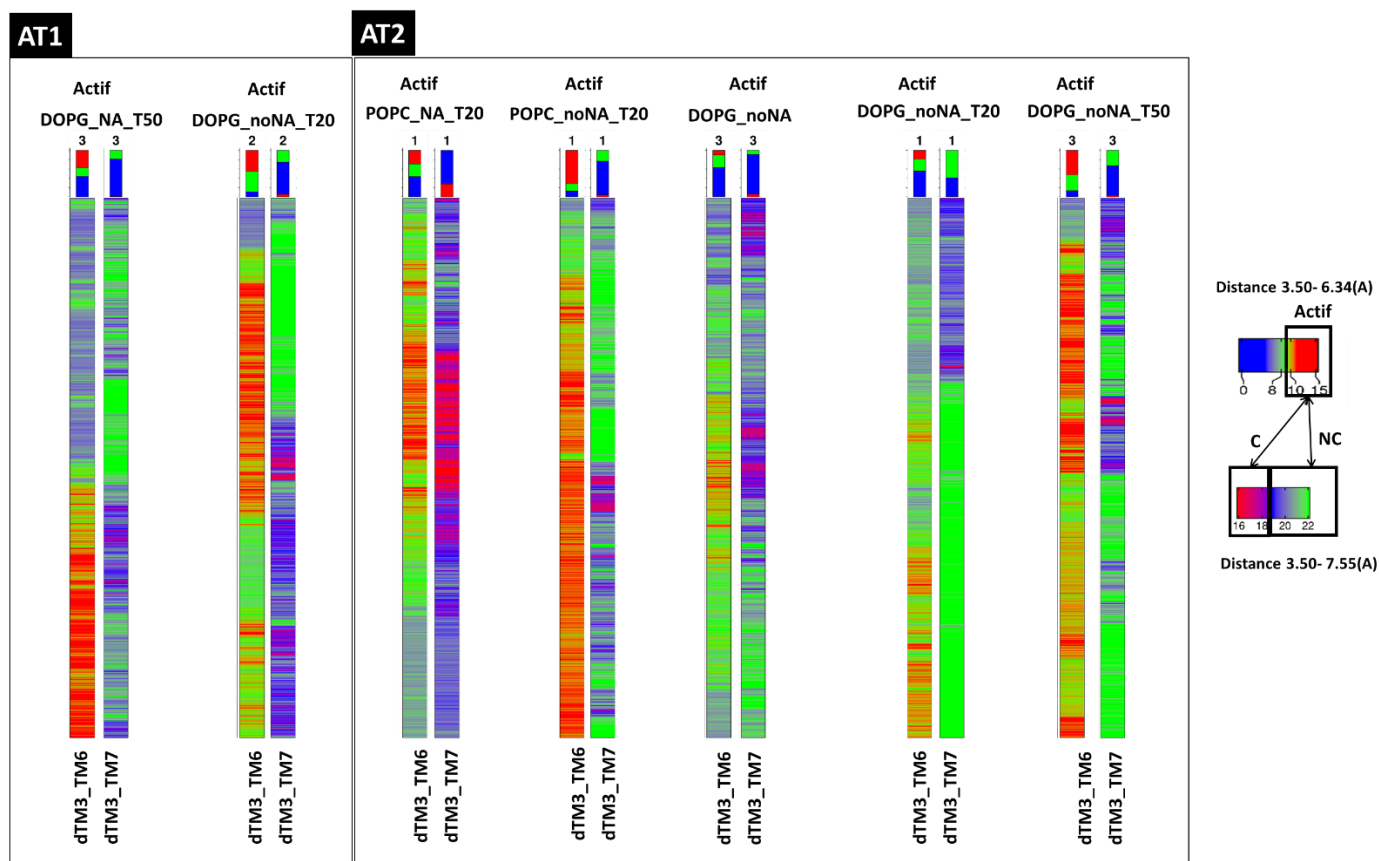


Figure 59 : Bilan des activations observées lors des simulations dynamiques classique.

Les conformations sont visualisées par les heatmaps indiquant par un code couleur les distances TM3-TM6 et TM3-TM7.

7.5. Effets de la tension sur la coordination sodium

Les résultats précédents montrent une tendance de la tension à faciliter les changements conformationnels des récepteurs AT1 et AT2. L'ion sodium étant un modulateur allostérique négatif de l'activation [41], nous avons analysé, sur l'ensemble des simulations classiques, la coordination du sodium pour tester l'hypothèse que la tension pourrait modifier celle-ci. Pour cela, nous avons calculé le pourcentage d'interaction de l'ion sodium avec les atomes électronégatifs protéiques du site allostérique (résidus en position 2.50, 3.35, 3.39, 7.45, 7.46 et 7.49) au cours de toute la trajectoire (distance inférieur à 2.8 Å).

Pour AT1, l'application d'une tension positive ou négative n'induit pas de modification du mode de fixation du sodium comparée à l'absence de tension (NPT). Le sodium interagit essentiellement avec les chaînes latérales de D2.50, S3.39, N7.45 et N7.46 avec des variations au niveau de l'interaction avec le résidu N7.49 (Figure 60).

De manière générale, pour AT2, l'ion sodium interagit essentiellement avec D2.50, N3.35, S3.39 et N7.45 (Figure 60). L'absence d'interaction avec le résidu 7.46 résulte de la différence de séquence entre AT1 et AT2 à cette position. En effet, le résidu 7.46 est une serine pour AT2 et une asparagine pour AT1 humain. Cette différence modifie la liaison avec le résidu 7.46 (Figure 60).

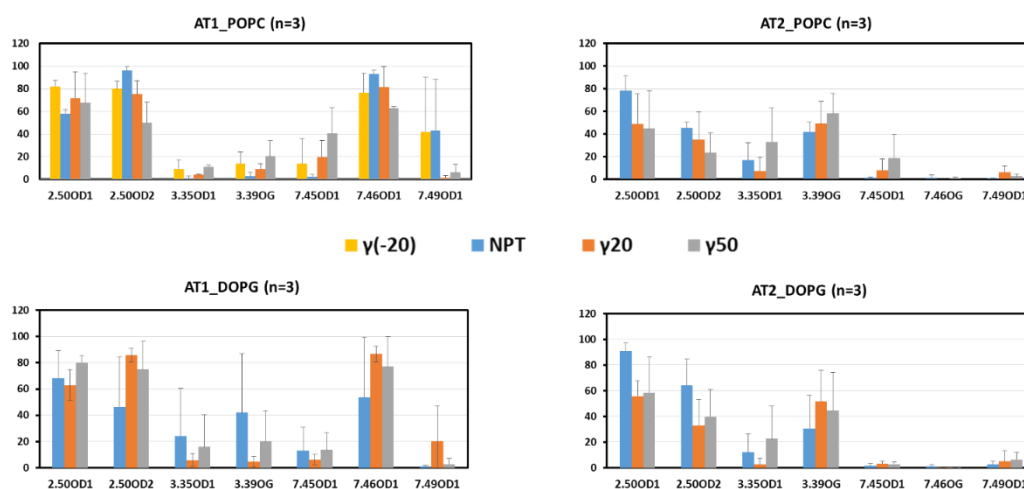


Figure 60 : Mode de liaison de l'ion sodium sous l'effet de la tension chez AT1 et AT2
Pourcentage d'occupation du sodium dans les sites de coordination avec une distance NA oxygène <math>< 2.8\text{\AA}</math>

De manière globale, nous n'observons pas de différence significative dans le mode de coordination de l'ion sodium par l'effet de la tension ou l'effet des lipides. Néanmoins les incertitudes étant importantes (Figure 60), pour valider cette observation, nous avons calculé la composante électrostatique de l'énergie libre de l'interaction de l'ion sodium avec les différents récepteurs dans différentes bicouches lipidiques et sous différentes tensions en utilisant Delphi [145]. Les calculs de la différence d'énergie électrostatique entre les complexes récepteurs-sodium et les éléments séparés (ΔG) ont été réalisés comme décrits dans la section Méthodes de l'article de l'annexe II.1. Pour les ΔG , il y a une différence nette entre AT1 et AT2, ce qui est conforme avec les résultats obtenus dans notre étude évolutive d'AT1 et AT2 [81]. Néanmoins ni la tension, ni les lipides n'affectent l'interaction avec le sodium.

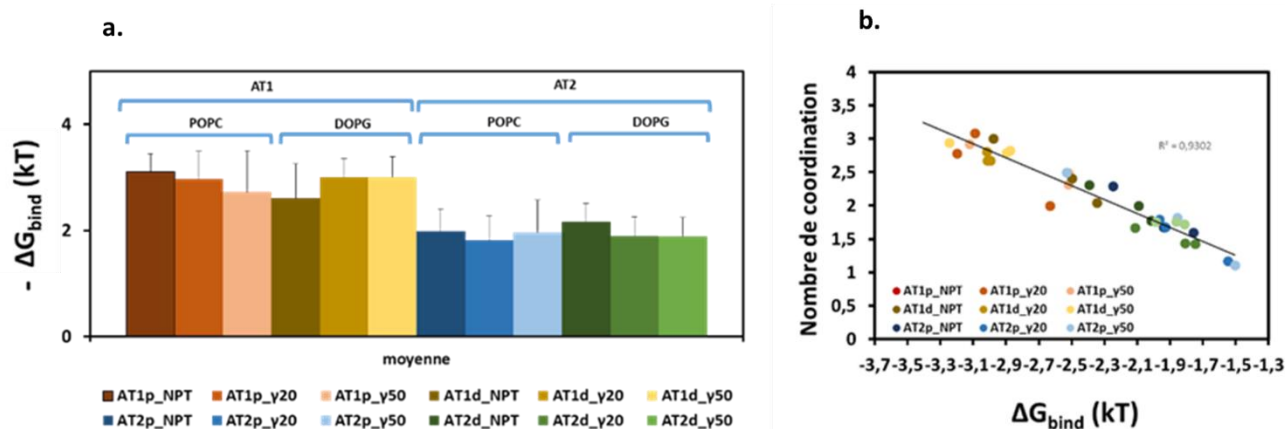


Figure 61 : Énergie libre électrostatique de l'interaction de l'ion sodium avec les récepteurs AT1 et AT2.
(a) Histogrammes de l'énergie libre électrostatique de l'interaction de l'ion sodium avec les récepteurs AT1 et AT2 dans des bicouches de POPC et DOPG (n=3); (b) Énergie libre électrostatique en fonction du nombre de coordination moyen du sodium aux ligands protéiques. Les points représentent les trajectoires individuelles.

8. Analyse des transitions conformationnelles par SMD

Les changements conformationnels observés en dynamique moléculaire classique sont rares, aléatoires et leur observation est très limitée par le temps de calcul. Afin de mieux comparer différentes conditions (nature des lipides, présence de ligands et contraintes mécaniques) et dévoiler le mécanisme d'activation des récepteur AT1 et AT2 de manière reproductible, nous avons mis en place des simulations de dynamique moléculaire dirigée (SMD) en utilisant des contraintes basées sur le principe du quaternion (voir la partie Méthode paragraphe 5.3.2.).

Dans cette méthode, nous avons appliqué une force de traction sur l'extrémité N-terminale de TM6 avec une constante de 5000 kcal / (mol*rad²) pendant 20ns pour entraîner l'ouverture de TM6. Cette étape de SMD a été poursuivie par 10ns de simulations dynamiques classiques (cMD) afin d'équilibrer le système. Les différentes conditions ont été testées avec 6 réplicas.

Toutes les analyses sont reportées sous forme de heatmaps dans la figure S6.

8.1. Récepteur AT1

8.1.1. Observations générales

L'observation d'une structure active non canonique chez AT1 par l'action des lipides anioniques nous a incitées à comprendre les étapes mécanistiques d'activation de ce récepteur. Pour ce but, nous avons réalisé des simulations dirigées en forçant l'hélice 6 à s'ouvrir et en observant le comportement de l'hélice 7. Dans un premier temps, ces simulations ont été réalisées en présence de l'ion sodium pour stabiliser les structures et dans une bicouche lipidique POPC, tout en utilisant une force de traction 5000 kcal/mol/rad² (n=6).

Afin de mieux décrire les mouvements de TM6 et TM7, nous avons mesuré les distances TM3-TM6 (distance entre les CA de 3.50 et 6.35) et TM3-TM7 (distance entre les CA de 3.50 et 7.55). Le mouvement en avant de l'hélice 6 se traduit par une augmentation de la distance TM3-TM6. De la même façon le mouvement en avant ou en arrière de l'hélice 7 se traduit par l'augmentation ou la diminution de la distance TM3-TM7.

Les SMD décrivent de façon très reproductibles (n=6) les étapes nécessaires pour réaliser une transition complète vers une structure active du récepteur AT1 au bout de 30 ns (20ns SMD +10ns CMD). En effet, la carte 2D (Figure 62) des distances TM3-TM6 et TM3-TM7 montrent que, pour AT1, une augmentation simultanée de ces distances est observée principalement au début de la simulation (la distance TM3-TM6 augmente de 7 à 13 Å, la distance TM3-TM7 augmente de 19 à 24 Å). Cela traduit le mouvement en avant des hélices 6 et 7 vers la structure non canonique (points bleues et roses). Dès que l'hélice 6 atteint une ouverture de 12±1 Å, la distance TM3-TM7 diminue de 23 Å à 14 Å traduisant le recul de l'hélice 7 vers la position active observée dans l'activation canonique d'AT1 (point rouges) (Figure 62).

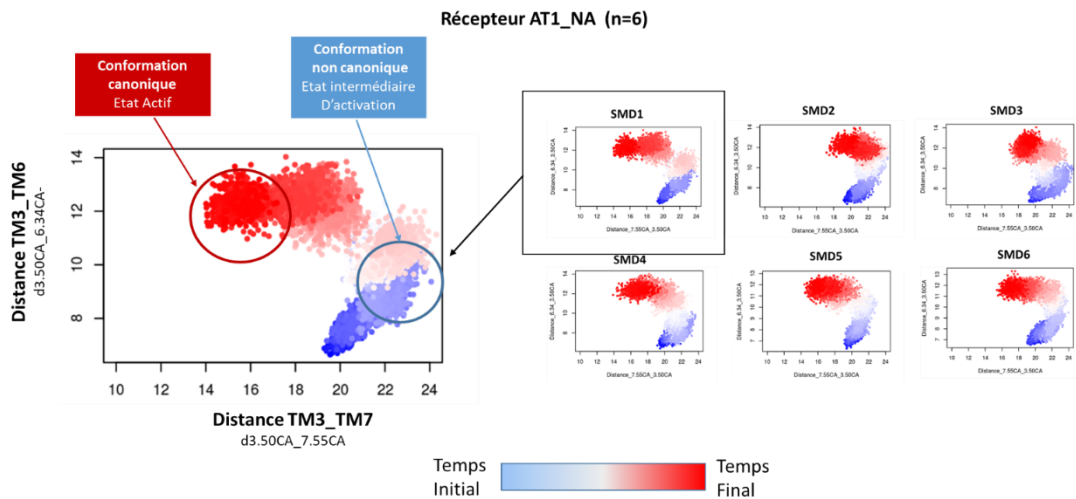


Figure 62 : Transition conformationnelle d'AT1 par SMD dans une bicouche de POPC.
Le dégradé des couleurs indique le temps, du bleu (t=0) vers le rouge (t=30ns).

8.1.2. Les liaisons H

Pour chercher la cause de cet état intermédiaire d'activation, nous avons examiné les interactions électrostatiques entre l'hélice 6 et l'hélice 7 pendant l'activation du récepteur.

Nous avons observé que, chez AT1, il existe une interaction ionique très forte entre la lysine K7.58 de TM7 chargée positivement et l'acide aspartique D6.32 de l'hélice 6 chargé négativement (Figure 63a).

Cette interaction reste stable durant les 10 premières ns des SMD correspondant à l'état inactif mais aussi à l'état préactif où on observe le mouvement en avant de TM7 parallèlement avec le début de l'ouverture de TM6. Au-delà de 10 ns de SMD, cette liaison est interrompue à cause de l'augmentation de la distance entre TM6 et TM7. Ceci coïncide avec le début du recul de l'hélice 7 vers sa position active, induisant ainsi la structure active canonique du récepteur AT1 (Figure 63b). Ces résultats suggèrent fortement que l'interaction entre D6.32 et K7.58 est responsable à l'initiation de la structure non canonique.

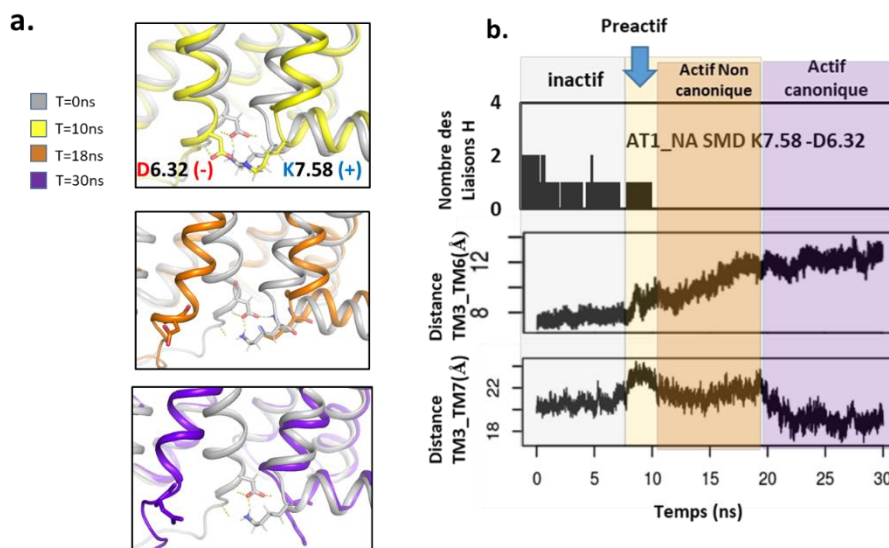


Figure 63 : Rôle des liaisons H dans la formation de la structure non canonique d'AT1.

(a) Liaisons H entre D6.32 et K7.58 d'AT1 dans les formes inactive, préactive, active non canonique et active canonique. (b) Exemple d'implication des liaisons H dans les mouvements de TM6 et TM7 pour AT1 au cours du temps : de haut au bas les figures présentent l'évolution du nombre des liaisons H, l'évolution des distances entre TM3 et TM6 et entre TM3 et TM7. Les couleurs représentent les différents états d'activation d'AT1.

8.1.3. Validation par mutagenèse *in silico*

Afin de confirmer le rôle des liaisons H entre les résidus 6.32 et 7.58 dans la formation de la structure non canonique d'AT1, nous avons muté ces résidus en alanine *in silico* puis nous avons réalisé des simulations SMD suivant le même protocole que précédemment (n=6). Les mutations D6.32A et K7.58A permettent de supprimer les interactions entre ces positions. Les cartes 2D décrivant les mouvements de l'hélice 7 (distance TM3-TM7) et 6 (distance TM3-TM6) montrent, chez le double mutant AT1 D6.32A K7.58A, une différence du profil des mouvements de TM7 et TM6 par rapport au récepteur sauvage. La plupart des SMD montrent un profil similaire à celui d'AT2 (voir plus loin) avec l'absence de la structure non canonique (Figure 64a,c). Nous avons aussi observé une instabilité structurale d'AT1 mutant avec un débobinage de la fin de l'hélice 6 (Figure 64d). Ces résultats confirment l'importance des interactions entre D6.32 et K7.58 dans le mécanisme d'activation d'AT1, la stabilité du récepteur ainsi que la formation de la structure non canonique.

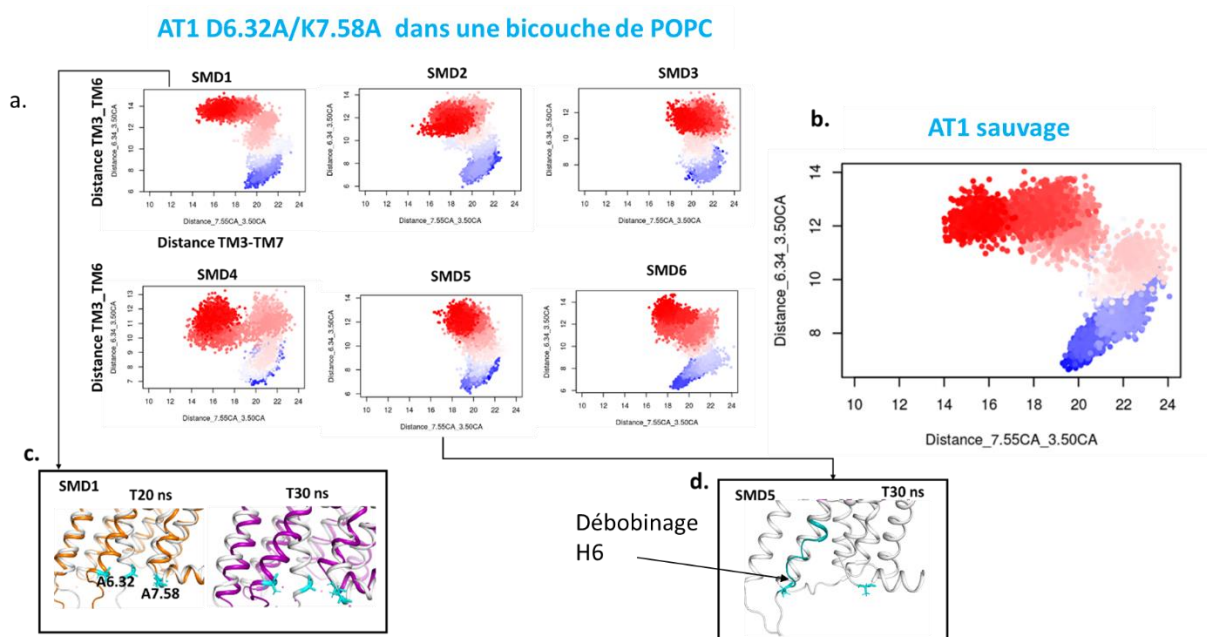


Figure 64 : Transition conformationnelle d'AT1 D6.32A/K7.58A par SMD dans une bicouche de POPC.

(a) Cartes 2D décrivant les mouvements de TM7 et TM6 du mutant d'AT1 D6.32A et K7.58A (n=6) ; (b) Carte 2D décrivant les mouvements de TM7 et TM6 d'AT1 sauvage ; (c) Snapshots montrant l'absence d'interaction entre A6.32 et A7.58 ; (d) Snapshots montrant l'instabilité de TM6 à 30 ns au cours du réplica SMD5.

8.1.4. Comparaison avec le récepteur AT1 d'anguille

Le récepteur AT1 d'anguille diffère du récepteur humain au niveau de la nature des résidus 7.58 et 6.32 mais partage les mêmes charges électrostatiques. En effet, dans AT1 d'anguille, le résidu 7.58 est une arginine chargée positivement et le résidu 6.32 est un acide aspartique chargé négativement. Nous avons alors modélisé le récepteur AT1 d'anguille et puis nous avons réalisé des simulations SMD suivant le même protocole que précédemment (n=6). Les cartes 2D décrivant les mouvements de l'hélice 7 et 6 montrent, pour AT1 d'anguille, une différence du profil des mouvements de TM7 et TM6 par rapport au récepteur humain. En effet, le mouvement en avant de l'hélice 7 est moins important pour AT1 d'anguille (pour AT1 humain la distance TM3-TM7 augmente de 19 à 24 Å maximum alors que pour AT1 d'anguille cette distance ne dépasse pas les 22 Å) (Figure 65a,b). L'analyse visuelle des SMD montre que, bien que l'interaction électrostatique entre les résidus 7.58 et 6.32 existe chez AT1 d'anguille et maintient bien TM7 avec TM6 à l'état inactif (t=0) et préactif (t=20ns), le couplage du mouvement en avant de TM7 et TM6 est moins important. Ceci peut être expliqué par

la longueur des chaînes latérales de ces deux résidus R7.58 et E6.32. Etant plus longues que dans le récepteur humain, il y a assez d'espace pour que TM6 avance sans entrainer TM7 en avant (Figure 65c). Ces résultats confirment bien l'importance de la nature des résidus 6.32 et 7.58 dans le mécanisme d'activation d'AT1, en particulier pour le mouvement de l'hélice 7 et la formation de la structure non canonique.

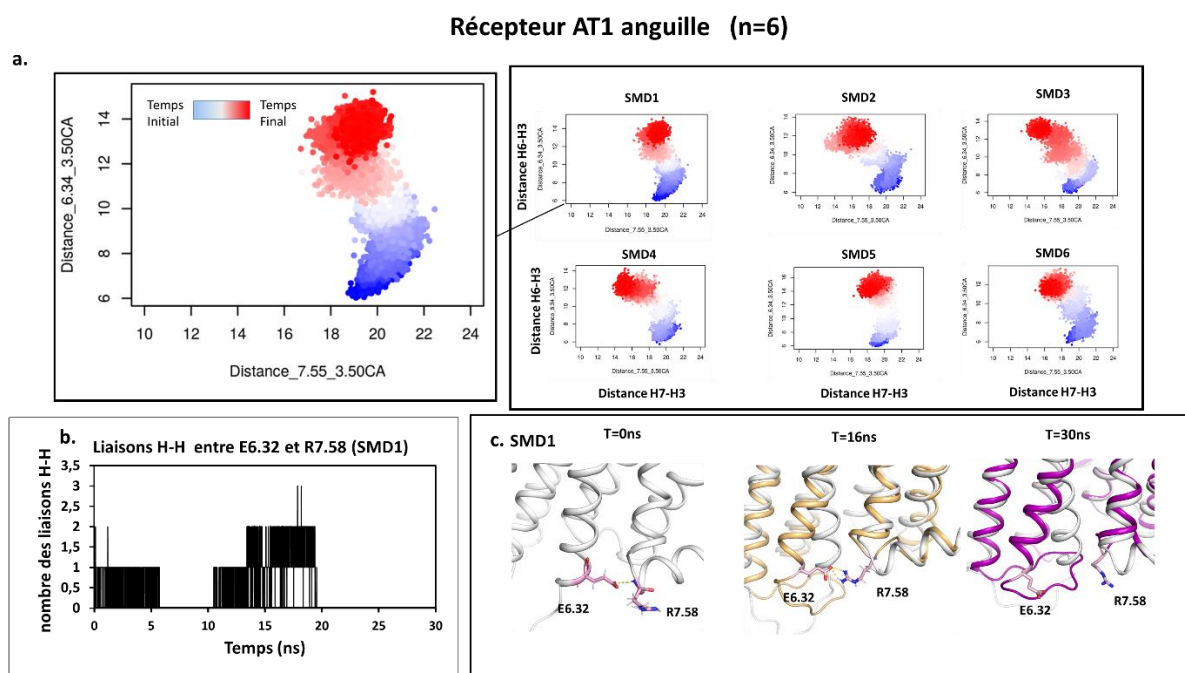


Figure 65 : Transition conformationnelle d'AT1 d'anguille par SMD dans une bicouche de POPC. (a) Cartes 2D décrivant les mouvements de TM6 et TM7 du récepteur d'anguille (n=6) ; (b) Exemple d'évolution du nombre des liaisons H entre E6.32 et R7.58 ; (c) Snapshots montrant l'interaction entre E6.32 et R7.58 à t=0, t=20ns et t=30ns.

8.2. Récepteur AT2

Nous avons étudié les transitions conformations d'AT2 humain inséré dans une membrane de POPC, par SMD, en suivant le même protocole que précédemment (20 ns SMD suivie de 10 ns cMD). Chez AT2, la distance TM3-TM6 augmente tandis que la distance TM3-TM7 diminue de façon simultanée. Cela traduit un mouvement en avant de l'hélice 6 et un mouvement en arrière de l'hélice 7 jusqu'à l'obtention de la structure active canonique d'AT2 (Figure 66a).

Dans le cas d'AT1, l'interaction entre les résidus 6.32 et K7.58 est responsable de l'initiation de la structure non canonique d'AT1. Dans le cas d'AT2, cette liaison est absente (Figure 66b,c) à cause des mutations des acides aminés à ces positions. En effet les résidus 6.32 et 7.58 sont respectivement une glutamine et une asparagine qui sont neutres et ne forment pas de liaisons H entre elles (Figure 66c). Ceci explique le recul plus rapide de l'hélice 7 et l'absence d'une structure non canonique chez AT2.

Nous pouvons noter que les cartes 2D confirment que, à l'état inactif, la distance entre TM3 et TM6 est légèrement plus petite pour AT2 que pour AT1 (Figure 62 et 66).

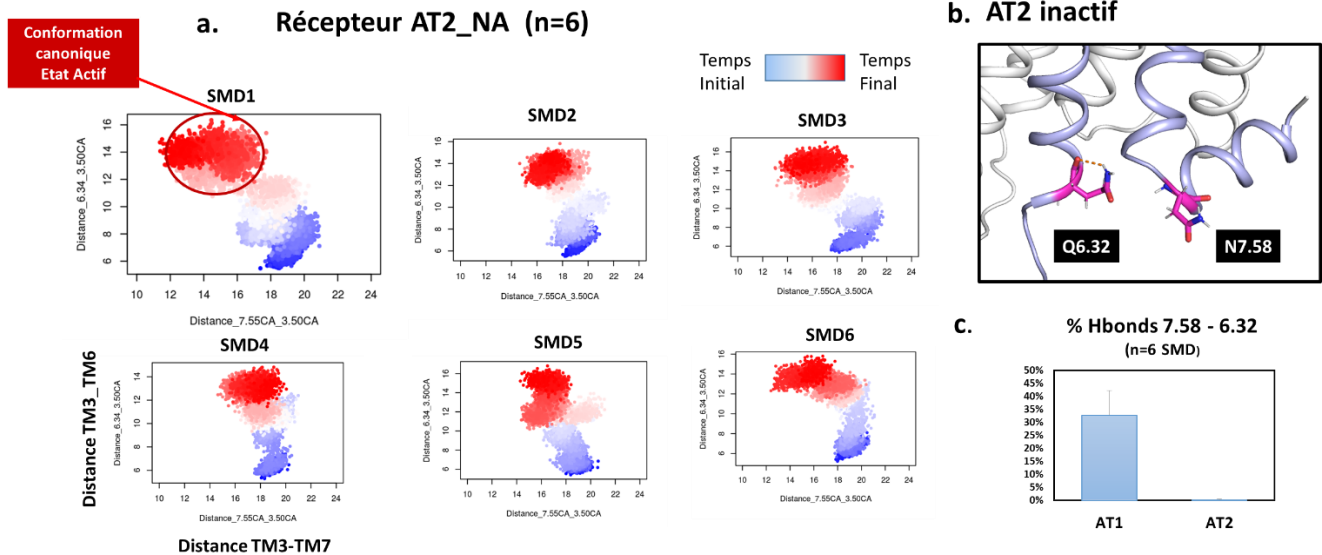


Figure 66 : Transition conformationnelle d'AT2 par SMD dans une bicouche de POPC.

(a) Cartes 2D comparant les mouvements de TM7 et TM6 d'AT2 humain dans une membrane POPC ; (b) Absence de liaison H entre Q6.32 et N7.58 chez AT2 ; (c) Histogramme présentant le pourcentage de liaisons H entre les résidus 7.58 et 6.32 chez AT1 et AT2. Les données représentent la moyenne et la déviation standard des 6 réplicas.

8.3. Mécanismes proposés

Nos observations structurales et les cartes 2D d'AT1 dans des conditions optimales et stables (force 5000 kcal/mol/rad², AT1 couplé au sodium dans une bicouche lipidique POPC) (Figure 67) mettent en évidence les étapes nécessaires pour réaliser une transition vers une structure canonique active:

- TM6 et TM7 étant maintenus par les interactions entre D6.32 et K7.58, un mouvement de l'un vers l'extérieur entraîne le mouvement de l'autre, d'où le mouvement simultané initial vers l'extérieur des hélices 6 et 7.
- La structure non canonique apparaît comme une étape intermédiaire dans l'activation des récepteurs avec l'hélice 6 en position active et l'hélice 7 en position inactive.
- L'hélice 6 continue à avancer ce qui entraîne la rupture de la liaison entre D6.32 et K7.58.
- L'hélice 7 effectue alors un mouvement vers l'intérieur, conduisant à l'apparition de la structure active canonique.

Ces résultats confirment l'hypothèse que la structure non canonique est une structure intermédiaire d'activation.

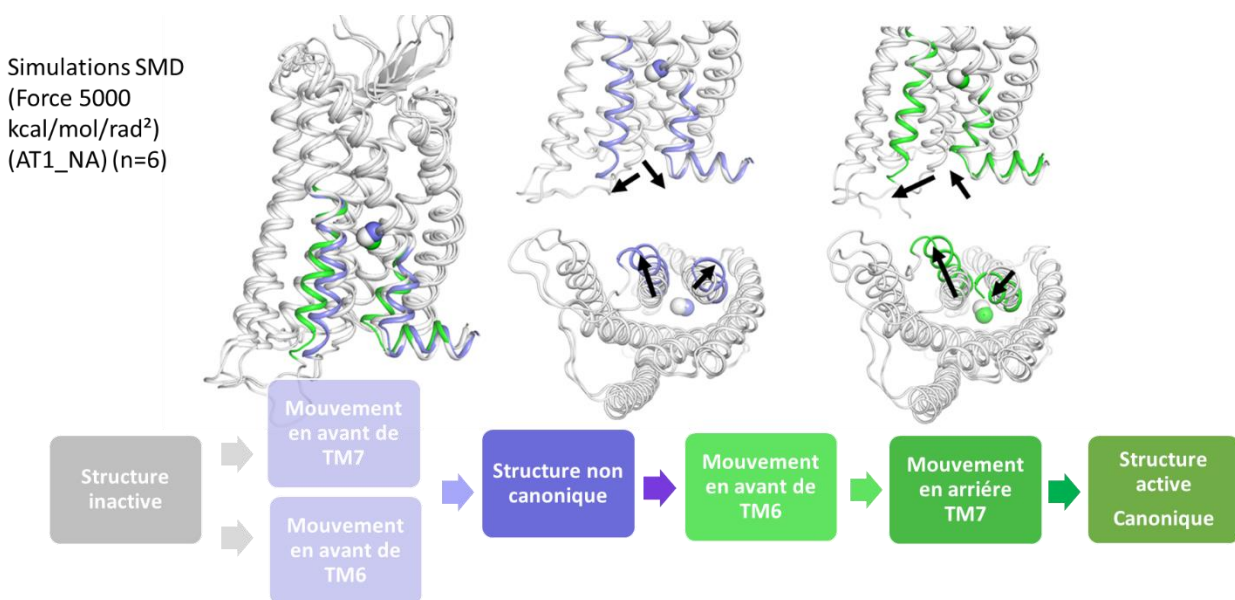


Figure 67 : Etapes mécanistiques d'activation du récepteur AT1 dans une bicouche de POPC.

La structure blanche présente l'état initial, la structure bleue présente l'état intermédiaire et la structure en vert présente l'état final. Les flèches noires présentent le mouvement de TM6 et TM7.

Pour AT2 et dans les mêmes conditions, les étapes mécanistiques sont les suivantes :

- A partir de la conformation inactive, un mouvement en avant de TM6 est observé avec un recul léger de TM7.
- A une ouverture maximale de TM6, un recul important de TM7 est observé, induisant ainsi une transition vers la forme d'activation canonique du récepteur.

Nous avons remarqué la quasi absence ou l'apparition très timide de la structure intermédiaire non canonique chez AT2 (Figure 68).

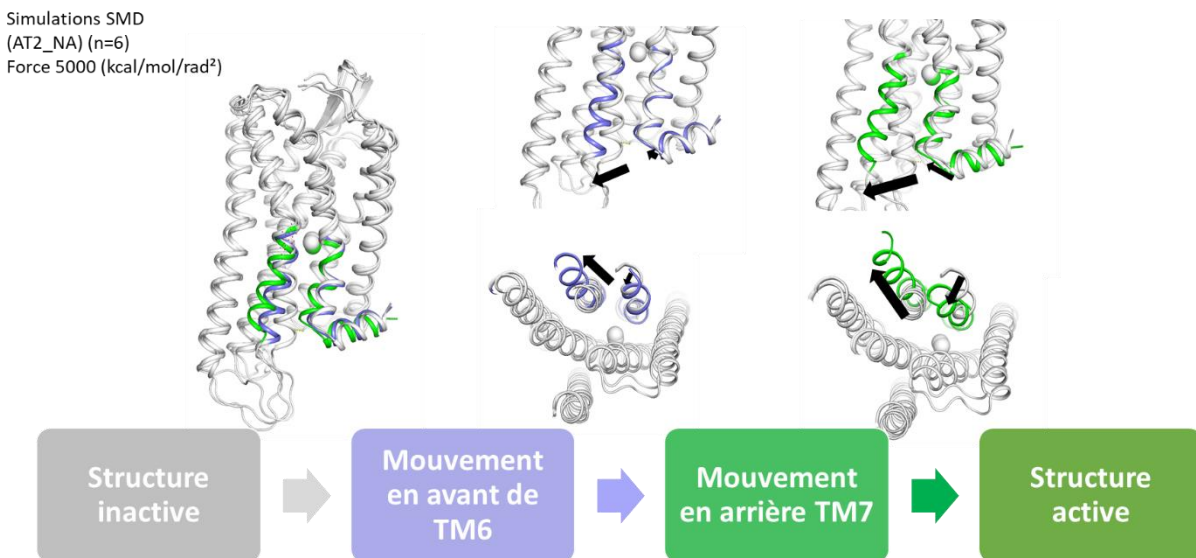


Figure 68 : Etapes mécanistiques d'activation du récepteur AT2 dans une bicouche lipidique POPC.

La structure blanche présente l'état initial, la structure bleue présente l'état intermédiaire et la structure en vert présente l'état final. Les flèches noires présentent le mouvement de TM6 et TM7.

8.4. Les transitions conformationnelles d'AT1 dans une bicouche de DOPG

Afin de comprendre l'effet des lipides DOPG sur les transitions conformationnelles d'AT1, nous avons réalisé des SMD d'AT1 lié au sodium dans une bicouche lipidique de DOPG dans des conditions optimales (force de traction de 5000 kcal/mol/rad²). Les cartes 2D décrivant les mouvements des hélices 6 et 7 montrent l'hétérogénéité des mouvements de l'hélice 7 entre les répliques (Figure 69a). Pour expliquer cette instabilité de TM7, nous avons cherché l'effet des DOPG sur la liaison ionique D6.32 - K7.58. Ainsi, nous avons quantifié les liaisons H de K7.58 avec les lipides et avec les résidus protéiques. Dans l'exemple montré sur la figure 69, au début des SMD (les 10 premières nanosecondes), comme observé pour AT1 avec les POPC, K7.58 interagit avec D6.32 (Figure 69b,c). Au-delà de 10 ns, cette liaison étant fragilisée par les forces de traction de TM6, la lysine 7.58 interagit avec les DOPG (Figure 69b,c) et maintient cette interaction tout le long de la simulation. Cette interaction est reproductible dans les 6 répliques et, par conséquent, peut influencer les mouvements de l'hélice 7 vers l'arrière ou vers l'avant. Ces résultats mettent en évidence que certains lipides peuvent agir comme modulateurs allostériques du récepteur AT1.

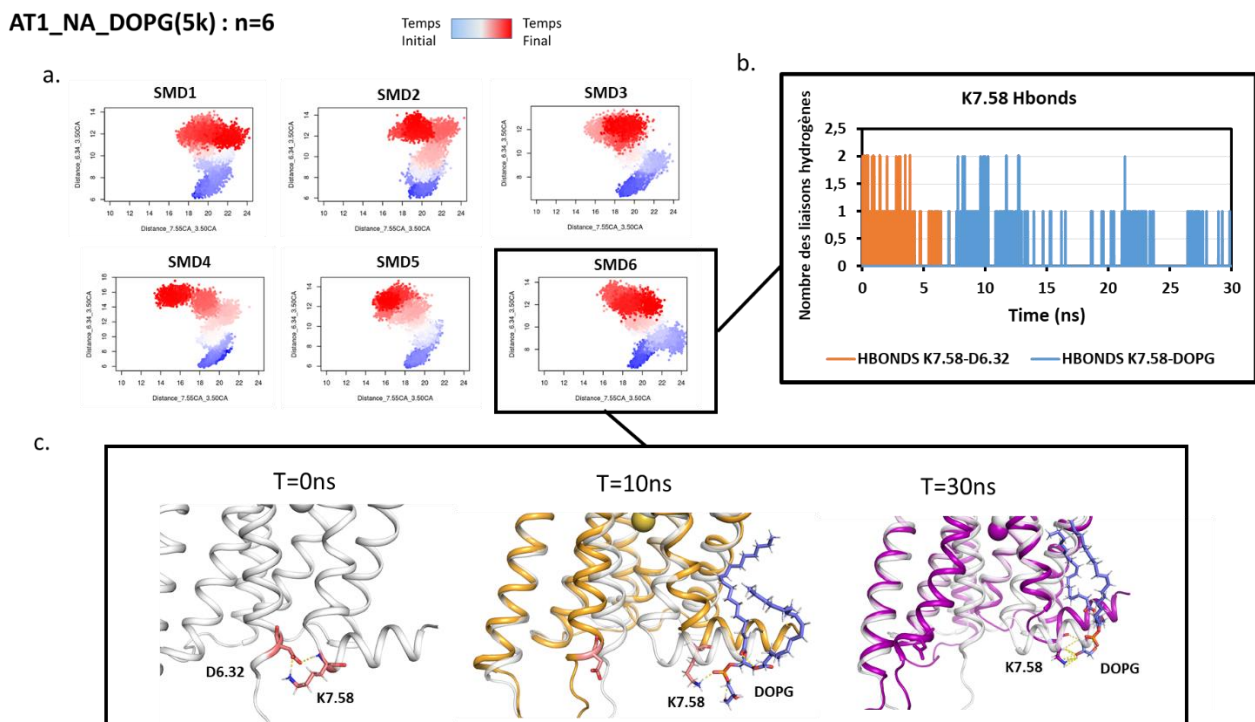


Figure 69 : Transition conformationnelle d'AT1 par SMD dans une bicouche de DOPG.

(a) Cartes 2D décrivant les mouvements de TM6 et TM7 d'AT1 lié au sodium dans une bicouche lipidique de DOPG (n=6) ;
 (b) Liaisons H de K7.58 avec D6.32 (en orange) et avec DOPG (en bleu) en cours du temps ; (c) Snapshots montrant les interactions de la lysine K7.58 à t = 0, t = 10ns et t = 30ns.

8.5. Les transitions conformationnelles d'AT1 sous l'effet de la tension

Afin d'examiner l'effet du stress mécanique sur les étapes mécanistiques d'activation d'AT1, nous avons réalisé des SMD d'AT1 lié au sodium dans une bicouche de POPC dans les mêmes conditions que précédemment (n=6)

en rajoutant une tension de surface de 20 dyne/cm. Les cartes 2D décrivant les mouvements des hélices 6 et 7 montrent 3 simulations ayant une structure finale non canonique et 3 simulations ayant une structure active canonique (Figure 70). Ces résultats montrent que le stress mécanique sous forme d'étirement peut avoir un effet sur les mouvements de l'hélice 7 favorisant parfois l'activation du récepteur vers une structure non canonique même en absence de lipides anioniques.

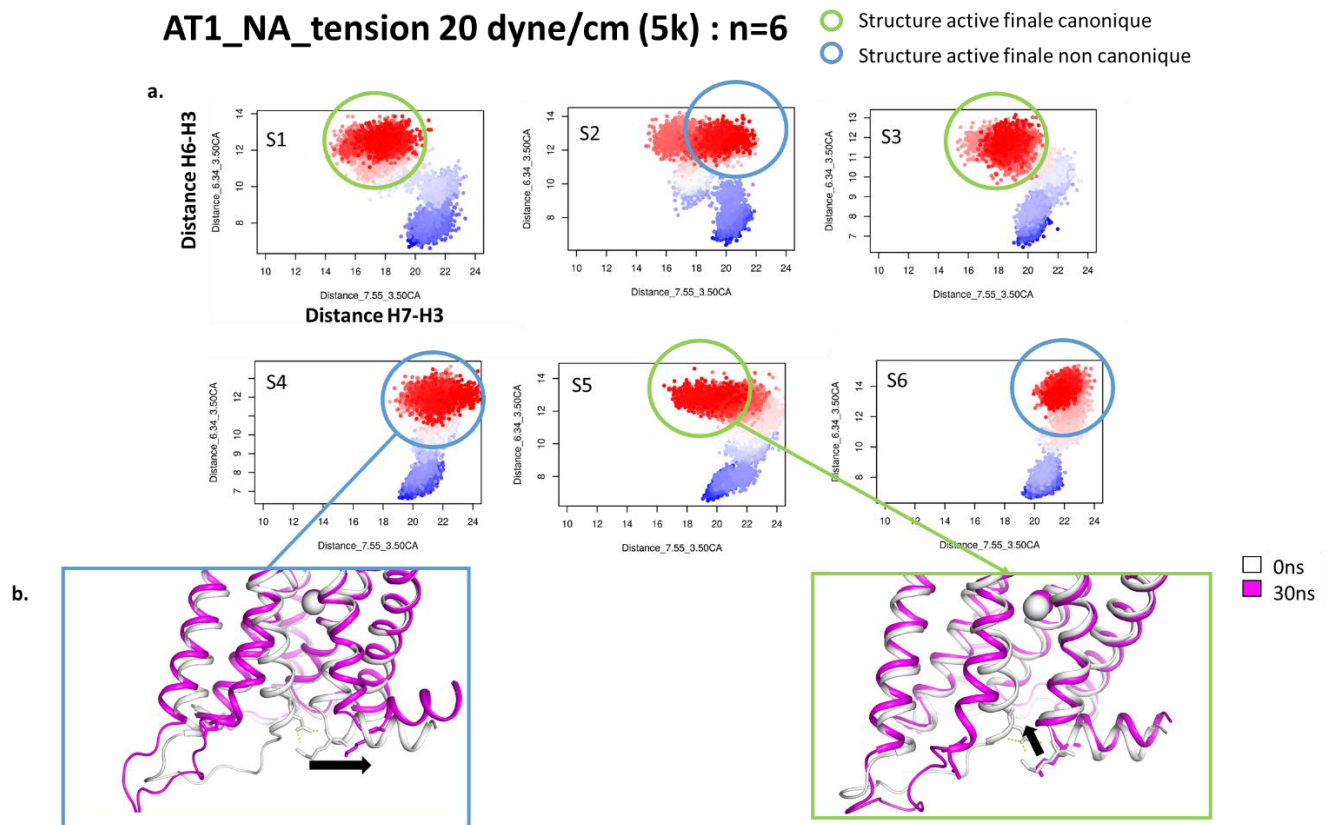


Figure 70 : Transition conformationnelle d'AT1 dans une bicouche de POPC sous tension de 20 dyne/cm. (a) Cartes 2D décrivant les mouvements de TM7 et TM6 d'AT1 sous la tension de 20 dyne/cm (n=6) ; (b) Snapshots de l'état initial (en blanc) et final (en magenta) des simulations 4 et 5.

Discussion

L'hypothèse de base de ce projet est que l'environnement vasculaire aussi bien dans des conditions physiologiques que pathologiques peut modifier la régulation du système rénine-angiotensine au niveau du fonctionnement moléculaire des récepteurs AT1 et AT2. Cet environnement dépend de divers facteurs physiques, notamment le stress mécanique auquel sont soumises les cellules de l'endothélium, ainsi que de facteurs biochimiques comme la composition lipidique de la membrane. L'objectif de ce projet est donc d'approfondir la compréhension des mécanismes moléculaires de l'activation d'AT1 et AT2 sous contraintes mécaniques et dans divers environnement lipidiques.

En utilisant une approche *in silico* de modélisation et de simulations de dynamique moléculaire, nous avons visé à répondre aux questions suivantes :

Des facteurs externes (stress mécanique ou lipides) modifient-ils la conformation inactive des récepteurs AT1 et à l'AT2, en absence de ligand?

Lors des simulations dynamiques d'AT1 et AT2 réalisées dans différentes conditions de stress mécanique ou d'environnement lipidique, nous avons observé quelques changements de conformations (Table 4) que nous avons détaillés dans les résultats. Ici, nous nous intéressons à l'influence de ces facteurs sur la structure des récepteurs inactifs. L'analyse des RMSD et RMSF ne permet pas de mettre en évidence des différences significatives de la structure globale de ces récepteurs selon les différentes conditions de simulations (Fig S1). Nous avons donc réalisé plusieurs analyses complémentaires pour déterminer si des différences plus fines pouvaient exister.

Structure globale :

Dans le but de chercher l'effet de la tension membranaire sur la conformation inactive des récepteur AT1 et AT2, nous avons mesuré le nombre des molécules d'eau dans le domaine transmembranaire. Dans les conformations inactives, nous n'observons pas d'augmentation significative du nombre moyen des molécules d'eau sauf à haute tension (50 dyne/cm) en l'absence de sodium lié au récepteur. Dans ce cas, la présence accrue d'eau apparait comme liée à une instabilité du récepteur dans les conditions de simulations. Lors de l'application d'une tension plus modérée, nous n'avons pas observé de changement notable du nombre de molécules d'eau. Dans ces conditions (20 dyne/cm), nous pouvons conclure que l'effet de la tension ou l'étirement ne se traduit pas par un relâchement direct de la structure qui se traduirait par une augmentation des molécules dans le domaine transmembranaire.

Interaction avec l'ion sodium :

Le sodium a un rôle de régulateur allostérique négatif dans divers GPCRs de la famille A [44]. Nous avons cherché à déterminer s'il y avait des changements significatifs dans la coordination de l'ion de sodium par l'effet du stress mécanique ou par l'action des changements de la composition lipidique. Aucune différence significative ni de la coordination du sodium (Figures 60 et S3), ni de l'énergie électrostatique d'interaction de celui-ci avec les récepteurs (Figure 61) en fonction de la tension appliquée ou de la nature des lipides n'a pu être mise en évidence.

Nous pouvons toutefois noter une observation intéressante lors de l'activation transitoire du récepteur AT2 en présence du sodium à une tension de 20 dyne/cm. Cette activation n'a pas montré une sortie du sodium comme décrit pour certains RCPG [146] mais un changement de coordination avec l'ion sodium qui interagit

avec S3.39 durant la période d'activation tout en perdant les interactions avec la chaîne latérale de D2.50 par l'effet écran des molécules d'eau. Cet événement étant unique dans nos simulations, nous ne pouvons pas relier ce changement de coordination à l'effet de la tension. Néanmoins, il suggère que la réorganisation des interactions du sodium avec le récepteur pourrait être une condition d'activation du récepteur. Cette réorganisation serait liée à l'entrée de molécules d'eau dans le domaine transmembranaire.

Liaisons H :

Les simulations dynamiques dirigées montrent l'importance des liaisons électrostatiques entre les résidus 6.32 et 7.58 dans la stabilisation de la structure inactive et l'apparition de la structure active non canonique chez AT1. Nous avons alors repris nos simulations dynamiques classiques et mesuré le nombre moyen des liaisons H entre les résidus 6.34 et 7.58 et entre ces résidus et les lipides dans différentes conditions (Fig .S4)

Il n'est pas possible de conclure qu'il existe un effet général dû à la tension ou aux lipides sur ces interactions. Néanmoins, en reprenant les cMD individuellement (Figure S5), pour AT1, la liaison H entre les résidus 6.32 et 7.58 semble être présente dans toutes les simulations dynamiques où le récepteur est resté à l'état inactif ou préactif. Par contre, dans les 2 simulations dynamiques qui montrent une activation dans une membrane de DOPG, nous observons la disparition de cette liaison dès le début de la simulation. Cela suggère que la disparition de cette liaison peut faciliter l'activation du récepteur AT1 par des conditions externes telles que la tension et les lipides anioniques.

Des facteurs mécaniques facilitent-ils les changements conformationnels ?

Nos résultats de simulations dynamiques classiques montrent que dans, une bicouche de POPC, des facteurs mécaniques peuvent induire des changements de conformation chez AT2 et des instabilités chez AT1.

- AT1 étant très stable dans nos conditions d'étude, il n'est pas facile de conclure un effet direct de la tension. Cette limitation peut être reliée aux conditions des simulations dynamiques classiques qui sont courtes (220ns) et en nombre limité (n=3) et ne présentent pas des conditions réelles où l'intervalle de temps est beaucoup plus long (millisecondes ou plus) d'où l'intérêt d'adopter une méthode dirigée.

- Pour AT2, l'application d'une tension de surface modérée (20 dyne/cm) suffit pour induire un changement conformationnel vers un état actif, y compris en présence d'un sodium fixé au site allostérique (Figures 55 & 57)

Ceci concorde avec les études précédentes sur AT2 qui suggèrent que ce récepteur est facilement activable [94]. Ainsi, à ce jour, les structures cristallographiques résolues d'AT2 sont des structures actives [98].

Des lipides facilitent-ils les changements conformationnels ?

Nous avons mis en évidence que le stress mécanique, par le biais de changement de la tension de surface membranaire, peut rendre les lipides plus fluides donc plus mobiles. Cela peut faciliter les interactions lipides-protéines, en particulier pour les lipides chargés.

Plusieurs observations suggèrent que les lipides anioniques peuvent faciliter l'activation des RCPG [16], [57] Nous avons testé cette hypothèse en insérant AT1 et AT2 dans une bicouche de DOPG.

AT2 est facilement activable en absence de l'ion sodium dans une bicouche lipidique anionique. En effet, nous avons observé des activations pour des valeurs de tension de 20 et 50 dyne/cm et même en absence de tension.

Concernant AT1, alors qu'en l'absence de tension, sa structure reste stable, l'application d'un étirement de 20 dyne/cm lorsque AT1 est inséré dans une bicouche de DOPG induit un changement conformationnel (Figure 52).

Une analyse détaillée de la transition conformationnelle révèle plusieurs étapes :

1. L'activation est initiée par l'étirement de H8 dû à l'interaction d'un résidu Lys (K294) avec une molécule de DOPG. L'étirement de H8 est couplé à un léger mouvement de TM7 vers l'extérieur.
2. Ce mouvement permet l'entrée d'eau à l'intérieur du récepteur, induisant l'ouverture progressive de TM6 avec une augmentation de la distance entre TM1 et TM6 de 22 à 34 Å.
3. L'ouverture de la TM7 est maintenue par des liaisons entre H8 et DOPG.

Ces résultats mettent en évidence le rôle crucial de l'hélice 8 dans l'activation des RCPG par le stress mécanique. Conformément avec les résultats de Erdogmus *et al.*, l'hélice 8 d'AT1 est un motif structural essentiel pour détecter les forces mécaniques et les convertir en réponses moléculaires puis vasculaires, contribuant ainsi à l'autorégulation du tonus des myogénique [72].

Quelles sont les étapes du mécanisme d'activation d'AT1 et d'AT2 affectées par les facteurs externes que sont le stress mécanique et les lipides de l'environnement ?

Nos simulations dynamiques classiques ont permis de mettre en évidence l'importance du rôle des lipides dans la régulation des récepteurs de l'AngII. Nous avons montré que, sous tension, les DOPG facilitent la transition d'AT1 vers une structure non canonique, similaire à celle observée dans la structure cristallographique du récepteur NRTR1, soutenant fortement l'hypothèse que la transition observée correspond bien à une étape intermédiaire de l'activation du récepteur (Figure 71).

Effet de l'environnement lipidique sur l'activation d'AT1

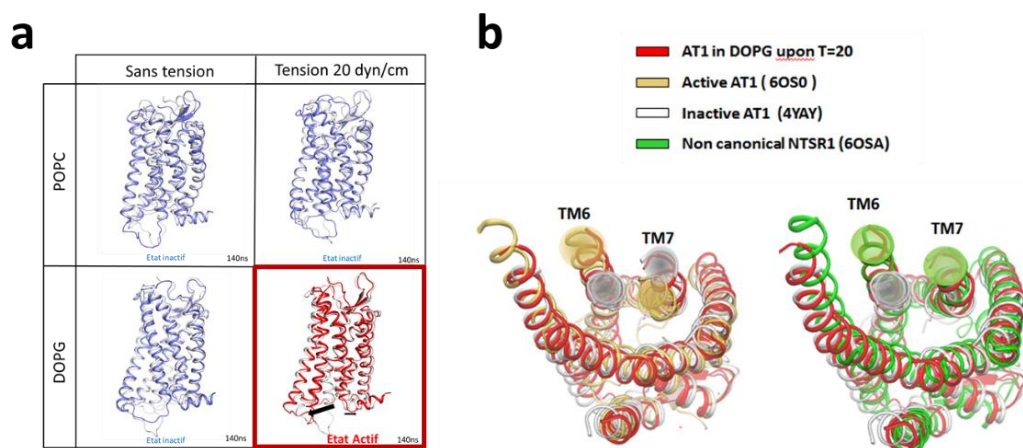


Figure 71: Effet des lipides anioniques sur AT1.

(a) Comparaison des réponses d'AT1 à un stress mécanique dans une bicouche de POPC ou de DOPG. On observe une transition vers un état actif uniquement dans DOPG en présence de stress mécanique (20 dyne/cm) ; (b) Comparaison de la structure obtenue (rouge) avec AT1 inactive (blanc), AT1 active (jaune) et la structure non-canonique de NTSR1 (vert). La structure d'AT1 dans une bicouche de DOPG sous tension se superpose à la structure active non-canonique du récepteur de la neurotensine 1 (NTSR1). Elle se caractérise par une ouverture vers l'extérieur des hélices 6 et 7.

Cette structure est aussi observé par Suomivuori et al [143] dans certaines de leurs simulations dynamiques où TM7 adopte parfois une conformation de type inactif, avec le motif NPXXY plus éloigné du faisceau hélicoïdal, tandis que l'extrémité intracellulaire de TM6 reste dans la position extérieure des conformations actives et alternatives canoniques. Une autre étude récente sur le récepteur AT1 montre que lors de simulations longues (300 μ s), un état intermédiaire clé est identifié dans la voie d'activation d'AT1 [147]. Dans cet état intermédiaire, un site potentiel de régulation allostérique « P6 » caché entre TM7 et H8 est observable lors du mouvement ascendant de H8 [147] (Figure 72). La perturbation directe de ce site par des mutations ponctuelles *in silico* ou *in vitro* a influencé le site d'interaction des protéines G et de la β -arrestine perturbant ainsi leur activité [147].

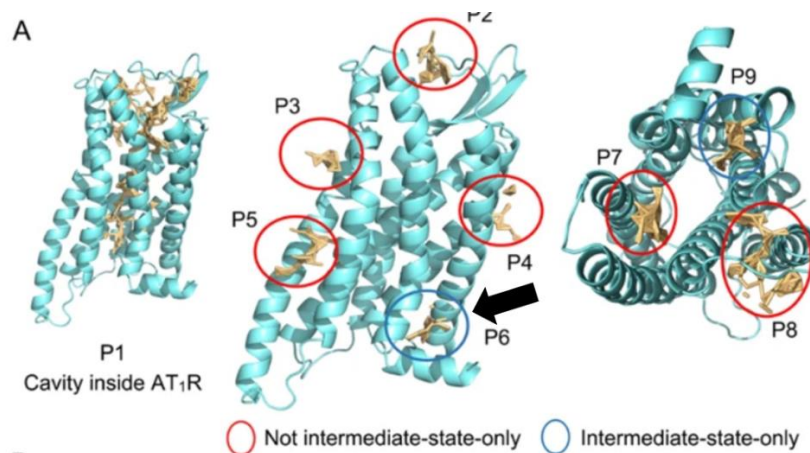


Figure 72 : Sites allostériques potentiels d'AT1.

Ces sites ont été identifiés chez AT1 dans l'état intermédiaire et hors de l'état intermédiaire d'activation [147].

Les simulations de dynamique moléculaire dirigées valident la présence d'une structure intermédiaire chez AT1 et expliquent les étapes mécanistiques de l'activation des récepteurs AT1 et AT2 ainsi que la déstabilisation apportée par les DOPG et le stress mécanique. En effet, les SMD mettent en évidence que la structure non canonique d'AT1 est une structure intermédiaire de l'activation et que les interactions électrostatiques entre TM6 et TM7 (D6.32 et K7.58) contribuent à l'initiation de cette structure (Figure S6).

Ces interactions peuvent être perturbées par la nature des lipides environnants et le stress mécanique, facilitant l'activation du récepteur AT1 dans ces conditions. En effet les lipides anioniques peuvent agir sur les hélices TM7 et H8, en particulier au niveau des résidus chargés positivement (K7.58 et K7.61). La réorganisation des liaisons H, entraînant la rupture de l'interaction entre D6.32 et K7.58, pourrait ainsi faciliter l'activation en présence de lipides anioniques. De même, l'activation du récepteur AT1 par les forces d'étirement pourrait être expliquée par l'étirement de H8 perturbant les interactions électrostatiques entre TM6 et TM7 et facilitant ainsi l'activation.

Contrairement à AT1, AT2 ne montre pas un état intermédiaire d'activation. Cela pourrait faciliter l'activation d'AT2 en absence de l'angiotensine II.

Quelles sont les différences entre AT1 et AT2 qui peuvent jouer un rôle dans la réponse différentielle de ces deux récepteurs aux facteurs externes ?

AT1 et AT2 partagent approximativement 30% d'homologie. Bien qu'ils couplent le même agoniste naturel, l'angiotensine II, la réponse physiologique est différente et même opposée [148]. Dans nos simulations dynamiques, nous avons remarqué qu'AT2 est plus facilement activable qu'AT1 par les conditions externes, comme il est plus facilement activable par des dérivés de l'AngII [148].

Cette réponse différentielle entre les deux récepteurs peut être expliquée au moins en partie par la différence au niveau du verrouillage ionique ou « ionic lock » du motif DRY. Pour AT1, le verrouillage ionique implique D7.49, R3.50 et N6.30 qui forme des liaisons hydrogène entre les hélices 3 et 6. Pour AT2, le verrouillage ionique est composé de 2 résidus seulement, D3.49 et R3.50, car le résidu 6.30 est une arginine. Il y a donc une répulsion entre R3.50 et R6.30 qui peut favoriser la libération spontanée de l'hélice TM6 (Figure 73).

Une deuxième différence se situe au niveau du site de fixation du sodium. Le résidu 7.46 est une serine pour AT2 et une asparagine pour AT1. Cela peut contribuer à l'activation d'AT2 sans la sortie du sodium mais avec une modification de sa coordination permise par l'arrivée de molécules d'eau. La coordination entre le sodium et D2.50 s'affaiblit avec l'effet d'écran des molécules d'eau et la coordination avec S3.39 est favorisée. Ce phénomène observé dans le cas d'AT2 lors de pré activation ou d'activation est similaire à celui observé lors de l'activation de CXCR4 [146]. Il révèle que la pénétration de l'eau et les changements du mode de liaison du sodium sont des facteurs clés du mécanisme d'activation.

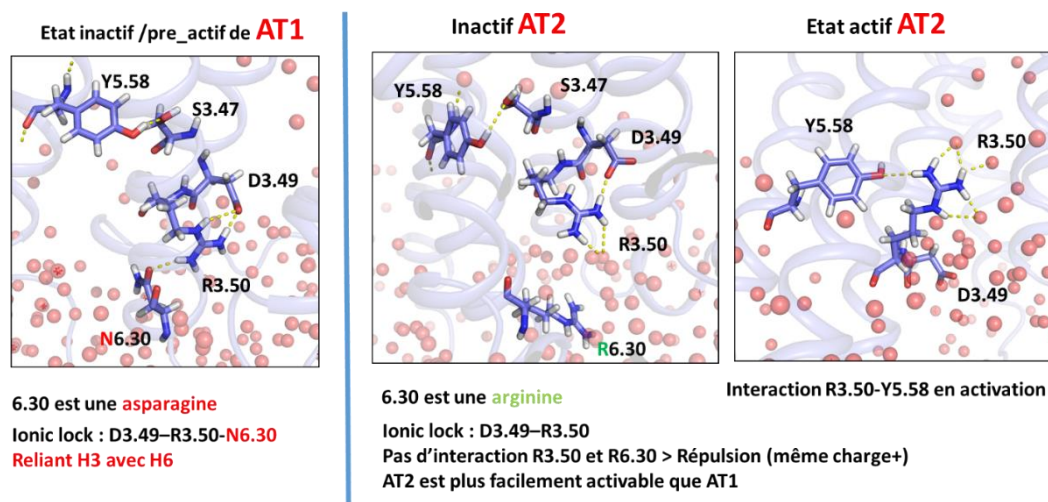


Figure 73 : Comparaison du verrouillage ionique entre AT1 et AT2

Une 3ème différence réside dans les interactions entre les hélices 6 et 7. Chez AT1, la forte interaction électrostatique entre les résidus D6.32 et K7.58 semble jouer un rôle dans la stabilisation de la structure inactive ainsi que dans les étapes mécanistiques d'activation et l'apparition de la structure non canonique. Les liaisons H entre ces résidus ne sont pas significativement modifiées par les conditions externes (DOPG/tension) et sont similaires entre les récepteurs AT1 humain et d'anguille qui partagent les mêmes charge au niveau de ces positions (Figure 74a). Cependant, dans le cas du récepteur d'anguille, la présence d'une arginine en position 7.58 augmente significativement les interactions de ce résidu avec les lipides. De même, la présence des lipides anioniques DOPG augmentent l'interaction du résidu 7.58 avec les lipides (Figure 74b). Cela peut perturber les mouvements de TM7 vers des structures non canoniques comme observés dans les SMD (Figure 69).

Concernant l'effet de la tension sur les liaisons H entre D6.32 et K7.58, bien que les moyennes ne soient pas significativement différentes du fait des larges déviations standard observées en présence de tension (Figure 74a), nous pouvons noter que cette dispersion même du taux de liaisons H pourrait être corrélée à la variabilité des transitions conformationnelles aboutissant à une structure non canonique finale dans plusieurs replicas (Figure 70).

Dans le cas d'AT2, des liaisons H entre les résidus 6.32 et 7.58 sont absentes (Figure 74a). Cependant, le résidu Q6.32 interagit fortement avec les POPC, ce qui pourrait faciliter le mouvement en avant de TM6 et l'activation du récepteur (Figure 74c).

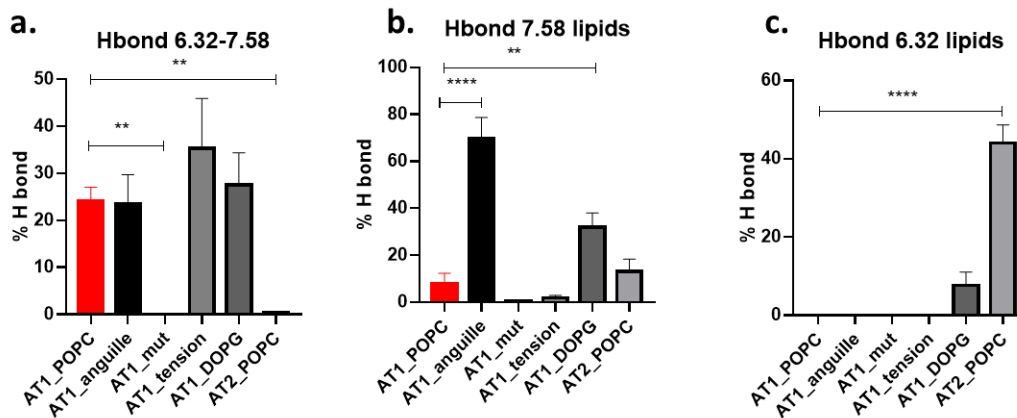


Figure 74 : Liaisons H des résidus 6.32 et 7.58 dans les simulations dynamiques dirigées.

(a) Taux de liaisons H entre les résidus 6.32 et 7.58; (b) Taux de liaisons H entre le résidu 7.58 et les lipides ; (c) Taux de liaisons H entre le résidu 6.32 et les lipides. Les liaisons H ont été mesurées lors des SMD (n=6). Les histogrammes présentent les moyennes et l'erreur standard obtenue à partir des moyennes des 6 réplicas par condition. Les résultats sont comparés à AT1 dans une bicouche de POPC. Test statistique Anova à 2 facteurs, test Dunnett avec $\alpha = 0.05$.

Finalement, l'analyse des séquences des récepteurs AT1 et AT2 terrestres et aquatiques montre que, dans le cas d'AT1, l'acide aspartique en position 6.32 et la lysine en position 7.58 sont apparus avec la transition à la vie terrestre. Ces résidus sont rares à ces positions chez les RCPG de classe A humains, suggérant une spécificité du récepteur AT1 terrestre. Dans le cas d'AT2, il n'y a pas de différence entre les récepteurs des animaux terrestres et aquatiques avec l'asparagine en position 7.58 et une lysine ou une arginine en position 6.32, un profil qui semble aussi assez fréquent dans la class A des RCPG humain (Figure 75).

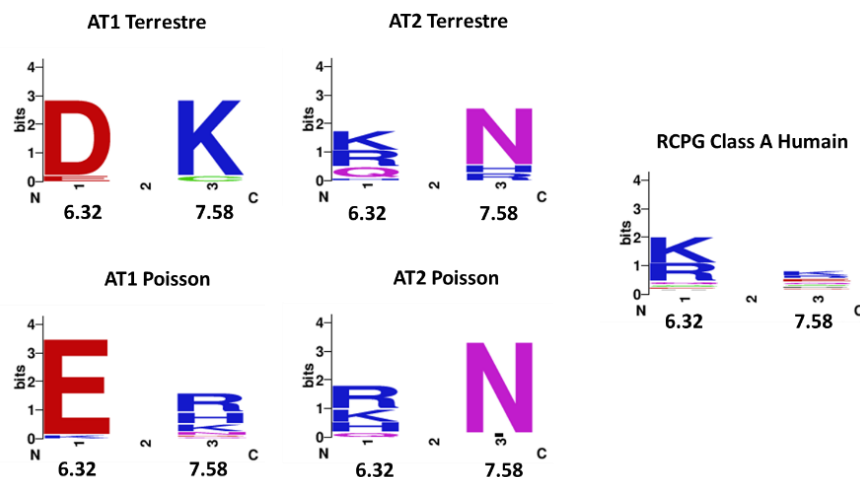


Figure 75 : Logoplots des résidus 6.32 et 7.58 des récepteurs AT1 et AT2 terrestres et aquatiques.

Le logoplot de ces positions pour un ensemble non-redondant de RCPG humains est donné pour comparaison.

Les réponses différentielles de ces deux récepteurs peuvent être importantes pour leur mécanisme d'action *in vivo*, où les facteurs mécaniques et la composition lipidique peuvent être modifiés dans différentes conditions physiopathologiques.

Conclusion & perspectives

L'objectif de ce projet est d'approfondir la compréhension des mécanismes moléculaires conduisant à l'activation des récepteurs AT1 et AT2 en absence du ligand naturel, l'angiotensine II, en particulier par l'effet du stress mécanique, et d'éclaircir le rôle de l'eau, de l'ion sodium et de la composition lipidique dans les changements conformationnel de ces récepteurs.

Pour répondre à cet objectif, nous avons combiné des méthodes de simulations dynamiques classiques et dirigées.

Les simulations dynamiques classiques montrent des changements conformationnels fréquents du récepteur AT2 dans différentes conditions. Pour AT1, étant plus stable, l'observation des changements conformationnels est plus rare, aléatoire, et nécessite des conditions plus drastiques. Des conformations actives non canoniques d'AT1 sont observées uniquement dans une bicouche de DOPG, lors de l'application soit d'une tension de 50 dyne/cm lorsqu'AT1 est lié au sodium, soit d'une tension plus faible de 20 dyne/cm en l'absence de sodium.

Les simulations dirigées que nous avons réalisées complètent nos simulations de dynamique moléculaire classique et fournissent un cadre reproductible pour étudier l'influence de facteurs externes.

En particulier, cette méthode révèle le mécanisme d'activation d'AT1 et l'importance de la liaison D6.32-K7.58 dans la stabilisation du récepteur. Cette liaison étant positionnée du côté intracellulaire des hélices TM6 et TM7, elle peut être perturbée par différentes facteurs externes tel que les lipides anioniques et l'étirement de l'hélice H8.

De plus, ces résultats fournissent une compréhension plus profonde de l'activation du récepteur AT1 à un niveau atomique et suggèrent des pistes pour la conception de modulateurs allostériques du récepteur AT1 avec un large éventail d'applications dans le domaine des RCPG.

Pour aller plus loin, il sera intéressant de tester l'effet d'autres lipides tel que le cholestérol et l'oméga3 ainsi que de tester des conditions plus physiologiques et l'effet des ligands (angiotensine II), des antagonistes (Canderstan) ou des agonistes biaisés vers les betas arrestines (TRV023/026) en présence du stress mécanique.

Il conviendra aussi, pour AT1, d'étudier la stabilité de la conformation non-canonique par des simulations de dynamiques moléculaires extensives, en présence ou non d'angiotensine II. Enfin, il faudra aussi étudier diverses compositions membranaires (en particulier des taux de DOPG variables) sur la transition vers la conformation non-canonique et la stabilité de celle-ci

En effet, la mise en évidence d'une conformation stable d'AT1, observable uniquement de certaines conditions (stress mécanique, présence accrue de lipides anioniques dans l'environnement) pourrait permettre de cibler le récepteur AT1 dans certains environnements pathologiques.

Il conviendra aussi d'étudier plus profondément le récepteur AT2, pour comprendre comment des facteurs externes agissant différenciellement sur AT1 et AT2 peuvent contribuer à déréguler le système rénine-angiotensine dans des conditions pathologiques.

Il convient aussi de faire des études expérimentales pour tester l'implication des résidus D6.32-K7.58 dans la stabilité du récepteur à l'état inactif par des mutations

Bibliographie

- [1] K. Sriram and P. A. Insel, "G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs?," *Mol. Pharmacol.*, vol. 93, no. 4, pp. 251–258, 2018, doi: 10.1124/MOL.117.111062.
- [2] K. Azushima, N. Morisawa, K. Tamura, and A. Nishiyama, "Recent Research Advances in Renin-Angiotensin-Aldosterone System Receptors," *Current Hypertension Reports*, vol. 22, no. 3. Curr Hypertens Rep, Mar. 01, 2020. doi: 10.1007/s11906-020-1028-6.
- [3] K. Nakamura, R. Okuyama, and Y. Kawakami, "Renin-angiotensin system in the tumor microenvironment," in *Advances in Experimental Medicine and Biology*, vol. 1277, Springer, 2020, pp. 105–114. doi: 10.1007/978-3-030-50224-9_7.
- [4] K. Verma, M. Pant, S. Paliwal, J. Dwivedi, *et al.*, "An Insight on Multicentric Signaling of Angiotensin II in Cardiovascular system: A Recent Update," *Frontiers in Pharmacology*, vol. 12. Frontiers Media S.A., p. 2175, Aug. 20, 2021. doi: 10.3389/fphar.2021.734917.
- [5] T. Kawai, S. J. Forrester, S. O'Brien, A. Baggett, *et al.*, "AT1 receptor signaling pathways in the cardiovascular system," *Pharmacol. Res.*, vol. 125, pp. 4–13, 2017, doi: 10.1016/J.PHRS.2017.05.008.
- [6] T. Kawai, S. J. Forrester, S. O'Brien, A. Baggett, *et al.*, "AT1 receptor signaling pathways in the cardiovascular system," *Pharmacol. Res.*, vol. 125, no. Pt A, p. 4, 2017, doi: 10.1016/J.PHRS.2017.05.008.
- [7] R. M. Carey, "AT2 receptors: Potential therapeutic targets for hypertension," *American Journal of Hypertension*, vol. 30, no. 4. Oxford Academic, pp. 339–347, Apr. 01, 2017. doi: 10.1093/ajh/hpw121.
- [8] A. Sadybekov and V. Katritch, "Breaking the Enigma Code of Angiotensin II Type 2 Receptor Signaling," vol. 28, no. 4, pp. 390–392, 2020, doi: 10.1016/j.str.2020.03.004.
- [9] L. M. Wingler, M. A. Skiba, C. McMahon, D. P. Staus, *et al.*, "STRUCTURAL BIOLOGY Angiotensin and biased analogs induce structurally distinct active conformations within a GPCR," American Association for the Advancement of Science, 2020. doi: 10.1126/science.aay9813.
- [10] Y. Zou, H. Akazawa, Y. Qin, M. Sano, *et al.*, "Mechanical stress activates angiotensin II type 1 receptor without the involvement of angiotensin II," *Nat. Cell Biol.*, vol. 6, no. 6, pp. 499–506, 2004, doi: 10.1038/ncb1137.
- [11] V. G. Barauna, F. C. Magalhaes, L. C. G. Campos, R. I. Reis, *et al.*, "Shear stress-induced Ang II AT1 receptor activation: G-protein dependent and independent mechanisms," *Biochem. Biophys. Res. Commun.*, vol. 434, no. 3, pp. 647–652, 2013, doi: 10.1016/j.bbrc.2013.04.005.
- [12] U. Storch, M. M. y Schnitzler, and T. Gudermann, "G protein-mediated stretch reception," *Am. J. Physiol. Circ. Physiol.*, vol. 302, no. 6, pp. H1241–H1249, 2012, doi: 10.1152/ajpheart.00818.2011.
- [13] W. Li, X. Yu, F. Xie, B. Zhang, *et al.*, "A Membrane-Bound Biosensor Visualizes Shear Stress-Induced Inhomogeneous Alteration of Cell Membrane Tension," *iScience*, vol. 7, pp. 180–190, 2018, doi: 10.1016/j.isci.2018.09.002.
- [14] C. Neale, H. D. Herce, R. Pomès, and A. E. García, "Can Specific Protein-Lipid Interactions Stabilize an Active State of the Beta 2 Adrenergic Receptor?," *Biophys. J.*, vol. 109, no. 8, pp. 1652–1662, 2015, doi: 10.1016/j.bpj.2015.08.028.
- [15] S. Inagaki, R. Ghirlando, J. F. White, J. Gvozdenovic-Jeremic, *et al.*, "Modulation of the interaction

- between neurotensin receptor NTS1 and Gq protein by lipid," *J. Mol. Biol.*, vol. 417, no. 1–2, pp. 95–111, 2012, doi: 10.1016/j.jmb.2012.01.023.
- [16] A. Bruzzese, J. A. R. Dalton, and J. Giraldo, "Insights into adenosine A2A receptor activation through cooperative modulation of agonist and allosteric lipid interactions," *PLoS Comput. Biol.*, vol. 16, no. 4, 2020, doi: 10.1371/journal.pcbi.1007818.
- [17] D. M. Rosenbaum, S. G. F. Rasmussen, and B. K. Kobilka, "The structure and function of G-protein-coupled receptors," *Nature*, vol. 459, no. 7245, p. 356, 2009, doi: 10.1038/NATURE08144.
- [18] D. Wacker, R. C. Stevens, and B. L. Roth, "How ligands illuminate GPCR molecular pharmacology," *Cell*, vol. 170, no. 3, p. 414, 2017, doi: 10.1016/J.CELL.2017.07.009.
- [19] V. Katritch, V. Cherezov, and R. C. Stevens, "Diversity and modularity of G protein-coupled receptor structures," *Trends Pharmacol. Sci.*, vol. 33, no. 1, pp. 17–27, 2012, doi: 10.1016/J.TIPS.2011.09.003.
- [20] D. Kamato, L. Thach, R. Bernard, V. Chan, *et al.*, "Structure, Function, Pharmacology, and Therapeutic Potential of the G Protein, G α /q,11," *Front. Cardiovasc. Med.*, vol. 2, p. 14, 2015, doi: 10.3389/FCVM.2015.00014/BIBTEX.
- [21] E. Ghosh, P. Kumari, D. Jaiman, and A. K. Shukla, "Methodological advances: The unsung heroes of the GPCR structural revolution," *Nature Reviews Molecular Cell Biology*, vol. 16, no. 2. Nature Publishing Group, pp. 69–81, Jan. 15, 2015. doi: 10.1038/nrm3933.
- [22] R. Fredriksson, M. C. Lagerström, L. G. Lundin, and H. B. Schiöth, "The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints," vol. 63, no. 6, pp. 1256–1272, 2003, doi: 10.1124/MOL.63.6.1256.
- [23] H. B. Schiöth and R. Fredriksson, "The GRAFS classification system of G-protein coupled receptors in comparative perspective," in *General and Comparative Endocrinology*, May 2005, vol. 142, no. 1–2 SPEC. ISS., pp. 94–101. doi: 10.1016/j.ygcen.2004.12.018.
- [24] N. Tuteja, "Signaling through G protein coupled receptors," *Plant Signaling and Behavior*, vol. 4, no. 10. Taylor & Francis, pp. 942–947, 2009. doi: 10.4161/psb.4.10.9530.
- [25] H. C. Huang and P. S. Klein, "The Frizzled family: receptors for multiple signal transduction pathways," *Genome Biol.*, vol. 5, no. 7, p. 234, 2004, doi: 10.1186/GB-2004-5-7-234.
- [26] Q. Zhou, D. Yang, M. Wu, Y. Guo, *et al.*, "Common activation mechanism of class a GPCRs," *Elife*, vol. 8, 2019, doi: 10.7554/eLife.50279.
- [27] J. Pelé, J. M. Bécu, H. Abdi, and M. Chabbert, "Bios2mds: An R package for comparing orthologous protein families by metric multidimensional scaling," *BMC Bioinformatics*, vol. 13, no. 1, pp. 1–7, 2012, doi: 10.1186/1471-2105-13-133.
- [28] J. Pelé, H. Abdi, M. Moreau, D. Thybert, *et al.*, "Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors," *PLoS One*, vol. 6, no. 4, pp. 1–10, 2011, doi: 10.1371/journal.pone.0019094.
- [29] J. Devillé, J. Rey, and M. Chabbert, "An indel in transmembrane helix 2 helps to trace the molecular evolution of class A G-protein-coupled receptors," *J. Mol. Evol.*, vol. 68, no. 5, pp. 475–489, 2009, doi: 10.1007/s00239-009-9214-9.
- [30] A. Tiss, R. Ben Boubaker, D. Henrion, H. Guissouma, *et al.*, "Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom," in *Methods in Molecular Biology*, vol. 2315,

Methods Mol Biol, 2021, pp. 73–97. doi: 10.1007/978-1-0716-1468-6_5.

- [31] G. M. Hu, T. L. Mai, and C. M. Chen, "Visualizing the GPCR Network: Classification and Evolution," *Sci. Rep.*, vol. 7, no. 1, pp. 1–15, 2017, doi: 10.1038/s41598-017-15707-9.
- [32] J. A. Ballesteros and H. Weinstein, "Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors," *Methods Neurosci.*, vol. 25, no. C, pp. 366–428, 1995, doi: 10.1016/S1043-9471(05)80049-7.
- [33] V. Isberg, C. De Graaf, A. Bortolato, V. Cherezov, *et al.*, "Generic GPCR residue numbers - Aligning topology maps while minding the gaps," *Trends in Pharmacological Sciences*, vol. 36, no. 1. NIH Public Access, pp. 22–31, 2015. doi: 10.1016/j.tips.2014.11.001.
- [34] N. R. Latorraca, A. J. Venkatakrisnan, and R. O. Dror, "GPCR dynamics: Structures in motion," *Chemical Reviews*, vol. 117, no. 1. pp. 139–155, 2017. doi: 10.1021/acs.chemrev.6b00177.
- [35] B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, *et al.*, "Action of Molecular Switches in GPCRs - Theoretical and Experimental Studies," *Curr. Med. Chem.*, vol. 19, no. 8, p. 1090, 2012, doi: 10.2174/092986712799320556.
- [36] H. E. Kato, Y. Zhang, H. Hu, C. M. Suomivuori, *et al.*, "Conformational transitions of a neurotensin receptor 1-Gi1 complex," *Nature*, vol. 572, no. 7767, pp. 80–85, 2019, doi: 10.1038/s41586-019-1337-6.
- [37] R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, *et al.*, "Activation mechanism of the β 2-adrenergic receptor," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 46, pp. 18684–18689, 2011, doi: 10.1073/pnas.1110499108.
- [38] A. J. Venkatakrisnan, A. K. Ma, R. Fonseca, N. R. Latorraca, *et al.*, "Diverse GPCRs exhibit conserved water networks for stabilization and activation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 8, pp. 3288–3293, 2019, doi: 10.1073/pnas.1809251116.
- [39] É. Bertalan, S. Lešnik, U. Bren, and A. N. Bondar, "Protein-water hydrogen-bond networks of G protein-coupled receptors: Graph-based analyses of static structures and molecular dynamics," *J. Struct. Biol.*, vol. 212, no. 3, p. 107634, 2020, doi: 10.1016/J.JSB.2020.107634.
- [40] M. Louet, M. Casiraghi, M. Damian, M. G. Costa, *et al.*, "Concerted conformational dynamics and water movements in the ghrelin G protein-coupled receptor," *Elife*, vol. 10, 2021, doi: 10.7554/ELIFE.63201.
- [41] W. H. Weng, Y. T. Li, and H. J. Hsu, "Activation-Induced Conformational Changes of Dopamine D3 Receptor Promote the Formation of the Internal Water Channel," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, 2017, doi: 10.1038/s41598-017-13155-z.
- [42] S. Yuan, S. Filipek, K. Palczewski, and H. Vogel, "Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway," *Nat. Commun.*, vol. 5, no. 1, pp. 1–10, 2014, doi: 10.1038/ncomms5733.
- [43] B. Zarzycka, S. A. Zaidi, B. L. Roth, and V. Katritch, "Harnessing ion-binding sites for GPCR pharmacology," *Pharmacol. Rev.*, vol. 71, no. 4, pp. 571–595, 2019, doi: 10.1124/pr.119.017863.
- [44] V. Katritch, G. Fenalti, E. E. Abola, B. L. Roth, *et al.*, "Allosteric sodium in class A GPCR signaling," *Trends Biochem. Sci.*, vol. 39, no. 5, pp. 233–244, 2014, doi: 10.1016/j.tibs.2014.03.002.
- [45] C. B. Pert, G. Pasternak, and S. H. Snyder, "Opiate agonists and antagonists discriminated by receptor binding in brain," *Science (80-.)*, vol. 182, no. 4119, pp. 1359–1361, 1973, doi:

10.1126/science.182.4119.1359.

- [46] C. Bihoreau, C. Monnot, E. Davies, B. Teutsch, *et al.*, "Mutation of Asp74 of the rat angiotensin II receptor confers changes in antagonist affinities and abolishes G-protein coupling.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 11, p. 5133, 1993, doi: 10.1073/PNAS.90.11.5133.
- [47] A. Massink, H. Gutiérrez-De-Terán, E. B. Lenselink, N. V. O. Zacarías, *et al.*, "Sodium ion binding pocket mutations and adenosine A2A receptor function," *Mol. Pharmacol.*, vol. 87, no. 2, pp. 305–313, 2015, doi: 10.1124/mol.114.095737.
- [48] J. D. Violin, D. G. Soergel, G. Boerrigter, J. C. Burnett, *et al.*, "Conformational switch of angiotensin II type 1 receptor underlying mechanical stress-induced activation," *EMBO Rep.*, vol. 23, no. 12, pp. 233–244, 2016, doi: 10.1111/micc.12293.
- [49] G. Fenalti, P. M. Giguere, V. Katritch, X. P. Huang, *et al.*, "Molecular control of δ -opioid receptor signalling," *Nature*, vol. 506, no. 7487, pp. 191–196, 2014, doi: 10.1038/nature12944.
- [50] B. Selvam, Z. Shamsi, and D. Shukla, "Universality of the Sodium Ion Binding Mechanism in Class A G-Protein-Coupled Receptors," *Angew. Chemie - Int. Ed.*, 2018, doi: 10.1002/anie.201708889.
- [51] P. Ghanouni, H. Schambye, R. Seifert, T. W. Lee, *et al.*, "The effect of pH on β_2 adrenoceptor function. Evidence for protonation- dependent activation," *J. Biol. Chem.*, vol. 275, no. 5, pp. 3121–3127, 2000, doi: 10.1074/jbc.275.5.3121.
- [52] B. Taddese, M. Deniaud, A. Garnier, A. Tiss, *et al.*, "Evolution of chemokine receptors is driven by mutations in the sodium binding site," *PLoS Comput. Biol.*, vol. 14, no. 6, pp. 1–25, 2018, doi: 10.1371/journal.pcbi.1006209.
- [53] R. Dawaliby, C. Trubbia, C. Delporte, M. Masureel, *et al.*, "Allosteric regulation of G protein-coupled receptor activity by phospholipids," *Nat. Chem. Biol.*, vol. 12, no. 1, pp. 35–39, 2016, doi: 10.1038/nchembio.1960.
- [54] A. J. Y. Jones, F. Gabriel, A. Tandale, and D. Nietlispach, "Structure and Dynamics of GPCRs in Lipid Membranes: Physical Principles and Experimental Approaches," *Molecules*, vol. 25, no. 20, 2020, doi: 10.3390/MOLECULES25204729.
- [55] J. Jakubík and E. E. El-Fakahany, "Allosteric Modulation of GPCRs of Class A by Cholesterol," *Int. J. Mol. Sci.*, vol. 22, no. 4, pp. 1–18, 2021, doi: 10.3390/IJMS22041953.
- [56] B. I. Sejdiu and D. P. Tieleman, "Lipid-Protein Interactions Are a Unique Property and Defining Feature of G Protein-Coupled Receptors," *Biophys. J.*, vol. 118, no. 8, pp. 1887–1900, 2020, doi: 10.1016/J.BPJ.2020.03.008.
- [57] A. Bruzzese, C. Gil, J. A. R. Dalton, and J. Giraldo, "Structural insights into positive and negative allosteric regulation of a G protein-coupled receptor through protein-lipid interactions," *Sci. Rep.*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-22735-6.
- [58] D. S. Im, "Omega-3 fatty acids in anti-inflammation (pro-resolution) and GPCRs," *Progress in Lipid Research*, vol. 51, no. 3. Elsevier Ltd., pp. 232–237, Jul. 2012. doi: 10.1016/j.plipres.2012.02.003.
- [59] R. Guixà-González, M. Javanainen, M. Gómez-Soler, B. Cordobilla, *et al.*, "Membrane omega-3 fatty acids modulate the oligomerisation kinetics of adenosine A2A and dopamine D2 receptors," *Sci. Rep.*, vol. 6, no. January, pp. 1–13, 2016, doi: 10.1038/srep19839.
- [60] D. Y. Oh, S. Talukdar, E. J. Bae, T. Imamura, *et al.*, "GPR120 Is an Omega-3 Fatty Acid Receptor

- Mediating Potent Anti-inflammatory and Insulin-Sensitizing Effects," *Cell*, vol. 142, no. 5, pp. 687–698, 2010, doi: 10.1016/j.cell.2010.07.041.
- [61] C. P. Cardozo, "Mechanotransduction: Overview," in *Encyclopedia of Bone Biology*, Academic Press, 2020, p. 217. doi: 10.1016/b978-0-12-801238-3.62233-x.
- [62] J. M. Kefauver, A. B. Ward, and A. Patapoutian, "Discoveries in structure and physiology of mechanically activated ion channels," *Nature*, vol. 587, no. 7835. Nature Publishing Group, pp. 567–576, Nov. 25, 2020. doi: 10.1038/s41586-020-2933-1.
- [63] C. S. Pereira and P. H. Hünenberger, "Effect of trehalose on a phospholipid membrane under mechanical stress," *Biophys. J.*, vol. 95, no. 8, pp. 3525–3534, 2008, doi: 10.1529/biophysj.108.131656.
- [64] S. Marullo, S. Doly, K. Saha, H. Enslin, *et al.*, "Mechanical GPCR Activation by Traction Forces Exerted on Receptor N-Glycans," *ACS Pharmacology and Translational Science*, vol. 3, no. 2. American Chemical Society, pp. 171–178, Apr. 10, 2020. doi: 10.1021/acsptsci.9b00106.
- [65] J. ichi Sadoshima, Y. Xu, H. S. Slayter, and S. Izumo, "Autocrine release of angiotensin II mediates stretch-induced hypertrophy of cardiac myocytes in vitro," *Cell*, vol. 75, no. 5, pp. 977–984, 1993, doi: 10.1016/0092-8674(93)90541-W.
- [66] M. Chachivilis, Y. L. Y.-L. Zhang, and J. A. Frangos, "G protein-coupled receptors sense fluid shear stress in endothelial cells," *Proc. Natl. Acad. Sci.*, vol. 103, no. 42, pp. 15463–15468, 2006, doi: 10.1073/pnas.0607224103.
- [67] J. Wang, K. Hanada, C. Gareri, and H. A. Rockman, "Mechanoactivation of the angiotensin II type 1 receptor induces β -arrestin-biased signaling through Gai coupling," *J. Cell. Biochem.*, vol. 119, no. 4, pp. 3586–3597, 2018, doi: 10.1002/jcb.26552.
- [68] M. Mederos Y Schnitzler, U. Storch, and T. Gudermann, "AT1 receptors as mechanosensors," *Curr. Opin. Pharmacol.*, vol. 11, no. 2, pp. 112–116, 2011, doi: 10.1016/j.coph.2010.11.003.
- [69] S. Abdul-Majeed and S. M. Nauli, "Dopamine receptor type 5 in the primary cilia has dual chemo- and mechano-sensory roles," *Hypertens. (Dallas, Tex. 1979)*, vol. 58, no. 2, pp. 325–331, 2011, doi: 10.1161/HYPERTENSIONAHA.111.172080.
- [70] M. M. Y Schnitzler, U. Storch, S. Meibers, P. Nurwakagari, *et al.*, "Gq-coupled receptors as mechanosensors mediating myogenic vasoconstriction," vol. 27, no. 23, 2008, doi: 10.1038/EMBOJ.2008.233.
- [71] A. Makino, E. R. Prossnitz, M. Bünemann, M. W. Ji, *et al.*, "G protein-coupled receptors serve as mechanosensors for fluid shear stress in neutrophils," *Am. J. Physiol. Cell Physiol.*, vol. 290, no. 6, 2006, doi: 10.1152/AJPCELL.00576.2005.
- [72] S. Erdogmus, U. Storch, L. Danner, J. Becker, *et al.*, "Helix 8 is the essential structural motif of mechanosensitive GPCRs," *Nat. Commun.*, vol. 10, no. 1, pp. 1–15, 2019, doi: 10.1038/s41467-019-13722-0.
- [73] Y. L. Zhang, J. A. Frangos, and M. Chachivilis, "Mechanical stimulus alters conformation of type 1 parathyroid hormone receptor in bone cells," *Am. J. Physiol. - Cell Physiol.*, vol. 296, no. 6, p. C1391, 2009, doi: 10.1152/ajpcell.00549.2008.
- [74] A. D. Ozkan, T. Gettas, A. Sogata, W. Phaychanpheng, *et al.*, "Mechanical and chemical activation of GPR68 probed with a genetically encoded fluorescent reporter," *J. Cell Sci.*, vol. 134, no. 16, 2021, doi:

10.1242/JCS.255455.

- [75] J. Xu, J. Mathur, E. Vessières, S. Hammack, *et al.*, "GPR68 senses flow and is essential for vascular physiology," *Cell*, vol. 173, no. 3, p. 762, 2018, doi: 10.1016/J.CELL.2018.03.076.
- [76] R. Busch, A. Strohbach, M. Pennewitz, F. Lorenz, *et al.*, "Regulation of the endothelial apelin/APJ system by hemodynamic fluid flow," *Cell. Signal.*, vol. 27, no. 7, pp. 1286–1296, 2015, doi: 10.1016/J.CELLSIG.2015.03.011.
- [77] C. Sassoli, F. Pierucci, S. Zecchi-Orlandini, and E. Meacci, "Sphingosine 1-phosphate (S1P)/ S1P receptor signaling and mechanotransduction: Implications for intrinsic tissue repair/regeneration," *International Journal of Molecular Sciences*, vol. 20, no. 22. Multidisciplinary Digital Publishing Institute (MDPI), Nov. 02, 2019. doi: 10.3390/ijms20225545.
- [78] U. Stroth and T. Unger, "The renin-angiotensin system and its receptors," in *Journal of Cardiovascular Pharmacology*, 1999, vol. 33, no. SUPPL. 1. doi: 10.1097/00005344-199900001-00005.
- [79] T. Unger, "The role of the renin-angiotensin system in the development of cardiovascular disease," in *American Journal of Cardiology*, Jan. 2002, vol. 89, no. 2 SUPPL. 1, pp. 3–9. doi: 10.1016/S0002-9149(01)02321-9.
- [80] C. Dasgupta and L. Zhang, "Angiotensin II receptors and drug discovery in cardiovascular disease," *Drug Discovery Today*, vol. 16, no. 1–2. NIH Public Access, pp. 22–34, Jan. 2011. doi: 10.1016/j.drudis.2010.11.016.
- [81] R. Ben Boubaker, A. Tiss, D. Henrion, H. Guissouma, *et al.*, "Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota," *PLOS Comput. Biol.*, vol. 18, no. 2, p. e1009732, 2022, doi: 10.1371/journal.pcbi.1009732.
- [82] A. J. Balmforth, A. J. Lee, P. Warburton, D. Donnelly, *et al.*, "The conformational change responsible for AT1 receptor activation is dependent upon two juxtaposed asparagine residues on transmembrane helices III and VII," *J. Biol. Chem.*, vol. 272, no. 7, pp. 4245–4251, 1997, doi: 10.1074/jbc.272.7.4245.
- [83] H. Unal and S. S. Karnik, "Constitutive activity in the angiotensin II type 1 receptor: Discovery and applications," *Adv. Pharmacol.*, vol. 70, pp. 155–174, 2014, doi: 10.1016/B978-0-12-417197-8.00006-7.
- [84] Y. H. Feng, L. Zhou, R. Qiu, and R. Zeng, "Single Mutations at Asn295 and Leu305 in the Cytoplasmic Half of Transmembrane α -Helix Domain 7 of the AT1 Receptor Induce Promiscuous Agonist Specificity for Angiotensin II Fragments: A Pseudo-Constitutive Activity," *Mol. Pharmacol.*, vol. 68, no. 2, pp. 347–355, 2005, doi: 10.1124/MOL.105.011601.
- [85] H. Zhang, H. Unal, R. Desnoyer, G. W. Han, *et al.*, "Structural basis for ligand recognition and functional selectivity at angiotensin receptor," *J. Biol. Chem.*, vol. 290, no. 49, pp. 29127–29139, 2015, doi: 10.1074/jbc.M115.689000.
- [86] G. Turu, A. Balla, and L. Hunyady, "The Role of β -Arrestin Proteins in Organization of Signaling and Regulation of the AT1 Angiotensin Receptor," *Front. Endocrinol. (Lausanne)*, vol. 0, p. 519, 2019, doi: 10.3389/FENDO.2019.00519.
- [87] S. M. Dewire and J. D. Violin, "Biased ligands for better cardiovascular drugs: Dissecting g-protein-coupled receptor pharmacology," *Circ. Res.*, vol. 109, no. 2, pp. 205–216, 2011, doi: 10.1161/CIRCRESAHA.110.231308.

- [88] L. M. Wingler, M. Elgeti, D. Hilger, N. R. Latorraca, *et al.*, "Angiotensin Analogs with Divergent Bias Stabilize Distinct Receptor Conformations," *Cell*, vol. 176, no. 3, pp. 468-478.e11, 2019, doi: 10.1016/j.cell.2018.12.005.
- [89] H. Zhang, H. Unal, C. Gati, G. W. Han, *et al.*, "Structure of the angiotensin receptor revealed by serial femtosecond crystallography," *Cell*, vol. 161, no. 4, pp. 833-844, 2015, doi: 10.1016/j.cell.2015.04.011.
- [90] C. M. Suomivuori, N. R. Latorraca, L. M. Wingler, S. Eismann, *et al.*, "Molecular mechanism of biased signaling in a prototypical G protein-coupled receptor," *Science (80-.)*, vol. 367, no. 6480, pp. 881-887, 2020, doi: 10.1126/science.aaz0326.
- [91] R. M. Carey, X. H. Jin, and H. M. Siragy, "Role of the angiotensin AT2 receptor in blood pressure regulation and therapeutic implications," in *American Journal of Hypertension*, 2001, vol. 14, no. 6 II, pp. 98S-102S. doi: 10.1016/s0895-7061(01)02076-3.
- [92] L. C. Matavelli and H. M. Siragy, "AT2 receptor activities and pathophysiological implications," *J. Cardiovasc. Pharmacol.*, vol. 65, no. 3, pp. 226-232, 2015, doi: 10.1097/FJC.0000000000000208.
- [93] S. S. Karnik, H. Unal, J. R. Kemp, K. C. Tirupula, *et al.*, "International Union of Basic and Clinical Pharmacology. XCIX. Angiotensin Receptors: Interpreters of Pathophysiological Angiotensinergic Stimuli [corrected].," *Pharmacol. Rev.*, vol. 67, no. 4, pp. 754-819, 2015, doi: 10.1124/pr.114.010454.
- [94] S. Miura and S. S. Karnik, "Ligand-independent signals from angiotensin II type 2 receptor induce apoptosis," *EMBO J.*, vol. 19, no. 15, p. 4026, 2000, doi: 10.1093/EMBOJ/19.15.4026.
- [95] E. R. Porrello, L. M. D. Delbridge, and W. G. Thomas, "The angiotensin II type 2 (AT2) receptor: An enigmatic seven transmembrane receptor," *Front. Biosci.*, vol. 14, no. 3, pp. 958-972, 2009, doi: 10.2741/3289.
- [96] E. Kaschina, P. Namsolleck, and T. Unger, "AT2 receptors in cardiovascular and renal diseases," *Pharmacol. Res.*, vol. 125, pp. 39-47, 2017, doi: 10.1016/J.PHRS.2017.07.008.
- [97] H. Zhang, G. W. Han, A. Batyuk, A. Ishchenko, *et al.*, "Structural basis for selectivity and diversity in angiotensin II receptors," *Nature*, vol. 544, no. 7650, pp. 327-332, 2017, doi: 10.1038/nature22035.
- [98] H. Asada, A. Inoue, F. M. Ngako Kadji, K. Hirata, *et al.*, "The Crystal Structure of Angiotensin II Type 2 Receptor with Endogenous Peptide Hormone," *Structure*, vol. 28, no. 4, pp. 418-425.e4, 2020.
- [99] J. A. Bevan and D. Henrion, "Pharmacological implications of the flow-dependence of vascular smooth muscle tone," *Annual Review of Pharmacology and Toxicology*, vol. 34, no. 1. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, pp. 173-190, Apr. 28, 1994. doi: 10.1146/annurev.pa.34.040194.001133.
- [100] N. Yasuda, H. Akazawa, Y. Qin, Y. Zou, *et al.*, "A novel mechanism of mechanical stress-induced angiotensin II type 1-receptor activation without the involvement of angiotensin II," in *Naunyn-Schmiedeberg's Archives of Pharmacology*, Jun. 2008, vol. 377, no. 4-6, pp. 393-399. doi: 10.1007/s00210-007-0215-1.
- [101] M. Mederos y Schnitzler, U. Storch, and T. Gudermann, "Mechanosensitive Gq/11 Protein-Coupled Receptors Mediate Myogenic Vasoconstriction," *Microcirculation*, vol. 23, no. 8, pp. 621-625, 2016, doi: 10.1111/micc.12293.
- [102] G. Nickenig, "Central role of the AT1-receptor in atherosclerosis," *J. Hum. Hypertens.*, vol. 16, pp. S26-

S33, 2002, doi: 10.1038/sj.jhh.1001436.

- [103] M. E. Johansson, B. Fagerberg, and G. Bergström, "Angiotensin type 2 receptor is expressed in human atherosclerotic lesions," *JRAAS - J. Renin-Angiotensin-Aldosterone Syst.*, vol. 9, no. 1, pp. 17–21, 2008, doi: 10.3317/jraas.2008.005.
- [104] Y. Chao, L. Zhu, X. Qu, J. Zhang, *et al.*, "Inhibition of angiotensin II type 1 receptor reduced human endothelial inflammation induced by low shear stress," *Exp. Cell Res.*, vol. 360, no. 2, pp. 94–104, 2017, doi: 10.1016/j.yexcr.2017.08.030.
- [105] G. Faria-Costa, A. Leite-Moreira, and T. Henriques-Coelho, "Cardiovascular effects of the angiotensin type 2 receptor," *Revista Portuguesa de Cardiologia*, vol. 33, no. 7–8. Sociedade Portuguesa de Cardiologia, pp. 439–449, Jul. 01, 2014. doi: 10.1016/j.repc.2014.02.011.
- [106] M. M. P. Aponte and G. S. Francis, "Effect of angiotensin-converting enzyme inhibitors and angiotensin receptor antagonists in atherosclerosis prevention," *Curr. Cardiol. Rep.*, vol. 14, no. 4, pp. 433–442, 2012, doi: 10.1007/s11886-012-0275-9.
- [107] M. Chen, R. Preston Mason, and T. N. Tulenko, "Atherosclerosis alters the composition, structure and function of arterial smooth muscle cell plasma membranes," *BBA - Mol. Basis Dis.*, vol. 1272, no. 2, pp. 101–112, 1995, doi: 10.1016/0925-4439(95)00073-D.
- [108] K. S. Heo, K. Fujiwara, and J. I. Abe, "Shear stress and atherosclerosis," *Molecules and Cells*, vol. 37, no. 6. Korean Society for Molecular and Cellular Biology, pp. 435–440, 2014. doi: 10.14348/molcells.2014.0078.
- [109] C. Souilhol, J. Serbanovic-Canic, M. Fragiadaki, T. J. Chico, *et al.*, "Endothelial responses to shear stress in atherosclerosis: a novel role for developmental genes," *Nature Reviews Cardiology*, vol. 17, no. 1. Nature Research, pp. 52–63, Jan. 01, 2020. doi: 10.1038/s41569-019-0239-5.
- [110] B. Ramkhalawon, J. Vilar, D. Rivas, B. Mees, *et al.*, "Shear stress regulates angiotensin type 1 receptor expression in endothelial cells," *Circ. Res.*, vol. 105, no. 9, pp. 869–875, 2009, doi: 10.1161/CIRCRESAHA.109.204040.
- [111] A. Daugherty, D. L. Rateri, H. Lu, T. Inagami, *et al.*, "Hypercholesterolemia stimulates angiotensin peptide synthesis and contributes to atherosclerosis through the AT1A receptor," *Circulation*, vol. 110, no. 25, pp. 3849–3857, 2004, doi: 10.1161/01.CIR.0000150540.54220.C4.
- [112] S. Kiriakidi, C. Chatzigiannis, C. Papaemmanouil, A. G. Tzakos, *et al.*, "Interplay of cholesterol, membrane bilayers and the AT1R: A cholesterol consensus motif on AT1R is revealed," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 110–120, 2021, [Online]. Available: <https://doi.org/10.1016/j.csbj.2020.11.042>
- [113] R. Guixà-González, J. L. Albasanz, I. Rodriguez-Espigares, M. Pastor, *et al.*, "Membrane cholesterol access into a G-protein-coupled receptor," *Nat. Commun.*, vol. 8, 2017, doi: 10.1038/ncomms14505.
- [114] S. Ishikane and F. Takahashi-Yanaga, "The role of angiotensin II in cancer metastasis: Potential of renin-angiotensin system blockade as a treatment for cancer metastasis," *Biochemical Pharmacology*, vol. 151. Elsevier Inc., pp. 96–103, May 01, 2018. doi: 10.1016/j.bcp.2018.03.008.
- [115] S. Rodrigues-Ferreira and C. Nahmias, "G-protein coupled receptors of the renin-angiotensin system: New targets against breast cancer?," *Frontiers in Pharmacology*, vol. 6, no. FEB. Frontiers Media S.A., 2015. doi: 10.3389/fphar.2015.00024.

- [116] A. Preetha, R. Banerjee, and N. Huilgol, "Surface activity, lipid profiles and their implications in cervical cancer," 2005. doi: 10.4103/0973-1482.19600.
- [117] M. R. Zanotelli and C. A. Reinhart-King, "Mechanical forces in tumor angiogenesis," in *Advances in Experimental Medicine and Biology*, vol. 1092, 2018, pp. 91–112. doi: 10.1007/978-3-319-95294-9_6.
- [118] L. Yang, X. Cui, N. Zhang, M. Li, *et al.*, "Comprehensive lipid profiling of plasma in patients with benign breast tumor and breast cancer reveals novel biomarkers," *Anal. Bioanal. Chem.*, vol. 407, no. 17, pp. 5065–5077, 2015, doi: 10.1007/s00216-015-8484-x.
- [119] S. Beloribi-Djefaflija, S. Vasseur, and F. Guillaumond, "Lipid metabolic reprogramming in cancer cells," *Oncogenesis*, vol. 5, no. 1, pp. e189–e189, 2016, doi: 10.1038/oncsis.2015.49.
- [120] A. Šali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *J. Mol. Biol.*, vol. 234, no. 3, pp. 779–815, 1993, doi: 10.1006/jmbi.1993.1626.
- [121] T. U. Consortium, "The UniProt Consortium UniProt: the universal protein knowledgebase in 2021." *Nucleic Acids Res.* 49:D1 (2021). [Online]. Available: <https://www.uniprot.org/help/uniprotkb>
- [122] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, *et al.*, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000, doi: 10.1093/NAR/28.1.235.
- [123] M. Hernández-Rodríguez, M. C. Rosales-Hernández, J. E. Mendieta-Wejebe, M. Martínez-Archundia, *et al.*, "Current Tools and Methods in Molecular Dynamics (MD) Simulations for Drug Design," *Curr. Med. Chem.*, vol. 23, no. 34, pp. 3909–3924, 2016, doi: 10.2174/0929867323666160530144742.
- [124] S. S. Ali and S. Bharadwaj, "Prospects for Detecting the 326.5 MHz Redshifted 21-cm HI Signal with the Ooty Radio Telescope (ORT)," *Journal of Astrophysics and Astronomy*, 2014. https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part1.html
- [125] M. A. González, "Force fields and molecular dynamics simulations," *École thématique la Société Française la Neutron.*, vol. 12, pp. 169–200, 2011, doi: 10.1051/sfn/201112009.
- [126] M. Ikeguchi, "Partial rigid-body dynamics in NPT, NPAT and NPγT ensembles for proteins and membranes," *J. Comput. Chem.*, vol. 25, no. 4, pp. 529–541, 2004, doi: 10.1002/JCC.10402.
- [127] T. Broemstrup and N. Reuter, "Molecular Dynamics Simulations of Mixed Acidic/Zwitterionic Phospholipid Bilayers," *Biophys. J.*, vol. 99, no. 3, pp. 825–833, 2010, doi: 10.1016/J.BJP.2010.04.064.
- [128] M. Sotomayor and K. Schulten, "Molecular dynamics study of gating in the mechanosensitive channel of small conductance MscS," *Biophys. J.*, vol. 87, no. 5, pp. 3050–3065, 2004, doi: 10.1529/biophysj.104.046045.
- [129] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 43–56, 1995, doi: 10.1016/0010-4655(95)00042-E.
- [130] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, *et al.*, "The Amber biomolecular simulation programs," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, Dec. 2005. doi: 10.1002/jcc.20290.
- [131] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, *et al.*, "Scalable molecular dynamics on CPU and GPU architectures with NAMD," *J. Chem. Phys.*, vol. 153, no. 4, 2020, doi: 10.1063/5.0014475.
- [132] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996, doi: 10.1016/0263-7855(96)00018-5.

- [133] M. Moradi, "Exploring Complex Conformational Transition Pathways," *NAMD Tutorials*, 2017, [Online]. Available: <http://www.ks.uiuc.edu/Training/Tutorials/>
- [134] I, "Steered Molecular Dynamics," *J. Phys. Chem. B*, vol. 115, no. 5, pp. 1211–1219, 2011, doi: 10.1021/jp107933k.
- [135] M. Moradi and E. Tajkhorshid, "Mechanistic picture for conformational transition of a membrane transporter at atomic resolution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 47, pp. 18916–18921, 2013, doi: 10.1073/pnas.1313202110.
- [136] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, *et al.*, "Comparative Protein Structure Modeling Using Modeller," *Curr. Protoc. Bioinforma.*, vol. 15, no. 1, p. Unit, 2006, doi: 10.1002/0471250953.bi0506s15.
- [137] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, 2008, doi: 10.1002/jcc.20945.
- [138] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, *et al.*, "Bio3d: An R package for the comparative analysis of protein structures," *Bioinformatics*, vol. 22, no. 21, pp. 2695–2696, 2006, doi: 10.1093/bioinformatics/btl461.
- [139] Z. Gu, R. Eils, and M. Schlesner, "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," *Bioinformatics*, vol. 32, no. 18, pp. 2847–2849, 2016, doi: 10.1093/bioinformatics/btw313.
- [140] "The PyMOL Molecular Graphics," *Version 2.0 Schrödinger, LLC.*, [Online]. Available: <https://pymol.org/2/support.html>
- [141] R. Guixà-González, I. Rodríguez-Espigares, J. M. Ramírez-Anguita, P. Carrió-Gaspar, *et al.*, "MEMBPLUGIN: Studying membrane complexity in VMD," *Bioinformatics*, vol. 30, no. 10, pp. 1478–1480, 2014, doi: 10.1093/bioinformatics/btu037.
- [142] B. G. Levine, J. E. Stone, and A. Kohlmeyer, "Fast analysis of molecular dynamics trajectories with graphics processing units-Radial distribution function histogramming," *J. Comput. Phys.*, vol. 230, no. 9, pp. 3556–3569, 2011, doi: 10.1016/J.JCP.2011.01.048.
- [143] L. M. Wingler and R. J. Lefkowitz, "Conformational Basis of G Protein-Coupled Receptor Signaling Versatility," *Trends in Cell Biology*, vol. 30, no. 9, pp. 736–747, 2020. doi: 10.1016/j.tcb.2020.06.002.
- [144] E. H. Schneider, D. Schnell, A. Strasser, S. Dove, *et al.*, "Impact of the DRY motif and the missing 'ionic lock' on constitutive activity and G-protein coupling of the human histamine H 4 receptor," *J. Pharmacol. Exp. Ther.*, vol. 333, no. 2, pp. 382–392, 2010, doi: 10.1124/jpet.109.163220.
- [145] C. Li, Z. Jia, A. Chakravorty, S. Pahari, *et al.*, "DelPhi Suite: New Developments and Review of Functionalities," *J. Comput. Chem.*, vol. 40, no. 28, pp. 2502–2508, 2019, doi: 10.1002/jcc.26006.
- [146] B. Taddese, A. Garnier, H. Abdi, D. Henrion, *et al.*, "Deciphering collaborative sidechain motions in proteins during molecular dynamics simulations," *Sci. Rep.*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-72766-1.
- [147] S. Lu, X. He, Z. Yang, Z. Chai, *et al.*, "Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design," *Nat. Commun.*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-25020-9.
- [148] M. Paz Ocaranza, J. A. Riquelme, L. García, J. E. Jalil, *et al.*, "Counter-regulatory renin-angiotensin

system in cardiovascular disease," *Nat. Rev. Cardiol.* 2019 172, vol. 17, no. 2, pp. 116–129, 2019, doi: 10.1038/s41569-019-0244-8.

Annexe I

Résultats complémentaires

Figure S1 : RMSD et RMSF des simulations dynamiques classiques.

Figure S2 : Heatmaps des distances entre les hélices 3 et 7 lors des simulations dynamiques classiques.

Figure S3 : Heatmaps de la coordination de l'ion sodium avec les résidus protéiques de son site lors des simulations dynamiques classiques.

Figure S4 : Tableau des moyennes et écarts type des liaisons H entre les résidus 6.32 et 7.58 lors des simulations dynamiques classiques.

Figure S5 : Heatmaps des liaisons H entre les résidus 6.32 et 7.58 lors des simulations dynamiques classiques.

Figure S6 : Heatmaps des simulations dynamiques dirigées.

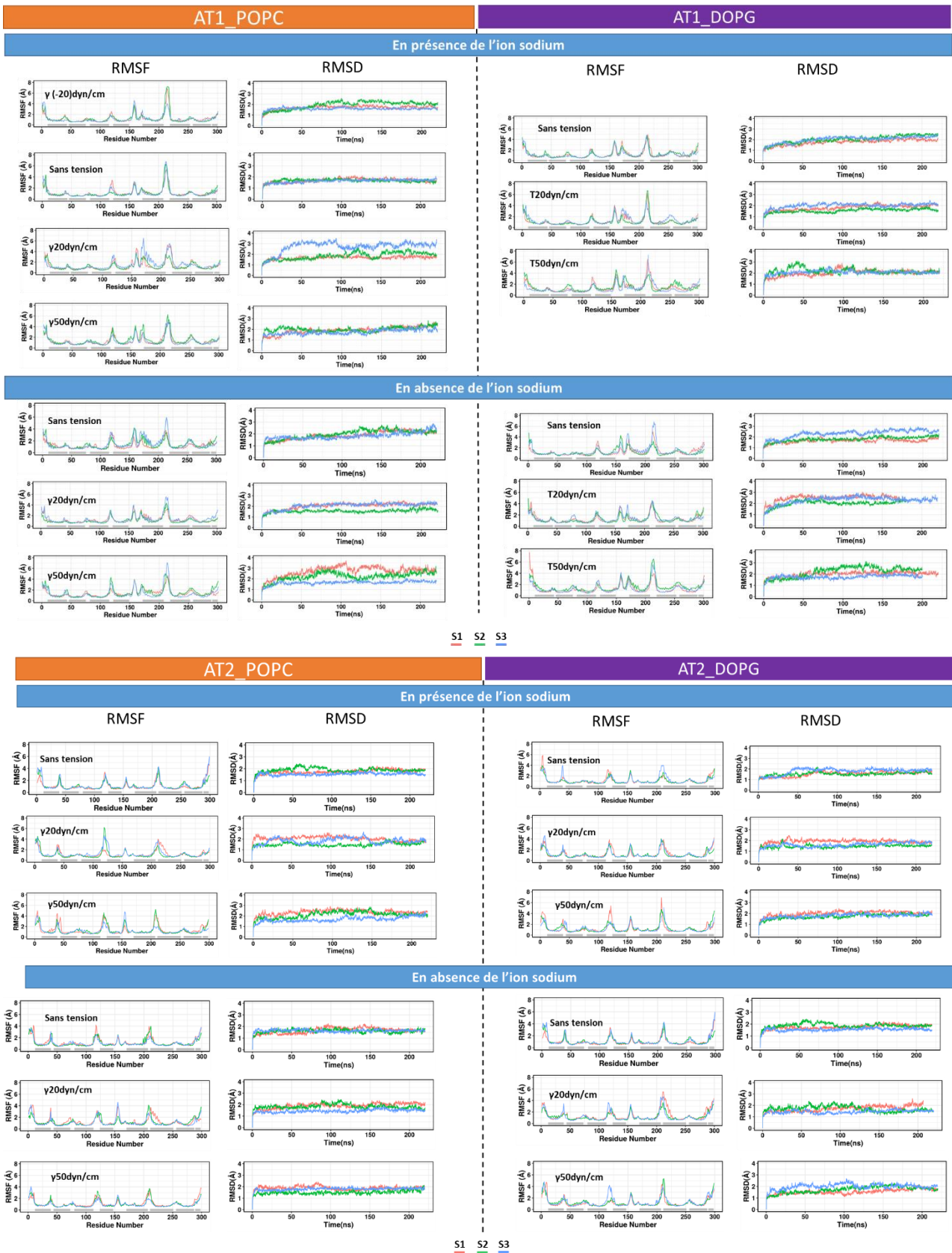


Figure S1 : RMSD et RMSF des simulations dynamiques classiques.

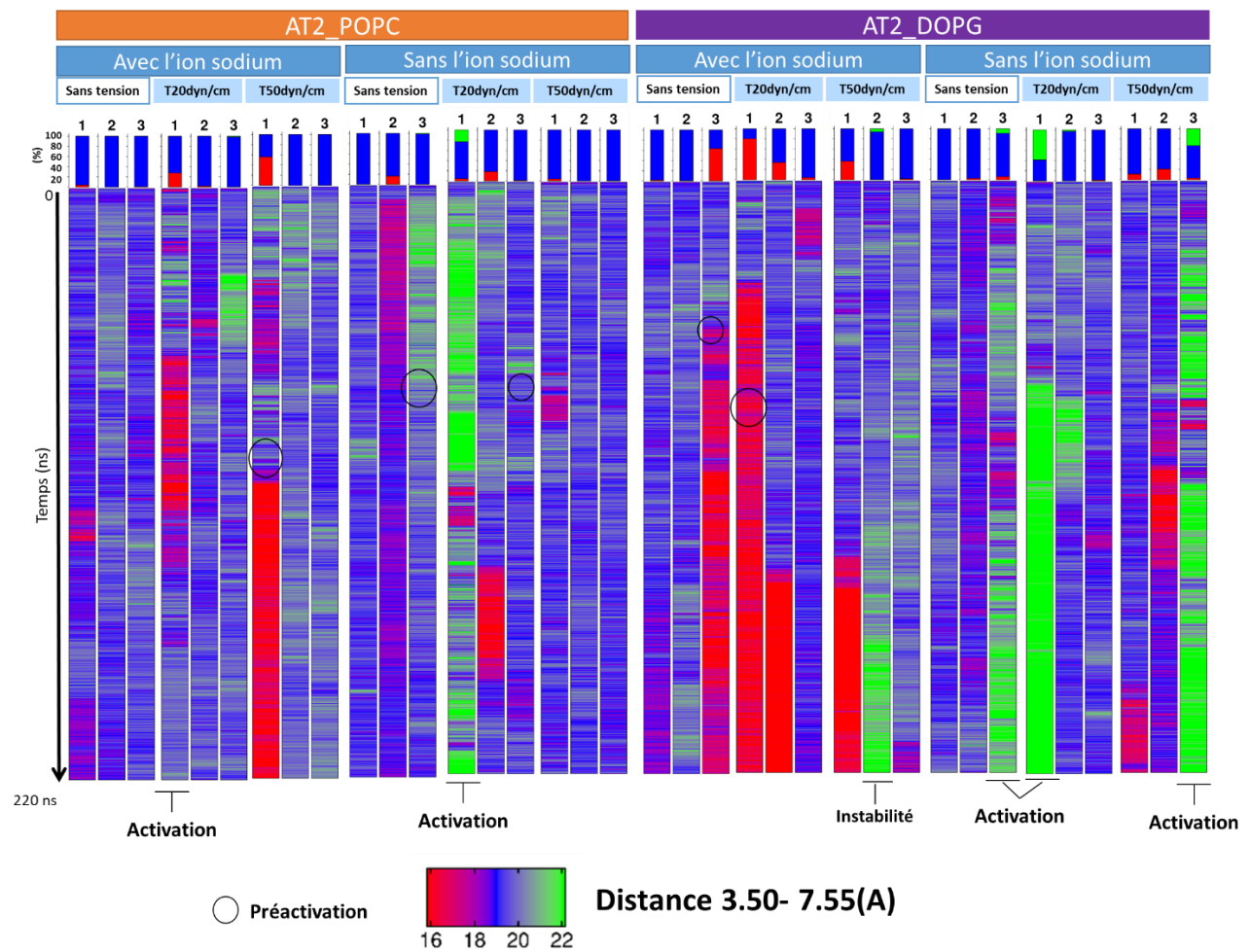
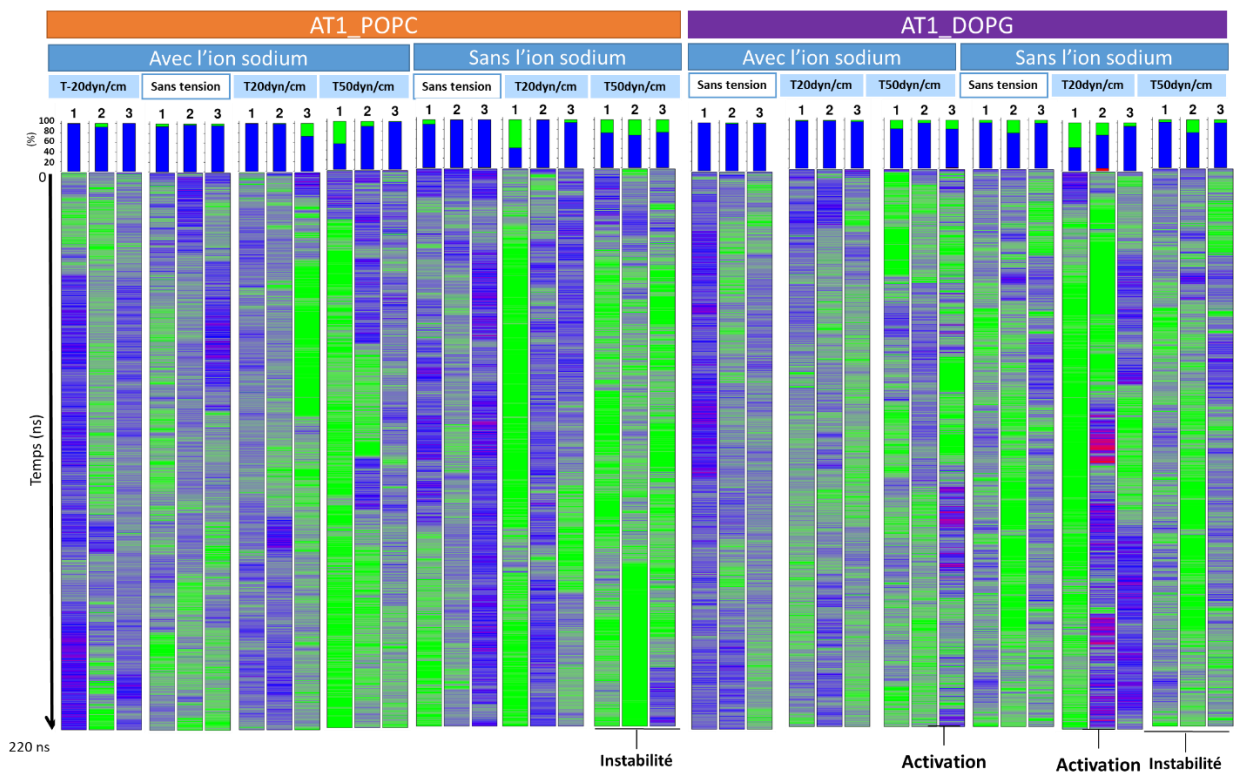


Figure S2 : Heatmaps des distances entre les hélices 3 et 7 (Distance 3.50-7.55) lors des simulations dynamiques classiques.

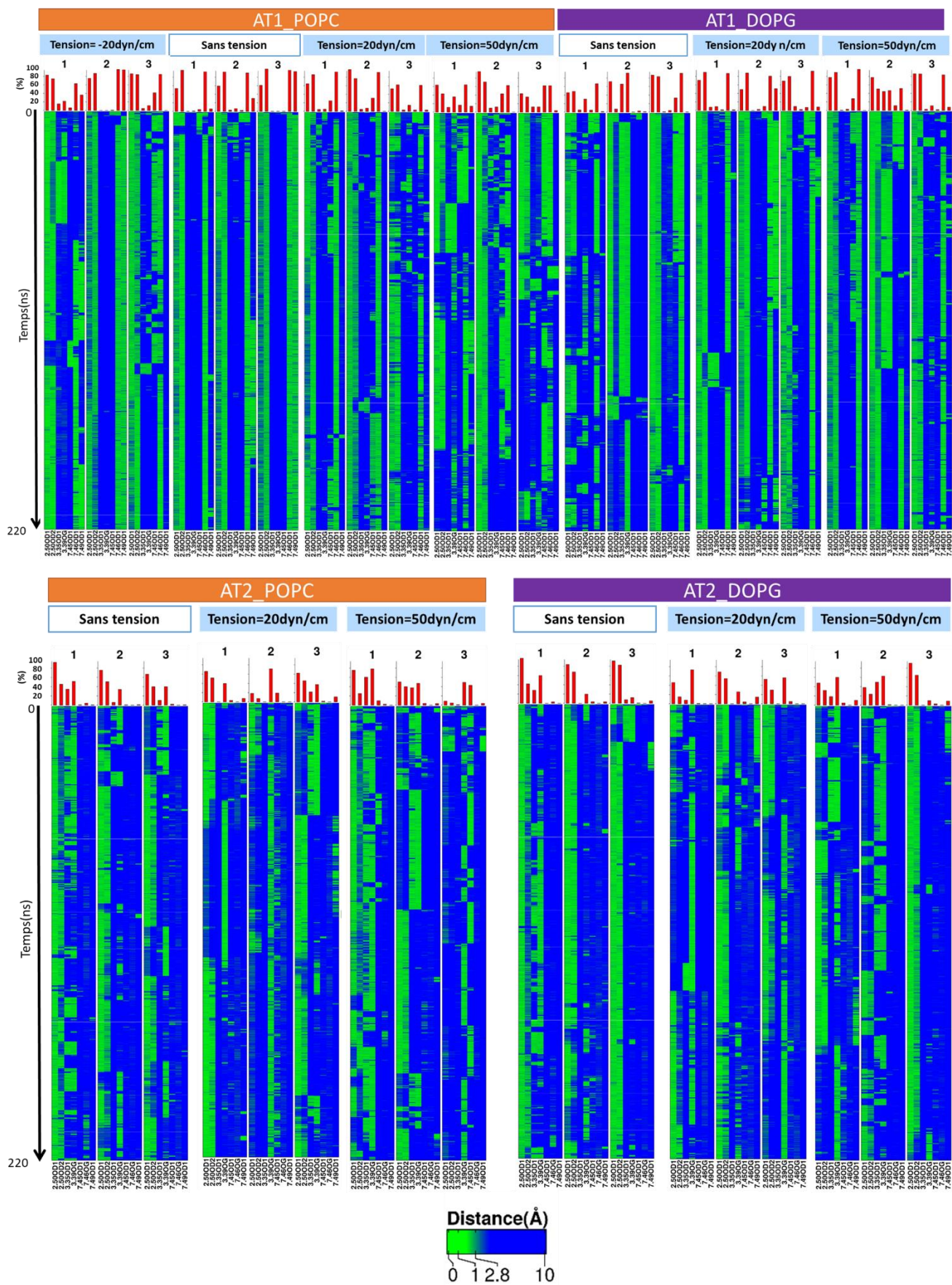


Figure S3 : Heatmaps des distances entre l'ion sodium et les résidus protéiques de son site lors des simulations dynamiques classiques.

Une distance inférieure à 2.8 Å indique une coordination.

Moyennes \pm SD	AT1_POPC							AT1_DOPG					
	AVEC l'ion sodium				sans l'ion sodium			AVEC l'ion sodium			sans l'ion sodium		
	γ (-20)	NPT	γ 20	γ 50	NPT	γ 20	γ 50	NPT	γ 20	γ 50	NPT	γ 20	γ 50
% HBONDS (n=3)													
6.32_7.58	70,3 \pm 5,7	68,5 \pm 8,9	49,3 \pm 32,9	46,03 \pm 26,1	51,43 \pm 24,2	54,41 \pm 39,0	23,18 \pm 14,9	55,64 \pm 12,1	61,34 \pm 9,6	43,27 \pm 14,6	36,2 \pm 3,1	18,5 \pm 20,2	42,4 \pm 7,4
7.58_lipid	11,6 \pm 2,5	4,87 \pm 2,6	5,14 \pm 0,4	4,36 \pm 1,9	7,05 \pm 3,8	10,32 \pm 8,5	0,91 \pm 0,4	18,84 \pm 9,0	21,10 \pm 6,0	6,58 \pm 2,7	33,3 \pm 3,8	21,9 \pm 9,1	1,9 \pm 2,5
6.32_lipid	0,0 \pm 0,0	0,0 \pm 0,0	4,7 \pm 8,1	0,0 \pm 0,0	0,0 \pm 0,0	0,0 \pm 0,0	0,0 \pm 0,0	1,0 \pm 1,4	0,1 \pm 0,2	0,2 \pm 0,1	4,3 \pm 5,0	3,1 \pm 3,0	0,1 \pm 0,2

Figure S4 : Tableau des moyennes et écarts type des liaisons H entre les résidus 6.32 et 7.58 lors des simulations dynamiques classiques (n=3).

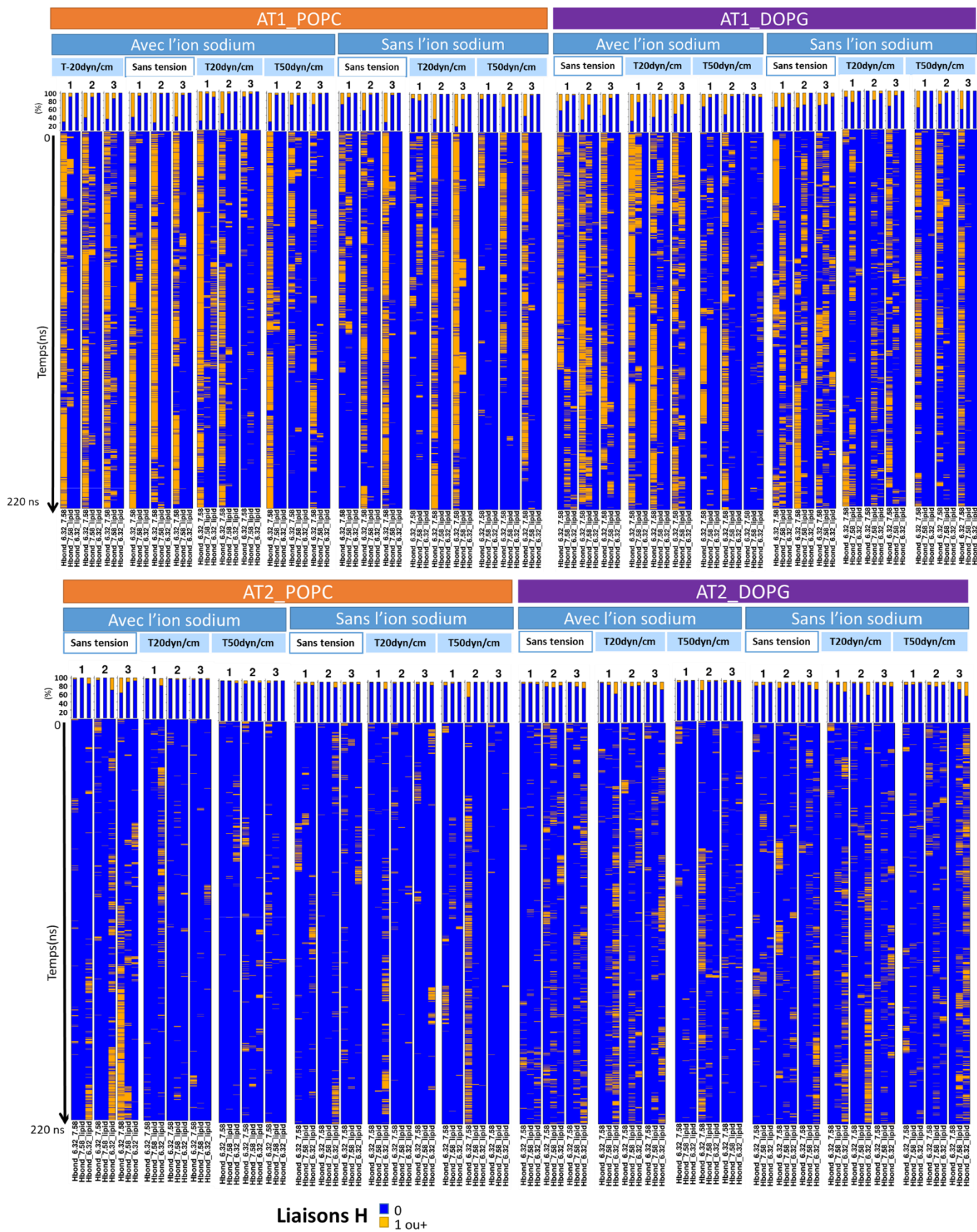
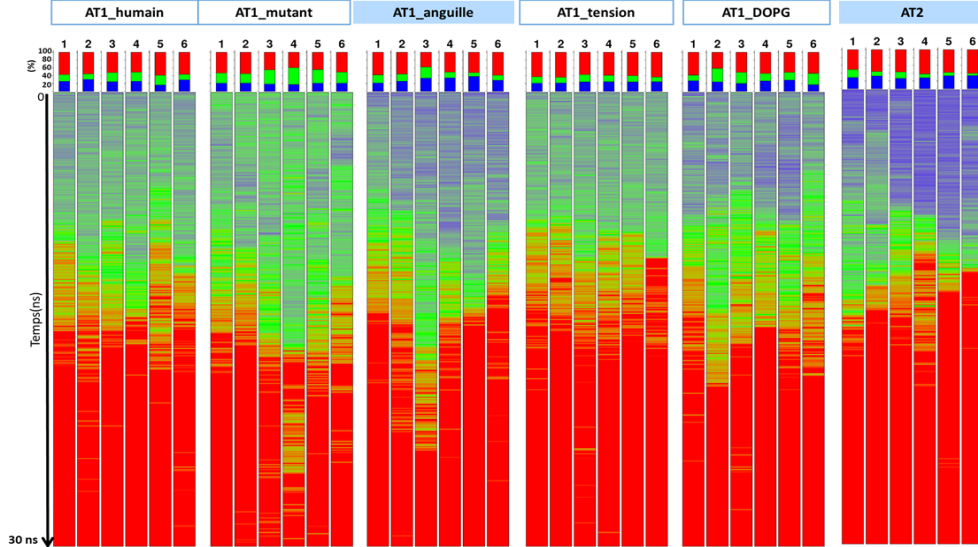
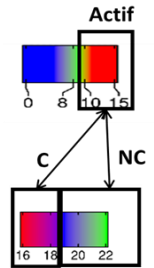


Figure S5 : Heatmaps des liaisons H entre les résidus 6.32 et 7.58 lors des simulations dynamiques classiques

Distance TM3-TM6

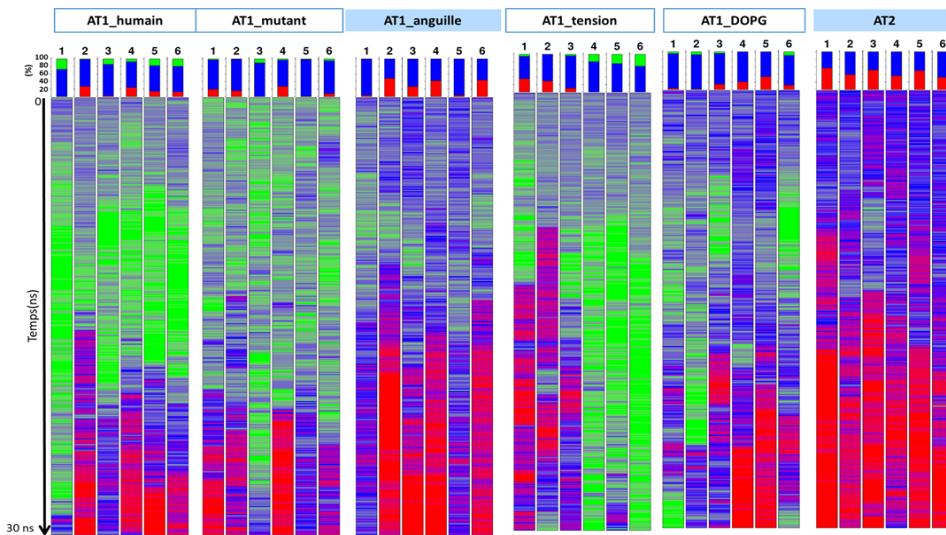


Distance 3.50- 6.34(A)



Distance 3.50- 7.55(A)

Distance TM3-TM7



Liaisons H

- 0
- 1 ou+

Liaisons H

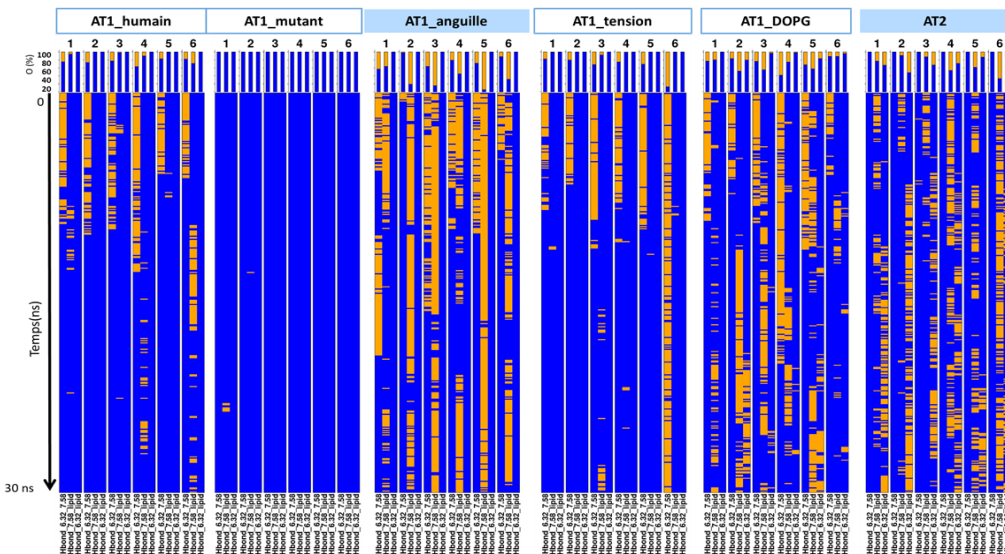


Figure S6 : Heatmaps des simulations dynamiques dirigées
Heatmaps des distances TM3-TM6 et TM3-TM7 ; Heatmap des liaisons H entre les résidus 6.32 et 7.58.

Annexe II

Publications

Annexe II.1 : Article : « Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota »

Annexe II.2 : Chapitres méthodologiques

Annexe II.3 : Abstracts d'autres publications

Annexe II.1

Article :

[Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota](#) **Ben Boubaker R***, Tiss A*, Henrion D, Guissouma H, Chabbert M (2022) PLOS Computational Biology 18(2): e1009732. <https://doi.org/10.1371/journal.pcbi.1009732> * Equal participation

Contexte et principales contributions

Ce projet s'intéresse au rôle de l'ion sodium dans le mécanisme d'activation des récepteurs de l'angiotensine II, AT1 et AT2. Pour mieux comprendre les spécificités de ces récepteurs, nous avons tout d'abord effectué une analyse phylogénétique qui a révélé une évolution en miroir des récepteurs AT1 et AT2. Chacun des deux récepteurs est divisé en deux clades, séparant les récepteurs de poissons des espèces terrestres. Cette analyse a également révélé que des mutations marquantes ont eu lieu au niveau ou à proximité du site de liaison du sodium dans les deux récepteurs AT1 et AT2.

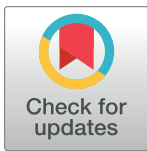
Les calculs électrostatiques et les simulations de dynamique moléculaire confirment la fixation de l'ion sodium à l'AT1 humain malgré la présence atypique d'une asparagine en position 7.46 du site de fixation du sodium. La comparaison des modes de liaison au sodium dans les récepteurs AT1 et AT2 sauvages et mutés de l'homme et de l'anguille indique que le contrôle allostérique par le sodium des récepteurs AT1 et AT2 a évolué au cours de la transition du poisson à l'amniote. En particulier, la mutation S7.46N d'AT1 est spécifique des amniotes et est reflétée par une mutation L3.46M dans AT2. Nous avons ainsi montré que ces mutations survenues lors de la transition vers la vie terrestre modifient le mode de fixation du sodium et les propriétés fonctionnelles des récepteurs AT1 et AT2.

Cette étude a donné lieu à un article intitulé « Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota » qui vient d'être publié dans Plos Comput Biol (doi: 10.1371/journal.pcbi.1009732).

Ce projet a été initié dans le cadre du sujet de thèse de Dr Asma Tiss soutenue en 27/06/2021. Ce projet étant un sujet complexe et nécessitant un aspect technique poussé, j'ai fortement contribué à l'élaboration des simulations dynamiques et leur procédure ainsi qu'au développement des scripts nécessaires pour accomplir les différentes analyses et la visualisation des données. J'ai aussi pris en charge une partie des demandes des examinateurs de l'article qui a nécessité près de 6 mois de travail. En conséquence, une grande partie de mon temps de thèse a été consacrée à ce projet. Ma contribution y ayant été équivalente à celle d'Asma Tiss, nous sommes toutes les deux co-premiers auteurs de l'article de Plos Comput Biol.

RESEARCH ARTICLE

Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota

Rym Ben Boubaker¹ , Asma Tiss^{1,2} , Daniel Henrion¹ , Hajer Guissouma², Marie Chabbert^{1*} ¹ CNRS UMR 6015 – INSERM U1083, Laboratoire MITOVASC, Université d'Angers, Angers, France, ² INSAT de Tunis, Université de Carthage, Carthage, Tunisie These authors contributed equally to this work.* marie.chabbert@univ-angers.fr

Abstract

In vertebrates, the octopeptide angiotensin II (AngII) is an important *in vivo* regulator of the cardiovascular system. It acts mainly through two G protein-coupled receptors, AT1 and AT2. To better understand distinctive features of these receptors, we carried out a phylogenetic analysis that revealed a mirror evolution of AT1 and AT2, each one split into two clades, separating fish from terrestrial receptors. It also revealed that hallmark mutations occurred at, or near, the sodium binding site in both AT1 and AT2. Electrostatics computations and molecular dynamics simulations support maintained sodium binding to human AT1 with slow ingress from the extracellular side and an electrostatic component of the binding free energy around -3kT, to be compared to around -2kT for human AT2 and the δ opioid receptor. Comparison of the sodium binding modes in wild type and mutated AT1 and AT2 from humans and eels indicates that the allosteric control by sodium in both AT1 and AT2 evolved during the transition from fish to amniota. The unusual S7.46N mutation in AT1 is mirrored by a L3.36M mutation in AT2. In the presence of sodium, the N7.46 pattern in amniota AT1 stabilizes the inward orientation of N3.35 in the apo receptor, which should contribute to efficient N3.35 driven biased signaling. The M3.36 pattern in amniota AT2 favours the outward orientation of N3.35 and the receptor promiscuity. Both mutations have physiological consequences for the regulation of the renin-angiotensin system.

 OPEN ACCESS

Citation: Ben Boubaker R, Tiss A, Henrion D, Guissouma H, Chabbert M (2022) Evolutionary information helps understand distinctive features of the angiotensin II receptors AT1 and AT2 in amniota. *PLoS Comput Biol* 18(2): e1009732. <https://doi.org/10.1371/journal.pcbi.1009732>

Editor: Bert L. de Groot, Max Planck Institute for Biophysical Chemistry, GERMANY

Received: June 7, 2021

Accepted: December 8, 2021

Published: February 24, 2022

Copyright: © 2022 Ben Boubaker et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data have been deposited at the Mendeley Data repository (<https://data.mendeley.com>) and are accessible under doi:10.17632/y6sh9yfyjt.1 (computational data) and doi:10.17632/pfr6765y73.1 (sequence data).

Funding: This work was supported by the grant 100567 from GENCI (Grand équipement national de calcul intensif) to MC, institutional grants from CNRS, INSERM, and the University of Angers to DH, a fellowship from the University of Carthage (Tunisia) to AT and a fellowship from the University

Author summary

The analysis of protein sequences from different species can reveal interesting trends in the structural and functional evolution of a protein family. Here, we analyze the evolution of two G protein-coupled receptors, AT1 and AT2, which bind the angiotensin II peptide and are important regulators of the cardiovascular system. We show that these receptors underwent a mirror evolution. Specific mutations at, or near, the sodium binding pocket occurred in both AT1 and AT2 during the transition to terrestrial life. We carried out electrostatics computations and molecular dynamics simulations to decipher the details of

of Angers to RBB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the sodium binding mode in eel and human receptors, as prototypes of fish and amniota receptors. Our results indicate that sodium binding is kinetically slow but thermodynamically stable. Comparison of the sodium binding modes in eel and human receptors reveals that an unusual mutation in the sodium binding pocket of AT1 is critical for biased signaling of amniota AT1 whereas a mutation in AT2 promotes promiscuity of amniota AT2. In turn, these data indicate that a few mutations at a strategic position (here the sodium binding pocket) are an efficient way to gain functional evolution.

Introduction

The renin-angiotensin system (RAS) is an important *in vivo* regulator of multiple cardiovascular and renal functions in vertebrates [1–3]. It initiates with the cleavage of the angiotensinogen by renin into the decapeptide angiotensin I that is then cleaved by the angiotensin converting enzyme 1 (ACE1) into the octapeptide angiotensin II (AngII). AngII is subsequently cleaved into angiotensin 1–7 (Ang1-7) by the angiotensin converting enzyme 2 (ACE2). A variety of other angiotensin derivatives are produced by diverse enzymes. The RAS system is composed of two axes. In the “classical” axis, AngII binds the type 1 AngII receptor (AT1) to induce most known effects of AngII, including vasoconstriction and increased blood pressure, anti-natriuresis, hypertrophy, fibrosis and inflammation [4,5]. Dysregulation of this axis promotes cardiac and vascular damages and AT1 blockers are widely used to fight hypertension. The “counter-regulatory” axis involves the type 2 AngII receptor (AT2) and the MAS receptor, through AngII (AT2) and Ang1-7 (AT2, MAS). These receptors promote vasodilation, natriuresis, anti-inflammation, anti-fibrosis, and anti-proliferative responses that counterbalance AT1 effects [3,6,7].

AT1 and AT2 are two G protein-coupled receptors (GPCRs) from class A that share 30% sequence identity and similar affinity for AngII but have very different properties, expression profiles and functions [4]. AT1 is expressed in all organs. When it is stimulated by AngII, it activates both G proteins and β -arrestins. A variety of AT1 ligands induce biased signaling with preferential activation of G proteins or β -arrestins [8,9]. AT2 is widely expressed during development whereas, in the adult life, it is expressed only in a few organs, except in pathological conditions where its expression is up-regulated [5,10,11]. Direct AT2 activation of the conventional GPCR effectors (G proteins and β -arrestins) remains controversial and the primary signaling pathways involved in AT2 signaling are unknown [12–14]. A better knowledge of these pathways is important as stimulation of AT2 in pathological situations might be an interesting alternative to the blockade of AT1 [10].

Several structures of AT1 [15–18] and AT2 [19–21] in complex with different ligands have been resolved. AT1 has been crystallized in both inactive and active states. AT2 has been crystallized in active states only, which supports the hypothesis of a “relaxed” conformation that would be prone to activation [22]. Comparison of the active structures of AT1 in complex with AngII and biased agonists provides clues to the biased responses of AT1 [16] that match those observed in large-scale molecular dynamics simulations [23]. Briefly, two conformations of the active state are differentially stabilized by β -arrestin-biased, balanced, and G protein-biased ligands. Two asparagine residues located at positions 3.35 and 7.46 (Ballesteros’ numbering [24]) play a key role in this mechanism. The inward/outward orientation of N3.35 acts as an allosteric switch between the balanced canonical and the β -arrestin-biased alternative conformations, whereas reorganization of the N7.46 H-bonding pattern helps transmit the

information to the intracellular side [16,23]. These two residues are located at an allosteric sodium binding site that is highly conserved among class A GPCRs.

Sodium binding is an essential element of receptor activation in most class A GPCRs. It acts as a negative allosteric regulator, by stabilizing the inactive structure, and might be involved in the activation mechanism [25,26]. Sodium coordination involves a strictly conserved aspartic acid at position 2.50 and polar residues at positions 3.35, 3.39, 7.45, 7.46, or 7.49. An asparagine at position 3.35 is frequent in GPCRs, whereas an asparagine at position 7.46 is unusual [25,27]. In the inactive structures of human AT1, N7.46 is H-bonded to N3.35 [17,18]. It was suggested that the N7.46 pattern might prevent sodium binding while stabilizing the inactive structure by interacting with N3.35 [16]. By contrast, human AT2 has a conventional sodium binding site, with the N3.35 and S7.46 patterns.

Since residues from the sodium pocket have a functionally important role in AT1 activation, a better knowledge of the history of this site is mandatory to understand the role of this unusual sodium pocket, not only in the activation mechanism of AT1, but also in the balance between AT1 and AT2 activation. This prompted us to analyze the evolutionary information contained in the extend sequences of both AT1 and AT2.

The evolutionary approach revealed a mirror evolution of AT1 and AT2. Both AT1 and AT2 are split in two clades that separate fish from terrestrial receptors. Mutations of residues at, or close to, the sodium binding site occurred specifically in amniota receptors. In particular, the N7.46 pattern in AT1 is an innovation of amniota. To understand the effects of the observed mutations, we carried out molecular dynamics (MD) simulations and electrostatics computations to determine (1) the sodium ingress pathway to the canonical binding site in AT1 and (2) the characteristics of the sodium binding mode in AT1 and AT2 receptors from humans and eels. The mutations specific of amniota AT1 and AT2 receptors have physiological consequences for the regulation of the renin-angiotensin system that will be discussed.

Results

Molecular evolution of the angiotensin II receptors

Data mining from Uniprot (see [Methods](#)) provided a set of 318 “clean” sequences of AngII receptors. Based on a preliminary tree, the sequences were assigned to either the AT1 or AT2 sub-family (203 and 115 members, respectively). These receptors presented strong sequence biases. Only 28% of AT1 pairs and 51% of AT2 pairs had a sequence identity lower than 50% ([Fig 1A](#)). To obtain a non-redundant set of vertebrate receptors, we clustered the 318 sequences using the NRDB program [28], with a threshold of 90% sequence identity. We obtained a total of 75 clusters: 37 clusters for the AT1 sequences and 38 clusters for the AT2 sequences. The 120 AT1 and 68 AT2 sequences from mammals clustered into 3 and 5 groups, respectively. The clusters containing human AT1 and AT2 included 115 and 39 sequences, respectively.

We selected a representative sequence from each cluster, based on sequence and genome qualities, to obtain a non-redundant (NR) sequence set. In this set, 51% and 74% of, respectively, AT1 and AT2 pairs had less than 50% identity ([Fig 1A](#)). Using the aligned sequences of the NR set, we built a Neighbour Joining phylogenetic tree of the AngII receptors. Starting from a single AngII receptor more closely related to AT1 in jawless fish (lamprey), duplication in an ancestor of jawed fishes led to an almost symmetrical evolution of the AT1 and AT2 receptors, hinting at their complementary role in RAS control ([Fig 1B](#)). Each AT1 and AT2 sub-tree possesses two major clades: one for fishes and one for terrestrial animals. There are two exceptions to this pattern: (1) The coelacanth AT1 and AT2 sequences are clearly part of the “terrestrial” clades, in the vicinity of the batracian sequences. Their position is consistent with the

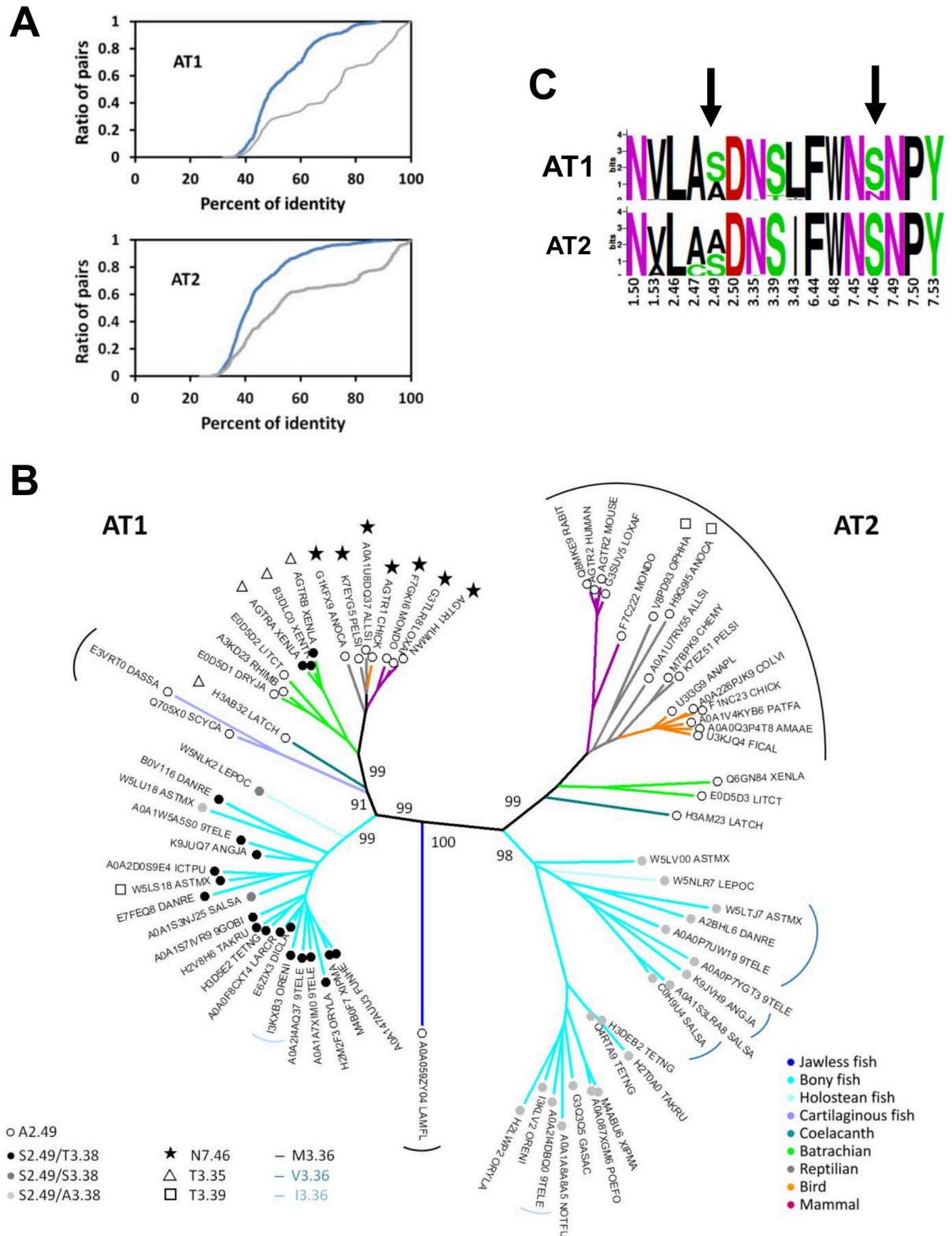


Fig 1. Evolution of the angiotensin II receptors. (A) Cumulative distribution function of the pairwise identities for AT1 (top) and AT2 (bottom), in the Uniprot sets (grey lines) and in the non-redundant sets (blue lines); (B) Neighbor-joining tree of the non-redundant set of AT1 and AT2 receptors (500 bootstraps). The bootstrap values in percent are indicated for main branches. The color code of the tree branches refers to the phylum (jawless fish: dark blue, bony fish: cyan, holostean fish: light cyan, cartilaginous fish: slate, coelacanth: deepteal,

batrachian: green, reptilian: grey, bird: orange, mammal: purple). The external labels indicate mutations in the sodium binding site as compared to the canonical sodium binding site characterized by N3.35, S3.39 and S7.46 (N7.46: stars, T3.35: open triangles, T3.39: open squares). The internal labels indicate the residues at positions 2.49 and 3.38 (A2.49: open circle, S2.49/T3.38: black circle, S2.49/S3.38: dark grey circle, S2.49/A3.38: light grey circle). The circular arcs indicate the residue at position 3.36 (black: M3.36, dark blue: V3.36, light blue: I3.36, no arc: L3.36); (C) Sequence logo plots of the sodium binding site in the NR set of AT1 and AT2. For these plots, all the residues lining up the sodium binding site were taken into account [23]. The arrows highlight the mutations at positions 2.49 and 7.46.

<https://doi.org/10.1371/journal.pcbi.1009732.g001>

coelacanth status of “living fossil”, in an intermediary position in the tree of life; (2) Only AT1 sequences are present in available cartilaginous fish genomes (rays, sharks), as detailed in Methods. These sequences are closer to terrestrial than to fish sequences. Interestingly, this feature was observed for the full genome of sharks [29].

Human AT1 possesses an unusual sequence pattern (N7.46) at the canonical sodium binding site. To determine when this pattern emerged, we analyzed the residues lining the sodium binding site [25,26]. After verifying that all the sequences from each cluster presented the same pattern, we reported key positions on the phylogenetic tree (Fig 1B). Most residues are highly conserved between and within the AT1 and AT2 sets (Fig 1C). In particular, residues usually involved in sodium coordination (D2.50, N3.35, S3.39 and N7.45) are highly conserved. A difference is observed at position 3.43 with Leu and Ile for AT1 and AT2, respectively. In both sets, the position 2.49 differentiates fish (S2.49) from terrestrial (A2.49) receptors (Fig 1B and 1C). The C2.47 pattern is present only in some AT2 sequences from fishes. Noteworthy, as observed in most GPCRs, a serine residue at position 7.46 is present in most AT1 and AT2 receptors (Fig 1B). An Asn residue at this position is observed only in amniota (reptilians, birds, and mammals). This indicates that this specificity of AT1 has been acquired during the transition from an aqueous to a terrestrial environment.

A “canonical” sodium binding site with two Asn residues at positions 3.35 and 7.45 and two Ser residues at positions 3.39 and 7.46 is thus observed in AT2 receptors from any species and in AT1 receptors from fishes and some amphibians (Neobatrachia clade). The N7.46 pattern is observed specifically in amniota whereas other coordinating residues are conserved. Notably, a N3.35T mutation of the residue facing S7.46 is observed in the coelacanth and batracians from the pipoidea clade (*Xenopus*), suggesting that evolution might have led to different solutions for a similar problem.

Properties of D2.50 in AT1 and AT2

The deprotonation of D2.50 is a key element of sodium binding to GPCRs. We thus wondered whether the mutations in the sodium binding site of AT1 and AT2 could alter the pKa of D2.50. To answer this question, we selected AT1 and AT2 from *Anguilla japonica* (eel) as prototypes of the fish receptors. The receptors were modeled from the crystallographic structures of human AT1 and AT2 (see Methods). We computed the pKa of D2.50 in AT1 and AT2 from humans (AT1h, AT2h) and eels (AT1e, AT2e) and in the human AT1 variant with the N7.46S mutation (AT1m). The pKa of D2.50 in the δ opioid receptor (OPRD) and in the chemokine receptor CXCR4 were also computed as controls. To calculate pKas, we used the DelphiPKa program with the model of Gaussian representation of atomic densities proposed by Alexov and coworkers [30–32]. This representation yields a smooth Gaussian-based dielectric function without defining a molecular surface. Electrostatic computations depend on two parameters: the width of the Gaussian function, σ , and the protein dielectric value, ϵ_{in} . We compared a wide range of parameters (σ varying from 0.5 to 1, ϵ_{in} varying from 2 to 10) for AT1h and OPRD in search of pairs compatible with sodium binding to OPRD (see Methods). Each pair of parameters led to similar pKa values for AT1 and OPRD (S1 Fig), indicating that D2.50 has the same protonation state in both receptors. Noteworthy, sodium binding to OPRD is not

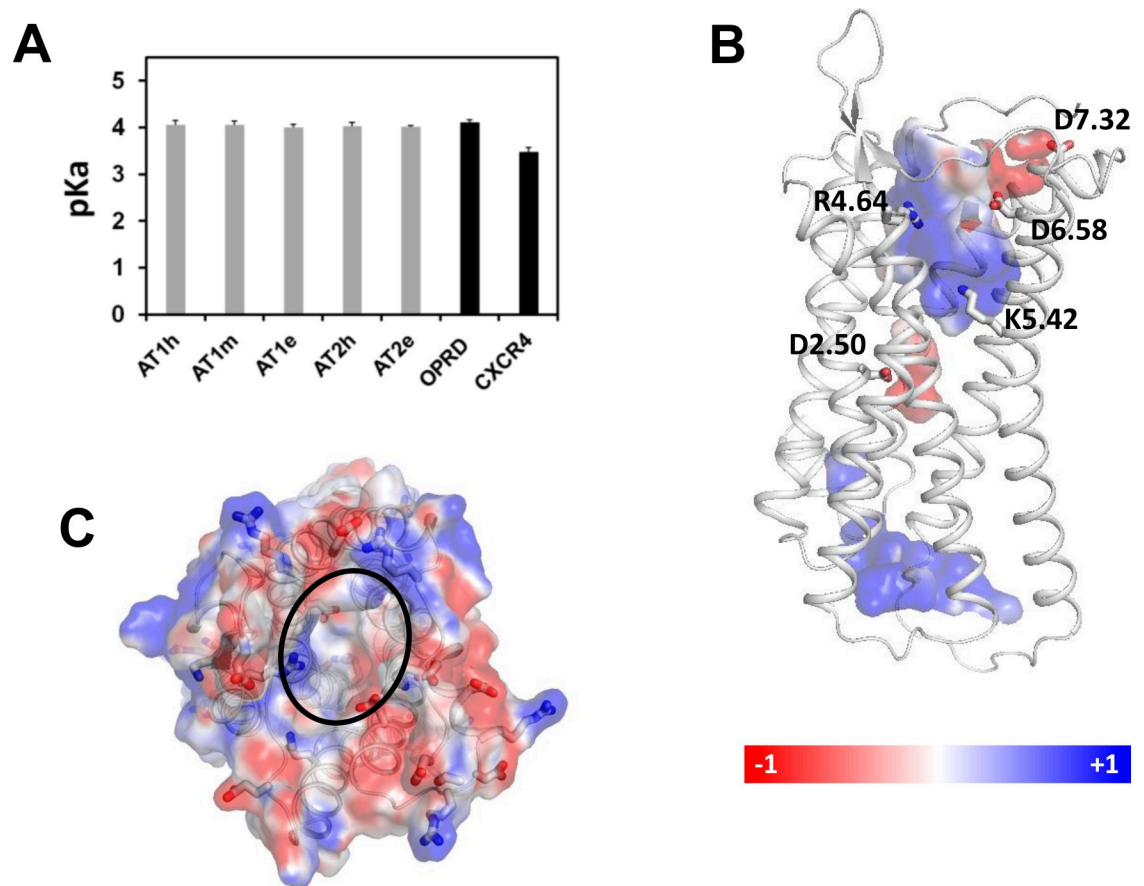


Fig 2. Electrostatic properties of AT1 and AT2. (A) pKas of residue D2.50 in AT1 and AT2 receptors and in the AT1m variant. The pKas of D2.50 in CXCR4 and OPRD are given for comparative purpose; (B) Electrostatic potential of the AT1h internal cavities (side view). The conserved charged side chains interacting with AngII (R4.64, K5.42, D6.58 and D7.32) and D2.50 are shown as sticks; (C) Electrostatic potential of the extracellular surface of AT1h (top view). The surfaces are colored from red (-1eV) to blue (+1eV).

<https://doi.org/10.1371/journal.pcbi.1009732.g002>

compatible with the default values for internal groups in a hydrophobic environment ($\sigma = 0.9$, $\epsilon_{in} = 4$), which lead to neutral pKa (7.3 ± 1.0), but supports the default values for external groups at protein surface ($\sigma = 0.7$, $\epsilon_{in} = 8$), which lead to a pKa of 4.1 ± 0 . These latter parameters are consistent with the location of D2.50 facing a water filled cavity. Using these parameters, we obtained similar values for the pKa of D2.50 in AT1 and AT2 from both humans and eels and in the AT1m variant (Fig 2A). This indicates that the presence of N7.46 in the sodium binding site does not alter the physico-chemical properties of D2.50 in human AT1. Notably, we observed a significant decrease in the pKa of D2.50 in CXCR4 due to the presence of a histidine at position 7.45. As a consequence of the pKa of about 4 for D2.50 in AT1h, the sodium pocket has a negative electrostatic potential (Fig 2B) and should bind a sodium ion in case this one reaches the allosteric site.

Sodium ingress pathway

These findings prompted us to initiate a kinetics study to determine whether the sodium ion could reach the allosteric site. We modeled AT1h without sodium and embedded the receptor into a hydrated POPC bilayer, in the presence of NaCl, both on the extracellular and

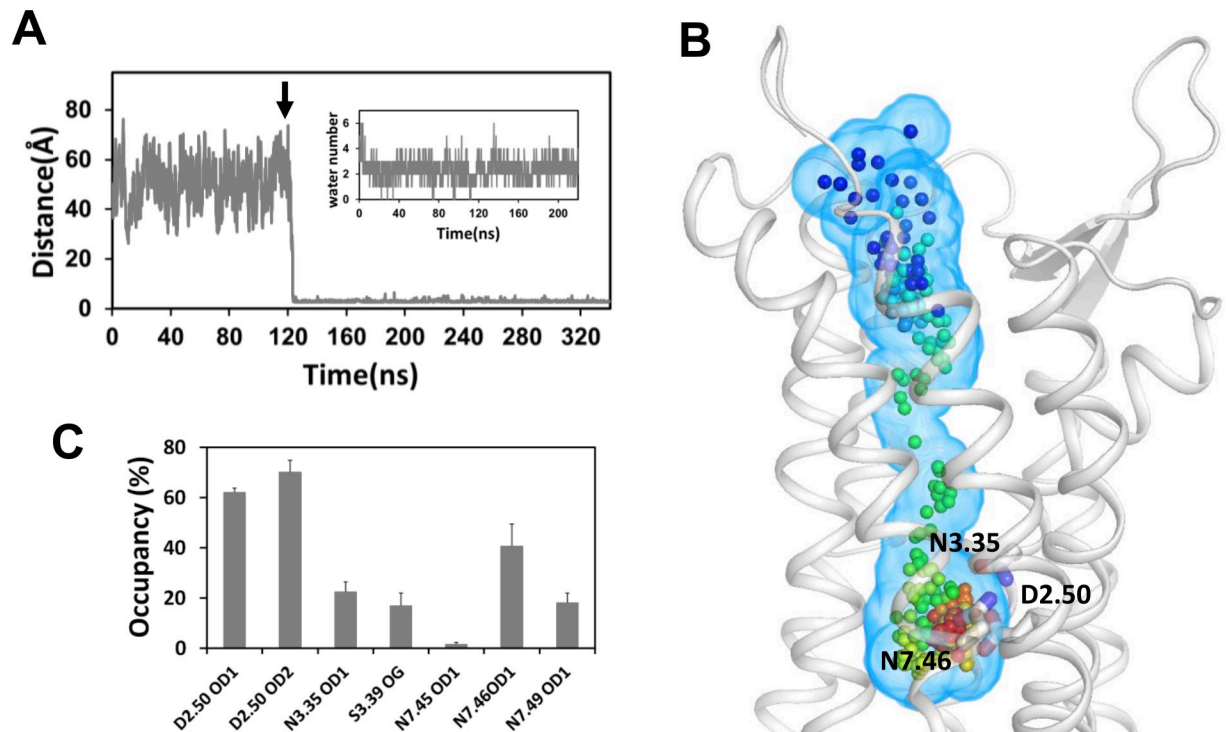


Fig 3. Kinetics of sodium binding to AT1h. (A) Trajectory of the sodium ion reaching the canonical binding site in AT1h, measured by the distance between the sodium ion and the CG atom of D2.50. The arrow indicates the beginning of the accelerated MD simulations after 120 ns of classical MD simulation. The insert represents the number of water molecules in the first coordination shell of the sodium ion from the beginning of the acceleration; (B) Equidistant snapshots of the sodium pathway to the canonical sodium binding site, superposed on the ribbon representation of AT1h (white). The snapshots were extracted every 8 ps, from 1.7 ns to 4.8 ns after the beginning of the acceleration. The rainbow color from blue to red indicates the time. The volumetric map in light blue indicates the volume occupied by the ion during its pathway to the allosteric site. Side chains of D2.50, N3.35 and N7.46 are shown as sticks; (C) Coordination of the sodium ion with protein atoms after desolvation (data are indicated as average \pm standard deviations of three slots in the 20–220 ns range of the aMD trajectory).

<https://doi.org/10.1371/journal.pcbi.1009732.g003>

intracellular water layers. These conditions allow testing the possibility of ingress from either the extracellular or the intracellular side, since both pathways have been observed [33]. We used CXCR4 and OPRD as controls. We did observe fast sodium binding, from the extracellular side, within 10 and 100 ns of classical MD simulations for, respectively, CXCR4 and OPRD, in the presence of 0.15M NaCl (S2 Fig). These data are consistent with previous observations by others [33]. Binding of a sodium ion to AT1h was not as easy and the same conditions (0.15M NaCl, classical MD simulations) failed for AT1h on the sub-microsecond timescale. Since our aim was the feasibility of sodium binding and not the details of the sodium entrance that might necessitate tens of microseconds [33], we used “harsh” conditions, by combining high sodium concentration (1M NaCl) and accelerated MD simulations (see Methods), in an attempt to speed up the kinetics.

After a 120 ns long classical MD step, we initiated acceleration. Under these conditions, we could observe the ingress of a sodium ion into the receptor internal cavity, and its pathway towards the allosteric pocket (Fig 3A and 3B). This step was completed in less than 10 ns of accelerated simulations and was followed by the desolvation of the ion from 5–6 to 2–3 water molecules in the first shell. After desolvation, the sodium ion was coordinated mainly to the D2.50 OD1 and OD2 atoms and to the N7.46 OD1 atom (Fig 3C), as observed under steady-state conditions (see below).

To understand the slow kinetics of sodium ingress, we analyzed the electrostatic potential of the receptor. The extracellular surface of AT1h displayed a complex electrostatic pattern whereas the orthosteric internal cavity was highly dipolar (Fig 2B and 2C). This polarity is due to strictly conserved positive (K5.42 and R4.64) and negative (D6.58 and D7.32) residues that interact with, respectively, the negative carboxyl terminal group and the positive Arg2 side chain of AngII in both AT1 [16] and AT2 [19]. Consequently, the sodium ion ingress is hindered by the strong electrostatic barrier due to K5.42 and R4.64, but remains possible. Similar electrostatic potentials (S3 Fig) were observed in AT1 and AT2 from both eels and humans, and in the AT1m variant for which sodium binding has been evidenced experimentally [34]. Noteworthy, the fast sodium ingress in CXCR4 and OPRD correlates with the negative potential of their extracellular cavity (S3 Fig).

Sodium binding mode in the human AT1 and AT2 receptors

We carried out MD simulations of the human AT1 and AT2 receptors modeled in the inactive state, with a sodium ion located at the canonical site. In both models, N3.35 was initially oriented inward, facing the sodium site, in the *trans* orientation, as observed in the inactive structures of AT1 [17,18]. In Fig 4, we show typical trajectories and analyses for AT1h and AT2h. We monitored sodium motion within its binding site, the RMSD of the ion from its initial position, the distances to the putative protein ligands, the orientation of N3.35, and corresponding histograms (Fig 4A–4D). Based on distance distributions, the threshold for interaction was set at 2.8 Å.

In AT1h (Fig 4, left panels), the initial position of the sodium ion near N3.35 was not stable and the sodium rapidly moved by about 3 Å towards a position in the vicinity of D2.50 and N7.46. In this position, the sodium ion interacted with the OD1 and OD2 atoms of D2.50 and the OD1 atom of N7.46. From this position, the sodium ion transiently escaped to interact with N3.35, N3.39, N7.45, or N7.49. These interactions were similar to those observed for the sodium ion that reached the allosteric site from the extracellular compartment (Fig 3C). The interactions depended on both the sodium position and the dihedral orientation of the side chains. As examples, the snapshots in Fig 4E show that transient interactions of the ion with N3.35 occurred only when the χ_1 and χ_2 dihedral angles of N3.35 were, respectively, in the *trans* and *g+* conformations.

In addition to interacting with protein atoms, the sodium ion also interacted with neighbour water molecules to complete the first coordination shell. These water molecules bridged the sodium ion to either electronegative atoms of the sodium cavity or to other water molecules to form a second coordination shell. A typical snapshot of sodium coordination in AT1h is shown in Fig 5A. The sodium ion directly interacted with the D2.50 OD1, D2.50 OD2, and N7.46 OD1 atoms and with three water molecules. N3.35 did not interact directly with the sodium ion, but through a water molecule, thus contributing to the second coordination shell.

The sodium trajectory was markedly different in AT2h (Fig 4, right panels). Initially, the sodium ion remained near the modeling position (1–2 Å) where it interacted with the D2.50 OD1, N3.35 OD1, and S3.39 OG atoms. However, after about 120 ns, N3.35 underwent an outward rotamerization, modifying the sodium binding cavity (Fig 4A). The sodium moved up to 4–5 Å from the initial position, first towards N7.49 (coordination with N7.49 and D2.50) and then back towards S3.39 in a position where it was coordinated mainly to the D2.50 OD1 atom and, intermittently, to the S3.39 OG atom. Sodium coordination was completed by water. The snapshot in Fig 5B, extracted after the outward rotamerization of N3.35, shows the sodium ion coordinated to the D2.50 OD1 atom and four water molecules.

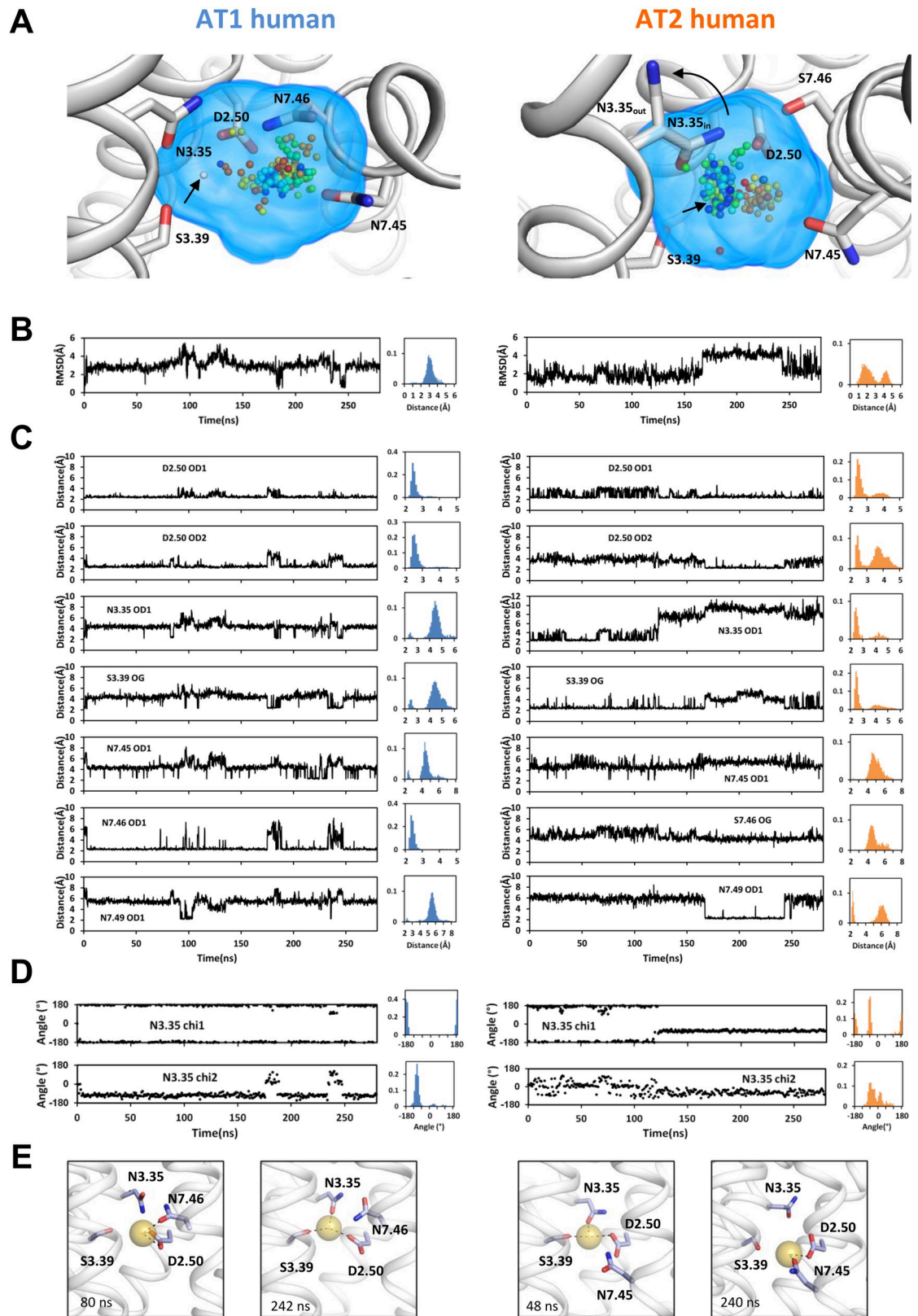


Fig 4. Sodium coordination in representative trajectories of AT1h (left) and AT2h (right). (A) Positioning of the sodium ion as a function of time (from blue to red), during a 280 ns long trajectory, superposed on the receptor initial conformation (white)

ribbon and sticks). The initial position of the sodium ion is indicated by a white sphere and an arrow. For AT2h, N3.35 is shown in both the initial inward and final outward orientations; (B) Root-mean square deviations (RMSD) of the sodium ion, and histograms of the deviations; (C) Distances of the sodium ion to the putative protein ligands and corresponding histograms; (D) Dihedral angles of N3.35 and corresponding histograms; (E) Snapshots showing how side chain rotamerization affects sodium coordination. For AT1, the left snapshot (80 ns) shows interaction of the sodium ion with D2.50 and N7.46. N3.35 is in the *trans* χ_1 and *g*- χ_2 conformation. The right snapshot (242 ns) shows the interaction of the sodium ion with N3.35 upon the transient rotamerization of the N3.35 χ_2 angle. For AT2, the left snapshot (48 ns) shows N3.35 in the inward, *trans* orientation and the sodium ion interacting with D2.50, N3.35, and S3.39. The right snapshot (240 ns) shows N3.35 in the outward, *g*- orientation and the sodium ion interacting with D2.50 and N7.45.

<https://doi.org/10.1371/journal.pcbi.1009732.g004>

Evolution of the sodium binding mode

To decipher the consequences of the mutations observed in the sodium binding site during the evolution of the AT1 and AT2 receptors, we analyzed the coordination of the sodium ion in a variety of AT1 and AT2 receptors. In addition to human receptors, we investigated AT1 and AT2 from eels, as prototypes of fish receptors, three variants and OPRD. The N7.46S mutation in AT1h (AT1m) reversed the evolutionary observed S7.46N mutation. In the chimeric AT1h variant (AT1c), a fragment of transmembrane helix 3 (TM3), from position 3.25 to 3.40, was replaced by the equivalent sequence of AT2h. The AT1c variant has a constitutive activity not observed for AT1 [35]. The S2.49A AT1e variant aimed at analyzing the importance of S2.49 in the sodium binding mode of eel receptors. Finally, OPRD was chosen for comparative purposes because its sodium cavity is similar to AT1 and AT2 and its sodium binding mode has been studied [26,36]. The residues lining the sodium cavity of these receptors are reported in Fig 6A.

As previously, in all the initial models, the sodium ion was positioned near D2.50 and N3.35, as observed in the crystal structure of OPRD, and N3.35 was oriented inward. All the simulations were replicated at least 3 times starting from the same construct. AT1h and AT2h were replicated 6 times from two constructs (see Methods). In all the simulations, the receptors were stable during the trajectory (S4 Fig). The RMSD deviations of the TM helices increased up to around 2 Å at the beginning of the simulations and then reached a plateau after 20–40 ns of simulations. High RMSF were observed for loops whereas TM helices were stable, especially on the intracellular side. The distance between TM3 and TM6, which is the main marker of receptor activation, ranged between 6 and 8 Å, indicating stable inactive structures.

The RMSD of the sodium ion revealed differences, both within and between simulations sets (S5 Fig). Highest RMSD were observed for AT1h and AT1c (3.2 ± 0.9 Å and 2.8 ± 1.1 Å, respectively), the two receptors with an Asn residue at position 7.46. For AT2h, the RMSD increased after outward motion of N3.35 as observed in Fig 4 (2.6 ± 1.1 Å). The RMSDs of the sodium ion in the other receptors were $\leq 2 \pm 1$ Å. In individual runs, the sodium position fluctuated by about 1 Å, with at least partial overlap between runs of the same receptor (S5 Fig). The only exception was observed for AT2e with two narrow sub-sites (RMSD around 0.5 Å) that did not interconvert during the simulations.

Analysis of sodium coordination was based on the distances between the sodium ion, electronegative protein groups and water oxygens, to determine residues and water molecules contributing to the first and second coordination shells (see Methods). Heatmaps of the distances between the sodium ion and putative ligands are given for each individual trajectory as S6 Fig. Because the sodium environment is dynamical with switches of neighbour groups or water molecules between the first and second shells, we also considered both shells together (“total” shell). The results are summarized in Fig 6B and 6C.

The average number of coordination to protein ligands in the first shell decreased from 2.8 ± 0.2 for AT1h and AT1c to 2.0 ± 0.3 for AT1e and AT2h, to be compared to 1.6 ± 0.2 for

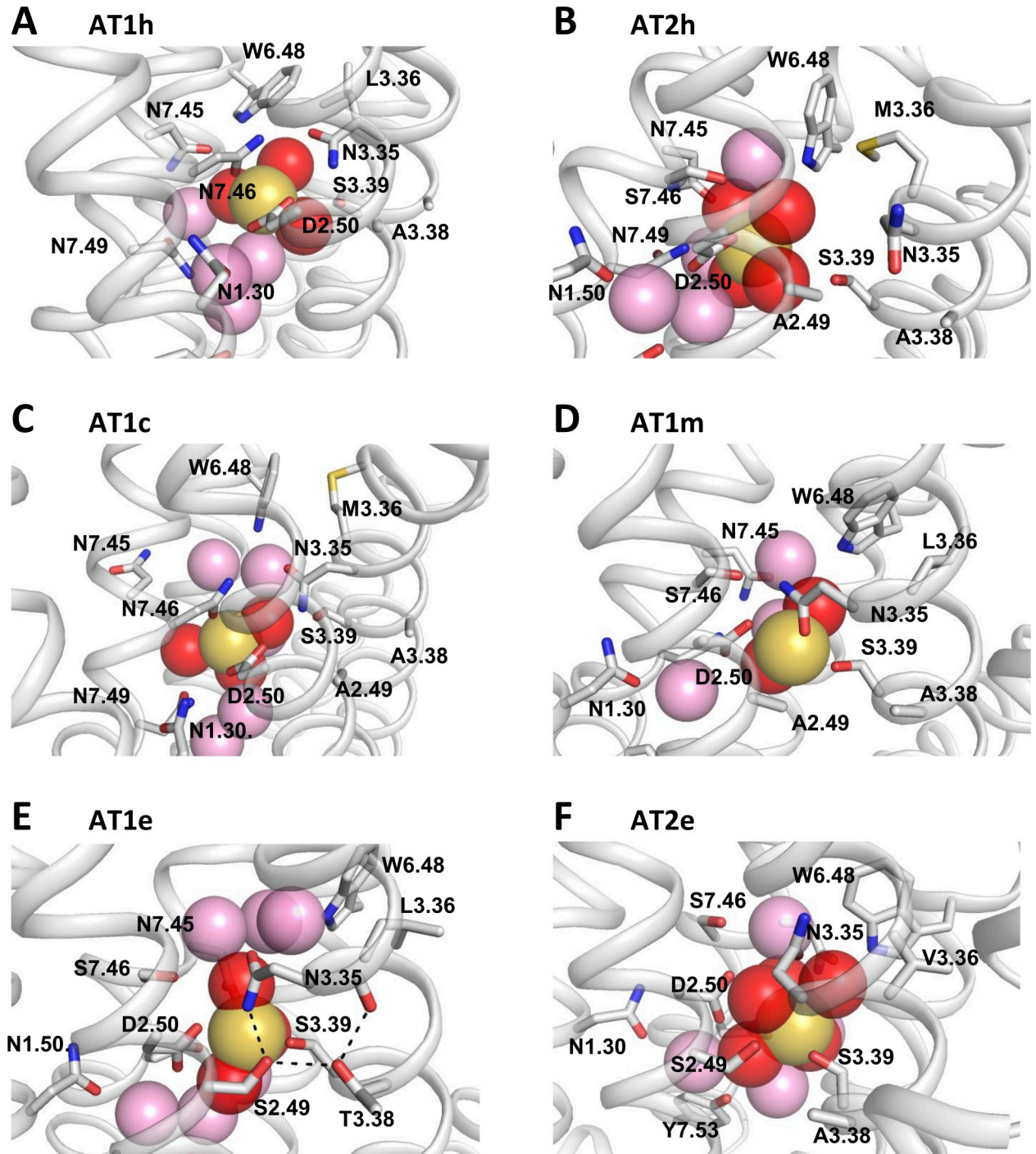


Fig 5. Snapshots of the sodium binding site in AT1 and AT2. (A) In the AT1h snapshot, the sodium ion is coordinated to the D2.50 OD1 and OD2 atoms, to the N7.46 OD1 atom, and to three water molecules; (B) In the AT2h snapshot, the sodium ion is coordinated to the D2.50 OD1 atom and to four water molecules. N3.35 is in the outward orientation; (C) In the AT1c snapshot, the sodium ion is coordinated to the D2.50 OD2 and N7.46 OD1 atoms, and to three water molecules; (D) In the AT1m snapshot, the sodium ion is coordinated to the D2.50 OD1, N3.35 OD1, and S3.39 OG atoms, and to two water molecules; (E) In the AT1e snapshot, the sodium ion is coordinated to the D2.50 OD2 and S3.39 OG atoms, and to three water molecules. The dash lines indicate the H-bonds linking the N3.35 ND2, S3.39 OG, T3.38 OG1, and F3.31 O atoms; (F) In the AT2e snapshot, the sodium ion is coordinated to the S3.39 OG atom and to five water molecules. N3.35 is in the *g+* orientation ($\chi_{1} = 87^{\circ}$). D2.50 is in the second shell. In (A-F), the sodium ion is shown as a yellow sphere. Water molecules participating to the first coordination shell are shown as red spheres. Other water molecules within 6 Å from the sodium ion are shown as pink spheres.

<https://doi.org/10.1371/journal.pcbi.1009732.g005>

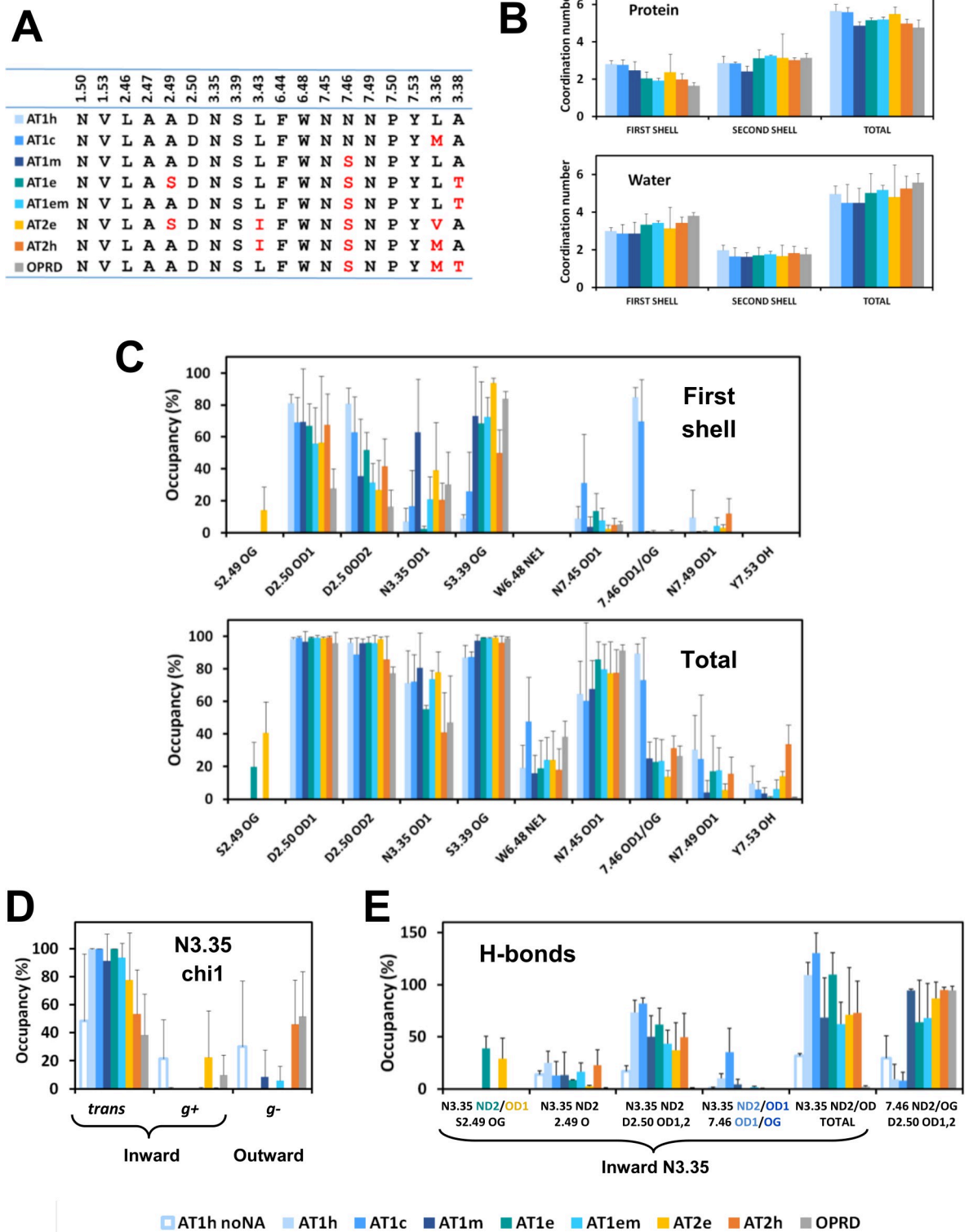


Fig 6. Coordination of the sodium ion in AT1 and AT2 receptors. (A) Residues lining the sodium binding cavity and neighbour positions 3.36 and 3.38 in the receptors under scrutiny; (B) Number of coordination of the sodium ion to protein ligands (top) and water molecules (bottom) in the first shell, in the second shell, and in either shell; (C) Average occupancy of the protein ligands in the first shell (top) and in either the first or the second shell (bottom); (D) Distribution of the chi1 dihedral angle of N3.35 observed during the

simulations; (E) Occupancy of the H-bonds stabilizing the sodium binding site. For N3.35 H-bonds, only frames with inward N3.35 were taken into account. In (B-E), the error bars represent standard error of the mean among replica. The receptors are indicated by the following color code: AT1h: light blue, AT1c: middle blue, AT1m: dark blue, AT1e: deep teal, AT1em: cyan, AT2e: yellow, AT1h: orange, OPRD: grey.

<https://doi.org/10.1371/journal.pcbi.1009732.g006>

OPRD (Fig 6B). The first shell was completed by two to five water molecules. This variability is exemplified in Fig 5 where the number of water molecules in the first shell is 2 for AT1m, 3 for AT1h, AT1e, and AT1c, 4 for AT2h, and 5 for AT2e. The average number of water molecules in the first shell varied from 3.0 ± 0.2 for AT1h to 3.4 ± 0.3 for AT2h and 3.8 ± 0.2 for OPRD. The second shell included around three protein ligands and two water molecules without significant differences between the receptors under scrutiny (Fig 6B). When both shells were taken into consideration, the total number of coordination to protein groups ranged from 5.7 ± 0.4 for AT1h to 5.0 ± 0.2 for AT2h and 4.8 ± 0.4 for OPRD. In any case, four to six water molecules were included in either the first or the second shell, indicating that the presence of an Asn residue at position 7.46 does not significantly reduce the size of the allosteric site or the number of water molecules in the environment of D2.50.

The occupancy rate of the putative ligands is reported in Fig 6C and highlights the specificities of the sodium binding mode in each receptor. Representative snapshots illustrating the variety of sodium coordination are reported in Fig 5.

Both in the AT1h receptor and in the AT1c variant, the sodium ion interacted preferentially with D2.50 (two coordinations) and N7.46 (Fig 6C). In both cases, N3.35 remained in the *trans* orientation (S7 Fig) and, along with S3.39, it participated in the second coordination shell (Fig 6C). A different binding mode was observed for the AT1m variant. In this case, the sodium ion interacted preferentially with the D2.50 OD1, N3.35 OD1, and S3.39 OG atoms. This is similar to the binding mode in the AT2h receptor with inward orientation of N3.35 (Fig 4) and corresponds to the binding mode observed in the crystal structure of OPRD [36].

The AT1e receptor has the same S7.46 pattern as the N7.46S variant but presented a different binding mode, with no coordination to N3.35 in the first shell. In this receptor, a network of polar interactions involving N3.35, D2.50, S2.49, and T3.38 stabilized the position of the N3.35 OD1 atom at about 4 Å of the sodium ion (Fig 5E). Thus, N3.35 did not participate in the first coordination shell but was part of the second shell. In the AT1em variant with an alanine at position 2.49, direct coordination of the sodium ion to N3.35 was restored (Fig 6C).

The AT2e receptor presented unique features among receptors with two narrow sub-sites (S5 Fig). In sub-site 1 (replicas 1 and 3), the sodium interacted preferentially with the D2.50 OD1, N3.35 OD1, and S3.39 OG atoms (occupancy of $80 \pm 5\%$, $56 \pm 1\%$, and $95 \pm 4\%$, respectively). Additional coordination to the S2.49 OG or the D2.50 OD2 atom was also possible. Sub-site 2 (replica 2) was highly unusual with N3.35 in the *g+* orientation, pointing towards W6.48 (Fig 5F). The sodium ion interacted directly with S3.39 only. It also interacted through water molecules with the S2.49 OG, D2.50 OD1 and OD2, N3.35 OD1, W6.48 NE1, and N7.45 OD1 atoms (occupancy of 60, 89, 86, 87, 44, and 54%, respectively).

In OPRD, the interaction of the sodium ion observed in the sodium structure with N3.35 [36] was lost during the MD simulations, due to N3.35 rotamerization, as observed by Filizola et co-workers [37]. In particular, the outward rotation of N3.35 in OPRD was observed in the three runs that have been carried out (S7 Fig). This behaviour was similar to AT2h for which the outward rotation of N3.35 was observed in five out of the six AT2h replicas (S7 Fig). Sodium binding in OPRD was dominated by interactions with S3.39 and, to a lesser extent, with D2.50.

Stability of the sodium binding site through H-bonds

The H-bonds between the polar residues of the sodium binding site contribute to the stability of the receptor. With this regard, comparison of the H-bonding pattern of human AT1 in the presence and absence of sodium is essential to understand the role of the sodium ion in AT1 regulation. Thus, we carried out MD simulations of AT1h in the absence of sodium to determine whether the H-bonding pattern observed between N3.35 and N7.46 in the crystal structures of inactive AT1 [17,18] could stabilize the structure of the sodium-free apo receptor, under dynamical conditions.

We compared the H-bonds involving positions 3.35 and 7.46 in the receptors under scrutiny. As previously pointed out, N3.35 can undergo an outward rotamerization. In this latter orientation, it cannot interact with residues from the sodium binding site (Fig 5B). Thus we monitored the χ_1 angle of N3.35 in all the simulations (S7 Fig) and analyzed the N3.35 H-bonding pattern in the inward orientation only.

A summary of the orientation of N3.35 during the simulations is reported in Fig 6D. In four receptors, AT1h, AT1c, AT1e, and AT2e, we did not observe outward rotamerization of N3.35. N3.35 remained oriented inward, mainly in the *trans* conformation. A high propensity of N3.35 for the outward orientation was observed only in AT2h and OPRD. In addition, an outward rotamerization was observed in (1) one out of five trajectories of the AT1m variant after 150 ns and (2) one out of three trajectories of the AT1em variant after 220 ns. The *g+* conformation, where N3.35 points towards W6.48, was transiently observed in several trajectories, but, in the presence of sodium, was stable only in the AT2e receptor (S7 Fig). Finally, the simulations of AT1h in the absence of bound sodium revealed a variety of orientations for N3.35. A stable *g+* orientation was observed in the first replica, an outward *g-* reorientation was observed after 40 ns in the second replica, whereas the third replica presented a stable *trans* conformation (Figs 6D and S7).

In the next step, we analyzed the H-bonding pattern of N3.35 in the inward orientation only, to determine whether the outward rotamerization was a consequence of impaired H-bonds (Fig 6E). Most contributions came from D2.50. However, when a serine was present at position 2.49, it contributed significantly to additional H-bonds. The highest total scores were larger than 100% (on average, the side chain of N3.35 was involved in at least one H-bond). They were obtained for AT1h, AT1c and AT1e that have either the N7.46 or the S2.49 pattern. The N7.46S mutation in AT1h induced a decrease in the H-bonding pattern of N3.35 from a total of $109 \pm 12\%$ to a total of $64 \pm 42\%$ (for four out of five AT1m trajectories, the total was $52 \pm 11\%$). In AT1e, the S2.49A mutation also induced a strong decrease in the H-bonds involving N3.35 from $110 \pm 21\%$ to $62 \pm 21\%$. Finally, in the absence of sodium, there was a collapse in the H-bonds involving N3.35 from $109 \pm 12\%$ to $32 \pm 2\%$. This indicates that *both* the sodium ion and N7.46 are necessary to stabilize the inward orientation of N3.35 in the AT1h receptor.

As observed for AT1e, in AT2e, S2.49 was also involved in H-bonds with N3.35 (Fig 6E). However, the results were an average of two different cases. In sub-site 1, N3.35 was strongly involved in H-bond interactions with D2.50 and S2.49 (total occupancy of $95 \pm 23\%$). In sub-site 2 (Fig 5F), the occupancy of these H-bonds was low (23%). In AT2h, the total H-bond occupancy of inward N3.35 averaged at $73 \pm 32\%$, which is similar to AT2e and AT1m but contrasts with OPRD (Fig 6E). This indicates that the strong propensity of N3.35 for the outward orientation in AT2h is not related to the absence of H-bonds in the inward orientation. The presence of a methionine at the neighbour position 3.36 was proposed to destabilize AT2h [20]. To support this assumption, in one of the AT2h trajectories, we did observe a bump between M3.36 and W6.48, just preceding the outward rotamerization of N3.35 (S8 Fig).

To summarize the H-bonding pattern of N3.35 observed in the AT1 and AT2 receptors, the inward orientation of N3.35 was always stabilized by H-bond interactions, although at various extents. Thus, in each of the receptors, the collapse of these H-bonds upon the outward rotation of N3.35 should destabilize the receptor structure and favour activation, as observed for CXCR4 [38].

For position 7.46, we aimed at comparing the changes in the H-bonding patterns due to the S7.46N mutation. We thus took into account the entire trajectories. H-bonds between residue N7.46 and N3.35 were marginal in all the simulations, except in one AT1c trajectory. This was also the case for AT1h in the absence of sodium. In this latter case, N7.46 was involved only in H-bonds with D2.50 ($30 \pm 21\%$). In the receptors with a serine at position 7.46, this residue was involved in H-bonds with D2.50 ($> 90\%$ for AT1m, AT2h and OPRD; $> 60\%$ for AT1e, AT1em and AT2e). At least for AT1m, these additional H-bonds do not compensate the detrimental effect of the N7.46S mutation on the H-bonding pattern of N3.35 [34]. A hypothesis for this marginal effect might be maintained H-bonds between S7.46 and D2.50 in the active state, as observed in the 5UNF and 5UNG AT2h structures.

Stability of the sodium binding site through electrostatic interactions

We computed the electrostatic component of the binding free energy of the sodium ion to the different receptors using Delphi. Electrostatic computations were carried out as described in Methods on representative snapshots from MD simulations (every 8 ns of each trajectory). This procedure allows not only an estimate of the average electrostatic free energy but also of its variability within the allosteric cavity during the simulations.

Fig 7A reports boxplots of the electrostatic free energy of sodium binding to the receptors under scrutiny. The medians ranged from -3.0 ± 0.2 kT for AT1h to -2.2 ± 0.3 kT for AT2h but extreme values could reach -1 or -4 kT. The averages were similar to the medians (from 2.9 ± 0.4 for AT1h to 2.2 ± 0.5 for AT2h). For a deeper insight into the variability of the results, for each trajectory, we reported the average free energy of binding as a function of the average number of coordination of the sodium ion to protein ligands in the first shell. As seen in Fig 7B, these parameters were linearly correlated with a coefficient of determination R^2 of 0.90. We also considered the total number of coordination in both the first and second shells. However, in this latter case, the correlation was modest (R^2 equal to 0.32).

In most cases, replicas of the same receptors led to similar number of coordination and free energy of binding. The only exception was observed for AT2e. Sub-site 1 displayed high electrostatic free energy of binding (around -3 kT). Sub-site 2 displayed low electrostatic free energy of binding (1.8 kT). These findings reinforced observations based on the N3.35 H-bonding pattern. Nevertheless, the water-mediated H-bonding pattern in sub-site 2 (Fig 5F) was strong enough to maintain this conformation during the 280 ns of the simulations.

The detailed analysis of the distribution of the binding free energy and the resulting cumulative index (Fig 7C and 7D) revealed significant differences between some of the receptors. In particular, for AT1, the electrostatic free energy of sodium binding decreased in the following order: AT1h $>$ AT1m $>$ AT1e $>$ OPRD. For AT2, the electrostatic component of the free energy of sodium binding decreased in the following order: AT2e $>$ AT2h \approx OPRD. This order was still more marked when only the high affinity site of AT2e was taken into account (AT2e* in Fig 7D).

The electrostatic component is only a part of the total free energy of sodium binding. This one includes components from the reorganization of the receptor and of water, which are exemplified by, respectively, the H-bonding network of N3.35 induced by sodium (Fig 6E) and

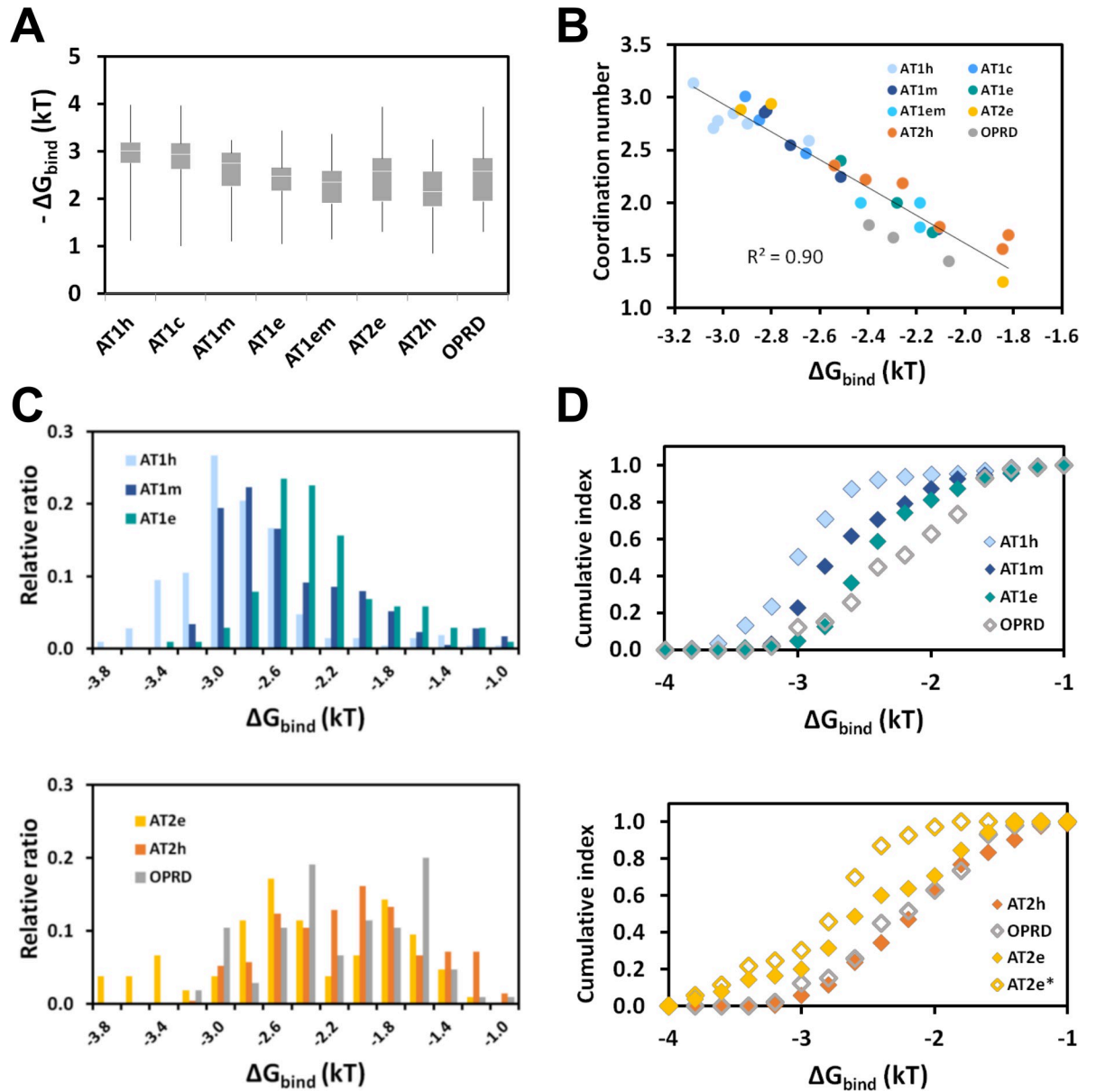


Fig 7. Electrostatic free energy of binding. (A) Boxplots of the electrostatic free energy of sodium binding. The white lines indicate the median. The top and bottom boxes indicate the second and third quartiles, respectively. The whiskers indicate the maximal and minimal values; (B) Dependence of the electrostatic free energy of binding on the number of coordination to protein ligands in the first shell. The dots represent individual trajectories; (C) Distributions of the electrostatic free energy of sodium binding to AT1 receptors (top) and to AT2 receptors and OPRD (bottom); (D) Cumulative index of the electrostatic free energy of sodium binding to AT1 (top) and to AT2 receptors (bottom). OPRD (grey diamonds) is shown for comparative purpose. The receptors are indicated by the following color code: AT1h: light blue, AT1c: middle blue, AT1m: dark blue, AT1e: deepteal, AT1em: cyan; AT2e: yellow, AT1h: orange, OPRD: grey. In (A, C, D), the data take into account all the replica of each receptor, except for the open yellow diamonds in (D) corresponding to the high affinity sub-site (replicas 1 and 3) of AT2e (AT2*).

<https://doi.org/10.1371/journal.pcbi.1009732.g007>

by the desolvation of the sodium ion observed during the kinetics simulations (Fig 3A). These components should have important entropic contributions. Nevertheless, in spite of inherent limits, the estimate of the electrostatic free energy of sodium binding to AT1 and AT2 yields values that are consistent with OPRD.

Discussion

Most class A GPCRs possess a highly conserved sodium binding site. The best known effect of sodium is its role as a negative allosteric modulator (NAM), with a strong stabilization of the inactive state. However, the role of the sodium ion is not limited to this “classical” NAM effect. Presence of sodium has dual effects on both reduced basal signaling and increased receptor stimulation by full agonists (reviewed in [26]). Some MD studies observed a transfer of a sodium ion through the membrane during receptor activation, due to the intracellular egress of the sodium upon activation [39]. Coupling of the sodium transfer from the extracellular to the intracellular side of the membrane is likely to make GPCRs sensitive to the electrostatic potential of the membrane [40,41]. Finally, mutations at the sodium binding site can induce biased or impaired signaling, suggesting a role as an allosteric cofactor of class A GPCR signaling [26].

In view of the putative sodium functions and of the role of residues within the sodium binding site of AT1 in the allosteric control of biased signaling [16,23], the question of sodium binding to the AT1 receptor is of primary importance to understand the activation mechanism of this receptor. To answer this question, we first deciphered the evolutionary history of the AT1 and AT2 receptors, with special focus on the sodium binding site (Fig 1). While the sodium site in AT2 was maintained through evolution with few exceptions (Fig 1), the sodium site of AT1 underwent an unusual S7.46N mutation during the transition to the terrestrial life, suggesting that sodium binding might be impaired [16].

Arguments against maintained sodium binding to human AT1 rely on the absence of significant negative allosteric effects of sodium on receptor basal activity [34] and agonist affinity [16,18]. However, as detailed in a recent review [26], classical NAM effects on agonist binding have not been observed for at least one receptor (the β 1 adrenergic receptor) whose sodium-bound structure has been resolved. Reversely, absence of sodium in crystallographic structures of putative sodium binding receptors is frequent.

Arguments supporting maintained sodium binding to human AT1 rely on several lines of evidence:

1. The N7.46S AT1h variant displays a constitutive activity at low but not at high sodium concentration. Such a behavior is usually interpreted as strong evidence for the stabilization of the inactive state by sodium binding [26].
2. The pK_as of the D2.50 residue in AT1 and AT2 receptors from humans and eels are similar to those computed for OPRD (around 4, see Fig 2A). Consequently, for all the receptors under scrutiny, D2.50 should be negatively charged at neutral pH. The sodium cavity is large enough to contain at least four or five water molecules (Figs 5 and 6), indicating the absence of steric hindrance to sodium binding in AT1h.
3. A sodium ion can actually reach its binding site in human AT1 from the extracellular site (Fig 3). The binding kinetics is very slow as compared to OPRD and CXCR4 (Figs 3 and S2), due to the positive potential of strictly conserved K5.42 (Fig 2B). The kinetics should be similarly slow in any AT1 or AT2 receptor, including the N7.46S AT1 variant (S3 Fig). A consequence of the strongly hindered sodium entrance (low on-rate) is that the sodium dissociation (off-rate) should also be very slow, except upon activation which opens an egress pathway towards the intracellular side. This suggests that the apparent lack of sodium effect on AT1h in experimental data might result from kinetics factors.
4. Comparative analysis of the sodium coordination in the different receptors support the hypothesis of an “efficient” sodium binding mode in AT1h, with 3 interactions with protein

groups that maintain the ion in the close vicinity of D2.50 (Fig 6). This agrees with the electrostatic free energy of sodium binding around -3 kT for AT1h, significantly higher than the value measured for OPRD, around -2kT (Fig 7).

Taken together, these results strongly support the assumption that a functional sodium binding site has been maintained in AT1 throughout evolution. This is also the case for AT2 whose sodium binding site is similar to OPRD (Figs 6 and 7). However, the sodium binding modes markedly differ between eels and humans for both AT1 and AT2, in link with hallmark mutations in the evolution of these receptors during the transition to terrestrial life.

The presence of an asparagine at position 7.46 is a unique feature of amniota AT1. Several important consequences result from this pattern. The first consequence can be inferred from the promiscuity of the N7.46S AT1h variant [34]. The N7.46S variant can be activated by Ang1-7, AngIV, and Ang5-8. Quoting Feng, “The promiscuous agonist specificity displayed by N295S [...] in this study could be disastrous if a similar mutation were to occur in AT1 receptors in nature” [34]. Actually, in nature, this is the reverse mutation which occurred in amniota (Fig 1B), preventing AT1 induced Gq activation by these peptides.

The second consequence can be inferred from the mechanism of biased signaling of AT1 recently proposed [16,23]. This mechanism relies on the orientation of N3.35 that acts as an allosteric switch between β -arrestin and G protein signaling. Extensive MD simulations have shown that the inward orientation of N3.35 yields an “alternative” conformation that prevents G protein binding, whereas the outward conformation yields the canonical conformation allowing both G protein and β -arrestin binding. Relative stabilities of the alternative and canonical conformations were estimated by replica exchange MD simulations. In classical MD, the conformations were observed sequentially starting from the inward orientation [23]. Thus, the stability of the inward orientation of N3.35 in the inactive state should contribute to efficient β -arrestin-biased signaling. We show in this study that the N7.46 pattern stabilizes the inward conformation of N3.35 in inactive AT1, *in the presence of sodium* (Figs 6D and S7). The stabilizing role of N7.46 is highlighted by comparing the H-bonding pattern in wild type AT1h and in the AT1m and AT1c variants (Fig 6E). This is especially remarkable for the AT1c variant. The H-bonds linking N3.35 to D2.50 maintain the inward orientation in the AT1c variant on the sub-microsecond timescale in spite of the destabilizing M3.36 pattern (Fig 6D and 6E).

The properties of eel AT1 cannot be straightforwardly deduced from the N7.46S AT1h variant. In eel AT1, the inward orientation of N3.35 is stabilized by an H-bonding network linking D2.50, N3.35, S2.49, and T3.38 (Figs 5E and 6E). The unusual S2.49/T3.38 pattern (only three occurrences in human GPCRs) is highly conserved in fish AT1 (Fig 1B), suggesting that it might be important for the structure-function relationship of this receptor. Presently, experimental data on eel or, more generally, on fish AT1 receptors are missing. They would be of great help to better understand the specificities of human AT1.

The sodium binding mode of AT2h presents strong similarities with OPRD (Fig 6). The sodium cavities are almost identical, with a single difference at position 3.43 which is not involved in sodium coordination (Ile and Leu for AT2h and OPRD, respectively), whereas position 3.36 is a methionine. In both cases, we observe a propensity of N3.35 for the outward orientation and a similar sodium binding mode, resulting in similar electrostatic free energy of binding for the sodium ion (Fig 7).

Full characterization of AT2h is difficult in the absence of response in conventional GPCR essays [4,13]. However, the pharmacological profiles based on binding affinities reveal that AT2 lacks the specificity of AT1 for AngII as compared to either synthetic or natural AngII derivatives [22,42,43]. The promiscuous binding of AT2 to AngII derivatives suggests a

“relaxed” conformation by analogy with constitutive AT1 variants [22]. The presence of a methionine at position 3.36 has been proposed to facilitate activation of AT2 by AngII derivatives missing the bulky terminal Phe-8 [19]. The outward orientation of N3.35 in the inactive receptor (Fig 6D) prevents the H-bonding network stabilizing the inactive state (Fig 6E) and should contribute to the “relaxed” AT2h conformation.

By contrast with the “relaxed” conformation of human AT2, eel AT2 appears strongly stabilized by sodium binding due to both electrostatic interactions and H-bonding patterns (Figs 6E and 7D). Ang1-7, which can bind to and activate human AT2 [22,42–44], does not activate eel AT2 [45]. These data support the assumption that evolution favoured the promiscuous behaviour of AT2 in amniota.

Further experimental data on fish receptors are mandatory to understanding the physiological consequences of the mutations observed in AT1 and AT2 receptors during the transition to terrestrial life. It would be of primary importance to determine whether fish AT1 is prone to biased signaling. In any case, thanks to the N7.46 pattern, amniota AT1 has been tailored by evolution for efficient β -arrestin-biased signaling. Biased signaling of AT1 was first discovered using *synthetic* ligands [46,47]. The therapeutic potential of biased AT1 ligands was soon acknowledged because β -arrestin signaling is cardioprotective and enhances cardiac performance [48–50]. More generally, development of biased ligands and understanding of underlying mechanisms are an intense research area for drug discovery. Search of *naturally* biased signaling did not receive such attention [51]. However, the evolutionary driven AT1 propensity for bias signaling prompts questions on the underlying endogenous ligand or mechanism. A putative candidate would be Ang1-7. This peptide is an agonist of AT2 and of MAS [44] and a β -arrestin-biased agonist of AT1 [52,53]. Cardioprotective effects of Ang1-7 are in part related to its activity as a biased agonist of AT1 [53]. Another possibility would be mechanotransduction. Indeed, AT1 responds to mechanical stimuli in the absence of agonist [54]. The underlying mechanisms involve β -arrestin signaling [55–57]. Most interestingly, the Frank-Starling law of the heart (increased cardiac filling leads to enhanced contractility) depends on the β -arrestin-biased activation of AT1 by mechanical factors [58]. This mechanism is of fundamental importance in the physiology of vertebrates [59].

The transition from aqueous to terrestrial environment required dramatic changes in the cardiovascular physiology [60]. Fishes have a single circulatory system with a two chamber heart. Terrestrial life requires uptake of oxygen from air, which led to the emergence of lungs, of two circulatory systems, and of a heart with three (amphibian) or four chambers (amniota). The two circulatory systems have very different blood pressure, in particular for amniota which need high pressure in systemic circulation to fight gravity and low pressure in lungs to favor oxygen exchange [60]. These physiological changes were accompanied by major changes in the renin-angiotensin system, with the emergence of novel receptors such as MAS and the development of the counter-regulatory axis of the RAS. The conventional activation of a G protein by AT2, observed in eels [45], has been lost in mammals [4,13] whereas novel receptors, such as MAS, emerged in terrestrial animals [61] and participate in the counter-regulatory arm of the RAS [3,44].

The key mutations observed in AT1 and AT2 (Fig 1) accompanied the transition to terrestrial life. They highlight how a few mutations at a strategic site participate in protein functional evolution. Comparative studies on AT1 and AT2 receptors from different species will lead to a deeper understanding of human cardiovascular regulation and physiology.

Methods

Sequence retrieval, clustering and analysis

AngII receptor sequences were obtained from Uniprot (release-2018-09) [62]. To avoid misclassification issues, we carried out an initial exhaustive request using the keywords “angiotensin” and “IPR000276” (Interpro identifier for class A GPCRs), with a sequence length ≥ 260 . This led to 550 hits. To eliminate false positives or truncated sequences, a multiple sequence alignment was carried out with ClustalX [63], and then a manual selection was carried out using GenDoc [64]. We obtained a set of 318 “clean” sequences that could be classified into 203 AT1 and 115 AT2 sequences based on a preliminary tree.

The over representation of highly similar mammalian sequences in this set (120 mammalian AT1 sequences out of 203, 68 mammalian AT2 sequences out of 115) prompted us to build a non-redundant (NR) set using the perl script NRDB.pl [28]. This program builds clusters with a sequence identity threshold of 90%. This approach led to 37 and 38 clusters from the AT1 and AT2 sequence sets, respectively. Among them, there were 3 and 5 mammalian clusters for AT1 and AT2, respectively. The aligned sequences of each cluster were visually inspected. The selection of the representative sequences was based on the quality of the sequence (full sequences with longest termini were privileged) and of the genome (well-known species from reference genomes were privileged). The alignments of the clusters were used to verify that the sequence patterns observed in the NR sequences were also valid for the whole cluster.

A Neighbour-Joining (NJ) phylogenetic tree of the aligned representative sequences of the clusters was built with the MEGA5 software [65] using 500 replicates for bootstrapping. The logo plots were obtained from the Weblogo site [66].

Noteworthy, there was no sequence of AT2 from cartilaginous fishes reported in Uniprot. To determine whether the absence of AT2 sequences was due to sequencing problems or to a real evolutionary difference, we extended the search of AT2 sequences to Genbank. In the two additional shark genomes reported, we found AT2-like sequences that should be non-functional, either because of an AngII binding impairing K5.42I mutation (*Rhincodon typus*) or of an N-terminal truncation (*Chiloscyllium punctatum*). No other example of such mutations was observed in the entire set of reported AT2 sequences, supporting the assumption that functional AT2 gene was lost in cartilaginous fishes.

Molecular modeling

The inactive state of AT1h with bound sodium was modeled from the sodium-free 4YAY [17] and 4ZUD [18] structures of AT1 and from the sodium-bound structure of OPRD (4N6H) [36], using the homology modeling software MODELLER [67]. The sodium-bound AT1h was modeled from residue 17 to 317 with 7 water molecules in the sodium pocket positioned by homology with 4N6H. The C-terminal helix 8 (H8) has been modeled by homology with 4NH6 because in the 4YAY and 4ZUD structures, H8 is either missing or bent relatively to the membrane plane. The bent orientation of H8 led to swing-saw motion of this helix and of TM7 during MD simulations [68]. The apo model of AT1h without bound sodium was obtained by replacing the sodium ion by a water molecule. The model of the N7.46S AT1h variant was prepared from the model of AT1h using the Charmm-Gui interface [69]. The model of the chimeric AT1c variant was prepared using MODELLER from the model of AT1h by substituting the sequence from residue 3.25 to 3.40 by the equivalent positions in AT2h [35]:

AT1h: CKIASASVSFNLYASV

AT2h: CKVFGSFLTLNMFASI

Two models of inactive human AT2 were built using MODELLER from the active-like 5UNG AT2 structure [21] and from the inactive AT1h model. The intracellular halves of TM5, TM6 and TM7 and the C-terminal H8 from the 5UNG structure were truncated and replaced by the equivalent parts of the AT1h model, including the sodium ion and the water molecules. The two models differed in the length of TM6. In model A, TM6 had the same length as in AT1h. In model B, α -helical restraints were added to the modeling procedure to insure that TM6 length was identical to the length in the 5UNG structure. In either case, the AT2h models were built from residue 35 to 333 and contained a sodium ion and 7 water molecules in the sodium pocket. In addition, N3.35 was modeled in the *trans* orientation as observed in the inactive AT1 structures (PDB 4YAY and 4ZUD) and faced the sodium binding site.

Models of sodium-bound AT1e and AT2e from *Anguilla japonica* were built with MODELLER by homology with the models of AT1h and AT2h (model B). The S2.49A AT1e variant was obtained from the Charmm-Gui interface. For OPRD, we used the 4N6H structure of OPRD with bound sodium and water molecules [36]. A point mutation in this structure was reversed to the wild type sequence with MODELLER.

Molecular dynamics simulations

The models were prepared for molecular dynamics simulations using the Charmm-Gui interface [69]. The models were embedded into a palmitoyl-oleoyl-phosphatidyl-choline (POPC) bilayer with 60 lipids on each layer, and then solvated with a TIP3P model for water molecules with all atoms represented explicitly. For simulations aimed at analyzing the sodium binding mode in models with a sodium ion in the allosteric site, the charges were neutralized by adding chloride ions. For simulations aimed at analyzing the sodium pathway towards the binding pocket, the sodium ion was replaced by a water molecule and the charges were neutralized by adding either 0.15 M or 1M NaCl. In all the simulations, the D2.50 residue located within the sodium binding site was negatively charged, in agreement with the pKa of this residue that was calculated with DelphiPKa [32].

Molecular dynamics simulations of the receptor models embedded into the hydrated POPC bilayer were carried out using NAMD 2.13 MD software [70] and the Charmm36 parameter set [71, 72]. They were performed using the HPC resources of IDRIS, granted by GENCI (www.genici.fr).

The entire assembly was subjected to energy minimization for 5000 steps to remove close contacts between atoms. Equilibration of protein–membrane system was carried out as previously described [27]. The equilibration protocol started with a first 1 ns phase with six steps in which harmonic restraints on protein and lipids were gradually taken off to achieve a smooth relaxation. This phase was followed by visual inspection of the resulting structure, and then by a 20 ns phase carried out under the same conditions as the production run to achieve stable conditions. No restraint was added to the sodium ion. In the first two equilibration steps of the first phase, the NVT ensemble at 310 K and a time step of 1 fs were used. The following equilibration and production steps were carried out at constant temperature (310 K) and pressure (1 atmosphere), using a 2 fs time-step for integration. The Particle Mesh Ewald method (PME) was used to calculate the electrostatic contribution to non-bonded interactions. The cutoff distance for van der Waals interactions was 12 Å with switch starting at 10 Å. The SHAKE algorithm was applied to the system. Each trajectory lasted a total of 280 ns.

For each receptor, the simulations (including the minimization/equilibration steps) were repeated at least three times from the same build (five replicas for AT1m). For AT1h, there were 2 builds with three replicas (runs 1, 2, 3 for build A, runs 4, 5, 6 for build B). For AT2h, there were 2 models/builds with three replicas (runs 1, 2, 3 for model A, runs 4, 5, 6 for model B). The other receptors had a single build.

The accelerated molecular dynamics simulations were based on the biased potential MD approach developed by McCammon and co-workers [73,74]. In this technique, the true potential energy landscape is modified by adding a bias potential that enhances the escape rates from potential wells, yielding accelerated molecular motions and enhanced conformational sampling. We used the dual boost method as implemented in NAMD [75]. This method works by adding two boosts: (1) the dihedral boost [74] is applied to the dihedral energy of the system and accelerates rotamerizations, and (2) the potential boost [73] is applied to the electrostatic and van der Waals energies of the system, including water and ions. Both boosts are necessary for efficient sampling of protein conformational space [76]. Boosts are applied only when the corresponding energies are below a threshold [73,74]. In the “soft” protocol that we use [76], the threshold energies are set to the average energies, E_{dihed_avg} and E_{pot_avg} , computed from classical MD simulations and the acceleration factors are calculated according to:

$$\alpha_{dihed} = \lambda \times \frac{E_{dihed_avg}}{5} \quad (1)$$

$$\alpha_{pot} = \lambda \times N \quad (2)$$

where the acceleration parameter λ is set at 0.3 and N represents the number of atoms in the system. The 20–120 ns range of classical MD simulations was used to calculate the acceleration parameters. The snapshot obtained after 120 ns was used to initiate the accelerated simulation.

Trajectories were graphically analyzed with VMD [77] and PYMOL (DeLano Scientific LLC, San Francisco, USA). Quantitative analyses were carried out with home-developed scripts, written in tcl, for use with the VMD software, and in R using functions from the Bio3D R package [78]. The cut-offs for sodium coordination were set as follows: (1) For water, 3 Å and 6 Å for the first and second shells, respectively; (2) For protein ligands, 2.8 Å for the first shell and 6 Å for the second shell with the additional requirement of a first shell water molecule within 3 Å. H-bonds were measured with the H-bond facility of VMD with cutoffs of 30° and 3.5 Å. For H-bonds involving N3.35, the trajectories with outward rotamerization of N3.35 within the first 50 ns were removed from the analysis. For the other trajectories, only the snapshots with inward orientation of N3.35 were taken into account. Heatmaps were prepared with R using the ComplexHeatmap package [79]. Volumetric maps were computed with VMD. PYMOL was used for figure preparation.

Electrostatics potential and pKa computations

Electrostatics potentials were computed by a Poisson-Boltzmann approach with the Delphi program [80], using the smooth Gaussian dielectric function developed by Alexov and al [30]. Similarly, the pKa values of the receptor ionizable residues were computed with DelphiPKa using the same Gaussian dielectric function [31,32]. Both Delphi and DelphiPKa have been installed locally. In the Gaussian representation of atomic density resulting into a smooth dielectric function, two parameters have to be defined: the width of the Gaussian function, σ , and the internal permittivity of the protein, ϵ_{in} . The default values for external ionizable groups are $\epsilon_{in} = 8$ and $\sigma = 0.7$, whereas the default values for internal groups are $\epsilon_{in} = 4$ and $\sigma = 0.9$. In these studies, we aimed at determining the electrostatic properties of D2.50, located at the sodium binding pocket, in AT1 and AT2. The sodium pocket is lined by several polar groups, including the aspartic acid D2.50, and is wide enough to contain a sodium ion and several water molecules. To determine adequate parameters, we compared the pKa of D2.50 in human AT1 and OPRD, as a function of a variety of ϵ_{in} and σ values. For each pair, analyses were done on seven equally spaced frames from MD trajectories. All the combinations of ϵ_{in}

and σ values led to similar pKa values for D2.50 in OPRD and AT1h (S1 Fig). Most of them led to a pKa value < 6 , consistent with deprotonation of D2.50 in physiological conditions for both receptors. This was not the case for the default values for protein hydrophobic interior that led to a value of 7.3 ± 1.0 for the pKa of D2.50 in OPRD, which is not consistent with the observed affinity of the sodium ion for this receptor [26]. By contrast, the default values for ionizable groups on protein surface led to a D2.50 pKa value of 4.1, which supports a negatively charged aspartic acid interacting with the sodium ion. Consequently, the latter set ($\epsilon_{in} = 8$ and $\sigma = 0.7$) was used for both electrostatic potentials and pKa calculations. For each receptor investigated, pKa computations were done on a representative set of seven frames from MD trajectories.

The electrostatics potential maps were calculated with the Gaussian-based method on the receptors placed at the center of a cube filled with an aqueous medium, implicitly modeled by a dielectric permittivity ϵ_{ex} of 80. The axes of the cube were oriented along the principal axis of the solute. The edge size of the cube, L , depends on the maximal dimension of the solute L_p ($L_p = L * 0.7$). The space was discretized with 1.8 grid points by angstrom. The permittivity of the receptor, ϵ_{in} , was 8 and the Gaussian width, σ , was 0.7. The ionic strength was set at 0.15 M. The atom charges and sizes used for the calculations were based on the default Charmm parameters provided by Delphi (Charmm22).

The electrostatic component of the free energy of sodium binding to the receptors under scrutiny was computed with Delphi. We used the grid energy difference method [81] and the Gaussian dielectric function with the same parameters as previously ($\epsilon_{in} = 8$ and $\sigma = 0.7$). The ionic strength was 0.15 M. The grid energy difference method requires three Delphi runs, for the sodium-bound receptor, the sodium-free receptor and the sodium ion. The separated elements and the complex must be kept in the same position in the cube used for electrostatics computations. The electrostatic component of the binding free energy was computed from the difference in the total grid energy (electrostatic energy computed at each grid point):

$$\Delta G_{\text{bind}} = E_{\text{grid}}(\text{receptor-Na}) - E_{\text{grid}}(\text{receptor}) - E_{\text{grid}}(\text{Na}) \quad (3)$$

To obtain representative sets of the sodium position in its site, we extracted snapshots at equidistant intervals (35 snapshots separated by 8 ns) from each trajectory, yielding at least 100 snapshots by receptor. All the snapshots of the sodium-bound receptor were superimposed on the initial model and saved as PDB files, using VMD. The PDB files for the sodium-free receptor and for the sodium ion were obtained from the sodium-bound receptor PDB files with shell scripts. The size grid (233 points) was large enough to insure an occupancy $< 70\%$ in each case.

Supporting information

S1 Fig. Comparison of D2.50 pKa in OPRD and AT1.

(TIF)

S2 Fig. Binding kinetics of the sodium ion to CXCR4 and OPRD.

(TIF)

S3 Fig. Electrostatics potentials of the internal cavities of AT1, AT2, CXCR4 and OPRD.

(TIF)

S4 Fig. Stability of the receptors during MD simulations.

(TIF)

S5 Fig. Comparison of the sodium trajectories during MD simulations.

(TIF)

S6 Fig. Heatmaps of the distances between the sodium ion and the putative ligands.
(TIF)

S7 Fig. Heatmaps of the N3.35 chi1 dihedral angle.
(TIF)

S8 Fig. Outward rotamerization of N3.35 in AT2h.
(TIF)

Acknowledgments

We thank Dr Guillaume Lebon (Institut de génomique fonctionnelle, Montpellier, France) for critical reading of the manuscript and stimulating discussion.

Author Contributions

Conceptualization: Marie Chabbert.

Data curation: Marie Chabbert.

Formal analysis: Rym Ben Boubaker, Asma Tiss.

Funding acquisition: Daniel Henrion, Marie Chabbert.

Investigation: Rym Ben Boubaker, Asma Tiss.

Methodology: Marie Chabbert.

Project administration: Marie Chabbert.

Resources: Daniel Henrion.

Software: Rym Ben Boubaker, Marie Chabbert.

Supervision: Hajer Guissouma, Marie Chabbert.

Validation: Rym Ben Boubaker, Asma Tiss, Marie Chabbert.

Visualization: Rym Ben Boubaker, Asma Tiss, Marie Chabbert.

Writing – original draft: Marie Chabbert.

Writing – review & editing: Daniel Henrion, Hajer Guissouma, Marie Chabbert.

References

1. Balakumar P, Jagadeesh G. A century old renin-angiotensin system still grows with endless possibilities: AT1 receptor signaling cascades in cardiovascular pathophysiology. *Cellular signalling*. 2014; 26(10):2147–60. <https://doi.org/10.1016/j.cellsig.2014.06.011> PMID: 25007996
2. Forrester SJ, Booz GW, Sigmund CD, Coffman TM, Kawai T, Rizzo V, et al. Angiotensin II Signal Transduction: An Update on Mechanisms of Physiology and Pathophysiology. *Physiological reviews*. 2018; 98(3):1627–738. <https://doi.org/10.1152/physrev.00038.2017> PMID: 29873596
3. Paz Ocaranza M, Riquelme JA, Garcia L, Jalil JE, Chiong M, Santos RAS, et al. Counter-regulatory renin-angiotensin system in cardiovascular disease. *Nature reviews Cardiology*. 2020; 17(2):116–29. <https://doi.org/10.1038/s41569-019-0244-8> PMID: 31427727
4. Karnik SS, Unal H, Kemp JR, Tirupula KC, Eguchi S, Vanderheyden PM, et al. International Union of Basic and Clinical Pharmacology. XCIX. Angiotensin Receptors: Interpreters of Pathophysiological Angiotensinergic Stimuli [corrected]. *Pharmacological reviews*. 2015; 67(4):754–819. <https://doi.org/10.1124/pr.114.010454> PMID: 26315714
5. Oparil S, Schmieder RE. New approaches in the treatment of hypertension. *Circulation research*. 2015; 116(6):1074–95. <https://doi.org/10.1161/CIRCRESAHA.116.303603> PMID: 25767291

6. Li XC, Zhang J, Zhuo JL. The vasoprotective axes of the renin-angiotensin system: Physiological relevance and therapeutic implications in cardiovascular, hypertensive and kidney diseases. *Pharmacological research*. 2017; 125(Pt A):21–38. <https://doi.org/10.1016/j.phrs.2017.06.005> PMID: 28619367
7. McFall A, Nicklin SA, Work LM. The counter regulatory axis of the renin angiotensin system in the brain and ischaemic stroke: Insight from preclinical stroke studies and therapeutic potential. *Cellular signalling*. 2020; 76:109809. <https://doi.org/10.1016/j.cellsig.2020.109809> PMID: 33059037
8. Rajagopal S, Ahn S, Rominger DH, Gowen-MacDonald W, Lam CM, Dewire SM, et al. Quantifying ligand bias at seven-transmembrane receptors. *Molecular pharmacology*. 2011; 80(3):367–77. <https://doi.org/10.1124/mol.111.072801> PMID: 21610196
9. Sauliere A, Bellot M, Paris H, Denis C, Finana F, Hansen JT, et al. Deciphering biased-agonism complexity reveals a new active AT1 receptor entity. *Nature chemical biology*. 2012; 8(7):622–30. <https://doi.org/10.1038/nchembio.961> PMID: 22634635
10. Kaschina E, Namsolleck P, Unger T. AT2 receptors in cardiovascular and renal diseases. *Pharmacological research*. 2017; 125(Pt A):39–47. <https://doi.org/10.1016/j.phrs.2017.07.008> PMID: 28694144
11. Matavelli LC, Siragy HM. AT2 receptor activities and pathophysiological implications. *Journal of cardiovascular pharmacology*. 2015; 65(3):226–32. <https://doi.org/10.1097/FJC.000000000000208> PMID: 25636068
12. de Gasparo M, Siragy HM. The AT2 receptor: fact, fancy and fantasy. *Regulatory peptides*. 1999; 81(1–3):11–24. [https://doi.org/10.1016/s0167-0115\(99\)00023-3](https://doi.org/10.1016/s0167-0115(99)00023-3) PMID: 10395404
13. Porrello ER, Delbridge LM, Thomas WG. The angiotensin II type 2 (AT2) receptor: an enigmatic seven transmembrane receptor. *Frontiers in bioscience*. 2009; 14:958–72. <https://doi.org/10.2741/3289> PMID: 19273111
14. Sadybekov A, Katritch V. Breaking the Enigma Code of Angiotensin II Type 2 Receptor Signaling. *Structure*. 2020; 28(4):390–2. <https://doi.org/10.1016/j.str.2020.03.004> PMID: 32268077
15. Wingler LM, McMahan C, Staus DP, Lefkowitz RJ, Kruse AC. Distinctive Activation Mechanism for Angiotensin Receptor Revealed by a Synthetic Nanobody. *Cell*. 2019; 176(3):479–90 e12. <https://doi.org/10.1016/j.cell.2018.12.006> PMID: 30639100
16. Wingler LM, Skiba MA, McMahan C, Staus DP, Kleinhenz ALW, Suomivuori CM, et al. Angiotensin and biased analogs induce structurally distinct active conformations within a GPCR. *Science*. 2020; 367(6480):888–92. <https://doi.org/10.1126/science.aay9813> PMID: 32079768
17. Zhang H, Unal H, Gati C, Han GW, Liu W, Zatspein NA, et al. Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell*. 2015; 161(4):833–44. <https://doi.org/10.1016/j.cell.2015.04.011> PMID: 25913193
18. Zhang H, Unal H, Desnoyer R, Han GW, Patel N, Katritch V, et al. Structural Basis for Ligand Recognition and Functional Selectivity at Angiotensin Receptor. *The Journal of biological chemistry*. 2015; 290(49):29127–39. <https://doi.org/10.1074/jbc.M115.689000> PMID: 26420482
19. Asada H, Horita S, Hirata K, Shiroishi M, Shiimura Y, Iwanari H, et al. Crystal structure of the human angiotensin II type 2 receptor bound to an angiotensin II analog. *Nature structural & molecular biology*. 2018; 25(7):570–6. <https://doi.org/10.1038/s41594-018-0079-8> PMID: 29967536
20. Asada H, Inoue A, Ngako Kadji FM, Hirata K, Shiimura Y, Im D, et al. The Crystal Structure of Angiotensin II Type 2 Receptor with Endogenous Peptide Hormone. *Structure*. 2020; 28(4):418–25 e4. <https://doi.org/10.1016/j.str.2019.12.003> PMID: 31899086
21. Zhang H, Han GW, Batyuk A, Ishchenko A, White KL, Patel N, et al. Structural basis for selectivity and diversity in angiotensin II receptors. *Nature*. 2017; 544(7650):327–32. <https://doi.org/10.1038/nature22035> PMID: 28379944
22. Miura S, Karnik SS. Angiotensin II type 1 and type 2 receptors bind angiotensin II through different types of epitope recognition. *Journal of hypertension*. 1999; 17(3):397–404. <https://doi.org/10.1097/00004872-199917030-00013> PMID: 10100078
23. Suomivuori CM, Latorraca NR, Wingler LM, Eismann S, King MC, Kleinhenz ALW, et al. Molecular mechanism of biased signaling in a prototypical G protein-coupled receptor. *Science*. 2020; 367(6480):881–7. <https://doi.org/10.1126/science.aaz0326> PMID: 32079767
24. Sealfon SC, Chi L, Ebersole BJ, Rodic V, Zhang D, Ballesteros JA, et al. Related contribution of specific helix 2 and 7 residues to conformational activation of the serotonin 5-HT2A receptor. *The Journal of biological chemistry*. 1995; 270(28):16683–8. <https://doi.org/10.1074/jbc.270.28.16683> PMID: 7622478
25. Katritch V, Fenalti G, Abola EE, Roth BL, Cherezov V, Stevens RC. Allosteric sodium in class A GPCR signaling. *Trends in biochemical sciences*. 2014; 39(5):233–44. <https://doi.org/10.1016/j.tibs.2014.03.002> PMID: 24767681
26. Zarzycka B, Zaidi SA, Roth BL, Katritch V. Harnessing Ion-Binding Sites for GPCR Pharmacology. *Pharmacological reviews*. 2019; 71(4):571–95. <https://doi.org/10.1124/pr.119.017863> PMID: 31551350

27. Taddese B, Deniaud M, Garnier A, Tiss A, Guissouma H, Abdi H, et al. Evolution of chemokine receptors is driven by mutations in the sodium binding site. *PLoS computational biology*. 2018; 14(6): e1006209. <https://doi.org/10.1371/journal.pcbi.1006209> PMID: 29912865
28. Holm L, Sander C. Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics*. 1998; 14(5):423–9. <https://doi.org/10.1093/bioinformatics/14.5.423> PMID: 9682055
29. Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, et al. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nature ecology & evolution*. 2018; 2(11):1761–71. <https://doi.org/10.1038/s41559-018-0673-5> PMID: 30297745
30. Li L, Li C, Zhang Z, Alexov E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *Journal of chemical theory and computation*. 2013; 9(4):2126–36. <https://doi.org/10.1021/ct400065j> PMID: 23585741
31. Pahari S, Sun L, Basu S, Alexov E. DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate. *Proteins*. 2018; 86(12):1277–83. <https://doi.org/10.1002/prot.25608> PMID: 30252159
32. Wang L, Li L, Alexov E. pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa. *Proteins*. 2015; 83(12):2186–97. <https://doi.org/10.1002/prot.24935> PMID: 26408449
33. Selvam B, Shamsi Z, Shukla D. Universality of the Sodium Ion Binding Mechanism in Class A G-Protein-Coupled Receptors. *Angewandte Chemie*. 2018; 57(12):3048–53. <https://doi.org/10.1002/anie.201708889> PMID: 29405531
34. Feng YH, Zhou L, Qiu R, Zeng R. Single mutations at Asn295 and Leu305 in the cytoplasmic half of transmembrane alpha-helix domain 7 of the AT1 receptor induce promiscuous agonist specificity for angiotensin II fragments: a pseudo-constitutive activity. *Molecular pharmacology*. 2005; 68(2):347–55. <https://doi.org/10.1124/mol.105.011601> PMID: 15901848
35. Feng YH, Miura S, Husain A, Karnik SS. Mechanism of constitutive activation of the AT1 receptor: influence of the size of the agonist switch binding residue Asn(111). *Biochemistry*. 1998; 37(45):15791–8. <https://doi.org/10.1021/bi980863t> PMID: 9843384
36. Fenalti G, Giguere PM, Katritch V, Huang XP, Thompson AA, Cherezov V, et al. Molecular control of delta-opioid receptor signalling. *Nature*. 2014; 506(7487):191–6. <https://doi.org/10.1038/nature12944> PMID: 24413399
37. Shang Y, LeRouzic V, Schneider S, Bisignano P, Pasternak GW, Filizola M. Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions. *Biochemistry*. 2014; 53(31):5140–9. <https://doi.org/10.1021/bi5006915> PMID: 25073009
38. Taddese B, Garnier A, Abdi H, Henrion D, Chabbert M. Deciphering collaborative sidechain motions in proteins during molecular dynamics simulations. *Scientific reports*. 2020; 10(1):15901. <https://doi.org/10.1038/s41598-020-72766-1> PMID: 32985550
39. Yuan S, Vogel H, Filipek S. The role of water and sodium ions in the activation of the mu-opioid receptor. *Angewandte Chemie*. 2013; 52(38):10112–5. <https://doi.org/10.1002/anie.201302244> PMID: 23904331
40. Vickery ON, Carvalheda CA, Zaidi SA, Pislakov AV, Katritch V, Zachariae U. Intracellular Transfer of Na(+) in an Active-State G-Protein-Coupled Receptor. *Structure*. 2018; 26(1):171–80 e2. <https://doi.org/10.1016/j.str.2017.11.013> PMID: 29249607
41. Vickery ON, Machtens JP, Tamburrino G, Seeliger D, Zachariae U. Structural Mechanisms of Voltage Sensing in G Protein-Coupled Receptors. *Structure*. 2016; 24(6):997–1007. <https://doi.org/10.1016/j.str.2016.04.007> PMID: 27210286
42. Bosnyak S, Jones ES, Christopoulos A, Aguilar MI, Thomas WG, Widdop RE. Relative affinity of angiotensin peptides and novel ligands at AT1 and AT2 receptors. *Clinical science*. 2011; 121(7):297–303. <https://doi.org/10.1042/CS20110036> PMID: 21542804
43. Bouley R, Perodin J, Plante H, Rihakova L, Bernier SG, Maletinska L, et al. N- and C-terminal structure-activity study of angiotensin II on the angiotensin AT2 receptor. *European journal of pharmacology*. 1998; 343(2–3):323–31. [https://doi.org/10.1016/s0014-2999\(97\)01549-5](https://doi.org/10.1016/s0014-2999(97)01549-5) PMID: 9570482
44. Karnik SS, Singh KD, Tirupula K, Unal H. Significance of angiotensin 1–7 coupling with MAS1 receptor and other GPCRs to the renin-angiotensin system: IUPHAR Review 22. *British journal of pharmacology*. 2017; 174(9):737–53. <https://doi.org/10.1111/bph.13742> PMID: 28194766
45. Wong MK, Takei Y. Angiotensin AT2 receptor activates the cyclic-AMP signaling pathway in eel. *Molecular and cellular endocrinology*. 2013; 365(2):292–302. <https://doi.org/10.1016/j.mce.2012.11.009> PMID: 23174758
46. Ahn S, Shenoy SK, Wei H, Lefkowitz RJ. Differential kinetic and spatial patterns of beta-arrestin and G protein-mediated ERK activation by the angiotensin II receptor. *The Journal of biological chemistry*. 2004; 279(34):35518–25. <https://doi.org/10.1074/jbc.M405878200> PMID: 15205453

47. Wei H, Ahn S, Shenoy SK, Karnik SS, Hunyady L, Luttrell LM, et al. Independent beta-arrestin 2 and G protein-mediated pathways for angiotensin II activation of extracellular signal-regulated kinases 1 and 2. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(19):10782–7. <https://doi.org/10.1073/pnas.1834556100> PMID: 12949261
48. DeWire SM, Violin JD. Biased ligands for better cardiovascular drugs: dissecting G-protein-coupled receptor pharmacology. *Circulation research*. 2011; 109(2):205–16. <https://doi.org/10.1161/CIRCRESAHA.110.231308> PMID: 21737816
49. Violin JD, Lefkowitz RJ. Beta-arrestin-biased ligands at seven-transmembrane receptors. *Trends in pharmacological sciences*. 2007; 28(8):416–22. <https://doi.org/10.1016/j.tips.2007.06.006> PMID: 17644195
50. Kim KS, Abraham D, Williams B, Violin JD, Mao L, Rockman HA. beta-Arrestin-biased AT1R stimulation promotes cell survival during acute cardiac injury. *American journal of physiology Heart and circulatory physiology*. 2012; 303(8):H1001–10. <https://doi.org/10.1152/ajpheart.00475.2012> PMID: 22886417
51. Kenakin T. Biased Receptor Signaling in Drug Discovery. *Pharmacological reviews*. 2019; 71(2):267–315. <https://doi.org/10.1124/pr.118.016790> PMID: 30914442
52. Galandrin S, Denis C, Boularan C, Marie J, M'Kadmi C, Pilette C, et al. Cardioprotective Angiotensin-(1–7) Peptide Acts as a Natural-Biased Ligand at the Angiotensin II Type 1 Receptor. *Hypertension*. 2016; 68(6):1365–74. <https://doi.org/10.1161/HYPERTENSIONAHA.116.08118> PMID: 27698068
53. Teixeira LB, Parreiras ESLT, Bruder-Nascimento T, Duarte DA, Simoes SC, Costa RM, et al. Ang-(1–7) is an endogenous beta-arrestin-biased agonist of the AT1 receptor with protective action in cardiac hypertrophy. *Scientific reports*. 2017; 7(1):11903. <https://doi.org/10.1038/s41598-017-12074-3> PMID: 28928410
54. Mederos y Schnitzler M, Storch U, Gudermann T. AT1 receptors as mechanosensors. *Current opinion in pharmacology*. 2011; 11(2):112–6. <https://doi.org/10.1016/j.coph.2010.11.003> PMID: 21147033
55. Wang J, Hanada K, Gareri C, Rockman HA. Mechanoactivation of the angiotensin II type 1 receptor induces beta-arrestin-biased signaling through Galphai coupling. *Journal of cellular biochemistry*. 2018; 119(4):3586–97. <https://doi.org/10.1002/jcb.26552> PMID: 29231251
56. Tang W, Strachan RT, Lefkowitz RJ, Rockman HA. Allosteric modulation of beta-arrestin-biased angiotensin II type 1 receptor signaling by membrane stretch. *The Journal of biological chemistry*. 2014; 289(41):28271–83. <https://doi.org/10.1074/jbc.M114.585067> PMID: 25170081
57. Rakesh K, Yoo B, Kim IM, Salazar N, Kim KS, Rockman HA. beta-Arrestin-biased agonism of the angiotensin receptor induced by mechanical stress. *Science signaling*. 2010; 3(125):ra46. <https://doi.org/10.1126/scisignal.2000769> PMID: 20530803
58. Abraham DM, Davis RT 3rd, Warren CM, Mao L, Wolska BM, Solaro RJ, et al. beta-Arrestin mediates the Frank-Starling mechanism of cardiac contractility. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(50):14426–31. <https://doi.org/10.1073/pnas.1609308113> PMID: 27911784
59. Jacob R, Dierberger B, Kissling G. Functional significance of the Frank-Starling mechanism under physiological and pathophysiological conditions. *European heart journal*. 1992; 13 Suppl E:7–14. https://doi.org/10.1093/eurheartj/13.suppl_e.7 PMID: 1478214
60. Schulte K, Kunter U, Moeller MJ. The evolution of blood pressure and the rise of mankind. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association—European Renal Association*. 2015; 30(5):713–23. <https://doi.org/10.1093/ndt/gfu275> PMID: 25140012
61. Bader M, Alenina N, Andrade-Navarro MA, Santos RA. MAS and its related G protein-coupled receptors, Mrgprs. *Pharmacological reviews*. 2014; 66(4):1080–105. <https://doi.org/10.1124/pr.113.008136> PMID: 25244929
62. UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*. 2019; 47(D1):D506–D15. <https://doi.org/10.1093/nar/gky1049> PMID: 30395287
63. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23(21):2947–8. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
64. Nicholas KB, Nicholas HB, Deerfield DW. GeneDoc: analysis and visualization of genetic variation. *EMBNEW News*. 1999; 4(14):3.
65. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*. 2011; 28(10):2731–9. <https://doi.org/10.1093/molbev/msr121> PMID: 21546353
66. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome research*. 2004; 14(6):1188–90. <https://doi.org/10.1101/gr.849004> PMID: 15173120

67. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*. 1993; 234(3):779–815. <https://doi.org/10.1006/jmbi.1993.1626> PMID: 8254673
68. Tiss A, Ben Boubaker R, Henrion D, Guissouma H, Chabbert M. Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom. *Methods Mol Biol*. 2021; 2315:73–97. https://doi.org/10.1007/978-1-0716-1468-6_5 PMID: 34302671
69. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*. 2008; 29(11):1859–65. <https://doi.org/10.1002/jcc.20945> PMID: 18351591
70. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*. 2005; 26(16):1781–802. <https://doi.org/10.1002/jcc.20289> PMID: 16222654
71. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *Journal of chemical theory and computation*. 2012; 8(9):3257–73. <https://doi.org/10.1021/ct300400x> PMID: 23341755
72. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B*. 2010; 114(23):7830–43. <https://doi.org/10.1021/jp101759q> PMID: 20496934
73. de Oliveira CA, Hamelberg D, McCammon JA. On the application of accelerated molecular dynamics to liquid water simulations. *The journal of physical chemistry B*. 2006; 110(45):22695–701. <https://doi.org/10.1021/jp062845o> PMID: 17092018
74. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*. 2004; 120(24):11919–29. <https://doi.org/10.1063/1.1755656> PMID: 15268227
75. Wang Y, Harrison CB, Schulten K, McCammon JA. Implementation of Accelerated Molecular Dynamics in NAMD. *Computational science & discovery*. 2011; 4(1). <https://doi.org/10.1088/1749-4699/4/1/015002> PMID: 21686063
76. Miao Y, Nichols SE, Gasper PM, Metzger VT, McCammon JA. Activation and dynamic network of the M2 muscarinic receptor. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(27):10982–7. <https://doi.org/10.1073/pnas.1309755110> PMID: 23781107
77. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996; 14(1):33–8, 27–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
78. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. 2006; 22(21):2695–6. <https://doi.org/10.1093/bioinformatics/btl461> PMID: 16940322
79. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016; 32(18):2847–9. <https://doi.org/10.1093/bioinformatics/btw313> PMID: 27207943
80. Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, et al. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC biophysics*. 2012; 5:9. <https://doi.org/10.1186/2046-1682-5-9> PMID: 22583952
81. Panday SK, Shashikala MHB, Koirala M, Pahari S, Chakraborty A, Peng Y, et al. Modeling electrostatics in molecular biology: A tutorial of Delphi and associated resources. *Living J Comp Mol Sci* 2019; 1(2):10841. <https://doi.org/https://doi.org/10.33011/livecoms.1.2.10841>

Annexe II.2

Chapitres méthodologiques

Chapitre 1: [Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom](#). Tiss A*, **Ben Boubaker R***, Henrion D, Guissouma H, Chabbert M. *Methods Mol Biol.* 2021; 2315:73-97. doi: 10.1007/978-1-0716-1468-6_5. PMID: 34302671. *Equal participation.

Chapitre 2: Homology modeling in the twilight zone: Improved accuracy by sequence space analysis. **Ben Boubaker R**, Tiss A, Henrion D and Chabbert M. *Methods Mol Biol* (in press).

Contexte et principales contributions

Durant la période de confinement dû à la crise sanitaire en 2020, j'ai contribué à la rédaction de deux chapitres de modélisation dans la série *Methods in Molecular biology dont je suis co-premier et premier auteur*. Le premier chapitre porte sur la modélisation par homologie des RCPG. Il est intitulé « Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom » et a été publié en juin 2021 [29]. Il prend le récepteur de la bradykinine comme cas d'étude. Le deuxième chapitre porte sur la modélisation par threading et prend la phéromone PRF1 de la salamandre comme cas d'étude. Il est intitulé « Homology modeling in the twilight zone: Improved accuracy by sequence space analysis ». Il a été accepté fin 2020 et est actuellement sous presse dans un volume intitulé « Homology Modeling » et édité par le Dr S. Filipek..

Lors de la rédaction du premier chapitre sur la modélisation des RCPG, j'ai contribué par une inspection de toutes les structures cristallographique des RCPG, permettant d'identifier les RCPG ayant des structures résolues ainsi que de caractériser ces structures (états d'activation, présence ou absence de l'ion sodium, résolutions des boucles...etc.). Cela a permis aussi d'identifier la longueur de toutes les boucles résolues et ainsi les patrons potentiels pour modéliser les boucles du récepteur cible. Ainsi, j'ai développé un script R pour l'analyse des longueurs de boucle des RCPG humains et leur visualisation sous forme d'histogramme.

Dans le cadre du deuxième chapitre de modélisation, j'ai contribué à la mise au point d'une nouvelle version du package Bios2mds, un package R dédié à l'analyse de séquences biologiques par multidimensional scaling (MDS). Cette nouvelle version nous a permis de visualiser l'espace de séquence de la phéromone PRF1 de la salamandre et de détecter des séquences intermédiaires entre CNTF et PRF1, permettant ainsi la modélisation de la protéine PRF1 à partir de la structure cristallographique de CNTF. J'ai aussi contribué à la réalisation et à la comparaison des modèles de PRF1 réalisés avec des outils automatiques (Robetta avec l'option RosettaCM, Phyre2 et I-TASSER de Zhang Lab) basés sur des méthodes de threading, avec le modèle personnalisé obtenu avec Modeller

Chapter 5

Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom

Asma Tiss, Rym Ben Boubaker, Daniel Henrion, Hajer Guissouma, and Marie Chabbert

Abstract

With 700 members, G protein-coupled receptors (GPCRs) of the rhodopsin family (class A) form the largest membrane receptor family in humans and are the target of about 30% of presently available pharmaceutical drugs. The recent boom in GPCR structures led to the structural resolution of 57 unique receptors in different states (39 receptors in inactive state only, 2 receptors in active state only and 16 receptors in different activation states). In spite of these tremendous advances, most computational studies on GPCRs, including molecular dynamics simulations, virtual screening and drug design, rely on GPCR models obtained by homology modeling. In this protocol, we detail the different steps of homology modeling with the MODELLER software, from template selection to model evaluation. The present structure boom provides closely related templates for most receptors. If, in these templates, some of the loops are not resolved, in most cases, the numerous available structures enable to find loop templates with similar length for equivalent loops. However, simultaneously, the large number of putative templates leads to model ambiguities that may require additional information based on multiple sequence alignments or molecular dynamics simulations to be resolved. Using the modeling of the human bradykinin receptor B1 as a case study, we show how several templates are managed by MODELLER, and how the choice of template(s) and of template fragments can improve the quality of the models. We also give examples of how additional information and tools help the user to resolve ambiguities in GPCR modeling.

Key words Homology modeling, MODELLER, GPCR, Membrane receptor, Loop modeling, Computational biology

1 Introduction

Class A (rhodopsin-like) G protein-coupled receptors (GPCRs) form the largest family of transmembrane receptors in the human genome [1, 2]. They include about 300 non-olfactory receptors classified into a dozen of subfamilies (Fig. 1) and 400 olfactory receptors. These receptors allow the transfer of information from an

Asma Tiss and Rym Ben Boubaker contributed equally to this work.

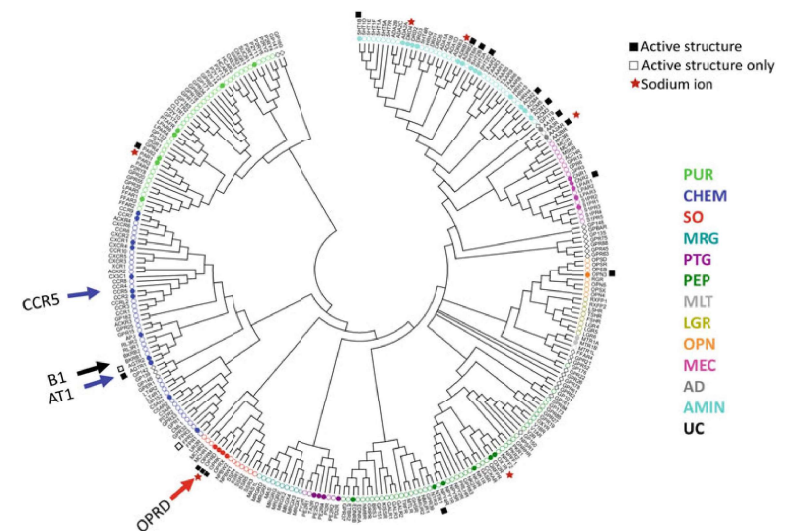


Fig. 1 Evolutionary tree of human class A GPCRs, indicating structurally resolved receptors by subfamily. The color code of the circles depends on the GPCR subfamily (PUR: light green, CHEM: blue, SO: red, MRG: teal, PTG: violet, PEP: dark green, MLT: light gray, LGR: khaki, OPN: orange, MEC: magenta; AD: dark gray; AMIN: cyan; UC: black). The circle is open when the receptor has no structure and closed when the receptor has at least one structure. Receptors with at least one active state structure are indicated by squares (black squares for both inactive and active state structures, open squares for active state structures only). Receptors with a sodium-bound structure are indicated by red stars. For the resolved structures, orthologous receptors from any organism are considered. The receptor names in the tree correspond to the UniProt identification name without the "HUMAN" extension. The arrows indicate the B1 target and the CCR5, AT1, and OPRD templates that are discussed in the text

extracellular signal to the cell cytoplasm. The extracellular signal is usually a ligand that, after binding to the receptor, induces a conformational change from an inactive to an active conformation, which in turn binds to and activates effector proteins such as G proteins and arrestins. GPCRs participate in numerous physiopathological processes and are the target of about 30% of presently available drugs [3].

The pharmacological importance of GPCRs explains the considerable effort spent to resolve their molecular structure. These receptors share a common fold comprising seven transmembrane helices (TM), with highly conserved anchor positions in each helix [4]. The first structure of a GPCR, rhodopsin, was resolved in 2000 [5]. The resolution of a second receptor, the β_2 adrenergic receptor, required 7 years [6]. Since then, several technical locks were broken, and new GPCR structures from different families followed almost non-stop. Now in 2020, the structures of 57 unique class A

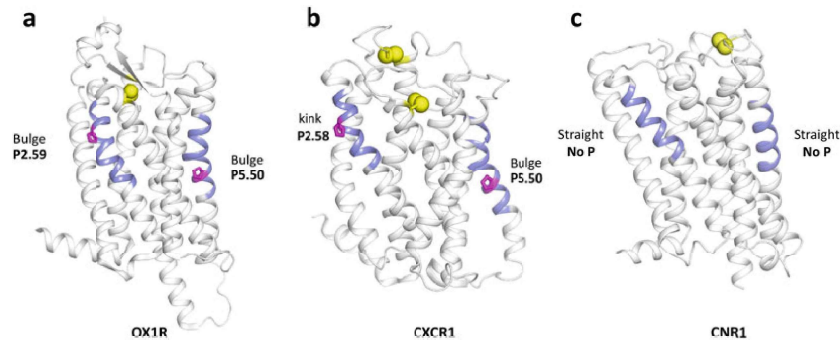


Fig. 2 Structural diversity of class A GPCRs. (a) The orexin receptor 1, OX1R, has a bulge in TM2 (P2.59 pattern), a bulge in TM5 (P5.50 pattern), and a disulfide bond linking the extracellular terminus of TM3 to ECL2; (b) The receptor CXCR1 has a kink in TM2 (P2.58 pattern), a bulge in TM5 (P5.50 pattern), the TM3-ECL2 disulfide bond, and an additional disulfide bond linking the extracellular terminus of TM7 to the N-terminus of the receptor; (c) The cannabinoid receptor 1, CNR1, has two straight TM2 and TM5 helices (no proline residue in these helices) and an unusual disulfide bond in ECL2. In the three cases, the receptors are in an inactive state. The TM2 and TM5 proline residues are magenta. The PDB codes are 6TOD (OX1R), 2NLN (CXCR1) and 5TGZ (CNR1)

receptors (all of them being non-olfactory receptors), totalizing 48 more than 300 structures in different complexes or activation 49 states, have been resolved by X-ray crystallography, serial femtosec- 50 ond crystallography, cryo-electron microscopy, or solid state NMR, 51 providing a deeper understanding of the mechanism of action of 52 GPCRs [7–10]. Nevertheless, in spite of this avalanche of struc- 53 tures, about 80% of the non-olfactory GPCRs still do not have 54 resolved structures. 55

Among the 57 class A receptors with at least one structure, 56 39 receptors have been resolved in inactive state only, two receptors 57 in active state only, and 16 receptors in different activation states. 58 These structures reveal the structural diversity of the transmem- 59 brane fold (Fig. 2) and have evidenced the large conformational 60 change occurring upon activation (e.g., the type 1 angiotensin II 61 receptor, AT1, in Fig. 3a), with a pivotal motion of TM6 that opens 62 an intracellular cavity allowing effector binding [11]. They have 63 also revealed an allosteric binding site for the sodium ion (e.g., the 64 δ opioid receptor, OPRD, in Fig. 3b) that acts as an allosteric 65 modulator [12] and should be taken into account in molecular 66 docking to GPCRs [13]. Inactive states of GPCRs are adapted for 67 drug design or virtual screening of antagonists/inverse agonists, 68 whereas active states should be adapted for drug design of agonists 69 [14, 15]. Design of biased agonists (agonists specific of a signaling 70 pathway) raises an additional level of complexity [16]. Deorphan- 71 sation of orphan receptors often relies on virtual screening 72

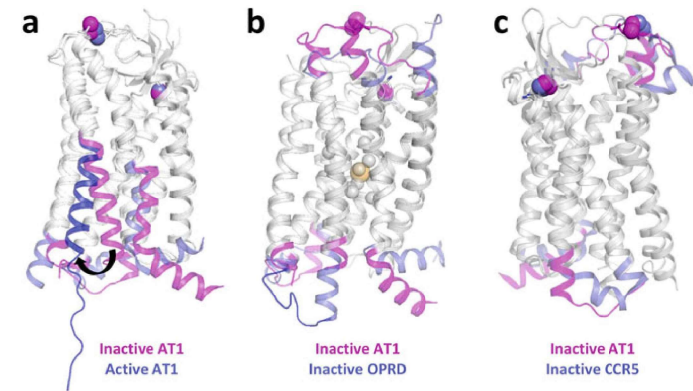


Fig. 3 Comparison of templates for B1 modeling. The inactive structure of AT1 (PDB 4YAY) is superposed on the structure of (a) active AT1 (PDB 6O50), (b) inactive OPRD (PDB 4N6H), and (c) inactive CCR5 (PDB 4MBS). The structures are shown as white ribbons with differences highlighted in magenta for inactive AT1 and slate for the other structures. The sulfur atoms of the disulfide bonds are shown as magenta and slate balls for inactive AT1 and the other receptors, respectively. In (a), the arrow indicates the pivotal motion of TM6 upon activation. The tilted orientation of H8 in the inactive structure is not observed in the active structure. In (b), the sodium ion and the neighbor water molecules present in the structure of OPRD are shown as yellow and gray spheres, respectively. The orientation highlights the tilted orientation of H8 in AT1 and the structure of ICL3 in OPRD. In (c), the orientation highlights the difference in the structures of ICL2 and the tilted orientation of the TM6 extracellular terminus in CCR5

[17, 18]. In addition, understanding the mechanisms of action of 73 GPCRs requires molecular dynamics (MD) simulations of recep- 74 tors in different activation states. These studies still rely heavily on 75 molecular modeling. Thus, even in the age of the GPCR structure 76 boom, molecular modeling by homology is still necessary. 77

This chapter discusses basic and advanced features of molecular 78 modeling of class A GPCRs with the homology-based MODEL- 79 LER software [19, 20]. The concept of homology modeling is 80 based on evolution. Proteins are homologous when they share a 81 common ancestor, which results in structure and, to a lesser extent, 82 to sequence and function similarity [21]. The unknown structure of 83 the target protein is modeled from the known structure of (at least) 84 one homologous protein (the template) and the sequence align- 85 ment of the target versus the template(s). With keeping in mind 86 that homology modeling programs always give a model, the main 87 questions concern the quality of the model(s) and the way(s) to 88 improve them. The inputs have to be carefully prepared and the 89 outputs carefully evaluated, with critical assessment of available 90 information and assumptions. 91

Difficulties of homology modeling depend on the sequence 92 similarities between the target and the template. It is generally 93 assumed that, above a threshold of 30% of identity, homology 94 modeling may be quite straightforward, even if caveats can occur 95 [22]. Below this limit, modeling usually becomes increasingly diffi- 96 cult because of structural variations. In spite of their common fold, 97 each GPCR structure is unique [7–10], with structural variations to 98 adapt to the variety of ligands (Fig. 2). It is thus mandatory to 99 carefully select template(s) to correctly translate sequence into 100 structural similarities. Moreover, as active structures are far less 101 frequent than inactive ones (Fig. 1), modeling of active states 102 requires special attention. 103

Here, we detail the procedure of molecular modeling of 104 rhodopsin-like GPCRs and the customization of the modelling 105 process that is possible with standalone MODELLER. We empha- 106 size the importance of evolutionary information [23–26] for tem- 107 plate selection and model customization. Using the human 108 bradykinin receptor B1 with bound sodium ion in the inactive 109 state as a case study, we show how MODELLER manages multiple 110 templates and how the quality of the resulting model(s) can be 111 improved by the careful choice of multiple template(s) and of 112 template fragment(s). 113

2 Materials 114

2.1 Hardware 115

A computer running Linux/Unix, Apple Mac OS X (10.6 or later), 116 or Microsoft Windows (XP or later). 117

2.2 Software 118

The MODELLER 9.23 program [19, 27] can be downloaded and 119 installed from <http://salilab.org/modeller>. It is written in Fortran 120 90 and uses Python for its control language. Thus, all input scripts 121 to MODELLER are Python scripts. In addition to MODELLER, 122 several tools are required:

1. A text editor capable of outputting plain text files, such as the 123 free and open-source software `gedit`, available for Linux, 124 Windows and Mac OS X. 125
2. A molecular viewing tool, such as the PyMOL Molecular Gra- 126 phics System, Schrodinger, LLC, or UCSF Chimera [28]. The 127 structural alignment tools they include are usually sufficient for 128 comparison of templates and models. 129
3. A multiple sequence alignment (MSA) program, such as Clus- 130 talW [29], T-COFFEE [30], or MUSCLE [31]. 131
4. A visual software for MSA editing. We recommend GeneDoc 132 (Multiple Sequence Alignment Editor, Analyzer, and Shading 133 Utility) developed at Pittsburgh Supercomputing Center's 134 National Resource for Biomedical Supercomputing [32]. 135

2.3 Input Files 136

MODELLER needs three types of input files: 137

1. The .py script file written in Python. 138
2. At least one .pdb file containing the structure of one template. 139
3. The .ali file indicating the alignment between the target and the 140 template(s). This alignment file has a specific format for 141 MODELLER. 142 143

2.4 Additional Tools and Websites 144

1. For template structures: The Protein Data Bank (PDB) 144 (<https://www.rcsb.org/>). 145
2. For the search of homologous templates: direct mining of the 146 PDB with the Blastp utility in UniProt (<https://www.uniprot.org/>), Expasy (<https://www.expasy.org/>), or NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). 147 148 149
3. For general information on receptors: UniProt, including 150 Swiss-Prot with reviewed entries and UniProtKB with auto- 151 matic entries. Swiss-Prot centralizes functional information on 152 proteins with detailed annotations that are curated by experts. 153
4. For secondary structure prediction: JPred4 [33], accessible at 154 <http://www.compbio.dundee.ac.uk/jpred4/index.html>. 155
5. For quality check of the model: PROCHECK [34] accessible at 156 <https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>. 157 158
6. For building non-redundant homologous sequence sets: the 159 `nrdb90.pl` perl script [35] that can be found at <ftp://biodisk.org/Program/Perl/Bioperl/nrdb90.pl>. 160 161 162

2.5 GPCR-Specific Websites 163

Two websites gather invaluable, updated information on GPCRs: 164

1. The GPCRDB database [10, 36], accessible at <https://gpcrdb.org>, which gathers sequences, structures, genetic variations, 165 mutational data, and structure-based alignments, with a classi- 166 fication based on ligands. 167
2. The Zhang Lab website at the university of Michigan (<https://zhanglab.cmb.med.umich.edu>) with online services for 168 GPCRs (GPCR-EXP: Experimental structures, GPCR-RD: 169 Experimental restraints, GLASS: GPCR-ligand association 170 database). 171 172 173

3 Methods 174

The general principles of homology modeling by satisfaction of 175 spatial restraints as implemented in MODELLER are detailed in 176 **Note 1**. Here we detail the steps necessary for homology modeling 177 of class A GPCRs in different activation states with MODELLER. 178 For clarity purpose, as a case study, we will model the human 179

bradykinin receptor B1 (UniProt access code: BKRB1_HUMAN, 180
thereafter B1), in an inactive state with a bound sodium ion at the 181
allosteric binding site. 182

3.1 Gaining Information on the Target Receptor

Before initiating the modeling project, gather available information 183
on the target. In addition to literature, check UniProt. In its 184
reviewed part, each entry includes not only core data (amino acid 185
sequence, name, citations, etc.) but also classifications, cross- 186
references, mutational data, natural variants, amino acid modifica- 187
tions (modified residues, glycosylation, lipidation, disulfide bonds, 188
etc.), binary interactions, subunit structure, and, possibly, 3D 189
structures or models. In addition to UniProt, the GPCRDB data- 190
base [36] and the Zhang Lab website [37] provide updated, 191
GPCR-specific information (see Materials). 192

3.2 Gaining Information on the Receptor Subfamily

Class A receptors are classified into a dozen of evolutionary-based 194
subfamilies that make consensus in the literature [2, 25]. Sequence 195
identity is usually around 25–30% within subfamilies, but may be as 196
low as 15% between subfamilies. In addition, around 20 receptors 197
cannot be related to any subfamily and are unclassified (UC). 198
Evolutionary classification gives information on several sequence 199
and structural patterns (*see Note 2*). A first pattern concerns two 200
transmembrane helices, TM2 and TM5, which have variable pro- 201
line motifs and, thus, structural variability [24, 25, 38], as exem- 202
plified in Fig. 2. A second pattern concerns the disulfide bonds 203
stabilizing the receptor extracellular domain. 204

In addition, careful analysis of the multiple sequence alignment 205
of the target sequence with receptors from its subfamily may reveal 206
unusual patterns that should be taken into account in the modeling 207
procedure and/or the subsequent computational studies, in partic- 208
ular (1) indels in transmembrane helices [24, 38], (2) mutations 209
in the highly conserved sequence motifs of class A GPCRs (DRY in 210
TM3, CWXP in TM6, NPXXY in TM7) [4], and (3) mutations in 211
the sodium binding site [12, 26]. 212

For example, in our case study, the receptor B1 belongs to the 213
subfamily of chemotactic (CHEM) receptors (Fig. 1). It is closely 214
related to a set of receptors for vasoactive peptide receptors (the 215
bradykinin receptor B2, the angiotensin II receptors AT1 and AT2, 216
the apelin receptor) and to the chemokine receptors. As these 217
receptors, it possesses the P2.58/P5.50 proline pattern character- 218
ized by a proline kink in TM2 and a proline bulge in TM5. In 219
addition, as most chemotactic receptors, it possesses two disulfide 220
bonds between TM3 and extracellular loop 2 (ECL2) and between 221
the N-terminus and ECL3. Please note that, in UniProt, it is 222
classified as class A GPCR ([IPR000276](#) GPCR_Rhodpsn), brady- 223
kinin receptor B1 ([IPR001186](#) Brdyknn_1_rcpt), and bradykinin 224
receptor ([IPR000496](#) Brdyknn_rcpt which includes B1 and B2 225
receptors). The intermediary subfamily level is not provided by 226
UniProt (*see Note 3*) but may be found in the literature [2, 24]. 227

3.3 Searching for Suitable Template(s)

The primary requirement for homology modeling is the identifica- 228
tion of at least one known structure with similarity to the target 229
sequence to be used as a template. Now, numerous structures of 230
GPCRs are available and the choice of the “best” template(s) has to 231
be done carefully, keeping in mind that the selection of the closest 232
homolog based on sequence identity does not guaranty model 233
accuracy [39, 40]. Nevertheless, as each subfamily presents unique 234
structural features that have to be considered for modeling, it is 235
recommended to select at least one template from the same sub- 236
family, with similar proline and cysteine patterns. This should be 237
possible in most cases except for the LGR (leucine-rich repeat 238
GPCRs) and MRG (Mas-related GPCRs) subfamilies for which 239
no structure has been resolved (Fig. 1). 240

Search of close homologs with resolved structures in the Pro- 241
tein Data Bank can be carried out straightforwardly by blasting the 242
target sequence from UniProtKB (hits with 3D structures or mod- 243
els), Expasy or NCBI (hits with 3D structures only). Homology 244
modeling may be based on several templates, rather than a single 245
one. Indeed, the use of several templates approximately equidistant 246
from the target sequence, with a weighting based on sequence 247
similarity, generally increases the model accuracy [41]. 248

For our test case B1, the closest hits (in decreasing order of 249
similarity) are AT1 (6 structures), AT2 (5), CCR5 (6), CCR9 (1), 250
APJ (1), and then, the opioid receptors (6, 4, and 4 structures for 251
delta, kappa, and mu, respectively). All these receptors have the 252
same TM2 and TM5 proline patterns as B1 and all, except the 253
opioid receptors, have the double disulfide bonds in the extracellu- 254
lar domain. All, except AT2, have inactive state structures. AT1, 255
AT2, and the opioid receptors have active state structures. Among 256
these structures, only the δ opioid receptor (OPRD) has an inactive 257
structure with a sodium ion bound at the allosteric binding site 258
[42]. In these structures, N- and C-termini and loops may be 259
missing, because of truncation, intrinsic disorder, or replacement 260
by a fusion protein for crystallization purposes. 261

3.4 Selecting Suitable Template(s)

Template selection depends on (1) the target state to be modeled 262
(see the large outward motion of TM6 that differentiates the 263
inactive from the active state of AT1 in Fig. 3a), (2) the resolution 264
of the N- and C-termini and of the loops, and (3) the inclusion of 265
bound ligands (receptor agonists or antagonists, ions, lipids, and 266
water molecules). The experimental resolution of the crystal struc- 267
tures is usually not the most crucial factor. By contrast, the mod- 268
ifications done for crystallization (mutations, nanobodies, and 269
insertions) and quaternary/crystal contacts may have an impact 270
on the receptor structure and have to be carefully evaluated. The 271
specific sequence/structure patterns of GPCRs also need careful 272
evaluation. It is necessary to check what is “normal” or not in the 273
structures under scrutiny, such as the curvature of helices, 274
275
276

additional bends or kinks, unusual distortions (π or 3_{10} helices), unusual structural motifs, and helix orientations, in particular for the C-terminal helix 8 (H8).

Figure 3 displays four putative templates for the modeling of B1: AT1 (4YAY) [43], CCR5 (4MBS) [44], and OPRD (4N6H) [42] in inactive states, and AT1 in an active state (6OSO) [45]. Among these templates, only OPRD has a resolved ICL3 loop and a bound sodium ion with coordination water, but this receptor does not possess the second disulfide bond between the N-terminus and ECL3. Inactive AT1 has an unstructured ICL2 and an unconventional, tilted positioning of the C-terminus (also observed in CCR5, but not in active AT1). In CCR5, active AT1 and OPRD, ICL2 is structured as a helix. In addition, CCR5 has an unusual outward orientation of the C-terminus of TM6. Thus, different templates will be necessary to model the entire sequence of B1, but their selection will affect the resulting models.

3.5 Mining and Analyzing Receptor Homologs

In many cases, ambiguities on “best” templates are observed. For example, in the case of B1, should we model ICL2 with an α -helix, as observed in the OPRD, CCR5, and active AT1 templates, or as a coil as observed in the inactive AT1 template? Other examples of ambiguities can be found when the target receptor does not possess one of the anchor prolines in the TM helices [38] or when there are insertions/deletions in loops of the target versus the template(s). In ambiguous cases, analysis of orthologous or paralogous sequence sets may be informative [24, 38]. To build these sets, use the InterPro identifiers (*see* Note 3) and mine UniprotKB. The subsequent analysis may be facilitated by building a nonredundant set with sequence identities lower than 90%. This can be carried out easily with the nrdb90.pl perl script [35]. The multiple sequence alignment of the resulting set should be carefully checked and, if necessary, manually corrected with Genedoc [32]. This procedure considerably reduces the number of orthologous sequences and, subsequently, allows better visualization of key evolutionary events. In some cases, secondary structure prediction using JPred4 [33] and customized MSA may be informative (see example below).

3.6 Including Ligand or Nonprotein Residues

If the template contains a ligand, water molecules, ions, or other nonprotein residues (anything marked as HETATM in the PDB file), MODELLER can include them into the generated model. By default, all HETATM records are ignored. They are read when the `env.io.hetatm` and `env.io.water` Booleans are set to `TRUE`. Ions and water molecules are indicated by “i” or “w” in the alignment and participate in the refinement step. The unrecognized residues or ligands are indicated with the BLK (“.”) residue type (both in the template and the target sequences) to copy them as rigid bodies into the model. The atom coordinates are transferred, but the BLK residues are static and do not participate in the refinement steps.

The HETATM records are read from the templates, in the order they are written in the PDB file. Thus, they must be indicated by the appropriate symbol (“.”, “w”, or “i”) in the *same order* as in the alignment. If the template includes extra HETATM ligands that must not be modeled in the target, manually delete them in the PDB file or align them with a gap (“-”) in the target sequence. If a chain break (“/”) is added immediately before the “.” residues in the alignment, this will force the ligands to have a different chain identifier (ID). If you model a ligand peptide, you need to add a chain break (“/”) between the receptor and the peptide ligand, resulting in different chain IDs.

To model the sodium binding site in an inactive receptor, use as template the sodium binding site of another receptor with similar binding mode [12]. Depending on the similarities with the other templates, you may prefer using the entire receptor or the sodium binding fragment including residues from TM2, TM3, and TM7. Most importantly, do not forget to include water molecules in the sodium binding site (Fig. 3b). They will not be present in the model if you do not indicate them. Be cautious when you select water molecules to be included. A strong overlap between the preliminary target model and a water molecule may lead to the crash of the MODELLER job. In that case, carefully check the water molecules to be included.

3.7 Building Receptor Chimera

In several cases, it may be interesting to use different templates for different parts of the target. In addition to the sodium binding site detailed above, two other cases are worth mentioning:

1. *Modeling an active target*: Comparison of active/inactive structures reveals that the largest structural changes during activation occur in TM5, TM6, and TM7 [6]. As active templates are less frequent than inactive ones (Fig. 1), an active state target may be better modeled as a chimera of (parts of) closely related inactive templates and farther related active templates. This may be done by using either (a) both an active template for the overall fold (TM1 to TM7) and an inactive template for the most stable part (TM1 to TM4) or (b) inactive templates for TM1 to TM4 and active templates for TM5 to TM7. In this latter case, however, it will be necessary to include the active template TM3 to correctly adjust the orientation of the active and inactive templates. The precise determination of the active and inactive template regions to be used requires careful visual inspection.
2. *Modeling loops*: This is a difficult step in a modeling procedure because loops may be highly variable in sequences and structures, and they are frequently missing in the template structure (s). MODELLER proposes two functions (`loopmodel` and `dope_loopmodel`) to automatically generate/refine loops

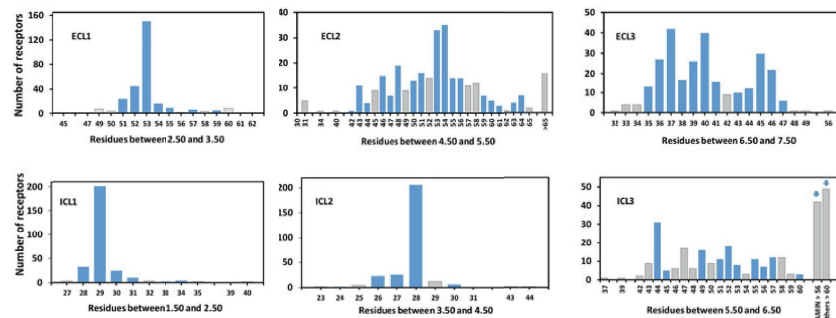


Fig. 4 Statistical analysis of the loop lengths in human GPCRs. The length is measured as the number of residues between the anchor residues $n.50$ present in each helix n . Blue bars indicate that at least one resolved loop of the indicated length is present in the available GPCR structures. Otherwise, bars are gray. The arrows for the long ICL3 loops indicate the resolution of two loops with 87 residues for the PEP receptor OX1R (6TP3) and 78 residues for the AMIN receptor ADRB2 (6MXT)

[46]. However, the loops obtained with these methods may markedly differ from structures revealed in GPCR structures. Now that a large set of GPCR structures are available, it is highly preferable to search for resolved equivalent loops with the same length and use them as templates. Figure 4 summarizes, for each loop, the lengths that have a structural template. In case of ambiguities, SS predictions based on judicious MSA may help. In addition, we note that insertions in loops have frequently an α -helical structure and may result in a protruding TM helix. A first example is given by the orexin receptor 1 (6TP3) in Fig. 2a, in which the long ECL3 (87 residues between the residues 5.50 and 6.50) is structured as a protruding N-terminal part of TM6. A second example is given by CXCR1, AT1, and CCR5 (Figs. 2 and 3) for which the long ECL3 is structured as a protruding N-terminal part of TM7.

3.8 Preparing Template File(s)

After selecting appropriate templates, prepare your PDB template file(s). MODELLER allows the user to select the first and last position of a *single contiguous* segment of the template to be used in the modeling procedure, but this selection is not possible in case of *discontinuous* segments. In this latter case, due for example to insertion of a fusion protein in the template, there is no alternative to manual editing of the pdb file to avoid long and unmanageable insertions in the alignment file. Breaks in the structure have to be indicated with a “/” symbol in the alignment file. In any case, even when this is not strictly necessary, manual editing of the PDB file(s) to excise the regions of the template that will not be used in the modeling procedure is recommended.

3.9 Aligning Model Sequence with the Template(s) and Preparing .ali File

This is a strategic part of the modeling procedure that has to be done very carefully. Several points have to be kept in mind:

1. Identification of the best templates generally involves the alignment of the target sequence with a set of available template sequences. However, the “best” alignments obtained by automatic alignment programs depend on the parameters used and may not be “optimal.” They do not consider user’s additional information that may improve the alignment and the resulting model(s). Thus, manual corrections of the alignment may be necessary.
2. In MODELLER, the alignment file creates *spatial restraints* (see Note 1). Thus, it may be useful to “de-align” residues to remove spatial restraints, increase flexibility and let MODELLER deal with stereochemical restraints only. The easiest way to remove structural restraints is adding gaps in the .ali file.
3. The template sequences in the .ali and .pdb files must *exactly* match. Thus, unresolved parts in the .pdb file must be removed from the sequence in the .ali file.
4. The .ali file has a special format. For each sequence, the first line gives the sequence name in the pir format (>P1; name). The third line gives the sequence and ends by the terminating character “*”. Each chain break must be indicated by a single “/”.
5. The second line gives information on the nature of the sequence and the region to be used. There must be 10 fields, separated by 9 colons (“:”). The first field indicates if the sequence is a template (structure, structureX, structureN..) or the target (sequence), the second field indicates the sequence name, the fields 7–10 are optional, but the colon characters “:” are mandatory. The fields 3–6 indicate first position and chain, last position and chain. If the template is a single contiguous segment, simply specify here the beginning and ending residues that will be used for modeling. Examples can be found in the MODELLER tutorial. However, when templates contain non-contiguous segments, the easiest way to avoid problems is to edit the templates in order to remove any atom or heteroatom that will not be used in the modeling procedure and then to use the “.” character between the colons as in “structure: name: : : : : : : : : : : .” This indicates that all the residues of the template .pdb file have to be read.

3.10 Adding or Suppressing Restraints

Two commands in MODELLER allow either adding restraints (`restraints.add()`) or removing restraints (`restraints.unpick()`). Different types of restraints can be added in a MODELLER script. However, for GPCR modeling, most useful ones are:

1. Secondary structure restraints. They allow extension of a TM helix, modeling of helix 8 which is often not resolved in templates, structuration of a missing loop as an α -helix or forcing the β -strand structure of ECL2.
2. Dihedral constraints to reorient side chains or favor interactions.
3. Distance restraints to maintain or create interactions.

Disulfide bonds not present in the template(s) may be added by the `special_patches()` command.

Suppressing restraints is also possible, in particular for structural restraints built from the alignment. This can be done with the `unpick(*atom_ids)` command. Alternatively, it can be obtained by de-aligning the alignment with the introduction of gaps. For example, before adding dihedral restraints on a sidechain, the residue can be de-aligned to avoid conflicts between restraints. It is worth to note that these commands work on the atoms or residues of the *target*, so that their correct identifiers in the target (and not in the template) have to be provided.

3.11 Model Building

This is the easiest step in the procedure. Select the number of models and the refinement procedure of models, and, optionally of loops (not recommended in GPCR modeling except for short loops or loop regions). Among the five procedures proposed (“none,” “very.fast,” “fast,” “slow,” “very.slow”), the “fast” and “slow” options give optimal results. From 20 to 50 models usually ensure a good representability of the available conformational space. A good compromise is to initiate the modeling procedure with 20 models/fast refinement option and to terminate with 50 models/slow refinement option, after optimization of restraints.

3.12 Evaluating Models

Once a set of models has been generated, there are different ways to assess them. The DOPE [47] score provided by MODELLER allows a comparison of model and template profiles and the visualization of putative problematic regions in the alignment. However, for GPCRs, when several models are generated from the same alignment, the DOPE scores are very similar, and the MODELLER molpdf scores [19] that indicate the violations of the restraints are more discriminative. Several steps have to be carried out to evaluate models:

1. Check the log file from the modeling run for runtime errors and restraint violations. The global molpdf scores are indicated at the end of the log file, which allows determining the “best” models that will be used for further analysis.
2. Visually inspect the models to ensure that no coarse mistake has been done in the alignment file or in the script. Superpose all the models to visualize the conformational space available, especially for loops.

3. Now focus on the “best” models (typically 5 out of 20 or 50 generated models). The sums of the violations for each residue are indicated at the B factor position in the .pdb files of the models (*B999*.pdb). Visualize them on the model structures. For regions with higher violations, inspect the *V999* files that give the type of restraints that is violated. If necessary, modify the alignment or the restraints accordingly.
4. Additional methods can be used to assess the quality of the model, for example, PROCHECK [34] that verifies the stereochemistry of the model.
5. In the selection of the “optimal” model used for further computational studies, be very careful to the orientation of key residues or motifs such as W4.50, W6.48, or DRY in TMH3. Be also very careful to the positions of N- and C-termini and to the structures of the loops.

3.13 Return to Our Test Case

Here we analyze the molecular models of the human B1 receptor, with bound sodium ion, obtained from different templates. The aim is highlighting MODELLER practice to help future users in template selection. In Fig. 3, we have shown four representative structures of homologs: AT1 in the inactive state (closest homolog), OPRD (resolved ICL3 and closest homolog with bound sodium ion), CCR5 (homolog with resolved ECL2), and AT1 in an active state (resolved ECL2). Among these putative templates, we removed CCR5 because of the unusual outward positioning of TM6 on the extracellular side.

First, we used inactive AT1 (4YAY) and OPRD (4N6H) as templates. Since, in OPRD, there is no disulfide bond linking the N-terminus and ECL3, we aligned the sequences of the templates with the target, except for the N-terminus and ECL3. Thus, both templates are used for ICL1, ICL2, ECL1, ECL2, and H8. Nevertheless, in the resulting models of B1 (Fig. 5a), the ICL1, ICL2 and ECL1 loops of the target match those of the AT1 template only (closest template). The ICL3 and ECL3 loops match their unique templates, respectively, OPRD and AT1. The models differ strongly in the modeling of the ECL2 hairpin, since there is no template for this segment. Interestingly, concerning H8, MODELLER does not privilege one template over the other one, and the resulting models have either the tilted orientation observed in AT1 (4YAY) or the horizontal orientation observed in OPRD (4N6H), but not an average orientation.

In the second procedure, we aimed at improving the modeling of the ECL2 hairpin (Fig. 5b). As the length between TMH4 and the ECL2 cysteine is identical in B1 and AT1, we selected a fragment of active AT1 (6OS0) in which ECL2 is resolved. The fragment includes TM3 and TM4 for proper positioning, ICL2 (in helical conformation) and ECL2 up to the Cys residue. With

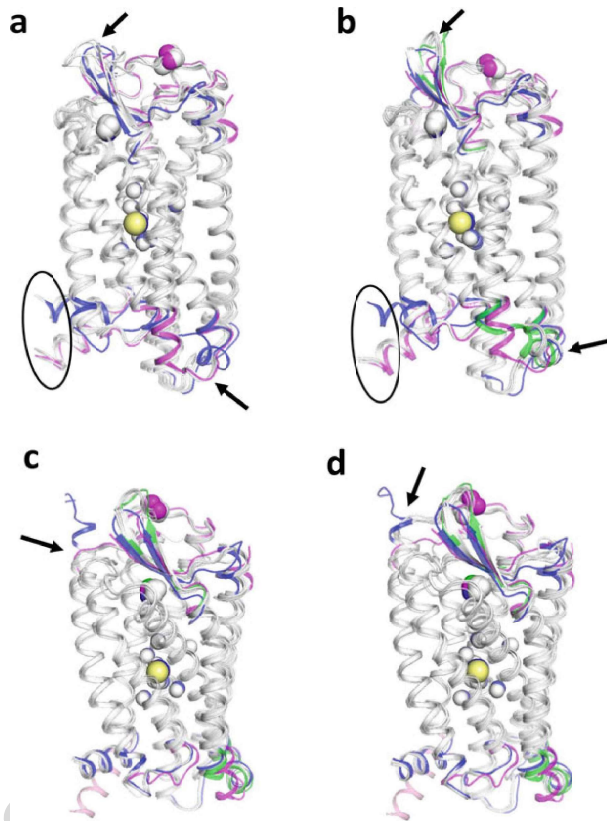


Fig. 5 Influence of templates on the resulting B1 models. In each panel, the top five models (out of 20) obtained with MODELER are superposed with the templates used in modeling procedure. For clarity purpose, all the structures are shown as white ribbon, except the regions of interest in the templates. In (a), B1 is modeled from OPRD (blue) and inactive AT1 (magenta). In (b), B1 is modeled from OPRD (blue), inactive AT1 (magenta), and a fragment from active AT1 encompassing TM3, ICL2, TM4, and ECL2 up to the Cys residue (green). In (c), B1 is modeled as in (b) except that the C-terminus of AT1, in light pink, has been “de-aligned.” In (d), B1 has been modeled as in (c), except for the N-terminus. In (a–c), the N-terminus has been modeled from the AT1 template only. In (d), the OPRD template has been included, resulting in an additional helical turn at the N-terminus of TM1

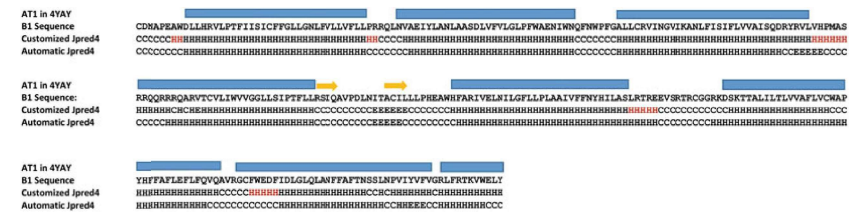


Fig. 6 Secondary structure predictions for B1 using JPred4. The automatic Jpred4 prediction is based on automatic BLAST search starting from the B1 sequence. The customized prediction is based on the MSA of 52 human receptors with the two disulfide bonds in the extracellular domain as observed in B1. The SS drawing corresponds to the experimental SS of AT1 in the 4YAY structure

the addition of this fragment, there are three templates for ICL2. Now, MODELLER privileges the helical structure of ICL2 and the tilted orientation of H8 in the B1 models.

These different models rise the issues of the orientation of H8 and of the structure of ICL2. Concerning H8, the influence of H8 modeling on subsequent MD simulations is detailed in **Note 4**, with the example of AT1, and we recommend to privilege the horizontal orientation of H8. Concerning ICL2, we note that, in both active AT1 and inactive CCR5, ICL2 has a helical structure. To help resolve the uncertainty between the helical and coil conformations of ICL2, we carried out secondary structure prediction using JPred4 [33]. Figure 6 displays SS predictions for B1 using either automatic BLAST search or the MSA of the 52 human sequences that share two properties. First, they belong to the evolutionary related CHEM and PUR subfamilies [24] and, second, they possess the double disulfide bonds in the extracellular domain. The former approach is not informative. By contrast, the latter approach with a customized MSA predicts a helical structure for ICL2, supporting the helical conformation.

In the third procedure, the C-terminus of inactive AT1 was “de-aligned” to exclude it from structural restraints in MODELLER. In this case, the resulting models have a helical ICL2 and a horizontal H8 (Fig. 5c). Finally, in the fourth procedure, we were concerned with the modeling of the N-terminus from the cysteine (first residue in our models) to TMH1. This segment in B1 is one residue longer than AT1 or CCR5. In the first to third procedures, it was modeled with the insertion of one residue compared to the AT1 template. In this fourth procedure, since SS predictions suggest an N-terminal extension of TM1, we checked whether the length of this segment in B1 is compatible with a helical extension. Thus, we also aligned the N-terminus of OPRD in the alignment file. In this latter case, the B1 models do have an additional helical turn at the extracellular side of TM1 without violations of stereochemical restraints (Fig. 5d).

3.14 Concluding Remarks

In this chapter, we have shown how the choice of the template(s) is determinant for the resulting target models and how MODELLER deals with multiple templates. Rather than an average conformation, MODELLER estimates the probability of each conformation and may propose several clearly different models, as exemplified by the H8 orientation in B1 modeling. We emphasize the importance of generating and comparing several models to estimate the conformational space compatible with the restraints provided by the template(s) and the alignment.

We have also shown that carefully selected receptor fragments can greatly improve modeling. Template-based combination of fragments is a powerful approach for molecular modeling of GPCRs that has been implemented in several websites for automatic GPCR modeling, such as GPCR-I-TASSER [48] and GPCR-SFFE [49]. The automatic approaches are very efficient for the modeling of the transmembrane domain but do not provide to the user the possibility to use target-specific information to guide modeling, especially for loops, now that a large variety of templates is available (Fig. 4). Understanding the details of the homology modeling procedure by MODELLER will help the user to make rational choices that will allow improving the quality of customized GPCR models.

4 Notes

1. Modeling by satisfaction of spatial restraints.

The modeling program MODELLER works by satisfying restraints on the structure of the target sequence [19, 20]. The program is designed to take into account:

- (a) *Spatial homology-derived restraints*: Distance and dihedral angle restraints on the target sequence are derived from the template structure and the sequence alignment between target and template. Distance restraints are obtained by assuming that the corresponding distances between *aligned residues in the alignment file* are similar. Dihedral angle restraints are also derived from *aligned residues* in the alignment file. The forms of the restraints are based on empirical knowledge of structural evolution in homologous protein families.
- (b) *Stereochemical restraints*: Additional restraints based on the CHARMM22 force field are added to enforce proper stereochemistry [50].
- (c) *User added restraints*: Additional structural or functional information on the target that cannot be derived from the templates, such as a suspected disulfide, cross-linking

restraints, site directed-mutagenesis results, and predicted secondary structure, can be included as additional restraints. Most commonly additional restraints for GPCRs are distance restraints, dihedral angle restraints, and secondary structure restraints (α -helix restraints for helix termini and β -strand/sheet restraints for extracellular loop 2). Information can also be obtained from careful examination of sequence alignment of the target orthologs which may provide general knowledge about receptor specificities. Adding restraints can be easily done with the `special_restraints()` and `special_patches()` functions of MODELLER and may markedly improve the quality of the target structure.

- (d) *User removed restraints*: MODELLER can unselect all the restraints on specified atoms with the `restraints.unpick()` command. It is also very easy to remove alignment-based spatial restraints by modifying the alignment with the addition of gaps.

Finally, all the restraints are combined in an objective function that is optimized in Cartesian space by using the variable target function method [51] to obtain the model. Then the model is refined by using conjugate gradients and simulated annealing [19]. For a receptor with about 350 residues and 2500 atoms, MODELLER builds and optimizes about 50,000 restraints. Several slightly different models can be calculated by varying the initial structure with a random shift in the position of each atom. This allows gaining information on regions with putative structural flexibility or variability.

2. GPCR classification.

GPCRs are present in fungi, amoeba, animals, and possibly plants [23, 52]. They have known a stupendous evolutionary success in animals where they highly diversified. Species-specific subfamilies make difficult a general classification among the animal reign but, in vertebrates, the GRAFS classification into five families has been widely adopted [2, 52]. Among these five families, the rhodopsin-like or class A is the largest one with 700 out of the 800 total human receptors. These 700 receptors include 400 olfactory and 300 nonolfactory receptors. Finally, the 300 human nonolfactory receptors can be further classified into a dozen of evolutionary based subfamilies, to which can be added the UC class of "unclassified" receptors. The nomenclature for these subfamilies is as follows: PUR (purinergic receptors), CHEM (chemotaxic receptors, including vasoactive peptide receptors), SO (somatostatin and opioid receptors), PEP (peptide receptors), MLT (melatonin receptors), OPN (opsins), PTG (prostaglandin receptors), MEC (melanocortin,

EDG, and cannabinoid receptors), AD (adenosine receptors), AMIN (amine receptors), MRG (MAS-related receptors), and LGR (leucine-rich repeat receptors). However, the evolutionary relationships between these subfamilies are not obvious. The classification into four α , β , γ , and δ groups [2] has been widely adopted in the literature but is not consistent with the order of subfamily apparition during evolution [24, 52]. Based upon a multidimensional scaling analysis, we proposed a model of radial evolution of GPCRs from ancestral peptide receptors with three main evolutionary pathways [25].

The model of radial evolution provides a framework to rationalize the variable proline patterns in TM2 and TM5 observed in class A GPCRs and is consistent with the order of subfamily apparition. Peptide (PEP) receptors are among the most ancient GPCRs. Most PEP receptors are characterized by proline residues at position 2.59 and 5.50 and bulges in TM2 and TM5, as observed for the orexin receptor 1 [53] shown in Fig. 2a. The first evolutionary pathway is related to the deletion of one residue in TM2 and led the SO, CHEM, and PUR receptors. This deletion occurred in an ancestor of the SO receptors that led to the CHEM and PUR receptors by divergence [24]. The SO, CHEM, and PUR subfamilies are characterized by the P2.58 and P5.50 proline patterns, corresponding to a kink in TM2 and a bulge in TM5, respectively, as observed for CXCR1 [54] (Fig. 2b). The second pathway is related to the divergence of amine (AMIN) and adenosine (AD) receptors, characterized by the same sequence and structural patterns as PEP receptors. The third pathway corresponds to independent evolution of several subfamilies with mutations of the proline residues in TM2 and TM5, which are often correlated. These mutations can lead to straight TM2 and TM5, as observed in the cannabinoid receptor 1 [55] from the MEC subfamily (Fig. 2c). Prostaglandin receptors (PTG) present an example of sequence variability for TM2 (P2.59 or no proline) with a conserved bulge observed in PE2R3 (P2.59), PD2R (P2.59) and TA2R (noP), whereas the absence of proline in TM5 led to a bulged TM5 in PD2R and to straight helices in PE2R3 and TA2R [56–58]. The MRG and LGR subfamilies have no proline residues in TM2 and TM5 and no structurally resolved member. Thus, no a priori hypothesis can be done on the structure of these helices. Combination of cellular biology experiments, extensive sequence analysis, and molecular dynamics simulations were necessary to propose a bulged TM2 and a straight TM5 in TSHR, a LGR receptor [38].

In addition to the proline pattern in TM2 and TM5, a second sequence pattern presents interesting features in class A GPCRs. It concerns the disulfide bond(s) stabilizing the receptor extracellular domain. A first disulfide bond links the N-terminal part of TM3 to the extracellular loop 2 (ECL2) and is present in most subfamilies, except MEC and MRG. A second disulfide bond links the N-terminus to an extracellular extension of TM7 in some purinergic and chemotactic receptors. Alternative disulfide bonds have been observed in MEC receptors, whereas an additional disulfide bond is frequently found in ECL3 of amine receptors. Thus, each subfamily presents unique structural features which have to be considered for modeling.

It is worth noting that GPCRDB [59] uses a classification based on the chemical nature of the ligands because it is focused on molecular docking and drug design. However, this may be confusing for homology modeling because similar ligands may bind receptors from different subfamilies. For example, the lysophosphatidic acid receptors LPAR1-3 and LPAR4-6 are members of the MEC and of the PUR subfamilies, respectively, with no proline in TM2 and TM5 for LPAR1-3 and two proline residues at P2.58 and P5.50 for LPAR4-6.

3. Mining GPCRs in UniProt.

In Uniprot, class A GPCRs can be easily mined by searching entries with the correct family identifiers. They are identified as PF00001 (7tm_1) in Pfam, IPR000276 (GPCR_Rhodpsn) or IPR017452 (GPCR_Rhodpsn_7TM) in InterPro, PS00237 (G_PROTEIN_RECEP_F1_1) or PS50262 (G_PROTEIN_RECEP_F1_2) in PROSITE. In InterPro, the IPR017452 identifier is broader than the IPR000276 identifier and includes a number of taste and vomeronasal receptors. In PROSITE, the PS50262 identifier, based on sequence profiles, is of higher quality than the PS00237 identifier, based on motifs. The three identifiers PF00001, IPR000276, and PS50262 are equivalent. Olfactory receptors can be identified as IPR000725 (Olfact_rcpt).

When the user wishes to go deeper in the classification tree, InterPro provides several useful levels of classification. For example, the angiotensin II receptors that include type 1 and type 2 have the reference IPR000248 (ATII_rcpt), type 2 has the reference IPR000147 (ATII_AT2_rcpt), allowing hierarchical selection. However, the intermediary level of classification corresponding to the 12 subfamilies is usually considered neither in UniProt nor in GPCRDB, based on the nature of the ligand (*see Note 2*). Lists of receptors belonging to the 12 subfamilies that make consensus can be found in the literature [2, 24].

4. Positioning of helix 8.

The user should never avoid critical assessment of the template(s). For example, in the structure of inactive AT1 shown in Fig. 3 (4YAY), the orientation of H8 is tilted. Such tilted orientation has been observed in other GPCR structures, e.g., for CCR5 (4MBS). However, it has not been observed in other structures of AT1, such as the recently resolved active structures of AT1 in complex with angiotensin II (6OS0) or angiotensin derivatives (6OD1, 6OS1, 6OS2) [45]. This suggests that the tilted orientation observed in the inactive structure of AT1 might result from artifacts, due to truncation of the C-terminus and experimental conditions.

To answer this question, we carried out molecular dynamics simulations of AT1 (Fig. 7). Starting from the 4YAY structure of AT1, we built two models of the receptor, the first one with H8 positioned as in the 4YAY structure and the second one with H8 positioned as observed in the OPRD 4N6H structure. The models were inserted into a POPC bilayer using the charmm-gui interface (www.charmm-gui.org) and then underwent a short equilibration step (1 ns procedure with progressive release of structural constraints) followed by 280 ns of production run with NAMD [60]. When H8 was modeled in an orientation parallel to the membrane bilayer, it remained stable in this orientation during the 280 ns of the simulations. When the starting structure had a tilted H8 orientation, this one was not stable. H8 underwent a large seesaw motion that induced strong perturbations not only of H8 but also of the intracellular sides of TM6, TM7, and TM1 (Fig. 7c). These results highlight the importance of the orientation of H8 on the stability of MD simulations and strongly suggests that the canonical horizontal orientation of H8 should be privileged in molecular modeling.

Acknowledgments

This study was supported by institutional grants from INSERM, CNRS and University of Angers. This work was granted access to HPC resources of IDRIS (GENCI grant 100567 to MC). MC is supported by CNRS. AT is supported by a fellowship from the University of Carthage (Tunisia). RB is supported by a fellowship from the University of Angers (France).

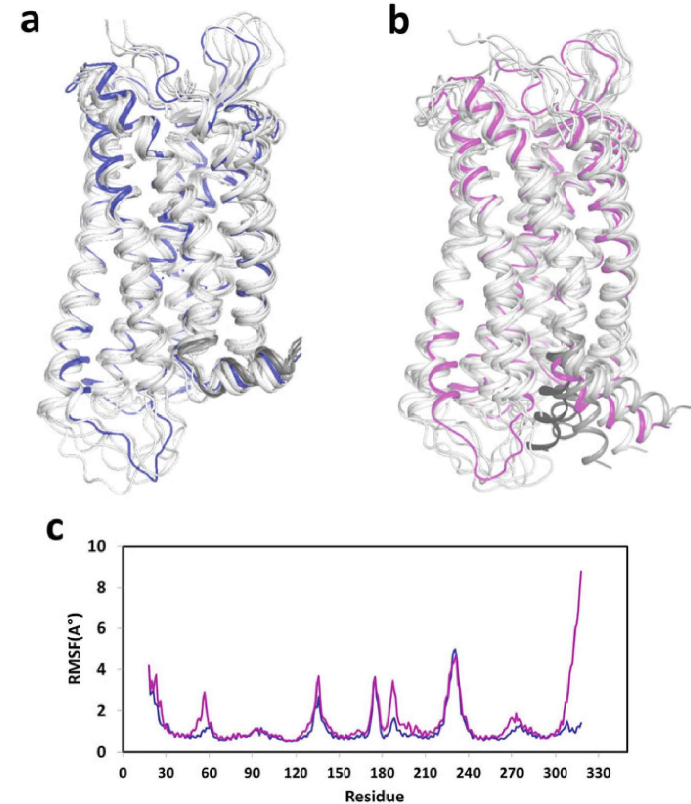


Fig. 7 Consequences of the orientation of H8 on MD simulations of the AT1 receptor. MD simulations were run for 280 ns with NAMD, starting from an AT1 model with horizontal H8 (a) and tilted H8 (b). The starting conformations are blue in (a) and magenta in (b). Representative snapshots from the beginning to the end of the production run indicate the conformational space sampled by H8 during the simulations. The snapshots are shown as white ribbons, with increasing graying of the C-terminus from white to dark gray with the simulation time. In (c), comparison of RMSF of AT1 when the trajectory was started with the horizontal (blue) and tilted orientation (magenta) of H8

References

1. Bockaert J, Pin JP (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J* 18(7):1723–1729. <https://doi.org/10.1093/emboj/18.7.1723>
2. Fredriksson R, Lagerstrom MC, Lundin LG, Schiöth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* 63(6):1256–1272. <https://doi.org/10.1124/mol.63.6.1256>
3. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5(12):993–996. <https://doi.org/10.1038/nrd2199>

- 816 4. Scalfon SC, Chi L, Ebersole BJ et al (1995)
817 Related contribution of specific helix 2 and
818 7 residues to conformational activation of the
819 serotonin 5-HT_{2A} receptor. *J Biol Chem* 270
820 (28):16683–16688. <https://doi.org/10.1074/jbc.270.28.16683>
- 822 5. Palczewski K, Kumasaka T, Hori T et al (2000)
823 Crystal structure of rhodopsin: a G protein-coupled
824 receptor. *Science* 289 (5480):739–745. <https://doi.org/10.1126/science.289.5480.739>
- 827 6. Cherezov V, Rosenbaum DM, Hanson MA
828 et al (2007) High-resolution crystal structure
829 of an engineered human beta₂-adrenergic G
830 protein-coupled receptor. *Science* 318
831 (5854):1258–1265. <https://doi.org/10.1126/science.1150577>
- 833 7. Xiang J, Chun E, Liu C et al (2016) Successful
834 strategies to determine high-resolution structures
835 of GPCRs. *Trends Pharmacol Sci* 37
836 (12):1055–1069. <https://doi.org/10.1016/j.tips.2016.09.009>
- 838 8. Garcia-Nafria J, Tate CG (2020) Cryo-electron
839 microscopy: moving beyond x-ray crystal structures
840 for drug receptors and drug development. *Annu Rev Pharmacol Toxicol*
841 60:51–71. <https://doi.org/10.1146/annurev-pharmtox-010919-023545>
- 844 9. Katritch V, Cherezov V, Stevens RC (2012)
845 Diversity and modularity of G protein-coupled
846 receptor structures. *Trends Pharmacol Sci* 33
847 (1):17–27. <https://doi.org/10.1016/j.tips.2011.09.003>
- 849 10. Munk C, Mutt E, Isberg V et al (2019) An
850 online resource for GPCR structure determination
851 and analysis. *Nat Methods* 16 (2):151–162. <https://doi.org/10.1038/s41592-018-0302-x>
- 854 11. Rasmussen SG, DeVree BT, Zou Y et al (2011)
855 Crystal structure of the beta₂ adrenergic
856 receptor-Gs protein complex. *Nature* 477
857 (7366):549–555. <https://doi.org/10.1038/nature10361>
- 859 12. Katritch V, Fenalti G, Abola EE et al (2014)
860 Allosteric sodium in class A GPCR signaling.
861 *Trends Biochem Sci* 39(5):233–244. <https://doi.org/10.1016/j.tibs.2014.03.002>
- 863 13. Margiotta E, Deganutti G, Moro S (2018) Could
864 the presence of sodium ion influence the accuracy
865 and precision of the ligand-posing in the human
866 A_{2A} adenosine receptor orthosteric binding site
867 using a molecular docking approach? Insights
868 from Dockbench. *J Comput Aided Mol Des* 32
869 (12):1337–1346. <https://doi.org/10.1007/s10822-018-0174-2>
- 871 14. Nygaard R, Frimurer TM, Holst B et al (2009)
872 Ligand binding and micro-switches in 7TM
receptor structures. *Trends Pharmacol Sci* 30
873 (5):249–259. <https://doi.org/10.1016/j.tips.2009.02.006>
- 876 15. Congreve M, Dias JM, Marshall FH (2014)
877 Structure-based drug design for G
878 protein-coupled receptors. *Prog Med Chem*
879 53:1–63. <https://doi.org/10.1016/B978-0-444-63380-4.00001-9>
- 881 16. Shonberg J, Lopez L, Scammells PJ et al
882 (2014) Biased agonism at G protein-coupled
883 receptors: the promise and the challenges—a
884 medicinal chemistry perspective. *Med Res Rev*
885 34(6):1286–1330. <https://doi.org/10.1002/med.21318>
- 887 17. Diaz C, Angeloz-Nicoud P, Pihan E (2018)
888 Modeling and deorphanization of orphan
889 GPCRs. *Methods Mol Biol* 1705:413–429.
890 https://doi.org/10.1007/978-1-4939-7465-8_21
- 891 18. Stockert JA, Devi LA (2015) Advancements in
892 therapeutically targeting orphan GPCRs. *Front
893 Pharmacol* 6:100. <https://doi.org/10.3389/fphar.2015.00100>
- 896 19. Sali A, Blundell TL (1993) Comparative protein
897 modelling by satisfaction of spatial
898 restraints. *J Mol Biol* 234(3):779–815.
899 <https://doi.org/10.1006/jmbi.1993.1626>
- 900 20. Webb B, Sali A (2016) Comparative protein
901 structure modeling using MODELLER. *Curr
902 Protoc Protein Sci* 86:2.9.1–2.9.37. <https://doi.org/10.1002/cpps.20>
- 904 21. Devos D, Valencia A (2000) Practical limits of
905 function prediction. *Proteins* 41(1):98–107.
906 [https://doi.org/10.1002/1097-0134\(20001001\)41:1<98::AID-PROT120>3.0.CO;2-S](https://doi.org/10.1002/1097-0134(20001001)41:1<98::AID-PROT120>3.0.CO;2-S)
- 909 22. Sanchez R, Sali A (1997) Advances in comparative
910 protein-structure modelling. *Curr Opin Struct Biol* 7(2):206–214. [https://doi.org/10.1016/s0959-440x\(97\)80027-9](https://doi.org/10.1016/s0959-440x(97)80027-9)
- 912 23. Chabbert M, Castel H, Pele J et al (2012)
913 Evolution of class A G-protein-coupled receptors:
914 implications for molecular modeling. *Curr Med Chem* 19(8):1110–1118. <https://doi.org/10.2174/092986712799320600>
- 917 24. Deville J, Rey J, Chabbert M (2009) An indel
918 in transmembrane helix 2 helps to trace the
919 molecular evolution of class A G-protein-coupled
920 receptors. *J Mol Evol* 68(5):475–489.
921 <https://doi.org/10.1007/s00239-009-9214-9>
- 924 25. Pele J, Abdi H, Moreau M et al (2011) Multi-
925 dimensional scaling reveals the main evolutionary
926 pathways of class A G-protein-coupled
927 receptors. *PLoS One* 6(4):e19094. <https://doi.org/10.1371/journal.pone.0019094>
- 928 26. Taddese B, Deniaud M, Garnier A et al (2018)
929 Evolution of chemokine receptors is driven by
930 mutations in the sodium binding site. *PLoS
931 Comput Biol* 14(6):e1006209. <https://doi.org/10.1371/journal.pcbi.1006209>
- 933 27. Webb B, Sali A (2016) Comparative protein
934 structure modeling using MODELLER. *Curr
935 Protoc Bioinformatics* 54:5.6.1–5.6.37.
936 <https://doi.org/10.1002/cpbi.3>
- 937 28. Pettersen EF, Goddard TD, Huang CC et al
938 (2004) UCSF chimera—a visualization system
939 for exploratory research and analysis. *J Comput
940 Chem* 25(13):1605–1612. <https://doi.org/10.1002/jcc.20084>
- 943 29. Larkin MA, Blackshields G, Brown NP et al
944 (2007) Clustal W and Clustal X version 2.0.
945 *Bioinformatics* 23(21):2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- 946 30. Notredame C, Higgins DG, Heringa J (2000)
947 T-Coffee: a novel method for fast and accurate
948 multiple sequence alignment. *J Mol Biol* 302
949 (1):205–217. <https://doi.org/10.1006/jmbi.2000.4042>
- 951 31. Edgar RC (2004) MUSCLE: multiple
952 sequence alignment with high accuracy and
953 high throughput. *Nucleic Acids Res* 32
954 (5):1792–1797. <https://doi.org/10.1093/nar/gkh340>
- 956 32. Nicholas KB, Nicholas HB Jr, Deerfield DWI
957 (1999) GeneDoc: analysis and visualization of
958 genetic variation. *EMBNEWNEWS* 4:14
- 959 33. Drozdetskiy A, Cole C, Procter J, Barton GJ
960 (2015) JPred4: a protein secondary structure
961 prediction server. *Nucleic Acids Res* 43(W1):
962 W389–W394. <https://doi.org/10.1093/nar/gkv332>
- 964 34. Laskowski RA, Rullmann JA, MacArthur MW
965 et al (1996) AQUA and PROCHECK-NMR:
966 programs for checking the quality of protein
967 structures solved by NMR. *J Biomol NMR* 8
968 (4):477–486. <https://doi.org/10.1007/bf00228148>
- 969 35. Holm L, Sander C (1998) Removing near-
970 neighbour redundancy from large protein
971 sequence collections. *Bioinformatics* 14
972 (5):423–429. <https://doi.org/10.1093/bioinformatics/14.5.423>
- 973 36. Isberg V, Mordalski S, Munk C et al (2016)
974 GPCRdb: an information system for G protein-
975 coupled receptors. *Nucleic Acids Res* 44(D1):
976 D356–D364. <https://doi.org/10.1093/nar/gkv1178>
- 977 37. Chan WK, Zhang H, Yang J et al (2015)
978 GLASS: a comprehensive database for experi-
979 mentally validated GPCR-ligand associations.
980 *Bioinformatics* 31(18):3035–3042. <https://doi.org/10.1093/bioinformatics/btv302>
- 981 38. Chantreau V, Taddese B, Munier M et al
982 (2015) Molecular insights into the transmem-
983 brane domain of the thyrotropin receptor. *PLoS
984 One* 10(11):e0142250. <https://doi.org/10.1371/journal.pone.0142250>
- 989 39. Castleman PN, Sears CK, Cole JA et al (2019)
990 GPCR homology model template selection
991 benchmarking: global versus local similarity
992 measures. *J Mol Graph Model* 86:235–246.
993 <https://doi.org/10.1016/j.jmgm.2018.10.016>
- 994 40. Costanzi S, Skorski M, Deplano A et al (2016)
995 Homology modeling of a class A GPCR in the
996 inactive conformation: a quantitative analysis of
997 the correlation between model/template
998 sequence identity and model accuracy. *J Mol
999 Graph Model* 70:140–152. <https://doi.org/10.1016/j.jmgm.2016.10.004>
- 1000 41. Srinivasan N, Blundell TL (1993) An evalua-
1001 tion of the performance of an automated pro-
1002 cedure for comparative modelling of protein
1003 tertiary structure. *Protein Eng* 6(5):501–512.
1004 <https://doi.org/10.1093/protein/6.5.501>
- 1005 42. Fenalti G, Giguere PM, Katritch V et al (2014)
1006 Molecular control of delta-opioid receptor sig-
1007 nalling. *Nature* 506(7487):191–196. <https://doi.org/10.1038/nature12944>
- 1008 43. Zhang H, Unal H, Gati C et al (2015) Structure
1009 of the angiotensin receptor revealed by
1010 serial femtosecond crystallography. *Cell* 161
1011 (4):833–844. <https://doi.org/10.1016/j.cell.2015.04.011>
- 1012 44. Tan Q, Zhu Y, Li J et al (2013) Structure of the
1013 CCR5 chemokine receptor-HIV entry inhibitor
1014 maraviroc complex. *Science* 341
1015 (6152):1387–1390. <https://doi.org/10.1126/science.1241475>
- 1016 45. Winkler LM, Skiba MA, McMahon C et al
1017 (2020) Angiotensin and biased analogs induce
1018 structurally distinct active conformations
1019 within a GPCR. *Science* 367(6480):888–892.
1020 <https://doi.org/10.1126/science.aay9813>
- 1021 46. Fiser A, Sali A (2003) Modeller: generation
1022 and refinement of homology-based protein
1023 structure models. *Methods Enzymol*
1024 374:461–491. [https://doi.org/10.1016/S0076-6879\(03\)74020-8](https://doi.org/10.1016/S0076-6879(03)74020-8)
- 1025 47. Shen MY, Sali A (2006) Statistical potential for
1026 assessment and prediction of protein struc-
1027 tures. *Protein Sci* 15(11):2507–2524.
1028 <https://doi.org/10.1110/ps.062416606>
- 1029 48. Zhang J, Yang J, Jang R, Zhang Y (2015)
1030 GPCR-I-TASSER: a hybrid approach to G
1031 protein-coupled receptor structure modeling
1032 and the application to the human genome.
1033 *Structure* 23(8):1538–1549. <https://doi.org/10.1016/j.str.2015.06.007>

- 1043 49. Worth CL, Kreuchwig F, Tiemann JKS et al (2017) GPCR-SSFE 2.0-a fragment-based molecular modeling web tool for class A G-protein coupled receptors. *Nucleic Acids Res* 45(W1):W408–W415. <https://doi.org/10.1093/nar/gkx399>
- 1049 50. MacKerell AD, Bashford D, Bellott M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616. <https://doi.org/10.1021/jp973084f>
- 1054 51. Braun W, Go N (1985) Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J Mol Biol* 186(3):611–626. [https://doi.org/10.1016/0022-2836\(85\)90134-2](https://doi.org/10.1016/0022-2836(85)90134-2)
- 1059 52. Fredriksson R, Schioth HB (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol Pharmacol* 67(5):1414–1425. <https://doi.org/10.1124/mol.104.009001>
- 1064 53. Rappas M, Ali AAE, Bennett KA et al (2020) Comparison of orexin 1 and orexin 2 ligand binding modes using x-ray crystallography and computational analysis. *J Med Chem* 63(4):1528–1543. <https://doi.org/10.1021/acs.jmedchem.9b01787>
- 1070 54. Park SH, Das BB, Casagrande F et al (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491(7426):779–783. <https://doi.org/10.1038/nature11580>
- 1075 55. Hua T, Vemuri K, Pu M et al (2016) Crystal structure of the human cannabinoid receptor CB1. *Cell* 167(3):750–762. e714. <https://doi.org/10.1016/j.cell.2016.10.004>
- 1079 56. Fan H, Chen S, Yuan X et al (2019) Structural basis for ligand recognition of the human thromboxane A2 receptor. *Nat Chem Biol* 15(1):27–33. <https://doi.org/10.1038/s41589-018-0170-9>
- 1084 57. Wang L, Yao D, Deepak R et al (2018) Structures of the human PGD2 receptor CRTH2 reveal novel mechanisms for ligand recognition. *Mol Cell* 72(1):48–59.e4. <https://doi.org/10.1016/j.molcel.2018.08.009>
- 1089 58. Morimoto K, Suno R, Hotta Y et al (2019) Crystal structure of the endogenous agonist-bound prostanoid receptor EP3. *Nat Chem Biol* 15(1):8–10. <https://doi.org/10.1038/s41589-018-0171-8>
- 1094 59. Isberg V, Mordalski S, Munk C et al (2017) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* 45(5):2936. <https://doi.org/10.1093/nar/gkw1218>
- 1099 60. Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802. <https://doi.org/10.1002/jcc.20289>

Homology modeling in the twilight zone: Improved accuracy by sequence space analysis

Rym Ben Boubaker, Asma Tiss, Daniel Henrion and Marie Chabbert*

UMR CNRS 6015 – INSERM 1083, Laboratoire MITOVASC, Université d’Angers, Angers,
France

* To whom correspondence should be addressed

i. **Running head** (60 characters): Homology modeling in the twilight zone

ii. **Summary**

The analysis of the relationship between sequence and structure similarities during the evolution of a protein family has revealed a limit of sequence divergence for which structural conservation can be confidently assumed and homology modeling is reliable. Below this limit, the twilight zone corresponds to sequence divergence for which modeling requires specific “threading” methods. Initially, the term “threading” applied to molecular modeling methods that searched the compatibility between a sequence and a fold. Now, this term applies to homology modeling when a putative template can be found only by profile-profile mining methods. In this case, the challenge relies on the identification of a template that shares not only a common ancestor (homology) but also a conserved structure with the query. As both homology and structural conservation are transitive properties, mining of sequence databases followed by multidimensional scaling (MDS) of the query sequence space can reveal intermediary sequences to infer homology and structural conservation between the query and the template. Here, as a case study, we studied the plethodontid receptivity factor isoform 1 (PRF1) from *Plethodon jordani*, a member of a pheromone protein family present only in lungless salamanders and weakly related to cytokines of the IL6 family. A variety of threading methods led to the cytokine CNTF as a template. Sequence mining, followed by phylogenetic and MDS analysis, provides missing links between PRF1 and CNTF and allows reliable homology modeling. In addition, we compare

automated modeling through web servers to a customized model to show how modeling can be improved by user-added information.

iii. **Key Words:**

Molecular modeling, threading, twilight zone, profile-profile mining, cytokine, plethodontid receptivity factor

1 Introduction

Since the resolution of the myoglobin structure in 1958 [1], the number of protein structures deposited in the Protein Data Bank [2] has increased exponentially to reach more than 160 000 structures in 2020. These structures led to a better understanding of protein functions and mechanisms of action. They have paved the way to computational approaches for rational drug design, search of targetable allosteric sites, better understanding of structure-function relationships, and so on. However, in spite of the huge advances in the field, the sequence space increases much more than structural space and computational approaches towards many proteins still rely on molecular modeling.

Presently, based on available structural information and deposited structures, proteins can be classified into four categories: (1) proteins with resolved structures, (2) proteins with closely related structurally resolved homologs that can be straightforwardly modeled by homology, (3) proteins within the twilight zone, for which structural information can be reached in spite of the absence of close structurally resolved homologs, and (4) proteins within the dark zone that lack similarity to any known structure and are inaccessible to homology modeling [3].

These dark proteins present unexpected features, such as short sequence length, low evolutionary reuse, disulfide bonding and few known interactions with other proteins. In the human genome, about 40% of the protein sequences (including full proteins and protein regions) are dark, 40% have a resolved structure or a closely related template and 20% are in the twilight zone [4].

The initial studies from the 1990s on the relationship between sequence and structural evolution remain valid today. In a hallmark paper, Sander and Schneider [5] determined a curve describing the limit of confidence between evolution of the sequence and conservation of the structure. This study was extended by Rost [6] who corroborated the main result: a

length dependent cutoff line separates close homologs with structural conservation from remote homologs with unknown structural similarity. The length of the aligned sequence is a key factor for confidence level in structural conservation. A cutoff of about 20% for a sequence of 200 amino acids or longer separates the safe zone for homology modeling from the twilight zone (Fig. 1). This does not mean that homology modeling is not possible in the twilight zone, nor that modeling errors are not possible in the safe zone, but this points to the additional difficulties that arise in the twilight zone.

Insert Fig. 1 here.

Homology modeling is based on evolution. Proteins do not arise from scratch and can be classified into families. Homologous proteins within a family evolved from a common ancestor and share sequence, structural, and, to a lesser extent, functional similarities [7]. The structural conservation within a family is the keystone of molecular modeling. Using known structures of homologous proteins (the templates) and a multiple sequence alignment between the query and the templates, homology modeling programs such as MODELLER [8] build restraints and optimize the query structure from the template structures. In the safe zone, the sequence identity between the query and putative templates is high enough to infer that these proteins belong to the same family and share a common structure. By contrast, in the twilight zone, the similarities may arise from chance, convergence or common ancestry, which raises several issues:

- (1) Finding templates by straightforward sequence-sequence comparison methods is not possible. This point has led to the development of “threading” methods based either on compatibility of sequence and fold or on profile-based search (see below).
- (2) Additional information may be necessary to find evidence of common ancestry (homology) between template and query and infer the conservation of the structure. Indeed, very low identity rates may correspond to divergent or convergent evolution. In addition,

random alignment leads to 5% sequence identity whereas, in highly divergent families, sequence identities can be as low as 8% (see an example in Fig. 2).

(3) Alignment between query and template(s) may be difficult. Indeed, the cutoff also separates easy from tricky alignments which may alter the quality of the homology modeling. Alignments are greatly improved by the use of multiple sequence alignment methods [9,10] but their accuracy remains challenging at low sequence identities.

Thus, the challenge of molecular modeling in the twilight zone relies on the recognition of correct templates and generation of accurate sequence-template alignments [6]. In this chapter, for clarity purpose, we will use, as a case study, the plethodontid receptivity factor isoform 1 (PRF1) from *Plethodon jordani*. This protein, for which structural and evolutionary data are missing, was discovered in 1999. It is a member of a pheromone protein family present only in lungless salamanders, with weak similarity with cytokines of the IL6 family [11]. We will show how sequence analysis methods, in particular multidimensional scaling (MDS), support the homology and structural conservation between PRF1 and cytokines of the IL6 family, by revealing intermediary sequences. We will also show how modeling of PRF1 can be improved by a variety of techniques.

2 Materials

2.1 Databases

1. UniProt (<https://www.uniprot.org/>) is a comprehensive resource of protein sequences and functional information [12]. It is composed of the manually curated SwissProt and of the automatically annotated trEMBL repositories. It contains not only protein sequences but also additional information including related 3D structures or models, and identifiers of the protein family in different family databases such as Pfam [13] and InterPro [14].

2. The Protein Data Bank (PDB, accessible at <https://www.rcsb.org/>) is the repository of biological macromolecular structures [2,15].
3. SCOP (Structural classification of proteins) [16] is a repository of protein folds, based on an initial classification into five structural classes: all alpha, all beta, alpha/beta, alpha+beta and small proteins.

2.2 Sequence analysis

1. NRDB90.pl [17] is a perl script aimed at clusterizing sequences based on sequence identity to build non-redundant sets. It can be downloaded from <ftp://biodisk.org>.
2. Different programs such as CLUSTAL [18], MUSCLE [19,20] and T-COFFEE [21] can be used to perform multiple sequence alignments (MSA).
3. The EXPRESSO program from the T-COFFEE suite [22] provides a multiple sequence alignment based on structural alignments, which may be a useful initial step for aligning proteins with low identities (<http://tcoffee.crg.cat/apps/tcoffee/do:expresso>).
4. MSA can be manually edited using Genedoc [23], a program aimed at editing and analyzing MSA through a graphical interface and available at <https://genedoc.software.informer.com/>. Subsequent phylogenetic analysis can be performed by the user-friendly MEGA software (Molecular Evolutionary Genetics Analysis) [24] available at <https://www.megasoftware.net/>.
5. The R package `Bios2mde` [25] is aimed at analyzing MSAs by multidimensional scaling and provides tools for user-friendly visualization of the sequence space (see Note 1).

2.3 Template mining

A variety of programs can be used to mine homologs in sequence databases. The choice of the program depends on the putative sequence identities between the query and the hits. Sequence - sequence comparison programs are adequate in the safe zone whereas sophisticated profile -

profile searches are adapted to the twilight zone. Initially, “threading” referred to the search of a template by analyzing the compatibility of a sequence with a protein fold. Presently, “threading” refers to any method searching a template by sequence-profile or profile-profile comparison.

Here is a non-exhaustive list of sequence database mining programs:

1. BLAST (Basic Local Alignment Search Tool) [26], based on local sequence similarity, allows fast sequence-sequence comparison.
2. PSI-BLAST (Position-Specific Interactive BLAST) [27] is based on sequence-profile comparison. It derives a position-specific scoring matrix (PSSM) from the multiple sequence alignment of sequences detected above a given score threshold using protein-protein BLAST.
3. HMMER [28] is a sequence-profile comparison method based on profile hidden Markov models (HMMs) (<https://toolkit.tuebingen.mpg.de/tools/hmmer>).
4. Phyre2 (Protein Homology/analogy Recognition Engine V 2.0) [29] performs its searches by mining a database of profile HMMs, one for each known 3D structure (<http://www.sbg.bio.ic.ac.uk/~phyre2/>).
5. HHPred [30-32] performs HMM-HMM profile searches in sequence databases to find homologs (<https://toolkit.tuebingen.mpg.de/tools/hhpred>).
6. LOMETS (local meta-threading-server) [33] performs template searches using eleven different threading methods (see Note 2). Starting from a query sequence, LOMETS works in three steps: (1) Building a sensitive (or deep) MSA, (2) Threading the deep MSA by individual programs, and (3) Ranking templates with a specific scoring function which takes into account normalized Z-scores and sequence identities. For each method, the normalized Z-scores differentiate good/bad templates (threshold of 1) (<https://zhanglab.ccmb.med.umich.edu/LOMETS/>).

7. SUPERFAMILY finds a protein fold based on a collection of hidden Markov models, which represent structural protein domains at the SCOP superfamily level [34,35]. SUPERFAMILY is a “true” threading program, aimed at finding a protein fold (<https://supfam.org/SUPERFAMILY/>).

2.4 Secondary structure prediction

When searching information for a protein without close structurally resolved homologs, prediction of secondary structure (SS) may yield useful information. Best performances for SS prediction are obtained with programs based on multiple sequence alignment profiles and neural networks, such as PSIPRED [36,37] (<http://bioinf.cs.ucl.ac.uk/psipred/>) and SPIDER3 [38] (<https://sparks-lab.org/server/spider3/>).

2.5 Automated structure prediction

Several web servers can perform automated structural prediction from a query sequence. They are based on threading methods to find templates, and then they use different methods for the subsequent modeling steps. Here, we list three automated 3D prediction servers compared in this Chapter:

1. Phyre2 [29] : After template detection by mining a HMM database, the subsequent molecular modeling step is carried out with MODELLER based on the resulting sequence alignment between query and templates (<http://www.sbg.bio.ic.ac.uk/~phyre2/>).
2. I-TASSER (Iterative Threading ASSEMBly Refinement) [39-41] is a hierarchical approach to protein structure prediction. It first identifies structural templates by comparing the best hits from the “best” ten out of fourteen threading methods with LOMETS, and then it builds full-length atomic models by iterative template-based fragment assembly simulations (<https://zhanglab.cmb.med.umich.edu/I-TASSER/>).

3. ROBETTA [42] is a protein structure prediction service. In the option for Rosetta Comparative Modeling (RosettaCM) [42], four independent methods (see Note 3) are used to detect templates and generate sequence alignments, and then models are built from template hybridization (<https://robetta.bakerlab.org>).

2.6 Customized structure prediction

For users who wish to build their own models, the MODELLER program [8] builds molecular models of the query from the template structure(s) by minimizing structural, stereochemical and user-defined restraints. The structural restraints are based on the structure of the template(s) and the alignment between query and template(s). To customize models in order to match structural and functional requirements, expert information can be introduced by (1) adding user-defined restraints such as distance between two residues or secondary structure elements in the modeling procedure and (2) by combining user-selected templates and template fragments.

2.7 Model validation

With threading methods, the metrics to compare models and templates must be more sensitive to the global fold similarity than to local structural variation. This is not the case of the traditional root-mean-square deviation (RMSD). The TM-score (see Note 4) has been specifically designed to solve this problem [43]. A threshold of 0.5 differentiates proteins with similar fold from proteins with different fold [44]. TM-scores can be calculated after structural alignment with TM-align [45] at the Zhang lab server (<https://zhanglab.cmb.med.umich.edu/TM-align/>).

2.8 Graphical analysis

Graphical analysis of templates and models can be carried out by a variety of molecular visualization programs, such as Chimera [46] or Pymol (<https://pymol.org/>). Note that the low identity rates between template and query sequences in the twilight zone prevent the use of sequence-based structural superposition functions, such as the `align` function in Pymol.

3 Method

3.1 Our case study

As a case study, we chose a small protein, the plethodontid receptivity factor isoform 1 (PRF1) from *Plethodon jordani* (UniProt entry: Q9PUJ2_PLEJO) [11]. This 215 amino acid protein (including a 23 amino acid peptide signal) is a courtship pheromone produced by males to increase female receptivity (see Note 5). When PRF1 was discovered in 1999, it was acknowledged that its sequence was weakly related (around 16% sequence identity) to the ciliary neurotrophic factor (CNTF) and cardiotrophin-1 (CTF1), two cytokines of the interleukin-6 (IL6) family [11]. Since then, two additional cytokines of the IL6 family have been discovered: cardiotrophin-2 (CTF2, absent in humans) [47] and the cardiotrophin-like cytokine factor 1 (CLCF1) [48]. CTF2 and CLCF1 have, respectively, 26% and 20% sequence identity with PRF1.

In mammals, the IL6 family of cytokines includes IL6, interleukin-11 (IL11), leukemia Inhibitory factor (LIF), oncostatin M (ONCM), CNTF, CTF1, CTF2 and CLCF1. Albeit the sequence identities can be as low as 8%, these cytokines share a common four-helix bundle fold with an up-up-down-down topology (Fig. 2). In addition, they signal through the gp130 receptor subunit and share similar binding sites with cognate receptors (see Note 6) [49-52]. Crystal structures have been resolved for five cytokines from the IL6 family: IL6 (1ALU [53], 5FUC [54]), IL11 (4MHL [55]), CNTF (1CNT [56]), LIF (1EMR, 1LKI [57], 2Q7N [58],

and ONCM (1EVS [59]). No crystal or NMR structure has been reported to date for PRF1 or cardiotrophin-like cytokines.

Insert Fig. 2 here.

3.2 Template search by PDB mining

In InterPro, PRF1 is described as belonging to the *4_helix_cytokine-like_core* superfamily ([IPR009079](#)) and to the *PRF/Cardiotrophin-like* family ([IPR010681](#)). Additional information from the PRF1 sequence was searched for using the SUPERFAMILY assignment server [34]. SUPERFAMILY predicts that PRF1 is in the class of *All alpha proteins*, and belongs to the fold/superfamily of *4-helical cytokines* with an E-value of $3 \cdot 10^{-55}$ (see Note 7). It also suggests an “uncertain” classification for the family level as *Long-chain cytokine* with an E-value of only 0.005.

The next step to find a template was the mining of the PDB in search of homologs. Using the mature sequence of PRF1 (residues 24-215) as a query, we performed different searches (Table 1):

1. Straightforward mining of the PDB with BLAST: this search led to no hit
2. Sequence - profile search: A PSI-BLAST search seeded with the PRF1 sequence, followed by selection of hits on most query sequence (>60%), led to CNTF as a hit with an unreliable E-value of 5.9
3. Profile HMM search: Profile search using the HMMER program [28] was carried out on the HHpred server. The search led to two hits, CNTF and LIF, as putative templates with very significant E-values of 10^{-10} or lower
4. Profile HMM - profile HMM search: the HHpred algorithm led to six hits with E-values lower than 0.1: CNTF, LIF, ONCM, IL11, GCSF (granulocyte colony-stimulating factor)

and IL6. Among them, GCSF (E-value of $8 \cdot 10^{-13}$) is a 4-helical cytokine that does not belong to the IL6 family, but shares the same up-up-down-down four-helix bundle fold.

3.3 Template search with LOMETS

Finally, a comparison of 11 threading methods was carried out with the LOMETS server [33] (Table 2). All the methods, but one, classified CNTF or LIF as best hits. Only CEthreader, which is a contact-based method, privileged prolactin (PDB 1RW5). This growth factor shares the four-helix bundle fold of the IL6 cytokines. Most additional hits include cytokines of the IL6 family (ONCM, IL11, IL6) or cytokines/growth factors with same four-helix bundle fold (prolactin, lactogen, IL23). However, several methods also found IL1Ra, the interleukin-1 receptor antagonist (PDB 1ILR) [60] that has a beta barrel fold (see Note 8). This finding serves as a reminder of how cautious users need to be when analyzing threading results.

3.4 Sequence space investigation

As exemplified by IL1Ra in LOMETS results (Table 2), finding a template by threading does not prove that there is homology, i.e. a common ancestor, between the template and the query, nor that the structure is conserved. The identity of 16% between the PRF1 query and the CNTF or LIF templates is positioned in the twilight zone (Fig. 1). However, homology and structural conservation are transitive properties. Investigation of the sequence space of the query analogs may reveal sequences homologous to both query and template(s) in the safe zone and consequently validate the template. Indeed, finding intermediates is an efficient strategy to reduce false positives [6,30].

To investigate the query sequence space, several steps must be carried out:

1. Blast search of the query homologs in sequence databases. Here, using PRF1 as a query in UniProt vertebrate sequences, we obtained 602 hits with E-value lower than 10. Among them, 190 sequences corresponded to salamander receptivity factors and shares with PRF1 sequence identities larger than 60%. Among very significant hits (E-value $< 10^{-10}$), several sequences are known cytokines: human and mouse CLCF1, human and mouse CTF1, mouse CTF2. Interestingly, the sequence of chicken CNTF was identified with an E-value of 5 and a sequence identity of 20%.
2. Building of a non-redundant (NR) set of the hits. This step considerably reduces the number of sequences investigated by suppressing non informative, highly similar sequences. Here, we built a NR set of PRF1 hits (id $< 90\%$), first by selecting sequences with length between 150 and 360 aa, and then, by clustering these sequences using the NRDB90.pl script [17]. The MSA of the NR set was carried with Clustal [18] and manually edited with Genedoc [23]. At this stage, a few additional truncated sequences were removed. This led to a non-redundant set of 125 aligned sequence sequences to which we added the human CNTF sequence. Computation of sequence identities on this alignment revealed that the closest neighbors to PRF1 are proteins from coelacanth (H3AGE1_LATCH, 31% id), turtle (K7FE02_PELSI, M7BAK3_CHEMY, 28%) and alligator (A0A3Q0FUK7_ALLSI, 27%).
3. Building a phylogenetic tree of the NR set. Using the alignment of the NR set, a Neighbor Joining tree was built with the MEGA software [24] (Fig. 3a). This tree visualizes the PRF, CTF1, CTF2, CLCF1 and CNTF sub-families along with the closest neighbors of PRF1.
4. Visualizing the sequence space of the non-redundant set. Using the `bios2mds` R package [25], we could visualize the sequence space of the PRF1 homologs by multidimensional scaling (Fig. 3b). The first two dimensions are driven by the differences

between CTF1, CTF2 and CLCF1 cytokines. CNTFs and PRFs are projected towards the center of the first two dimensions. To better visualize the relationship between these proteins, we projected the sequences onto the third and fourth dimensions. In this case, we observed the closest neighbors of PRF1 at intermediary positions between PRFs and CNTFs. Indeed, these sequences are above the twilight threshold for human and/or chicken CNTF, strongly supporting the assumption of homology between PRF1 and CNTF (Fig. 2b).

Insert Fig. 3 here.

3.5 Template analysis

As several structures of IL6 cytokines are available, careful graphical analysis may give information on conserved and variable elements that should be taken into account for modeling. Models are less reliable in the variable parts and more reliable in variable parts. Fig. 4a displays the crystal structure of human CNTF superimposed with structure of LIF, IL6, ONCM and IL11. The superposition was carried out with the `super` function of Pymol, based on structures because the low sequence identities between these cytokines prevented the use of the `align` function based on sequence identities.

Insert Fig. 4 here.

The general topology of the four-helix bundle is conserved but specific properties are observed. LIF has shorter helices A and D on the “bottom” side but otherwise does not present marked differences. Shorter helices A and D are also observed for IL6. In addition, IL6 has shorter helices B and C on the “top” side and a 12 residue long α -helix in loop CD. ONCM has a strong distortion in helix A, an insertion with two helical turns in loop BC and two helical elements in loop AB. A helical element is also observed in loop BC of IL11 and distorts the C-terminus of helix B.

3.6 Sequence alignment

MODELLER uses spatial restraints based on the alignment between query and template(s) to build a model. The alignment step is thus crucial for the quality of the model. In the twilight zone, straightforward alignment of the query and the template should be avoided as they do not lead to reliable alignment. It was early acknowledged that MSA methods lead to improved quality of the alignment [9]. Here are some tips:

1. First, use large homolog sets to build an MSA. Then, manually inspect the MSA, compare it to structural alignment of the templates and correct it if necessary with Genedoc [23]. This should limit template/query alignment errors.
2. You may use EXPRESSO [22] to obtain an initial sequence alignment of all the putative templates based on their structural alignment. Compare the sequence alignment to the superimposed structures and correct if necessary.
3. Subsequently, you may use this template profile as a seed to align the query and its homologs.
4. Visual inspection of MSA is mandatory to insure that the alignment is correct.
5. In any case, never forget that sequence alignment may be the limiting step of homology modeling.

3.7 Loop modeling

Long loops are challenging to reliably model due to combination of different factors: (1) they have high sequence and structural variability, (2) they are frequently missing in templates (e.g. the CNTF template in which loops AB and CD are missing and partly missing, respectively), and (3) the functions developed for loop modeling, such as `model_loop` in MODELLER [61], are suited to short loops but not long loops. Two options are possible to

improve modeling of long loops: (1) SS prediction programs may suggest SS elements in these long loops that can be introduced in the modeling procedure; and (2) loops with the same (or, if not possible, similar) size in related proteins may be used as templates.

In our case study, the long loops joining helices A and B, and helices C and D are highly variable and they are missing or partially missing in all the templates except LIF (Fig. 5). However, we note that the loops AB of LIF, ONCM and IL6 present helical elements. In loops CD, a long helix is observed only for IL6. To gain information on putative secondary elements in the long loops of PRF1, we used the SS prediction programs PSIPRED [36] and SPIDER3 [38]. Both programs predicted a helical element at the C-terminus of loop AB that should be taken into account for modeling (Fig. 5). In addition, careful scrutiny of the templates indicates the strong conservation of a proline residue from loop AB at the cleft between helices B and D (Fig. 4b). This position might tether the loop AB at the cytokine surface and should be an element to take into account for PRF1 modeling.

Insert Fig. 5 here.

3.8 Model building

In this last step, we compare the models of PRF1 obtained with three automated modeling procedures through web servers (Phyre2, RosettaCM and I-TASSER) to a customized model built with MODELLER. Concerning the automated models, the four-helix bundle fold was obviously recognized in the three cases, although RosettaCM did not provide information on the templates used. I-TASSER uses the first hits from the best ten out of fourteen threading programs: CNTF (best hit for four programs), LIF (best hit for five programs) and ONCM (best hit for one program). In Phyre2, CNTF, LIF, IL11 and ONCM were found with a confidence level $\geq 99.8\%$.

The customized PRF1 model was built with MODELLER with the following parameters:

1. LIF (PDB 1EMR) and CNTF (PDB 1CNT) were used as templates after structural comparison (Fig. 4).
2. Loop CD was modeled by homology with LIF, with the two residue insertion positioned as indicated in Fig. 5.
3. Loop AB of PRF1 contains a six residue insertion compared to LIF. Thus, the N-terminus of loop AB (up to the anchor proline) was built from the LIF template and the insertion was positioned at the C-terminus with a helical restraint corresponding to the helical prediction.
4. PRF1 Phe-70 is homologous to CNTF Trp-64 which is crucial for receptor alpha binding [62-64]. We did not constraint this residue in the modeling procedure but checked its orientation in the models as a criterion for structural validation.

Fig. 6 displays the three automated models superimposed on the customized MODELLER model. No significant difference was observed for the four-helix bundle core, but the loops AB were markedly different. Among automated models, only the Phyre2 model predicted a position of Pro-61 similar to the customized model. By contrast, only RosettaCM predicted a helical element at the C-terminus of loop AB. However, the exposed orientation of Phe-70, similar to the orientation of Trp-64 in CNTF, was observed only in the customized model, supporting the positioning of the helical element predicted by SPIDER3 and PSIPRED.

Insert Fig. 6 here.

The TM-scores between models and templates (Table 3) can assess the quality of the models. The higher scores obtained by the customized and the Phyre2 models with the CNTF and LIF templates, respectively, reflect the templates favored by MODELLER in the modelling procedure, while the I-TASSER and RosettaCM work by template based fragment assembly or model hybridization, and do not match a template as closely as the customized and Phyre2

models. In any case, the high TM-scores are consistent with similar folds and corroborate the “threadability” of PRF1 [65].

3.9 Concluding remarks

This chapter provides an example of homology modeling in the twilight zone. It is important to note that all the threading methods used to find templates led to similar results (Tables 1 and 2). This is a general observation that threading methods succeed or fail together for molecular modeling of proteins in the twilight zone [65]. Indeed, “best” templates of any method, including CEthreader, share the up-up-down-down four-helix bundle fold of the IL6 family. Nevertheless, several methods found a hit with a beta barrel fold. Caution is thus always recommended when threading methods must be used. Here, the TM-scores obtained for the PRF1 models (Table 3) indicate high “threadability”, which is consistent with the four-helix bundle fold recognized as “best” hit by the different threading methods. The sequence space analysis corroborates the homology between the query and the templates by finding intermediate sequences linking PRF1 and CNTF. The most difficult task remains the modeling of the long loops. In this case, additional information provided by the user may improve the quality of the model, based on structural or functional criteria. Finally, “unthreadability” appears as a property inherent to protein fold [65]. Exploration of the dark zone will require the development of *ab initio* approaches to deduce the three dimensional structure of a protein from physics and physico-chemical principles alone.

4 Notes

1. Multidimensional scaling is a multivariate analysis method that transforms a distance matrix into points in a low dimensional space. The distances between points in the resulting space are as close as possible to the distances in the original matrix [66]. When

applied to an MSA, this method allows the 2D or 3D visualization of the sequences (the sequence space) based on the identity or on the similarity matrix of the sequences in the MSA.

2. LOMETS [33] compares eleven threading methods. Four of these methods (CEthreader, MUSTER, Neff-MUSTER and PPAS) have been developed by Zhang and coworkers [33,67,68]. The methods are based on contact searches (CEthreader), sequence profile – sequence profile searches (MUSTER, Neff-MUSTER, FFAS-3D [69], PPAS, PROSPECT2 [70], SP3 [71], SparksX [72]) and profile HMM – profile HMM searches (HHsearch [31], HHpred [30], and PRC [73]).
3. In RosettaCM, the four independent threading methods used to detect templates are RaptorX [74], HHpred [30], Sparks-X [72], and Map align [75].
4. TM-score is a metric for assessing the topological similarity of protein structures. It gives more weight to smaller deviations than to larger deviations and is thus more sensitive to the global fold than traditional root mean square deviations. In addition, it is scaled by sequence length. TM-scores have values in [0,1] range, with 1 indicating a perfect match between two structures. Scores are higher than 0.5 for structures with similar folds and lower than 0.17 for randomly chosen unrelated proteins [41,43]. Interestingly, a TM-score threshold of 0.4 can also be used to differentiate “threadable” from “non threadable” proteins. The “threadability” of a protein appears related to inherent properties of the protein fold [42].
5. The *Plethodontidae* or lungless salamanders have complex mating behaviors and courtship rituals. PRF1 is produced in a gland on the male’s chin and delivered to the female during courtship by scratching or slapping. A male using the scratching mode of delivery administers the pheromone by wiping his mental gland across the female’s dorsum while scraping her skin to deliver the pheromone directly into the female’s

circulatory system. In the slapping delivery mode, the male slaps its gland directly across the female's nares to deliver the pheromone to chemoreceptors in the vomeronasal organ [76].

6. The cytokines of the IL6 family [49-51] are long chain four-helix bundle proteins. In spite of low sequence identities (8-20%), these cytokines share a common fold with an up-up-down-down topology for the four A to D helices. In mammals, the family includes IL6, IL11, LIF, ONCM, CNTF, CTF1, CTF2 (absent in humans) and CLCF1. These cytokines recruit two receptors, including the common gp130 receptor and a second variable receptor (gp130, LIFR or OSMR). The two receptors transactivate each other to trigger a cellular response through JAK/STAT pathways. In addition, a third receptor (the α receptor) without intracellular part may be recruited to increase the affinity. These receptors bind three conserved sites of the cytokines (Fig. 2). The site I located on helix D and loop AB binds the α receptor, site II located on helices A and C binds gp130 and site III at the top of helix D binds either gp130, LIFR or OSMR. The signature of LIFR binding is an FxxK motif at the N-terminus of helix D.
7. Expect-value or E-value describes the number of hits "expected" by chance when searching a database of a particular size. It depends on both the matching score and the length of the query. The closer to 0 the E-value is, the more significant the hit is.
8. Six threading programs (CEthreader, SparksX, FFAS3D, HHsearch, SP3, PPAS), based on three different search methods, find the interleukin-1 receptor antagonist (IL1Ra) as a putative "good" hit. This protein (PDB: 1ILR) has a very different beta trefoil structure [60]. The beta trefoil fold consists of six beta hairpins, each formed by two beta strands. Together they form a beta barrel with a triangular cap and an approximate three-fold symmetry.

Acknowledgments: This study was supported by institutional grants from INSERM, CNRS and University of Angers. MC is supported by CNRS. RB is supported by a fellowship from the University of Angers (France). AT is supported by a fellowship from the University of Carthage (Tunisia).

References

1. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181:662-666. doi:10.1038/181662a0
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235-242. doi:10.1093/nar/28.1.235
3. (!!! INVALID CITATION !!!)
4. Perdigo N, Rosa A (2019) Dark Proteome Database: Studies on Dark Proteins. *High Throughput* 8 doi:10.3390/ht8020008
5. Sander C, Schneider R (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 9:56-68. doi:10.1002/prot.340090107
6. Rost B (1999) Twilight zone of protein sequence alignments. *Protein Eng* 12:85-94. doi:10.1093/protein/12.2.85
7. Devos D, Valencia A (2000) Practical limits of function prediction. *Proteins* 41:98-107.
8. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779-815. doi:10.1006/jmbi.1993.1626
9. Wallace IM, Blackshields G, Higgins DG (2005) Multiple sequence alignments. *Curr Opin Struct Biol* 15:261-266. doi:10.1016/j.sbi.2005.04.002
10. Kemena C, Notredame C (2009) Upcoming challenges for multiple sequence alignment methods in the high-throughput era. *Bioinformatics* 25:2455-2465. doi:10.1093/bioinformatics/btp452
11. Rollmann SM, Houck LD, Feldhoff RC (1999) Proteinaceous pheromone affecting female receptivity in a terrestrial salamander. *Science* 285:1907-1909. doi:10.1126/science.285.5435.1907
12. UniProt C (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506-D515. doi:10.1093/nar/gky1049
13. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427-D432. doi:10.1093/nar/gky995
14. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY, Finn RD (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47:D351-D360. doi:10.1093/nar/gky1100

15. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Dalenberg K, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranovic V, Guzenko D, Hudson BP, Kalro T, Liang Y, Lowe R, Namkoong H, Peisach E, Periskova I, Prlic A, Randle C, Rose A, Rose P, Sala R, Sekharan M, Shao C, Tan L, Tao YP, Valasatava Y, Voigt M, Westbrook J, Woo J, Yang H, Young J, Zhuravleva M, Zardecki C (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res* 47:D464-D474. doi:10.1093/nar/gky1004
16. Andreeva A, Kulesha E, Gough J, Murzin AG (2020) The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res* 48:D376-D382. doi:10.1093/nar/gkz1064
17. Holm L, Sander C (1998) Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics* 14:423-429. doi:10.1093/bioinformatics/14.5.423
18. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam F, Wallentin F, Wallace JM, Willm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948. doi:10.1093/bioinformatics/btm404
19. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47:W636-W641. doi:10.1093/nar/gkz268
20. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi:10.1186/1471-2105-5-113
21. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205-217. doi:10.1006/jmbi.2000.4042
22. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 34:W604-608. doi:10.1093/nar/gkl092
23. Nicholas KB, Jr NHB, Deerfield DWI (1999) GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* 4:14.
24. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739. doi:10.1093/molbev/msr121
25. Pele J, Becu JM, Abdi H, Chabbert M (2012) Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. *BMC Bioinformatics* 13:133. doi:10.1186/1471-2105-13-133
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410. doi:10.1016/S0022-2836(05)80360-2
27. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402. doi:10.1093/nar/25.17.3389
28. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755-763. doi:10.1093/bioinformatics/14.9.755
29. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845-858. doi:10.1038/nprot.2015.053
30. Soding J, Remmert M (2011) Protein sequence comparison and fold recognition: progress and good-practice benchmarking. *Curr Opin Struct Biol* 21:404-411. doi:10.1016/j.sbi.2011.03.005
31. Soding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951-960. doi:10.1093/bioinformatics/bti125
32. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* 430:2237-2243. doi:10.1016/j.jmb.2017.12.007
33. Zheng W, Zhang C, Wuyun Q, Pearce R, Li Y, Zhang Y (2019) LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Res* 47:W429-W436. doi:10.1093/nar/gkz384
34. Pandurangan AP, Stahlhake J, Oates ME, Smithers B, Gough J (2019) The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res* 47:D490-D494. doi:10.1093/nar/gky1130
35. Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313:903-919. doi:10.1006/jmbi.2001.5080
36. Buchan DWA, Jones DT (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* 47:W402-W407. doi:10.1093/nar/gkz297
37. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195-202. doi:10.1006/jmbi.1999.3091
38. Heffernan R, Yang Y, Paliwal K, Zhou Y (2017) Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. *Bioinformatics* 33:2842-2849. doi:10.1093/bioinformatics/btx218
39. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725-738. doi:10.1038/nprot.2010.5
40. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12:7-8. doi:10.1038/nmeth.3213
41. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40. doi:10.1186/1471-2105-9-40
42. Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, Thompson J, Baker D (2013) High-resolution comparative modeling with RosettaCM. *Structure* 21:1735-1742. doi:10.1016/j.str.2013.08.005
43. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* 57:702-710. doi:10.1002/prot.20264
44. Xu J, Zhang Y (2010) How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* 26:889-895. doi:10.1093/bioinformatics/btq066
45. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302-2309. doi:10.1093/nar/gki524
46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-1612. doi:10.1002/jcc.20084
47. Derouet D, Rousseau F, Alfonsi F, Froger J, Hermann J, Barbier F, Perret D, Diveu C, Guillet C, Preisser L, Dumont A, Barbado M, Morel A, deLapeyriere O, Gascan H, Chevalier S (2004) Neurotrophin, a new IL-6-related cytokine signaling through the ciliary neurotrophic factor receptor. *Proc Natl Acad Sci U S A* 101:4827-4832. doi:10.1073/pnas.0306178101
48. Senaldi G, Varnum BC, Sarmiento U, Starnes C, Lile J, Scully S, Guo J, Elliott G, McNinch J, Shaklee CL, Freeman D, Manu F, Simonet WS, Boone T, Chang MS (1999) Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the IL-6 family. *Proc Natl Acad Sci U S A* 96:11458-11463. doi:10.1073/pnas.96.20.11458
49. Heinrich PC, Behrmann I, Haan S, Hermanns HM, Muller-Newen G, Schaper F (2003) Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J* 374:1-20. doi:10.1042/BJ20030407
50. Huisling MO, Kruiswijk CP, Flik G (2006) Phylogeny and evolution of class-I helical cytokines. *J Endocrinol* 189:1-25. doi:10.1677/joe.1.06591
51. Rose-John S (2018) Interleukin-6 Family Cytokines. *Cold Spring Harb Perspect Biol* 10 doi:10.1101/cshperspect.a028415

52. Sims NA (2015) Cardiotrophin-like cytokine factor 1 (CLCF1) and neuropoietin (NP) signalling and their roles in development, adulthood, cancer and degenerative disorders. *Cytokine Growth Factor Rev* 26:517-522. doi:10.1016/j.cytogfr.2015.07.014
53. Somers W, Stahl M, Seehra JS (1997) 1.9 A crystal structure of interleukin 6: implications for a novel mode of receptor dimerization and signaling. *EMBO J* 16:989-997. doi:10.1093/emboj/16.5.989
54. Adams R, Burnley RJ, Valenzano CR, Qureshi O, Doyle C, Lumb S, Del Carmen Lopez M, Griffin R, McMillan D, Taylor RD, Meier C, Mori P, Griffin LM, Wernery U, Kinne J, Rapecki S, Baker TS, Lawson AD, Wright M, Ettorre A (2017) Discovery of a junctional epitope antibody that stabilizes IL-6 and gp80 protein:protein interaction and modulates its downstream signaling. *Sci Rep* 7:37716. doi:10.1038/srep37716
55. Putoczki TL, Dobson RC, Griffin MD (2014) The structure of human interleukin-11 reveals receptor-binding site features and structural differences from interleukin-6. *Acta Crystallogr D Biol Crystallogr* 70:2277-2285. doi:10.1107/S1399004714012267
56. McDonald NQ, Panayotatos N, Hendrickson WA (1995) Crystal structure of dimeric human ciliary neurotrophic factor determined by MAD phasing. *EMBO J* 14:2689-2699.
57. Robinson RC, Grey LM, Staunton D, Vankelecom H, Vernallis AB, Moreau JF, Stuart DJ, Heath JK, Jones EY (1994) The crystal structure and biological function of leukemia inhibitory factor: implications for receptor binding. *Cell* 77:1101-1116. doi:10.1016/0092-8674(94)90449-9
58. Huyton T, Zhang JG, Luo CS, Lou MZ, Hilton DJ, Nicola NA, Garrett TP (2007) An unusual cytokine:Ig-domain interaction revealed in the crystal structure of leukemia inhibitory factor (LIF) in complex with the LIF receptor. *Proc Natl Acad Sci U S A* 104:12737-12742. doi:10.1073/pnas.0705577104
59. Deller MC, Hudson KR, Ikemizu S, Bravo J, Jones EY, Heath JK (2000) Crystal structure and functional dissection of the cytoskeletal cytokine oncostatin M. *Structure* 8:863-874. doi:10.1016/S0969-2126(00)00176-3
60. Schreuder HA, Rondeau JM, Tardif C, Soffientini A, Sarubbi E, Akesson A, Bowlin TL, Yanofsky S, Barrett RW (1995) Refined crystal structure of the interleukin-1 receptor antagonist. Presence of a disulfide link and a cis-proline. *Eur J Biochem* 227:838-847. doi:10.1111/j.1432-1033.1995.tb20209.x
61. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* 374:461-491. doi:10.1016/S0076-6879(03)74020-8
62. Panayotatos N, Radziejewska E, Acheson A, Somogyi R, Thadani A, Hendrickson WA, McDonald NQ (1995) Localization of functional receptor epitopes on the structure of ciliary neurotrophic factor indicates a conserved, function-related epitope topography among helical cytokines. *J Biol Chem* 270:14007-14014. doi:10.1074/jbc.270.23.14007
63. Perret D, Guillet C, Elson G, Froger J, Plun-Favreau H, Rousseau F, Chabbert M, Gauchat JF, Gascan H (2004) Two different contact sites are recruited by cardiotrophin-like cytokine (CLC) to generate the CLC/CLF and CLC/sCNTFRalpha composite cytokines. *J Biol Chem* 279:43961-43970. doi:10.1074/jbc.M407686200
64. Plun-Favreau H, Elson G, Chabbert M, Froger J, deLapeyriere O, Lelievre E, Guillet C, Hermann J, Gauchat JF, Gascan H, Chevalier S (2001) The ciliary neurotrophic factor receptor alpha component induces the secretion of and is required for functional responses to cardiotrophin-like cytokine. *EMBO J* 20:1692-1703. doi:10.1093/emboj/20.7.1692
65. Skolnick J, Zhou H (2017) Why Is There a Glass Ceiling for Threading Based Protein Structure Prediction Methods? *J Phys Chem B* 121:3546-3554. doi:10.1021/acs.jpcc.6b09517
66. Abdi H (2007) Metric multidimensional scaling. In: *Encyclopedia of Measurement and Statistics*. Salkind NJ, editor. Sage. Thousand Oaks (CA). pp. 598-605.
67. Wu S, Zhang Y (2008) MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* 72:547-556. doi:10.1002/prot.21945
68. Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* 35:3375-3382. doi:10.1093/nar/gkm251
69. Xu D, Jaroszewski L, Li Z, Godzik A (2014) FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* 30:660-667. doi:10.1093/bioinformatics/btt578
70. Xu Y, Xu D (2000) Protein threading using PROSPECT: design and evaluation. *Proteins* 40:343-354.
71. Zhou H, Zhou Y (2005) Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. *Proteins* 58:321-328. doi:10.1002/prot.20308
72. Yang Y, Faraggi E, Zhao H, Zhou Y (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* 27:2076-2082. doi:10.1093/bioinformatics/btr350
73. Madera M (2008) Profile Comparer: a program for scoring and aligning profile hidden Markov models. *Bioinformatics* 24:2630-2631. doi:10.1093/bioinformatics/btn504
74. Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7:1511-1522. doi:10.1038/nprot.2012.085
75. Ovchinnikov S, Park H, Varghese NU, Huang PS, Pavlopoulos GA, Kim DE, Kamisetty H, Kyrpides NC, Baker D (2017) Protein structure determination using metagenome sequence data. *Science* 355:294-298. doi:10.1126/science.aah4043
76. Palmer CA, Watts RA, Gregg RG, McCall MA, Houck LD, Highton R, Arnold SJ (2005) Lineage-specific differences in evolutionary mode in a salamander courtship pheromone. *Mol Biol Evol* 22:2243-2256. doi:10.1093/molbev/msi219

Figure Captions

Fig. 1: Schematic representation of the dark, twilight and safe zones for molecular modeling of a protein as a function of the sequence identity and aligned length between the query and the template. Above the yellow line (drawn from [5]), the light grey zone indicates the safe zone of homology modeling. Below the yellow zone, the dark grey zone indicates the twilight zone for which molecular modeling becomes increasingly difficult because sequence identity does not infer structural conservation. In the twilight zone, templates cannot be found with BLAST and threading methods must be used. When they fail, the dark zone is reached. Note that any two proteins have at least 5% sequence identity (dashed blue line).

Fig. 2: Cytokines of the IL6 family. (a) General up-up-down-down topology of the four-helix bundle fold of this cytokine family. The helices are numbered from A to D. The positions of

the three conserved sites of interaction with the cognate receptors are indicated. Site II interacts with gp130, site III interacts with either gp130, LIFR or OSMR, while site I can interact with a third, specific, “ α ” receptor. See Note 6 for details; (b) Sequence identities between PRF1, its closest homologs and the cytokines of the IL6 family. The color code indicates the reliability of the sequence identity (green: safe zone, yellow: transition zone, red: twilight zone).

Fig. 3: Evolutionary information on PRF1. (a) NJ tree of PRF1 homologs (500 bootstraps); (b) Sequence space of the PRF1 homologs with projection of the sequences onto the first and second dimensions (top) and onto the third and fourth dimensions (bottom) of the MDS analysis. The PRF1 homologs were obtained by mining UniProt with BLAST, followed by clustering to obtain a non-redundant set. The MDS analysis was carried out with the `bios2mds` package. In (a) and (b), the color code is as follows: PRF proteins: red, CNTF: orange, CTF1: dark blue, CTF2: magenta, CLCF1: green, the closest four homologs: black, others: grey. In (a), labels indicate PRF1 (red circle), human CTF1 (blue circle), murine CTF2 (cyan circle), human CLCF1 (green circle), human CNTF (orange diamond), chicken CNTF (orange circle), K7FE02_PELSI (black circle), M7BAK3_CHEMY (black square), H3AGE1 LATCH (black triangle) and A0A3Q0FUQ7_ALLSI (black diamond). In (b), only black labels for closest PRF1 homologs are shown.

Fig. 4: Comparison of different templates. In (a), a ribbon representation of CNTF (white) is superimposed with LIF (orange), IL6 (cyan), ONCM (magenta) and IL11 (blue). Regions structurally different as compared to CNTF are indicated by a darker color for clarity purpose. In (b), zooming on the AB loop reveals the conserved positioning of a proline residue (shown as stick) in the cleft between helices B and D in LIF, IL6, ONCM and IL11. Note that the

loop AB is absent in the CNTF structure. The PDB files are 1CNT (CNTF), 1EMR (LIF), 5FUC (IL6), 1EVS (ONCM) and 4MHL (IL11).

Fig. 5: Alignment of PRF1 versus a structural alignment of the templates. The initial alignment based on structure superposition was obtained with EXPRESSO and manually refined. The SS predictions were obtained by SPIDER3 and PSIPRED from the sequence of PRF1. Residues in blue indicate the sequences of the peptide signal. Residues in italic are missing in the PDB files. The homologous CNTF Trp-64 and PRF1 Trp-70 residues are highlighted in yellow on black background because this position may be involved in the binding of a specific “ α ” receptor.

Fig. 6: Comparison of the PRF1 models. In (a), the customized MODELLER model (white and yellow) is superimposed with the best models obtained with RosettaCM (blue), I-TASSER (green) and Phyre2 (pink). Pro-61 and Phe-70 are indicated by sticks. The yellow ribbon in the MODELLER model indicates the user-added helical restrains. In (b), the zooming on the C-terminal part of loop AB reveals that the superposition of the phenyl ring of CNTF Trp-64 (cyan) with PRF1 Phe-70 (yellow) in the customized MODELLER model is not observed in the other three models.

TABLE 1**PDB mining using the PRF1 sequence as a query**

Search method	Program	Hits ¹	E-value
Sequence based	BLAST	No hit	
Sequence profile based	PSI-BLAST	CNTF	5.9
Profile HMM based	HMMER	CNTF	7 10 ⁻²⁰
		LIF	4 10 ⁻¹⁰
Profile HMM – Profile HMM based	HHpred	CNTF	7 10 ⁻³²
		LIF	3 10 ⁻³¹
		IL11	5 10 ⁻²⁸
		ONCM	8 10 ⁻²⁶
		<i>GCSK</i>	8 10 ⁻¹³
		IL6	2 10 ⁻⁷

¹ For clarity purpose, only the proteins (and not the PDB numbers) are indicated. Italics fonts indicate a growth factor with the same four-helix bundle fold as the IL6 family.

TABLE 2**Comparison of the LOMETS threading programs using the PRF1 sequence as a query**

Program ¹	Method ²	Top hit ³	Additional hits with $Z_n > 1$ ^{3,4}
HHpred	HMM	CNTF	LIF, IL11, ONCM
CEthreder	Contact	<i>Prolactin</i>	IL11, CNTF, <i>lactogen</i> , LIF, IL1Ra
SparksX	Profile	CNTF	LIF, IL11, ONCM, IL1Ra , IL6, <i>prolactin</i>
FFAS3D	Profile	CNTF	LIF, ONCM, IL11, <i>GCSF</i> , IL6, IL1Ra , <i>prolactin</i> , <i>lactogen</i>
MUSTER	Profile	LIF	CNTF, ONCM, IL11
Neff-MUSTER	Profile	LIF	CNTF
HHsearch	HMM	LIF	CNTF, ONCM, IL11, <i>GCSF</i> , IL6, IL1Ra
SP3	Profile	LIF	CNTF, ONCM, IL11, IL6, <i>GCSF</i> , IL1Ra , <i>prolactin</i>
PPAS	Profile	CNTF	LIF, ONCM, IL11, <i>GCSF</i> , IL6, IL1Ra , <i>IL23</i>
PROSPECTOR2	Profile	CNTF	LIF, ONCM, IL11
PRC	HMM	LIF	CNTF, ONCM, IL11

¹ The programs are ranked as determined by LOMETS. See Note 2 for references.

² HMM corresponds to profile HMM – profile HMM based searches; profile corresponds to sequence profile – sequence profile based searches.

³ For clarity purpose, only the proteins (and not the PDB numbers) are indicated.

⁴ Normalized Z-scores (Z_n) indicate the quality of the hits. They are considered “good” above the threshold of 1. The hits are sorted by decreasing Z_n . When several hits correspond to the same protein (different origins, conditions or methods), only the first hit is indicated. Italics fonts indicate four-helix bundle cytokines/growth factors that do not belong to the IL6 family but share the same fold. The bold fonts for IL1Ra highlight a hit with a beta barrel fold.

TABLE 3

TM-scores of the PRF1 models

Method	TM-score ¹	
	CNTF	LIF
I-TASSER	0.71	0.76
RosettaCM	0.69	0.69
Phyre2	0.68	0.88
Customized	0.78	0.85

¹ The TM-scores between the PRF1 models and the CNTF and LIF templates were computed using the length of PRF1 as reference.

Figure 1

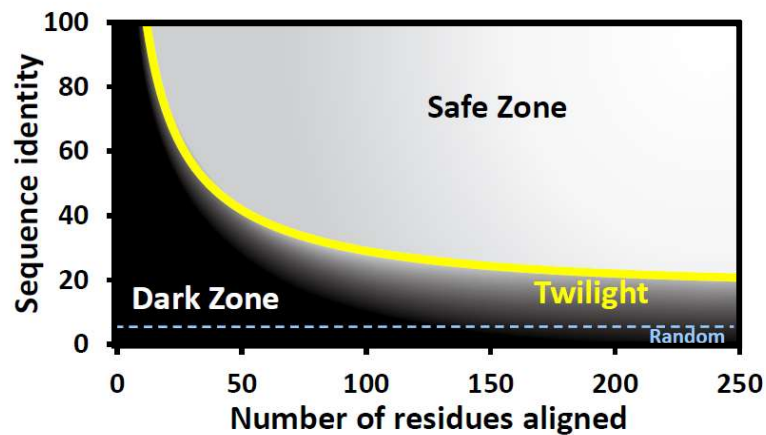


Figure 2

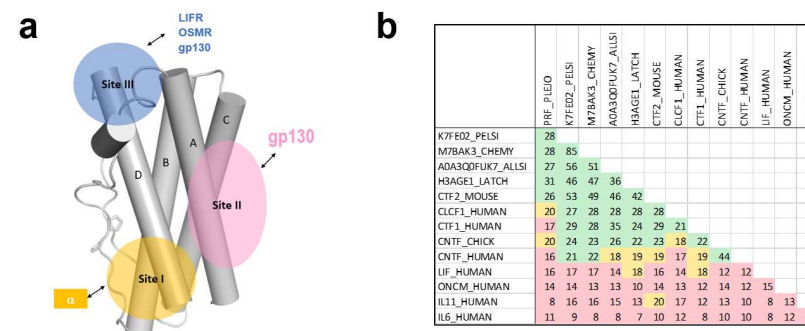
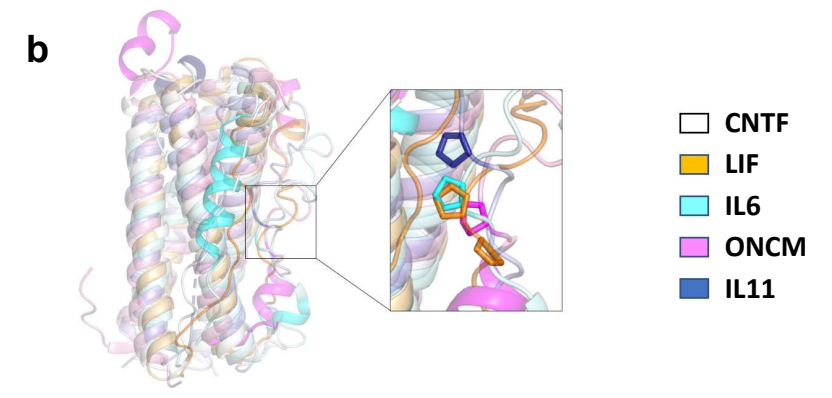
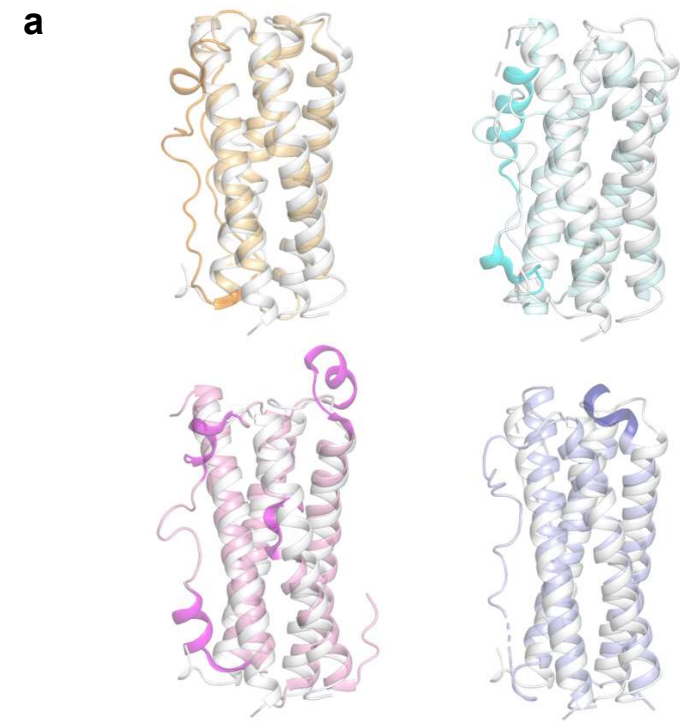
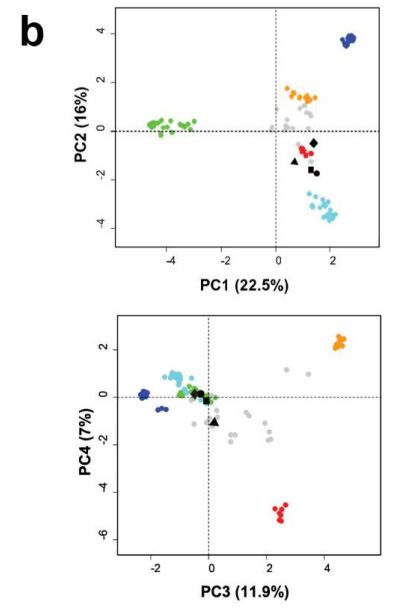
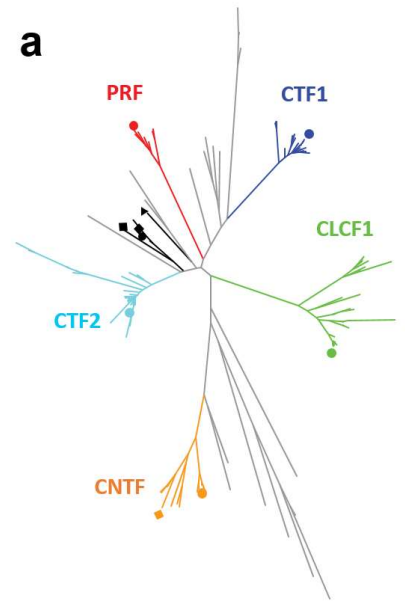


Fig. 3

Fig. 4



Annexe II.3

Autres publications

Abstract 1: [G Protein-coupled Receptors in Radioiodine-refractory Thyroid Cancer in the Era of Precision Medicine](#). Suteau V, Seegers V, Munier M, **Ben Boubaker R**, Reyes C, Gentien D, Wery M, Croué A, Illouz F, Hamy A, Rodien P, Briet C. *J Clin Endocrinol Metab*. 2021 Jul 13;106(8):2221-2232

Abstract 2: [Identification of dysregulated expression of G protein coupled receptors in endocrine tumors by bioinformatics analysis: potential drug targets](#). Suteau V, Munier M, **Ben Boubaker R**, Wery M, Henrion D, Rodien P, Briet C. *Cells* 2022, 11, 703.

Contexte et principales contributions

Travailler dans une équipe multidisciplinaire avec des médecins et des biologistes m'a permis d'assister mes collègues par la conception des outils nécessaires à l'analyse de leurs données cliniques, ce qui a résulté en deux publications dont je suis co-auteur.

Dans ces 2 articles j'ai conçu les scripts R nécessaires pour filtrer les données expérimentales et bio-informatiques et les représenter en forme de heatmaps ou volcanoplots afin d'identifier les RCPG potentiellement d'intérêt thérapeutique pour le traitement du cancer et qui peuvent être ciblés par des médicaments approuvés à partir de bases de données pharmacologiques.

Ces articles sont représentés en forme d'abstract dans ce rapport mais sont accessible en ligne sous les liens suivants :

- doi: 10.1210/clinem/dgab343.PMID: 34000025
- <https://doi.org/10.3390/cells11040703>

G Protein–coupled Receptors in Radioiodine–refractory Thyroid Cancer in the Era of Precision Medicine Get access

Valentine Suteau ✉, Valérie Seegers, Mathilde Munier, Rym Ben Boubaker, Cécile Reyes, David Gentien, Méline Wery, Anne Croué, Frédéric Illouz, Antoine Hamy ... [Show more](#)

The Journal of Clinical Endocrinology & Metabolism, Volume 106, Issue 8, August 2021, Pages 2221–2232, <https://doi.org/10.1210/clinem/dgab343>

Published: 17 May 2021 **Article history** ▼

Abstract

Context

Radioiodine–refractory thyroid cancers have poor outcomes and limited therapeutic options (tyrosine kinase inhibitors) due to transient efficacy and toxicity of treatments. Therefore, combinatorial treatments with new therapeutic approaches are needed. Many studies link G protein–coupled receptors (GPCRs) to cancer cell biology.

Objective

To perform a specific atlas of GPCR expression in progressive and refractory thyroid cancer to identify potential targets among GPCRs aiming at drug repositioning.

Methods

We analyzed samples from tumor and normal thyroid tissues from 17 patients with refractory thyroid cancer (12 papillary thyroid cancers [PTCs] and 5 follicular thyroid cancers [FTCs]). We assessed GPCR mRNA expression using NanoString technology with a custom panel of 371 GPCRs. The data were compared with public repositories and pharmacological databases to identify eligible drugs. The analysis of prognostic value of genes was also performed with TCGA datasets.

Results

With our transcriptomic analysis, 4 receptors were found to be downregulated in FTC (*VIPR1*, *ADGRL2/LPHN2*, *ADGRA3*, and *ADGRV1*). In PTC, 24 receptors were deregulated, 7 of which were also identified by bioinformatics analyses of publicly available datasets on primary thyroid cancers (*VIPR1*, *ADORA1*, *GPRC5B*, *P2RY8*, *GABBR2*, *CYSLTR2*, and *LPAR5*). Among all the differentially expressed genes, 22 GPCRs are the target of approved drugs and some GPCRs are also associated with prognostic factors.

Discussion



For the first time, we performed GPCR mRNA expression profiling in progressive and refractory thyroid cancers. These findings provide an opportunity to identify potential therapeutic targets for drug repositioning and precision medicine in radioiodine–refractory thyroid cancer.

Keywords: [Thyroid cancer](#), [papillary, thyroid cancer](#), [follicular](#), [G protein–coupled receptors](#), [molecular targeted therapies](#), [prognostic factor](#)

Issue Section: [Clinical Research Articles](#)

Article

Identification of Dysregulated Expression of G Protein Coupled Receptors in Endocrine Tumors by Bioinformatics Analysis: Potential Drug Targets?

Valentine Suteau ^{1,2,*,†} , Mathilde Munier ^{1,2,3,†} , Rym Ben Boubaker ² , Méline Wery ⁴, Daniel Henrion ² , Patrice Rodien ^{1,2,3,5}  and Claire Briet ^{1,2,3,5}

- ¹ Département d'Endocrinologie-Diabétologie Nutrition, Centre Hospitalier Universitaire (CHU) d'Angers, 49933 Angers, France; mathilde.munier@univ-angers.fr (M.M.); parodien@chu-angers.fr (P.R.); claire.briet@chu-angers.fr (C.B.)
- ² Laboratoire MITOVASC, UMR CNRS 6015, INSERM 1083, Université d'Angers, 49100 Angers, France; rym.benboubaker@univ-angers.fr (R.B.B.); daniel.henrion@univ-angers.fr (D.H.)
- ³ Centre de Référence des Maladies Rares de la Thyroïde et des Récepteurs Hormonaux, Centre Hospitalier Universitaire (CHU) d'Angers, 49933 Angers, France
- ⁴ Faculté de Santé, Université d'Angers, SFR ICAT, 49100 Angers, France; meline.wery@univ-angers.fr
- ⁵ Centre de Référence des Maladies Rares de l'Hypophyse, Centre Hospitalier Universitaire (CHU) d'Angers, 49933 Angers, France
- * Correspondence: valentine.suteau@chu-angers.fr
- † These authors contributed equally to this work.



Citation: Suteau, V.; Munier, M.; Ben Boubaker, R.; Wery, M.; Henrion, D.; Rodien, P.; Briet, C. Identification of Dysregulated Expression of G Protein Coupled Receptors in Endocrine Tumors by Bioinformatics Analysis: Potential Drug Targets? *Cells* **2022**, *11*, 703. <https://doi.org/10.3390/cells11040703>

Academic Editor: Hiroshi Miyamoto

Received: 14 January 2022

Accepted: 15 February 2022

Published: 17 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Many studies link G protein-coupled receptors (GPCRs) to cancer. Some endocrine tumors are unresponsive to standard treatment and/or require long-term and poorly tolerated treatment. This study explored, by bioinformatics analysis, the tumoral profiling of the GPCR transcriptome to identify potential targets in these tumors aiming at drug repurposing. Methods: We explored the GPCR differentially expressed genes (DEGs) from public datasets (Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA)). The GEO datasets were available for two medullary thyroid cancers (MTCs), eighty-seven pheochromocytomas (PHEOs), sixty-one paragangliomas (PGLs), forty-seven pituitary adenomas and one-hundred-fifty adrenocortical cancers (ACCs). The TCGA dataset covered 92 ACCs. We identified GPCRs targeted by approved drugs from pharmacological databases (ChEMBL and DrugBank). Results: The profiling of dysregulated GPCRs was tumor specific. In MTC, we found 14 GPCR DEGs, including an upregulation of the dopamine receptor (*DRD2*) and adenosine receptor (*ADORA2B*), which were the target of many drugs. In PGL, seven GPCR genes were downregulated, including vasopressin receptor (*AVPR1A*) and PTH receptor (*PTH1R*), which were targeted by approved drugs. In ACC, *PTH1R* was also downregulated in both the GEO and TCGA datasets and was the target of osteoporosis drugs. Conclusions: We highlight specific GPCR signatures across the major endocrine tumors. These data could help to identify new opportunities for drug repurposing.

Keywords: G protein-coupled receptors; endocrine tumors; paraganglioma; pheochromocytoma; adrenocortical cancer; medullary thyroid cancer; pituitary adenoma; drug repurposing

1. Introduction

G protein-coupled receptors (GPCRs) are the largest family of membrane receptors involved in many types of cellular responses. The GPCR family represents approximately 4% of the human genes, with more than 800 members [1]. GPCRs are involved in important functions, such as cardiac function, hormone regulation, immune responses and neurotransmission. Abnormal expression or activity is associated with several human diseases [2]. As a result, GPCRs are considered as therapeutic targets in many diseases. Drugs that target GPCRs account for about 34% of the current medicines, making it the

RÉSUMÉ

Résumé : Les récepteurs de l'angiotensine II de type 1 (AT1) et de type 2 (AT2) sont deux récepteurs qui appartiennent à la superfamille des RCPG. Ces récepteurs sont impliqués fortement dans l'homéostasie cardiovasculaire et la pression artérielle. En absence de leur ligand naturel, certains RCPG sont activables par des forces mécaniques (forces de cisaillement ou étirement membranaire) ou par des lipides anioniques tels que les DOPG mais nous ignorons encore comment ces facteurs externes modifient les mécanismes moléculaires d'activation des récepteurs de l'angiotensine. Nous avons donc réalisé des simulations de dynamique moléculaire classique et dirigée des récepteurs AT1 et AT2 en modélisant le stress mécanique sous forme de tension de surface, mimant ainsi l'effet d'étirement. Nous avons examiné les mécanismes d'activation de ces deux récepteurs en absence et en présence de l'ion sodium, un régulateur allostérique négatif, ainsi que dans des environnements lipidiques différents (DOPG/POPC). Nos simulations dynamiques classiques montrent que l'activation du récepteur AT1 par l'effet des lipides anioniques DOPG couplé à une tension de 20 dyn/cm conduit à l'adoption d'une conformation active particulière « non-canonique ». En utilisant les simulations dynamiques dirigées, nous avons montré les étapes mécanistiques d'activation des récepteurs AT1 et AT2 et nous avons pu ainsi identifier la structure non canonique d'AT1 comme une structure intermédiaire d'activation. Cette structure est modulée par les interactions électrostatiques entre les hélices 6 et 7, en particulier les interactions entre D6.32 et K7.58. Par conséquent, toute perturbation de ces interactions par des facteurs externes peut initier l'activation du récepteur d'AT1 en absence de son ligand. Ainsi, l'activation du récepteur AT1 par les lipides anioniques peut être expliquée par l'interaction des DOPG avec les résidus chargés positivement de l'hélice 7 et 8 tel que les lysines K7.58 et K7.61. Ces résultats mettent en évidence un site allostérique qui pourrait être ciblé en drug design.

mots-clés : AT1, AT2, RCPG, dynamique moléculaire, stress mécanique.

ABSTRACT

Angiotensin II type 1 (AT1) and type 2 (AT2) receptors are two receptors that belong to the GPCR superfamily. These receptors are strongly involved in cardiovascular homeostasis and blood pressure. In the absence of their natural ligand, some GPCRs can be activated by mechanical forces (shear forces or membrane stretching) or by anionic lipids such as DOPGs, but the effects of these external factors on the AT1 and AT2 molecular mechanisms of activation are still unknown. To answer these questions, we performed classical and steered molecular dynamics simulations of AT1 and AT2 receptors by modeling mechanical stress as surface tension, thus mimicking stretching. We analyzed the activation mechanisms of these two receptors in the absence and presence of the sodium ion, a negative allosteric regulator, and in different lipid environments (DOPG/POPC). Our classical dynamics simulations show that activation of the AT1 receptor by the effect of anionic DOPG lipids coupled to a surface tension of 20 dyn/cm leads to the adoption of a particular "non-canonical" active conformation. Steered dynamics simulations reveal the mechanistic steps of AT1 and AT2 activation and thus identify the non-canonical structure of AT1 as an intermediate structure in the activation process. This structure is modulated by electrostatic interactions between helices 6 and 7, in particular the interactions between D6.32 and K7.58. Therefore, any disruption of these interactions by external factors may initiate AT1 receptor activation in the absence of ligand. Thus, activation of the AT1 receptor by anionic lipids can be explained by the interaction of DOPGs with positively charged residues of helix 7 and 8 such as K7.58 and K7.61. These results highlight an allosteric site that could be targeted in drug design.

keywords : AT1, AT2, GPCR, molecular dynamics simulations, mechanical stress.