

Modélisation moléculaire de la perception de la saveur sucrée : approches structurales et dynamiques

Jean-Baptiste Cheron

► To cite this version:

Jean-Baptiste Cheron. Modélisation moléculaire de la perception de la saveur sucrée : approches structurales et dynamiques. Autre. Université Côte d'Azur, 2017. Français. <NNT : 2017AZUR4066>. <tel-01677089>

HAL Id: tel-01677089

<https://tel.archives-ouvertes.fr/tel-01677089>

Submitted on 8 Jan 2018

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

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



École doctorale Sciences Fondamentales et Appliquées
Unité de recherche : Institut de Chimie de Nice

THÈSE DE DOCTORAT

Présentée en vue de l'obtention du
grade de docteur en Chimie
de
L'Université Côte d'Azur par

JEAN-BAPTISTE CHÉRON

Modélisation moléculaire de la perception de la saveur sucrée : approches structurales et dynamiques.

Co-dirigée par Serge Antonczak, Professeur des Universités
et par Sébastien Fiorucci, Maître de Conférence Habilité à Diriger des Recherches,
Institut de Chimie de Nice.

Soutenue le 20/09/2017

Devant le jury composé de :

Francine	Acher	Directrice de Recherche au CNRS, Université Paris Descartes	Examinatrice
Serge	Antonczak	Professeur des Universités, Université Côte d'Azur	Directeur de thèse
Alexandre	de Brevern	Directeur de Recherche à l'INSERM, Université Paris Diderot	Rapporteur
Philippe	Derreumaux	Professeur des Universités, Université Paris Diderot	Rapporteur
Sébastien	Fiorucci	Maître de Conférence, Université Côte d'Azur	Co-Directeur de thèse
Fabrice	Neiers	Maître de Conférence, Université de Bourgogne	Examineur

« La vraie faute est celle qu'on ne corrige pas. »

Confucius

UNIVERSITÉ CÔTE D'AZUR

Résumé

Faculté des Sciences

Département de Chimie

Institut de Chimie de Nice

Modélisation moléculaire de la perception de la saveur sucrée : approches structurales et dynamiques.

La consommation excessive et chronique de sucre est un facteur de risque pour l'apparition de pathologies telles que le diabète de type II ou l'obésité. Une des solutions pour répondre à cet enjeu majeur de santé publique, tout en conservant le plaisir de la saveur sucrée, consiste en l'utilisation d'édulcorants en substitution du sucre. Actuellement, un certain nombre d'édulcorants sur le marché présentent arrière-goût amer ou sont sujets à débat quant à leurs effets sur la santé. Un des objectifs de ces travaux de thèse consiste à proposer de nouveaux édulcorants grâce à des approches rationnelles *in silico*. Un modèle statistique a été établi sur la base des structures chimiques et a permis d'identifier de nouveaux édulcorants d'origine naturelle. Ensuite, la reconstruction par homologie du récepteur à la saveur sucrée et l'étude des sites de liaison apportent des indices, à l'échelle atomique, qui permettront d'identifier ou même de concevoir de nouveaux édulcorants. L'étude dynamique d'un récepteur de la même famille (Récepteur Couplé aux Protéines G (RCPG) de classe C) a permis d'émettre une hypothèse sur le mécanisme d'activation, phénomène important pour la compréhension de la conversion du signal chimique en signal électrique.

Mots clés : RCPG, classe C, Édulcorants, Pouvoir Sucrant, QSAR, Modélisation Moléculaire.

Molecular modeling of the sweet taste perception : structural and dynamic approaches.

Sugar overconsumption is a risk factor for pathologies such as type II diabetes or obesity. Sweeteners consumption is used to overcome this public health issue. Indeed, they have low caloric value but still preserve the pleasure of sweet taste. Currently, number of sweeteners are commercially available, but they present a bitter aftertaste or there is a debate about their safety. One aim of this work was to propose new intense sweeteners using computational modeling strategies. Through a statistical approach to predict the sweetness based on the chemical structure of already known sweeteners, new natural compounds have been identified. Furthermore, the structural study of the homology model of the sweet taste receptor provides some clues to design new sweeteners. The molecular dynamic study of a class C G-protein coupled receptor gives the first molecular hypothesis of the activation process.

Keywords : GPCR, Class C, Sweeteners, Sweetness, QSAR, Molecular Modeling.

Remerciements

En premier lieu, je tiens à remercier mes directeurs de thèse, le Dr. Sébastien Fiorucci et le Pr. Serge Antonczak. Merci pour m'avoir accueilli dans votre équipe, pour vos précieux conseils, la confiance et la liberté que vous m'avez accordées. Je n'oublie pas le Pr. Jérôme Golebiowski pour son regard bienveillant sur mes travaux de thèse et pour l'évolution de l'infrastructure informatique dont j'ai pu profiter. J'ai toujours eu plaisir à venir travailler, chaque jour de ces trois années, grâce à votre soutien et votre bonne humeur.

Je tiens à remercier :

- le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche pour avoir financé cette thèse.
- l'Université Côte d'Azur et le département de chimie de m'avoir donné l'opportunité d'obtenir un contrat d'enseignement.
- l'Institut de Chimie de Nice pour m'avoir accueilli dans ses locaux.
- le *Giract* et la *Gen Foundation* pour avoir aidé à financer mes projets de recherches.
- le COST Action GLISTEN et l'ISOT pour les subventions de déplacement en congrès scientifique.

J'exprime ma profonde gratitude au Pr. Philippe Derreumaux et au Dr. Alexandre de Brevern qui ont accepté d'être rapporteur de ma thèse. Je remercie également le Dr. Francine Acher et le Dr. Fabrice Neiers pour avoir examiné ces travaux.

J'ai une pensée particulière pour Claire et Julien m'ont accueilli chaleureusement à mon arrivée au laboratoire. Je tiens également à remercier Jérémie, pour ta spontanéité et ton humour. Merci pour ta confiance, ton soutien dans les projets scientifiques et la maintenance informatique. Je n'oublie pas Xiaojing Cong et Caroline Bushdid (alias Carole Bushid) pour leur bonne humeur. Plus largement je remercie toute l'équipe, j'ai appris énormément d'un point de vue professionnel et personnel : il y a l'avant et l'après *chemosim* !

Merci à Antoine, Florian, Daniel et Pierrick pour tous les bons moments passés à Nice. Plus généralement, merci pour l'accueil à l'ensemble du 3^{ème} et j'ai une pensée pour toutes les personnes de l'ICN que j'ai pu côtoyer au cours des « séminaires et repas des non permanents » ces dernières années. Je souhaite aussi remercier les Nantais qui sont venus me visiter à Nice, mention spéciale à Gaby, Seb et F. Bourbon.

Je remercie sincèrement mes parents, pour m'avoir toujours soutenu dans mes choix et dans mes études. Si j'en suis là aujourd'hui, c'est grâce à vous et à vos encouragements. Sans oublier ma sœur, j'espère que vous serez fier de moi.

Merci à Nathalie et Xavier, notamment pour ces séjours au ski.

Merci à vous tous d'être venue nous voir, d'avoir quitté le soleil nantais ou parisien pour profiter de la pluie niçoise.

Enfin, mes derniers mots iront à Julia, ma femme. Merci pour m'avoir suivi en Norvège puis à Nice puis ... Je te suis reconnaissant pour ton écoute, ainsi que pour ton soutien quotidien qui m'a largement aidé à me consacrer à ma thèse.

Table des matières

Résumé	iii
Remerciements	v
Table des figures	xi
Liste des tableaux	xiii
Liste des Abréviations	xv
1 Introduction	1
1.1 Le goût	2
1.1.1 Sens chimiques	2
1.1.2 Une histoire de goût	4
1.2 De la molécule sapide à l'intégration du signal	6
1.2.1 Biologie du système gustatif	6
1.2.2 Chimie des 5 saveurs	9
1.3 Focus sur la perception sucrée	15
1.3.1 Hédonicité et enjeux sociétaux	15
1.3.2 Molécules sucrées	17
1.3.3 Récepteur	25
1.4 Conclusion	30
2 De la structure chimique au pouvoir sucrant	31
2.1 Préambule : à la poursuite de l'édulcorant « parfait »	32
2.2 Base de données de molécules « sucrées »	33
2.2.1 316 composés et leur pouvoir sucrant	33
2.2.2 Structures chimiques	34
2.2.3 Diversité structurale	35
2.3 Approche statistique « ligand-based »	37
2.3.1 Méthodologie pour l'étude de la relation structure chimique - pouvoir sucrant	37
2.3.2 Construction du modèle de prédiction	40
2.3.3 Application : prédiction du pouvoir sucrant de composés naturels	45
2.4 Conclusion	51
3 Structure du récepteur T1R2/T1R3	53
3.1 Préambule : une échelle atomique	53
3.2 Des atomes à la « protéine »	54
3.3 Données disponibles pour la modélisation du récepteur T1R2-T1R3	57

3.3.1	La PDB	58
3.3.2	Les structures 3D de RCPG	58
3.3.3	Données de mutagenèse dirigée	59
3.4	Construction des modèles de T1R2 et T1R3	59
3.5	Description moléculaire de T1R2 et T1R3	66
3.5.1	VFD	66
3.5.2	CRD	67
3.5.3	TMD	68
3.6	Conclusion	74
4	Relation structure-fonction de T1R2-T1R3	77
4.1	Préambule : Stratégie d'étude par domaine	77
4.2	Préparation des systèmes moléculaires	78
4.2.1	Paramétrisation des dynamiques moléculaires	79
4.2.2	Protocole des simulations de dynamique moléculaire	82
4.3	Domaine extracellulaire	84
4.3.1	Préparation des systèmes	84
4.3.2	Stabilité des systèmes	85
4.3.3	Interactions ligand-récepteur	86
4.3.4	Décomposition de l'énergie d'interaction par résidu	88
4.3.5	Analyses des sites de liaison des disaccharides et dipeptides	95
4.3.6	Conclusion	96
4.4	Domaine transmembranaire	97
4.4.1	Préparation des systèmes	97
4.4.2	Entrée du cyclamate dans le TMD de T1R3	98
4.4.3	Étude de la stabilité des DM	99
4.4.4	Résidus clés du site d'interaction	100
4.4.5	Conclusion	102
5	Première étape de la perception, l'activation du récepteur	107
5.1	Préambule : vers une compréhension de l'activation du récepteur.	107
5.2	Hypothèse du mode d'activation pour les RCPG de classe C	109
5.2.1	Du VFD à la protéine G, hypothèse pour le mGluR	109
5.2.2	Cinétique d'activation	111
5.2.3	La partie transmembranaire du mGluR5, un prototype de RCPG de classe C	111
5.3	Méthodologie des simulations de dynamique moléculaire du mGluR5	113
5.3.1	Structure cristallographique	114
5.3.2	Préparation du mGluR5 pour les simulations de dynamique moléculaire	115
5.3.3	Paramétrisation des dynamiques moléculaires	115
5.3.4	Protocole des simulations de dynamique moléculaire	116
5.4	La partie transmembranaire du mGluR5	116
5.4.1	Étude de la stabilité des simulations de dynamique moléculaire	116
5.4.2	Analyse en composantes principales	117
5.4.3	Analyse des interrupteurs moléculaires	123
5.4.4	Schéma de l'hypothèse d'activation du verrou ionique	129

5.5 Conclusion	130
6 Conclusion générale	133
A Données supplémentaires	139
B Méthodes et principes	145
B.1 Les méthodes d'apprentissage	145
B.1.1 Forêts aléatoires	145
B.1.2 Séparateurs à vaste marge	146
B.2 Modélisation par homologie	148
B.2.1 Alignement de séquences en AA	148
B.2.2 Modélisation du récepteur	150
B.2.3 Validation du modèle	150
B.3 <i>Docking</i> moléculaire	151
B.4 Méthodes et principes de la dynamique moléculaire	151
B.4.1 Un modèle comme objet d'étude	151
B.4.2 Champ de force AMBER	152
B.4.3 Les conditions aux limites périodiques	153
B.4.4 Image minimum et rayon de troncature	154
B.4.5 Minimisation de l'énergie potentielle	155
B.4.6 Principe de la dynamique moléculaire	158
B.4.7 Ensemble thermodynamique	162
B.5 Estimation de l'énergie d'interaction	163
Bibliographie	165

Table des figures

1.1	Chronologie	4
1.2	Anatomie du système gustatif	7
1.3	Schéma représentant les cinq familles de récepteurs gustatifs.	10
1.4	Représentation de Fischer des aldoses et cétooses C3 à C6.	20
1.5	Réaction d'hémi-acétalisation et liaison osidique.	21
1.6	Structures chimiques d'édulcorants sur le marché.	24
1.7	Arbre phylogénique des RCPG	27
1.8	Schéma du dimère T1R2-T1R3	29
2.1	Diversité structurales des composés de la SweetenersDB	35
2.2	Propriétés physico-chimiques des composés de la SweetenersDB.	38
2.3	Schéma du protocole de sélection de nouveaux composés sucrés.	41
2.4	Corrélations mesurées pour l'optimisation des RF en brute force.	44
2.5	Performances des modèles de prédictions.	46
2.6	Espace chimique de la SuperNatural II	49
2.7	Histogramme des logPS prédits sur les composés de la SuperNatural II	52
2.8	Proportions des familles chimiques mises en évidence par le modèle QSAR à partir la base de données de composés naturels.	52
3.1	Diagramme de Venn qui regroupe les AA par leurs propriétés physico-chimiques	55
3.2	Structure secondaire	56
3.3	Représentation de structures cristallographiques du mGluR.	60
3.4	Alignement des séquences en AA des VFD des RCPG de classe C	62
3.5	Alignement des séquences en AA des CRD des RCPG de classe C	63
3.6	Alignement des séquences en AA des TMD des RCPG de classe C	65
3.7	Modèles par homologie des VFD de T1R2 et T1R3	67
3.8	Représentation schématiques des AA importants du TS sur les modèles de T1R2 et T1R3	70
3.9	Modèles par homologie du TMD du STR	72
3.10	Comparaison des modèles par homologie de T1R3 selon les alignements de séquence en AA et le patron utilisé.	74
4.1	Ion Mg ²⁺ présent dans la structure cristallographique du mGluR1 (1EWK)	80
4.2	Vérification de la convergence des étapes de pré-production.	85
4.3	Convergence du RMSD du VFD de T1R2 et des ligands lors de l'étape de production.	87
4.4	Décomposition par résidus des énergies d'interaction du VFD de T1R2 avec les disaccharides.	90

4.5	Décomposition par résidus des énergies d'interaction du VFD de T1R2 avec les dipeptides.	91
4.6	Représentation des sites d'interaction ligand-récepteur du VFD de T1R2 pour chaque cluster.	93
4.7	Représentation de la densité des molécules d'eau (en bleu) au cours des trajectoires.	94
4.8	Représentation en deux dimensions des sites d'interaction du VFD de T1R2 pour les disaccharides et les dipeptides.	95
4.9	Représentation en trois dimensions des sites d'interaction du VFD de T1R2 pour les disaccharides et les dipeptides.	96
4.10	Les deux sites d'interaction du TMD de T1R3 avec le cyclamate estimés par <i>docking</i>	98
4.11	Cyclamate entrant dans la cavité haute par <i>brute force</i>	99
4.12	RMSD du <i>backbone</i> du TMD de T1R3 (bleu) et des atomes lourds du cyclamate (orange).	100
4.13	Contributions énergétiques des AA des cavités haute et basse pour le cyclamate dans le TMD de T1R3.	103
4.14	Sites d'interaction du cyclamate dans le TMD de T1R3.	105
5.1	Schéma représentant l'hypothèse des réarrangements structuraux lors de l'activation des RCPG de classe C de J.P. Pin et P. Rondard [137].	110
5.2	Hypothèse du mécanisme et de la cinétique d'activation des RCPG de classe C.	112
5.3	Ligands utilisés avec le mGluR5.	114
	(a) 33-DFBA	114
	(b) MPEP	114
5.4	Séquences en AA du TMD du mGluR5.	115
5.5	RMSD du mGluR5 et des ligands 33-DFBA et MPEP lors des simulations de dynamique moléculaire.	118
5.6	Analyse en composantes principales réalisée sur l'ensemble des simulations de dynamique moléculaire.	121
5.7	Informations extraites de la première composante principale (CP1).	122
5.8	Distances minimales mesurées entre les atomes des AA du verrou ionique.	124
5.9	Distances mesurées entre les atomes $C\alpha$ des AA du verrou ionique.	125
5.10	Résumé des distances minimales entre les AA du verrou ionique selon chaque simulation de dynamique moléculaire.	127
5.11	Résumé des distances entre les atomes $C\alpha$ des AA du verrou ionique selon chaque simulation de dynamique moléculaire.	128
5.12	Schéma de l'activation du verrou ionique pour le mGluR5.	129
B.1	Représentation schématique des méthodes de prédictions RF et SVM.	147
B.2	Représentation des composantes d'un champ de force.	153
B.3	Représentation tridimensionnelle du principe des conditions aux limites périodiques.	154
B.4	Schéma de l'image minimum et de la sphère de troncature dans l'utilisation des conditions aux limites périodiques.	155
B.5	Algorithme de dynamique moléculaire	159

Liste des tableaux

2.1	Partition des données pour le modèle de classification « sucré » / « amer »	50
2.2	Tableau de contingence	50
(a)	Théorique	50
(b)	Apprentissage	50
(c)	Validation externe	50
3.1	Liste des structures expérimentales des RCPG de classe C.	59
3.2	Liste des AA importants du STR : AA du site de liaison du VFD et TMD ainsi que les AA formant les interrupteurs moléculaires.	75
4.1	Résumé de l'utilisation des champs de force et des paramètres des ligands.	81
4.2	Énergie d'interaction des ligands avec le VFD de T1R2	88
4.3	Résumé des données de mutagenèse dirigée pour le VFD de T1R2	92
4.4	Résumé des données de mutagenèse dirigée pour le TMD de T1R3 avec le cyclamate	104
5.1	Distances entre les AA des TM3, TM6 et TM7.	120
A.1	Base de données des molécules dont le pouvoir sucrant est connu.	140
A.2	Données de mutagenèse dirigée	142
A.3	Pourcentage d'identité de séquences en acide-aminés	144

Liste des Abréviations

33-DFBA	3,3'-DiFluoroBenzAldazine
7tm	7 hélices transmembranaires
ACP	Analyse en Composante Principale
ADME-Tox	Absorption, Distribution, Métabolisme et Élimination - Toxicité
BitterDB	« Bitter DataBase » ou Base de Donnée de molécules amères
CP	Composante Principale
CRD	« Cystein Rich Domain » ou Domaine Riche en Cystéine
D-Trp	D-Tryptophane
EC50	Concentration Efficace médiane
ECL	« ExtraCellular Loop » ou Boucle Extra-Cellulaire
GDP/GTP	Guanosine Di ou Tri Phosphate
GMS	Glutamate MonoSodique
HB	« Hydrophobic Barrier » ou barrière hydrophobe
ICL	« IntraCellular Loop » ou Boucle Extra-Cellulaire
IL	« Ionic Lock » ou verrou ionique
logP	coefficient de partage octanol/eau
logPS	logarithme du Pouvoir Sucrant
Mcc	Coefficient de corrélation de Matthews
mGluR	Récepteur Métabotrope au Glutamate
MMGBSA	« Molecular Mechanics energies combined with the Generalized Born and Surface Area continuum solvation »
MPEP	2-Methyl-6-(PhenylEthynyl)-Pyridine hydrochloride
NHDC	Néohesperidine DiHydroChalcone
NPT	Nombre d'atome, Volume et Température constants
NVT	Nombre d'atome, Pression et Température constants
PDB	« Protein Data Bank »
PKD2L1	« Polycystic Kidney Disease 2-Like 1 protein »
PS	Pouvoir Sucrant
QSAR	« Quantitative Structure-Activity Relationship »
R/A(o-f)	Repos/ Actif(ouvert-fermé)
R²	Coefficient de détermination
RCPG	Récepteur couplé aux Protéines G
RF	« Random Forest »
RMSD	« Root Mean Square Deviation »
STR	« Sweet Taste Receptor »
SuperNatural II	Base de Donnée de molécules naturelles
SuperSweet	Base de Donnée de molécules apparentées à des sucres
SVM	« Support Vector Machine » ou machine à vecteur support
SweetenersDB	Base de Donnée de molécules sucrées à pouvoir sucrant connu
T1R[123]	« Taste Receptor type 1 member [1, 2, 3] »
T2R	« Taste Receptor type 2 »
Tc	Coefficient de Tanimoto
TM1 - 7	Hélice transmembranaire 1 à 7
TMD	« TransMembrane Domain » ou domaine transmembranaire
TS	« Transmission Switch » ou commutateur de transmission
VFD	« Venus Flytrap Domain »

Chapitre 1

Introduction

« Les cinq sens donnés par la nature sont la seule mesure du réel, notre seule possibilité de comprendre et d'interpréter le monde. »

Bessa Vugo

Ce chapitre est basé sur la publication suivante :

Chéron, J.-B., Golebiowski, J., Briand, L., Antonczak, S., Fiorucci, S. (2017).
Ces molécules qui éveillent nos papilles. Actualité Chimique.

1.1 Le goût

1.1.1 Sens chimiques

- **Cinq sens pour discerner le monde**

L'homme possède cinq sens, la vue, l'ouïe, le toucher, l'odorat et le goût. Ils se divisent en deux catégories selon l'origine du stimulus. Les trois premiers sont des sens que l'on peut qualifier de « physiques ». En effet, la vue permet l'acquisition d'un signal électromagnétique, l'ouïe, des vibrations mécaniques - ondes sonores - et le toucher, des variations de pression. L'odorat et le goût, quant à eux, sont des sens dits « chimiques ». La perception associée à ces deux sens provient de l'interaction d'une ou plusieurs molécules avec les récepteurs situés au niveau de l'organe sensoriel dédié, le nez ou la langue. Si ces cinq sens sont les plus couramment cités, l'homme possède d'autres sens dits « inconscients ». La proprioception permet de situer ses membres dans l'espace, l'équilibration informe du déplacement du corps *via* l'oreille interne, la nociception avertit d'une douleur et enfin la thermoception renseigne sur l'absence ou la présence de chaleur.

- **La complexité des sens chimiques**

Les sens « physiques » ont été les plus étudiés. En effet, la compréhension des phénomènes physiques tel que les ondes sonores, les vibrations électromagnétiques et la pression à l'origine des perceptions sensorielles y a largement contribué. Au contraire, l'étude de l'odorat et du goût ont longtemps été mis de côté, notamment à cause de la difficulté à caractériser l'origine moléculaire de ces sens. Cette disparité est perceptible d'un point de vue étymologique. La description d'une saveur ou d'une odeur se rapporte souvent à un objet du quotidien. Il apparaît plus difficile de verbaliser ce que l'on goûte ou ce que l'on sent que ce qui est vu ou entendu. Un autre exemple, nommer les pathologies associées aux sens physiques semble évident (surdité, cécité, analgésie),

mais il est moins intuitif de citer celles liées aux sens chimiques telle que l'anosmie ou l'ageusie.

• Le goût, du sapide à l'insipide

Le goût est un sens complexe et l'utilisation à mauvais escient de ce terme l'illustre parfaitement. L'emploi courant et erroné de l'expression « goût d'un aliment » est lié à une perception multi-sensorielle comprenant les saveurs (détection des molécules sapes par la langue) à laquelle s'ajoutent des sensations olfactives (rétronasale) et somesthésiques (thermiques, tactiles et proprioceptives) [1–5]. En réalité, l'ensemble des sensations gustatives, olfactives et somesthésiques sont désignées par le terme de « flaveur » (issu de l'anglais *flavour* et du latin *flare*) [6]. La saveur et l'arôme en sont respectivement les composantes gustatives et olfactives. Quant à la sensibilité somesthésique, elle permet d'évaluer la texture et la température d'un aliment ou encore de ressentir le pétillant, le piquant, l'astringent... La définition du goût, au sens strict, correspond à la perception de molécules sapes par l'organe sensoriel dédié, la langue.

Cinq saveurs sont aujourd'hui admises : acide, amère, salée, sucrée et umami. L'umami est la plus récente et est associée à la perception de certains acides aminés (AA). Elle est caractéristique de la sauce soja ou de certains fromages comme le parmesan.

Avec seulement cinq saveurs, il est difficile de caractériser - ou de verbaliser - ce que l'on peut ressentir lorsque la saveur perçue est inqualifiable. L'eau, par exemple, peut être considérée comme « insipide » mais une préférence pour certaines eaux est observée [7, 8] et des bars à eaux se développent pour proposer de telles dégustations. Contrairement à l'olfaction, pour lequel le prix Nobel de physiologie et de médecine a été décerné à Richard Axel et Linda Buck en 2004 pour leurs découvertes des gènes codant pour les récepteurs olfactifs et l'organisation du système olfactif [9], l'ensemble des récepteurs associés à la perception gustative n'est pas encore connu. La découverte de nouveaux récepteurs gustatifs, possédant une spécificité et une affinité pour

une nouvelle classe de molécules permettra de définir de nouvelles saveurs, enrichissant de fait le vocabulaire associé à ce sens. À celles déjà connues, on peut citer notamment les saveurs nouvellement proposées : « gras » (oléogustus) ou « kokumi » permettant de détecter, respectivement, les acides gras [10] et le calcium [11].

1.1.2 Une histoire de goût

• À la découverte de la sémantique

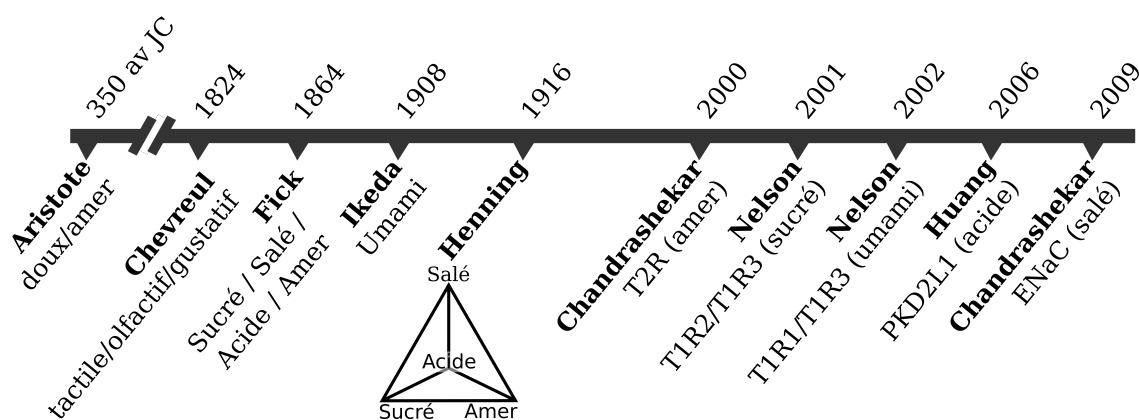


FIGURE 1.1 – Chronologie

La définition du goût a largement évolué depuis l'antiquité (figure 1.1). Dans l'œuvre « De l'âme », Aristote distingue deux saveurs principales et opposées, le doux et l'amer et considère l'onctueux, le salé, l'aigre, l'âpre, l'astringent et l'acide comme des sensations dérivant de ces deux saveurs élémentaires. En 1751, Carl von Linné différenciait dix qualités gustatives « l'humide, le sec, l'acide, l'amer, le gras, l'astringent, le sucré, l'aigre, le muqueux et le salé ». En 1824, Michel-Eugène Chevreul distinguait les dimensions tactiles, olfactives et gustatives. Ce n'est qu'en 1864 que quatre saveurs « primaires » ou « fondamentales » : le sucré, le salé, l'acide et l'amer sont associées à des récepteurs gustatifs par A. Fick. La saveur umami, mot d'origine japonaise signifiant savoureux, est définie comme cinquième saveur primaire par K. Ikeda en 1908.

Le nombre de saveurs fondamentales est toujours sujet à débat. Aujourd'hui, une saveur est reconnue s'il existe un « récepteur gustatif » associé. Ils ont été mis en évidence à partir des années 2000 (figure 1.1). Le dogme du caractère additif des saveurs « primaires » (par analogie aux couleurs) a longtemps persisté. Cependant, les sens chimiques sont des sens complexes et certaines molécules possèdent une double modalité (la saccharine est par exemple sucrée et amère). Cette subtilité ne permet pas de catégoriser strictement les saveurs et les connaissances actuelles convergeraient vers une redéfinition du goût d'un aliment à l'aide d'un continuum de qualités gustatives.

- **Aujourd'hui, place à la gastronomie et aux saveurs**

L'odeur ou l'aspect d'aliments peuvent s'avérer trompeurs et leur ingestion peut être bénéfique ou se révéler dangereuse. La détection de la qualité et/ou de la quantité de molécules sapides devient alors essentielle pour assurer le bon fonctionnement de l'organisme, voire sa survie. Le sens du goût permet de nous renseigner sur le caractère comestible et nutritif d'un aliment. Les cinq saveurs aujourd'hui reconnues (saveurs primaires) peuvent être associées à des catégories de nutriments. La perception de la saveur sucrée informe sur l'apport énergétique des nutriments, la saveur salée sur la présence de minéraux indispensables au fonctionnement de notre organisme tandis que l'amertume renseigne sur la présence de poisons potentiels. L'acidité est quant à elle un marqueur de maturité des fruits ou de la présence de microorganismes potentiellement dangereux pour l'organisme dans nos aliments [12]. Enfin, il est supposé que l'umami renseigne sur la présence d'AA, important pour notre organisme.

Il est rare aujourd'hui d'utiliser le sens du goût pour sélectionner de la nourriture comestible. Lié à cette fonction primitive, il faut évoquer le mécanisme dit d'alliesthésie. Ce terme décrit la relation entre la valence (plaisir ou déplaisir) d'un stimulus et son effet sur l'organisme. Un stimulus bénéfique pour l'organisme sera généralement associé à une valence positive. Les saveurs sucrée et salée en sont les représentants, leur consommation est perçue comme favorable par notre cerveau et peut conduire à

des excès. À l'inverse la saveur amère - associée aux poisons - engendre un mécanisme de dégoût qui peut être détourné par accoutumance.

L'impact des saveurs, des arômes et la texture des aliments sur le comportement alimentaire sont bien connus. Le plaisir de se nourrir apparaît comme crucial dans nos sociétés. À tel point que l'ajustement de chaque ingrédient est important et permet d'influencer le consommateur, allant jusqu'à préparer les plats en laboratoire [13, 14]. L'art culinaire et la chimie se rejoignent avec la gastronomie moléculaire, discipline apparue dans les années 1980. S'il paraît facile de déterminer la valence hédonique lors de la consommation d'un aliment, les mécanismes moléculaires associés à la perception des saveurs restent encore méconnus.

1.2 De la molécule sapide à l'intégration du signal

1.2.1 Biologie du système gustatif

La surface de la langue possède des papilles gustatives qui varient en forme et localisation. Une dizaine de papilles caliciformes forment un « V » sur la partie postérieure de la langue, une dizaine de papilles foliées sont localisées sur les parties latérales et environ 10 000 papilles fongiformes sont réparties sur la langue avec une plus forte densité sur la partie antérieure [15] (figure 1.2). Ces papilles contiennent des bourgeons gustatifs composés de plusieurs strates cellulaires en « forme d'oignons ». Les plus externes sont les cellules de soutien, elles forment la paroi du bourgeon et permettent de délimiter un pore pour les échanges avec le milieu buccal. À l'intérieur, les cellules sensorielles ont une extrémité apicale avec des microvillosités permettant d'augmenter leur surface de contact. Elles reposent sur une couche de cellules dites « basales » non différenciées permettant de régénérer les cellules sensorielles, soumises à un stress permanent, environ tous les 10 jours [16]. La cellule sensorielle contient les récepteurs gustatifs. L'identification de ces récepteurs dans les années 2000 (figure 1.1) a largement contribué à une meilleure compréhension des premières étapes moléculaires de

la perception gustative [17].

Les premiers mis en évidence sont les récepteurs aux saveurs amère, sucrée et umami au début des années 2000 [18–20] ; puis les récepteurs aux saveurs acide et salé [21, 22]. On distingue deux types de chimiorécepteurs : les récepteurs « canaux » qui permettent l'entrée d'ions dans la cellule et les récepteurs métabotropes (ou couplés aux protéines G - RCPG) qui transmettent le signal sans autoriser l'entrée des molécules sapides dans la cellule (figure 1.3). Selon la nature du récepteur, des voies de signalisation différentes sont sollicitées pour mener à l'activation du neurone afférent (figure 1.2).

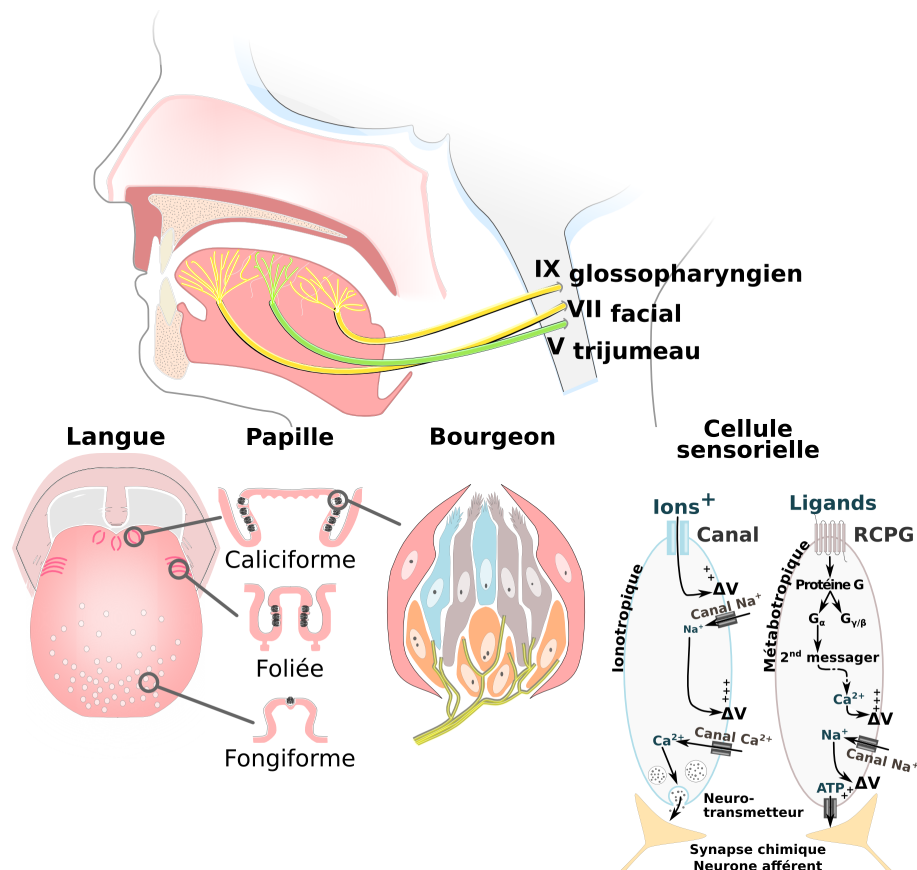


FIGURE 1.2 – Anatomie du système gustatif

Les récepteurs « canaux » permettent la perception des saveurs salée et acide.

Pour la saveur salée, l'entrée d'ions alcalins se fait par les canaux ENaC (*Epithelial Sodium Channel*) [21], principalement sélectifs au sodium, tandis que le canal à proton

PKD2L1 (*Polycystic Kidney Disease 2-Like 1 protein*) [22] est impliqué dans la perception de la saveur acide. Toute cellule a un potentiel membranaire, dit de « repos » dû à la différence de concentration en ions de part et d'autre de la membrane plasmique. La particularité des cellules sensorielles (excitables) est de pouvoir faire varier ce potentiel membranaire pour transférer un signal électrique. La diffusion des ions à travers les canaux, modifiant alors la polarité de la cellule sensorielle, est à l'origine du message électrique qui sera transmis au neurone afférent et interprété par le cerveau. Les ions Na^+ ou H^+ entrent dans la cellule et entraînent une accumulation de charge positive à la face interne de la membrane plasmique. Ce transfert de charge conduit à l'activation de canaux sensibles à la dépolarisation de la membrane, le signal est alors amplifié par une entrée massive d'ions Na^+ dans la cellule. Le signal se propage jusqu'à la partie basale où la synapse est présente. Des canaux calciques sensibles aux variations de potentiel membranaire vont alors s'activer et faire entrer du Ca^{2+} dans la cellule [23]. Les protéines calcium-dépendantes du complexe SNAR permettent l'exocytose du neurotransmetteur et activent le neurone post-synaptique (figure 1.2).

Les saveurs amère, sucrée et umami sont dépendantes de l'activation de RCPG.

Ces récepteurs partagent tous un motif structural caractéristique de sept hélices transmembranaires capable de transférer une information chimique externe à l'intérieur de la cellule. Les RCPG subissent un réarrangement structural suite à leur interaction avec une molécule sapide. Ce signal conduit à l'activation d'une protéine G hétérotrimérique (notamment la G-gustducine [24]), couplée au récepteur, par l'échange d'une molécule de GDP (Guanosine diphosphate) en GTP (Guanosine triphosphate). Le complexe de protéine G se scinde alors en $G\alpha$ et $G\beta/\gamma$. Cette dissociation entraîne l'activation de multiples effecteurs ayant pour but l'activation de canaux sodiques voltage-dépendants [23]. Au niveau de la synapse, l'entrée d'ions Na^+ et une dépolarisation suffisante permettent l'activation des canaux voltage-dépendants perméants à l'ATP ayant le rôle de neurotransmetteur pour activer le neurone afférent (figure 1.2).

Les neurones afférents permettent la transmission de l'information par les paires de nerfs crâniens VII (nerf facial) et IX (nerf glossopharyngien) vers le bulbe rachidien, le thalamus et enfin vers les centres du goût formés par le cortex gustatif et l'insula pour être traitée puis analysée de manière consciente et inconsciente [25, 26].

Les récepteurs de la saveur umami et sucrée sont des hétérodimères de RCPG de type T1R (*Taste Receptor type 1*). T1R3 (*T1R member 3*) est une sous-unité commune aux deux récepteurs. Le récepteur à l'umami est constitué des sous-unités T1R1 (*T1R member 1*) et T1R3, le récepteur à la saveur sucrée (nommé STR par la suite, pour *Sweet Taste Receptor*) de T1R2 (*T1R member 2*) et T1R3. La présence de multiples sites de liaison sur ces récepteurs leur permet d'interagir avec des structures chimiques très variées. Les récepteurs métabotropes, plus particulièrement le STR, seront plus largement détaillés dans le chapitre 1.3.3. Enfin, il existe une famille de gènes (T2R ou *Taste Receptor type 2*) codant pour les récepteurs de la saveur amère. L'Homme compte environ 25 gènes fonctionnels et à titre de comparaison, la souris en dispose de 33 [27]. Un récepteur peut reconnaître plusieurs types de molécules amères et une molécule amère peut être reconnue par plusieurs récepteurs. La combinatoire d'association est extrêmement importante et varie en fonction des espèces. Cette combinatoire explique la grande diversité structurale de molécules sapides pouvant être détectées par notre système gustatif (chapitre 2.2.3 p 35). Des facteurs externes peuvent moduler le signal et ajouter de la subtilité au système de détection des saveurs. La composition chimique de la salive joue notamment un rôle primordial [28].

1.2.2 Chimie des 5 saveurs

• Salé

D'un point de vue chimique, les sels sont des composés neutres formés par l'association de cations et d'anions. La saveur salée fait penser immédiatement au sel de table ou chlorure de sodium, NaCl. Toutefois, d'autres halogénures d'alcalins possèdent un

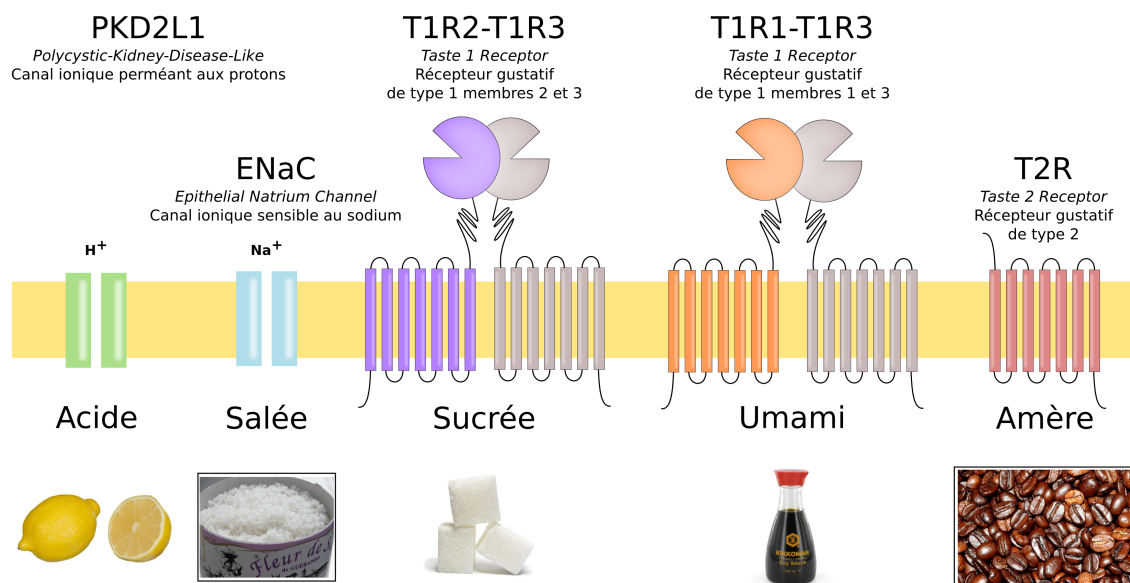


FIGURE 1.3 – Schéma représentant les cinq familles de récepteurs gustatifs.

Les saveurs acide et salée impliquent des canaux ioniques alors que les saveurs sucrée, umami et amère sont dépendantes de l'activation de récepteurs couplés à une protéine G (RCPG).

goût salé [29]. Ils sont cependant entachés d'un arrière-goût souvent amer qui augmente avec la concentration en sel. L'intensité de la saveur salée est non seulement dépendante de la nature du cation mais également du contre-ion. En effet, ce dernier joue un rôle sur la solubilité du cation et donc sur sa disponibilité dans la salive [30], mais peut aussi interagir avec les voies de transduction du signal dans la cellule sensorielle et moduler la réponse au stimulus chimique. D'autres composés comme NH_4Cl ou $CaCl_2$ peuvent être perçus salés [31]. Cependant le consensus général est que leur goût est complexe, dans le sens où ils possèdent un arrière-goût amer, parfois acide, en fonction de leur concentration. Le seuil de perception du chlorure de sodium (molécule salée de référence) est de $10^{-2} \text{ mol.L}^{-1}$. Dans le cas de $NaCl$, le seuil de perception est relativement élevé, car la présence d'ions dans les nutriments est cruciale à notre organisme afin de réguler les équilibres ioniques (homéostasie interne). Le sel est associé à une valence positive à faible concentration et joue le rôle d'exhausteur de goût. À plus haute concentration en revanche, le mécanisme d'alliesthésie vu précédemment

(chapitre 1.1.2) s'inverse et le sel devient aversif. Ce mécanisme de défense mobilise les cellules sensorielles de l'amer et de l'acide, qui sont, elles, naturellement associées à des réponses aversives [32]. C'est au travers de cette double stratégie de reconnaissance que nous régulons la quantité de sel ingérée à chaque repas. Il n'existe pas, à ce jour, d'alternative au sel NaCl qui possèderait le même profil sensoriel. Pourtant, l'omniprésence de cet exhausteur de goût (un BigMac : 1010 mg, un plat préparé de nouille : 4520 mg) [33] inquiète les organismes de santé publique puisqu'il est à l'origine de nombreuses pathologies cardio-vasculaires [34].

• Acide

La saveur acide est liée à la détection de protons (H^+). Elle est liée à la définition de Brønsted-Lowry de la fonction acide, groupement susceptible de libérer un proton. Le nombre de molécules appartenant à l'espace chimique de la saveur acide est difficile à déterminer selon cette définition. Pour un composé donné, l'acidité perçue correspond au rapport des seuils de perception entre ce composé et l'acide chlorhydrique, choisi comme référence. Le seuil de détection de ce dernier est de l'ordre de $10^{-4} \text{ mol.L}^{-1}$. De manière intuitive, on imagine pouvoir relier l'acidité perçue d'un composé à son pKa. En réalité aucune corrélation ne peut être établie, confirmant notamment que des facteurs supramoléculaires entrent en jeu. Il a été montré que d'autres protéines réagissent à l'acidification du milieu intracellulaire suite à la diffusion passive d'acide faible dans la cellule et modulent le seuil d'excitabilité membranaire [35]. Cette complexité biologique et la multitude de protagonistes impliqués fait de l'acidité la saveur la moins bien comprise. Au-delà de cette complexité moléculaire, cette modalité est aussi fortement dépendante de la grande variabilité inter-individuelle dans les populations. Les facteurs tels que la quantité de salive produite, son pouvoir tampon à travers la concentration en anhydrase carbonique ou son pH [36] font varier l'intensité perçue d'un composé acide qui y serait dilué.

• Amer

Les molécules amères présentent une grande diversité structurale, allant de composés très simples comme le cyanure d'hydrogène (HCN) ou le chlorure de potassium (KCl) jusqu'à des composés polycycliques de masse molaire supérieure à 900 g.mol⁻¹. On trouve, par exemple, des composés de la famille des flavonoïdes, des catéchines ainsi que des peptides modifiés [37]. L'amertume mobilise un circuit d'aversion qui lui a valu son nom, tiré du latin *amarus* signifiant « désagréable, pénible, déplaisant ». De nombreux composés toxiques sont en effet fortement amers, tel que la strychnine, un alcaloïde naturel létal à faible concentration. Cependant, le lien entre toxicité et amertume n'est pas évident. En dépit de cette aversion naturelle, certains composés inoffensifs (la caféine du café ou la théobromine du chocolat) sont appréciés pour cette saveur s'ils ne sont pas présents en quantité trop importante [38]. L'analyse du génome de différentes espèces fait plutôt apparaître un lien entre régime alimentaire et capacités perceptives. En d'autres termes, les espèces comme les carnivores, qui peuvent par exemple se permettre de rejeter les plantes de leur régime alimentaire, ont une plus grande sensibilité aux composés amers que les herbivores [39]. Chez l'homme, les seuils de détection de ces molécules sont les plus faibles parmi l'ensemble des molécules sapides. Celui de strychnine est de l'ordre de 10⁻⁷ mol.L⁻¹, celui du benzoate de dénatonium, un des composés les plus amers, atteint 10⁻⁸ mol.L⁻¹. Ce dernier est utilisé dans la composition de pesticides, antigels ou cosmétiques comme agent répulsif dissuadant l'ingestion. L'espace chimique connu des molécules à saveur amère compte environ un millier de composés [37]. La perception de l'amertume mobilise 25 types de récepteurs, soulignant le rôle stratégique qu'elle tient parmi les modalités sensorielles. Cet espace chimique n'est pas intégralement découvert en regard du nombre de récepteurs qui interviennent dans ces liens chemo-génomiques.

• Umami

L'umami, dont le terme provient du japonais signifiant « savoureux » ou « appétissant », est la plus jeune saveur quant à sa description et son approbation par la communauté scientifique. L'espace chimique associé à la perception de la saveur umami est donc relativement restreint à ce jour. Originellement caractérisée chez l'homme par la perception des sels de L-glutamate, et plus particulièrement le glutamate monosodique (GMS), cette saveur est définie au sens large par la perception de quelques AA (L) et certains di- et tri-peptides. Ce n'est que plus tard que deux nucléotides, l'inosine et la guanosine monophosphate (5'-IMP et GMP), ont été découverts pour leur effet synergique avec le glutamate. Le seuil de détection de la saveur umami est 10^{-3} mol.L⁻¹ pour le SMS, valeur relativement élevée par rapport à la perception des molécules amères par exemple. C'est pourquoi cette saveur se trouve dans les plats riches en glutamate ou protides tels que la sauce soja, certains fromages comme le parmesan, les viandes, les poissons et crustacés.

L'acide L-glutamique est la forme neutre du glutamate naturellement présent dans notre alimentation. Il fait partie des 20 AA naturels et est essentiel au bon fonctionnement de notre organisme. C'est un acide faible (pKa de la chaîne latérale valant ≈ 4.2) et l'intensité de la saveur umami diminue si le glutamate est mis en présence d'un acide plus fort. C'est donc le glutamate et pas l'acide glutamique qui déclenche la saveur umami. On trouve le glutamate libre à l'état naturel dans les aliments protéinés d'origine végétale ou animale (par exemple 1,2 g pour 100 g de parmesan ou de sauce soja). Sa consommation est de l'ordre de 5 à 12 g par jour en Europe [40] sous sa forme naturelle à laquelle il faut ajouter environ 1 g par jour en provenance de plats industriels contenant du glutamate monosodique (GMS). Cette consommation additionnelle dépasse les 4 g par jour dans les pays asiatiques. Le SMS est utilisé comme exhausteur de goût sous forme d'additif alimentaire (E621) dans de nombreuses préparations industrielles (plats préparés, en particulier les plats asiatiques, charcuteries/viandes, biscuits apéritifs, ou encore produits déshydratés). Cette consommation importante de

GMS est au centre d'un intense débat scientifique et sociétal [40, 41]. Le glutamate est un des principaux neurotransmetteurs du système nerveux central et on soupçonne qu'une concentration trop élevée en glutamate au niveau des fentes synaptiques est à l'origine de troubles neurologiques. Il existe chez certaines personnes une réaction d'hypersensibilité appelée « syndrome du restaurant chinois » qui consiste en l'apparition quelques minutes après l'ingestion d'un repas riche en GMS, de sensations de brûlures, de pressions faciales, des maux de tête, des nausées, ainsi que des douleurs thoraciques [42]. Certaines études ont cependant dénoncé le manque de rigueur dans les protocoles expérimentaux et l'absence de données cliniques pour tirer de telles conclusions [43, 44]. Par ailleurs, lorsque la dose journalière est inférieure à 16 g par kg, le glutamate est considéré sans risque pour la santé par les organismes de santé publique. Le consensus actuel dans la communauté scientifique est que de nouvelles études doivent être menées pour mieux comprendre les effets du GMS à haute dose sur l'organisme.

• Sucré

En regard des autres saveurs, le seuil de détection de la saveur sucrée est relativement élevé. Pour fonctionner, un organisme a tout intérêt à reconnaître les nutriments à haute teneur en sucre et donc à fort apport calorique. On la retrouve naturellement la saveur sucrée dans des aliments comme les fruits ou le miel et elle est fortement liée à une valence positive. La famille chimique des sucres (chapitre 1.3.2 p 18) n'est cependant pas la seule à provoquer une saveur sucrée [45]. Les molécules qui ne sont pas des sucres mais ayant un goût sucré sont appelés des édulcorants. Leur avantage est d'être peu - ou pas - calorique et explique leur succès commercial. Ils peuvent être d'origine naturelle ou de synthèse. Parmi les composés naturels on retrouve les très grandes familles des terpènes, ainsi que des phénols et, plus étonnant, certains AA. La majorité des phénols à saveur sucrée sont des polyphénols dérivés de flavonoïdes glycosylés. Parmi les terpènes, on distingue des composés aglycones monoterpéniques,

sesquiterpéniques et surtout les composés diterpéniques dérivés du noyau du stéviol et triterpéniques de type saponine. L'engouement pour le stéviolside, dérivé glycosylé du stéviol, présent dans la plante *Stevia rebaudiana* en fait un exemple de choix. Parmi les édulcorants de synthèse, on n'oubliera pas de citer les plus utilisés par l'industrie agroalimentaire ces dernières décennies comme la saccharine, le cyclamate de sodium ou l'acésulfame de potassium. Leurs découvertes ont, pour bon nombre d'entre eux, été le fruit du hasard. C'est pourquoi leurs structures chimiques sont très variées et sont différentes des édulcorants naturels, quelques exemples sont donnés dans le chapitre 1.3.2. La saveur sucrée et plus particulièrement la caractérisation structurale des sucres et des édulcorants est un des sujets de cette thèse et sera plus largement décrit par la suite.

1.3 Focus sur la perception sucrée

1.3.1 Hédonicité et enjeux sociétaux

- **Appétence de la saveur sucrée, un caractère inné**

Les sucres sont indispensables pour le bon développement de l'organisme et cette saveur a donc une forte valence hédonique positive [46]. L'appréciation de la saveur sucrée est considérée comme innée, à l'inverse de la saveur amère qui engendre un phénomène « réflexe » d'aversion. Des études faites chez le nourrisson montrent que dès la naissance il peut distinguer différentes concentrations de boissons sucrées. Une réponse physiologique globale est également observée avec une augmentation du rythme cardiaque, une décontraction du visage et un sourire ressemblant à un sentiment de plénitude [47]. De plus, chez le nourrisson, le sucre permet de réduire la douleur. De l'eau glucosée à 30% est utilisée pour provoquer un état « analgésique » et pratiquer des actes médicaux [48]. Même si la valence hédonique positive est universelle, l'appréciation quantitative est dépendante des individus et de l'âge. Il est supposé que les

enfants ont un seuil de détection plus élevé et une préférence accrue pour les besoins de leur croissance. Les raisons et les mécanismes sont toujours méconnus.

• Une envie insatiable de sucre

La valence hédonique positive prend tout son sens d'un point de vue génétique et évolutif, elle a permis la sélection et la consommation de denrées alimentaires caloriques nécessaires à la survie. De nos jours, l'accès facilité à des composés sucrés a détourné cette fonction primitive et a conduit à une consommation excessive et chronique de sucre. Il a été montré que la consommation de sucre tel que le saccharose permet l'activation du circuit de la récompense, aussi emprunté par les drogues [49, 50]. Un comportement d'addiction a même été énoncé et reste sujet à débat [51]. Les comportements et les conséquences associés à la consommation de sucre sont même comparés au tabac ou à l'alcool [52].

Il est intéressant de distinguer les effets des sucres et édulcorants sur les systèmes hédonique et homéostatique [53]. Le système homéostatique comprend des détecteurs qui permettent de définir les besoins énergétiques afin de mieux réguler les apports en sucre. Le système hédonique, comme énoncé plus haut, est associé au plaisir et à la récompense. Le saccharose active ces deux systèmes, il engendre à la fois un sentiment de récompense et de satiété. Cependant, des effets différents sont observés avec les édulcorants [53]. D'une manière générale, ils engendrent une sensation de plaisir, mais ils ne sont pas tous aussi efficaces pour diminuer l'appétit. Le cerveau analyse de manière distincte la valence hédonique positive de la saveur sucrée et celle de l'apport énergétique.

• L'alimentation et la surconsommation : un enjeu de santé publique

Le saccharose est largement utilisé dans l'alimentation humaine, notamment sous forme ajoutée. Mais sa consommation en excès dans notre société est un piège. Les rapports de l'OMS du 4 mars 2015 et du 11 octobre 2016 appellent à réduire la consommation de sucres libres (ajoutés) car elle est « *un facteur important de l'augmentation du nombre*

de cas d'obésité et de diabète dans le monde ». Ces sucres libres sont le principal combat des organismes de santé et de scientifiques qui tirent la sonnette d'alarme face à cette surconsommation. Ils dénoncent les pratiques de l'industrie agroalimentaire qui use et abuse de saccharose caché dans les plats préparés, boissons, ...[52] C'est un produit peu coûteux et apprécié des consommateurs pour les raisons énoncées précédemment. Par exemple, une canette de soda sucré contient jusqu'à 10 cuillères à café de saccharose (environ 40 g) alors que l'OMS recommande une consommation journalière de sucres libres de 25 g, soit environ 6 cuillères à café. En 2012, 1,5 millions de décès ont été référencés à cause de diabète et que 2,2 millions de plus sont liés aux conséquences de l'hyperglycémie (maladies cardiovasculaires et l'obésité).

En résumé, la surconsommation de saccharose est un facteur de risque pour des pathologies telles que les caries, le diabète de type II, les maladies cardiovasculaires et l'obésité. La réduction de la quantité de sucre dans l'alimentation est un enjeu de santé publique, à tel point qu'une politique fiscale visant à taxer les produits sucrés a été instaurée en France en 2012 (« taxe soda »). L'identification d'édulcorants à faible apport calorique est donc un axe de recherche stratégique pour satisfaire les consommateurs et les différents acteurs de l'industrie agroalimentaire. Un des objectifs de cette thèse est d'utiliser des données structurales (sucres, édulcorants et le STR) pour mettre en évidence de nouveaux édulcorants. Cependant, les édulcorants n'ont pas les mêmes effets physiologiques que le saccharose, notamment sur le sentiment de satiété. Il faut être conscient qu'au-delà de ces alternatives, la responsabilisation du consommateur est primordiale et nécessaire pour lutter contre les pathologies, comme le montre les nombreux rapports de l'OMS.

1.3.2 Molécules sucrées

• Pouvoir sucrant, une métrique dans la perception de la saveur sucrée

L'intensité de perception des molécules sucrées est appelée « pouvoir sucrant » (PS). Il est important de définir ce terme, qui sera largement utilisé dans ce manuscrit. Le

PS est une mesure semi-quantitative, il est défini comme étant le rapport entre une masse - ou une concentration - de saccharose et d'une substance sucrée à intensité de perception égale. Le saccharose est la molécule de référence pour mesurer le PS d'une molécule, en général sa concentration de référence correspond à sa limite de détection qui est de 10^{-2} mol.L⁻¹. Plus simplement, plus le pouvoir sucrant d'un composé est élevé, moins il faut en utiliser pour obtenir la même perception que du sucre de table. Ainsi, selon la définition, le PS du saccharose vaut 1.

• Les sucres

Les monosaccharides.

Les sucres, aussi appelés glucides, sont des composés chimiques - aldéhyde ou cétone - polyhydroxylés. Ils sont qualifiés par le suffixe « ose ». Il faut distinguer deux grandes familles, les cétooses et les aldoses, possédant de 3 à 6 carbones pour les oses les plus courants. Au-delà de ce nombre, ils sont rencontrés dans les lipopolysaccharides des parois bactériennes. Les oses les plus simples sont l'érythrulose (C₄H₈O₄) et le glycéraldéhyde (C₃H₆O₃) pour chacune des familles, respectivement. Le carbone portant la fonction aldéhyde est noté C₁ et celui portant la fonction cétone C₂. L'orientation du C_{n-1} définit la série D ou L (figure 1.4). Les oses de série D sont prépondérants dans la nature. Les structures peuvent alors se complexifier, selon le nombre de carbones présents dans le squelette (triose, tetrose, pentose, ...). Chacun des carbones porteur d'une fonction hydroxyle est un centre de chiralité. L'ensemble des diastéréoisomères explique la grande diversité des sucres. On peut distinguer les énantiomères étant des sucres images l'un de l'autre dans un miroir et les épimères qui diffèrent par un seul carbone asymétrique (par exemple, sur la figure 1.4, le mannose est un épimère en C2 du glucose).

Les oses linéaires sont en équilibre avec les formes cyclisées qui sont produites par une réaction d'hémi-acétalisation. Le carbone de la fonction cétone ou aldéhyde se lie avec la fonction hydroxyle portée par le carbone n ou n-1 formant ainsi des cycles à

5 (furanose) ou à 6 (pyranose) atomes. L'orientation de la fonction portée par le C₁ (aldose) ou le C₂ (cétose) détermine l'anomère α ou β , respectivement en dessous ou au-dessus du cycle comme représenté sur la figure 1.5. D'un point de vue sensoriel, ces composés chimiques sont généralement associés à un PS (de 0,32 pour le rhamnose à 1,43 pour le fructose [54]). Ce sont ces sucres simples sous formes « ose » (le fructose ou le glucose) qui sont assimilés par notre organisme.

Structures complexes et nomenclature.

Le carbone anomérique¹ est réactif et est susceptible de former des liaisons osidiques, créant ainsi des oligomères d'oses tels que les oligosaccharides (2 à 10 oses). Ils peuvent être linéaires ou ramifiés, augmentant la complexité de ces structures. Un exemple de disaccharide est donné dans la figure 1.5. Le saccharose est formé par deux monomères, où le carbone C₁ du glucose (Glc) est lié au carbone C₂ du fructose (Fru). Le type d'ose impliqué, sa forme cyclisée et l'orientation de cette liaison permet de définir la nomenclature des oligosaccharides. Dans cet exemple, le saccharose est donc nommé α -D-glucopyranosyl-(1 $\rightarrow$ 2)- β -D-fructofuranoside ou Glc α (1 $\rightarrow$ 2) β Fru. La combinatoire associée à la formation des sucres traduit la complexité structurale et la difficulté d'établir un lien entre structure et PS. D'un point de vue biologique, ces oligosaccharides doivent être clivés par des enzymes pour être assimilés. Ce clivage est opéré dès la mise en bouche par les enzymes salivaires pour les plus grands polysaccharides et dans le tractus intestinal pour les plus courts (sucrase, lactase, etc). Il faut alors comprendre que les sucres sont perçus à différents PS et que leur indice glycémique² varie. Le fructose par exemple, est quasiment aussi sucré que le saccharose à concentrations équivalentes (PS de 1,43) mais possède un indice glycémique faible (19%). Cette subtilité est prise en compte par l'industrie, car le consommateur reste en présence d'un sucre mais possédant de faibles propriétés caloriques.

1. Carbone rendu asymétrique après cyclisation, la position du groupement hydroxyle définit l'anomère α et β .

2. L'indice glycémique est une métrique pour classer les aliments en fonction de l'élévation du taux de sucre dans le sang (glycémie). Plus cet indice est élevé, plus le taux de sucre dans le sang est important après ingestion.

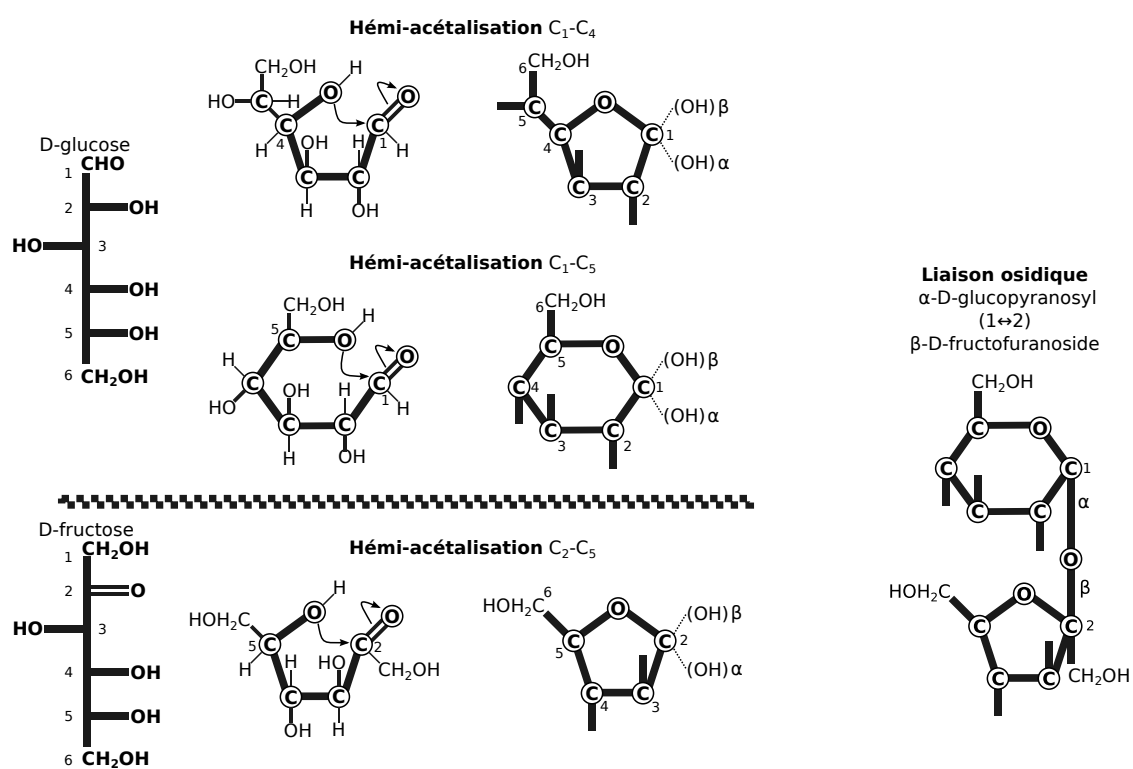


FIGURE 1.5 – Réaction d'hémi-acétalysation et liaison osidique.

Les sucres cyclisés sont représentés selon la projection de Haworth.

• Les édulcorants

Selon le dictionnaire de l'Académie française, un édulcorant est une « *substance qui a la propriété d'adoucir la saveur d'un médicament, d'une boisson* ». Cependant, cette dénomination est plus souvent utilisée pour désigner des composés qui sont perçus sucrés et apportant pas - ou peu - de calories. Grant E. DuBois énonçait en 2012 que plusieurs milliers de composés répondant à la définition d'un édulcorant étaient connus [55]. L'ensemble de ces composés présente une grande diversité structurale. Le record du PS revient à des composés possédant une fonction centrale de type guanidine, dont le lugduname avec une intensité sucrée 200 000 fois plus élevée que celle du saccharose. Ces derniers n'ont pas été découverts par hasard mais à l'occasion d'un vaste programme de recherche de substitut de l'aspartame à l'Université Claude Bernard de Lyon dans les années 1990 [56]. Les édulcorants de type guanidine ne sont pas encore approuvés en vue d'une mise sur le marché et ils ne le seront probablement jamais. Pour les agences de sécurité alimentaire, ils ne présentent pas d'autres avantages que leur PS important. Pour être susceptible d'être mis sur le marché, un édulcorant doit réunir six propriétés : sécurité, qualité gustative, stabilité, solubilité, coût et brevetabilité [57].

• Molécules sur la marché, quelques exemples

Aspartame (E951).

L'aspartame est un édulcorant de synthèse découvert par hasard en 1965 lors d'une synthèse de dérivés dipeptidiques [58]. C'est un ester méthylique de l'acide L-aspartique et de la L-phénylalanine (figure 1.6A). Il est l'édulcorant le plus utilisé aujourd'hui. Il apporte 4 calories par gramme (autant que le saccharose), cependant, avec un PS de 200, il a l'avantage d'être utilisé en plus faible quantité [59]. Il est le plus connu mais également le plus controversé pour ses effets sur la santé [60].

Sucralose (E955).

Le sucralose est un édulcorant artificiel intense non calorique avec un PS de 600. Il est structuralement proche du saccharose puisque c'est un dérivé chloré de celui-ci. Des atomes de chlore sont présents à la place des fonctions hydroxyles en 4, 1' et 6' (figure 1.6B). Le sucralose n'est pas clivé par les enzymes digestives et n'est donc pas métabolisé [59]. Découvert en 1976, il est l'édulcorant artificiel le plus coûteux à produire [55].

Saccharine (E954).

La découverte de la saccharine (ou sulfinide benzoïque, figure 1.6C) en 1878 est le fruit du hasard. Elle présente un PS compris entre 300 et 500. À l'origine utilisée comme médicament pour contrôler le diabète, elle a été utilisée comme édulcorant à cause des rationnements alimentaires durant la seconde guerre mondiale [59]. Ce fut le premier édulcorant commercialisé [55]. Cet édulcorant artificiel et non calorique procure un arrière-goût désagréable chez certains sujets.

Cyclamate (E952).

Découvert en 1937, le PS estimé pour le cyclamate est de 30 à 40 (figure 1.6D). Il est utilisé en combinaison d'autres édulcorants, notamment la saccharine, car il permet de limiter les arrières-goûts désagréables et entraîne une saveur sucrée persistante [55]. Sa commercialisation est actuellement interdite aux États-Unis [61].

Stevioside (E960).

Le stevioside est un diterpène glycosylé (figure 1.6E) extrait des feuilles de la plante *Stevia rebaudiana*. Elle est originaire d'Amérique du Sud et le goût sucré de ces feuilles est connue depuis des centaines d'années. L'identification du composé « stéviolside » date des années 1955 [55]. Malgré un arrière-goût de réglisse, c'est un édulcorant naturel non calorique ayant un PS compris entre 200 et 300, ce qui explique son succès industriel et commercial [59].

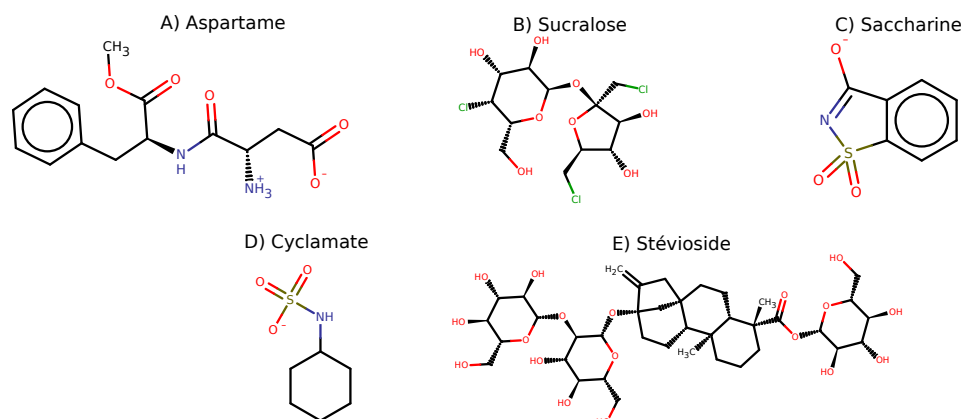


FIGURE 1.6 – Structures chimiques d'édulcorants sur le marché.

Il est intéressant de noter que des protéines (composés d'une centaine d'AA) peuvent, aussi, être perçues sucrées avec un PS allant jusqu'à 3 000. On peut citer, par exemple, la monelline, la thaumatine ou la brazzéine [62–65]. Tandis que d'autres, comme la miraculine ou la curculine, ont la capacité de modifier la perception des saveurs. Une solution acide consommée après celles-ci sera perçue sucrée [66].

La recherche d'édulcorants d'origines naturels ou de synthèses est toujours d'actualité puisque le profil sensoriel des composés actuellement sur le marché est souvent perturbé par un arrière-goût amer. C'est d'ailleurs le cas pour la saccharine ou les dérivés de flavonoïdes. À l'heure actuelle, les consommateurs tendent à privilégier les composés issus de substances naturelles.

• Objectifs : recherche de nouveaux édulcorants

La recherche de nouveaux composés, quel que soit le domaine d'application, a toujours été un défi pour les chercheurs. Les alternatives au sucre, pour pallier aux apports caloriques et aux conséquences sur la santé, ne font pas exception [55, 67]. Comme énoncé, au chapitre 1.3.2, nombreux sont les édulcorants découverts par hasard. Cependant, des programmes de recherches ont permis de mettre en évidence de nouveaux composés ou d'optimiser la perception sucrée de ceux déjà connus [68, 69]. Que ce soit,

par des approches rationnelles ou systématiques, toutes ces études soulignent la difficulté de conception inhérente à la diversité moléculaire des composés sucrés [70, 71]. De plus, la complexité du sens goût induit souvent une perception qui n'est pas seulement sucrée pour ces composés, autrement dit, leur profil sensoriel est loin de ressembler à celui du saccharose.

L'enjeu majeur est donc de proposer une solution pour une sélection rationnelle de composés potentiellement sucrés. Des études ont déjà tenté de proposer des modèles statistiques de prédiction du PS mais restent limitées à l'élaboration d'un modèle théorique (détaillées au chapitre 2). Deux approches seront abordées dans cette étude, une approche statistique basée seulement sur la connaissance de la structure des ligands et une autre, basée sur la mécanique moléculaire, tiendra compte du récepteur. L'établissement d'un protocole combinant plusieurs approches statistiques sera proposé pour sélectionner des composés sucrés, non toxiques et sans goût amer.

1.3.3 Récepteur

Les sucres et les édulcorants interagissent avec des récepteurs membranaires présents à la surface des cellules sensorielles de la langue (partie 1.2.1). Un récepteur est une protéine pouvant reconnaître une molécule afin de transmettre un message à l'intérieur de la cellule *via* une cascade de signalisation. Comme toute protéine, ils sont formés par un enchaînement de monomères que sont les AA et sont catégorisés par famille.

• Un récepteur couplé aux protéines G...

Les récepteurs impliqués dans la perception de la saveur sucrée sont des récepteurs transmembranaires appartenant à la superfamille des récepteurs couplés aux protéines G (ou métabotropiques). Les RCPG possèdent une structure caractéristique à sept hélices transmembranaires (7tm). La partie N-terminale de la protéine est extracellulaire

et la partie C-terminale intracellulaire. Ils permettent d'activer la voie de transduction d'un signal extérieur à la cellule sans diffusion de molécule. Comme leur nom l'indique, le passage du signal requière un couplage à un complexe hétérotrimérique appelé protéine G. La transduction du signal a été détaillée dans la partie 1.2.1 p 6.

Plus de 800 séquences de RCPG ont été référencées dans le génome humain (dont 396 récepteurs olfactifs) et ils constituent ainsi la plus grande famille de récepteurs membranaires [72, 73]. Les RCPG appartenant à cette superfamille varient en structure et en fonction. Ils sont retrouvés dans de nombreux processus biologiques tels que la vision, l'olfaction, la reconnaissance de neurotransmetteurs ou d'hormones, ... Dans un but de catégorisation, des approches de classification de RCPG ont été réalisées à partir d'études phylogénétiques. L'une des premières classifications (nomenclature A-F) a été proposée par Attwood et Findlay en 1994 [74]. Elle a pour vocation d'être exhaustive en regroupant les RCPG de tous les organismes vertébrés ou invertébrés, en six classes : classe A (rhodopsin-like), classe B (secretin receptor-like), classe C (metabotropic), classe D (récepteurs aux phéromones sexuelles), classe E (récepteur à l'AMP cyclique) et classe F (frizzled). La grande variabilité des séquences en AA de tous ces organismes complexifie grandement la classification. Plus récemment, une classification nommée GRAFS et basée sur 800 séquences de RCPG humains a été proposée [75]. La classification GRAFS répartit les RCPG humains en cinq classes [75] et peut être mis en relation avec la classification A-F [76] : *Rhodopsine-like* (classe A), *Secretine- et Adhesion-like* (classe B), *Glutamate-like* (classe C) et *Frizzled/Taste2* (classe F) (figure 1.7).

- ... de classe C

Les récepteurs aux saveurs sucrée et umami appartiennent à la famille des *Glutamate-like* ou RCPG de classe C [19, 20, 77–79]. Cette classe inclue également huit récepteurs métabotropiques au glutame (mGluR), deux récepteurs à l'acide γ -aminobutyrique

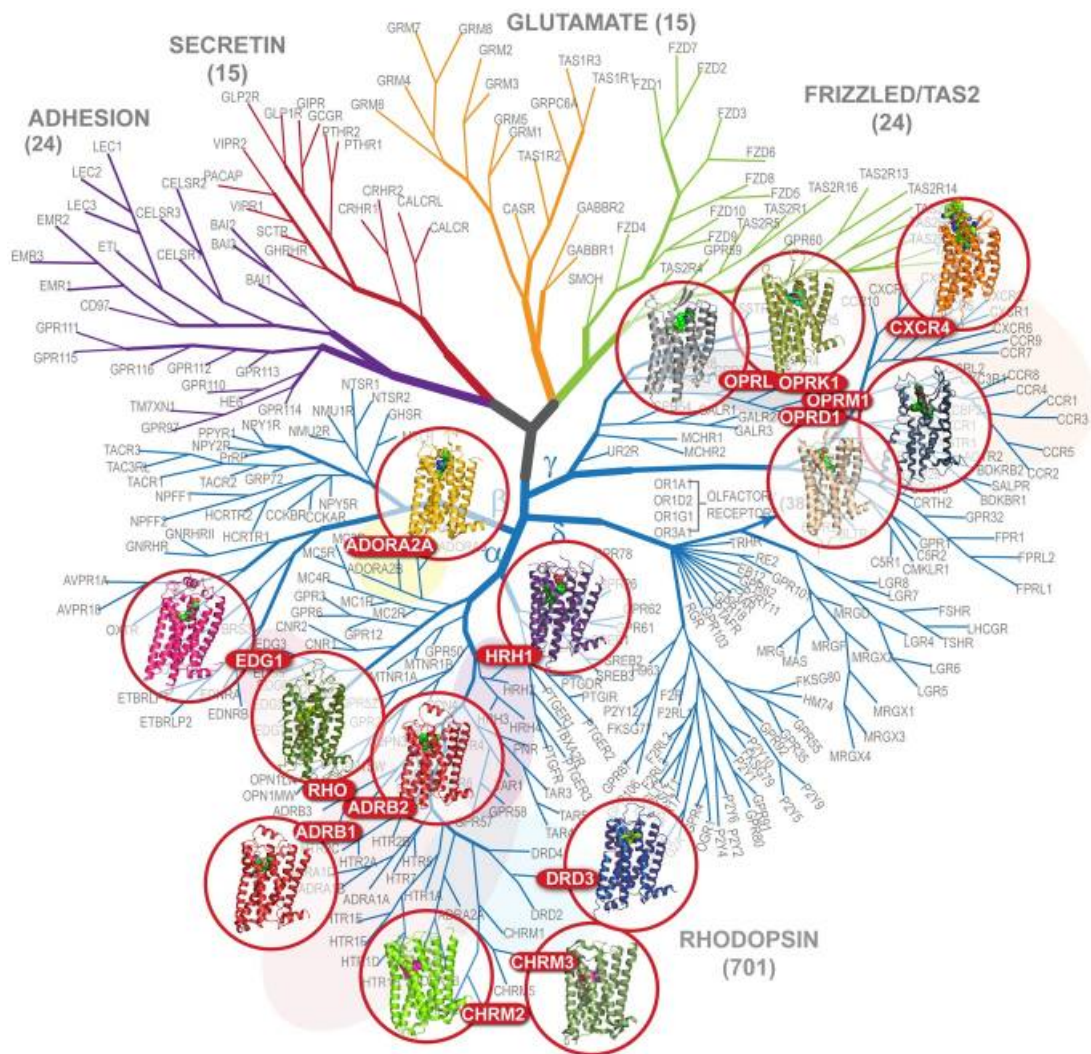


FIGURE 1.7 – Arbre phylogénique des RCPG

Superfamille des RCPG issu de la publication de Katritch *et al.* [73]. Actuellement, une centaine de structures expérimentales sont disponibles et la plupart sont des RCPG de classe A (Rhodopsine-like).

(GABA-B₁ et GABA-B₂), un récepteur sensible au calcium (CasR), un récepteur sensible aux AA (GRPC6A) et enfin sept récepteurs orphelins [78].

Les RCPG de classe C partagent une structure similaire. Ils possèdent trois domaines distincts (figure 1.8). La partie N-terminale du récepteur (ou ectodomaine extracellulaire) est appelée *Venus Flytrap Domain* (VFD). Elle est nommée ainsi pour sa ressemblance avec la *Dionaea muscipula* ou Dionée attrape-mouche. Ce domaine N-terminal a une similarité structurale avec les protéines bactériennes « LIVBP » (leucine isoleucine valine - *binding protein*). Le VFD permet la reconnaissance des ligands « naturels » (glutamate pour l'umami et saccharose pour le sucré). Il est lié à un domaine riche en cystéine (CRD) qui permet de connecter le VFD au domaine transmembranaire (TMD) caractéristique des RCPG.

L'assemblage des RCPG de classe C T1R2 et T1R3 (TAS1R2 et TAS1R3 pour les gènes) forme le récepteur à la saveur sucrée [19]. Plusieurs sites de liaisons ont été reportés [80] (figure 1.8). Les sucres ainsi que la plupart des édulcorants interagissent avec le VFD de T1R2 ou T1R3. Ensuite, chacun des TMD est en mesure de reconnaître des édulcorants ou des modulateurs allostériques. Même si ce n'est pas l'objet de cette thèse, il faut noter que les protéines perçues sucrées ont la capacité d'interagir avec le récepteur au niveau des CRD [65]. Les sites de liaison actuellement connus ont été référencés par G.E. Dubois [81] et sont représentés dans la figure 1.8. Même s'il existe un seul STR, la présence de ces multiples sites de liaison et la diversité des ligands qu'il est capable de reconnaître, augmentent la difficulté d'établir des relations structure/goût. Ceci reflète également la difficulté de description des mécanismes de perception de la saveur sucrée.

• Objectifs : Compréhension des mécanismes moléculaires

Cette partie porte sur la mise en évidence des AA clés impliqués dans la liaison des édulcorants. Des nombreuses études *in vitro*, par mutagénèse dirigée, ont permis d'identifier certains AA cruciaux. Cependant, peu de données structurales sont disponibles

pour analyser, d'un point de vue moléculaire, les modes de liaisons et le comportement dynamique de l'association ligand-récepteur.

Il est important de souligner que, en comparaison aux RCPG de classe A, les mécanismes moléculaires qui entrent en jeu lors de l'activation des RCPG de classe C ont été peu étudiés. Ainsi, les détails structuraux de la dynamique d'activation chez les classes A seront considérés comme un point de départ pertinent pour tenter d'élucider celui des classes C.

Du fait d'un mécanisme d'activation complexe, les enjeux sont ici divers et nombreux. L'identification des AA impliqués dans les sites d'interactions des molécules sucrées permettront, non seulement, une meilleure connaissance des mécanismes moléculaires mis en jeu mais aussi de permettre une conception rationnelle de nouveaux édulcorants. Par ailleurs, le mécanisme de transmission de l'information du site de reconnaissance à la partie intracellulaire du récepteur reste un défi majeur. L'observation de ce phénomène d'activation lors de la liaison d'un édulcorant sera une métrique pertinente en complément de la mesure de l'affinité ligand-récepteur.

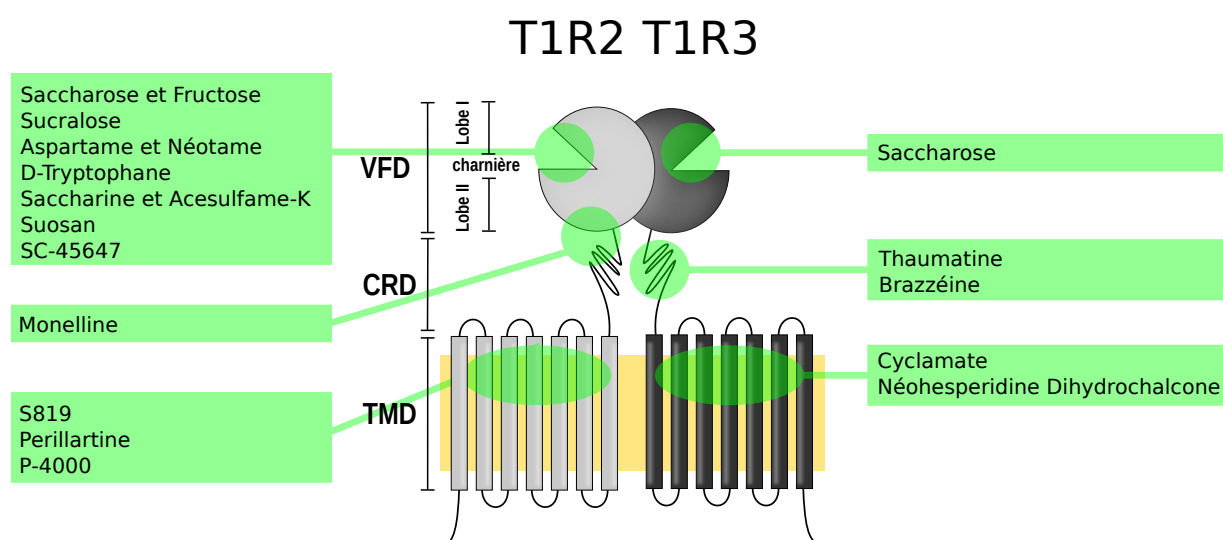


FIGURE 1.8 – Schéma du dimère T1R2-T1R3

1.4 Conclusion

L'objectif de cette thèse est d'apporter des indices sur la compréhension des mécanismes moléculaires impliqués dans la première étape de la perception de la saveur sucrée. La première étude consistera à établir un modèle statistique pour prédire le PS à partir d'une structure moléculaire. En effet, ces molécules partagent des propriétés physico-chimiques communes qui permettent au récepteur T1R2-T1R3, premier protagoniste de la perception, de les reconnaître.

Ensuite, la reconnaissance de ces molécules par le STR sera également abordée, pour définir quels sont les AA clés de ces interactions. Pour cela, la structure tridimensionnelle (3D) du STR est nécessaire. Des méthodes bioinformatiques seront utilisées pour pallier à l'absence de données structurales. L'obtention d'une structure 3D du STR qui récapitule l'ensemble des données expérimentales (notamment les données de mutagenèse dirigée) sera un point de départ robuste pour les études statiques et dynamiques visant à décrire les interactions ligand-récepteur.

L'interaction d'un sucre ou d'un édulcorant avec le STR est une étape nécessaire mais pas suffisante pour entraîner la perception sucrée d'une molécule. Une analogie peut être donnée à ces systèmes, ils sont comparables à des interrupteurs, la fixation d'une molécule doit engendrer un signal produit par le récepteur. Ce concept est applicable à bon nombre de systèmes biologiques. Les études dynamiques permettront de donner de nouveaux indices sur les étapes moléculaires dites d'« activation » du récepteur. Cependant, du fait du manque de données expérimentales et mécanistiques sur l'activation du STR, les informations disponibles sur le mGluR seront d'une grande aide. Ainsi, les simulations de dynamique moléculaire réalisées sur ce système serviront de références pour comprendre le mécanisme d'activation des RCPG de classe C.

Un large éventail d'outils bioinformatiques mis à disposition par la communauté scientifique sera utilisé pour mener à bien ce projet. Les descriptions théoriques des méthodes employées seront abordées dans ce manuscrit et seront plus largement détaillées en annexes.

Chapitre 2

De la structure chimique au pouvoir sucrant

Ce chapitre est basé sur la publication suivante :

Chéron, J.-B., Casciuc, I., Golebiowski, J., Antonczak, S., Fiorucci, S. (2017).
Sweetness prediction of natural compounds. *Food Chemistry*, 221, 1421–1425.
<http://doi.org/10.1016/j.foodchem.2016.10.145>

2.1 Préambule : à la poursuite de l'édulcorant « parfait »

Le sucre de table, ou saccharose, est omniprésent dans notre alimentation. Les enjeux associés à sa surconsommation, notamment sous forme « cachée », dans les plats préparés par exemple, ont largement été décrits dans le paragraphe 1.3 p 15 de l'introduction. L'utilisation d'édulcorants est une des alternatives au saccharose pour pallier aux conséquences que peut avoir un excès de sucre dans l'alimentation. Ce sont des molécules peu – ou pas – caloriques qui sont perçues sucrées. Elles peuvent être d'origine naturelle ou synthétique. Aujourd'hui, un grand nombre de composés répond à la définition d'édulcorant [55, 65, 71]. Cependant, toutes ces molécules n'ont pas le même profil chimique (solubilité, stabilité thermique, ...) et chémosensoriel que celui du saccharose. En effet, lors de la mise en bouche d'un édulcorant la saveur sucrée n'est pas la seule à être perçue. Par exemple, la saccharine qui est l'un des édulcorants les plus consommés, est à la fois perçue sucrée et amère [82]. De plus, certains sont sujets à des controverses pour leur impact sur la santé, tandis que d'autres n'ont jamais reçu d'autorisation de mise sur le marché [83]. Étant donné les problématiques énoncées, la découverte d'édulcorants est toujours un défi, autant pour des raisons de santé publique, que pour leur utilisation par l'industrie agroalimentaire.

La mise en relation de la structure chimique des sucres et des édulcorants avec la perception de la saveur sucrée a toujours été un défi. À cette fin, de nombreux modèles pharmacophoriques¹ ont été proposés ces dernières années. L'historique et les diverses évolutions de ces modèles pharmacophoriques de la saveur sucrée sont largement documentés dans la revue de G.E. Dubois de 2011 [67]. Le premier modèle, qui est aussi le plus connu, date de 1967. Ce modèle, nommé « A-H/B », développé par Shallenderger et Acree, stipule que tout composé sucré doit posséder un groupement donneur et un groupement accepteur de liaison hydrogène séparés par une distance

1. Un pharmacophore est une propriété – ou un ensemble de propriétés – partagée par des molécules ayant une même action biologique.

comprise entre 2,5 et 4 Å [84]. Depuis, divers modèles ont été proposés pour le compléter, notamment le modèle de Kier qui prend en compte le caractère hydrophobe porté par certains édulcorants [85]. Le plus complexe est connu sous le nom de *multiple points attachment theory*. Il date de 1996 et contient pas moins de neuf propriétés physico-chimiques [86]. L'ensemble de ces tentatives de description révèle la grande complexité structurale des molécules perçues sucrées.

L'objectif de ce chapitre est de proposer une approche *in silico* pour la recherche et la sélection de nouveaux composés perçus sucrés. La première étape consiste à établir une base de données de molécules perçues sucrées dont la structure chimique et le PS² sont déjà connus. Ensuite, un ensemble de propriétés structurales et physico-chimiques seront calculés sur ces composés. Une analyse de ces propriétés permettra d'extraire les grandes caractéristiques chimiques des sucres et édulcorants. Enfin, l'objectif sera de construire un modèle statistique de prédiction du PS. Les avantages de ces méthodes statistiques sont, à la fois, leur robustesse face à la variable à prédire (ici le PS) et le fait qu'elles se basent sur la seule connaissance des ligands. Ce modèle de prédiction sera utilisé pour sélectionner des composés d'intérêts dans une base de données de composés naturels. Pour finir, le protocole proposé est développé pour éliminer les composés perçus amers (souvent en arrière goût) et les composés présentant des groupements toxiques référencés.

2.2 Base de données de molécules « sucrées »

2.2.1 316 composés et leur pouvoir sucrant

La liste de molécules, dont le PS est connu, est établie à partir d'études précédentes [54, 87, 88]. Seuls les composés qui sont ni des mélanges (plus d'une molécule),

2. Pouvoir Sucrant : il est défini comme étant le rapport entre une masse - ou une concentration - de saccharose et d'une substance sucrée à intensité de perception égale.

ni des protéines sont conservés. Les PS associés varient de 0,2 à 225 000. L'échelle logarithmique en base 10 du PS est utilisée à cause du large intervalle des valeurs du PS et de leur asymétrie de répartition (noté logPS par la suite). Le tableau A.1 en annexe référence les composés sélectionnés pour cette étude. Cette base de données est notée par la suite `SweetenersDB`.

2.2.2 Structures chimiques

Dans un premier temps, le protocole pour prédire le logPS nécessite la connaissance de la structure chimique (en 2 ou 3 dimensions) des composés de la `SweetenersDB`.

2D

Une majorité des composés de la `SweetenersDB` est présent au format 2D « isomérique » dans la publication de Zhong *et al.* [54]. Le format SMILES décrit une molécule en 2D à l'aide d'une chaîne de caractères, mais il n'apporte aucune information sur les coordonnées des atomes dans l'espace. Contrairement à la forme canonique, la forme isomérique des SMILES permet de prendre en compte la chiralité et la conformation des doubles liaisons. Pour les autres composés, les SMILES isomériques sont téléchargés depuis la base de données PubChem [89]. Les états de protonation des composés sont déterminés pour un pH de 6,5, correspondant au pH salivaire physiologique [90].

3D

Les SMILES isomériques sont utilisés pour générer de manière automatique les structures 3D avec le logiciel Marvin (Marvin 14.9.29.0, 2014, Chemaxon). Le champ de force MMFF94 (*Merck Molecular Force Field*) [91] est utilisé. Pour chaque molécule, les trois conformères de plus basse énergie sont sélectionnés pour la suite de l'étude. Ici aussi, l'état de protonation est défini à pH 6,5.

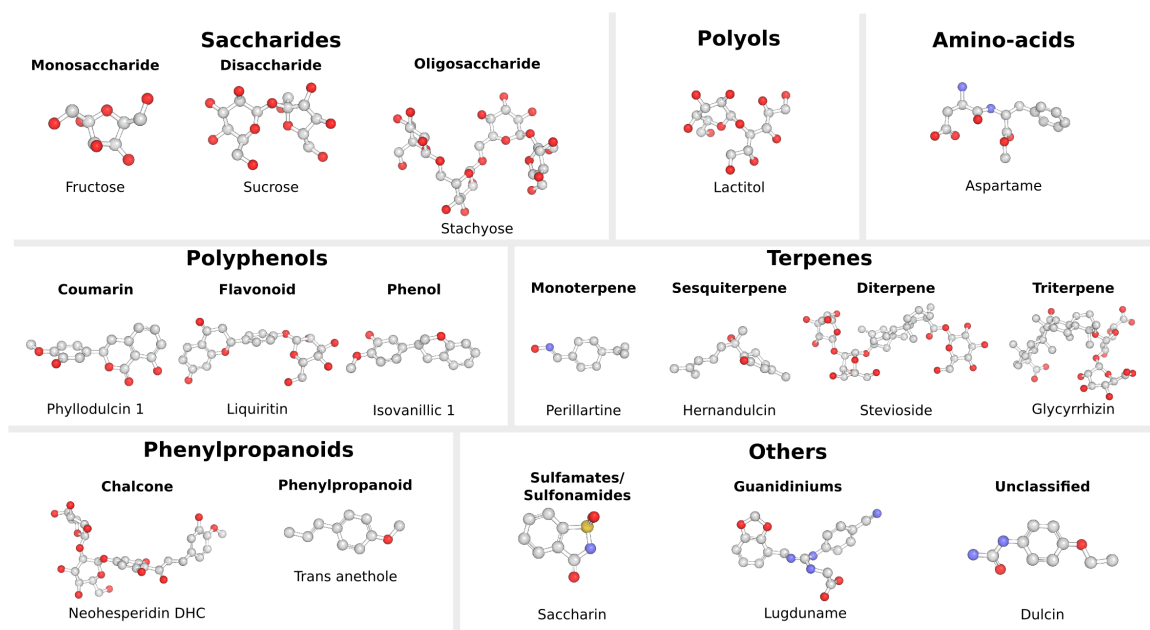


FIGURE 2.1 – Diversité structurale des composés de la SweetenersDB

2.2.3 Diversité structurale

Familles chimiques

Comme énoncé dans l'introduction, au paragraphe 1.3.2, les molécules perçues sucrées proviennent de familles chimiques variées (figure 2.1). La SweetenersDB reflète cette diversité structurale. La famille caractéristique de cette saveur est bien évidemment celle des sucres, représentée par des mono et des polysaccharides. Il faut noter qu'une partie des composés inclus dans cette famille ne sont pas des sucres naturels mais sont des dérivés chlorés et bromés du saccharose. Les polyols sont des composés polyhydroxylés qui ne sont pas des sucres, mais certains sont également perçus sucrés. Quelques acides-aminés libres, polypeptides et leurs dérivés forment également une sous-famille. 157 dérivés peptidiques issus de la publication de Zhong *et al.* [54] sont présents dans la SweetenersDB. Les composés terpéniques sont largement représentés ($\approx 10\%$ de la base de données), sous forme mono, di, tri ou sesquiterpènes. De manière contre intuitive, au vu du caractère largement polaire des sucres, les dérivés

terpéniques et les phenylpropanoïdes sont caractérisés par un noyau hydrophobe. Cependant, cette propriété est en accord avec les modèles pharmacophoriques énoncés précédemment. Il est à noter que d'autres composés de synthèse forment une classe à part dans la *SweetenersDB*, comme la saccharine ou le lugduname par exemple.

Masse molaire

Les composés de la *SweetenersDB* présentent des masses molaires allant de 120 à 1 300 g.mol⁻¹. Ils contiennent de 8 à 89 atomes lourds, dont $\approx 60\%$ sont des atomes de carbone. Les composés naturels (colorés en vert sur la figure 2.2) ont les masses molaires les plus élevées. Les composés ayant les masses molaires les plus importantes ne sont pas perçus les plus sucrés. Ils atteignent un plateau pour un logPS valant 3. Les composés les plus sucrés, quant-à eux, ont des masses molaires comprises entre 350 et 450 g.mol⁻¹.

Liaisons hydrogènes

En ce qui concerne la *SweetenersDB*, les molécules avec les logPS les plus élevés possèdent 4 donneurs et jusqu'à 10 accepteurs de liaisons hydrogènes. Les molécules naturelles sont celles qui peuvent établir le plus de liaisons hydrogènes, notamment par leur forte glycosylation. Dans une structure osidique, la distance entre deux fonctions $-OH$ consécutives, jouant le rôle de donneur et accepteur de liaisons hydrogènes est $\approx 3 \text{ \AA}$ et est en parfait accord avec le modèle « A-H/B » de Shallenberger et Acree.

Coefficient de partage octanol/eau

Le coefficient de partage octanol/eau, noté logP, est défini comme le ratio de concentration d'une substance X d'intérêt dans ces deux phases : $\log P = \log\left(\frac{[X]_{\text{octanol}}}{[X]_{\text{eau}}}\right)$. S'il est positif, la molécule X est plus soluble dans l'octanol que dans l'eau. Il informe sur la solubilité dans l'eau et, par extension, de l'hydrophobicité potentielle des molécules. De manière générale, les composés perçus sucrés doivent être solubles dans la salive. La moyenne du logP des composés de la *SweetenersDB* est de $0 \pm 2,1$. Cet écart-type est

représentatif de la grande diversité chimique des composés. Les logPS les plus importants sont observés pour des composés ayant un logP compris entre 3 et 3,5, autrement dit, ceux présentant un caractère plutôt hydrophobe.

Comme le montre la figure 2.2, les caractéristiques précédemment citées ne sont pas linéairement corrélées au logPS. Ainsi, les composés les plus performants d'un point de vue sensoriel présentent une composante hydrophobe, une masse molaire modérée, 4 donneurs et jusqu'à 10 accepteurs de liaisons hydrogènes. L'ensemble des propriétés énoncées ici sont en accord avec les modèles pharmacophoriques précédemment cités. Les molécules naturelles ne présentent pas les logPS les plus élevés, comme le montre leurs propriétés. Cela suggère que des édulcorants intenses restent à découvrir parmi les composés naturels.

2.3 Approche statistique « ligand-based »

2.3.1 Méthodologie pour l'étude de la relation structure chimique - pouvoir sucrant

Dans ce chapitre, il ne sera pas question d'utiliser les caractéristiques physico-chimiques des composés perçus sucrés pour établir un nouveau modèle pharmacophorique. De telles descriptions ont déjà été effectuées et permettent de définir les propriétés portées par les sucres et les édulcorants. Ici, l'objectif est d'utiliser ces informations physico-chimiques pour établir un modèle mathématique de prédiction du logPS. Un modèle fiable et adapté permettra de prédire cette variable sur de nouveaux composés. Cependant, la diversité moléculaire des composés, dont le logPS est connu, augmente la complexité pour mettre en place une telle approche. La stratégie envisagée est de calculer un ensemble de propriétés (ou descripteurs) sur les structures moléculaires des composés présents dans la `SweetenersDB` et de construire un modèle statistique

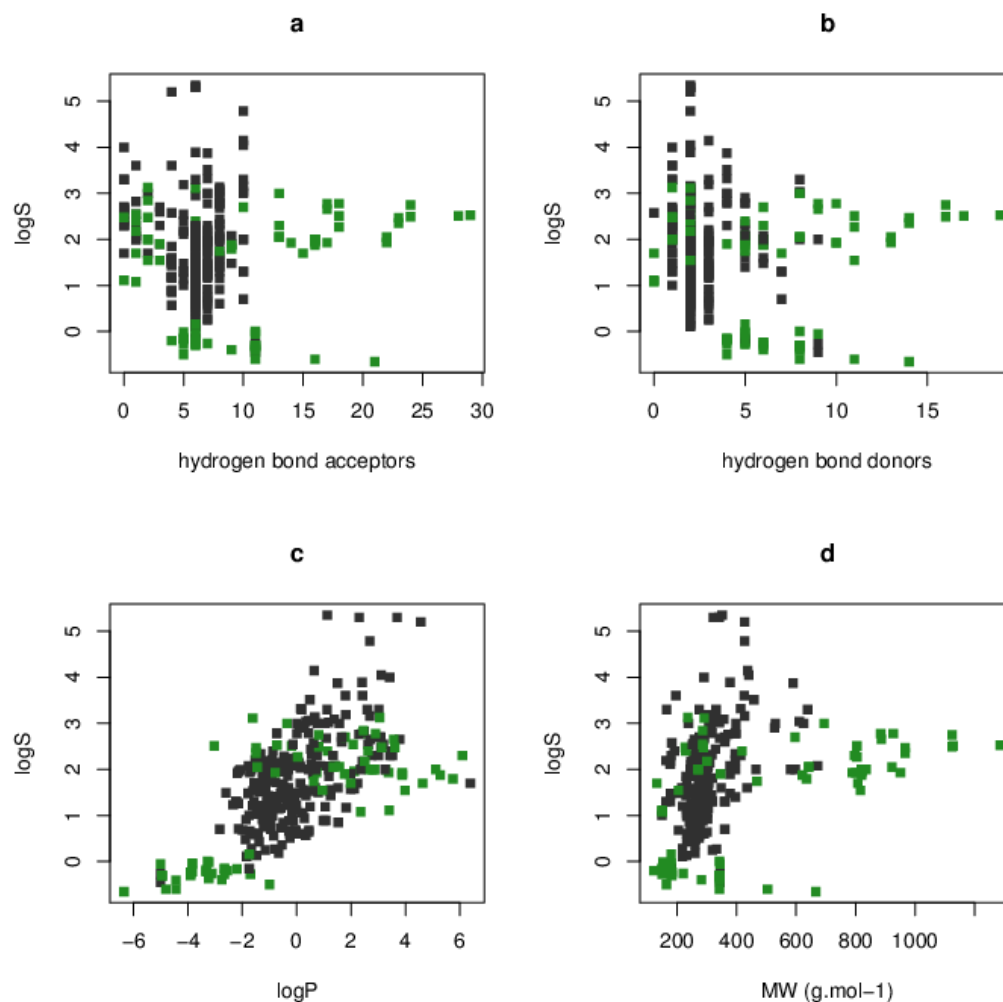


FIGURE 2.2 – Propriétés physico-chimiques des composés de la SweetenersDB.

Le nombre d'accepteurs (a) et de donneurs (b) de liaisons hydrogènes, le coefficient de partage octanol/eau (c), ainsi que la masse molaire (d) sont représentés en fonction du logPS. Les composés naturels sont représentés en vert et les composés de synthèse en noir.

de prédiction. Ainsi, le modèle sera en mesure de prédire rapidement et avec fiabilité le logPS à partir des mêmes descripteurs calculés sur des molécules d'intérêts.

• Principe du QSAR

QSAR est l'acronyme de *Quantitative Structure Activity Relationship* ou étude de relation structure-activité. Les algorithmes utilisés pour construire les modèles de prédictions ne sont pas seulement restreints à un emploi en biologie ou bioinformatique mais sont des méthodes statistiques applicables à n'importe quel jeu de données. Ces méthodes sont utilisées pour prédire une variable (ici le logPS) à partir d'un ensemble de données (dans ce cas des descripteurs physico-chimiques). La *SweetenersDB* contient 316 molécules avec leurs logPS, elle est utilisée pour construire le modèle statistique de prédiction. Deux sous-ensembles sont extraits, un premier (majoritaire) est utilisé pour construire le modèle *via* un algorithme de prédiction. Il est appelé « jeu d'apprentissage ». Le second permet de tester sa performance, il est appelé « jeu de validation externe ». Ce jeu de validation externe n'est jamais observé par le modèle d'apprentissage, mais les logPS prédits sont comparés avec les valeurs expérimentales pour tester la robustesse du modèle.

Il est possible de construire le modèle sur le jeu d'apprentissage complet ou alors d'utiliser des méthodes d'échantillonnage (type validations croisées) pour augmenter sa performance. Les méthodes des *k-fold* ou du *leave one out* peuvent être appliquées sur les N individus. Pour les *k-fold*, le modèle est réalisé k fois. Il est construit sur $N - k$ individus et la prédiction est réalisée sur les k restants. L'erreur associée au modèle final de prédiction est déterminée par la moyenne des k erreurs. La méthode du *leave one out* (ou «seul contre tous») est un cas particulier, où le modèle est construit sur $N - 1$ individus et prédit sur 1. De la même manière, le modèle final dérive d'une combinaison des modèles de prédictions pour réduire l'erreur quadratique moyenne globale. Dans le cas présent, la méthode du *leave one out* est utilisée.

Les méthodes statistiques dites « supervisées » sont celles dont la variable à prédire

est connue. Un grand nombre d’algorithmes de construction de modèle d’apprentissage sont disponibles. Le choix de la méthode dépend des données utilisées par le modèle et du type de variable à prédire. Comme le logPS est une variable continue, une méthode de régression est nécessaire. Le modèle le plus simple pour prédire une variable continue à partir d’un ensemble de descripteurs en entrée est la régression linéaire multiple. Son objectif est de pondérer l’ensemble des descripteurs pour établir une relation linéaire avec la variable à prédire. Ce modèle simpliste sous entend que la relation à établir entre les descripteurs et la variable à prédire est linéaire. Dans le cas de la prédiction du logPS des méthodes plus complexes sont nécessaires. Le détail des algorithmes utilisés, les Forêts Aléatoires (RF pour *Random Forest*) et les Machines à Vecteur de Support (SVM pour *Support Vector Machine*), est donné en annexe B.1.

2.3.2 Construction du modèle de prédiction

La première étape est de construire des modèles statistiques de prédiction à partir des composés de la *SweetenersDB*. Ensuite, les modèles doivent être comparés pour sélectionner le plus performant. Pour déterminer si le modèle est exhaustif, la structure chimique des molécules de la *SweetenersDB* est comparée à l’ensemble des composés sucrés déjà connus (indépendamment du PS). Le modèle statistique est alors utilisé pour prédire le logPS sur un jeu de données de molécules d’intérêts. La comparaison des structures chimiques avec la *SweetenersDB* permet de sélectionner les composés sur lesquels le modèle de prédiction peut être appliqué : c’est le domaine d’applicabilité. Le protocole global est schématisé dans la figure 2.3.

• Préparation des jeux de données et paramétrisation

Matrices de descripteurs

Un ensemble de 4 887 descripteurs physico-chimiques sont disponibles avec le logiciel Dragon [92]. Il permet de calculer, sans erreur pour l’ensemble de la *SweetenersDB*, 2 080 descripteurs 2D et 3 162 descripteurs 2D+3D. Pour les molécules en 3D, dont

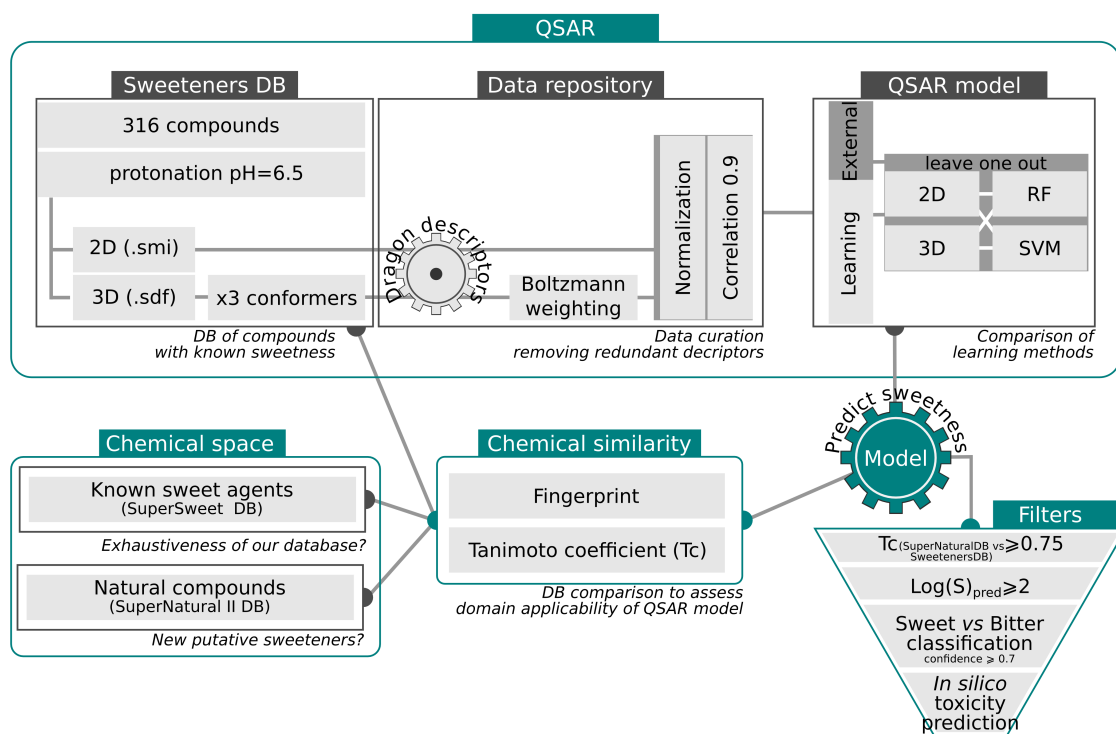


FIGURE 2.3 – Schéma du protocole de sélection de nouveaux composés sucrés.

Résumé de la méthode :

- Construction du modèle statistique de prédiction
- Comparaison des espaces chimiques et définition du domaine d'application
- Utilisation du modèle de prédiction du logPS et applications des filtres

plusieurs conformères ont été trouvés, les trois structures de plus basse énergie ont été conservées. Les valeurs des descripteurs correspondant aux conformères d'une même molécule sont pondérées par le pourcentage relatif de chaque conformère dans la statistique de Maxwell-Boltzmann, selon le rapport des populations N_1 d'énergie E_1 et N_2 d'énergie E_2 (avec $E_1 < E_2$) :

$$\frac{N_2}{N_1} = e^{-\frac{E_2 - E_1}{k_B \cdot T}} \quad (2.1)$$

avec k_B la constante de Boltzmann et T la température (300 K). Ainsi, un seul jeu de descripteurs caractérise une molécule. Les descripteurs des deux matrices sont normalisés entre -1 et 1. Étant donné le nombre de descripteurs, une redondance est présente dans la caractérisation des molécules. Pour s'affranchir de ce biais, un seul représentant parmi les descripteurs corrélés à plus de 90% est sélectionné. Les matrices finales contiennent 244 descripteurs pour les molécules décrites en 2D et 265 pour celles en 3D. Enfin, les matrices contenant les descripteurs des molécules de la *SweetenersDB* sont divisées en deux sous-ensembles par tirage aléatoire. Ces sous-ensembles sont identiques pour les deux jeux de données 2D et 3D. Ils contiennent 225 ($\approx 70\%$) et 91 ($\approx 30\%$) molécules pour former, respectivement, les jeux d'apprentissage et de validation externe.

Paramètres des modèles de prédiction

Les jeux d'apprentissage des deux représentations (2D et 3D) sont utilisés pour construire les modèles de prédiction du logPS avec les méthodes des RF et des SVM. Les modèles SVM sont construits en utilisant la méthode du *leave one out*. Une relation linéaire entre le logPS expérimental et le logPS calculé est attendue. Le calcul du coefficient de corrélation de Pearson, qui représente la somme des écarts à la moyenne selon l'équation :

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_{i=1}^N ((X_i - \bar{X})(Y_i - \bar{Y}))}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \cdot \sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (2.2)$$

permet de quantifier la linéarité de la relation.

Une optimisation des paramètres internes à chacun des algorithmes est réalisée en *brute-force*³. L'objectif est d'obtenir la meilleure prédiction, donc de maximiser le coefficient de corrélation, sur les jeux d'apprentissage et de validation externe. Pour la méthode des RF (annexe B.1), le nombre d'arbres varie de 100 à 1 000 avec un pas de 50. Le nombre de descripteurs utilisés par chaque arbre est également optimisé, en faisant varier ce paramètre de 5 à 55 avec un pas de 10. La figure 2.4 répertorie les valeurs de corrélation en fonction des paramètres à optimiser pour les jeux de données en 2D et 3D. Pour les descripteurs 2D, les corrélations les plus importantes sont 0,980 et 0,860 pour, respectivement, le jeu d'apprentissage et de validation externe. Elles sont observées pour le couple de paramètres de 700 arbres et de 25 descripteurs par arbre. Pour le modèle utilisant les descripteurs 3D, 250 arbres et l'utilisation de 15 descripteurs par arbre montrent les meilleures corrélations : 0,990 pour le jeu d'apprentissage et 0,872 pour le jeu de validation externe.

Pour la méthode des SVM (annexe B.1), les noyaux mathématiques linéaire, polynomial et radial sont testés et le polynomial en base 3 montre les meilleures performances. L'optimisation des paramètres « coût » (C) et « marge de tolérance » (ε) est effectuée pour cette méthode. Le coût varie de 0,1 à 2 avec un pas de 0,25 et la marge de tolérance varie de 0,001 à 0,015 avec un pas de 0,005. Pour chacune des approches, l'optimisation de ces paramètres n'influe que très peu sur la corrélation mesurée.

- 2D : $0,827 \pm 0,000$ pour le jeu d'apprentissage et $0,910 \pm 0,000$ pour le jeu de validation externe
- 3D : $0,829 \pm 0,000$ pour le jeu d'apprentissage et $0,919 \pm 0,000$ pour le jeu de validation externe

Les paramètres par défauts sont alors conservés, coût=1,0 et $\varepsilon=0,001$.

3. Une approche par *brute-force* consiste à explorer toutes les combinaisons possibles.

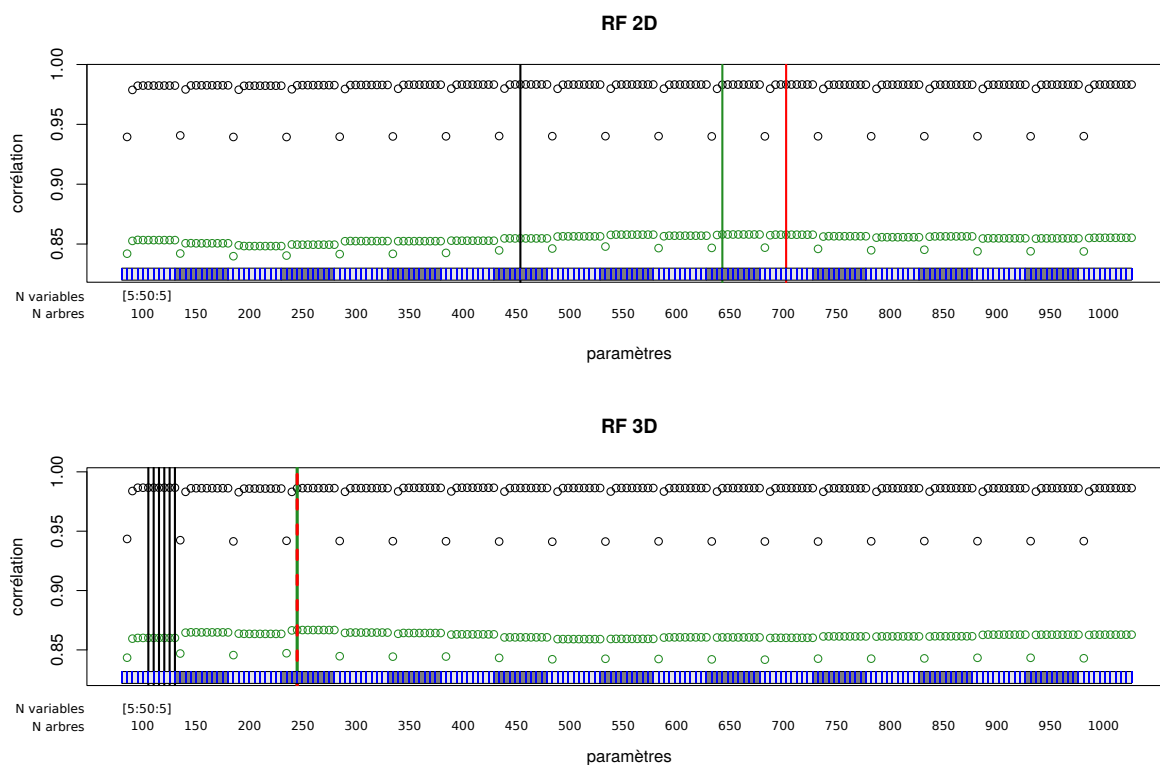


FIGURE 2.4 – Corrélations mesurées pour l'optimisation des RF en brute force.

L'optimisation de la méthode des RF sur les jeux de données 2D et 3D est représentée. Le nombre d'arbres varie de 100 à 1000 avec un pas de 50. Pour chacune de ces valeurs, le nombre de descripteurs utilisés pour construire les arbres varie de 5 à 50 avec un pas de 5. Les corrélations mesurées sur les jeux d'apprentissage et de validation externe sont représentées, respectivement, par des points noirs et verts. Les lignes noires correspondent aux meilleures corrélations pour le jeu d'apprentissage, les lignes vertes les meilleures corrélations pour le jeu de validation externe. Enfin la ligne rouge correspond au couple de paramètres sélectionnés pour maximiser les corrélations des jeux d'apprentissage et de validation externe.

• Performance des modèles

Les performances de chacune des approches avec les paramètres cités précédemment sont reportées dans la figure 2.5. Le coefficient de détermination R^2 est utilisé pour comparer les modèles entre eux. Le R^2 peut avoir plusieurs définitions. Ici, il correspond au coefficient de corrélation de Pearson élevé au carré. Dans ce cas, il ne reflète pas directement la relation de linéarité entre la variable expérimentale et la variable à prédire mais la performance du prédicteur.

L'ensemble des modèles de prédiction présente des valeurs de coefficient de détermination satisfaisantes ($R^2 > 0,8$) pour les jeux d'apprentissage et de validation externe. Les modèles basés sur les descripteurs 2D sont privilégiés pour des raisons de coût et de temps de calcul. En ce qui concerne la méthode RF, les valeurs prédites sur les composés du jeu de validation externe ayant un logPS élevé sont sous-estimées. Un plateau est atteint pour prédire les forts logPS – de potentiels édulcorants intenses – qui présentent un intérêt particulier pour cette étude. De plus, les R^2 élevés montrent que l'approche des SVM est performant sur le jeu de validation externe. En conclusion, le modèle SVM calculé sur les molécules décrites en 2D est conservé pour sélectionner de nouveaux composés à fort logPS dans une base de données de composés naturels.

2.3.3 Application : prédiction du pouvoir sucrant de composés naturels

Une base de données de composés naturels est utilisée pour prédire le logPS car, comme énoncé dans le paragraphe 2.2.3 de ce chapitre, ceux actuellement connus n'atteignent pas les performances des composés de synthèses. De manière plus concrète et à des fins pratiques, l'aspect naturel est de plus en plus important pour le consommateur et est donc privilégié dans cette étude.

• Comparaison des espaces chimiques

Deux bases de données sont récupérées :

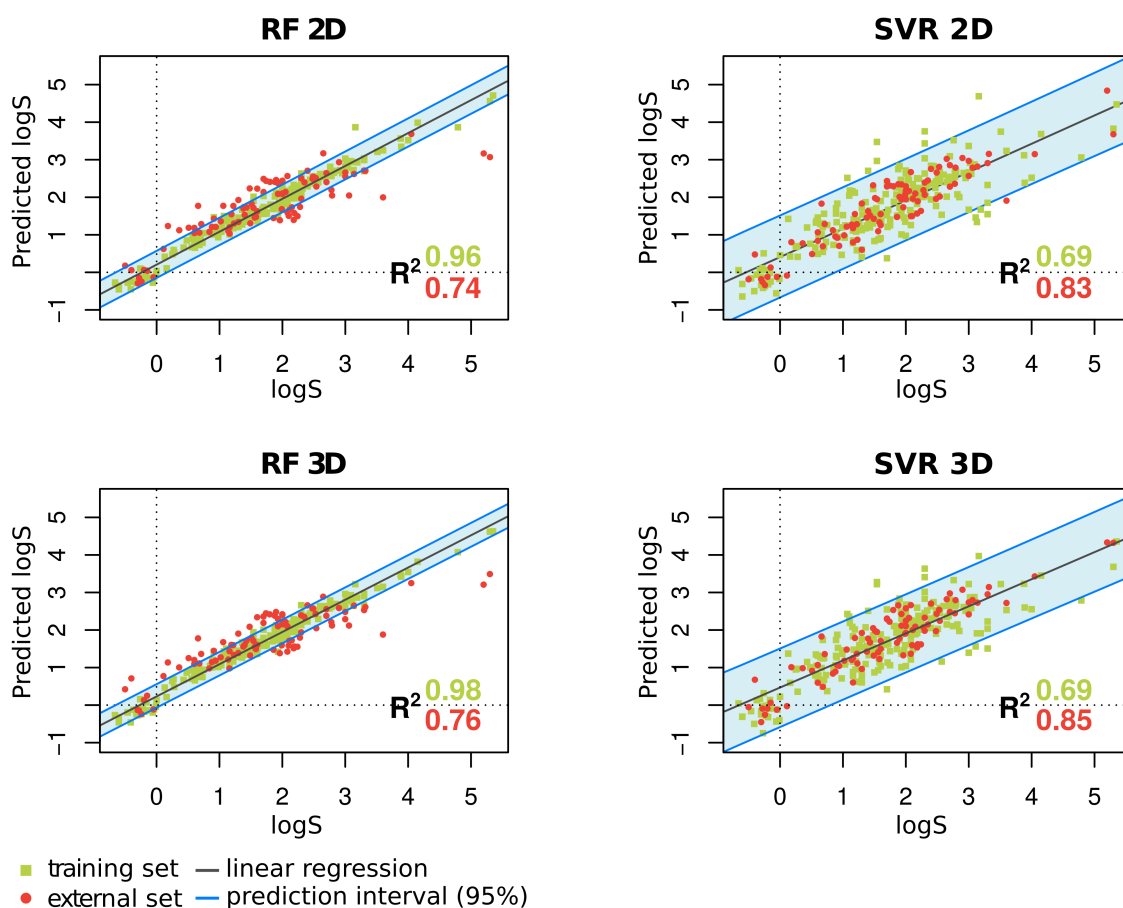


FIGURE 2.5 – Performances des modèles de prédictions.

Les modèles SVM et RF en 2D et 3D sont évalués sur leur performance à prédire le logPS. Les droites de corrélations et les valeurs des coefficients de détermination sont représentées pour les jeux d'apprentissage et de validation externe.

- la SuperSweet [45] qui contient la liste la plus exhaustive de molécules sucrées ou apparentées aux sucres (7 471 composés sont sans information sur leur PS)
- la SuperNatural II qui contient 325 102 molécules naturelles [93]

La comparaison des espaces chimiques de ces deux bases de données avec la SweetenersDB permet de i) définir si la SweetenersDB est représentative de l'ensemble des composés sucrées connues et ii) sélectionner les composés de la SuperNatural II suffisamment proche, d'un point de vue chimique, de la SweetenersDB pour appliquer le modèle QSAR. Les molécules de ces deux dernières bases de données sont récupérées au format 2D. Les *fingerprints* PubChem sont calculés sur l'ensemble des composés et la similarité par paire est déterminée avec le coefficient de Tanimoto (T_c).

Fingerprint

Les formats 2D sont utilisés pour calculer les *fingerprints*, ils correspondent à une succession de 0 et 1 renseignant sur l'absence ou la présence d'une propriété physico-chimique. Le format PubChem contient 881 bits, soit autant d'informations permettant de comparer deux structures chimiques. Les *fingerprints* sont calculés sur l'ensemble des molécules des trois bases de données.

Coefficient de Tanimoto

Le T_c permet de définir un indice de similarité entre 0 et 1. Il est défini, entre deux molécules A et B, telle que :

$$T_{c_{AB}} = \frac{c}{a + b - c} \quad (2.3)$$

avec a et b le nombre de bits valant 1 dans les *fingerprints* des molécules A et B, puis c le nombre de bits valant 1 en commun. Plus cette valeur est proche de 1 plus les structures chimiques sont similaires. La figure 2.6 représente les similarités par paire entre les trois différentes bases de données. Le seuil du T_c , pour considérer qu'au-delà de cette valeur deux structures font partie du même espace chimique, est fixé à 0,75.

La comparaison de la SuperSweet avec la SweetenersDB est représentée par les individus en noir sur la figure 2.6. Elle montre que 99,5% des composés de la SuperSweet ont un Tc supérieur à 0,75 avec la SweetenersDB. Ainsi, la SweetenersDB est représentative des composés sucrés actuellement référencés. Les quelques molécules de la SweetenersDB non représentées sont des AA (D-glutamine, D-tryptophane, ...), des nucléotides, ainsi que des dérivés fluorés du benzothiazol.

Les Tc calculés entre la SuperNatural II et la SweetenersDB permettent de sélectionner les composés utilisables par le modèle de prédiction, autrement dit de définir le domaine d'applicabilité. La projection des Tc (en gris) sur la figure 2.6 reflète la diversité structurale de la SuperNatural II. En fixant un seuil de 0,75 pour le Tc, 105 620 composés peuvent être utilisés par le modèle QSAR, soit 32% de la base de données initiale.

Application du modèle de prédiction

Le modèle statistique SVM 2D est appliqué pour prédire le logPS sur les 105 620 composés ayant passés le filtre de similarité. Environ 75 000 molécules sont prédites avec un logPS supérieur à 2, donc considérées comme de probables édulcorants intenses. Ces composés d'origine naturelle sont sélectionnés et des filtres supplémentaires sont appliqués afin d'éliminer ceux potentiellement perçus amers et ceux présentant des groupements connus pour leur toxicité.

• Évaluation de l'amertume et de la toxicité

Modèle de classification « sucré » / « amer »

Le premier filtre vise à éliminer les composés pouvant posséder les deux modalités. Un modèle statistique de classification est réalisé pour prédire la variable qualitative « sucrée » ou « amère ». La bitterDB contient la liste des composés actuellement connus pour être amer [37]. Les molécules au format SMILES sont récupérées depuis cette base de données et les descripteurs physico-chimiques 2D sont calculés. Ces derniers sont

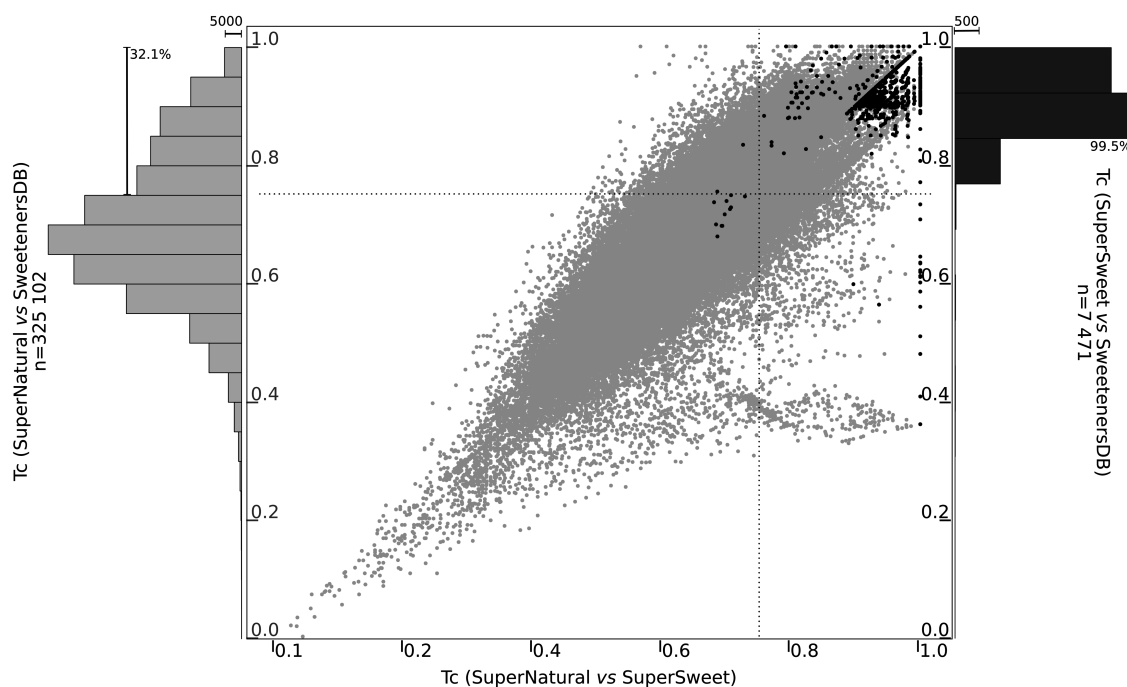


FIGURE 2.6 – Espace chimique de la SuperNatural II

concaténés avec la SweetenersDB, normalisés entre -1 et 1, puis un descripteur parmi ceux corrélés à plus de 90% est sélectionné. Le descripteur qualitatif « sucré » / « amer » permet de discriminer les molécules de chacune des bases de données. 316 composés définis comme « sucré » et 680 composés « amer » sont utilisés pour construire le modèle de prédiction. Ce jeu de données est divisé en jeu d'apprentissage (80%) et jeu de validation externe (20%), la répartition des classes est représentée dans le tableau 2.1. Dans ce cas, la variable à prédire par le modèle de prédiction est qualitative. Pour estimer la performance du modèle le coefficient de Matthews est calculé à partir des tableaux de contingence du jeu d'apprentissage et de validation externe :

$$Mcc = \frac{VP * VN - FP * FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (2.4)$$

Un prototype de tableau de contingence est représenté en 2.2A. Il permet de définir les individus correctement ou incorrectement prédits dans chacune des classes. Les

Mcc calculés à partir des tableaux de contingence des jeux d'apprentissage et de validation externe valent, respectivement, 0,99 et 0,84 pour la méthode des RF et 0,82 et 0,84 pour les SVM. Les RF apparaissent nettement plus performants. Les tableaux de contingence 2.2B et C, réalisés avec la méthode des RF, représentent les résultats pour, respectivement, le jeu d'apprentissage et de validation externe.

Le modèle de classification des RF est appliqué aux $\approx 75\,000$ composés de la SuperNatural II avec un logPS prédit supérieur à 2. Parmi eux, 21% des $\approx 75\,000$ composés sont sélectionnés, car ils sont classés dans la catégorie « sucrée » sans ambiguïté.

	sucré	amer	total
Apprentissage (80%)	253	543	796
Validation externe (20%)	63	137	199
total	316	680	996

TABLE 2.1 – Partition des données pour le modèle de classification « sucré » / « amer »

		Prédit	
		sucré	amer
Vrai	sucré	VP	FN
	amer	FP	VN
(A) Théorique			

		Prédit	
		sucré	amer
Vrai	sucré	253	0
	amer	3	540
(B) Apprentissage			

		Prédit	
		sucré	amer
Vrai	sucré	54	9
	amer	4	133
(C) Validation externe			

TABLE 2.2 – Tableau de contingence

Prédiction de la toxicité

Enfin, la toxicité est évaluée par des approches *in silico*. Le logiciel FAF-drugs3 [94] permet de calculer les propriétés ADME-Tox (Absorption, Distribution, Métabolisme et Élimination - Toxicité) de molécules à partir de leur représentation 2D. Les filtres

physico-chimiques de FAF-drugs3 ne sont pas utilisés, car ils se basent sur des propriétés ADME qui n'ont pas d'intérêts ici. Mais des règles sont appliquées à partir de la description de la `SweetenersDB` dans le paragraphe 2.2.3. Chacune des propriétés forme un filtre pour lequel les composés sélectionnés doivent être inclus, avec une marge de tolérance de 10%. Le filtre « Tox » du logiciel FAF-drugs3 permet d'éliminer les composés présentant des groupements chimiques toxiques, appelés toxicophores.

Finalement, 4 585 composés naturels ont passé l'ensemble des filtres. Ce sont de potentiels édulcorants intenses avec un logPS supérieur à 2. Ils sont prédits sucrés, non amers et ne présentent pas de groupement toxique connu.

2.4 Conclusion

Les résultats présentés lors de ce chapitre montrent que les méthodes de prédiction de type régression non linéaire sont pertinentes pour prédire le logPS. La méthode des SVM est choisie, car elle est performante sur le jeu de validation externe ($R^2=0,83$) et elle permet de prédire plus efficacement les composés avec un logPS élevé. La description des molécules au format 2D est privilégiée au modèle 3D pour des raisons de temps et coût de calcul puisqu'ils présentent des performances similaires. Le protocole de sélection de composés naturels potentiellement sucrés est décrit dans la figure 2.3. Le logPS est prédit sur les composés suffisamment proches d'un point de vue chimique de la `SweetenersDB`, donc inclus dans le domaine d'applicabilité du modèle QSAR. Les composés définis comme « édulcorants intenses » (si le logPS est supérieur à 2) sont sélectionnés. Un ensemble de filtres est appliqué pour éliminer les composés potentiellement amers ou toxiques, permettant de passer de 75 000 à 4 585 composés d'intérêts. Ainsi, les molécules issues de la `SuperNatural II` sélectionnées par le protocole ont des logPS allant de 2 à 4 (figure 2.7). La figure 2.8 représente les familles chimiques dont sont issues ces 4 585 composés. En tant que base de données de composés naturels, il n'est pas surprenant qu'une grande majorité des composés prédits sucrés soient

des di et triterpènes (environ 96,2%). Enfin, 3,4% sont des phenylpropanoïdes et 0,4% des molécules diverses (oses, AA, ...). Actuellement, 3 nouvelles molécules issues de cette recherche sont en cours de test *in vitro*.

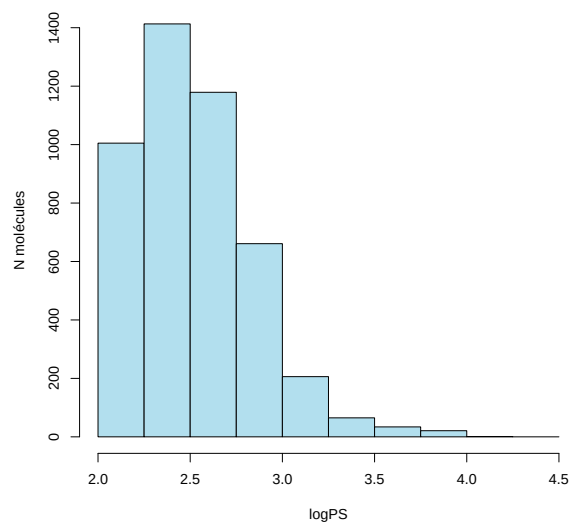


FIGURE 2.7 – Histogramme des logPS prédits sur les composés de la SuperNatural II

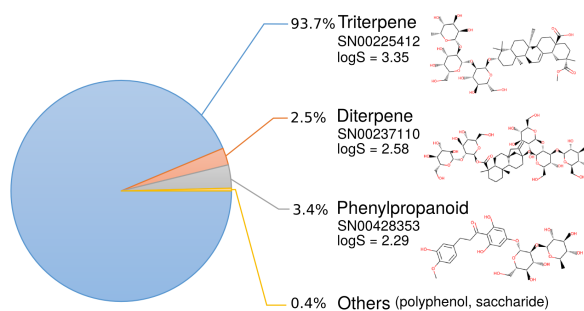


FIGURE 2.8 – Proportions des familles chimiques mises en évidence par le modèle QSAR à partir la base de données de composés naturels.

Chapitre 3

Structure du récepteur T1R2/T1R3

Ce chapitre est basé sur la publication suivante :

Chéron, J.-B., Golebiowski, J., Antonczak, S., & Fiorucci, S. (2017). The anatomy of mammalian sweet taste receptors. *Proteins : Structure, Function, and Bioinformatics*. <http://doi.org/10.1002/prot.25228>

3.1 Préambule : une échelle atomique

La détermination de la structure 3D des protéines est un enjeu majeur pour un grand nombre de disciplines, allant de la chimie à la biologie, notamment pour le développement de nouveaux ligands ou la compréhension de mécanismes moléculaires. Ces informations tridimensionnelles permettent une étude approfondie de la structure, au niveau atomique, pour déterminer et caractériser les sites d'interactions d'une molécule (ligand) avec son récepteur (protéine). Elles constituent également un point de départ pertinent pour des études dynamiques, que ce soit pour l'étude de réactions chimiques ou l'observation de phénomènes physiques.

Un accroissement du nombre de structures expérimentales disponibles est observé

depuis quelques années. En effet, les méthodes expérimentales pour l'élucidation structurale des protéines sont de plus en plus efficaces. La cristallographie aux rayons X reste la méthode la plus utilisée pour déterminer la structure de protéines telles que les RCPG. La première structure cristallographique de RCPG date des années 2000 [95]. Aujourd'hui, une centaine de structures de protéines de cette famille sont résolues. Les difficultés pour cristalliser les protéines membranaires résident dans leur expression [96], leur solubilisation en vue de leur cristallisation et dans l'existence de multiples états conformationnels [97].

Pour les RCPG de classe C (cf classification au chapitre 1.3.3 p 25), la présence de l'ectodomaine rend presque impossible l'utilisation des méthodes actuellement connues pour obtenir la structure du récepteur complet. L'élucidation structurale doit se faire par domaine (VFD¹, CRD², TMD³). À l'heure actuelle, aucune donnée structurale expérimentale n'est disponible pour le STR⁴. L'objectif de ce chapitre est d'exposer la stratégie de « modélisation par homologie » qui va permettre d'obtenir un modèle théorique. Dans un premier temps, les données structurales et les données de mutagenèse dirigée sur les RCPG de classe C seront collectées pour réaliser un alignement de séquences en AA. Cet alignement sera ensuite utilisé pour modéliser la structure du STR, qui est un point crucial pour décrire les AA clés par des approches statiques et des études dynamiques décrits au chapitre 3.

3.2 Des atomes à la « protéine »

Les protéines sont des biomolécules qui assurent une multitude de fonctions au sein de la cellule (structuration, transport, stockage, réception de signaux, ...). La majorité des protéines connues se compose de 20 briques élémentaires que sont les AA. Le carbone nommé C_α porte les fonctions chimiques communes, amine primaire et

-
1. *Venus Flytrap Domain*
 2. *Domaine Riche en Cytéine*
 3. *Domaine TransMembranaire*
 4. *Sweet Taste Receptor*

acide carboxylique. Les AA diffèrent par l'enchaînement des atomes qui constituent la chaîne latérale, elle aussi portée par le C_α . La chaîne latérale définit les propriétés physico-chimiques de l'AA. Une catégorisation selon leurs propriétés est reportée dans la figure 3.1.

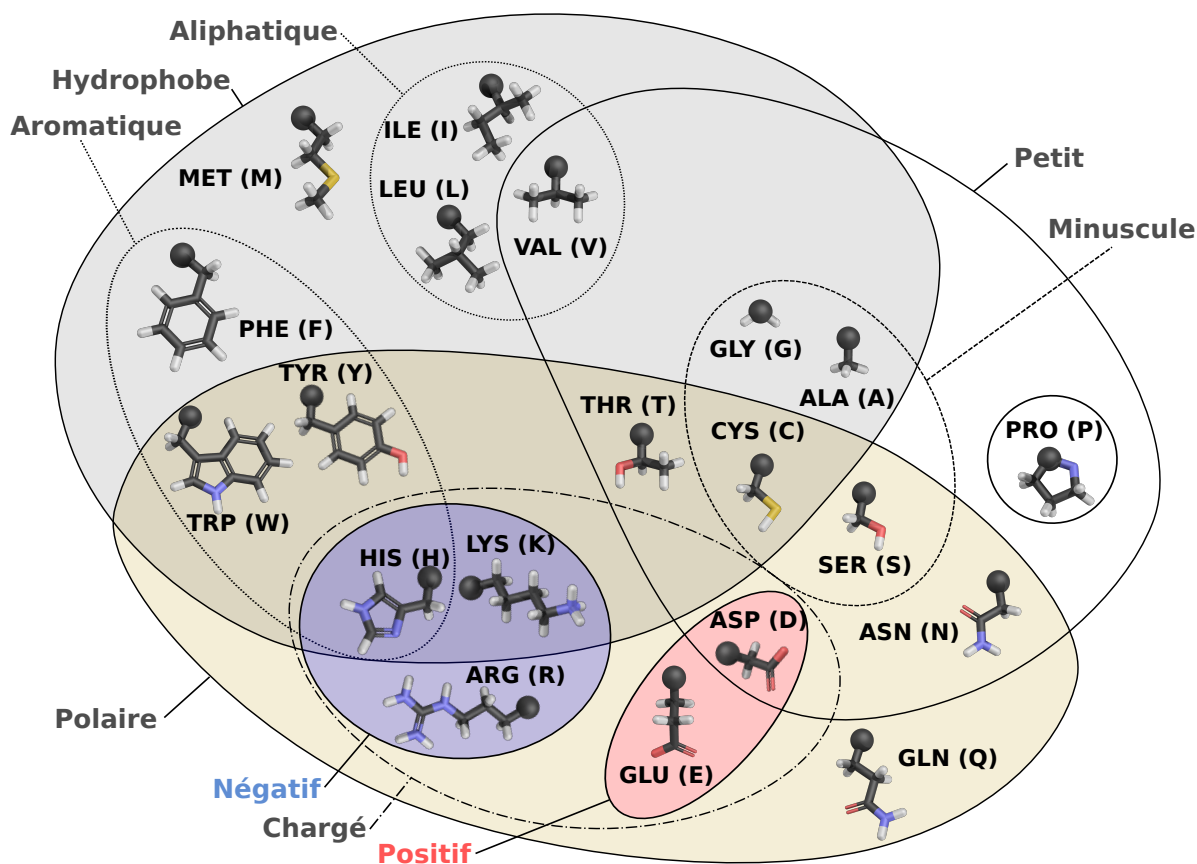
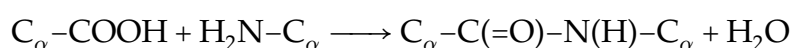


FIGURE 3.1 – Diagramme de Venn qui regroupe les AA par leurs propriétés physico-chimiques

Les AA sont représentés sous forme de « bâtons ». L'atome C_α de chaque AA est représenté par une sphère, les atomes des fonctions communes acides carboxyliques et amines primaires ne sont pas représentés. Les atomes de carbone, d'hydrogène, d'azote, d'oxygène et de soufre sont colorés, respectivement, en noir, blanc, bleu, rouge et jaune.

L'enchaînement des AA est réalisé par la formation de liaisons covalentes, dites « liaison peptidique ». Elles se forment par une réaction chimique entre l'acide carboxylique et l'amine primaire de deux AA pour former une fonction amide :



La configuration *trans* de la liaison peptidique est largement favorisée à cause de l'encombrement stérique des chaînes latérales. Cet assemblage d'AA permet de définir plusieurs niveaux d'organisation. La structure primaire correspond à la séquence en AA qui forme la protéine. Les propriétés physico-chimiques des AA permettent la formation d'interactions non liantes (liaisons hydrogène, interactions ioniques, hydrophobes, ...). Les AA interagissent entre eux pour établir une structure secondaire, premier niveau d'organisation tridimensionnelle, qui correspond aux feuillets et hélices (figure 3.2). L'agencement des structures secondaires permet le repliement global de la protéine pour obtenir sa structure tertiaire. Enfin, l'assemblage de plusieurs protéines forme la structure quaternaire. Il est important de noter que le repliement est influencé par le solvant. Dans le cas de l'eau, les AA apolaires auront tendance à se trouver au cœur de la protéine. De plus, si la protéine est enchâssée dans une membrane lipidique, ce sont les AA apolaires qui seront orientés vers l'extérieur.

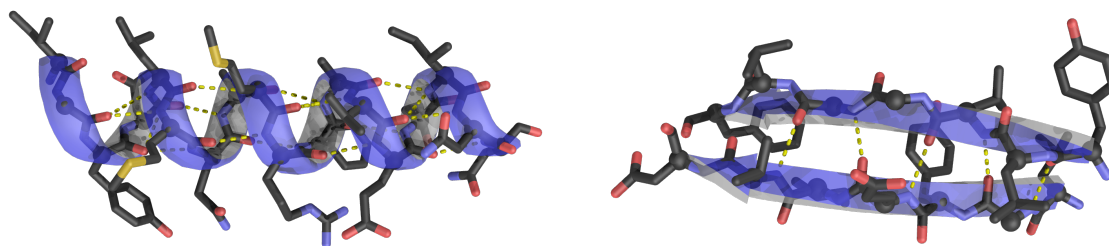


FIGURE 3.2 – Structure secondaire

Représentation d'une structure secondaire type « hélice » à gauche et « feuillet β » à droite. Les interactions non-liantes qui maintiennent la structure secondaire sont représentées en pointillés jaunes.

Dans le cas de récepteurs, comme le STR, le repliement des AA forme des cavités susceptibles d'accueillir des molécules. Il peut exister un ou plusieurs sites de liaison par protéine. La forme et la composition en AA de ces cavités permettent la reconnaissance de molécules spécifiques. La liaison d'une molécule peut avoir de multiples conséquences, selon le rôle de la protéine. Pour le STR, la fixation d'un sucre ou d'un édulcorant entraîne des modifications structurales qui permettent la transmission de l'information portée par le signal chimique à l'intérieur de la cellule.

3.3 Données disponibles pour la modélisation du récepteur T1R2-T1R3

Aucune donnée structurale expérimentale n'est disponible pour le STR. La compréhension du lien qui existe entre la séquence en AA et la structure d'une protéine est un défi dans le domaine de la recherche en bioinformatique. Ainsi, de nombreuses méthodes et de nombreux outils de prédiction sont aujourd'hui disponibles. Les méthodes peuvent être classées en quatre catégories [98] :

- *ab initio* : elles se basent sur les lois de la physique pour prédire le repliement des protéines.
- *ab initio* avec utilisation de données précompilées : elles utilisent en plus des informations tels que les motifs structuraux.
- par reconnaissance de repliement basée sur des notions d'évolution de séquences en AA.
- par homologie de séquences en AA (méthode utilisée pour cette étude).

Pour appliquer la méthode de modélisation par homologie, il est nécessaire d'avoir la structure expérimentale d'une protéine dont la séquence en AA est proche, d'un point de vue évolutionnaire, de celle à modéliser. À l'instar des autres méthodes, le premier postulat pour déterminer la structure d'une protéine est que, en théorie, seule la séquence en AA suffit. Le second est plus spécifique à la modélisation par homologie : il a été observé qu'au cours de l'évolution, la structure est plus conservée que la séquence. Des protéines dont les séquences en AA ont un pourcentage d'identité⁵ ou de similarité⁶ élevé possèdent la même structure 3D. Ainsi, pour obtenir un modèle par homologie il faut réaliser un alignement de séquences en AA entre la protéine d'intérêt et une ou plusieurs protéines dont les structures expérimentales sont connues. Ensuite, le modèle est construit à partir de ces patrons, puis sa géométrie doit être validée. Les

5. pourcentage d'AA identiques entre les deux séquences après alignement

6. pourcentage d'AA qui partagent les mêmes propriétés physico-chimiques après alignement

détails méthodologiques sont donnés en annexe B.2. Avant toute chose, il est nécessaire d'énumérer les données disponibles qui vont permettre de construire un modèle 3D du STR.

3.3.1 La PDB

La PDB (acronyme de *Protein Data Bank*, <http://www.rcsb.org/>) centralise l'ensemble des structures 3D de protéines et d'acides nucléiques résolues expérimentalement [99]. Elle contient $\approx 125\,000$ structures identifiées par un code à 4 caractères alphanumériques (nommé par la suite « code PDB »). À cette date, ≈ 180 structures de RCPG sont disponibles, incluant tous les organismes et toutes classes confondues. Ces données expérimentales constituent un point de départ robuste pour proposer des modèles de RCPG dont la structure est inconnue.

3.3.2 Les structures 3D de RCPG

Comme énoncé précédemment, la structure est plus conservée que la séquence au cours de l'évolution. Plus les protéines sont phylogénétiquement proches, plus le modèle proposé est pertinent. Pour obtenir un modèle du STR le plus réaliste possible, il est judicieux de choisir comme patron un RCPG et, de surcroît, de classe C. Il est important de rappeler que chaque sous-unité du STR est organisée en trois domaines : l'ectodomaine N-terminal extra-cellulaire comprenant le VFD et le CRD, ainsi la partie C-terminale transmembranaire (TMD) (figure 1.8 p 29). Les structures cristallographiques disponibles sont référencées dans le tableau 3.1. De nombreuses structures du VFD des mGluR et des GABA_{B1} et _{B2} ont été résolues depuis les années 2000. Seules quelques structures possèdent le CRD, région peu structurée qui fait le lien avec la partie transmembranaire. Enfin, la prédiction de la structure du TMD était jusqu'à présent réalisée en utilisant des RCPG de classe A comme patron [100–102]. Récemment, deux structures cristallographiques du TMD de RCPG de classe C ont été publiées et

semblent plus appropriées pour modéliser cette région du STR. Une comparaison des deux approches est exposée à la fin de ce chapitre.

Domaine	Récepteur	code PDB et référence
VFD ^a	mGluR1	[1EWK, 1EWT, 1EWV] [103] [1ISS, 1ISR] [104] [3KS9]*
	mGluR2	[5CNI, 5CNJ, 5CNK, 5CNM] [105]
	mGluR3	[2E4U, 2E4V, 2E4W, 2E4X, 2E4Y, 2E4Z] [106] [3SM9]*
	mGluR5	[3LMK]*
	mGluR7	[3MQ4]*
	GABA _{B1etB2}	[4MQE, 4MQF, 4MR7, 4MR8, 4MR9, 4MRM, 4MS1, 4MS3, 4MS4] [107]
TMD	mGluR1	[4OR2] [108]
	mGluR5	[4OO9, 5CGC, 5CGD] [109]
*pas de publication associée		

TABLE 3.1 – Liste des structures expérimentales des RCPG de classe C.

a. De nouvelles structures du VFD du mGluR ont été résolues pendant ces travaux de thèse.

3.3.3 Données de mutagenèse dirigée

Les données de mutagenèse dirigée permettent, tout d'abord, d'ajuster l'alignement de séquences en AA. Ensuite, elles présentent un grand intérêt pour valider le modèle par homologie en vérifiant si les AA en interaction avec les ligands sont correctement orientés dans la cavité de liaison. Le référencement de plus de 600 mutations (tableau en annexe A.2) a été effectué et a été reporté sur les alignements de séquences en AA des RCPG de classe C.

3.4 Construction des modèles de T1R2 et T1R3

Les séquences en AA des RCPG de classe C humains sont issues d'*Uniprot* [110]. Elles sont ensuite sub-divisées en domaine (VFD, CRD, TMD). Un premier alignement multiple est réalisé avec *Clustal Omega* et est ensuite ajusté manuellement à l'aide de contraintes expérimentales. Une première contrainte issue des structures secondaires

des données cristallographiques disponibles (tableau 3.1) est appliquée sur l'alignement. Enfin, quelques ajustements sont effectués pour satisfaire les données de mutagenèse dirigée référencées dans le tableau A.2 en annexe.

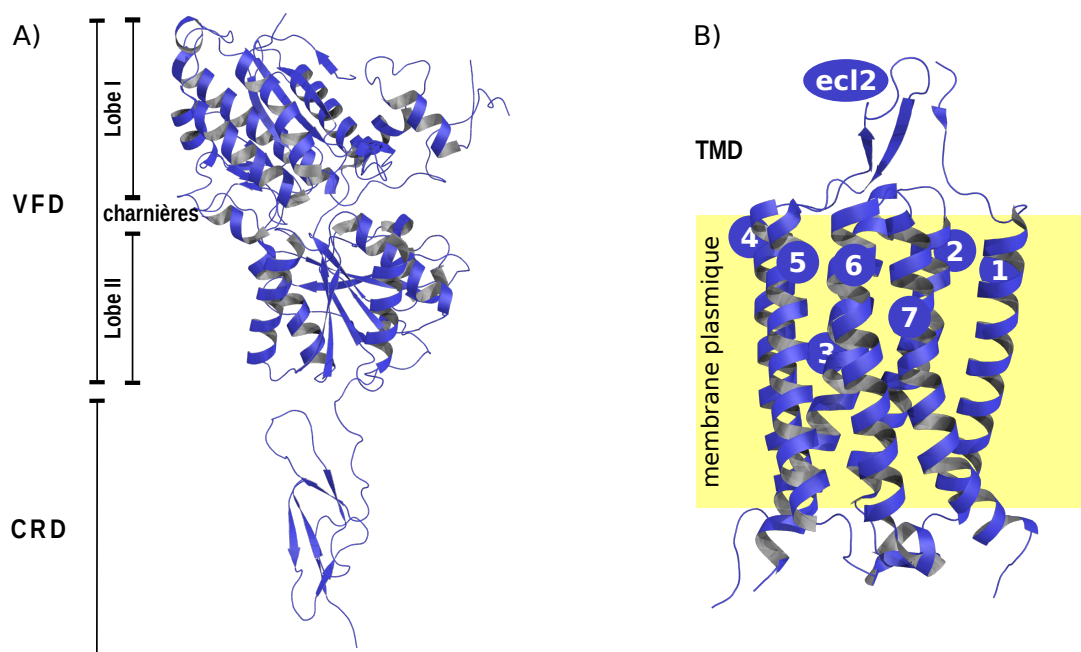


FIGURE 3.3 – Représentation de structures cristallographiques du mGluR

A) Structure de la partie extracellulaire du mGluR3, code PDB 2E4U
 B) Structure de la partie transmembranaire du mGluR1, code PDB 4OR2

VFD

Le VFD constitue un large domaine N-terminal du STR. Il est formé par une séquence d'environ 400 AA et, d'un point de vue structural, il présente deux lobes réunis par trois charnières (figure 3.3A). Les lobes supérieur et inférieur sont nommés respectivement lobe I et lobe II. C'est entre ces deux lobes que se trouve le site orthostérique, où se fixent les ligands « naturels » des RCPG de classe C. Le VFD existe sous deux formes, ouvert « o » ou fermé « f ». La forme « o » correspond à la forme libre du VFD et la forme « f » est stabilisée par l'interaction avec un agoniste [111]. Les VFD présentent deux à trois ponts disulfures intra-domaine. Une cystéine dans le lobe I des

mGluR permet de stabiliser les formes dimériques (homo- ou hétéro-dimères). Enfin, une dernière cystéine permet de lier le VFD au CRD.

L'alignement de séquences en AA des VFD est représenté dans la figure 3.4. La superposition structurale des données cristallographiques des VFD de mGluR1, 2, 3, 5, et 7, GABA_B1 et 2 permet de déterminer les régions où la topologie est conservée. Les séquences en AA qui définissent des régions structurées de type hélice α et feuillet β sont représentées, respectivement, en rouge et en vert sur l'alignement. L'alignement final révèle un pourcentage d'identité de séquence modéré entre les VFD :

- 29±1% d'identité de séquences entre les T1R et les mGluR
- 39% d'identité de séquences entre T1R2 et T1R3

L'ensemble des identités de séquences en AA associé à l'alignement de la figure 3.4 est référencé dans le tableau A.3 en annexe. Le recensement des données de mutagenèse dirigées permet d'ajuster l'alignement en prenant en compte l'aspect fonctionnel des AA. Les positions des AA impliqués dans la reconnaissance et la liaison des ligands sont extrapolées sur l'alignement de séquences en AA. Les AA cités par Wellendorph *et al.* [79] formant le site de liaison orthostérique des RCPG de classe C sont également référencés et utilisés pour ajuster l'alignement multiple. Cet alignement est utilisé par *Modeller* (annexe B.2) pour proposer des modèles par homologies des VFD de T1R2 et T1R3. Les structures cristallographiques du mGluR1 utilisées comme patrons sont 1EWT pour la forme « o » et 1EWK pour la forme « f ».

CRD

Le CRD fait suite au VFD (figure 3.3A). C'est un domaine qui ne présente pas de structure secondaire particulière mais qui, comme son nom l'indique, contient de nombreuses cystéines pour stabiliser sa structure par la formation de ponts disulfures. Pour une longueur d'environ 70 AA, il présente neuf cystéines qui permettent de former quatre ponts disulfures intra-domaine et un autre avec le VFD. Il faut noter que ce domaine est absent chez les récepteurs GABA_B.



Les structures secondaires sont soulignées en vert (feuillet) et en rouge (hélices). Elles ont été extrapolées à partir des structures cristallographiques de mGluR1, mGluR3, mGluR5, mGluR7 et GABA_B 1 et 2. Les interfaces protéine-protéine sont également issues des données de cristallographie de mGluR1, mGluR3, mGluR5 et GABA_B. Les AA appartenant aux lobes I et II sont soulignés en violet et en rose, respectivement. Les AA surlignés en noir correspondent à la cavité de liaison orthostérique proposée par Wellendorph et Bräuner-Osborne. Les données de mutagenèse dirigée réalisées sur T1R1, T1R2, mGluR1, mGluR2, mGluR3, mGluR4, mGluR5, mGluR8 et GABA_B qui ont un impact sur la fonction sont représentées en vert et celles qui n'ont pas d'impact en gris.

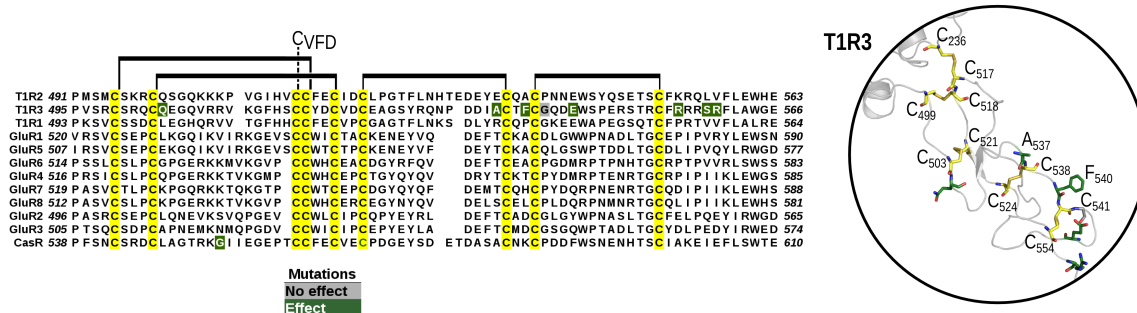


FIGURE 3.5 – Alignement des séquences en AA des CRD des RCPG de classe C

Malgré un pourcentage d'identité de séquence modéré, $31 \pm 2\%$ entre les T1R et les mGluR, l'alignement des CRD est fortement influencé par la présence des cystéines hautement conservées (figure 3.5). Les ponts disulfures sont contraints lors de la construction du modèle par homologie. La structure cristallographique du mGluR3 est utilisée comme patron.

TMD

Le TMD constitue la région prototypique de la superfamille des RCPG. Il contient ≈ 300 AA qui se replient pour former sept hélices qui traversent la membrane. Elles sont nommées TM1 à 7 (figure 3.3B). Ces structures secondaires sont reliées par des boucles généralement non structurées. Les boucles extracellulaires sont notées « ECL » et intracellulaire « ICL », comme suit :

(N-terminal) TM1 – ICL1 – TM2 – ECL1 – TM3 – ICL2 – TM4 – ECL2
– TM5 – ICL3 – TM6 – ECL3 – TM7 (C-terminal)

Un pont disulfure extrêmement conservé dans la superfamille des RCPG est retrouvé dans les classes C, il permet de lier l'ECL2 avec le haut de la TM3.

La topologie du TMD des RCPG, commun à toute classe, permet d'établir une nomenclature particulière. En effet, il est communément admis de définir un AA par sa

position dans la séquence de la protéine. Cependant, cette dénomination est peu judicieuse pour situer et comparer des AA dans plusieurs séquences. Ainsi, la numérotation dite de « Ballesteros-Weinstein », basée sur les AA les plus conservés des RCPG, a donc été définie par les auteurs du même nom. Elle consiste, pour chaque hélice, à définir l'AA le plus conservé dans un alignement de séquence en AA multiple et de lui attribuer le chiffre « .50 ». Le chiffre avant le « . » correspond à l'hélice où se situe l'AA. Par exemple, pour la TM1, le résidu X le plus conservé sera noté $X^{1.50}$. Ainsi, tous les résidus en amont ont un numéro inférieur à 50 et les résidus en aval supérieur à 50. Les boucles marquent la fin de cette numérotation qui est propre à chaque hélice. Cette nomenclature sera largement utilisée dans la suite de ce manuscrit.

L'alignement des séquences des AA des TMD est représenté dans la figure 3.6. Cette région prototypique des RCPG présente sept hélices transmembranaires qui sont reportées sur l'alignement. La topologie issue des données structurales des TMD résolues par cristallographie est une contrainte forte pour ajuster l'alignement de séquence. En effet, la présence de « trous » dans ces régions critiques a des conséquences sur l'orientation des chaînes latérales des AA. Un tour d'hélice α compte quatre AA, donc l'insertion d'un trou dans l'alignement décalera l'ensemble des AA d'un quart de tour. Les données de mutagenèse dirigée (tableau A.2 en annexe) permettent de réaliser quelques ajustements au niveau de la cavité de liaison et de proposer un alignement plus pertinent. Enfin, l'analyse de la structure du mGluR faite par Doré *et al.* [109] permet de contraindre l'alignement en se basant sur des régions clés impliquées dans la fonction du récepteur, appelées « interrupteurs moléculaires » (détaillées au chapitre 3.5.3). L'élucidation récente des structures de RCPG de classe C mGluR1 [108] et mGluR5 [109] en fait les candidats parfaits pour proposer un modèle du TMD du STR. La structure du mGluR1, code pdb 4OR2, est utilisée comme patron pour modéliser les TMD de T1R2 et T1R3. Une contrainte est appliquée pour lier les deux cystéines qui forment le pont disulfure canonique entre l'ECL2 et le haut de la TM3.

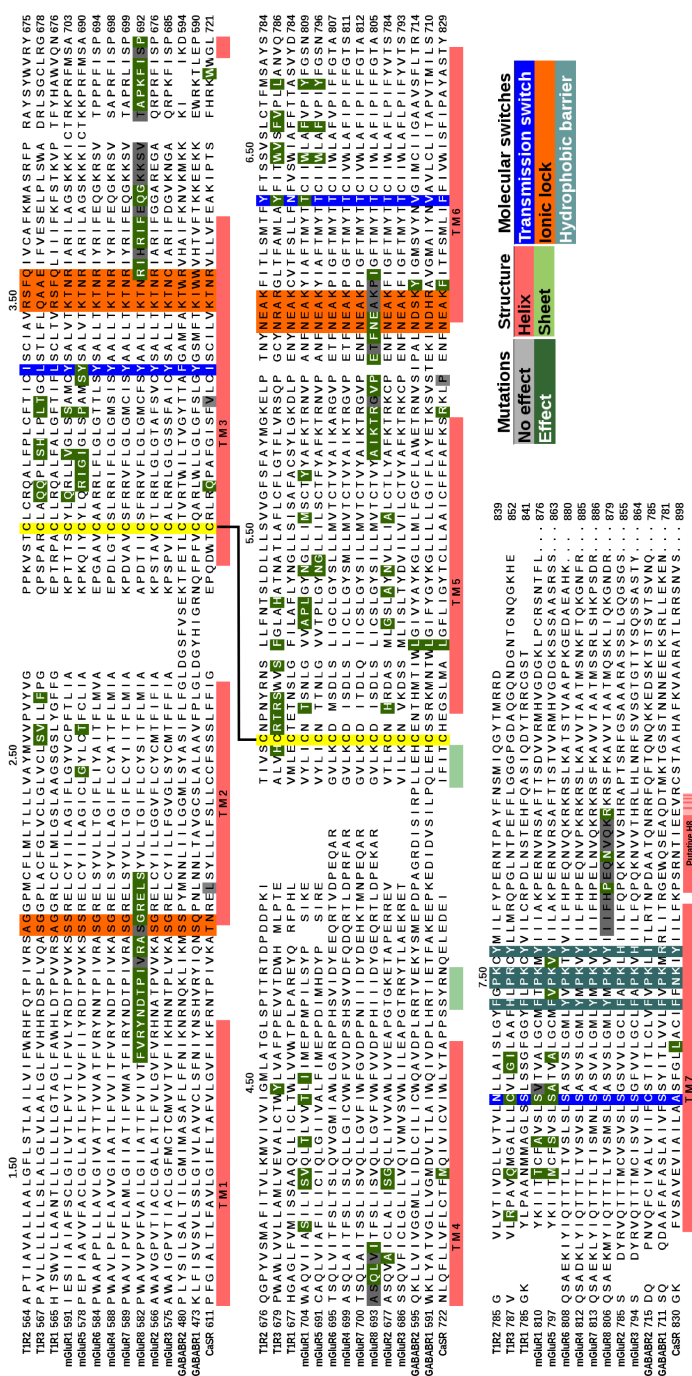


FIGURE 3.6 – Alignement des séquences en AA des TMD des RCPG de classe C

Les structures secondaires déterminées à partir des structures cristallographiques du mGluR1 et mGluR5 sont représentées en rouge (hélices) et vert (feuillets). Les AA dont les mutations qui ont un effet sur la fonction du récepteur sont surlignées en vert et celles qui n'en ont pas en gris. Les « interrupteurs moléculaires » sont représentés en orange, bleu et gris-bleu. La définition de Pin *et al.* de la numérotation de Ballesteros-Weinstein pour les RCPG de classe C est représentée.

3.5 Description moléculaire de T1R2 et T1R3

3.5.1 VFD

Le VFD est le domaine qui a été le plus étudié pour les RCPG de classe C, que ce soit pour les mGluR ou le STR. Afin de réaliser des études structurales, de nombreux alignements de séquences en AA ont été proposés pour construire un modèle par homologie du VFD [112–115]. Les zones les plus critiques de l'alignement sont présentes dans les régions non structurées. En se basant sur les données de mutagenèse dirigée et les structures secondaires issues des données cristallographiques, seulement quelques différences sont observées entre l'alignement proposé dans cette étude et ceux déjà publiés. Le résidu K65, par exemple, est impliqué dans le maintien du VFD à l'état fermé [114–116]. Pour satisfaire cet aspect fonctionnel, l'AA K65 précédemment aligné avec le résidu Q/E des mGluR est aligné ici avec les lysines présentes en $n - 2$. Aussi, plusieurs modèles et analyses structurales de T1R2 ont été effectués pour confirmer la position des AA Q328 et D213 dans l'alignement de séquences en AA (figure 3.4).

L'alignement et le modèle par homologie permettent de décrire les AA impliqués dans la liaison des sucres et des édulcorants. Les AA surlignés en noir dans la figure 3.4 définissent les AA formant la cavité de liaison, ils sont issus d'une étude de Wellendorph et Bräuner-Osborne [79]. Les résultats de mutagenèse dirigée sur T1R2 confirment l'importance de ces AA dans le site de liaison orthostérique.

Le VFD présente une structure bilobée avec une large cavité de liaison. En utilisant le mGluR1 en forme « o » comme patron, les modèles structuraux de T1R2 et T1R3 présentent une cavité ayant un volume de $\approx 4\,900\text{ \AA}^3$. Dans les deux cas, elles sont hydrophiles, avec un ratio de la surface polaire sur la surface totale de 45% pour T1R2 et de 50% pour T1R3. Cette grande polarité est en accord avec le type de ligand que peuvent fixer les VFD de T1R2 et T1R3, tel que les sucres, dont les propriétés physico-chimiques ont été énoncées dans le chapitre 1.3.2. Une cavité annexe a été mise en évidence à partir d'une étude structurale du mGluR4 [117]. Ce type de cavité, adjointe

à la principale, trouve toute son importance pour la conception de nouvelles molécules fixant de manière plus spécifique et efficace le récepteur. Cette dernière est retrouvée sur les modèles proposés et semble avoir une importance dans la liaison d'édulcorant, notamment pour T1R2, où la mutation S40A diminue la réponse à l'aspartame (figure 3.7).

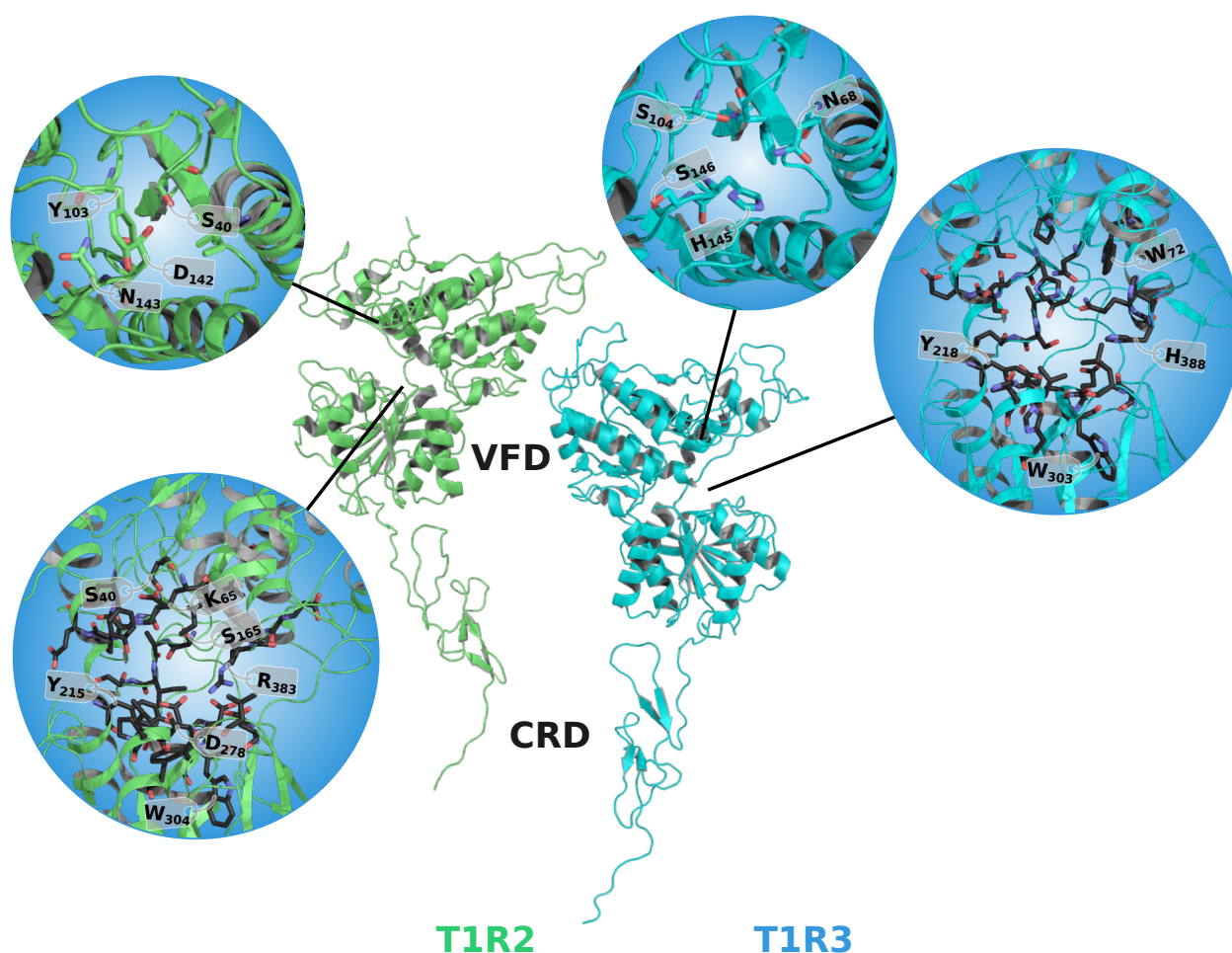


FIGURE 3.7 – Modèles par homologie des VFD de T1R2 et T1R3

Les AA mis en évidence sont les plus conservés parmi ceux décrits comme importants par les données de mutagenèse dirigée.

3.5.2 CRD

La structure du CRD est représentée dans la figure 3.5. Les ponts disulfures qui structurent cette région ont un agencement particulier. Le chevauchement des ponts

disulfures, ainsi que la liaison d'une cystéine au bas du VFD rigidifient fortement cette région. Deux d'entre eux font intervenir des cystéines consécutives dans la séquence en AA et permettent de restreindre les mouvements d'élongation du CRD. Malgré une séquence d'environ 70 AA, la structure du CRD qui ne présente pas de structure secondaire particulière reste compacte. Ces observations confirment ce qui a été suggéré, ces ponts disulfures servent à rigidifier le couplage VFD-TMD [79]. La fin de séquence du CRD, comptant une dizaine d'AA n'est, quant-à elle, pas contrainte et autorise quelques degrés de liberté dans l'association VFD-TMD.

3.5.3 TMD

Site de liaison

Le TMD des RCPG de classe C possède un site de liaison orthostérique. Une étude sur le mGluR5 [118] montre que ce domaine est fonctionnel *in vitro* même sans l'ectodomaine extracellulaire. Pour le STR, T1R3 est la seule sous-unité pour laquelle des données de mutagenèse dirigée sont disponibles [100–102]. Il est capable de lier des modulateurs allostériques négatifs (lactisole) et positifs (cyclamate et néohesperidine dihydrochalcone (NHDC)). Pour les petits ligands que sont le cyclamate et le lactisol, les données de mutagenèse montrent que la cavité de liaison du TMD de T1R3 est comprise entre les TM 3, 4, 6 et qu'elle comprend également l'ECL2 [100, 101]⁷. Pour le NHDC, les hélices 2, 4 et 7 s'ajoutent à cette cavité, incluant les AA : S₆₂₀^{2.52}, F₆₂₄^{2.56}, Y₆₉₉^{4.50} et C₈₀₁^{7.39} [102]. La construction de chimères humain/murin a montré que le TMD de T1R2 est aussi en mesure de lier des composés, tels que la périllartine, le P-4000 et le S-819 [81, 112]. L'absence de données de mutagenèse pour T1R2 rend impossible la description expérimentale de cette cavité de liaison. Cependant, pour la suite de la comparaison, elle peut être extrapolée à partir des AA décrits dans T1R3 et les mGluR. Des variations dans les AA orientés dans la cavité de liaison sont observées entre T1R2

7. La description de la cavité du TMD de T1R3 est plus largement détaillée dans le chapitre 4.

et T1R3 et peuvent expliquer la différence de reconnaissance des composés précédemment cités. Le résidu S₆₄₀^{3.36} dans T1R3 est remplacé par une F₆₃₇^{3.36} dans T1R2 suggérant un encombrement plus important de la cavité de liaison. Ainsi, les volumes des cavités estimés de T1R2 et T1R3 sont, respectivement, 210Å³ et 270Å³. Trois AA sont décrits comme point d'ancrage dans T1R3 : H₆₄₁^{3.37}, R₇₂₃^{ECL2} et F₇₇₈^{6.53}. Ils stabilisent la liaison du cyclamate par des interactions de type électrostatique et π - π *stacking* [100]. Les propriétés de ces AA sont absentes dans T1R2, car selon l'alignement, ils sont remplacés par les AA P₆₃₈^{3.37}, N₇₂₁^{ECL2} et L₇₇₆^{6.53}. L'ensemble de ces observations peuvent expliquer pourquoi les molécules fixant T1R2 et T1R3 ne peuvent pas partager les mêmes propriétés physico-chimiques.

Les interrupteurs moléculaires

Les interrupteurs moléculaires sont des motifs structuraux impliqués dans la transmission du signal induit par la liaison du ligand. Ces AA ou séquences courtes d'AA ont été largement décrits pour les RCPG de classe A [95, 119–121]. Des motifs similaires ont été identifiés chez les RCPG de classe C [122], notamment dans les récentes structures cristallographiques du mGluR [108, 109]. Ils ont été reportés dans l'alignement de séquences en AA sur la figure 3.4. Une forte conservation est observée pour les mGluR. Les motifs correspondants pour le STR sont extrapolés à partir de l'alignement de séquences en AA, ils sont au nombre de trois :

- Le commutateur de transmission ou « transmission switch » est constitué d'AA formant le fond de la cavité de liaison et sert de premier senseur de la fixation d'un ligand.
- Le verrou ionique ou « ionic lock » se trouve du côté intracellulaire et est formé d'AA polaires ou chargés dans les TM3 et TM6. Leur interaction a un rôle de verrou quand le RCPG est à l'état inactif.
- La barrière hydrophobe ou « hydrophobic barrier » est localisée dans la TM7 et

prévient l'entrée d'eau dans la partie intracellulaire du RCPG dans son état inactif. L'inondation de cette région permet le couplage de la protéine G et l'activation de la cascade de signalisation.

Le *Transmission Switch*, le *Ionic Lock* et l'*Hydrophobic Barrier* seront notés respectivement TS, IL et HB par la suite.

Le TS des RCPG de classe C est formé par une triade d'AA aux positions 3.44, 6.46 et 7.39, selon la nomenclature de Ballesteros-Weinstein. La figure 3.8 résume les AA impliqués dans cet interrupteur moléculaire selon le récepteur. Même si les AA diffèrent entre les mGluR et au sein des T1R, il est important de noter que les propriétés physico-chimiques des AA sont conservées. Par exemple, la tyrosine du mGluR dans la TM3 est présente dans la TM6 pour T1R2 et T1R3 (Y^{3.44} et Y^{6.46}, respectivement). Pour les RCPG de classe A, un seul AA est impliqué, il s'agit du W^{6.48}. Un AA tryptophane très conservé est également retrouvé dans les classes C (W^{6.50}) mais il ne délimite pas le fond de la cavité. Il est possible qu'il ait un autre rôle dans cette famille. Seul T1R2 ne présente pas de tryptophane en position 6.50 mais une sérine, suggérant qu'un aspect fonctionnel est peut-être manquant dans cette sous-unité du STR.

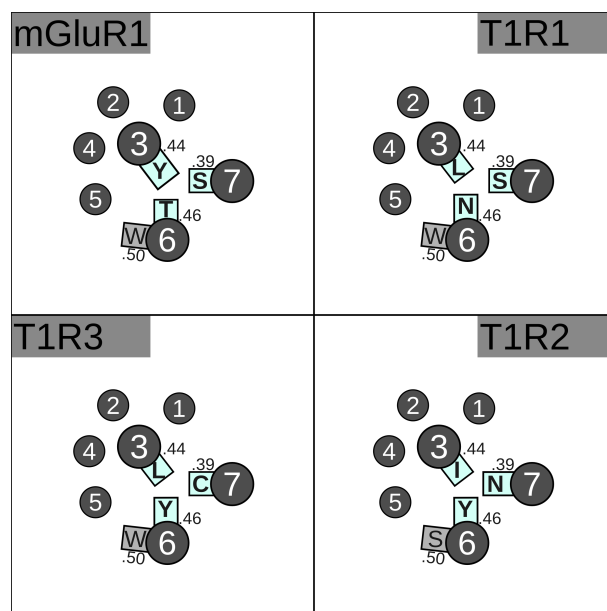


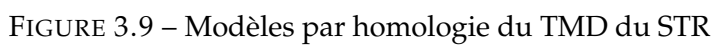
FIGURE 3.8 – Représentation schématisques des AA importants du TS sur les modèles de T1R2 et T1R3

Les motifs du IL sont très conservés chez les mGluR, les séquences en AA sont K^{3.50}TNR dans la TM3 et NEAK dans la TM6. Ce sont les chaînes latérales des AA K^{3.50} et E^{6.35} qui forment le verrou. Comme il a été montré dans les structures cristallographiques des mGluR, deux sérines de l'ICL1 et le résidu R^{3.53} s'ajoutent aux AA cités pour renforcer les interactions électrostatiques. Pour le STR, les AA correspondants sont représentés dans l'alignement de séquence (figures 3.6) et sur le modèle par homologie (figure 3.9). Le même schéma est retrouvé pour T1R2 et T1R1 avec les AA R^{3.50} et E^{6.35}. En revanche, la polarité de l'interaction est inversée pour T1R3 avec les AA Q^{3.50} et R^{6.35}. Même si l'alignement de séquences en AA peut révéler quelques variations, les modèles par homologie de T1R2 et T1R3 démontrent que des interactions similaires ont lieu entre la TM3 et la TM6 pour former le IL.

Dans les RCPG de classe C, un motif [F/Y] xPKxY conservé dans la TM7 forme la barrière hydrophobe. En se basant sur la structure du mGluR, l'acteur principal est la F ou la Y en position 7.48. Pour T1R3, une histidine est retrouvée à cette position. Pour les T1R, un autre AA aromatique semble renforcer cette barrière hydrophobe, la F^{7.47} pour T1R1/2 et la Y^{7.47} pour T1R3. Il est intéressant de noter que les AA impliqués dans les interrupteurs moléculaires sont présents dans les TM3, TM6 et TM7. Ces TM sont l'objet de l'étude réalisée dans le chapitre 5.

Choix du patron

Des modèles par homologie de T1R3 ont déjà été proposés dans des études précédentes [100–102]. L'absence de données structurales sur les RCPG de classe C a nécessité l'utilisation d'un RCPG de classe A comme patron, la rhodopsine bovine (code PDB 1F88) [95]. Deux alignements différents ont été réalisés pour construire les modèles [100, 102]. Dans les deux cas, un décalage des hélices transmembranaires est observé en comparant les deux modèles de T1R3 (figure 3.10), notamment pour les TM 2, 5 et 7. L'utilisation de la rhodopsine comme patron entraîne également une diminution du volume de la cavité de liaison. Le modèle basé sur la structure du mGluR conduit à



Les AA mis en évidence sont les plus conservés parmi ceux décrits comme importants par les données de mutagenèse dirigée. Ils représentent les AA de la cavité de liaison et des interrupteurs moléculaires.

une rotation d'un tour d'hélice de la TM5, sans conséquence sur la liste des AA impliqués dans la cavité de liaison (Y^{5.40} et H^{5.44}) mais important pour l'étude dynamique du récepteur. Enfin, un décalage d'un demi-tour d'hélice en fin de TM4 permet de replacer l'AA Y^{4.50} dans la cavité de liaison, mal positionnée avec un des alignements précédent [100]. Cet AA a été montré comme impliqué dans la liaison du NHDC [102].

Deux différences majeurs concernant les interrupteurs moléculaires sont à noter. Les études précédentes suggéraient que l'AA W^{6.50} des RCPG classe A soit aligné avec le W^{6.48} des classes C. Cependant, le faible pourcentage d'identité de séquence entre les classes A et le STR (14% avec la rhodopsine) ne permet pas d'affirmer que les interrupteurs moléculaires sont identiques. Dans les RCPG de classe C, l'AA W^{6.48} ne forme pas le fond de la cavité mais pointe vers la TM5. En se basant sur le mGluR, le modèle par homologie est donc largement affecté à cette position. De plus, chez les RCPG de classe A, le motif correspondant à la barrière hydrophobe est NPxxY [121]. La tendance des alignements précédents était d'aligner la Y^{7.53} des classes A avec la Y^{7.53} des classes C. Cependant, les structures cristallographiques du mGluR révèlent que c'est la première tyrosine (Y^{7.48}) du motif [F/Y]xPKxY qui est correctement orientée pour jouer le rôle de barrière hydrophobe. Cette différence dans l'alignement et dans le patron entraîne une grande variation dans la structure finale de la TM7 (figure 3.10).

L'utilisation d'un patron appartenant à la même famille de RCPG que le STR permet de réaliser un alignement de séquence en AA plus convaincant et d'obtenir un modèle structural fiable du STR.

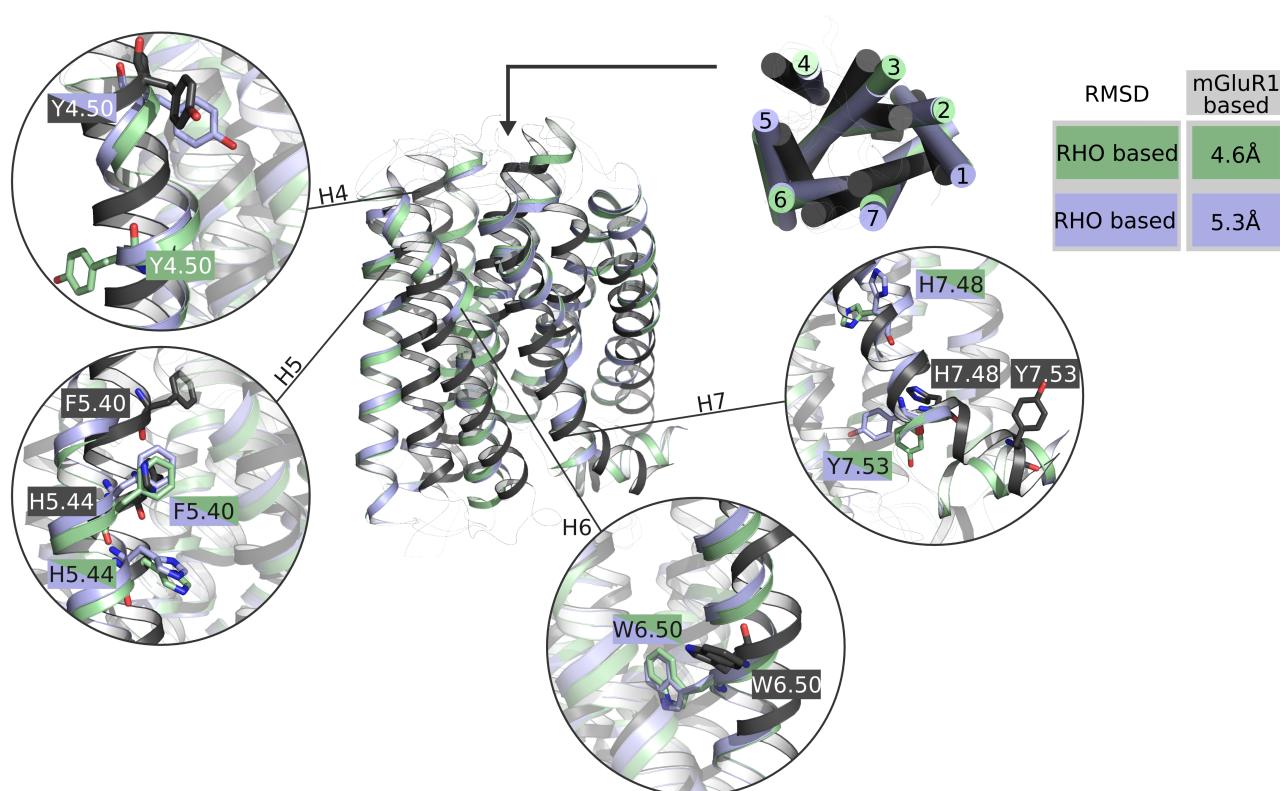


FIGURE 3.10 – Comparaison des modèles par homologie de T1R3 selon les alignements de séquence en AA et le patron utilisé.

En noir, le modèle par homologie proposé dans cette étude. Les modèles proposés par Jiang *et al.* [100] et Winnig *et al.* [102] sont représentés, respectivement, en vert et en bleu.

3.6 Conclusion

Le référencement de plus de 600 données de mutagenèse dirigée, ainsi que les topologies des RCPG de classe C extraites des structures cristallographiques ont permis d'ajuster les alignements pour proposer un modèle 3D fiable du STR. Des différences mineures sont à notées dans l'alignement de séquence en AA, donc dans la structure modélisée, du VFD. L'intérêt majeur de cette étude pour cette région est la définition des cavités de liaison, principale et annexe, de T1R2 et T1R3. Ce site de fixation orthostérique peut lier des molécules très variées d'un point de vue chimique. La suite de l'étude portera sur la caractérisation structurale de cette cavité de liaison par des études dynamiques. L'alignement des séquences en AA des CRD présente peu d'ambiguïté,

comme énoncé précédemment, les cystéines impliquées dans les ponts disulfures de cette région sont une contrainte forte. Enfin, le modèle par homologie du TMD de T1R3 est le premier basé sur une structure cristallographique d'un RCPG de classe C. La définition de la cavité de liaison et les interrupteurs moléculaires permettront de réaliser une étude approfondie sur la dynamique du récepteur lors de la fixation d'un ligand puis de son activation. La structure du STR proposée ici apporte une solution aux désaccords des modèles proposés dans les études précédentes.

	T1R2	T1R3
VFD	S40 K65 I67 L71 Y103 D142 N143 S144 E145 S165 A166 I167 S168 A187 H190 Y215 P277 D278 L279 Y282 E302 S303 W304 D307 V309 E382 R383 V384	P42 N68 W72 S104 H145 S146 S147 E148 G168 A169 S170 M171 D190 Q193 Y218 S276 V277 H278 H281 E301 A302 W303 S306 L308 H387 H388 Q389
TMD	<i>V</i> ^{2.52} ₆₁₇ <i>V</i> ^{2.53} ₆₁₈ <i>Y</i> ^{2.54} ₆₂₁ <i>R</i> ^{3.32} ₆₃₃ <i>Q</i> ^{3.33} ₆₃₄ <i>F</i> ^{3.36} ₆₃₇ <i>P</i> ^{3.37} ₆₃₈ <i>F</i> ^{3.40} ₆₄₁ <i>T</i> ^{3.41} ₆₄₂ <i>G</i> ^{4.50} ₆₉₆ <i>S</i> ^{<i>ECL2</i>} ₇₁₉ <i>N</i> ^{<i>ECL2</i>} ₇₂₁ <i>P</i> ^{<i>ECL2</i>} ₇₂₂ <i>N</i> ^{<i>ECL2</i>} ₇₂₃ <i>Y</i> ^{<i>ECL2</i>} ₇₂₄ <i>S</i> ^{5.39} ₇₂₇ <i>L</i> ^{5.40} ₇₂₈ <i>T</i> ^{5.44} ₇₃₂ <i>Y</i> ^{6.46} ₇₆₉ <i>S</i> ^{6.50} ₇₇₃ <i>V</i> ^{6.51} ₇₇₄ <i>L</i> ^{6.53} ₇₇₆ <i>C</i> ^{6.54} ₇₇₇ <i>M</i> ^{6.57} ₇₈₀ <i>V</i> ^{7.28} ₇₈₈ <i>D</i> ^{7.32} ₇₉₂ <i>N</i> ^{7.39} ₇₉₉ <i>A</i> ^{7.42} ₈₀₂ <i>I</i> ^{7.43} ₈₀₃	<i>S</i> ^{2.52} ₆₂₀ <i>V</i> ^{2.53} ₆₂₁ <i>F</i> ^{2.56} ₆₂₄ <i>Q</i> ^{3.32} ₆₃₆ <i>Q</i> ^{3.33} ₆₃₇ <i>S</i> ^{3.36} ₆₄₀ <i>H</i> ^{3.37} ₆₄₁ <i>L</i> ^{3.40} ₆₄₄ <i>T</i> ^{3.41} ₆₄₅ <i>Y</i> ^{4.50} ₆₉₉ <i>H</i> ^{<i>ECL2</i>} ₇₂₁ <i>R</i> ^{<i>ECL2</i>} ₇₂₃ <i>T</i> ^{<i>ECL2</i>} ₇₂₄ <i>R</i> ^{<i>ECL2</i>} ₇₂₅ <i>S</i> ^{<i>ECL2</i>} ₇₂₆ <i>S</i> ^{5.39} ₇₂₉ <i>F</i> ^{5.40} ₇₃₀ <i>H</i> ^{5.44} ₇₃₄ <i>Y</i> ^{6.46} ₇₇₁ <i>W</i> ^{6.50} ₇₇₅ <i>V</i> ^{6.51} ₇₇₆ <i>F</i> ^{6.53} ₇₇₈ <i>V</i> ^{6.54} ₇₇₉ <i>L</i> ^{6.57} ₇₈₁ <i>R</i> ^{7.28} ₇₉₀ <i>Q</i> ^{7.32} ₇₉₄ <i>C</i> ^{7.39} ₈₀₁ <i>G</i> ^{7.42} ₈₀₄ <i>I</i> ^{7.43} ₈₀₅
TS	<i>I</i> ^{3.44} ₆₄₅ <i>Y</i> ^{6.46} ₇₆₉ <i>N</i> ^{7.39} ₇₉₉	<i>L</i> ^{3.44} ₆₄₈ <i>Y</i> ^{6.46} ₇₇₁ <i>C</i> ^{7.39} ₈₀₁
IL	<i>A</i> ^{5.99} ₅₉₉ <i>G</i> ^{6.00} ₆₀₀ <i>R</i> ^{3.50} ₆₅₁ <i>S</i> ^{3.51} ₆₅₂ <i>F</i> ^{3.52} ₆₅₃ <i>Q</i> ^{3.53} ₆₅₄ <i>N</i> ^{6.34} ₇₅₇ <i>E</i> ^{6.35} ₇₅₈ <i>A</i> ^{6.36} ₇₅₉ <i>K</i> ^{6.37} ₇₆₀	<i>S</i> ^{6.02} ₆₀₂ <i>G</i> ^{6.03} ₆₀₃ <i>Q</i> ^{3.50} ₆₅₄ <i>A</i> ^{3.51} ₆₅₅ <i>A</i> ^{3.52} ₆₅₆ <i>E</i> ^{3.53} ₆₅₇ <i>N</i> ^{6.34} ₇₅₉ <i>R</i> ^{6.35} ₇₆₀ <i>A</i> ^{6.36} ₇₆₁ <i>R</i> ^{6.37} ₇₆₂
HB	<i>Y</i> ^{7.47} ₈₀₇ <i>F</i> ^{7.48} ₈₀₈ <i>P</i> ^{7.50} ₈₁₀ <i>K</i> ^{7.51} ₈₁₁ <i>Y</i> ^{7.53} ₈₁₃	<i>F</i> ^{7.47} ₈₀₉ <i>H</i> ^{7.48} ₈₁₀ <i>P</i> ^{7.50} ₈₁₂ <i>R</i> ^{7.51} ₈₁₃ <i>Y</i> ^{7.53} ₈₁₅

TABLE 3.2 – Liste des AA importants du STR : AA du site de liaison du VFD et TMD ainsi que les AA formant les interrupteurs moléculaires.

Les AA notés en italique sont extrapolés depuis l'alignement de séquences. TS, IL et HB signifient, respectivement, *Transmission Switch*, *Ionic Lock* et *Hydrophobic Barrier*.

Chapitre 4

Relation structure-fonction de T1R2-T1R3

Ce chapitre est basé sur un article en préparation :

Chéron, J.-B., Golebiowski, J., Antonczak, S., & Fiorucci, S. (en préparation).

Molecular dynamic simulations enlighten different ligand binding modes in the N-terminal domain of the sweet taste receptor.

4.1 Préambule : Stratégie d'étude par domaine

Au-delà des approches *ligand-based*, la recherche et le développement de nouveaux composés nécessite la connaissance structurale du ligand et du récepteur. L'étude statistique du chapitre 2 aborde les approches *ligand-based* et permet de sélectionner des composés d'intérêts de manière automatique. En revanche, cette méthode n'apporte pas – ou peu – d'information sur les interactions nécessaires pour la reconnaissance du ligand par le récepteur. La construction du modèle par homologie du STR dans le chapitre 3 permet d'aborder cette problématique. En effet, même si la structure chimique des ligands porte une information quant-à leur pouvoir sucrant, la première étape de

la perception sucrée nécessite l'interaction d'un sucre ou d'un édulcorant avec le STR. Il apparaît primordial de comprendre comment ces structures chimiques sont reconnues par le STR. La mise en évidence des mécanismes moléculaires qui entrent en jeu permet, à la fois, de décrire structurellement les sites de liaisons du STR et de concevoir de nouveaux édulcorants par des approches *receptor-based*.

Comme il a été énoncé dans les chapitres précédents, le STR présente plusieurs sites de liaison. L'objectif de ce chapitre est de caractériser les sites de reconnaissance du VFD de T1R2 et du TMD de T1R3. Le VFD de T1R2 est la cavité « naturelle » ; elle est capable d'accueillir des structures chimiques très variées [81]. Un intérêt particulier est porté aux disaccharides et dipeptides, car des données de mutagenèse dirigée sont disponibles pour le saccharose, le sucralose, l'aspartame et le neotame. De plus, les disaccharides permettent d'étudier spécifiquement la région de reconnaissance des sucres qui est essentielle dans la fixation de molécules glycosylées. Les dipeptides, quant à eux, possèdent un groupement hydrophobe et des groupements chargés. L'analyse des AA en interactions avec ces groupements, dont l'importance a déjà été citée dans le chapitre 2.2.3, présente un grand intérêt pour le développement rationnel de nouveaux édulcorants.

Les domaines transmembranaires de T1R2 et T1R3 sont aussi en capacité de fixer des édulcorants. T1R2 peut fixer le S819 et T1R3 le cyclamate [80, 112]. L'étude du TMD de T1R3 est effectuée ici, car des données de mutagenèse dirigée sont disponibles. La compréhension du site de reconnaissance de T1R3, dit « allostérique », permet de définir des résidus clés, à la fois pour le développement de nouveaux modulateurs et pour la mise en évidence des résidus critiques pour l'activation du récepteur.

4.2 Préparation des systèmes moléculaires

Dans ce chapitre, des simulations de dynamique moléculaire sont réalisées avec le modèle par homologie du STR construit précédemment. Cette approche est utilisée

pour modéliser la physique d'un système en fonction du temps. Il est alors possible d'étudier les résidus clés des interactions ligand-récepteur en prenant en compte le comportement du ligand dans sa cavité. La dynamique moléculaire est la méthode principale des chapitres 4 et 5. Ainsi, une description de ces lois est faite en annexe B.4.

Le point de départ de ces simulations est estimé par des approches de *docking*, dont la théorie est décrite en annexe B.3. Les espaces de recherche sont limités aux sites de liaison connus pour le VFD de T1R2 « fermé » et le TMD de T1R3. Pour chacun des *docking* réalisés, la pose de plus basse énergie et correspondant aux données expérimentales de mutagenèse dirigée est sélectionnée. Ensuite, le complexe ligand-récepteur est placé dans un environnement physiologique. Enfin, le système construit peut être soumis aux calculs de dynamique moléculaire.

4.2.1 Paramétrisation des dynamiques moléculaires

Le logiciel AMBER [123] est utilisé pour réaliser les calculs de dynamique moléculaire. Les paramètres de chacun des atomes doivent être définis (rayon, charge, ...), ainsi que leurs interactions par paire (annexe B.4.2). Pour cela, un ensemble de constantes, issues de données expérimentales ou de calculs quantiques, est présent dans le champ de force.

Des acides aminés au champ de force pour les protéines

Une mise à jour vers AMBER16 a été effectuée au cours de ces trois années de travaux. En fonction de la date de début des projets, les champs de force ff03.r1 ou ff14SB ont été utilisés. Ces deux champs de force sont des améliorations du ff99SB (*AMBER force field 99 Stony Brook*), reprenant les paramètres du ff99, spécialement développé pour les biomolécules. Le ff99SB a été optimisé pour pallier aux limitations du ff99. Les angles dièdres ont été re-paramétrisés, le traitement du *backbone* de la glycine corrigé et le gain énergétique dû à la formation de structures secondaires révisé. Le ff14SB est dans la continuité de ces optimisations et constitue la dernière mise à jour. Il optimise les

paramètres des angles dièdres du *backbone* et des chaînes latérales pour reproduire la géométrie de données expérimentales. Le ff03.r1, quant-à lui, dérive directement du ff99. Les modifications de ce champ de force sont une ré-estimation de la charge des atomes ainsi que des paramètres des angles dièdres pour limiter le repliement préférentiel en hélice. Il était utilisé historiquement au laboratoire avec AMBER9 et a permis de réaliser les premières simulations (TMD de T1R3). Pour chaque système, l'état de protonation des AA des protéines est défini avec H++ [124]. Pour le VFD de T1R2 un ion Mg^{2+} est ajouté dans une boucle (figure 4.1), comme dans les structures cristallographiques du mGluR [103].

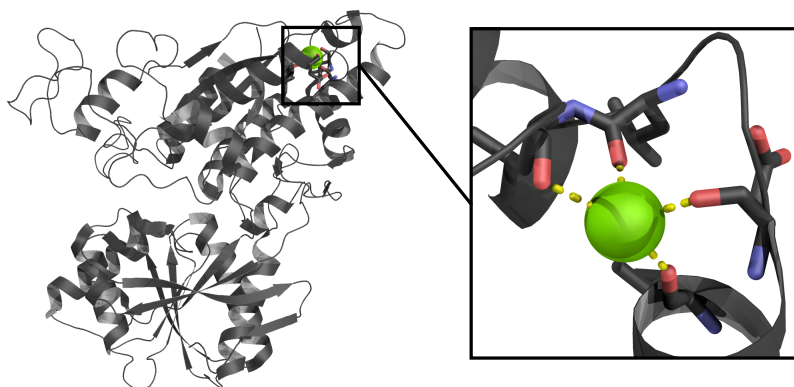


FIGURE 4.1 – Ion Mg^{2+} présent dans la structure cristallographique du mGluR1 (1EWK)

La protéine est représentée en noir. Les atomes de carbone, oxygène et azote sont colorés, respectivement, en noir, rouge et bleu. L'ion Mg^{2+} est représenté par une sphère verte.

Champs de force des sucres et des ligands non « standard »

Pour les simulations du TMD de T1R3, le ligand étudié est le cyclamate. Dans le cas du VFD de T1R2, les ligands sont des dipeptides ou des disaccharides. Les paramètres du saccharose proviennent d'un champ de force spécialement développé pour les sucres (GLYCAM 06j-1) et compatible avec AMBER [125]. Cette version permet une compatibilité avec le champ de force des protéines ff14SB. Pour les autres ligands, le champ

de force utilisé est GAFF2 (*general AMBER force field*) [126], car il est entièrement compatible avec les autres champs de force AMBER. De plus, il contient les paramètres nécessaires à la plupart des molécules organiques. Le programme *Antechamber* [127] est utilisé pour optimiser la géométrie des ligands et pour calculer les charges des atomes. Les méthodes appliquées pour assigner les charges sont résumées dans le tableau 4.1.

Type	T1R2-VFD	T1R3-TMD	mGluR5-TMD ^a
Protéine	ff14SB	ff03.r1	ff14SB
Eau-Ions	TIP3P	TIP3P	TIP3P
Lipides	–	lipid11	lipid14
Ligands	GLYCAM 06j-1 ou Antechamber GAFF2 Optimisation HF/6- 31G* et charges RESP [129]	Antechamber GAFF et charges am1- bcc [128]	Antechamber GAFF2 et charges am1- bcc [128]

TABLE 4.1 – Résumé de l'utilisation des champs de force et des paramètres des ligands.

^a. Chapitre suivant

Simulation avec membrane

Les simulations du TMD de T1R3 sont réalisées dans des conditions « physiologiques ». Le récepteur est enchassé dans une bicouche lipidique modélisée par des phospholipides de type « POPC » (1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine). L'orientation du récepteur dans la membrane a été prédit avec le serveur OPM [130]. Le champ de force lipid11¹, adapté de GAFF, est utilisé [131].

Représentation de l'eau

Le modèle d'eau à trois points « TIP3P » est pris en compte. Dans le cas de la simulation d'une protéine dans l'eau (VFD de T1R2), une boîte est constituée à 10 Å autour de la protéine. Si une membrane est présente, une couche d'eau de 10 Å est ajoutée selon

1. Dans le chapitre suivant, la mise à jour vers AMBER16 a permis d'utiliser le champ de force lipid14 avec ff14SB pour les simulations du TMD du mGluR. lipid14 constitue une importante mise à jour. Il combine lipid11 avec le développement de nouveaux paramètres pour les lipides (GAFFlipid).

l'axe z . Le champ de force utilisé pour les ions a été optimisé pour le solvant « TIP3P ». La neutralisation de la charge des systèmes se fait avec les ions Na^+ et Cl^- .

4.2.2 Protocole des simulations de dynamique moléculaire

Le protocole standard avant la période de production correspond à une minimisation de l'énergie, une thermalisation et une équilibration du système. Le module pmemd avec l'algorithme SHAKE [132], pour contraindre les liaisons impliquant des atomes d'hydrogène, est utilisé. L'approximation SHAKE permet d'effectuer les calculs avec un pas d'intégration de 2 fs. Le rayon de troncature pour estimer les interactions non-liantes est de 8 Å avec une mise à jour de la liste des paires tous les 15 pas. Les conditions limites périodiques sont appliquées (voir annexe B.4 pour plus de détails).

• Pré-équilibration du système

Lors de la construction du système de nombreuses approximations sont présentes. Le modèle réalisé avec *Modeller* peut être dans un état énergétique différent du minimum estimé par AMBER. De plus, l'ajout du solvant, des ions et du ligand nécessite de replacer la totalité du système dans son minimum énergétique. Cette étape est indispensable avant de réaliser les calculs de dynamique moléculaire. Elle permet la réorganisation des distances, des angles et des dièdres autour de leur point d'équilibre énergétique. Les encombrements stériques sont également supprimés pendant cette étape. Deux protocoles sont utilisés en fonction du type de système (avec ou sans membrane).

Protéine en solution aqueuse

Quatre cycles de minimisation et d'équilibration sous contraintes à 100 K sont effectués. Chaque cycle permet de diminuer la valeur de la contrainte appliquée sur le soluté de 15 à 0 kcal.mol⁻¹.Å⁻². Les étapes de minimisation de l'énergie correspondent à 5 000 pas de *steepest descent* et 5 000 pas de gradient conjugués. Les équilibrations sont réalisées pendant 20 ps dans les conditions NVT². L'algorithme PME (*Particle Mesh*

2. Le Nombre d'atome, le Volume et la Température sont constants

Ewald) est utilisé avec une grille de 1 Å et un seuil de troncature de 8 Å. L'intérêt de cette approche est de permettre au système de franchir les barrières faibles en énergie. Le système est relaxé en douceur et le solvant peut ainsi s'adapter autour de la protéine dont les coordonnées des atomes sont contraints.

Protéine membranaire

Dans ce cas, les minimisations énergétiques sous contraintes des différentes phases sont faites consécutivement. Une contrainte de $50 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ est appliquée sur la position des atomes en dehors de la phase à minimiser. Tout d'abord, la minimisation de l'énergie se fait sur les chaînes latérales de la protéine, ensuite la membrane et enfin, l'eau et les ions. Ensuite, une étape de thermalisation du solvant (membrane, eau et ions) à 310 K est effectuée en 20 ps. Elle est suivie d'une équilibration du solvant pendant 200 ps. À ce point, les lipides, l'eau et les ions se sont réorganisés autour de la protéine dont la position des atomes était contrainte. Les quatre cycles de minimisation et d'équilibration à 100 K, en diminuant la contrainte sur la protéine de 15 à $0 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$, sont appliqués. Dans ce cas, les étapes d'équilibrations se font dans les conditions NPT³ car, en présence d'une membrane, la pression qui est anisotrope⁴ doit être contrôlée.

• Thermalisation

Lors de la minimisation du système, l'énergie cinétique est nulle. La thermalisation correspond à une attribution progressive des vitesses avec un bain thermostaté. Sans membrane, elle est effectuée dans les conditions NVT de 100 K à 310 K, en 20 ps, en couplant le système à un thermostat de Langevin. En présence d'une membrane, les conditions NPT sont appliquées.

3. Le Nombre d'atome, la Pression et la Température sont constants

4. La pression appliquée n'est pas la même entre les axes x , y et l'axe z pour réguler la tension superficielle générée par la présence des phases lipidiques et aqueuses.

• Équilibration

Une courte équilibration de 100 ps à 310 K en NVT est effectuée dans le cas des simulations dans l'eau seulement, dans le but de stabiliser la température. Cette étape doit être très courte pour ne pas créer de « bulles de vide ». Elle est suivie d'une équilibration en NPT de 100 ps pour permettre à la densité de s'équilibrer. Pour les systèmes avec membrane, seule une équilibration en NPT de 40 ns est réalisée. Les temps d'équilibration sont différents entre les systèmes (avec ou sans membrane) car l'eau ajoutée au départ est pré-équilibrée, tandis que les lipides ne le sont pas. Les simulations en NPT sont effectuées à 1 bar (0,987 atm) et 310 K. Le barostat de Berendsen et le thermostat de Langevin sont appliqués.

• Production

Toutes les productions sont réalisées en condition NPT (1 bar et 310 K). La température et la pression sont, ici aussi, contrôlées par le thermostat de Langevin et le barostat de Berendsen. Dans le cas des simulations avec membrane, le contrôle de la pression est anisotrope. Les trajectoires des simulations de dynamique moléculaires peuvent être utilisées pour estimer l'énergie d'interaction entre un ligand et son récepteur. Les détails méthodologiques de la dynamique moléculaire et de l'estimation de l'énergie d'interaction sont donnés en annexe [B.4](#) et [B.5](#).

4.3 Domaine extracellulaire

4.3.1 Préparation des systèmes

Le VFD de T1R2 est le site de reconnaissance dit « orthostérique ». Il fixe les ligands naturels du STR (les sucres). Il a été montré que ce site est également capable d'accueillir divers édulcorants [[Dubois2016](#)]. L'objectif est de comprendre comment le VFD de T1R2 reconnaît des structures chimiques variées et de déterminer les AA formant les différents sites d'interactions. Pour l'étude du site de liaison du VFD de T1R2,

quatre systèmes sont préparés avec deux disaccharides, le saccharose et le sucralose, et deux dipeptides, l'aspartame et le neotame. Le protocole de simulation de dynamique moléculaire défini précédemment (dans une boîte d'eau) est appliqué. Pour chaque ligand, deux simulations de 150 ns sont réalisées.

4.3.2 Stabilité des systèmes

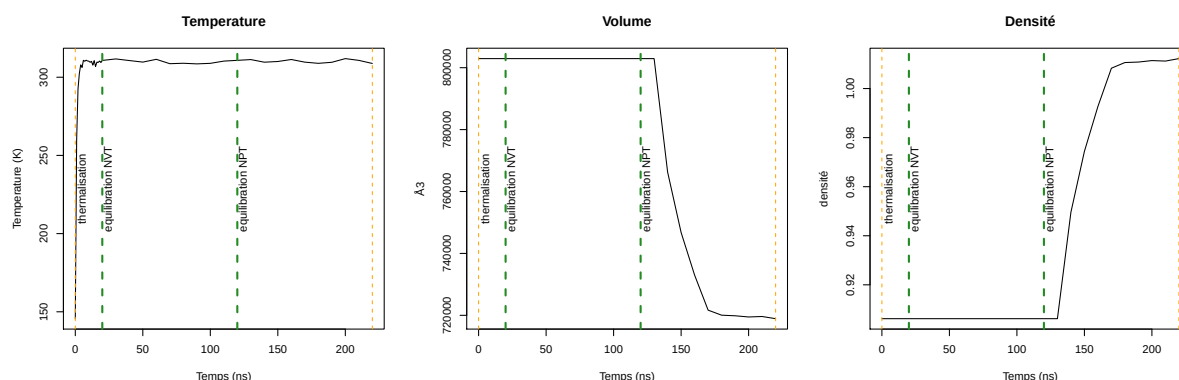


FIGURE 4.2 – Vérification de la convergence des étapes de pré-production.

Les 200 premières ns permettent la convergence de la température (310 K), du volume et de la densité du système moléculaire.

Les phases de minimisation, de thermalisation et d'équilibration sont cruciales pour obtenir un système à son point d'équilibre énergétique. La figure 4.2 montre que le protocole appliqué permet la convergence de la température à 310 K ainsi que de la densité à ≈ 1 . Lors des phases de thermalisation et d'équilibration en NVT, le volume des systèmes converge aussi à un point d'équilibre. Alors, l'étape de production peut être effectuée pour échantillonner le comportement des complexes ligand-récepteur.

Le RMSD (*Root Mean Square Deviation*) quantifie les fluctuations des positions atomiques au cours de la trajectoire. Il est défini, entre deux molécules A et B de N atomes de coordonnées x, y, z , tel que :

$$RMSD(A, B) = \sqrt{\frac{1}{N} \sum_{i=1}^N [(A_{xi} - B_{xi})^2 + (A_{yi} - B_{yi})^2 + (A_{zi} - B_{zi})^2]} \quad (4.1)$$

La figure 4.3 représente le RMSD pour chacune des trajectoires. La première structure de la trajectoire est utilisée comme point de référence. En quelques ns l'ensemble des systèmes ont atteint leur point d'équilibre d'un point de vue structural.

4.3.3 Interactions ligand-récepteur

Le calcul par la méthode MMGBSA permet d'estimer l'affinité ligand-récepteur (annexe B.5). Plus l'énergie d'interaction estimée est basse, plus le ligand en question est stable dans la cavité. Les deux simulations de chaque système sont concaténées, superposées aux atomes du squelette protéique de la première structure de l'étape de production et clusterisées selon la position du ligand. Une matrice de distance euclidienne, selon le RMSD du ligand, est construite et une méthode de classification hiérarchique est appliquée. Les groupes sont déterminés avec seuil de RMSD à 2 Å. Les deux clusters majoritaires sont sélectionnés pour effectuer les calculs MMGBSA et la décomposition par résidu. Le tableau 4.2 résume les valeurs de ΔH . Pour chaque famille chimique, le composé perçu le plus sucré (logPS le plus élevé) présente la plus forte affinité pour le récepteur : le sucralose et le neotame pour, respectivement, les disaccharides et les dipeptides. Le sucralose est un dérivé chloré du saccharose. Le chlore est uniquement accepteur de liaison hydrogène et il semble augmenter la force de l'interaction avec le récepteur. Le neotame, quant-à lui, présente deux cœurs hydrophobes, contre un pour l'aspartame. Cette propriété suggère que la contribution hydrophobique des dipeptides a une grande importance dans l'énergie d'interaction.

Il faut garder à l'esprit que la perception de la saveur sucrée n'est pas seulement due à la liaison du ligand. Même si cette étape est nécessaire, elle n'est pas suffisante. L'activation du récepteur, qui fait suite à la liaison, est le facteur déclencheur de la perception de la saveur sucrée. Cependant, la caractérisation du site de reconnaissance permet de comprendre les propriétés moléculaires des édulcorants intenses et peut aider pour la conception d'édulcorants plus performants.

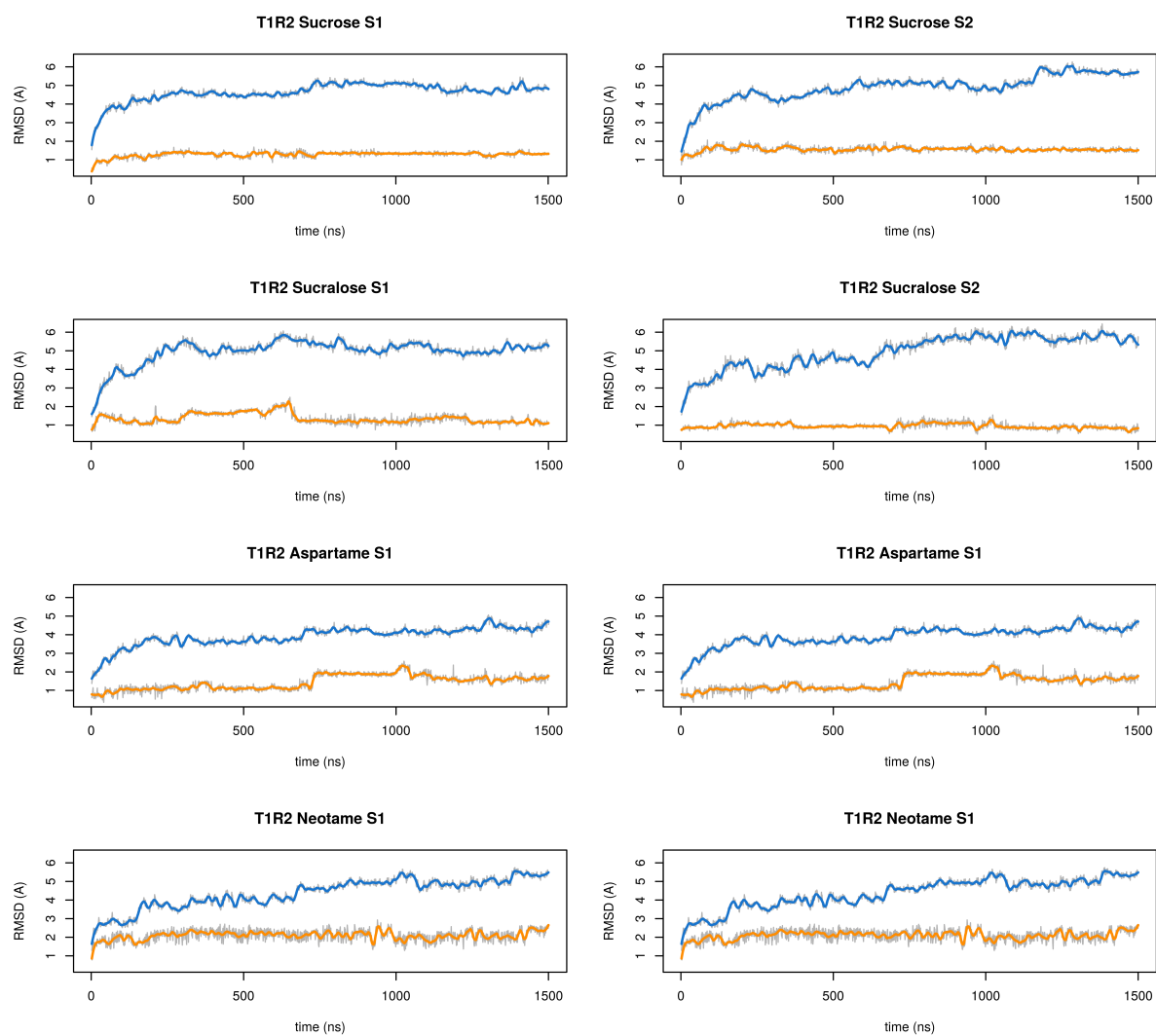


FIGURE 4.3 – Convergence du RMSD du VFD de T1R2 et des ligands lors de l'étape de production.

Le RMSD est calculé par rapport à la structure de la première fenêtre de l'étape de production. La variation du RMSD du VFD de T1R2 est représentée en bleu, celui du ligand en orange.

Molécule	logPS	pEC50	ΔH (kcal.mol ⁻¹)
Saccharose	0	1,65	-24,2
Sucralose	2,78	4,59	-29,2
Aspartame	2,25	3,91	-36,2
Neotame	3,89	6,4	-39,8

TABLE 4.2 – Énergie d'interaction des ligands avec le VFD de T1R2

Pour chaque, la moyenne des ΔH pondérée par le nombre de fenêtres des deux clusters majoritaires est donnée. Les pEC50^a sont issues des publications de Servant *et al.* [133], Masuda *et al.* [114] et Assadi-Porter *et al.* [134].

a. $pEC50 = -\text{LOG}(EC50)$. L'EC50 (concentration efficace médiane) est la concentration (mol.L⁻¹) où 50% de l'effet maximal est observé.

4.3.4 Décomposition de l'énergie d'interaction par résidu

La méthode MMGBSA permet d'estimer la contribution énergétique de chaque AA du récepteur en interaction avec le ligand. La somme de ces énergies constitue l'énergie totale de l'interaction ligand-récepteur. Comme énoncé dans l'équation B.11 en annexe, cette énergie par AA tient compte, à la fois, de l'énergie d'interaction et de la « pénalité » de désolvatation.

Les figures 4.4 et 4.5 représentent les valeurs (en kcal.mol⁻¹) de la contribution par AA pour, respectivement, les disaccharides et les dipeptides. Les résidus qui ont une contribution favorable, sélectionnés avec un seuil de -0,5 kcal.mol⁻¹, sont : S40, K65, V66, I67, Y103, D142, N143, Y164, S165, I167, Y215, P277, D278, L279, T280, Y282, E302, S303, W304, A305, A307, L310 et V384. La figure 4.6 met en évidence les résidus importants sur le représentant de chaque cluster.

Le tableau 4.3 référence les expériences *in vitro* et les effets des mutations. Une grande partie des AA mutés ayant un effet *in vitro* est retrouvée par la décomposition énergétique par résidu. Il est intéressant de noter que les six résidus Y164, T280, Y282, W304, A305 et L310 n'ont pas été testés par mutagenèse dirigée. Cette approche fait ressortir seulement les résidus qui interagissent directement avec le ligand et l'analyse de la cavité de liaison révèle la présence de molécules d'eau pontantes. La figure 4.7 représente la carte de densité des molécules d'eau autour des ligands au cours de la

simulation de dynamique moléculaire. Cela permet d'expliquer pourquoi des AA importants pour la liaison du ligand ne sont pas retrouvés par la méthode MMGBSA (tels que les AA S40 ou S144 surlignés en bleu dans le tableau 4.3).

L'analyse des trajectoires montre que le résidu K65 interagit avec la D142 et que la R383 interagit avec les AA D278 et E302. Ceci confirme ce qui a été énoncé dans des études précédentes, les interactions de ces AA permettent d'établir un contact entre lobe I et II à l'entrée de la cavité. En plus de la stabilisation des ligands dans la cavité de liaison, ces AA sont importants pour la fonction du récepteur.

Pour l'aspartame, les clusters majoritaires sur les simulations correspondent aux deux conformations citées dans une étude précédente [135]. Dans cet article, les auteurs ont analysé les résultats proposés par E. Maillet *et al.* [115] qui prédisent le site d'interaction de l'aspartame avec le VFD de T1R2 à partir de données de mutagenèse dirigée et de *docking*. Cependant, la critique faite dans la référence [135] est, qu'une telle approche ne prend pas en compte la flexibilité de l'aspartame. Les approches de dynamique moléculaire réalisées ici permettent d'explorer les deux états conformationnels de l'aspartame (figure 4.6) et réconcilient les résultats de C. Toniolo et P. Temussi [135] avec ceux de E. Maillet *et al.* [115].

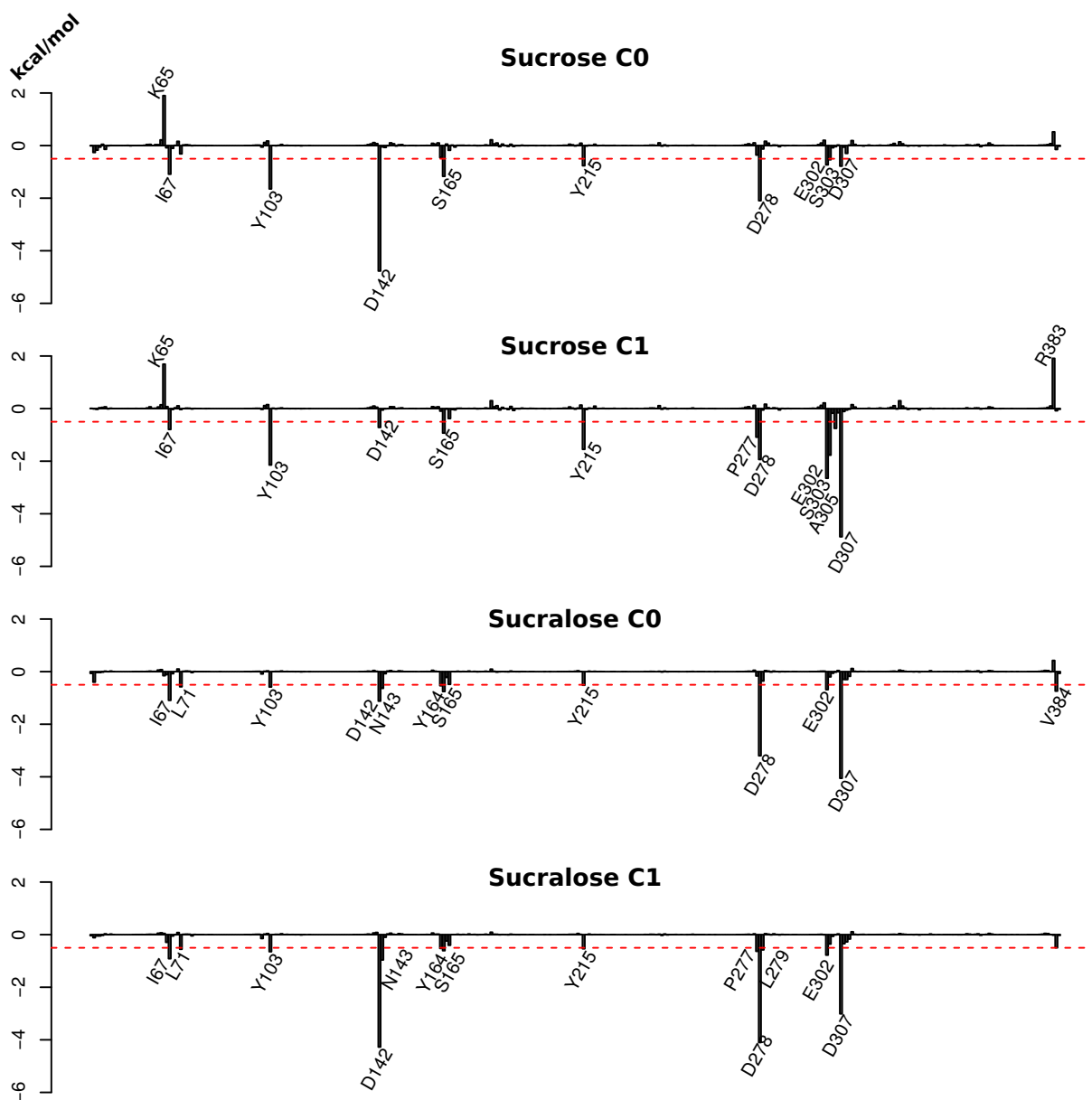


FIGURE 4.4 – Décomposition par résidus des énergies d'interaction du VFD de T1R2 avec les disaccharides.

L'analyse est effectuée sur les deux clusters majoritaires. Pour le saccharose les clusters C0 et C1 représentent, respectivement, 60% et 27% de la trajectoire. Pour le sucralose, ils représentent 35% et 27%.

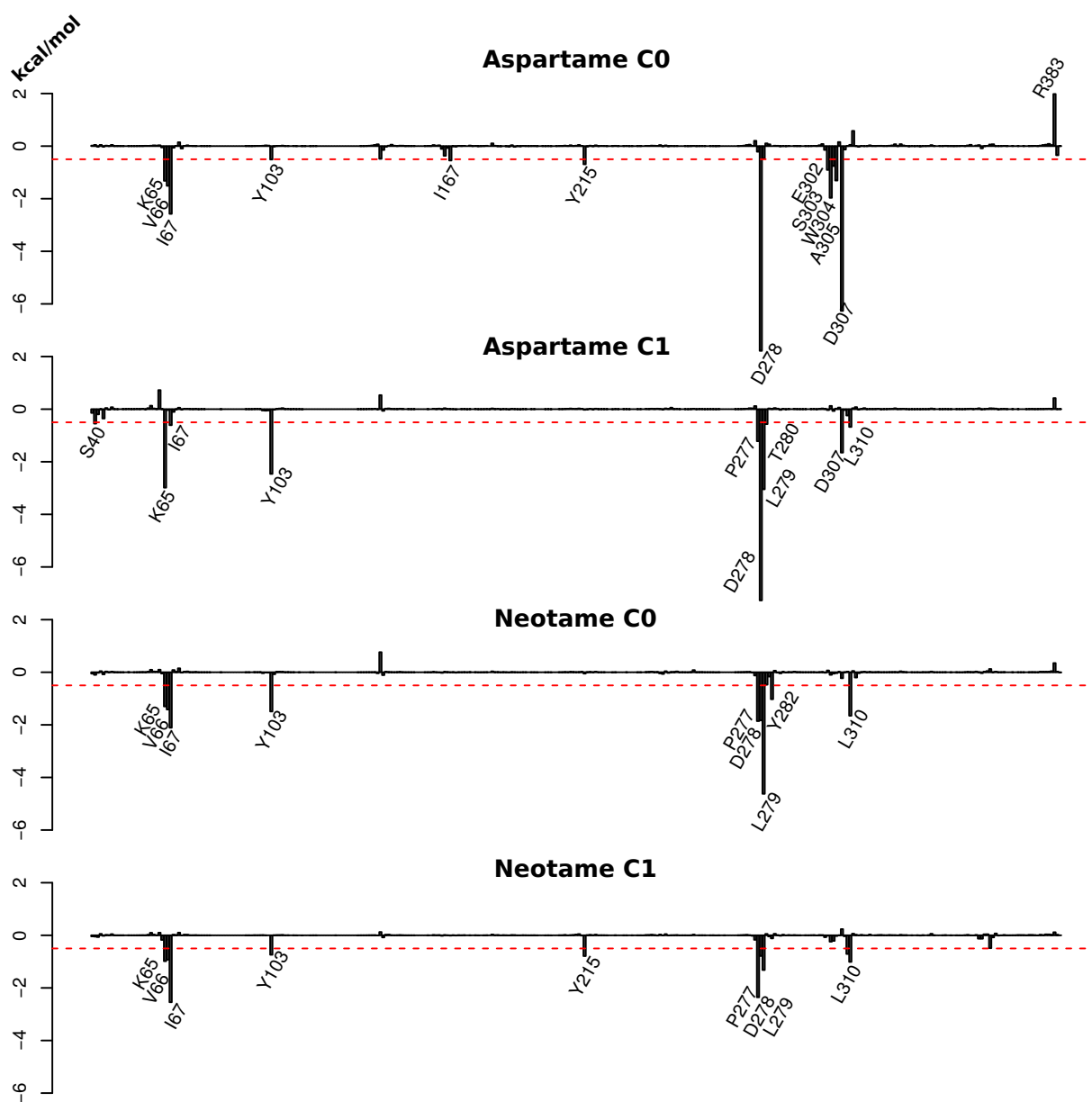


FIGURE 4.5 – Décomposition par résidus des énergies d'interaction du VFD de T1R2 avec les dipeptides.

L'analyse est effectuée sur les deux clusters majoritaires. Pour l'aspartame les clusters C0 et C1 représentent, respectivement, 47% et 26% de la trajectoire. Pour le neotame, ils représentent 45% et 41%.

Résidu	saccharose	sucralose	aspartame	neotame
S40	NR	NR	D	D
K65*	SE	D	D	
V66			D	D
I67			D	D
Y103	NR	NR	NR	
D142*	NR	NR	NR	
N143			SE	
S144	SE	SE	NR	D
Y164				
S165		SE	D	
I167	SE	SE		
Y215*			D	
P277	D	D		
D278*	NR	NR	D	
L279	D	D	**	
T280				
Y282				
E302	NR	NR	NR	D
S303			D	
W304				
A305				
D307*	NR	D	D	D
L310				
V384			SE	

TABLE 4.3 – Résumé des données de mutagenèse dirigée pour le VFD de T1R2

Les AA surlignés en verts sont retrouvés par MMGBSA en interaction directe avec le ligand. En bleu, les AA en interaction avec le ligand *via* un réseau d'interaction par liaisons hydrogène avec des molécules d'eau. Les abréviations NR (Non Répondant), D (Diminue la Réponse) et SE (Sans Effet) sont issues des données de mutagenèse dirigée (en alanine). Le symbole « * » dénote les AA aussi impliqués dans la fonction du récepteur. La mutation en alanine de l'AA L279 augmente légèrement la réponse du récepteur (noté « ** »). Pour le détail des références, voir le tableau [A.2](#).

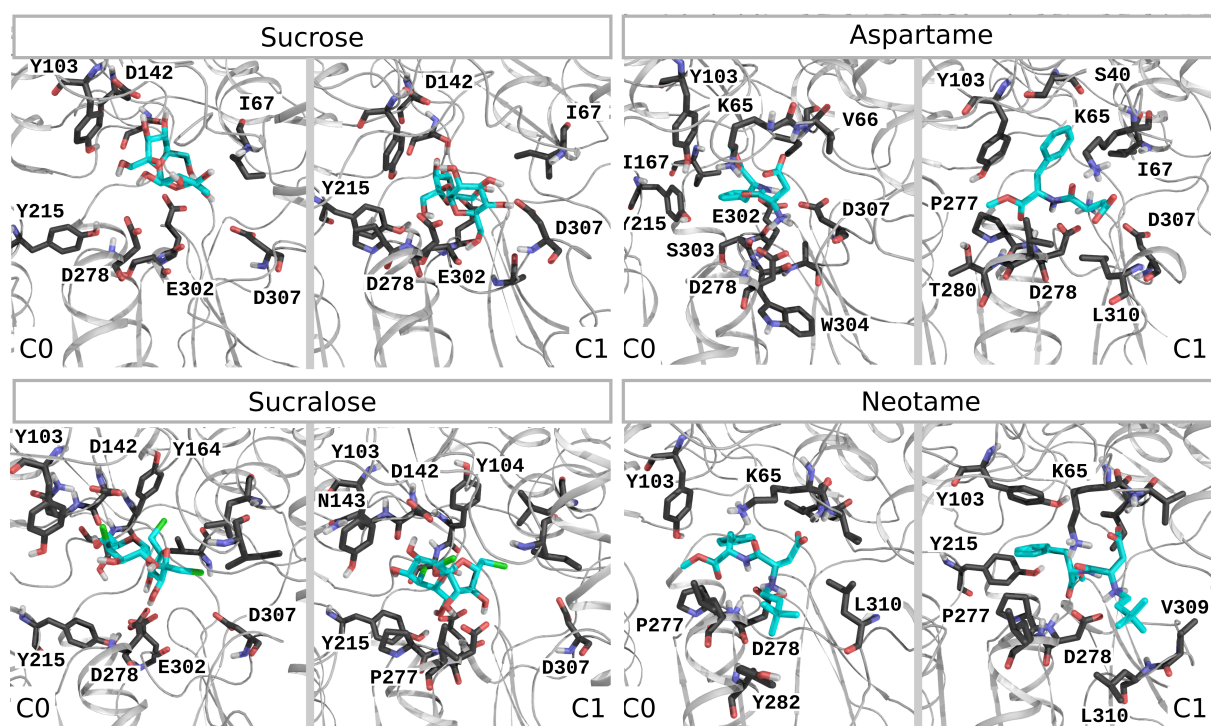


FIGURE 4.6 – Représentation des sites d'interaction ligand-récepteur du VFD de T1R2 pour chaque cluster.

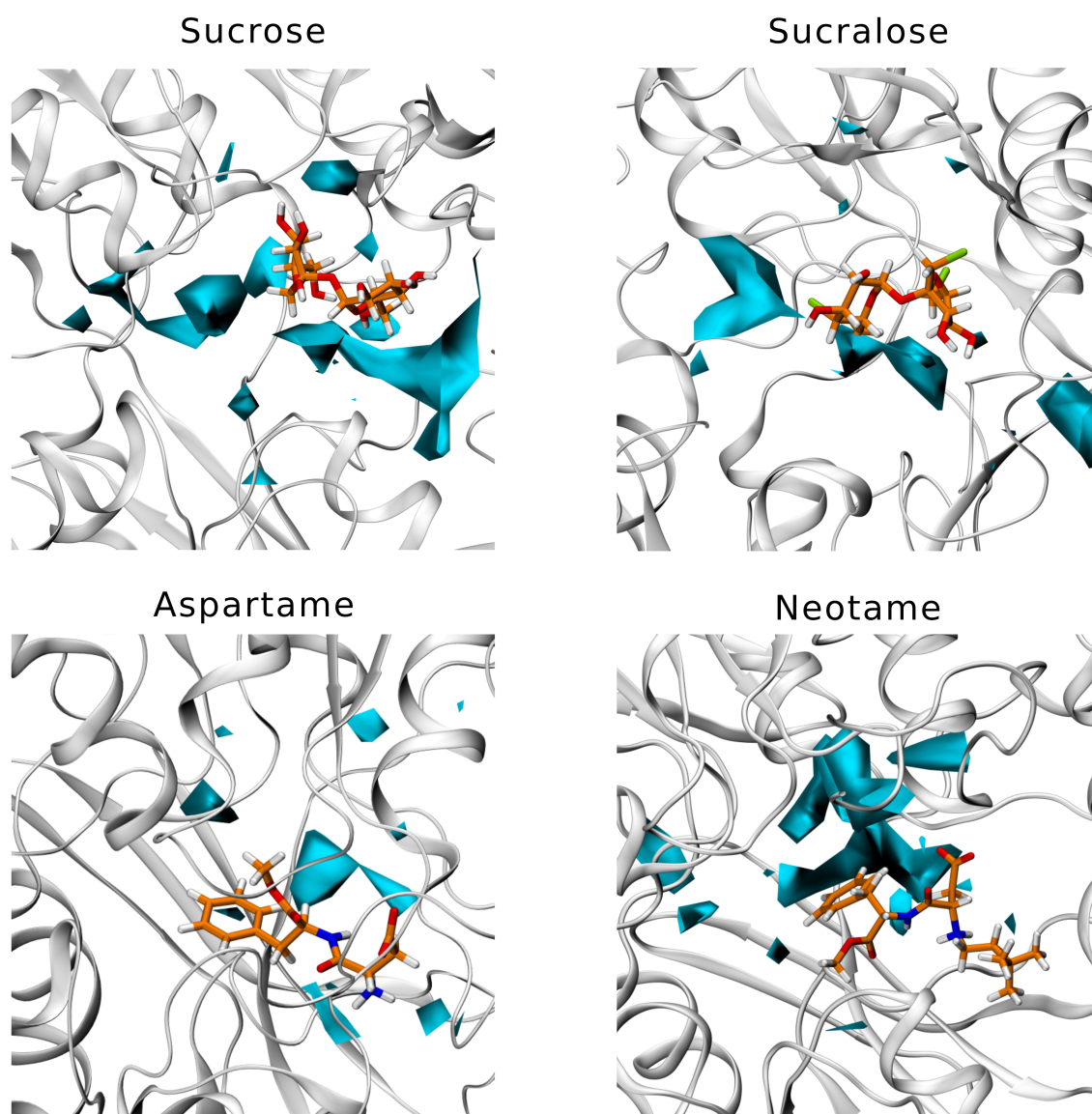


FIGURE 4.7 – Représentation de la densité des molécules d'eau (en bleu) au cours des trajectoires.

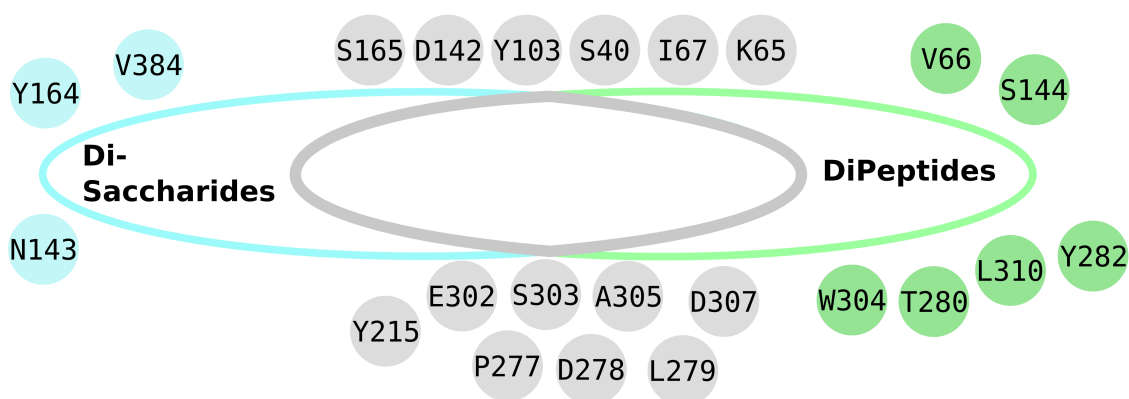


FIGURE 4.8 – Représentation en deux dimensions des sites d'interaction du VFD de T1R2 pour les disaccharides et les dipeptides.

Les AA notés ont été mis en évidence par l'approche MMGBSA. Ceux colorés en gris sont communs aux disaccharides et dipeptides. Les AA colorés en bleu et en vert sont, respectivement, spécifiques des disaccharides et dipeptides.

4.3.5 Analyses des sites de liaison des disaccharides et dipeptides

Le même schéma d'interaction est observé pour chacune des familles. La figure 4.8 résume les résidus en interaction avec les ligands. 14 résidus sont retrouvés dans les deux schémas d'interactions. Ils forment une cavité commune au centre du VFD de T1R2, comme le montre la figure 4.9A. Selon les résultats fournis par l'approche MMGBSA, les AA N143, Y164 et V384 apparaissent spécifiques pour la reconnaissance des disaccharides. La partie distincte des disaccharides est localisée dans le fond de poche de T1R2 au niveau du lobe I (figure 4.9A) et est plutôt polaire (figure 4.9B). À l'inverse, cette approche montre que les AA V66, S144, T280, Y282, W304 et L310 sont en interaction directe avec les dipeptides. Les dipeptides interagissent spécifiquement avec des résidus plutôt hydrophobes à l'entrée de la cavité (figures 4.9 A et B).

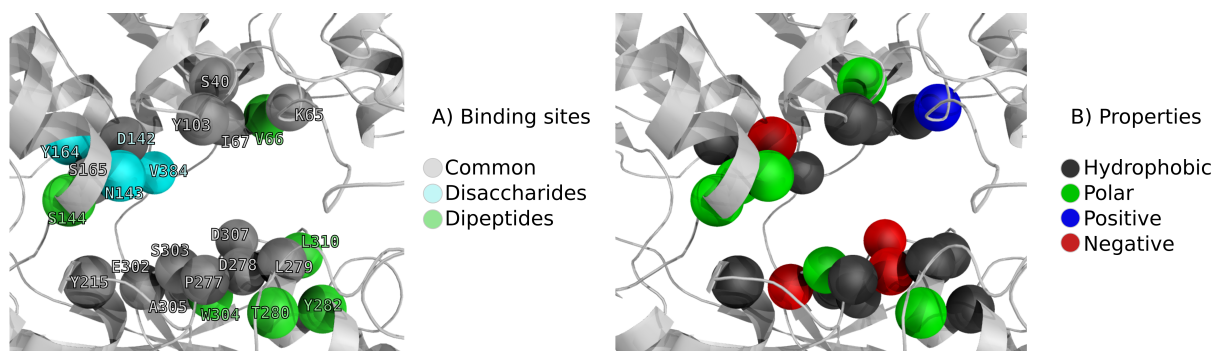


FIGURE 4.9 – Représentation en trois dimensions des sites d'interaction du VFD de T1R2 pour les disaccharides et les dipeptides.

En A) la liste des AA impliqués dans la reconnaissance des ligands. Les AA communs sont colorés en gris, en bleu et vert les AA spécifiques pour, respectivement, les disaccharides et les dipeptides. En B) les mêmes AA colorés en fonctions de leurs propriétés physico-chimiques.

4.3.6 Conclusion

Les résultats des simulations de dynamique moléculaire des complexes ligand-récepteur permettent de définir les résidus impliqués dans la reconnaissance des ligands. Pour le VFD de T1R2, la comparaison des sites d'interactions des disaccharides et dipeptides révèle une région commune et deux sites différenciables. Les disaccharides interagissent avec le fond de la cavité et plus particulièrement au niveau du lobe I. Les dipeptides, quant-à eux, sont en interaction avec des AA à l'entrée de la cavité. Le cœur hydrophobe et les groupements chargés des dipeptides sont impliqués dans cette reconnaissance différentielle. En plus des AA en interaction directe avec les ligands, la polarité de la cavité du VFD de T1R2 permet à des molécules d'eau pontantes de stabiliser les ligands. Les trajectoires montrent que les résidus K65, R383 du lobe I et D278, E302, D307 du lobe II sont en interactions lorsque le VFD est dans son état fermé [113]. Ils sont aussi impliqués dans le maintien des ligands, notamment des dipeptides [115]. Les AA mis en évidence lors de cette étude sont cruciaux pour la reconnaissance des ligands. Même si des « glucophores » traduisent les propriétés physico-chimiques des édulcorants, les données structurales permettent de caractériser les AA impliqués coté récepteur. Les AA déjà listés et ceux nouvellement énoncés peuvent servir de point de

départ pour concevoir, de manière rationnelle, de nouveaux édulcorants.

4.4 Domaine transmembranaire

Le TMD de T1R3 est connu pour interagir avec des modulateurs allostériques. Certains ligands peuvent même activer le récepteur [100, 102]. C'est pourquoi ce site peut aussi être nommé « site orthostérique secondaire ». Le cyclamate est l'un de ces ligands. Les modes d'interactions entre le cyclamate et le TMD de T1R3 ont été étudiés par mutagenèse dirigée et *docking* [100]. Les simulations de dynamique moléculaire vont permettre de compléter ces données, à la fois pour quantifier énergétiquement la contribution des AA impliqués dans la reconnaissance du cyclamate mais aussi pour mettre en évidence de nouveaux résidus clés. Des études complémentaires de mutagenèse dirigée vont appuyer les résultats proposés par les approches *in silico*.

La connaissance structurale de ce site de liaison présente un intérêt pour i) le développement de nouveaux modulateurs allostériques et ii) la compréhension des mécanismes moléculaires intervenant dans l'activation qui sera plus largement abordé dans le chapitre suivant.

4.4.1 Préparation des systèmes

Les données de mutagenèse dirigée montrent que les AA en interaction avec le cyclamate sont plus ou moins enfouis dans le TMD. Le *docking* du cyclamate dans le TMD de T1R3 révèle deux cavités (figure 4.10). Pour chaque système, deux simulations de dynamique moléculaire sont effectuées. Pour la pose haute, les trajectoires représentent 3,5 μ s et 4 μ s. Pour la pose basse, elles sont de 2 μ s et 1,7 μ s. Les simulations à partir de la pose haute présentent des temps plus longs pour tenter d'échantillonner le passage vers la pose basse.

Une simulation supplémentaire est réalisée en plaçant cinq molécules de cyclamate dans le solvant afin d'observer le phénomène d'entrée du ligand dans la cavité haute.

Une contrainte de distance est appliquée entre le centre de masse de chacun des ligands et la cystéine canonique de la TM3. Cette approche permet d'éviter que les molécules de cyclamate s'éloignent de la partie extra-cellulaire du récepteur et qu'elles n'interagissent pas avec la partie intra-cellulaire.

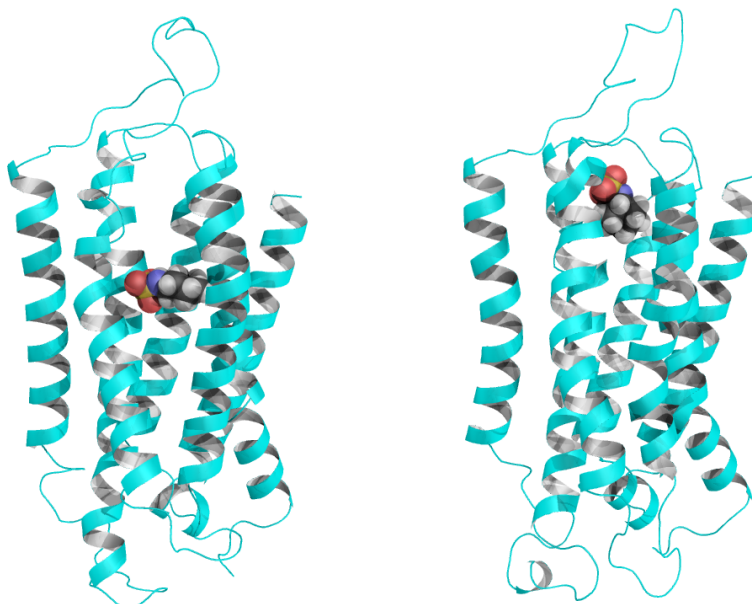


FIGURE 4.10 – Les deux sites d'interaction du TMD de T1R3 avec le cyclamate estimés par *docking*.

4.4.2 Entrée du cyclamate dans le TMD de T1R3

Cette simulation de dynamique moléculaire permet d'échantillonner le passage du ligand du solvant vers le TMD de T1R3. Au cours de cette trajectoire, trois molécules de cyclamate s'enchâssent dans la membrane à cause de leur caractère amphiphile et une autre reste dans l'eau. La dernière molécule entre dans la cavité haute du TMD de T1R3. En prenant comme référence la pose de *docking* haute, le calcul du RMSD montre que le cyclamate rejoint la cavité en 500 ns avec une valeur finale de 1,5 Å (calculé sur les atomes lourds du cyclamate, figure 4.11). Pendant son entrée dans la cavité, le ligand reste dans un vestibule pendant ≈ 100 ns. Il est intéressant de noter

que le chemin d'accès s'effectue entre les hélices 1, 2 et 7 et que l'ECL2 forme la porte d'entrée. Cette simulation confirme que le cyclamate peut rejoindre cette cavité haute.

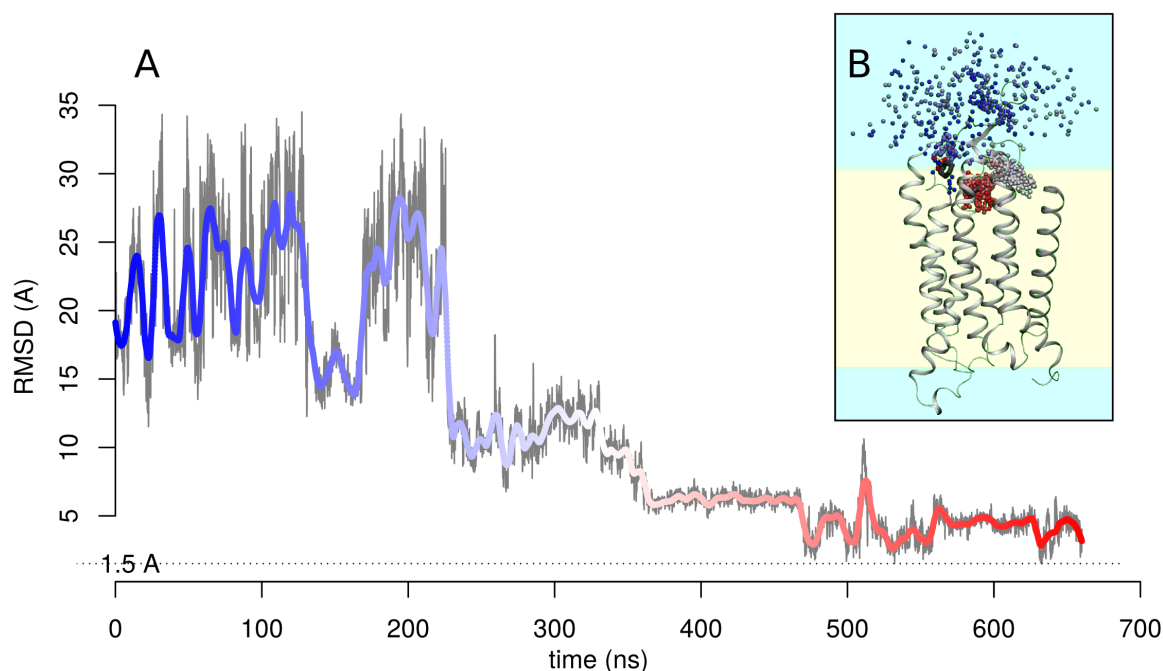


FIGURE 4.11 – Cyclamate entrant dans la cavité haute par *brute force*.

En A, le RMSD du cyclamate avec la pose de *docking* haute comme référence. En B, la visualisation de l'entrée. Le passage de la couleur bleue à rouge représente le temps de simulation.

4.4.3 Étude de la stabilité des DM

Deux simulations pour chaque pose sont effectuées. La figure 4.12 représente le RMSD du squelette de la protéine ainsi que celui des atomes lourds du cyclamate. La première fenêtre de la simulation est utilisée comme référence. Pour les quatre trajectoires réalisées, le RMSD se stabilise dans les premières ns. Quelques fluctuations sont observées pour le cyclamate. Dans le cas de la pose basse, elles sont dues à la rotation du groupement cyclohexyle du cyclamate. Dans le cas de la pose haute, il faut ajouter à cette rotation quelques mouvements de translations horizontaux dans la cavité (vers le vestibule cité précédemment).

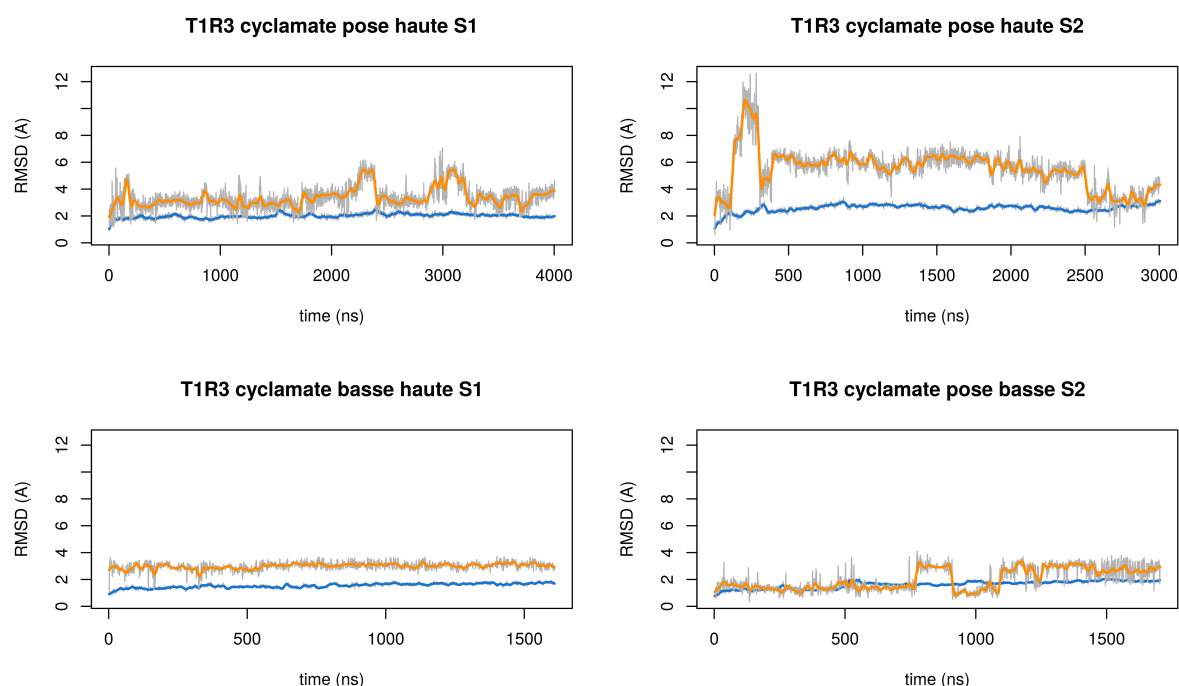


FIGURE 4.12 – RMSD du *backbone* du TMD de T1R3 (bleu) et des atomes lourds du cyclamate (orange).

4.4.4 Résidus clés du site d'interaction

Les trajectoires de chacune des poses sont concaténées, superposées aux atomes du squelette protéique de la première structure de l'étape de production et clusterisées selon le RMSD du ligand (comme pour le VFD de T1R2). Un seul cluster est trouvé pour la pose basse et deux pour la pose haute, dont un, largement majoritaire (90%). Le cluster majoritaire est sélectionné pour effectuer les calculs de l'énergie d'interaction par l'approche MMGBSA et la décomposition énergétique par résidu. Les énergies d'interaction estimées (ΔH) sont : $-31 \text{ kcal.mol}^{-1}$ pour la pose basse et $-32 \text{ kcal.mol}^{-1}$ pour la pose haute. En faisant l'approximation que le terme entropique soit équivalent entre les deux poses, cela revient à dire que le $\Delta\Delta G$ est inférieur à 1 kcal.mol^{-1} . Cela confirme la pertinence de ces deux points de départ pour réaliser des simulations de dynamique moléculaire.

La décomposition par résidu des énergies d'interaction définit les résidus impliqués dans la reconnaissance du cyclamate. Pour la pose haute, onze AA ont une contribution inférieure à $-0,5 \text{ kcal.mol}^{-1}$: P_{628}^{ECL1} , $R_{632}^{3.24}$, $C_{633}^{3.25}$, $Q_{636}^{3.28}$, $Q_{637}^{3.29}$, H_{721}^{ECL2} , C_{722}^{ECL2} , R_{725}^{ex2} , $S_{729}^{5.39}$, $F_{730}^{5.40}$ et $P_{791}^{7.29}$. Pour la pose basse, quatorze AA ont une contribution inférieure à $-0,5 \text{ kcal.mol}^{-1}$: $Q_{637}^{3.29}$, $S_{640}^{3.36}$, $H_{641}^{3.37}$, $L_{644}^{3.40}$, $H_{734}^{5.44}$, $N_{737}^{5.47}$, $Y_{771}^{6.46}$, $T_{774}^{6.49}$, $W_{775}^{6.50}$, $F_{778}^{6.53}$, $Q_{794}^{7.32}$, $A_{797}^{7.35}$, $L_{798}^{7.36}$ et $C_{801}^{7.39}$. Il est intéressant de noter que les résidus $Y_{771}^{6.46}$, $W_{775}^{6.50}$ et $C_{801}^{7.39}$ font partie des « interrupteurs moléculaires » de T1R3 et que seul le résidu $Q_{637}^{3.29}$ est commun aux deux modes d'interaction.

Les AA mis en évidence par l'approche MMGBSA dans ces deux cavités coïncident avec les données de mutagenèse dirigée énonçant les AA en interaction avec le cyclamate. Le tableau 4.4 référence les effets des mutants pour la reconnaissance du cyclamate. Les résidus dont les mutations ont un impact sur la réponse du récepteur sont : $Q_{636}^{3.28}$, $Q_{637}^{3.29}$, $H_{641}^{3.37}$, $L_{644}^{3.40}$, $T_{645}^{3.41}$, H_{721}^{ECL2} , R_{723}^{ECL2} , $F_{730}^{5.40}$, $H_{734}^{5.44}$, $Y_{771}^{6.46}$, $W_{775}^{6.50}$, $F_{778}^{6.53}$, $L_{782}^{6.57}$, $R_{790}^{7.28}$, $Q_{794}^{7.32}$, $C_{801}^{7.39}$ et $I_{805}^{7.43}$. Il n'est pas certain que tous ces AA interagissent directement avec le cyclamate, dans la mesure où leur mutation en alanine entraîne une diminution d'activité même en présence du D-Trp qui interagit avec le VFD de T1R3. Sur ces données, sept mutants ne répondent ni au cyclamate ni au D-tryptophane (D-Trp) : $L_{644}^{3.40}$ A, $T_{645}^{3.41}$ A, $H_{734}^{5.44}$ A, $Y_{771}^{6.46}$ A, $W_{775}^{6.50}$ A, $Q_{794}^{7.32}$ A, et $I_{805}^{7.43}$ A [100]. Ils traduisent une perte de la fonction du récepteur (représentés en gras dans le tableau 4.4). Ces observations sont en accord avec le fait qu'une partie de ces résidus, formant les interrupteurs moléculaires, sont critiques pour la fonction du récepteur.

Quatre AA, dont la mutation n'a aucun effet sur l'activité du récepteur, sont retrouvés par l'approche *in silico*. Les sérines $S_{729}^{5.39}$ et $S_{640}^{3.36}$ interagissent, respectivement, avec le cyclamate dans les cavités haute et basse. Il est possible que la mutation en alanine ne permette pas d'estimer correctement l'impact de cet AA dans la reconnaissance du cyclamate. La $L_{798}^{7.37}$, quant-à elle, est impliquée dans la pose basse *via* les atomes du *backbone*. Enfin, le mutant R_{725}^{ECL2} M n'entraîne pas de modification sur l'activité du

récepteur. Les propriétés physico-chimiques ne sont peut-être pas suffisamment différentes pour observer l'importance de cet AA. Des expériences de mutagenèse dirigée sur cet AA permettraient de valider son implication.

Les approches *in silico*, en considérant les deux cavités de reconnaissance, ont permis de mettre en évidence de nouveaux AA en interactions avec le cyclamate :

- pose haute : P_{628}^{ECL1} , $R_{632}^{3.24}$, $C_{633}^{3.25}$ et C_{722}^{ECL2}
- pose basse : $N_{737}^{5.47}$, $T_{774}^{6.49}$ et $A_{797}^{7.35}$

Cependant, le croisement des données *in vitro* et *in silico* laissent quelques doutes sur la réalité d'une cavité basse dans le TMD de T1R3. Même si certains mutants le confirment, la majorité des mutations des AA présents dans la cavité basse ont un impact à la fois sur le cyclamate et sur le D-Trp. Il semblerait que ces résidus ont un rôle dans la fonction du récepteur. Mais rien, d'après les expériences *in vitro*, ne garantit leur implication dans la reconnaissance du cyclamate. Cette hypothèse, qui semble valide d'un point de vue thermodynamique, devra être confirmée par les données de mutagenèse dirigée en cours de réalisation au *Monell Chemical Senses Center*.

4.4.5 Conclusion

Les études du TMD de T1R3 par *docking* et dynamique moléculaire ont permis d'explorer deux cavités d'interactions avec le cyclamate. Les données structurales et la décomposition de l'énergie d'interaction par résidu du TMD de T1R3, qui est le site « allostérique » (ou « orthostérique secondaire »), a révélé de nouveaux AA impliqués dans la reconnaissance du cyclamate. Ces AA sont soumis à des études de mutagenèse dirigée pour renforcer les données *in silico*. Cette étude apporte des informations importantes pour concevoir de nouveaux ligands en mesure d'interagir avec le TMD de T1R3 grâce à l'ajout d'AA et à la connaissance de la contribution énergétique pour chacun d'entre eux.

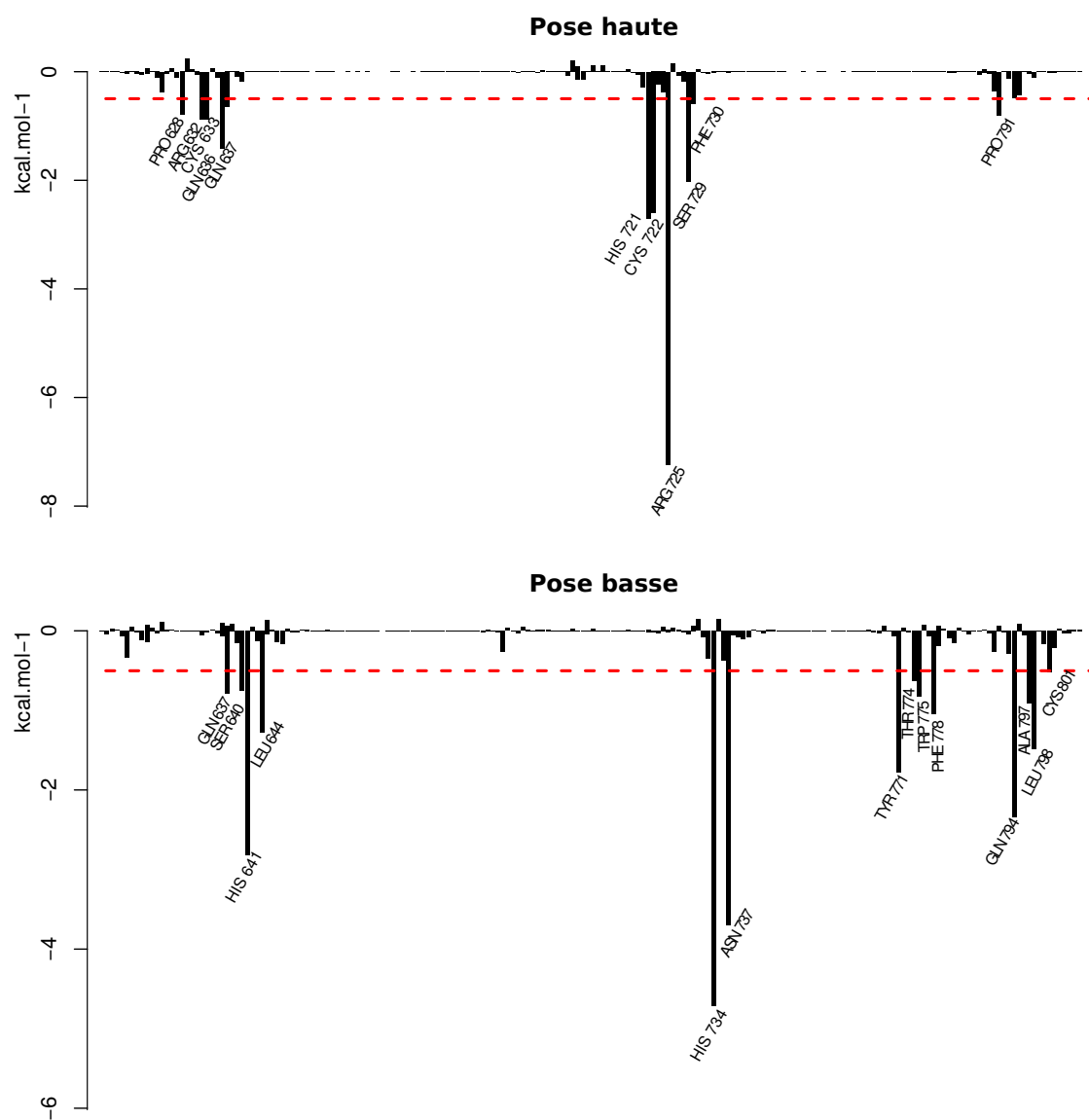


FIGURE 4.13 – Contributions énergétiques des AA des cavités haute et basse pour le cyclamate dans le TMD de T1R3.

Mutation	Position	Effet
Q636A	3.32	diminue activation
Q637E	3.33	diminue activation
S640A	3.36	sans effet
H641A	3.37	diminue activation
L644A	3.40	abolit activation
T645A	3.41	abolit activation
H721A	ex2	diminue activation
R723A	ex2	diminue activation
R725M	ex2	sans effet
S729A	5.39	sans effet
F730A	5.40	abolit activation
A733V	5.43	sans effet
H734A	5.44	abolit activation
A735I	5.45	sans effet
T739M	5.49	sans effet
Y771A	6.46	abolit activation
W775A	6.50	abolit activation
V776A	6.51	sans effet
F778A	6.53	diminue activation
V779A	6.54	sans effet
L782A	6.57	diminue activation
V788A	ex3	sans effet
L789Y	ex3	sans effet
R790A	7.28	abolit activation
Q794A	7.32	abolit activation
L798I	7.36	sans effet
L800V	7.38	sans effet
C801I	7.39	abolit activation
V802A	7.40	sans effet
I805A	7.43	abolit activation
A807V	7.45	sans effet
A808T	7.46	sans effet

TABLE 4.4 – Résumé des données de mutagenèse dirigée pour le TMD de T1R3 avec le cyclamate

Poses haute et basse en vert et bleu, respectivement. Les AA impliqués dans une perte globale de réponse (cyclamate et D-Trp) sont notés gras. Pour le détail des références, voir le tableau [A.2](#)

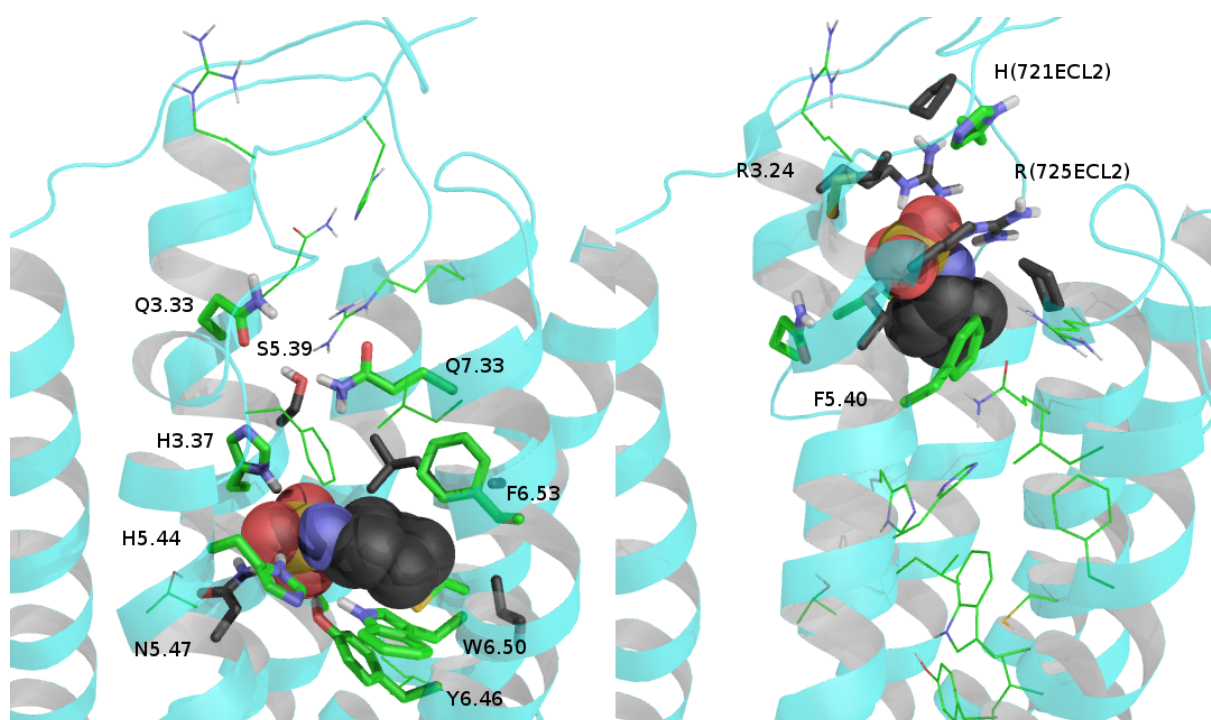


FIGURE 4.14 – Sites d'interaction du cyclamate dans le TMD de T1R3.

En « bâton » les résidus mis en évidence par l'approche MMGBSA. En vert ceux référencés par mutagenèse dirigée.

Chapitre 5

Première étape de la perception, l'activation du récepteur

« Les résultats attendus d'une simulation de dynamique moléculaire sont régis par deux préceptes : la foi et le bol. »

Pr. S. Antonczak

Ce chapitre est basé sur un article en préparation :

Chéron, J.-B., Golebiowski, J., Antonczak, S., & Fiorucci, S. (en préparation).

Molecular mechanisms underlying the activation of class C GPCR : mGluR5 as a test case.

5.1 Préambule : vers une compréhension de l'activation du récepteur.

Le fonctionnement, à l'échelle moléculaire, des RCPG de classe C est peu connu. Des données expérimentales sont disponibles pour comprendre les mécanismes globaux de leur activation. Pour le VFD, des structures cristallographiques sous forme

dimérique apportent une information sur les états actifs et inactifs (elles sont référencées dans le tableau 3.1 p 59). Ces données laissent peu de doutes sur le mode de fonctionnement du domaine extra-cellulaire. La résolution de la structure des CRD [106] permet, également, d'émettre des hypothèses sur le couplage de l'information du VFD au TMD. En effet, la rigidité de cette région suggère une transmission de l'information de manière « mécanique ». Le CRD agit comme un levier pour transmettre et amplifier le signal provenant du VFD. Enfin, le TMD est le moins connu et le moins étudié des domaines des RCPG de classes C. Quelques données biologiques permettent de comprendre le fonctionnement du TMD sous forme dimérique. Il est le domaine dont les données cristallographiques sont les plus récentes.

L'hypothèse du mécanisme global de l'activation a été mise en évidence au cours de divers travaux [136–138] et est décrite dans les paragraphes suivants. Cependant, les mécanismes moléculaires restent mal connus, en comparaison au RCPG de classe A [95, 119–121] dont les « interrupteurs moléculaires » ont largement été étudiés. L'objectif, ici, est de compléter les données expérimentales qui ont permis de proposer les mécanismes « macroscopiques » de l'activation des RCPG de classe C (décrits au paragraphe 5.2). Les simulations de dynamique moléculaire permettent de proposer un schéma, à l'échelle atomique, de la dernière partie de la transmission de l'information : le déverrouillage des « interrupteurs moléculaires ».

Dans ce chapitre, la problématique porte sur le fonctionnement des RCPG de classe C. L'importance de cette étude relève de leur implication dans divers processus biologiques (saveurs, jonctions neuronales, ...). Pour obtenir les résultats les plus pertinents possibles, le mGluR5 est utilisé comme objet d'étude, car sa structure cristallographique est disponible. Ainsi, les observations faites au cours de ce chapitre pourront être extrapolées à la famille des RCPG de classe C, incluant le récepteur à la saveur sucrée.

5.2 Hypothèse du mode d'activation pour les RCPG de classe C

5.2.1 Du VFD à la protéine G, hypothèse pour le mGluR

En condition physiologique *in vivo*, les RCPG de classe C se trouvent sous forme dimérique. Actuellement, aucune structure expérimentale complète n'est disponible. L'hypothèse du mode de fonctionnement des RCPG de classe C a été établie à partir de données cristallographiques par domaine et par des expériences *in vitro*.

Pour le VFD, le mode d'activation a été mis en évidence par analogie avec celui des LIVBP [139] et a largement été décrit par la suite [104, 111, 136, 138, 140–143]. Si le VFD interagit avec un agoniste, sa conformation « fermée » (f) est favorisée. À l'inverse, la forme apo adopte une conformation préférentielle « ouverte » (o) (cf chapitre 3.5.1). Sous forme dimérique, les structures cristallographiques ont révélé six états [141], en fonction de la conformation de chacun des partenaires. Par convention, l'état du récepteur dimérique au « repos » est noté (R) et « actif » noté (A). La forme (R) est caractérisée par une interface constituée uniquement des lobes I des deux partenaires. La forme (A), quant à elle, est définie par une interface constituée des lobes I et II. Les six états qui peuvent être définis sont : (Roo), (Rof), (Rff), (Aoo), (Aof) et (Aff). Il faut noter que la forme (Aoo) n'est pas stable, comme montré pour le mGluR1 [111], à cause de répulsions électrostatiques d'AA chargés qui se font face dans le lobe II. Cette répulsion empêche la formation d'un complexe actif (A) en absence d'agoniste.

En ce qui concerne le TMD, l'hypothèse de l'activation du dimère repose sur un réarrangement de l'interface protéine-protéine [136]. Xue *et al.* ont mis en évidence, par des études de transfert d'énergie entre molécules fluorescentes (FRET) et de *Cysteine-based cross-linking*, les TM impliqués dans les états actif et inactif du mGluR2 [137]. L'état inactif correspond à une interface protéine-protéine impliquant les TM4 et TM5 de chacun des partenaires. Quant à l'état actif, il fait intervenir les TM6 (figure 5.1)

après une rotation de $\approx 45^\circ$ de chaque TMD. Pour le TMD, deux structures cristallographiques ont été résolues récemment. Elles sont complexées avec des modulateurs allostériques négatifs [108, 109]. La structure cristallographique dont le code PDB est 4or2 est présente sous forme dimérique. Cependant, l'interface protéine-protéine est constituée par les TM1 de chaque partenaire et contient des molécules de cholestérol. Cette observation va à l'encontre des données *in vitro*, mais elle peut être due à l'absence du domaine extra-cellulaire, à un artefact de cristallisation ou à l'oligomérisation de plusieurs récepteurs [137].

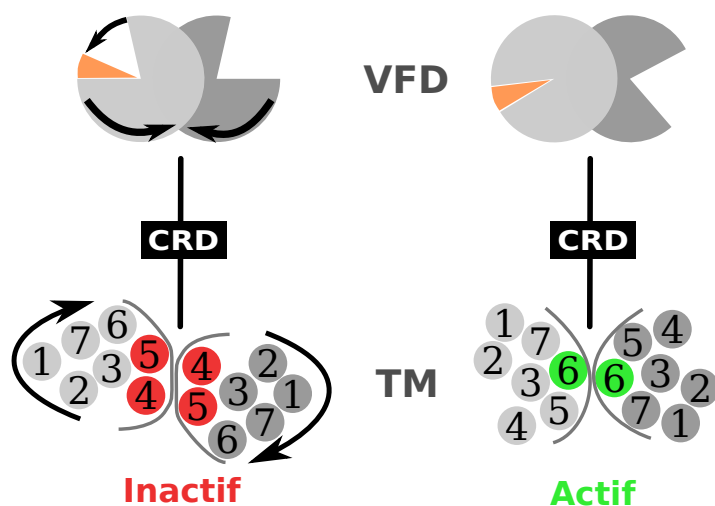


FIGURE 5.1 – Schéma représentant l'hypothèse des réarrangements structuraux lors de l'activation des RCPG de classe C de J.P. Pin et P. Rondard [137].

A partir d'un état de départ (Roo) :

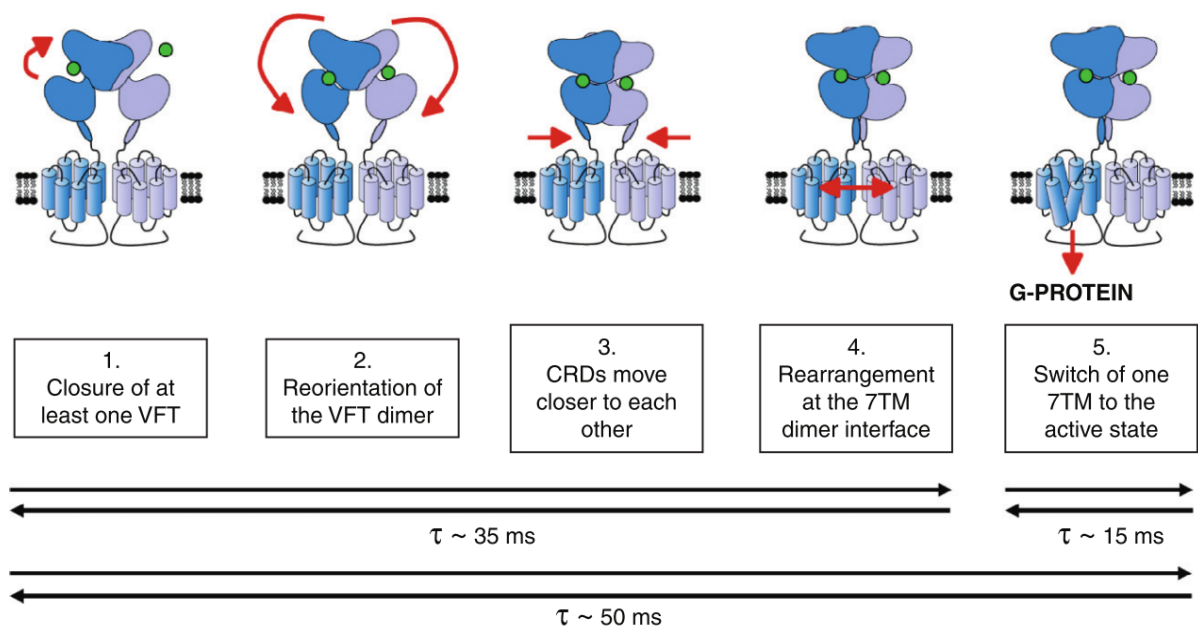
1. Stabilisation de l'état (f) du VFD par la fixation d'un agoniste dans le site orthostérique (en orange).
2. Rotation des deux VFD pour passer de l'état (Rof ou Rff) à (Aof ou Aff).
3. Transmission mécanique du signal d'activation du VFD au TMD par le CRD.
4. Réarrangement de l'interface protéine-protéine de TM4-5/TM4-5 à TM6/TM6.
5. Activation du TMD par « rupture » des interrupteurs moléculaires pour un couplage à la protéine G.

5.2.2 Cinétique d'activation

L'interaction d'un agoniste avec le site orthostérique de l'un, ou des deux, VFD permet l'activation du RCPG de classe C. La complexité structurale et le schéma d'activation cité précédemment impliquent une succession d'événements permettant de transmettre l'information à l'intérieur de la cellule. L'étude de P. Rondard et J.P. Pin [138] montre que le temps d'activation du dimère de mGluR est de l'ordre de la milliseconde (figure 5.2). Il est peu concevable d'étudier ce mécanisme à l'aide de simulations de dynamique moléculaire sur ces échelles de temps et sans contrainte [144]. Cependant, comme le schéma global d'activation est connu, il est intéressant de décrypter les mécanismes qui se mettent en place à l'échelle atomique. En ce qui concerne le VFD, comme énoncé précédemment dans ce chapitre, les données cristallographiques, ainsi que les études *in vitro* et *in silico* apportent beaucoup d'informations. Au contraire, les mécanismes moléculaires qui entrent en jeu durant la dernière étape (n°5 sur la figure 5.2) restent à élucider. La résolution récente de la structure du TMD du mGluR5 apporte les données expérimentales nécessaires pour l'étude des mécanismes impliqués dans l'activation du TMD des RCPG de classe C.

5.2.3 La partie transmembranaire du mGluR5, un prototype de RCPG de classe C

La cinétique d'activation et le nombre d'atomes constituant le récepteur ne permettent pas, aujourd'hui, de réaliser une étude à l'aide de simulations de dynamique moléculaire pour observer des phénomènes d'ampleurs, tel que l'activation du mGluR. Malgré tout, l'objectif de cette partie est de comprendre le fonctionnement moléculaire du TMD. Ce domaine a largement été étudié chez les RCPG de classe A. Le processus d'activation des RCPG de classe A a été mis en évidence à l'aide des structures cristallographiques des états actif et inactif [121, 145–148], ainsi qu'avec des simulations de dynamique moléculaire [149, 150].



Current Opinion in Pharmacology

FIGURE 5.2 – Hypothèse du mécanisme et de la cinétique d'activation des RCPG de classe C.

Image issue de la publication de P. Rondard et J.P. Pin [138].

Pour les RCPG de classe A, le mécanisme d'activation fait intervenir les « interrupteurs moléculaires » [151]. Les interactions des AA clés ont été décrites dans le paragraphe 3.5.3 p 69. En résumé, il est proposé que la fixation d'un agoniste dans le TMD change la conformation d'un AA conservé dans la TM6 : le commutateur de transmission (TS). Cet AA est généralement le W^{6.48}, selon la nomenclature des RCPG de classe A. Ce changement de conformation entraîne une modification de l'angle de courbure (*kink*) formé par la TM6. Le second interrupteur est le verrou ionique (IL) qui fait intervenir des AA de la TM3 et de la TM6. Les AA impliqués chez les RCPG de classe A sont E^{6.30} et D/E^{3.49} - R^{3.50} du motif (D/E)RY de la TM3. La rupture de ce verrou ionique serait la conséquence du changement de conformation du TS. Enfin, la barrière hydrophobe (HB) est formée par les AA du motif nPxxY de la TM7. La rupture d'un réseau d'interactions hydrogènes en amont de ce motif permet un décalage de la TM7 et une solvation de la partie intracellulaire du RCPG pour le couplage à la protéine G [148].

Pour étudier les RCPG de classe C, la stratégie employée est basée sur une étude de C. Goudet *et al.* [118]. Elle démontre que le mGluR5, dont la structure cristallographique a été résolue, se comporte comme un RCPG de classe A en absence de l'ectodomaine extracellulaire. Ainsi, l'interaction de ce récepteur tronqué avec un agoniste permet d'observer une réponse *in vitro*. Les ligands testés dans l'étude sont l'agoniste 3,3'-difluorobenzaldazine (33-DFBA) et l'agoniste inverse 2-méthyl-6-(phényléthynyl)-pyridine hydrochloride (MPEP), représentés dans la figure 5.3. Cette simplification - issue de données expérimentales - permet de diminuer le nombre d'atomes du système et d'étudier les comportements des « interrupteurs moléculaires » des RCPG de classe C, mis en évidence chez les RCPG de classe A [121, 152].

Des simulations de dynamique moléculaire du mGluR5 non complexé (apo), avec un agoniste (33-DFBA) ou avec un agoniste inverse (MPEP) vont être réalisées. Les formes actives et inactives seront extraites à l'aide d'une analyse en composantes principales sur les conformations de l'ensemble des trajectoires. Cette analyse permettra de définir les AA clés du processus d'activation. Des analyses de distances entre ces AA seront effectuées pour observer leur comportement au cours du temps et selon le système considéré. Enfin, une hypothèse du mécanisme d'activation des RCPG de classe C, basée sur le mGluR5, sera proposée.

5.3 Méthodologie des simulations de dynamique moléculaire du mGluR5

Le système membranaire du mGluR5 monomérique est utilisé pour étudier son comportement en interaction avec les ligands MPEP ou 33-DFBA. La structure cristallographique récemment élucidée sert de point de départ pour l'étude dynamique du récepteur. Cette dernière est complexée avec un modulateur allostérique négatif. La structure du mGluR5 a donc une conformation représentative de l'état inactif. Trois types de système sont préparés : un récepteur apo (sans ligand), un récepteur complexé

avec la molécule MPEP et un récepteur complexé avec la molécule 33-DFBA. Chacun des systèmes permet d'étudier le comportement du mGluR5 et sa tendance à adopter une conformation active ou inactive, en fonction du ligand.



Les atomes de carbones sont représentés en vert et rouge pour, respectivement, les molécules 33-DFBA et MPEP. Les atomes d'hydrogène, d'azote et de fluor sont colorés, respectivement, en blanc, bleu et cyan.

FIGURE 5.3 – Ligands utilisés avec le mGluR5.

5.3.1 Structure cristallographique

Les données expérimentales montrent que le mGluR5 se comporte comme un RCPG de classe A sans la partie N-terminale. Elles ont été obtenues sur le récepteur du rat [118]. Cette séquence en AA est utilisée pour réaliser les simulations de dynamique moléculaire. Les récepteurs mGluR5 de l'homme et du rat ont la même réponse lorsqu'ils interagissent avec les molécules 33-DFBA et MPEP [153]. Les séquences en AA des TMD pour les deux organismes présentent 99.6% d'identité (figure 5.4). Une structure cristallographique résolue récemment correspond au mGluR5 humain [109]. La seule différence avec celle du rat est la mutation $V^{1.60}I$, qui est prise en compte pour compléter la structure expérimentale du mGluR5. Elle présente des régions manquantes (non cristallisées) : l'ICL2 et l'ECL2. Le logiciel *Modeller* [154] est utilisé pour construire ces régions et introduire la mutation $V^{1.60}I$ à partir de la structure 4o09.

```

568 IPVQYLRWGDPEPIAAVVFACLGLLATLFVTVM1.60FIIFYRDTPVVKSSSRELC
619 YIILAGICLGYLCTFCLIAKPKQIYCYLQRIGIGLSPAMSYSALVTKTNRI
670 ARILAGSKKKICTKKPRFMSACAQLVIAFILICIQLGIIVALFIMEPPDIM
721 HDYPSIREVYILICNTTNLGVVTPPLGYNGLLILSCTFYAFKTRNV PANFNEA
772 KYIAFTMYTTCTIIWLAFVPIYFGSNYKIITMCFVSLSATVALGCMFVPKV
823 YIILAKPERN

```

FIGURE 5.4 – Séquences en AA du TMD du mGluR5.

Les AA colorés en vert sont ceux formant les hélices transmembranaires. Les régions colorées en orange sont absentes de la structure cristallographique. Enfin, l'AA 1.60 (en rouge) représente la seule différence entre la séquence en AA chez l'homme (V) et le rat (I).

5.3.2 Préparation du mGluR5 pour les simulations de dynamique moléculaire

Les molécules 33-DFBA et MPEP sont *dockées* dans le mGluR5. Seulement deux poses de *docking* sont proposées à cause de la rigidité structurale de ces deux composés et de la forme allongée de la cavité. Seule l'orientation selon le premier axe d'inertie du récepteur distingue ces deux poses des ligands dans la cavité. Celles choisies correspondent à la position proposée par Muhlemann *et al.* à partir d'études de mutagenèse dirigée et de *docking* [153].

Le protocole pour la préparation des systèmes avec membrane est similaire à celui réalisé précédemment (chapitre 4.2 p 78). Les trois systèmes (apo, avec 33-DFBA et avec MPEP) sont placés dans une membrane type POPC. L'orientation du récepteur dans la membrane a été prédite avec le serveur OPM [130].

5.3.3 Paramétrisation des dynamiques moléculaires

Les deux ligands 33-DFBA et MPEP sont préparés avec *AnteChamber* avec la méthode AM1-BCC pour assigner les charges. Les paramètres du champ de force GAFF2

sont appliqués pour les ligands. Pour la protéine et les lipides, ce sont respectivement, les champs de force FF14SB et lipid14 (les champs de force utilisés sont référencés dans le tableau 4.1 p 81).

5.3.4 Protocole des simulations de dynamique moléculaire

Le protocole des simulations dynamique moléculaire est identique à celui établi dans le paragraphe 4.2.2 p 82. Seul les temps de simulation des étapes de production changent. Pour la forme apo, deux simulations de 4 μ s sont effectuées. Pour la forme complexée au MPEP, les temps de simulations sont de 3,5 μ s et 1 μ s. Pour la forme complexée au 33-DFBA, une simulation de 5 μ s et 6 simulations de 1 μ s sont réalisées. Le détail de la méthodologie des approches de dynamique moléculaire est donné en annexe B.4.

5.4 La partie transmembranaire du mGluR5

5.4.1 Étude de la stabilité des simulations de dynamique moléculaire

Pour cette étude, 11 simulations de dynamique moléculaire ont été effectuées pour un temps simulé total de $\approx 23 \mu$ s. Pour chacune des simulations, les points de départ sont les systèmes équilibrés. Ainsi, chacune des trajectoires est réalisée avec des vitesses initiales différentes (déterminées avec une graine aléatoire) pour observer des comportements différents. La figure 5.5 représente les variations structurales de la protéine et des ligands. Le RMSD est toujours inférieur à 3 Å. Seules quelques variations du RMSD sont observées pour le MPEP, correspondant à la rotation du cycle phényle. Ces valeurs reflètent la stabilité globale au cours du temps mais aussi les variations de la structure du mGluR5 selon le système.

Les structures cristallographiques du mGluR5 sont complexées avec des modulateurs allostériques négatifs. La conformation de l'état inactif correspond à ces structures. Doré *et al.* fournissent une analyse détaillée des « interrupteurs moléculaires » en comparant les RCPG de classe A et C. Les distances entre les AA de la structure cristallographique 4o09 [109] impliqués dans l'IL sont résumées dans le tableau 5.1. Cependant, aucune structure active peut servir de référence pour comparer les conformations échantillonnées au cours des simulations de dynamique moléculaire.

Il apparaît difficile d'extraire une information, sans *a priori*, sur les trajectoires des différents complexes qui ont des comportements différents. L'analyse en composante principale (ACP) est une approche de choix pour résumer les variations structurales sur toutes les conformations échantillonnées au cours des simulations de dynamique moléculaire.

5.4.2 Analyse en composantes principales

L'ACP est une méthode statistique qui résume le plus pertinemment possible les données initiales. L'objectif est de réduire les dimensions à analyser. À partir d'un ensemble de données $n \times k$, avec n le nombre d'individus et k le nombre de variables, cette méthode définit k nouvelles variables appelées « composantes principales » (CP). Chacune des CP correspond à une combinaison linéaire des k variables établie à partir d'une matrice de variance-covariance. Les premières CP captent le plus d'information afin de maximiser la variance au sein des n individus. Sur des trajectoires de dynamique moléculaire elles permettent, notamment, d'extraire les mouvements décorrélés et les mouvements de plus grande amplitude. L'intérêt de l'ACP dans cette étude est d'identifier les structures les plus dissimilaires à partir des simulations de dynamique moléculaire.

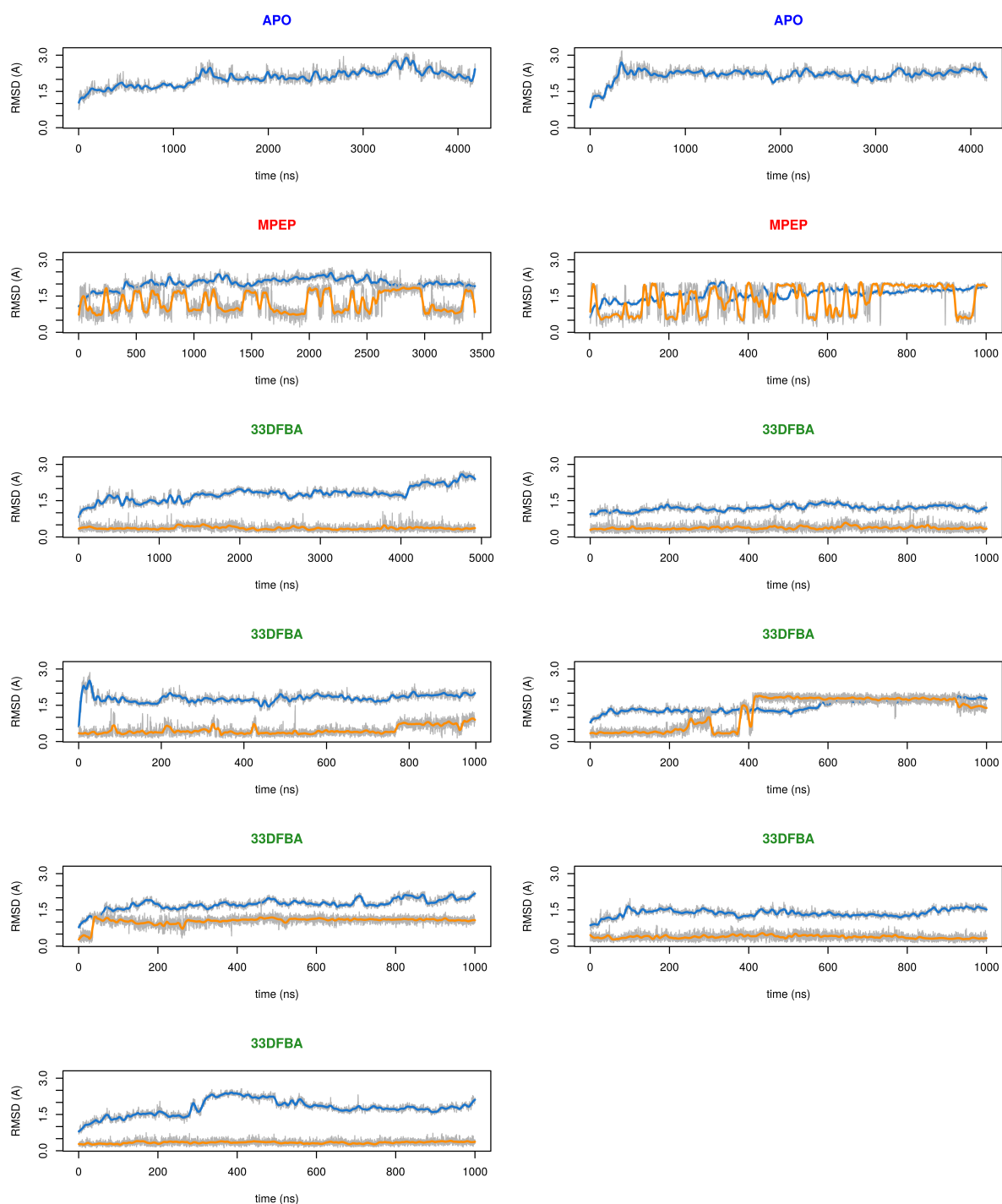


FIGURE 5.5 – RMSD du mGluR5 et des ligands 33-DFBA et MPEP lors des simulations de dynamique moléculaire.

Le RMSD est calculé en prenant comme référence la première structure de la trajectoire de dynamique moléculaire. La même échelle est utilisée pour le RMSD mais les échelles de temps sont différentes. Pour la protéine (en bleu), le RMSD est calculé sur les atomes du squelette (CA, C, O et N) des hélices transmembranaires. Pour le ligand (en orange), seul les atomes lourds sont pris en compte pour le calcul.

- **Application de l'ACP sur les trajectoires**

Comme énoncé dans la publication de Doré *et al.* [109], les TM3, TM6 et TM7 sont celles qui interviennent dans l'activation - à l'échelle atomique - des RCPG de classe C. Ces TM contiennent les interrupteurs moléculaires. L'étude de ces TM se fait sur l'ensemble des trajectoires des simulations de dynamique moléculaire concaténées (une fenêtre sur 100 est sélectionnée pour chaque trajectoire). Pour s'affranchir de la superposition de chaque conformation sur une structure de référence, l'ACP n'est pas réalisée sur les variations des coordonnées atomiques en x , y et z mais sur une matrice de distances par paire de C_α inter-hélice (TM3, 6 et 7). La matrice a une dimension de $N * D$, avec $N = 4\,977$ le nombre de fenêtres et $D = 2\,573$ le nombre de distances entre C_α . Cette matrice de distance est réduite en éliminant celles dont la variance est inférieure à 2 (soit un écart-type de ≈ 1.4 Å, considéré comme un mouvement de faible amplitude). La méthode de l'ACP est appliquée sur cette matrice finale (4 977 fenêtres $\times$ 85 distances).

- **L'ACP permet d'extraire les formes active et inactive**

La figure 5.6 montre que les 3 premières CP expliquent 75% de la variabilité de la matrice des distances. La première CP, à elle seule, capte 50% de la variabilité totale sur la matrice de distance. Elle permet de discriminer les conformations du récepteur en fonction de son interaction avec un ligand. Les conformations correspondant aux couleurs bleue, rouge et verte sont, respectivement, la forme apo, complexée au MPEP (agoniste inverse) et complexée au 33-DFBA (agoniste). La première CP distingue les conformations actives et inactives du mGluR5.

La représentation de chaque fenêtre pour chacune des simulations de dynamique moléculaire est faite dans la figure 5.7. Les structures les plus dissimilaires sont observées avec les ligands MPEP et 33-DFBA, respectivement en rouge et vert. Le RMSD des TM3, TM6 et TM7 entre la structure cristallographique et la forme active issue de la première CP est de 2,18 Å. Le RSMD entre la structure cristallographique et la forme

inactive est de 1,95 Å. Un verrou ionique est observable pour la forme inactive entre les chaînes latérales des AA K^{3.50} et E^{6.45}. À l'inverse, la forme active montre une rupture de l'interaction ionique entre la TM3 et la TM6. Cette analyse est détaillée dans le paragraphe 5.4.3.

Conformation	K ^{3.50} -E ^{6.45}		K ^{3.50} -K ^{7.51}	E ^{6.45} -K ^{7.51}	
	d _{Cα}	d _{minimale}	d _{Cα}	d _{Cα}	d _{minimale}
Forme inactive	9,3 Å	1,8 Å	13,9 Å	7,9 Å	1,9 Å
Forme active	10,5 Å	11,2 Å	18,9 Å	11,5 Å	2,1 Å
4009	11,2 Å	1,9 Å	16,3 Å	8,5 Å	1,8 Å

TABLE 5.1 – Distances entre les AA des TM3, TM6 et TM7.

d_{Cα} correspond à la distance entre les atomes C_α des AA d'intérêts. d_{minimale}, quant-à elle, est la distance inter-atomique minimale entre chacun des AA. d_{minimale} n'est pas définie sur la paire d'AA K^{3.50}-K^{7.51} à cause de la répulsion électrostatique des chaînes latérales. Les formes appelées « active » et « inactive » sont les conformations les plus extrêmes provenant de la première composante de l'ACP. Les distances qui correspondent à 4009 proviennent de la structure cristallographique du même nom.

L'ACP a permis d'extraire les structures les plus dissimilaires selon la première CP. Elle représente le mouvement de plus grande amplitude observé en bas des TM3, TM6 et TM7. Les distances entre les AA du IL (figure 5.7) sont résumées dans le tableau 5.1. L'interaction ionique entre les AA K^{3.50}-E^{6.45} peut être rompue, avec une différence de 9,4 Å, comme le montre les distances minimales calculées sur conformations les plus dissimilaires selon la première CP. L'interaction entre les AA E^{6.45}-K^{7.51} est conservée entre les deux états. La distance entre les atomes C_α des AA K^{3.50}-E^{6.45} augmente lors de l'activation du récepteur.

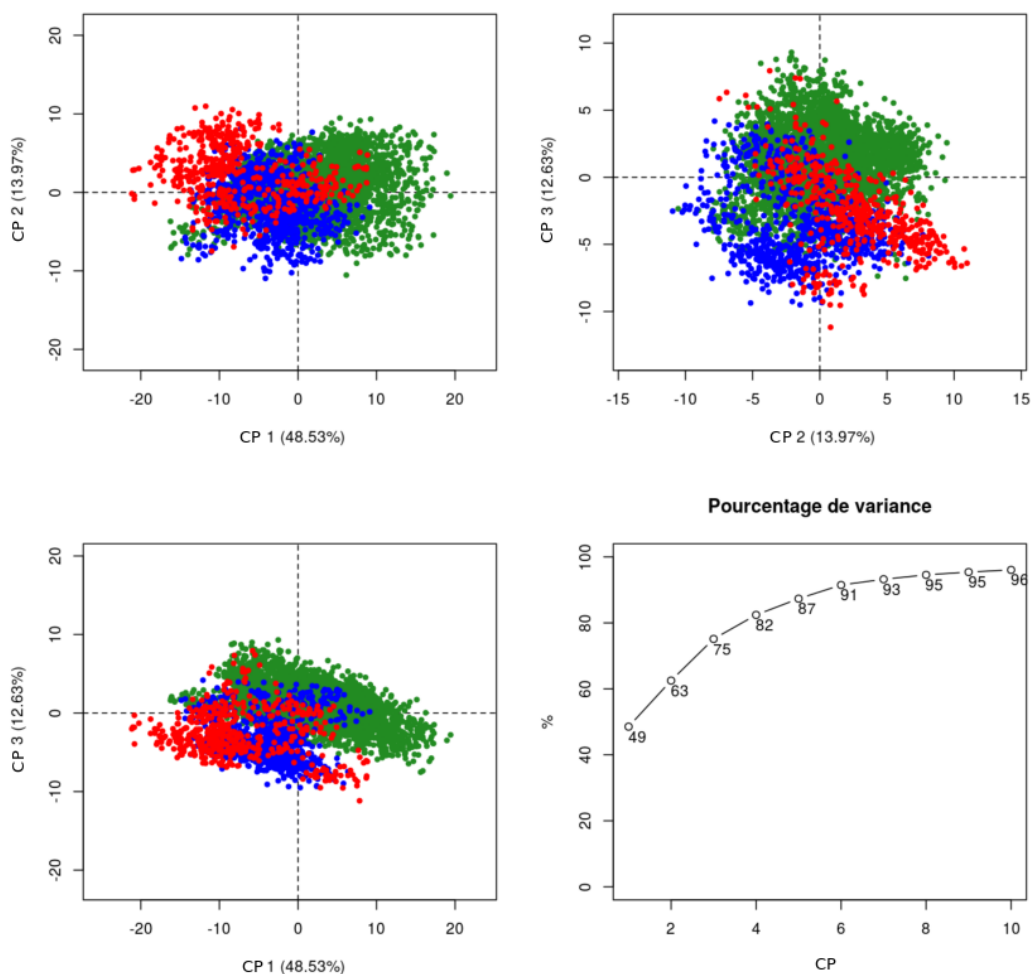


FIGURE 5.6 – Analyse en composantes principales réalisée sur l'ensemble des simulations de dynamique moléculaire.

Les composantes 1, 2 et 3 sont représentées selon les axes CP1, CP2 et CP3, ainsi que la variance expliquée par les 10 premières composantes principales. Chaque conformation correspond à un point sur la figure. Les couleurs correspondent au type de système moléculaire : mGluR5 complexé avec le ligand 33-DFBA en vert ($\approx 3\,500$ structures), mGluR5 complexé avec le ligand MPEP en rouge ($\approx 1\,000$ structures) et mGluR5 apo en bleu (≈ 500 structures).

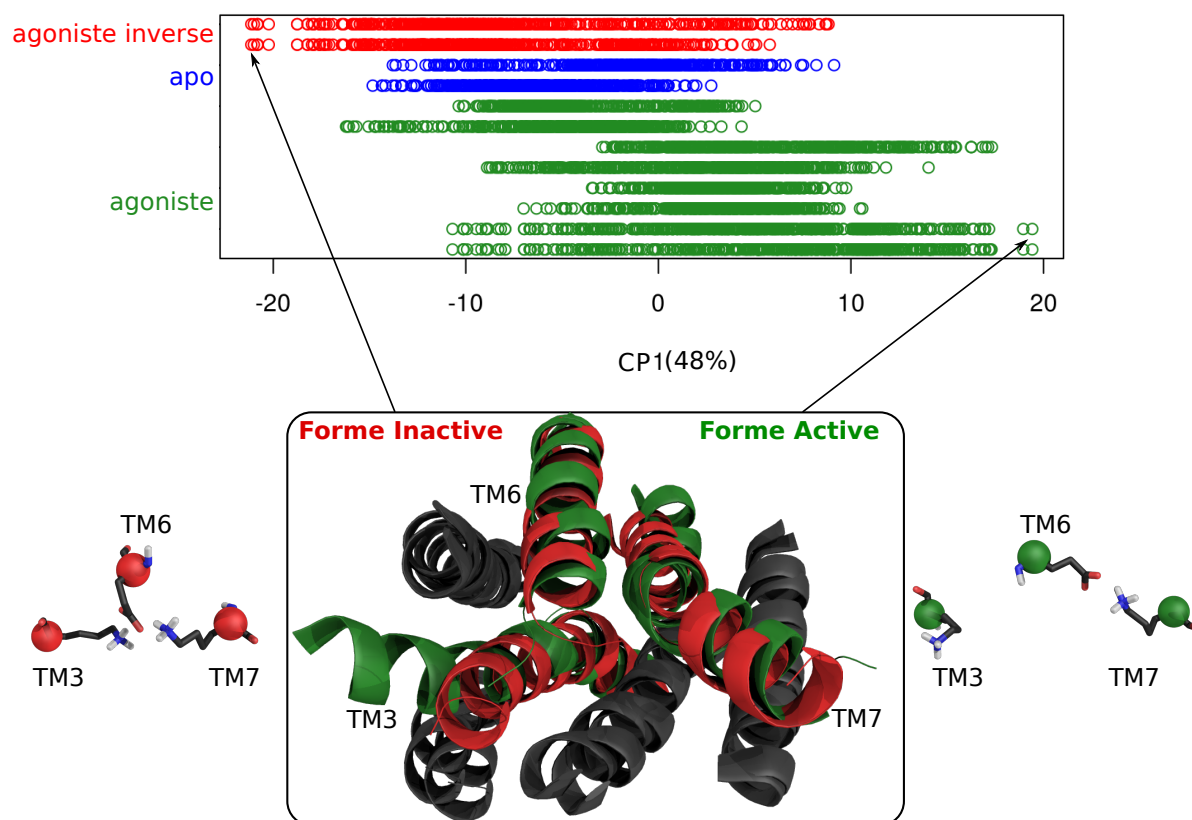


FIGURE 5.7 – Informations extraites de la première composante principale (CP1).

Chacune des trajectoires de simulation de dynamique moléculaire est représentée. Chacun des points correspond à une conformation issue des trajectoires des simulations de dynamique moléculaire. Chacune des lignes est une trajectoire unique : avec le ligand MPEP en rouge, avec le ligand 33-DFBA en vert et apo en bleu. Les structures les plus dissimilaires extraites de la première CP sont représentées en vert et rouge pour, respectivement, la forme active et inactive. Les AA formant le verrou ionique correspondant à ces structures sont également représentés.

5.4.3 Analyse des interrupteurs moléculaires

- **Les analyses de distance du verrou ionique sont en accord avec l'ACP**

L'ACP met en évidence que la plus grande variabilité associée aux états actif et inactif provient du verrou ionique. Les analyses montrent que les distances entre ces AA semblent être une métrique de l'activation du mGluR5 pour des études *in silico*. Les figures 5.8 et 5.9 représentent les variations de distance entre les AA de cette triade (K^{3.50}, E^{6.45} et K^{7.51}) au cours du temps sur une trajectoire représentative de chaque forme : apo, agoniste, agoniste inverse. Ces trajectoires ont permis de définir des seuils de distance pour distinguer les états actifs et inactifs au cours d'une trajectoire. Ces valeurs ont été choisies à partir de la distribution de chacune des distances représentées sur les figures 5.8 et 5.9, où les points bleus et gris correspondent, respectivement, à des conformations considérées comme actives et inactives.

Pour la simulation avec l'agoniste (33-DFBA), une rupture des interactions ioniques est observée, notamment entre les AA K^{3.50}-E^{6.45} (figure 5.8). Ce comportement n'est pas observé pour la forme apo et le mGluR5 complexé avec un agoniste inverse (MPEP). Cette rupture de l'interaction ionique entraîne un écartement des hélices transmembranaires 3, 6 et 7. La distance mesurée entre les atomes C_α des AA K^{3.50}-K^{7.51}, montre un écartement du bas des TM3 et TM7 (figure 5.9). Les variations de distances entre les TM3 et TM6 ainsi que TM6 et TM7 sont moins flagrantes, ce qui suggère un mouvement coordonné de la TM6 lors de l'écartement des TM3 et TM7.

Les distances mesurées entre les TM3, TM6 et TM7 sont en accord avec la première CP de l'ACP. En effet, au cours des trajectoires analysées, le mGluR complexé avec l'agoniste 33-DFBA explore plus souvent une conformation active. Ainsi, les variations de distances des AA du verrou ionique sont un marqueur de l'activation du mGluR5 au cours du temps.

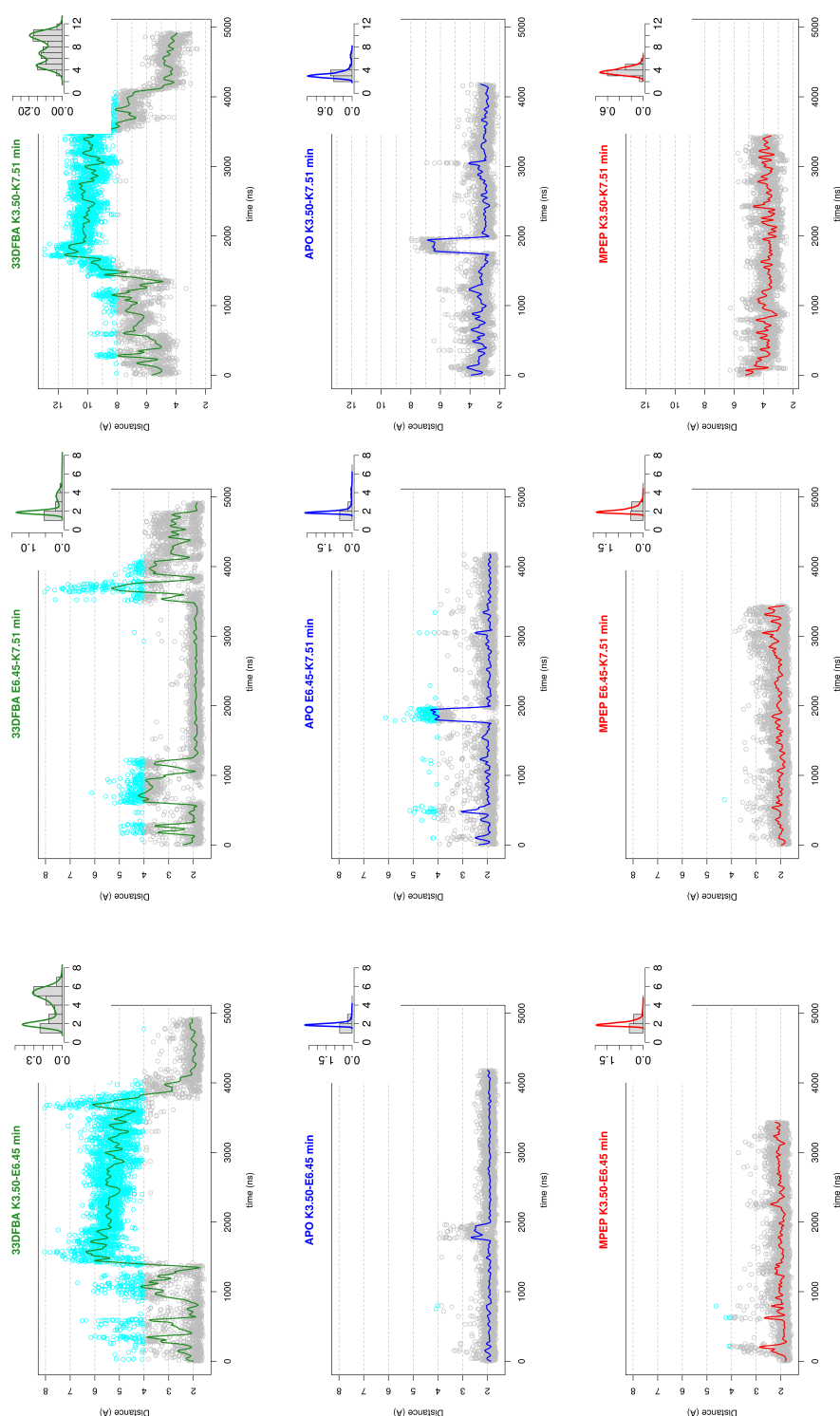


FIGURE 5.8 – Distances minimales mesurées entre les atomes des AA du verrou ionique.

Les points en gris et bleu sont des conformations considérées comme, respectivement, inactives et actives. Pour chaque distance, la répartition des conformations selon la distance est représentée en haut à droite.

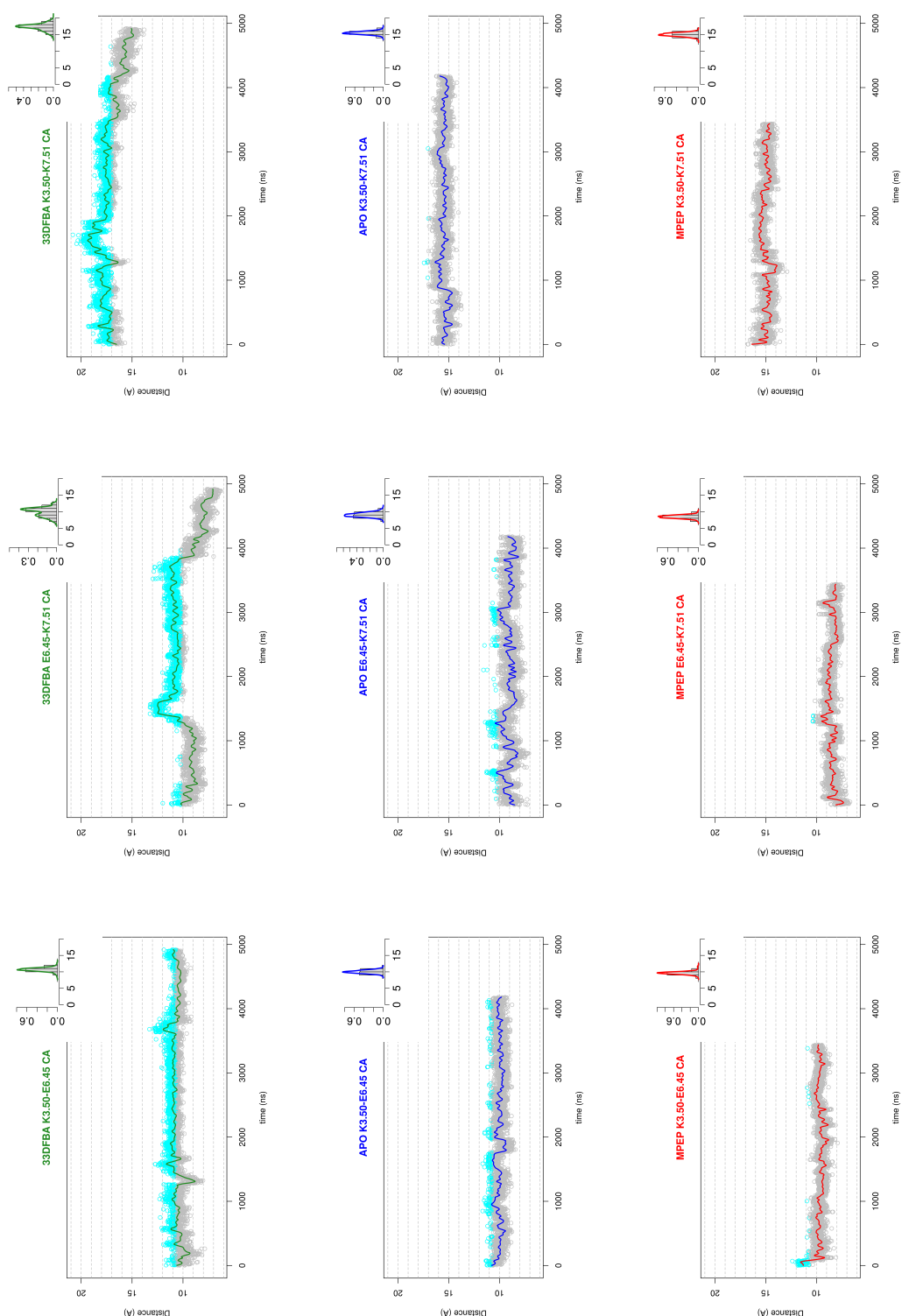


FIGURE 5.9 – Distances mesurées entre les atomes C α des AA du verrou ionique.

Les points en gris et bleu sont des conformations considérées comme, respectivement, inactives et actives. Pour chaque distance, la répartition des conformations selon la distance est représentée en haut à droite.

- **L'analyse du verrou ionique est une métrique de la forme active.**

L'analyse des mêmes distances, $K^{3.50}-E^{6.45}$, $E^{6.45}-K^{7.51}$ et $K^{3.50}-K^{7.51}$, est effectuée sur l'ensemble des trajectoires réalisées pour cette étude. Les seuils de distances définis à partir des figures 5.8 et 5.9 sont utilisés pour définir si une conformation est active ou inactive. Les figures 5.10 et 5.11 résument les formes considérées comme actives (colorées) ou inactives (grisées) au cours des simulations.

La figure 5.10 montre que, pour toutes les trajectoires, l'interaction ionique entre les AA $E^{6.45}-K^{7.51}$ est conservée quel que soit le complexe (peu de coloration violette). À l'inverse, l'interaction ionique formée par les AA $K^{3.50}-E^{6.45}$ est fréquemment rompue (coloration marron si distance > 4 Å) lorsque le mGluR5 est complexé avec l'agoniste 33-DFBA. La même constatation peut être faite avec les AA $K^{3.50}-K^{7.51}$, les chaînes latérales sont plus éloignées avec les formes complexées à la molécule 33-DFBA (coloration orange si distance > 8 Å).

La figure 5.11 représente les états actifs et inactifs selon les seuils de distance appliqués sur les atomes $C\alpha$, ainsi que l'angle de courbure de la TM6. L'angle de la TM6, qui contient l'AA $W^{6.50}$, semble être un marqueur de la forme active (coloration noir si l'angle est inférieur à 133°). La fermeture de cet angle correspond à une entrée du bas de la TM6 entre les hélices 6 et 7. Ensuite, le marqueur le plus probant est la distance entre la TM3 et la TM7 (en orange), comme montré par l'ACP. L'écartement de ces deux hélices laisse la place pour le bas de la TM6 qui suit le mouvement d'écartement de la TM3. Selon ces critères et comme pour l'étude des distances minimales, les formes complexées à la molécule 33-DFBA adoptent une conformation active dans la majorité des cas. Une faible proportion des conformations du récepteur, lors des simulations sous forme apo, est présente à l'état actif. Elle est encore plus faible s'il est complexé à l'agoniste inverse. L'exploration d'une telle conformation peut correspondre à la présence d'une activité « basale » du récepteur.

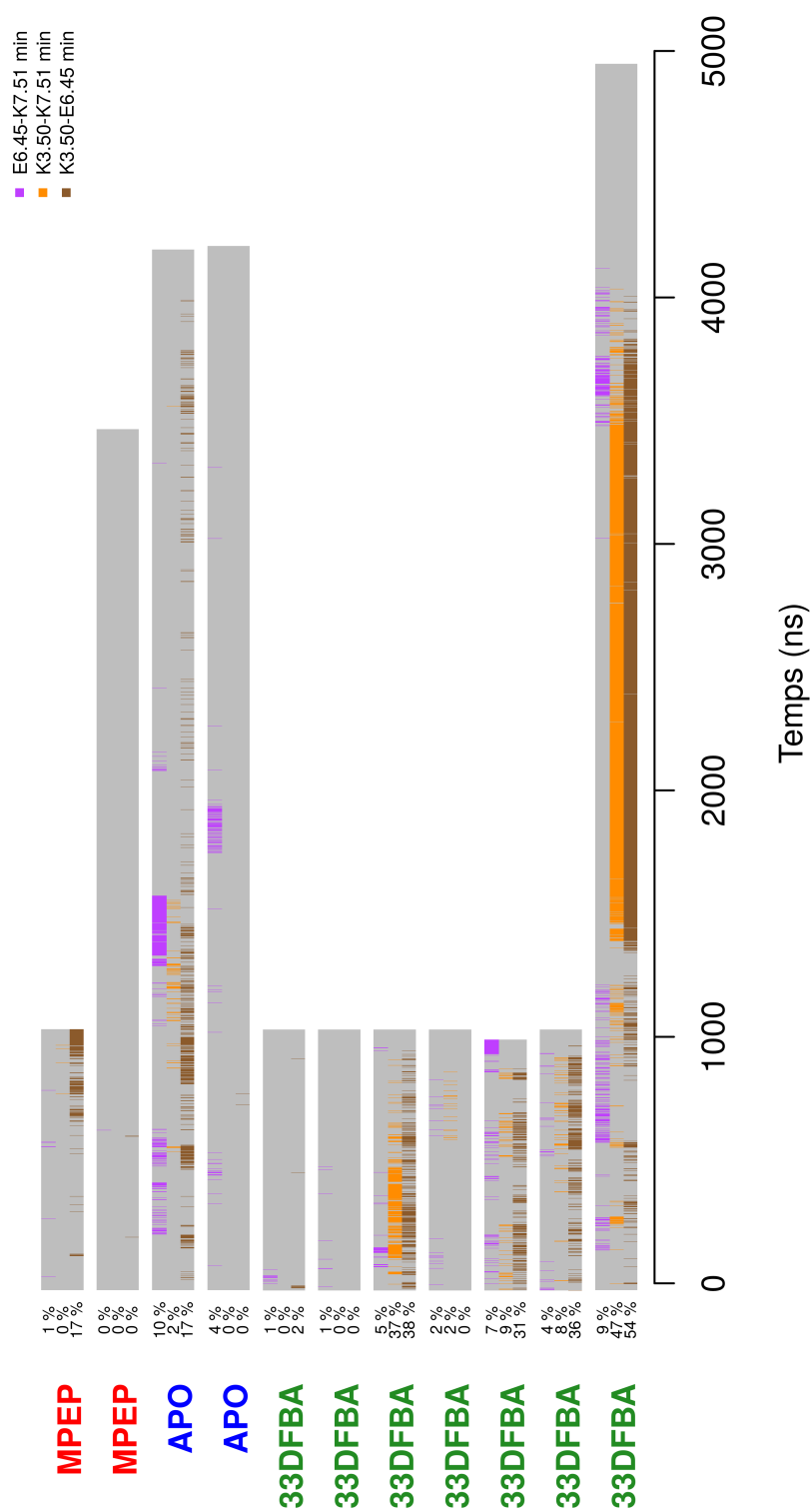


FIGURE 5.10 – Résumé des distances minimales entre les AA du verrou ionique selon chaque simulation de dynamique moléculaire.

Les régions colorées correspondent à la forme considérée comme active pour chacune des métriques. Pour chacune des simulations, les conformations sont colorées en violet et marron si, respectivement, les distances minimales entre les AA $\text{K}^{3.50}\text{-E}^{6.45}$ et $\text{E}^{6.45}\text{-K}^{7.51}$ sont supérieures à 4 Å. Elles sont colorées en orange si la distance minimale entre les AA $\text{K}^{3.50}\text{-K}^{7.51}$ est supérieure à 8 Å.

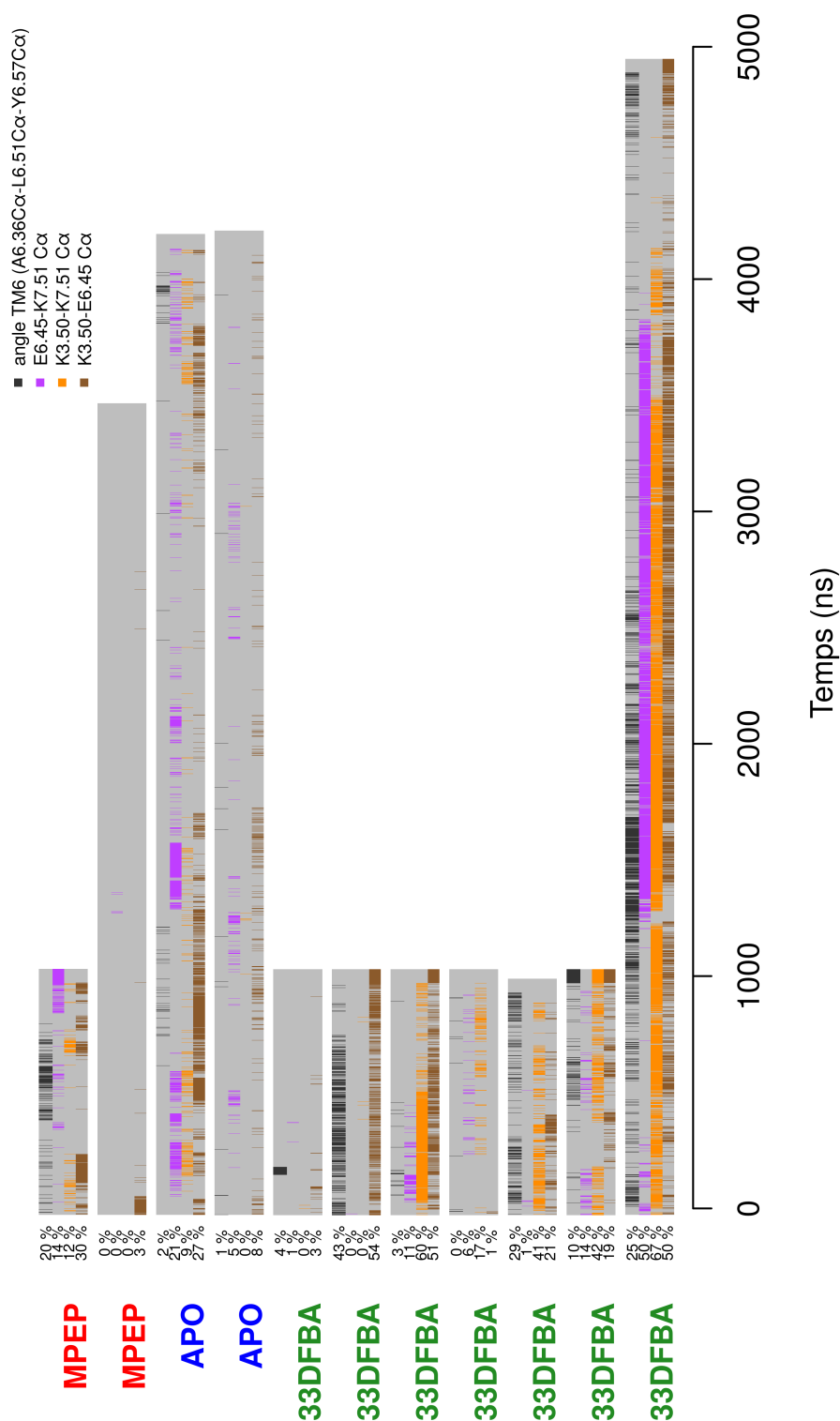


FIGURE 5.11 – Résumé des distances entre les atomes Cα des AA du verrou ionique selon chaque simulation de dynamique moléculaire.

Les régions colorées correspondent à la forme considérée comme active pour chacune des métriques. Pour chacune des simulations, les conformations sont colorées en violet et marron si, respectivement, les distances minimales entre les AA K^{3.50}-E^{6.45} et E^{6.45}-K^{7.51} sont supérieures à 10 Å. Elles sont colorées en orange si la distance minimale entre les AA K^{3.50}-K^{7.51} est supérieure à 17 Å. Enfin, la coloration noire représente les structures en voie d'activation avec un angle de courbure de la TM6 plus aigu que 133°.

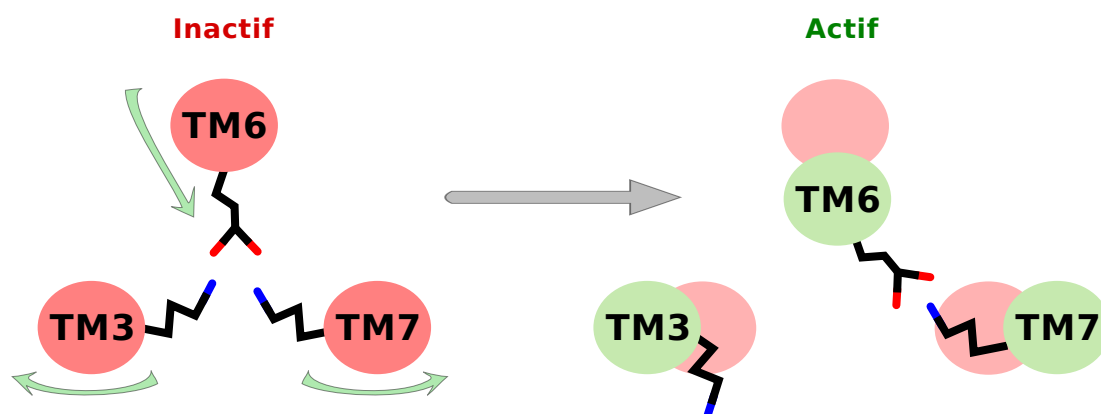


FIGURE 5.12 – Schéma de l'activation du verrou ionique pour le mGluR5.

Dans cette hypothèse, le signal précurseur serait la courbure de la TM6 avec un point d'inflexion au niveau de l'AA W^{6.50}. La courbure de la TM6 entraînerait une rupture du verrou ionique lors de l'activation. La chaîne latérale de l'AA E^{6.45} reste en interaction avec l'AA K^{7.51} tandis que l'interaction ionique avec l'AA K^{3.50} est rompue. La perte de cette interaction permet un écartement des TM3 et TM7 afin de laisser la place à la TM6.

5.4.4 Schéma de l'hypothèse d'activation du verrou ionique

À l'aide des données issues de l'ACP et des analyses de distance du verrou ionique, il est possible de proposer une hypothèse du mécanisme d'activation du mGluR5. La principale information apportée par l'ACP est que la partie intracellulaire du récepteur adopte différentes conformations. Les atomes C α des AA K^{3.50} et K^{7.51} s'écartent, laissant de la place à la TM6 pour se placer entre ces deux dernières. La chaîne latérale de l'AA E^{6.45} reste en contact avec l'AA K^{7.51}. La fermeture de l'angle de la TM6, avec un point d'inflexion au niveau de l'AA W^{6.50}, semble avoir lieu avant la rupture du verrou ionique et permet un mouvement du bas de l'hélice. L'AA W^{6.50} apporte une première piste pour la compréhension de la transmission du signal entre la cavité de liaison et la partie intracellulaire du récepteur. Il est intéressant de noter que la TM6 est celle qui intervient dans l'interface protéine-protéine de l'état actif du récepteur dimérique. Le déplacement du bas de cette hélice peut laisser de la place à l'interface protéine-protéine pour un couplage à la protéine G. La figure 5.12 représente l'hypothèse du changement conformationnel induit lors de l'activation du mGluR5.

5.5 Conclusion

Un ensemble de données expérimentales a été récolté pour obtenir un point de départ pertinent afin de réaliser une étude sur l'activation du mGluR5 :

- les données cristallographiques récentes du mGluR5 [108, 109]
- le fonctionnement du TMD du mGluR5 sans l'ectodomaine extracellulaire [118]
- les données de mutagenèse dirigée pour les molécules 33-DFBA et MPEP [153]
- une hypothèse du mécanisme d'activation du mGluR2 [137]

Les simulations de dynamique moléculaire réalisées sur le mGluR5 tronqué sous forme apo, complexé avec un agoniste (33-DFBA) ou un agoniste inverse (MPEP) ont permis d'émettre une hypothèse sur le schéma d'activation des RCPG de classe C. L'ACP réalisée sur l'ensemble des trajectoires a mis en évidence que la conformation de la région intracellulaire des TM3, TM6 et TM7 permet de discriminer l'état actif de l'état inactif. Les analyses de distances sur cette région, montrent que le verrou ionique se rompt lors de l'activation et confirme que le mode d'activation des RCPG de classe C est similaire à celui des RCPG de classe A. La stratégie visant à répliquer les simulations confirme l'hypothèse émise, car des comportements similaires sont observables pour chaque système.

Un réarrangement structural des TM3, TM6 et TM7 a lieu lors de l'activation. La rupture du verrou ionique entre la TM3 et la TM6 permet un écartement des atomes $C\alpha$ des TM3 et TM7. L'atome $C\alpha$ de la TM6 suit le mouvement de la TM3 et libère de l'espace à l'interface TM6-TM6 qui est caractéristique de la forme active du dimère de mGluR. Le précurseur de ce changement de conformation est le résidu W^{6.50}. Il est impliqué dans la conformation « coudée » de la TM6 et confirme l'hypothèse de son rôle crucial émise dans le chapitre 3.

Le « commutateur de transmission » impliquant les TM3, TM6 et TM7 forme le fond de la cavité de liaison. Il a, chez les RCPG de classe A, un rôle de « senseur » du ligand. Des analyses de distances réalisées sur le mGluR5 n'ont pas permis de mettre

en évidence différentes conformations entre les états actifs et inactifs. Des analyses complémentaires aideront, peut-être, à expliquer le rôle de ces AA en amont (d'un point de vue temporel) de l'activation du récepteur.

Des simulations de dynamique moléculaire complémentaires sont à réaliser, notamment du mGluR5 à l'état dimérique, pour comprendre les mécanismes qui se mettent en place à l'interface protéine-protéine lors de l'activation. Les AA correspondant au IL sur le récepteur à la saveur sucrée (STR) sont R^{3.50}, E^{6.45}, K^{7.51} pour T1R2 et Q^{3.50}, R^{6.45}, R^{7.51} pour T1R3. Les propriétés physico-chimiques des AA ne sont pas réparties de la même façon sur T1R3 que sur les autres RCPG de classe C. En effet, pour T1R3, c'est la TM3 qui porte l'AA chargé négativement. Ceci peut expliquer, en partie, pourquoi T1R2 est la seule sous-unité du STR qui couple l'information à la protéine G [80].

Chapitre 6

Conclusion générale

Les sucres sont une source d'énergie pour l'organisme et sont indispensables à la vie. Par conséquent, cette saveur a une forte valence hédonique positive et permet, depuis son origine, la sélection d'aliments riches en énergie. Aujourd'hui, l'accès aux denrées alimentaires est largement facilité et cette perception n'est plus associée à des fins de survie. Le plaisir engendré par la saveur sucrée peut conduire à une consommation excessive et chronique de sucre. Ce comportement est un facteur de risques menant à des pathologies telles que les caries, le diabète de type II, les maladies cardiovasculaires et l'obésité. La réduction de la quantité de sucre dans l'alimentation est devenue un enjeu majeur de santé publique.

Les édulcorants sont des composés perçus sucrés peu - ou pas- caloriques. L'intensité de perception des molécules sucrées est appelée « Pouvoir Sucrant » (PS). Il est défini comme le rapport entre une masse - ou une concentration - de saccharose et d'une substance sucrée à intensité de perception égale. Il est facile d'imaginer qu'un PS plus important permet de diminuer la quantité de molécules ingérées pour une même expérience sensorielle. Ces composés interagissent avec le récepteur à la saveur sucrée (STR pour *Sweet Taste Receptor*) qui est le premier protagoniste de la perception. Il est constitué de deux RCPG de classe C : T1R2 et T1R3, où plusieurs sites de liaisons ont été reportés. Les sucres, ainsi que la plupart des édulcorants, interagissent avec le *Venus Flytrap Domain* (VFD) de T1R2 ou T1R3. De plus, chacun des Domaines TransMembranaires (TMD) est en mesure de reconnaître des édulcorants ou des modulateurs

allostériques. La présence de ces multiples sites de liaison et la diversité structurale des composés que le STR est capable de reconnaître traduisent la difficulté d'établir des relations structure/goût.

L'enjeu majeur de cette étude est de proposer des pistes pour une sélection et une conception rationnelle de composés potentiellement sucrés. Cependant, peu de données structurales sont disponibles pour analyser, d'un point de vue moléculaire, les modes de liaisons et le comportement dynamique de l'association ligand-récepteur. Au cours de ces travaux de thèse, deux approches ont été abordées. La première est une approche statistique basée sur la seule connaissance de la structure des ligands. La seconde tient compte du récepteur et utilise des approches de mécanique moléculaire. Les études effectuées sur un modèle par homologie du STR (VFD et TMD) ont permis une meilleure description des sites d'interactions ligand-récepteur et ont mis en évidence de nouveaux AA clés. Ces résultats sont essentiels pour une conception rationnelle de nouveaux édulcorants. Ensuite, le mécanisme d'activation des RCPG de classe C a été étudié avec le TMD du mGluR5 (Récepteur Métabotrope au Glutamate de type 5). L'élucidation de la structure tridimensionnelle de ce récepteur est une avancée majeure dans la description moléculaire de cette famille et a permis, au cours de ce projet, d'en explorer le comportement dynamique. L'observation de ce phénomène d'activation, lors de la liaison d'une molécule, est une métrique pertinente en complément de la mesure de l'affinité ligand-récepteur pour prédire la qualité (agoniste, agoniste-inverse, antagoniste) d'un ligand potentiel.

La première partie de cette étude consistait à établir un protocole combinant plusieurs approches statistiques pour sélectionner des molécules sucrées, non toxiques et sans goût amer. Dans cet objectif, 316 composés avec un PS allant de 0,2 à 225 000 et provenant de familles chimiques variées ont été référencés dans une base de données (*SweetenersDB* disponible sur internet). Les études structurales ont montré que les

meilleurs édulcorants présentent un caractère hydrophobe, une masse molaire modérée (350 à 450 g.mol⁻¹), 4 donneurs et jusqu'à 10 accepteurs de liaisons hydrogène. La comparaison des espaces chimiques couverts par ces 316 composés avec la plus grande base de données de molécules apparentées à des sucres atteste de son exhaustivité. Des descripteurs moléculaires 2D et 2D+3D ont été calculés pour construire un modèle de prédiction du PS. Après sélection des descripteurs pertinents et comparaison des modèles, il est apparu que la méthode des machines à vecteurs de support (SVM) avec les descripteurs 2D est la moins coûteuse en temps de calcul et la plus performante, avec un coefficient de détermination de 0,69 sur le jeu d'apprentissage et 0,83 sur le jeu de validation. L'application de ce modèle sur plus de 300 000 composés naturels a permis de sélectionner des composés avec un PS supérieur à 100. Enfin, l'application d'un premier filtre pour éliminer les composés potentiellement amers (basé sur un modèle de classification), puis un second pour éliminer les molécules présentant des groupements toxiques connus, a conduit à la sélection de 4 585 composés d'intérêts. Lors de cette étape, un modèle de prédiction performant a été construit et son application permet de proposer une liste de molécules naturelles potentiellement sucrées. Cette approche a permis de sélectionner des nouvelles molécules jamais testées, qui sont actuellement en cours d'étude *in vitro*.

La seconde partie de cette étude a constitué un point crucial pour le reste des travaux. Dans la perspective d'utiliser le STR en complément des données structurales des ligands, il a fallu proposer une structure 3D. L'absence de donnée cristallographique a pu être palliée par la modélisation par homologie. Le recensement des structures expérimentales des RCPG de classe C, ainsi que plus de 600 données de mutagenèse dirigée ont servi à affiner l'alignement de séquences en AA. En effet, la structure secondaire et les AA clés, tels que ceux constituant les sites de liaison, ont été utilisées comme contraintes pour l'alignement de séquences en AA. Un modèle par homologie du STR a pu être proposé. Les analyses de cet alignement et de la structure du STR

permettent de décrire les sites de liaison et ont révélé les motifs structuraux importants du TMD. Le *Transmission Switch* est formé par une triade d'AA aux positions 3.44, 6.46 et 7.39 (respectivement I, Y et N pour T1R2 et L, Y et C pour T1R3). Le résidu W^{6.50} est également très conservé chez les RCPG de classe C. C'est l'analogue du W^{6.48} chez les RCPG de classe A, ce qui suggère un rôle important pour la fonction du récepteur. Le *Ionic Lock* implique les AA Q^{3.50} et R^{6.35} pour T1R3 et R^{3.50} et E^{6.35} pour T1R2. Enfin, le modèle confirme la présence d'un motif [F/Y] xPKxY conservé qui forme la barrière hydrophobe dans la TM7. Ainsi, ce modèle du STR constitue un point de départ robuste pour décrire la dynamique des ligands dans les sites de liaison et définir les AA clés tels que les interrupteurs moléculaires à l'aide d'approches de dynamique moléculaire par domaine.

La structure du VFD de T1R2 a été utilisée pour comprendre comment un seul site de liaison est en mesure d'interagir avec la plupart des composés sucrés connus, notamment les disaccharides et les dipeptides. L'identification des AA en interaction avec ces deux familles chimiques a aidé dans la compréhension du large spectre de reconnaissance des molécules par le STR. Des modes de liaison différents ont été observés entre les disaccharides et les dipeptides. La mise en commun des données de mutagenèse dirigée et des approches *in silico* a permis de définir la contribution énergétique de chaque AA. Les disaccharides interagissent avec des AA plutôt polaires enfouis dans le VFD, notamment avec les résidus N143, Y164 et V384. À l'inverse, les dipeptides vont interagir de façon spécifique avec des AA moins enfouis dans la cavité de liaison. Ils établissent des interactions avec, par exemple, les AA V66, S144, T280, Y282, W304 et L310.

Cette étude a également porté sur l'investigation du site de liaison du TMD de T1R3. Le *docking* moléculaire a estimé l'existence de deux cavités d'interaction avec le cyclamate, une haute et une basse. L'entrée du cyclamate dans la cavité haute du récepteur au cours d'une simulation de dynamique moléculaire met en évidence un

vestibule, ainsi que la voie d'entrée du cyclamate entre l'ECL2 et les TM1, TM2, TM7. Chacune des poses haute et basse ont été utilisées comme point de départ pour des simulations de dynamique moléculaire. La décomposition par résidu des énergies d'interaction a mis en évidence les AA clés de l'interaction ligand-récepteur. La combinaison des résultats des deux cavités permet de retrouver les AA provenant d'études de mutagenèse dirigée par Jiang *et al.* en 2005, mais aussi d'en sélectionner de nouveaux. Suite à ces observations, des études complémentaires de mutagenèse dirigée sont en cours sur les AA impliqués dans la reconnaissance du cyclamate dans la cavité basse et ceux qui présentent un intérêt pour la fonction du récepteur.

La dernière partie présentée au cours de ce projet consiste en l'étude du mécanisme d'activation des RCPG de classe C. Des approches expérimentales ont proposé le mécanisme global d'activation. Cependant, le fonctionnement de la dernière étape de l'activation des RCPG de classe C, impliquant la rupture des interrupteurs moléculaires, reste méconnu. Pour comprendre ces mécanismes, la structure cristallographique du TMD du mGluR5 a été utilisé comme prototype. Des simulations de dynamique moléculaire de ce récepteur sous forme non complexé, en interaction avec un agoniste puis un agoniste inverse ont été effectuées. Une analyse en composantes principales sur l'ensemble de ces trajectoires a mis en évidence les structures les plus dissimilaires, correspondant aux formes actives et inactives. Elles ont permis d'émettre une première hypothèse sur l'activation impliquant les TM3, TM6 et TM7. La fermeture de l'angle de la TM6, avec un point d'inflexion au niveau de l'AA W^{6.50}, a lieu avant la rupture du verrou ionique et permet un mouvement du bas de l'hélice 6. Ainsi, les atomes C α des AA K^{3.50} et K^{7.51} s'écartent, laissant de la place à la TM6 pour se placer entre ces deux dernières. Ces résultats apportent une première piste pour la compréhension de la transmission du signal entre la cavité de liaison et la partie intracellulaire du récepteur.

Les résultats de ces travaux de thèse constituent un modèle, au niveau atomique, de la première étape de la perception de la saveur sucrée. L'utilisation d'une approche statistique basée sur la structure des sucres et des édulcorants s'est révélée performante pour prédire le pouvoir sucrant. La conception d'un modèle par homologie du récepteur à la saveur sucrée et sa mise à disposition pour la communauté scientifique présente une avancée majeure pour les futurs travaux de modélisation moléculaire. Cela a d'ailleurs été l'objet des premières investigations illustrées dans ce manuscrit avec la description des cavités de liaison. Enfin, l'étude du mGluR5 permet d'apporter les premiers indices des mécanismes moléculaires qui ont lieu lors de l'activation des RCPG de classe C. L'ensemble de ces travaux apporte de nombreux outils et de nombreuses informations structurales pour concevoir, de manière rationnelle, de nouveaux édulcorants.

Annexe A

Données supplémentaires

TABLE A.1 – Base de données des molécules dont le pouvoir sucrant est connu.

Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)
1p-4p-6p-3Cl-sucrose	2	asp 103	1.4	asp 169	2.08	asp 4	0.89	asp 98	0.92	lugduname	5.35	taxifolin 3-acetate	1.9
1p-4p-2Cl-sucrose	1.48	asp 104	0.48	asp 17	1.19	asp 5	1.19	asp 99	1.16	maltitol	-0.06	telosmoside A	3
1p-6p-2Cl-sucrose	1.89	asp 105	1.21	asp 170	1.93	asp 52	0.27	aspartame	2.25	maltose	-0.41	trans-anethole	1.11
1p-Cl-sucrose	1.3	asp 106	1.58	asp 171	1.41	asp 53	0.25	baiyunoside	2.7	mannitol	-0.3	trans-cinnamaldehyde	1.7
4p-Br-4-1p-6p-3Cl-sucrose	3.51	asp 107	1.6	asp 172	1.97	asp 55	1.15	benzenepropanoic acid	3.6	mannose	-0.28	trehalose	-0.37
4p-F-4-1p-6p-3Cl-sucrose	3	asp 108	1.1	asp 173	1.44	asp 56	1.16	bernardame	5.3	mogroside II	1.92	tryptophan	1.54
4-1p-4p-6p-4Br-sucrose	3.88	asp 109	1.22	asp 174	0.71	asp 57	0.75	cyclamate	1.48	mogroside IV	2.49	xylitol	-0.01
4-1p-4p-6p-4Cl-sucrose	3.32	asp 11	1.54	asp 175	0.66	asp 58	0.13	cyclocarioside A	2.3	mogroside V	2.53	xylobiose	-0.4
4-1p-4p-3Cl-sucrose	2.34	asp 110	1.06	asp 176	1.51	asp 6	1.5	dulcin	2.3	naringin DHC	2	xylose	-0.28
4-1p-6p-3Br-sucrose	2.9	asp 112	0.66	asp 177	0.91	asp 60	1.6	dulcoside A	1.93	neoeriocitrin DHC	2	perrilartine	3.3
4-1p-6p-3Cl-sucrose	2.8	asp 12	1.56	asp 178	1.23	asp 61	2.11	erythritol	-0.2	neohesperidin DHC	3.04	perillaldehyde	1.08
4-1p-6p-3F-sucrose	1.6	asp 122	0.91	asp 179	1.36	asp 62	2.23	fructose	0.16	neotame	3.89	hernandulcin	3.13
4-1p-6p-3I-sucrose	2.08	asp 123	1.53	asp 18	1.58	asp 63	2.95	fucose	-0.16	osladin	2.65	monatine	3.11
4-1p-2Cl-sucrose	2.08	asp 124	1.43	asp 180	1.31	asp 64	2.15	galactitol	-0.24	phyllodulcin 1	2.85	carrelame	5.2
4-4p-6p-3Cl-sucrose	2.2	asp 125	1.65	asp 181	1.86	asp 65	2.15	galactose	-0.19	phyllodulcin 11	2.4	alitame	3.3
4-6-1p-6p-4Cl-sucrose	2.3	asp 126	1.81	asp 182	0.68	asp 66	2	gaudichaudioside A	1.74	phyllodulcin 2	2	P4000	3.6
4-Cl-sucrose	0.7	asp 128	1.16	asp 183	0.98	asp 67	0.28	glucose	-0.22	phyllodulcin 4	2.54		
4-F-1p-4p-6p-3Br-sucrose	3	asp 129	0.84	asp 184	1.02	asp 68	0.61	glycyrrhizin	1.96	phyllodulcin 5	2.18		
4-F-1p-4p-6p-3Cl-sucrose	2.78	asp 13	1.85	asp 185	1.56	asp 69	0.56	hematoxylin	2.08	phyllodulcin 6	2		
4-O-n-propyl glycosides DHC	3.3	asp 130	0.6	asp 186	1.2	asp 7	1.16	hesperinin glucoside DHC	2	phyllodulcin 7	2.48		
4b-10a-dimethyl-1-2-3-4-5-10-hexahydrofluorene-4a-6a-dicarboxylic acid	3.19	asp 131	0.68	asp 187	1.75	asp 70	1.16	homoeohesperidin DHC	3	phyllodulcin 8	2.48		
6p-Cl-sucrose	1.3	asp 132	0.64	asp 188	2.1	asp 71	1.03	isomalt	-0.27	polypodoside A	2.78		
6-1p-6p-3Cl-sucrose	1.4	asp 133	0.88	asp 189	1.95	asp 72	0.99	isomaltulose	-0.32	prunin DHC	1.6		
6-chloro-tryptophan	3	asp 134	0.73	asp 19	2.11	asp 73	1.36	isovanillic 1	2.3	psicose	-0.15		
D-tagatose	-0.04	asp 135	0.36	asp 190	0.9	asp 74	1.61	isovanillic 10	2.3	pterocaryoside A	1.8		
DHQ-3-SIFA-4p-methyl ether	2.6	asp 136	0.6	asp 191	0.68	asp 75	1.86	isovanillic 11	4	pterocaryoside B	1.88		
liquiritin	2.4	asp 137	0.68	asp 192	2.15	asp 76	1.85	isovanillic 12	2.65	raffinose	-0.6		
N-Acetyl-D-glucosamine	-0.26	asp 138	1.16	asp 193	2.15	asp 77	1.58	isovanillic 13	2.7	rebaudioside A	2.47		
Periandrin I	1.93	asp 139	1.22	asp 194	1.43	asp 78	1.98	isovanillic 15	2.3	rebaudioside B	2.51		
Periandrin II	2	asp 14	1.14	asp 197	2.03	asp 79	1.99	isovanillic 16	2	rebaudioside C	1.93		
Periandrin III	2	asp 140	0.18	asp 198	2.79	asp 8	1.57	isovanillic 17	1.7	rebaudioside D	2.51		

Suite page suivante

TABLE A.1 – Suite

Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)	Nom	log(PS)
Periandrin IV	2	asp 141	0.51	asp 199	3.06	asp 80	2.18	isovanillic 18	2.18	rebaudioside E	2.35		
abrusoside B	2	asp 142	0.88	asp 2	0.85	asp 81	0.6	isovanillic 19	2.7	rebaudioside F	2.05		
abrusoside C	1.7	asp 143	0.94	asp 20	2.3	asp 82	2.08	isovanillic 2	2.54	rhamnose	-0.5		
abrusoside D	1.88	asp 15	1.17	asp 200	2.92	asp 83	2.08	isovanillic 20	2.82	ribose	-0.15		
acesulfame potassium CH ₃ CH ₂ CH ₂ CH ₃	1.48	asp 153	0.43	asp 201	3.14	asp 84	2.28	isovanillic 21	1.9	rubusoside	2.06		
acesulfame potassium CH ₃ CH ₂ CH ₃	2.4	asp 154	0.11	asp 204	3.02	asp 85	1.68	isovanillic 22	2.3	saccharin	2.57		
acesulfame potassium CH ₃ CH ₂ H	1.3	asp 155	0.7	asp 207	2.76	asp 86	2.24	isovanillic 23	2.48	selligueain A	1.54		
acesulfame potassium CH ₃ CH ₃	2.11	asp 156	0.6	asp 208	3.16	asp 87	1.53	isovanillic 24	2.08	siamenoside I	2.75		
acesulfame potassium CH ₃ H	1.3	asp 157	1.3	asp 209	4.79	asp 88	2.13	isovanillic 25	2	sorbitol	-0.24		
acesulfame potassium H CH ₃	2.19	asp 159	1.04	asp 21	1.49	asp 89	1.7	isovanillic 26	1.7	sorbose	-0.22		
acesulfame potassium H CH ₃ CH ₂	2.18	asp 16	0.89	asp 211	2.32	asp 9	1.6	isovanillic 3	2.54	stachyose	-0.66		
acesulfame potassium H H	1	asp 160	1.49	asp 212	2.44	asp 90	1.23	isovanillic 4	2.65	steviolbioside	2.05		
albiziasaponin B	2.78	asp 161	1.92	asp 213	2.68	asp 91	1.99	isovanillic 6	2.7	stevioside	2.27		
arabinose	-0.16	asp 163	1.96	asp 214	3.09	asp 92	1.96	isovanillic 7	2.4	sucralose	2.78		
asp 1	0.57	asp 164	1.75	asp 215	4.05	asp 93	1.48	isovanillic 9	3.3	sucrononic acid	5.3		
asp 10	1.43	asp 165	2.21	asp 22	2.11	asp 94	0.58	lactitol	-0.46	sucrose	0		
asp 100	1.48	asp 166	1.72	asp 23	1.9	asp 95	1.71	lactose	-0.6	suosan	2.54		
asp 101	1.38	asp 167	1.26	asp 24	2.07	asp 96	1.3	lactulose	-0.3	superaspartame	4.15		
asp 102	1.38	asp 168	1.93	asp 3	1.56	asp 97	1.21	leucrose	-0.3	tagatose	-0.05		

TABLE A.2 – Données de mutagenèse dirigée

VFD		CRD	TMD		
T1R1	mGluR3	T1R3	T1R3	A ^{7.46} ₈₀₇ V [101]	G ^{5x42} ₇₃₀ I [155]
S148N [156]	S152D [157]	Q504K [158]	Q ^{3.32} ₆₃₆ A [100]	A ^{7.47} ₈₀₈ T [101]	A ^{5x45} ₇₃₃ T [155]
R151H [156]	S152H [157]	A537T [158]	S ^{3.36} ₆₄₀ A [100]	S ^{2.51} ₆₂₀ A [102]	N ^{5x47} ₇₃₅ D [159]
A170E [156]	R64A [160]	R556P [158]	H ^{3.37} ₆₄₁ A [100]	V ^{2.52} ₆₂₁ L [102]	N ^{5x47} ₇₃₅ D [155]
E174V [156]	R68A [160]	S559P [158]	L ^{3.40} ₆₄₄ A [100]	V ^{2.52} ₆₂₁ I [102]	A ^{5x52} ₇₄₀ I [159]
A302D [156]	Y150A [160]	R560K [158]	T ^{3.41} ₆₄₅ A [100]	F ^{2.55} ₆₂₄ L [102]	A ^{5x52} ₇₄₀ I [155]
R307T [156]	S151A [160]	I536F [161]	H ^{None} ₇₂₁ A [100]	Q ^{3.32} ₆₃₆ A [102]	N ^{5x47} ₇₃₅ D [162]
M320T [156]	T174A [160]	A537T [161]	R ^{None} ₇₂₃ A [100]	Q ^{3.33} ₇₃₅ E [102]	N ^{5x47} ₇₃₅ D [162]
K328Q [156]	D194A [160]	F540P [161]	S ^{None} ₇₂₆ A [102]	S ^{3.36} ₈₄₀ A [102]	N ^{5x47} ₇₃₅ D [162]
K377E [156]	Y222A [160]	G542N [161]	S ^{5.39} ₇₂₉ A [100]	H ^{3.37} ₆₄₁ A [102]	N ^{5x47} ₇₃₅ D [162]
K379G [156]	R277A [160]	E545Q [161]	F ^{5.40} ₇₃₀ L [100]	Y ^{4.49} ₆₉₉ L [102]	A ^{4x34} ₆₈₁ F [159]
D435K [156]	D301A [160]		F ^{5.40} ₇₃₀ A [100]	Y ^{4.49} ₆₉₉ F [102]	S ^{4x41} ₆₈₈ L [159]
K460E [156]	K389A [160]		F ^{5.40} ₇₃₀ C [100]	H ^{None} ₇₂₁ A [102]	S ^{4x41} ₆₈₈ L [155]
H71A [112]	Q306A [160]		F ^{5.40} ₇₃₀ Q [100]	R ^{None} ₇₂₃ A [102]	G ^{4x42} ₆₈₉ V [159]
T149A [112]	S278A [160]		F ^{5.40} ₇₃₀ S [100]	T ^{None} ₇₂₄ L [102]	A ^{5x45} ₇₃₃ T [159]
S172A [112]	mGluR4		F ^{5.40} ₇₃₀ T [100]	R ^{None} ₇₂₅ M [102]	N ^{5x47} ₇₃₅ D [162]
D192A [112]	K74Y [163]		F ^{5.40} ₇₃₀ W [100]	S ^{None} ₇₂₆ A [102]	mGluR5
Y220A [112]	R78A [164]		F ^{5.40} ₇₃₀ Y [100]	S ^{5.39} ₇₂₉ A [102]	G ^{2.49} ₆₂₈ F [165]
E301A [112]	S157A [164]		A ^{5.43} ₇₃₃ V [100]	F ^{5.40} ₇₃₀ L [102]	F ^{5.53} ₇₈₈ D [165]
S306A [112]	S159A [164]		H ^{5.44} ₇₃₄ A [100]	A ^{5.43} ₇₃₄ V [102]	A ^{7.41} ₇₃₄ V [166]
S385A [112]	S160A [164]		A ^{5.45} ₇₃₅ I [100]	H ^{5.44} ₇₃₄ A [102]	T ^{2.53} ₆₃₂ A [165]
T1R2	S181A [164]		T ^{5.49} ₇₃₉ M [100]	W ^{6.50} ₇₇₅ A [102]	T ^{2.53} ₆₃₂ A [165]
S144A [80]	T182A [164]		Y ^{6.46} ₇₇₁ A [100]	F ^{6.53} ₇₇₈ A [102]	T ^{2.53} ₆₃₂ P [165]
E302A [80]	K74A [167]		W ^{6.50} ₇₇₅ A [102]	V ^{6.54} ₇₇₉ A [102]	I ^{3.36} ₆₅₁ V [166]
Y103A [114]	K74Y [167]		W ^{6.50} ₇₇₅ A [100]	L ^{6.57} ₇₈₂ A [102]	I ^{3.36} ₆₅₁ F [165]
D142A [114]	K74N [167]		V ^{6.51} ₇₇₆ A [100]	R ^{7.29} ₇₉₀ Q [102]	I ^{3.36} ₆₅₁ F [165]
S144A [114]	K74Q [167]		F ^{6.53} ₇₇₈ A [100]	L ^{7.37} ₇₉₈ I [102]	I ^{3.36} ₆₅₁ F [165]
S165A [114]	R258A [167]		V ^{6.54} ₇₇₉ A [100]	L ^{7.39} ₈₀₀ F [102]	I ^{3.36} ₆₅₁ F [165]
P277A [114]	N286A [167]		L ^{6.57} ₇₈₂ A [100]	C ^{7.40} ₈₀₁ I [102]	P ^{3.40} ₆₅₅ S [166]
D278A [114]	E287A [167]		V ^{None} ₇₈₈ A [100]	C ^{7.40} ₈₀₁ L [102]	P ^{3.40} ₆₅₅ S [165]
E302A [114]	K405A [167]		L ^{None} ₇₈₉ Y [100]	G ^{7.43} ₈₀₄ A [102]	P ^{3.40} ₆₅₅ S [165]
D307A [114]	K317A [167]		R ^{7.29} ₇₉₀ Q [100]	G ^{7.43} ₈₀₄ V [102]	P ^{3.40} ₆₅₅ S [165]
E382A [114]	S313G [168]		R ^{7.29} ₇₉₀ H [100]	mGluR1	P ^{3.40} ₆₅₅ S [165]
R383H [114]	K74Y [168]		R ^{7.29} ₇₉₀ K [100]	F ^{6.53} ₈₀₁ A [169]	S ^{3.43} ₆₅₈ C [166]
S40A [115]	E287G [168]		R ^{7.29} ₇₉₀ Y [100]	Q ^{3.32} ₆₆₀ A [108]	Y ^{3.44} ₆₅₉ A [165]
Y103A [115]	K317R [168]		R ^{7.29} ₇₉₀ A [100]	V ^{3.36} ₆₆₄ I [108]	Y ^{3.44} ₆₅₉ A [165]
D142A [115]	R78A [170]		R ^{7.29} ₇₉₀ E [100]	I ^{4.51} ₇₂₅ F [169]	Y ^{3.44} ₆₅₉ A [165]
D142N [115]	K405A [170]		Q ^{7.33} ₇₉₄ A [100]	N ^{5.47} ₇₆₀ A [169]	Y ^{3.44} ₆₅₉ A [165]
S144A [115]	S157A [170]		L ^{7.37} ₇₉₈ I [100]	N ^{5.47} ₇₆₀ A [169]	N ^{5.47} ₇₄₇ A [165]
S165I [115]	G158A [170]		L ^{7.39} ₈₀₀ V [100]	N ^{5.47} ₇₆₀ A [169]	N ^{5.47} ₇₄₇ S [165]
Y215A [115]	T108A [170]		C ^{7.40} ₈₀₁ I [102]	N ^{5.47} ₇₆₀ A [169]	N ^{5.47} ₇₄₇ A [165]
Y215H [115]	S110A [170]		V ^{7.41} ₈₀₂ A [100]	T ^{6.46} ₇₉₄ A [108]	N ^{5.47} ₇₄₇ S [165]
D278A [115]	mGluR7		I ^{7.44} ₈₀₅ A [100]	W ^{6.50} ₇₉₈ A [169]	N ^{5.47} ₇₄₇ A [165]
D278N [115]	N74K [163]		A ^{7.46} ₈₀₇ V [100]	F ^{6.53} ₈₀₁ A [169]	N ^{5.47} ₇₄₇ S [165]
E302A [115]	GABA-B1		A ^{7.47} ₈₀₈ T [100]	F ^{6.53} ₈₀₁ A [169]	N ^{5.47} ₇₄₇ A [165]
E302Q [115]	Q188A [171]		S ^{2.51} ₆₂₀ A [102]	F ^{6.53} ₈₀₁ A [169]	N ^{5.47} ₇₄₇ S [165]
D307A [115]	E192A [171]		V ^{2.52} ₆₂₁ I [102]	F ^{6.53} ₈₀₁ A [169]	G ^{5.48} ₇₄₈ L [165]
D307N [115]	S246A [171]		F ^{2.55} ₆₂₄ L [102]	F ^{6.53} ₈₀₁ A [172]	G ^{5.48} ₇₄₈ L [165]
V309N [115]	S247A [171]		Q ^{3.32} ₆₃₆ A [102]	Y ^{6.57} ₈₀₅ A [169]	G ^{5.48} ₇₄₈ L [165]
R383A [115]	S249A [171]		Q ^{3.32} ₆₃₆ A [101]	Y ^{6.57} ₈₀₅ A [169]	W ^{6.50} ₇₈₅ A [165]
R383F [115]	T250A [171]		Q ^{3.33} ₆₃₇ E [102]	Y ^{6.57} ₈₀₅ A [169]	W ^{6.50} ₇₈₅ A [165]
R383K [115]	S268A [171]		S ^{3.36} ₆₄₀ A [102]	Y ^{6.57} ₈₀₅ A [169]	W ^{6.50} ₇₈₅ A [165]
mGluR1	S269A [171]		S ^{3.36} ₆₄₀ A [101]	Y ^{6.57} ₈₀₅ A [172]	W ^{6.50} ₇₈₅ A [165]
Y74A [173]	S270A [171]		H ^{3.37} ₆₄₁ A [102]	T ^{7.33} ₈₁₅ M [169]	F ^{6.53} ₇₈₈ A [165]

VFD		CRD	TMD		
R78A [173]	T310A [171]		H ^{3.37} ₆₄₁ A [101]	T ^{7.33} ₈₁₅ M [169]	F ^{6.53} ₇₈₈ W [165]
S164A [173]	Q312A [171]		I ^{3.40} ₆₄₄ A [101]	T ^{7.33} ₈₁₅ M [169]	F ^{6.53} ₇₈₈ A [165]
S165A [173]	Q313A [171]		T ^{3.41} ₆₄₅ A [101]	T ^{7.33} ₈₁₅ M [169]	F ^{6.53} ₇₈₈ W [165]
T188A [173]	T314A [171]		Y ^{4.49} ₆₉₉ L [102]	T ^{7.33} ₈₁₅ M [169]	F ^{6.53} ₇₈₈ W [165]
D208A [173]	T315A [171]		Y ^{4.49} ₆₉₉ F [102]	T ^{7.33} ₈₁₅ M [169]	F ^{6.53} ₇₈₈ A [165]
Y236A [173]	E316A [171]		H ^{None} ₇₂₁ A [102]	T ^{7.33} ₈₁₅ M [172]	F ^{6.53} ₇₈₈ W [165]
E292A [173]	E458A [171]		H ^{None} ₇₂₁ A [101]	A ^{7.36} ₈₁₈ S [108]	Y ^{6.57} ₇₉₂ A [165]
G293A [173]	E459A [171]		R ^{None} ₇₂₃ A [102]	S ^{7.40} ₈₂₂ A [108]	Y ^{6.57} ₇₉₂ F [165]
D318A [173]	T460A [171]		R ^{None} ₇₂₃ A [101]	S ^{3.40} ₆₆₈ P [108]	Y ^{6.57} ₇₉₂ A [165]
R323A [173]	Q464A [171]		T ^{None} ₇₂₄ L [102]	S ^{4.37} ₇₁₁ F [169]	Y ^{6.57} ₇₉₂ F [165]
K409A [173]	Y470A [171]		R ^{None} ₇₂₅ M [102]	S ^{4.41} ₇₁₅ C [169]	Y ^{6.57} ₇₉₂ A [165]
D318I [116]	T290K [174]		S ^{None} ₇₂₆ A [102]	V ^{4.42} ₇₁₆ I [169]	Y ^{6.57} ₇₉₂ F [165]
D322I [116]	GABA-B2		S ^{5.39} ₇₂₉ A [102]	T ^{4.45} ₇₁₉ G [169]	Y ^{6.57} ₇₉₂ A [165]
D324I [116]	Y690V [174]		S ^{5.39} ₇₂₉ A [101]	T ^{4.49} ₇₂₃ A [169]	M ^{7.33} ₈₀₂ T [166]
E325I [116]			F ^{5.40} ₇₃₀ L [102]	I ^{4.51} ₇₂₅ F [169]	S ^{7.36} ₈₀₅ A [166]
E328I [116]			F ^{5.40} ₇₃₀ L [101]	I ^{4.51} ₇₂₅ F [169]	S ^{7.40} ₈₀₉ F [165]
Y74K [163]			A ^{5.43} ₇₃₃ V [102]	I ^{4.51} ₇₂₅ F [169]	S ^{7.40} ₈₀₉ A [165]
mGluR2			A ^{5.43} ₇₃₃ V [101]	I ^{4.51} ₇₂₅ F [169]	S ^{7.40} ₈₀₉ F [165]
R74K [163]			A ^{5.43} ₇₃₃ E [101]	I ^{4.51} ₇₂₅ F [169]	S ^{7.40} ₈₀₉ A [165]
Y144F [157]			A ^{5.43} ₇₃₃ F [101]	T ^{N.e} ₇₄₈ A [108]	S ^{7.40} ₈₀₉ F [165]
A166S [157]			A ^{5.43} ₇₃₃ G [101]	A ^{5.42} ₇₅₅ T [169]	S ^{7.40} ₈₀₉ A [165]
S167A [157]			A ^{5.43} ₇₃₃ S [101]	P ^{5.43} ₇₅₆ S [108]	S ^{7.40} ₈₀₉ F [165]
S169A [157]			A ^{5.43} ₇₃₃ T [101]	I ^{5.44} ₇₅₇ V [175]	A ^{7.41} ₈₁₀ V [165]
S164A [157]			H ^{5.44} ₇₃₄ A [101]	I ^{5.44} ₇₅₇ V [176]	A ^{7.41} ₈₁₀ V [165]
R271A [157]			A ^{5.45} ₇₃₅ I [101]	N ^{5.47} ₇₆₀ A [169]	A ^{7.41} ₈₁₀ V [165]
S145A [157]			T ^{5.49} ₇₃₉ M [101]	N ^{5.47} ₇₆₀ A [169]	A ^{7.41} ₈₁₀ V [165]
R57A [157]			Y ^{6.46} ₇₇₁ A [101]	M ^{5.52} ₇₆₅ L [169]	A ^{7.44} ₈₁₃ F [165]
R57K [157]			W ^{6.50} ₇₇₅ A [102]	Y ^{5.56} ₇₆₉ F [169]	A ^{7.44} ₈₁₃ F [165]
R57Y [157]			W ^{6.50} ₇₇₅ A [101]	W ^{6.50} ₇₉₈ A [169]	A ^{7.44} ₈₁₃ F [165]
Y216A [157]			V ^{6.51} ₇₇₆ A [101]	W ^{6.50} ₇₉₈ A [169]	A ^{7.44} ₈₁₃ F [165]
Y216F [157]			F ^{6.53} ₇₇₈ A [102]	W ^{6.50} ₇₉₈ A [169]	V ^{7.49} ₈₁₉ T [166]
D295A [157]			F ^{6.53} ₇₇₈ A [101]	W ^{6.50} ₇₉₈ A [169]	T ^{2.53} ₆₃₂ A [165]
D295R [157]			V ^{6.54} ₇₇₉ A [102]	W ^{6.50} ₇₉₈ A [169]	T ^{2.53} ₆₃₂ P [165]
S148A [157]			V ^{6.54} ₇₇₉ A [101]	W ^{6.50} ₇₉₈ A [172]	T ^{2.53} ₆₃₂ P [165]
T168A [157]			I ^{6.57} ₇₈₂ A [102]	F ^{6.53} ₈₀₁ A [169]	T ^{2.53} ₆₃₂ A [165]
R183A [157]			I ^{6.57} ₇₈₂ A [101]	Y ^{6.57} ₈₀₅ A [169]	T ^{2.53} ₆₃₂ P [165]
Y144A [157]			V ^{None} ₇₈₈ A [101]	Y ^{6.57} ₈₀₅ A [169]	R ^{3.33} ₆₄₈ E [165]
Y144S [157]			I ^{None} ₇₈₉ Y [101]	T ^{7.33} ₈₁₅ M [108]	R ^{3.33} ₆₄₈ E [165]
Y144G [157]			R ^{7.29} ₇₉₀ Q [102]	T ^{7.33} ₈₁₅ A [108]	R ^{3.33} ₆₄₈ E [165]
D146S [157]			R ^{7.29} ₇₉₀ Q [101]	V ^{7.41} ₈₂₃ A [166]	R ^{3.33} ₆₄₈ E [165]
K377A [160]			Q ^{7.33} ₇₉₄ A [101]	mGluR2	I ^{3.34} ₆₄₉ L [166]
D188A [160]			L ^{7.37} ₇₉₈ I [102]	A ^{4x34} ₆₈₁ F [155]	G ^{3.35} ₆₅₀ L [166]
R271A [160]			L ^{7.37} ₇₉₈ I [101]	G ^{4x42} ₆₈₉ V [155]	Y ^{3.44} ₆₅₉ F [165]
L300A [160]			I ^{7.39} ₈₀₀ F [102]	H ^{None} ₇₂₃ V [159]	Y ^{3.44} ₆₅₉ F [165]
			I ^{7.39} ₈₀₀ V [101]	H ^{None} ₇₂₃ V [155]	Y ^{3.44} ₆₅₉ F [165]
			C ^{7.40} ₈₀₁ I [102]	G ^{5x42} ₇₃₀ I [159]	Y ^{3.44} ₆₅₉ F [165]
			V ^{7.41} ₈₀₂ A [101]		F ^{6.53} ₇₈₈ A [165]
			C ^{7.43} ₈₀₄ A [102]		Y ^{6.57} ₇₉₂ F [165]
			G ^{7.43} ₈₀₄ V [102]		M ^{7.33} ₈₀₂ T [165]
			I ^{7.44} ₈₀₅ A [101]		S ^{7.40} ₈₀₉ A [165]
					V ^{7.52} ₈₂₂ M [166]
					I ^{7.56} ₈₂₆ I [166]
					S ^{7.36} ₈₀₅ A [165]
					G ^{5.48} ₇₄₈ L [165]

(%)	T1R3	T1R1	mGluR1	mGluR5	mGluR6	mGluR4	mGluR7	mGluR8	mGluR2	mGluR3	GABAB2	GABAB1	CasR
T1R2	39	45	28	29	29	30	30	29	27	27	20	22	28
	42	46	31	31	32	32	34	32	31	31	X	X	38
	28	40	24	24	24	22	21	21	26	25	13	14	32
T1R3		40	28	29	29	30	29	29	30	29	19	22	28
		50	34	32	27	34	31	31	35	28	X	X	39
		31	24	23	23	21	19	20	25	24	13	14	30
T1R1			29	30	29	31	29	30	28	29	20	22	30
			30	30	27	28	27	24	28	26	X	X	35
			30	29	23	23	23	22	28	25	13	13	32
mGluR1				73	46	48	46	47	45	46	18	20	27
				78	46	46	46	47	54	49	X	X	35
				80	40	42	43	43	49	46	17	16	32
mGluR5					44	47	45	47	45	46	19	20	26
					42	49	47	46	54	50	X	X	32
					43	45	46	46	51	48	21	16	34
mGluR6						72	70	73	51	51	18	20	28
						68	58	65	39	43	X	X	32
						75	72	76	49	47	19	18	32
mGluR4							72	78	50	52	20	22	28
							74	74	42	45	X	X	36
							78	83	52	47	18	18	32
mGluR7								75	48	50	19	21	29
								78	43	46	X	X	36
								86	51	48	20	18	30
mGluR8									49	51	20	22	29
									45	45	X	X	36
									52	48	19	18	31
mGluR2										70	20	21	27
										66	X	X	34
										73	18	17	35
mGluR3											20	20	29
											X	X	28
											18	17	35
GABAB2												46	20
												X	X
												38	18
GABAB1													19
													X
													17

TABLE A.3 – Pourcentage d'identité de séquences en acide-aminés

Gris foncé : VFD ; Gris clair : CRD, Blanc : TMD

Annexe B

Méthodes et principes

B.1 Les méthodes d'apprentissage

Les méthodes d'apprentissage supervisées sont des techniques automatiques utilisées pour prédire une variable à partir d'un ensemble de données. Le modèle établit une relation mathématique (statistique) qui peut ensuite être utilisée sur des données externes.

B.1.1 Forêts aléatoires

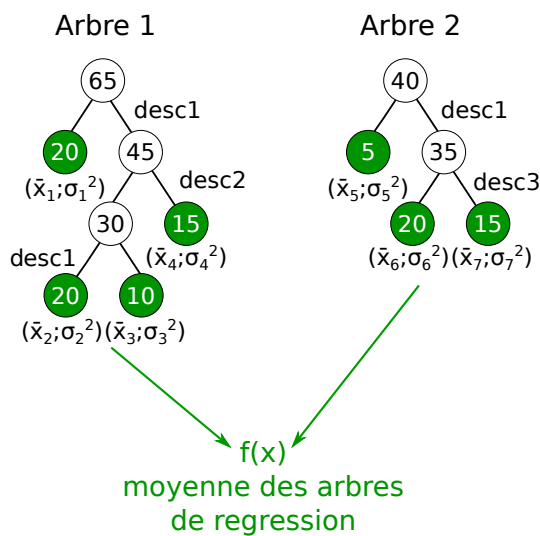
Le principe des Forêts aléatoires (ou *Random Forest*, noté RF) repose sur l'établissement de plusieurs arbres de prédiction, applicable en classification et régression, qui sont moyennés pour établir le modèle final. Pour chaque arbre, un sous-ensemble d'individus est tiré aléatoirement pour établir le modèle de prédiction. Chacun des arbres est alors construit en tirant aléatoirement un sous-ensemble de descripteurs. Par cette construction, les Forêts aléatoires possèdent intrinsèquement un algorithme d'échantillonnage connu sous le nom de *bagging*, abréviation de *bootstrap* et *aggregating*. Chacune de ces valeurs (nombre d'individus et de descripteurs par arbre) est à établir *a priori*. Un nœud de l'arbre caractérise un descripteur, il permet de diviser les individus par dichotomie selon la variable à prédire afin de les classer, selon leur moyenne ($\bar{x}$ sur la figure B.1A), en sous ensemble. Le descripteur choisi pour construire un nœud est tiré aléatoirement et avec remise. Dans le cas d'une variable continue, le critère de

séparation des individus pour un nœud est la minimisation d'une fonction de coût. Ici, elle correspond à la variance associée aux deux sous-ensembles : σ^2 sur la figure B.1A. Le critère d'arrêt, quant-à lui, correspond à la profondeur de l'arbre ou au nombre d'individus présents dans le nœud terminal. Les résultats des estimateurs sont ensuite agrégés et la valeur prédite correspond à une moyenne des estimations dans le cas de la régression. Cette fonction, qui combine les résultats de l'ensemble des arbres, est le modèle de régression non linéaire ($f(x)$, figure B.1A).

B.1.2 Séparateurs à vaste marge

Ces méthodes, aussi appelées machines à vecteur de support (*Support Vector Machine* pour SVM), ont été initialement développées pour des problèmes de classifications. Elles ont ensuite été étendues à la régression. Le principe est de définir un hyperplan permettant de relier chaque individu dans un espace multidimensionnel correspondant au nombre de descripteurs (figure B.1B). C'est l'établissement d'une marge de tolérance ε qui va permettre de définir l'hyperplan le plus pertinent. Dans le cas de la régression, le but est de maximiser les points à l'intérieur des marges. La pénalité associée aux individus à l'intérieur de cette marge est nulle, c'est pourquoi ce type de fonction est aussi appelée fonction de perte - ε insensible. Ainsi, plus la valeur ε est grande plus les erreurs sont tolérées lors de la construction du modèle. Pour les individus en dehors des marges, appelés « variable de ressort », une pénalité est appliquée. La fonction de « coût » de facteur C est linéairement dépendante de la distance ξ entre un point et la marge la plus proche. Le choix du noyau pour établir le modèle mathématique permettant de créer l'hyperplan trouve toute son importance pour prédire la variable d'intérêt.

A) Forêts Aléatoires



B) Machines à vecteurs de support

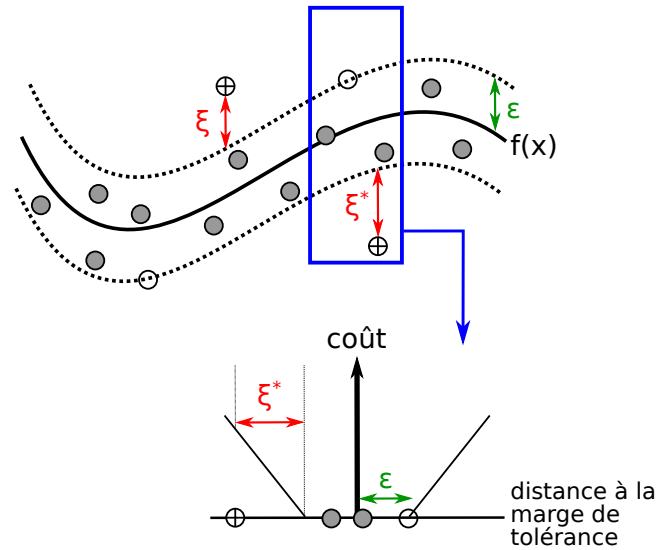


FIGURE B.1 – Représentation schématique des méthodes de prédictions RF et SVM.

A) Représentation schématique du fonctionnement de l'algorithme des RF. Avec cet exemple, les arbres sont construits avec, respectivement, 65 et 40 individus tirés aléatoirement. Chaque nœud correspond à la séparation des individus en sous-ensemble afin de minimiser la variance σ^2 . Les moyennes $\bar{x}$ dans les nœuds terminaux (en vert) sont utilisées pour construire l'estimateur final $f(x)$.

B) Représentation schématique du fonctionnement de l'algorithme des SVM. L'hyperplan correspond à la fonction $f(x)$ en trait noir plein et les marges de tolérance définies par ε sont en pointillées. Les points pleins sont compris dans la marge de tolérance. Les cercles sont en dehors et sont pénalisants pour le modèle. Pour ces individus, la projection de la distance ξ à la marge de tolérance est utilisée pour définir le coût associé. L'ensemble des individus et des paramètres permettent de définir l'équation de l'hyperplan le plus pertinent pour prédire la variable d'intérêt.

B.2 Modélisation par homologie

B.2.1 Alignement de séquences en AA

B.2.1.1 Le cas de deux séquences en AA

La première étape pour réaliser un modèle par homologie est de définir l'alignement de séquences en AA optimal entre la protéine à modéliser et le patron. Pour résoudre le problème de l'alignement de deux séquences en AA, la complexité de l'algorithme requis est de $O(n * m)$, avec n et m la longueur des séquences en AA. Il faut citer les algorithmes de programmation dynamique¹ « Needleman-Wunsch » et « Smith-Waterman » pour réaliser, respectivement, des alignements globaux ou locaux. Le but est de maximiser les AA identiques, ou partageant les mêmes propriétés physico-chimiques, et donc le score global de l'alignement. Ces approches sont basées sur des matrices de substitutions qui contiennent des scores par paire d'AA et qui pénalisent plus ou moins leur mutation. Les matrices les plus utilisées sont les BLOSUM (*BLOcks SUBstitution Matrix*) [177]. Celles-ci sont construites à partir de séquences courtes très conservées (blocs) de protéines homologues et qui ne contiennent pas d'insertions ou de délétions. Le score est alors défini selon l'équation $S_{ij} = \frac{\log(q_{ij})}{p_i \cdot p_j}$, avec q_{ij} la fréquence de substitution des AA i et j dans les blocs et $p_i \cdot p_j$ la probabilité qu'ils soient alignés par hasard. En d'autres termes, il correspond à la probabilité que deux AA i et j soient à la même position dans ces blocs, divisée par la probabilité d'alignement par hasard de ces deux AA dans un alignement de séquences de protéines non homologues. Il est possible de placer des « trous » (ou *gaps*) - associés eux aussi à un score de pénalité - au sein de ces alignements pour pallier aux insertions ou délétions d'AA entre les deux séquences.

1. résolution d'un problème complexe en sous ensemble de problèmes plus simples

B.2.1.2 Alignement de plus de deux séquences en AA

Les alignements de séquences multiples sont bénéfiques pour ajouter de la statistique sur la position des AA conservés. Dans ce cas, le problème posé est d'aligner N séquences (avec $N > 2$) de longueur L . La complexité pour la résolution exacte de ce problème, afin d'obtenir un alignement optimal, est $O(L^N)$. De nombreux algorithmes ont été développés pour réduire ce coût de calcul. Celui appliqué dans cette thèse est *Clustal Omega* [178]. Il utilise un modèle probabiliste basé sur des chaînes de Markov cachées (HMM)² pour résoudre l'alignement. La première étape de cet algorithme est basée sur la construction d'un « arbre guide » établit à partir d'un score de distance par paire de séquences³. Cet arbre est ensuite utilisé pour réaliser un alignement multiple progressif en agglomérant les séquences par similarité décroissante. À chaque itération, une séquence est alignée contre un profil HMM [180] donnant les probabilités d'apparition, d'insertions et de délétions des AA à chaque position, afin d'être ajouté à l'alignement multiple.

Cependant, ces algorithmes d'alignement de séquences en AA n'ont pas la prétention de proposer une solution exacte, car le résultat dépend de nombreux paramètres fixés *a priori* (matrice de score, pénalité, sous-ensemble, ...). Même s'ils sont très utiles pour obtenir une première hypothèse de travail, il est nécessaire d'ajuster manuellement la solution des algorithmes sus-cités.

Correction de l'alignement

Pour corriger l'alignement proposé, il est fréquent d'utiliser des données expérimentales. Ainsi, le recensement des données de mutagenèse dirigée et des données structurales (notamment la structure secondaire) permet de contrôler et d'ajuster les zones qui semblent peu pertinentes.

2. Un modèle de Markov est un processus statistique qui permet de prédire un état futur en fonction de l'état présent uniquement. Le terme « caché » fait référence à l'impossibilité de connaître le processus qui a donné le résultat, car ce n'est pas l'état qui est considéré mais une observation.

3. Le score est calculé à partir de la méthode des *k* - *uplet* [179], elle vise à comparer l'ensemble des mots composés de *k* lettres entre les deux séquences.

B.2.2 Modélisation du récepteur

Une fois que les patrons – dont les données structurales sont disponibles – ont été choisis et que l’alignement est optimal, la méthode de modélisation par homologie peut être appliquée. Le logiciel « *Modeller* » [154] est utilisé pour construire le modèle. L’objectif est d’extraire un ensemble d’informations structurales sur le (ou les) patron(s) qui seront utilisées pour définir les contraintes à appliquer lors de la construction du modèle. Les contraintes extraites sont, par exemple, des distances de liaisons hydrogènes, des angles dièdres sur le squelette et les chaînes latérales, ... La valeur et la forme des contraintes ont été calibrées d’après une analyse statistique sur des alignements par paires de structures homologues. Ces données empiriques sont stockées dans une base de données, sous forme de densité de probabilité. La fonction objective de *Modeller* vise à satisfaire ces contraintes spatiales et à minimiser l’énergie du système. L’optimisation de cette fonction objective est appliquée avec la méthode de minimisation du gradient conjugué et raffinée en dynamique moléculaire par recuit simulé. Des contraintes supplémentaires sont appliquées pour les ponts disulfures observés sur la structure du patron. En faisant varier les coordonnées de la structure initiale, *Modeller* peut proposer plusieurs modèles.

B.2.3 Validation du modèle

Les modèles proposés sont évalués avec le score DOPE (*Discrete Optimized Protein Energy*) [181]. Ce score est empirique, calibré sur un jeu de données de structures cristallographiques, et basé sur une fonction de densité de probabilité dépendante de la distance inter-atomique. L’énergie libre totale du système est estimée à partir des potentiels statistiques et sert de premier filtre pour sélectionner un modèle dont la conformation est la plus favorable. Les modèles de plus basse énergie proposés par *Modeller* sont, ensuite, vérifiés manuellement pour sélectionner ceux respectant les critères géométriques et énergétiques. La structure globale est comparée au patron, ainsi que

l'orientation des chaînes latérales des AA. Les régions non structurées sont analysées, à la fois leur entremêlement, et la présence de liaisons peptidiques *cis* (peu probable).

B.3 Docking moléculaire

Le point de départ des simulations de dynamique moléculaire est obtenu par *docking*. Les interactions ligand-récepteur sont définies par le logiciel Vina [182]. Un espace de recherche est établi sur le récepteur. Vina construit des grilles (un treillis tridimensionnel) dont les points sont espacés de 1 Å. Chaque point de la grille stocke l'énergie potentielle associée à un atome, ou un groupe fonctionnel, du ligand. La fonction de score de Vina, permettant d'estimer quelle pose est la plus favorable, est basée sur des données théoriques et empiriques. Cette fonction approxime une énergie d'interaction selon l'équation :

$$\begin{aligned} \Delta G_{liaison} = & \Delta G_{van\ der\ Waals} + \Delta G_{electrostatique} + \Delta G_{liaison\ hydrogene} \\ & + \Delta G_{desolvation} + \Delta G_{torsion} \end{aligned} \quad (B.1)$$

Les résultats, appelés « poses » de *docking*, sont clusterisés et triés selon l'énergie d'interaction estimée. Le logiciel Vina permet d'estimer l'affinité (score) en kcal.mol⁻¹ de chacune des poses proposées.

B.4 Méthodes et principes de la dynamique moléculaire

B.4.1 Un modèle comme objet d'étude

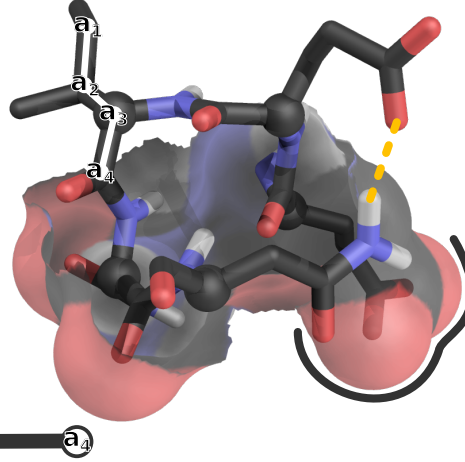
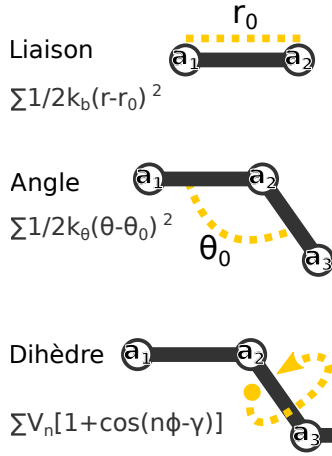
Un modèle consiste à simplifier la réalité tout en s'en rapprochant autant que possible. Un modèle ne vaut que par le degré de concordance entre les valeurs des mesures théoriques et celles effectuées au cours d'expériences réelles. On ne peut pas se satisfaire de bons résultats sur une seule expérience. La comparaison doit être menée

sur un ensemble d'expériences, ou d'exemples, judicieusement choisis pour tenter de représenter l'espace des valeurs possibles.

Les modèles moléculaires peuvent être de deux types : quantique ou classique. Le modèle quantique représente explicitement les électrons. Malgré sa précision, le temps de calcul nécessaire pour résoudre les équations qu'il implique n'est pas adapté aux systèmes de grande taille. Dans le modèle classique, les atomes, regroupant le noyau et les électrons, sont décrits comme des charges ponctuelles auxquelles une masse et un rayon sont associés. Les interactions s'exerçant entre les différents atomes sont modélisées au moyen d'un champ de force dont les paramètres sont issus de données spectroscopiques, de calculs de chimie quantique et optimisés sur des structures connues. Les énergies dérivant de ces forces sont séparées en deux groupes : les termes liés et non liés.

B.4.2 Champ de force AMBER

Les termes d'énergie liés sont des termes représentant les énergies de liaison, d'angle et de torsion. Les termes dits « non liés » regroupent les modèles d'interactions de type van der Waals et électrostatiques. Les énergies de liaison et d'angle sont calculées par des potentiels harmoniques (termes 1 et 2 de l'équation B.2). L'énergie de torsion, associée à la rotation autour d'une liaison, est décrite selon un potentiel périodique (3^{eme} terme de l'équation B.2). Enfin, les énergies associées aux interactions de van der Waals et électrostatiques sont décrites par des fonctions inversement proportionnelles à des puissances de la distance (dernier terme de l'équation B.2). L'énergie potentielle du système (V_{total}) est la somme de ces termes, comme montré dans l'équation B.2 et illustré dans la figure B.2.

Interactions liantes**Interactions non-liantes**

Electrostatique

$$\sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Van der Waals

$$\sum 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

FIGURE B.2 – Représentation des composantes d'un champ de force.

$$\begin{aligned}
 V_{total} = & \sum_{bonds} k_b(r - r_0)^2 \\
 & + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\
 & + \sum_{dihedrals} V_n[1 + \cos(n\phi - \gamma)] \\
 & + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]
 \end{aligned} \tag{B.2}$$

B.4.3 Les conditions aux limites périodiques

La prise en compte des limites d'un système est cruciale pour les simulations de dynamique moléculaire et autorise à calculer des grandeurs macroscopiques à partir d'une seule unité du système. Le recours aux conditions limites périodiques permet de s'affranchir des effets de bord durant une simulation. Cette méthode consiste en la réplique d'une boîte unitaire contenant le système selon les trois dimensions de l'espace. Les coordonnées des particules dans les images de la boîte élémentaire sont calculées en additionnant ou soustrayant des multiples entiers des dimensions de la

boite (figure B.3). Lorsqu'une molécule quitte une boite par un côté, elle est réintroduite par la face opposée. Le nombre de particules dans la boite unitaire reste donc constant au cours du temps.

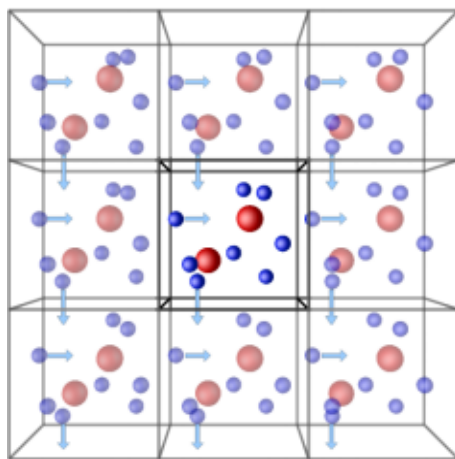


FIGURE B.3 – Représentation tridimensionnelle du principe des conditions aux limites périodiques.

B.4.4 Image minimum et rayon de troncature

L'utilisation des conditions aux limites périodiques nécessite d'introduire quelques subtilités dans le traitement des interactions inter-atomiques. Le calcul de la force d'interaction entre la particule i et j s'effectue entre l'image de j la plus proche de i (qui n'est pas obligatoirement située dans la même boite que i). On définit ainsi l'image minimale du système, comme illustrée sur la figure B.4.

L'utilisation d'un potentiel tronqué limite le calcul des forces interagissant sur la particule i au-delà d'une distance arbitrairement définie : le rayon de troncature. Cette distance ne doit pas dépasser la moitié de la distance du plus petit côté de la boite pour éviter les interactions entre une particule et son image dans une cellule adjacente.

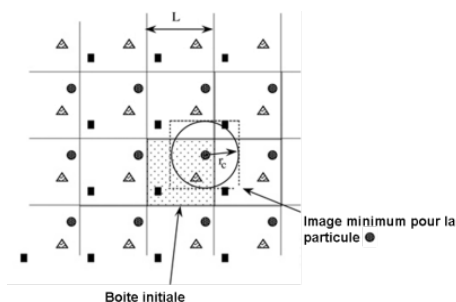


FIGURE B.4 – Schéma de l'image minimum et de la sphère de troncature dans l'utilisation des conditions aux limites périodiques.

B.4.5 Minimisation de l'énergie potentielle

La minimisation constitue une étape importante dans les protocoles de modélisation moléculaire. Elle permet une optimisation de géométrie complète pour corriger les défauts des structures initiales et pour adapter le système au champ de force utilisé. Cette étape nécessite la plupart du temps de combiner plusieurs méthodes de minimisation pour converger vers un minimum.

Dans le cas de la modélisation moléculaire, la fonction à minimiser correspond à l'hamiltonien, c'est-à-dire, la fonction qui permet de retrouver l'énergie de la molécule en fonction de sa géométrie. Pour une molécule de N atomes, la fonction est représentée dans $3N - 6$ dimensions. Soit $3N$ qui correspondent à N pour l'axe x , N pour l'axe y et N pour l'axe z . Les 6 dimensions en moins correspondent à la rotation (3 dimensions) et à la translation (3 dimensions) de cette molécule.

Comme son nom l'indique, la minimisation consiste à déplacer les atomes d'un point donné de l'hypersurface énergies-coordonnées. La molécule finira alors par se trouver dans un minimum local d'énergie suite à des incrémentations itératives. Cette méthode permet d'obtenir une conformation plus stable (qui possède une énergie plus basse que le point de départ) mais ne peut pas garantir d'obtenir le minimum global. La difficulté est de trouver la bonne direction à emprunter sur la surface d'énergie, afin de produire la géométrie ayant l'énergie la plus basse possible. Cette partie est basée

sur des approches mathématiques qui décrivent la façon d'évoluer sur la surface associée à une fonction. Plusieurs algorithmes de minimisation existent et leur combinaison permet d'éviter de rester dans un minimum local trop proche du point initial. Ces algorithmes sont basés sur des méthodes dérivatives, il est alors nécessaire de développer la fonction énergie potentielle selon un développement de Taylor.

B.4.5.1 Algorithme d'ordre 1

Les méthodes de minimisation d'ordre 1 utilisent la dérivée première (gradient) de l'énergie potentielle. Il en existe deux principales : le gradient simple et le gradient conjugué.

Gradient simple (*steepest descent*) Cette approche consiste à chercher la direction de plus grande pente selon laquelle l'énergie (fonction $f(\vec{r}_n)$) décroît le plus fortement. La direction de recherche (D_n) est donnée par l'opposée du signe du gradient de l'énergie (équation B.3a). Les coordonnées à l'itération $n + 1$ sont déterminées à partir des positions précédentes $\vec{r}_n$, de la direction opposée au gradient (équation B.3a, la plus grande pente) et un pas k . Le pas k est déterminé à chaque itération n dans le but de minimiser la fonction f dans le sens du gradient. Si l'énergie est plus basse que lors de la précédente itération, le pas est augmenté selon $k_{n+1} = 1,1 \cdot k_n$. Dans le cas contraire le pas est alors diminué selon $k_{n+1} = 0,5 \cdot k_n$. Cette méthode est efficace pour éliminer les mauvais contacts entre atomes. Elle est appliquée en première approche, car elle est peu coûteuse en temps et en mémoire. Néanmoins, elle connaît des problèmes de convergence lors de l'approche de minimum où des oscillations peuvent apparaître.

$$D_n = -\nabla f(\vec{r}_n) \quad (\text{B.3a})$$

$$\vec{r}_{n+1} = \vec{r}_n + k_n D_n \quad (\text{B.3b})$$

Gradient conjugué Cette approche peut être vue comme une extension de la méthode de « la plus grande pente » où le pas est déterminé de manière optimale, permettant de converger plus efficacement vers le minimum local. Cette méthode prend en compte les étapes précédentes D_{n-1} afin de définir plus précisément la direction de recherche D_n . En d'autres termes, elle permet de déterminer la direction de descente conjuguée à la direction de descente précédente (équation B.4). L'intérêt de cet algorithme est d'éviter une oscillation autour du minimum et d'accélérer la convergence puisque la direction est ajustée à chaque cycle afin d'obtenir la meilleure diminution d'énergie. Cet algorithme présente deux principaux inconvénients. Il est peu efficace pour éliminer les défauts structuraux comme des contacts entre atomes. D'autre part, la convergence ne pourra être atteinte si la structure initiale est trop distante du minimum d'énergie.

$$\vec{D}_n = -\nabla f(\vec{r}_n) + \frac{|\nabla f(\vec{r}_n)|^2}{|\nabla f(\vec{r}_{n-1})|^2} \vec{D}_{n-1} \quad (\text{B.4})$$

B.4.5.2 Algorithme d'ordre 2

Ces algorithmes, tel que la méthode Newton-Raphson, font intervenir la dérivée seconde de la fonction d'énergie par rapport aux coordonnées atomiques, ou Hessien. La méthode de Newton-Raphson est basée sur une approche du développement de Taylor au second ordre (équation B.5a). La série de Taylor appliquée au gradient donne alors l'équation B.5b. On recherche un minimum énergétique, c'est-à-dire un puits où la dérivée de ∇f s'annule. De manière itérative, il est alors facile de déterminer les nouvelles coordonnées à partir l'équation B.5c.

Comme la surface d'énergie potentielle n'est pas quadratique, plusieurs étapes sont nécessaires pour trouver un minimum. Cette méthode nécessite peu d'étapes pour converger à la condition que la structure initiale ne soit pas trop éloignée du minimum. Toutefois, le calcul de l'inverse du Hessien est fastidieux et nécessite des ressources en

mémoire plus importante que pour les méthodes d'ordre 1.

$$f(\vec{r}_{n+1}) \approx f(\vec{r}_n) + (\vec{r}_{n+1} - \vec{r}_n) \nabla f(r_n) + \frac{1}{2} (\vec{r}_{n+1} - \vec{r}_n)^2 \nabla^2 f(r_n) \quad (\text{B.5a})$$

$$\nabla f(\vec{r}_{n+1}) = \nabla f(r_n) + (\vec{r}_{n+1} - \vec{r}_n) \nabla^2 f(r_n) \quad (\text{B.5b})$$

$$\vec{r}_{n+1} = \vec{r}_n - \frac{\nabla f(\vec{r}_n)}{\nabla^2 f(\vec{r}_n)} \quad (\text{B.5c})$$

B.4.6 Principe de la dynamique moléculaire

Les simulations de Dynamique Moléculaire (DM) constituent un des outils les plus utilisés dans les études théoriques de biologie moléculaire. Cette méthode de calcul permet de suivre l'évolution temporelle du comportement d'un système moléculaire. Elle a démontré son apport dans l'étude des fluctuations et des changements de conformation de protéines ou d'acides nucléiques. Elle peut être utilisée pour étudier la structure ou les interactions moléculaires, mais aussi pour obtenir des grandeurs thermodynamiques de systèmes complexes par le biais d'outils de thermodynamique statistique.

Le défi est alors de prédire des grandeurs expérimentales, obtenues à partir d'une moyenne d'états, aux grandeurs obtenues par les simulations de dynamique moléculaire qui résultent d'une moyenne temporelle. La mécanique statistique se base sur l'hypothèse ergodique pour relier des données microscopiques à des grandeurs macroscopiques. Elle définit qu'un système, au bout d'un temps infini, parcourt la totalité de l'hypersurface d'énergie, ainsi la moyenne temporelle sur un système approche la moyenne sur tous les états du système. Donc, si l'échantillonnage est suffisamment important au cours de la simulation, on pourra alors déduire des grandeurs thermodynamiques comparables à des observations macroscopiques.

Une représentation schématique de l'algorithme de dynamique moléculaire est faite dans la figure [B.5](#).

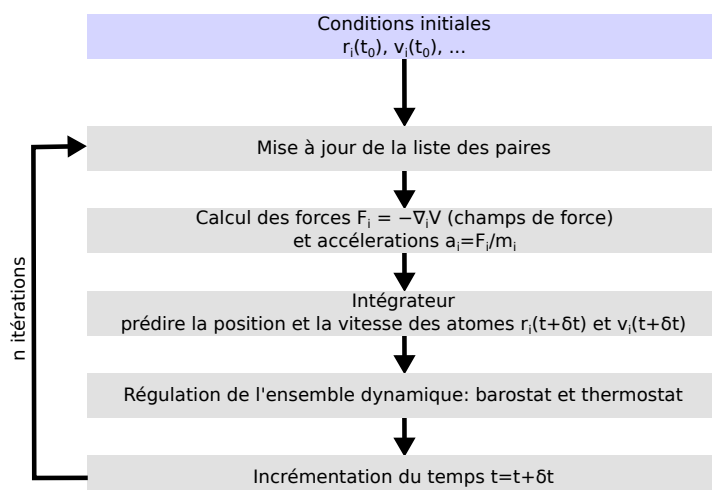


FIGURE B.5 – Algorithme de dynamique moléculaire

B.4.6.1 Données initiales pour les simulations de dynamique moléculaire

La position initiale des atomes lors d'une simulation de dynamique moléculaire est primordiale. Pour les protéines, les positions initiales des atomes sont idéalement issues des structures expérimentales déposées sur la *Protein Data Bank* (PDB). En l'absence de structure expérimentale, des approches de reconstruction par homologie de séquences peuvent être utilisées.

B.4.6.2 Mécanique classique : calcul des forces

À température ambiante (300 K), les structures biologiques sont très mobiles, il est alors nécessaire de simuler le mouvement des atomes. Ces simulations sont basées sur la deuxième loi de Newton (équation B.6a), où F_i est la force appliquée à la particule i , de m_i sa masse et avec $\ddot{r}_i$ son accélération. En connaissant les forces agissant sur l'ensemble des particules du système on peut alors calculer leur accélération. L'intégration des équations du mouvement mène à une trajectoire qui décrit les positions, les vitesses et les accélérations de chacune des particules au cours du temps. Cette méthode déterministe permet, à partir de positions et de vitesses connues à un instant donné, d'identifier l'état du système dans le futur. Les forces appliquées sur les particules sont

dérivées du potentiel moléculaire précédemment décrit, équation B.6b avec V issue de l'équation B.2 [183]. En combinant cette équation B.6a aux équations du mouvement de Newton on obtient alors la relation B.6c. Ainsi les équations de Newton relient la dérivée de l'énergie potentielle aux positions des atomes au cours du temps. Il existe alors plusieurs méthodes numériques pour résoudre cette équation et produire une trajectoire de simulation de dynamique moléculaire.

$$F_i = m_i \ddot{r}_i \quad (\text{B.6a})$$

$$F_i = -\nabla_i V = -\left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i} \right) \quad (\text{B.6b})$$

$$-\frac{dV}{dr_i} = m_i \ddot{r}_i \quad (\text{B.6c})$$

B.4.6.3 Intégration numérique de l'équation

Le principe de base de l'intégration des équations du mouvement consiste à développer en série de Taylor les expressions des positions ($\vec{r}$), vitesses ($\vec{v}$) et accélérations ($\vec{a}$), respectivement dans les équations B.7a, B.7b et B.7c.

$$\vec{r}(t + \delta t) = \vec{r}(t) + \vec{v}(t)\delta t + \vec{a}(t)\frac{\delta t^2}{2} + \dots \quad (\text{B.7a})$$

$$\vec{v}(t + \delta t) = \vec{v}(t) + \vec{a}(t)\delta t + \vec{b}(t)\frac{\delta t^2}{2} + \dots \quad (\text{B.7b})$$

$$\vec{a}(t + \delta t) = \vec{a}(t) + \vec{b}(t)\delta t + \vec{c}(t)\frac{\delta t^2}{2} + \dots \quad (\text{B.7c})$$

Les positions successives $\vec{r}(t + \delta t)$ sont obtenues en résolvant l'équation différentielle du second ordre grâce à l'algorithme Verlet [184], par exemple. Cet algorithme prend en compte les positions et accélérations aux temps $t - \delta t$ et t pour calculer les positions au temps $t + \delta t$ (équations B.8a et B.8b)

$$\vec{r}(t + \delta t) = \vec{r}(t) + \frac{d\vec{r}(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\vec{r}(t)}{dt^2}\delta t^2 + \frac{1}{6}\frac{d^3\vec{r}(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4) \quad (\text{B.8a})$$

$$\vec{r}(t - \delta t) = \vec{r}(t) - \frac{d\vec{r}(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\vec{r}(t)}{dt^2}\delta t^2 - \frac{1}{6}\frac{d^3\vec{r}(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4) \quad (\text{B.8b})$$

En additionnant les deux équations, on obtient alors la position au temps $t + \delta t$ (équation B.9a) et les vitesses au temps t (équation B.9b). L'algorithme de Verlet n'utilise pas les vitesses pour calculer les nouvelles positions (équation B.9a). Son utilisation ne nécessite pas d'importantes ressources de mémoires et permet un temps de calcul rapide. Cependant, sa précision reste relativement faible. Pour des méthodes plus précises, on peut citer l'algorithme *leap-frog* qui utilise les positions au temps t et les vitesses au temps $t + \frac{\delta t}{2}$, ainsi que l'algorithme Verlet vitesse qui permet de connaître simultanément les positions, les vitesses et les accélérations au temps t .

La détermination des coordonnées n'est possible qu'en connaissant la position précédente des atomes ($\vec{r}(t)$). L'attribution des vitesses initiales des atomes se fait selon une distribution de Maxwell-Boltzmann qui dépend du choix de la température.

$$\vec{r}(t + \delta t) = 2\vec{r}(t) - \vec{r}(t - \delta t) + \vec{a}(t)\delta t^2 + \mathcal{O}(\delta t^4) \quad (\text{B.9a})$$

$$\vec{v}(t) = \frac{\vec{r}(t + \delta t) - \vec{r}(t - \delta t)}{2\delta t} \quad (\text{B.9b})$$

Il faut que les algorithmes soient en mesure de décrire les mouvements moléculaires les plus rapides. Le choix de la valeur du pas d'intégration δt ne doit pas être supérieure à la fréquence de Nyquist du système. Cette fréquence correspond à la fréquence de vibration la plus élevée des liaisons interatomiques des systèmes biologiques (protéines), soit ≈ 1 fs qui est la fréquence de vibration de l'élongation d'une liaison X-H (avec X étant N, C, O, ...).

L'utilisation de l'algorithme SHAKE [132] permet d'augmenter le pas d'intégration et, par conséquent, diminuer le temps de calcul pour une simulation. Le principe est d'imposer une restriction sur les distances C-H, qui possèdent la fréquence de vibration la plus élevée. Cet algorithme définit une contrainte pour rétablir la longueur des liaisons C-H entre le pas t et $t + \delta t$. Ainsi, il est possible de fixer un pas d'intégration à 2 fs.

B.4.7 Ensemble thermodynamique

La résolution des équations du mouvement de Newton permet d'explorer l'ensemble dans lequel le nombre d'atomes N , le volume V et l'énergie E sont constants. Cet ensemble est appelé l'ensemble microcanonique et correspond à un système isolé. D'autres ensembles sont toutefois envisageables.

B.4.7.1 NVT ou canonique

Les simulations réalisées dans cet ensemble permettent de contrôler la température par couplage du système à un thermostat. Ce couplage entraîne l'apparition d'un nouveau terme dans les équations du mouvement. Plusieurs protocoles existent et ont pour objectif de limiter les écarts induits par ce terme correctif. En effet les trajectoires Newtoniennes impliquent la conservation de l'énergie et de la quantité de mouvement.

Le thermostat le plus utilisé pour les systèmes biologiques est le thermostat de Langevin qui applique une force aléatoire sur les particules à intervalles réguliers. Les vitesses des atomes sont alors contrôlées par l'ajout d'un terme de friction dans les équations du mouvement (équation B.10). Dans ce cas, Γ est le terme de friction et W est la force appliquée sur les atomes qui n'est ni corrélée dans le temps et ni aux autres atomes. Le principal avantage lors de l'utilisation de ce thermostat provient de la stabilité de la température, ce qui autorise un pas de temps supérieur à celui utilisé lors de simulation NVE.

$$F_i = m_i \ddot{r}_i = -\frac{dV}{dr_i} - m_i \Gamma_i \dot{r}_i + W_i(t) \quad (\text{B.10})$$

B.4.7.2 NPT ou isotherme-isobare

En réalisant les simulations dans l'ensemble NPT il est possible de comparer les grandeurs thermodynamiques calculées avec des données expérimentales réalisées à pression constante. C'est l'ensemble thermodynamique se rapprochant le plus des conditions de laboratoire en biologie. Le maintien de la température et de la pression

peut être assuré en couplant le système à un thermostat de Langevin et un barostat de Berendsen.

B.5 Estimation de l'énergie d'interaction

Les simulations des dynamiques moléculaires peuvent être utilisées pour estimer l'énergie libre de Gibbs (ou enthalpie libre) entre un état lié ($[RL]$) et un état non-lié $[R] + [L]$. Cette énergie est notée $\Delta G_{liaison}$.



L'aspect dynamique permet d'échantillonner temporellement les différentes interactions qui prennent place entre le ligand et le récepteur. La méthode MMGBSA (*Molecular Mechanics energies combined with the Generalized Born and Surface Area continuum solvation*) estime l'enthalpie libre d'interaction entre un ligand et une protéine, tel que :

$$\begin{aligned} \Delta G_{liaison} &= \langle G_{complexe} \rangle - \langle G_{proteine} \rangle - \langle G_{ligand} \rangle \\ G &= E_{MM} + H_{trans/rot} + G_{sol} - TS \end{aligned} \quad (B.11)$$

$\langle G_x \rangle$ correspond à la moyenne de l'énergie libre totale sur les fenêtres de la simulation de dynamique moléculaire. E_{mm} est l'énergie issue de la mécanique moléculaire, G_{sol} l'énergie de solvation et $H_{trans/rot}$ correspond à l'énergie des six degrés de liberté de rotation et translation ($H_{trans/rot} = 3kT \approx 1,8 kcal.mol^{-1}$ à 300 K). L'énergie de solvation est la somme des contributions électrostatiques (GB⁴) et hydrophobiques (SASA⁵). ΔS définit la contribution entropique. L'estimation de l'entropie, qui se fait par une analyse des modes normaux, a tendance à apporter une grande marge d'erreur dans le résultat final. Il peut être négligé pour des systèmes ayant des entropies similaires, notamment dans la comparaison de ligands ayant le même site d'interaction.

4. Modélisation implicite du solvant aqueux par le modèle de Born généralisé

5. Estimation de la surface accessible au solvant

Bibliographie

1. Faurion, A. Naissance et obsolescence du concept de quatre qualités en gustation. *Journal d'agriculture traditionnelle et de botanique appliquée* **35**, 21–40 (1988).
2. Jaubert, J.-N. Apports de la chimie dans les saveurs. *L'actualité Chimique* **11**, 13–14 (2000).
3. This, H. Goût, saveur, odeur, arôme ? *L'actualité Chimique* **332**, 9–11 (2009).
4. Bensafi, M. & Rouhy, C. Le goût, l'odeur et le cerveau. *Biofutur* **30**, 29–33 (2011).
5. Bushdid, C., March, C. A. D. & Antonczak, S. Ces molécules qui nous mènent par le bout du nez. *L'actualité Chimique*, 21–30 (2016).
6. Smith, B. Perspective : Complexities of flavour. *Nature* **486**, S6–S6 (2012).
7. O'MAHONY, M. Purity Effects and Distilled Water Taste. *Nature* **240**, 489–489 (1972).
8. Gilbertson, T. A., Baquero, A. F. & Spray-Watson, K. J. Water taste : the importance of osmotic sensing in the oral cavity. *Journal of water and health* **4 Suppl 1**, 35–40 (2006).
9. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors : A molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
10. Running, C. A., Craig, B. A. & Mattes, R. D. Oleogustus : The Unique Taste of Fat. *Chemical Senses* **40**, 507–516 (2015).
11. Ohsu, T. *et al.* Involvement of the Calcium-sensing Receptor in Human Taste Perception. *Journal of Biological Chemistry* **285**, 1016–1022 (2010).
12. Lindemann, B. Receptors and transduction in taste. *Nature* **413**, 219–25 (2001).
13. This, H. in *Columbia University Press* 377 (Columbia University Press, 2006).
14. Vega, C., Ubbink, J. & van der Linden, E. *The Kitchen as Laboratory : Reflections on the Science of Food and cooking* 279 (Columbia University Press, 2012).
15. Trivedi, B. P. Gustatory system : The finer points of taste. *Nature* **486**, S2–S3 (2012).
16. Faurion, A. Physiologie de la gustation. *Encyclopédie Médico-Chirurgicale*, 50–54 (2000).
17. Trivedi, B. P. Gustatory system : The finer points of taste. *Nature* **486**, S2–S3 (2012).
18. Chandrashekar, J. *et al.* T2Rs Function as Bitter Taste Receptors. *Cell* **100**, 703–711 (2000).
19. Nelson, G. *et al.* Mammalian Sweet Taste Receptors. *Cell* **106**, 381–390 (2001).

20. Nelson, G. *et al.* An amino-acid taste receptor. *Nature* **416**, 199–202 (2002).
21. Chandrashekar, J. *et al.* The cells and peripheral representation of sodium taste in mice. *Nature* **464**, 297–301 (2010).
22. Huang, A. L. *et al.* The cells and logic for mammalian sour taste detection. *Nature* **442**, 934–938 (2006).
23. Gilbertson, T. a., Damak, S. & Margolskee, R. F. The molecular physiology of taste transduction. *Current Opinion in Neurobiology* **10**, 519–527 (2000).
24. McLaughlin, S. K. & Margolskee, R. F. Gustducin is a taste-cell-specific G protein closely related to the transducins. *Nature* **357**, 563–569 (1992).
25. Taruno, A. *et al.* CALHM1 ion channel mediates purinergic neurotransmission of sweet, bitter and umami tastes. *Nature* **495**, 223–226 (2013).
26. Liman, E. R., Zhang, Y. V. & Montell, C. Peripheral Coding of Taste. English. *Neuron* **81**, 984–1000 (2014).
27. Bachmanov, A. A. & Beauchamp, G. K. Taste Receptor Genes. *Annual Review of Nutrition* **27**, 389–414 (2007).
28. Christensen, C. M., Brand, J. G. & Malamud, D. Salivary changes in solution pH : A source of individual differences in sour taste perception. *Physiology & Behavior* **40**, 221–227 (1987).
29. Murphy, C., Cardello, A. V. & Brand, J. G. Tastes of fifteen halide salts following water and NaCl : Anion and cation effects. *Physiology & Behavior* **26**, 1083–1095 (1981).
30. SCHIFFMAN, S. S. in *Biological and Behavioral Aspects of Salt Intake* 99–111 (Elsevier, 1980).
31. Tordoff, M. G. Some Basic Psychophysics of Calcium Salt Solutions. *Chemical Senses* **21**, 417–424 (1996).
32. Oka, Y., Butnaru, M., von Buchholtz, L., Ryba, N. J. P. & Zuker, C. S. High salt recruits aversive taste pathways. *Nature* **494**, 472–475 (2013).
33. Morris, M. J., Na, E. S. & Johnson, A. K. Salt craving : The psychobiology of pathogenic sodium intake. *Physiology & Behavior* **94**, 709–721 (2008).
34. He, F. J. & MacGregor, G. A. Reducing Population Salt Intake Worldwide : From Evidence to Implementation. *Progress in Cardiovascular Diseases* **52**, 363–382 (2010).
35. Ye, W. *et al.* The K⁺ channel K_{IR} 2.1 functions in tandem with proton influx to mediate sour taste transduction. *Proceedings of the National Academy of Sciences* **113**, E229–E238 (2016).
36. Liem, D. G. Sour Taste Preferences of Children Relate to Preference for Novel and Intense Stimuli. *Chemical Senses* **29**, 713–720 (2004).
37. Wiener, A., Shudler, M., Levit, A. & Niv, M. Y. BitterDB : a database of bitter compounds. en. *Nucleic Acids Research* **40**, D413–D419 (2012).
38. Drewnowski, A & Gomez-Carneros, C. Bitter taste, phytonutrients, and the consumer : a review. *The American journal of clinical nutrition* **72**, 1424–35 (2000).

39. Glendinning, J. I. Is the bitter rejection response always adaptive ? *Physiology & Behavior* **56**, 1217–1227 (1994).
40. Beyreuther, K *et al.* Consensus meeting : monosodium glutamate – an update. *European Journal of Clinical Nutrition* **61**, 928–928 (2007).
41. Williams, A. N. & Woessner, K. M. Monosodium glutamate ‘allergy’ : menace or myth ? *Clinical & Experimental Allergy* **39**, 640–646 (2009).
42. Kwok, R. H. M. Chinese-Restaurant Syndrome. *New England Journal of Medicine* **278**, 796–796 (1968).
43. Jinap, S & Hajeb, P. Glutamate. Its applications in food and contribution to health. *Appetite* **55**, 1–10 (2010).
44. Zhou, Y., Yang, M. & Dong, B. R. in *Cochrane Database of Systematic Reviews* (éd. Zhou, Y.) CD004357 (John Wiley & Sons, Ltd, Chichester, UK, 2012).
45. Ahmed, J. *et al.* SuperSweet—a resource on natural and artificial sweetening agents. *Nucleic Acids Research* **39**, D377–D382 (2011).
46. Ventura, A. K. & Mennella, J. A. Innate and learned preferences for sweet taste during childhood. *Current Opinion in Clinical Nutrition and Metabolic Care* **14**, 379–384 (2011).
47. Reed, D. R. & Knaapila, A. in *Progress in molecular biology and translational science* 213–240 (2010).
48. Stevens, B., Yamada, J., Ohlsson, A., Haliburton, S. & Shorkey, A. Sucrose for analgesia in newborn infants undergoing painful procedures. *The Cochrane database of systematic reviews* **7** (éd. Yamada, J.) CD001069 (2016).
49. Lenoir, M., Serre, F., Cantin, L. & Ahmed, S. H. Intense Sweetness Surpasses Cocaine Reward. *PLoS ONE* **2** (éd. Baune, B.) e698 (2007).
50. Madsen, H. B. & Ahmed, S. H. Drug versus sweet reward : greater attraction to and preference for sweet versus drug cues. *Addiction Biology* **20**, 433–444 (2015).
51. Avena, N. M., Rada, P. & Hoebel, B. G. Evidence for sugar addiction : Behavioral and neurochemical effects of intermittent, excessive sugar intake. *Neuroscience & Biobehavioral Reviews* **32**, 20–39 (2008).
52. Lustig, R. H., Schmidt, L. a. & Brindis, C. D. Public health : The toxic truth about sugar. *Nature* **482**, 27–29 (2012).
53. Murray, S., Tulloch, A., Criscitelli, K. & Avena, N. M. Recent studies of the effects of sugars on brain systems involved in energy balance and reward : Relevance to low calorie sweeteners. *Physiology & Behavior* **164**, 504–508 (2016).
54. Zhong, M., Chong, Y., Nie, X., Yan, A. & Yuan, Q. Prediction of Sweetness by Multilinear Regression Analysis and Support Vector Machine. *Journal of Food Science* **78**, S1445–S1450 (2013).
55. DuBois, G. E. & Prakash, I. Non-Caloric Sweeteners, Sweetness Modulators, and Sweetener Enhancers. *Annual Review of Food Science and Technology* **3**, 353–380 (2012).

56. Nofre, C. and Tinti, J.M. and Ouar, F. *Sweeteners derived from glycine and beta-alanine, process for sweetening various products and compositions containing such sweeteners* 1989.
57. DuBois, G. E. in *ACS Symposium Series* 444–462 (2008).
58. Mazur, R. H., Schlatter, J. M. & Goldkamp, A. H. Structure-taste relationships of some dipeptides. *Journal of the American Chemical Society* **91**, 2684–2691 (1969).
59. Magnuson, B. A., Carakostas, M. C., Moore, N. H., Poulos, S. P. & Renwick, A. G. Biological fate of low-calorie sweeteners. *Nutrition Reviews* **74**, 670–689 (2016).
60. Kirkland, D. & Gatehouse, D. “Aspartame : A review of genotoxicity data”. *Food and Chemical Toxicology* **84**, 161–168 (2015).
61. Services, H. *Cyclamate, Commissioner’s Decision*, 45 FR 61474, September 16, 19890 rapp. tech. 181 (1980), 61474–61530.
62. Kinghorn, a. D. & Soejarto, D. D. Intensely sweet compounds of natural origin. *Medicinal Research Reviews* **9**, 91–115 (1989).
63. Gibbs, B. F., Alli, I. & Mulligan, C. Sweet and taste-modifying proteins : A review. *Nutrition Research* **16**, 1619–1630 (1996).
64. Temussi, P. a. Natural sweet macromolecules : how sweet proteins work. *Cellular and Molecular Life Sciences* **63**, 1876–1888 (2006).
65. Behrens, M., Meyerhof, W., Hellfritsch, C. & Hofmann, T. Sweet and Umami Taste : Natural Products, Their Chemosensory Targets, and Beyond. *Angewandte Chemie International Edition* **50**, 2220–2242 (2011).
66. Koizumi, A. *et al.* Human sweet taste receptor mediates acid-induced sweetness of miraculin. *Proceedings of the National Academy of Sciences* **108**, 16819–16824 (2011).
67. DuBois, G. E. Validity of early indirect models of taste active sites and advances in new taste technologies enabled by improved models. *Flavour and Fragrance Journal* **26**, 239–253 (2011).
68. Acton, E. & Stone, H. Potential new artificial sweetener from study of structure-taste relationships. *Science* **193**, 584–586 (1976).
69. Dubois, G. E. Replicating the taste of sugar. *Food Technology* **65**, 30–39 (2011).
70. Morini, G., Bassoli, A. & Borgonovo, G. Molecular modelling and models in the study of sweet and umami taste receptors. A review. *Flavour and Fragrance Journal* **26**, 254–259 (2011).
71. Bassoli, A., Borgonovo, G. & Morini, G. Lost and found in sweeteners : forgotten molecules and unsolved problems in the chemistry of sweet compounds. *Flavour and Fragrance Journal* **26**, 269–273 (2011).
72. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
73. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-Function of the G Protein–Coupled Receptor Superfamily. *Annual Review of Pharmacology and Toxicology* **53**, 531–556 (2013).

74. Attwood, T. K. & Findlay, J. B. Fingerprinting G-protein-coupled receptors. *Protein engineering* **7**, 195–203 (1994).
75. Fredriksson, R. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology* **63**, 1256–1272 (2003).
76. Schiöth, H. B. & Fredriksson, R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. *General and Comparative Endocrinology* **142**, 94–101 (2005).
77. Zhao, G. Q. *et al.* The Receptors for Mammalian Sweet and Umami Taste. *Cell* **115**, 255–266 (2003).
78. Wellendorph, P., Johansen, L. D. & Brauner-Osborne, H. Molecular Pharmacology of Promiscuous Seven Transmembrane Receptors Sensing Organic Nutrients. *Molecular Pharmacology* **76**, 453–465 (2009).
79. Wellendorph, P. & Bräuner-Osborne, H. Molecular basis for amino acid sensing by family C G-protein-coupled receptors. *British Journal of Pharmacology* **156**, 869–884 (2009).
80. Xu, H. *et al.* Different functional roles of T1R subunits in the heteromeric taste receptors. *Proceedings of the National Academy of Sciences* **101**, 14258–14263 (2004).
81. DuBois, G. E. Molecular mechanism of sweetness sensation. *Physiology & Behavior* **164**, 453–463 (2016).
82. Kuhn, C. Bitter Taste Receptors for Saccharin and Acesulfame K. *Journal of Neuroscience* **24**, 10260–10265 (2004).
83. Suez, J. *et al.* Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature* **514**, 181–186 (2014).
84. Shallenberger, R. S. & Acree, T. E. Molecular Theory of Sweet Taste. *Nature* **216**, 480–482 (1967).
85. Kier, L. B. A Molecular Theory of Sweet Taste. *Journal of Pharmaceutical Sciences* **61**, 1394–1397 (1972).
86. Nofre, C. & Tinti, J.-M. Sweetness reception in man : the multipoint attachment theory. *Food Chemistry* **56**, 263–274 (1996).
87. Choi, Y.-H., Hussain, R. a., Pezzuto, J. M., Kingborn, A. D. & Morton, J. F. Abrusosides A-D, Four Novel Sweet-Tasting Triterpene Glycosides from the leaves of *Abrus precatorius*. *Journal of Natural Products* **52**, 1118–1127 (1989).
88. Zheng, J.-X., Rao, Z.-J. & Jia, C.-X. Study on the Relationship Between Structure and Sweetness of Sucrose Derivatives. *Food Science* **24**, 29–33 (2003).
89. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Research* **44**, D1202–D1213 (2016).
90. Aframian, D., Davidowitz, T. & Benoliel, R. The distribution of oral mucosal pH values in healthy saliva secretors. *Oral Diseases* **12**, 420–423 (2006).

91. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **17**, 490–519 (1996).
92. TALETE srl. Dragon. *Software for Molecular Descriptor Calculation* 2014.
93. Banerjee, P. *et al.* Super Natural II—a database of natural products. *Nucleic Acids Research* **43**, D935–D939 (2015).
94. Lagorce, D., Sperandio, O., Baell, J. B., Miteva, M. A. & Villoutreix, B. O. FAF-Drugs3 : a web server for compound property calculation and chemical library design. *Nucleic Acids Research* **43**, W200–W207 (2015).
95. Palczewski, K. Crystal Structure of Rhodopsin : A G Protein-Coupled Receptor. *Science* **289**, 739–745 (2000).
96. Lv, X. *et al.* In vitro expression and analysis of the 826 human G protein-coupled receptors. *Protein & Cell* **7**, 325–337 (2016).
97. Cherezov, V., Abola, E. & Stevens, R. C. in *Methods in molecular biology* (Clifton, N.J.) 141–168 (2010). arXiv : [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
98. Dorn, M., e Silva, M. B., Buriol, L. S. & Lamb, L. C. Three-dimensional protein structure prediction : Methods and computational strategies. *Computational Biology and Chemistry* **53**, 251–276 (2014).
99. Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242 (2000).
100. Jiang, P. *et al.* Identification of the Cyclamate Interaction Site within the Transmembrane Domain of the Human Sweet Taste Receptor Subunit T1R3. *Journal of Biological Chemistry* **280**, 34296–34305 (2005).
101. Jiang, P. *et al.* Lactisole Interacts with the Transmembrane Domains of Human T1R3 to Inhibit Sweet Taste. *Journal of Biological Chemistry* **280**, 15238–15246 (2005).
102. Winnig, M., Bufer, B., Kratochwil, N. a., Slack, J. P. & Meyerhof, W. The binding site for neohesperidin dihydrochalcone at the human sweet taste receptor. *BMC Structural Biology* **7**, 66 (2007).
103. Kunishima, N. *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–7 (2000).
104. Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H. & Morikawa, K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd³⁺. *Proceedings of the National Academy of Sciences* **99**, 2660–2665 (2002).
105. Monn, J. A. *et al.* Synthesis and Pharmacological Characterization of C4-(Thiotriazolyl)-substituted-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylates. Identification of (1R,2S,4R,5R,6R)-2-Amino-4-(1H-1,2,4-triazol-3-ylsulfanyl)bicyclo[3.1.0]hexane-2,6-dicarboxylic Acid (LY2812. EN. *Journal of medicinal chemistry* **58**, 7526–48 (2015).
106. Muto, T., Tsuchiya, D., Morikawa, K. & Jingami, H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. *Proceedings of the National Academy of Sciences* **104**, 3759–3764 (2007).

107. Geng, Y., Bush, M., Mosyak, L., Wang, F. & Fan, Q. R. Structural mechanism of ligand activation in human GABA(B) receptor. *Nature* **504**, 254–9 (2013).
108. Wu, H. *et al.* Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. *Science* **344**, 58–64 (2014).
109. Doré, A. S. *et al.* Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *en. Nature* **511**, 557–562 (2014).
110. UniProt : a hub for protein information. *Nucleic Acids Research* **43**, D204–D212 (2015).
111. Pin, J.-P., Galvez, T. & Prézeau, L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacology & Therapeutics* **98**, 325–354 (2003).
112. Zhang, F. *et al.* Molecular mechanism for the umami taste synergism. *Proceedings of the National Academy of Sciences* **105**, 20930–20934 (2008).
113. Zhang, F. *et al.* Molecular mechanism of the sweet taste enhancers. *Proceedings of the National Academy of Sciences* **107**, 4752–4757 (2010).
114. Masuda, K. *et al.* Characterization of the Modes of Binding between Human Sweet Taste Receptor and Low-Molecular-Weight Sweet Compounds. *PLoS ONE* **7** (éd. Meyerhof, W.) e35380 (2012).
115. Maillet, E. L. *et al.* Characterization of the Binding Site of Aspartame in the Human Sweet Taste Receptor. *Chemical Senses* **40**, 577–586 (2015).
116. Jiang, Y. *et al.* Elucidation of a Novel Extracellular Calcium-binding Site on Metabotropic Glutamate Receptor 1 α (mGluR1 α) That Controls Receptor Activation. *Journal of Biological Chemistry* **285**, 33463–33474 (2010).
117. Acher, F. C., Selvam, C., Pin, J.-P., Goudet, C. & Bertrand, H.-O. A critical pocket close to the glutamate binding site of mGlu receptors opens new possibilities for agonist design. *Neuropharmacology* **60**, 102–107 (2011).
118. Goudet, C. *et al.* Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. *Proceedings of the National Academy of Sciences* **101**, 378–383 (2004).
119. Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).
120. Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Current Opinion in Structural Biology* **21**, 541–551 (2011).
121. Trzaskowski, B *et al.* Action of molecular switches in GPCRs—theoretical and experimental studies. *Current medicinal chemistry* **19**, 1090–109 (2012).
122. Binet, V. *et al.* Common structural requirements for heptahelical domain function in class A and class C G protein-coupled receptors. *The Journal of biological chemistry* **282**, 12154–63 (2007).
123. Case, D. *et al.* AMBER 2016 University of California, San Francisco, 2016.

124. Gordon, J. C. *et al.* H++ : a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **33**, W368–W371 (2005).
125. Kirschner, K. N. *et al.* GLYCAM06 : A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* **29**, 622–655 (2008).
126. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157–1174 (2004).
127. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25**, 247–260 (2006).
128. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model : II. Parameterization and validation. *Journal of Computational Chemistry* **23**, 1623–1641 (2002).
129. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges : the RESP model. *The Journal of Physical Chemistry* **97**, 10269–10280 (1993).
130. Lomize, A. L., Pogozheva, I. D., Lomize, M. A. & Mosberg, H. I. Positioning of proteins in membranes : A computational approach. *Protein Science* **15**, 1318–1333 (2006).
131. Skjevik, Å. A., Madej, B. D., Walker, R. C. & Teigen, K. LIPID11 : A Modular Framework for Lipid Simulations Using Amber. *The Journal of Physical Chemistry B* **116**, 11124–11136 (2012).
132. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341 (1977).
133. Servant, G. *et al.* Positive allosteric modulators of the human sweet taste receptor enhance sweet taste. *Proceedings of the National Academy of Sciences* **107**, 4746–4751 (2010).
134. Assadi-Porter, F. M. *et al.* Key Amino Acid Residues Involved in Multi-Point Binding Interactions between Brazzein, a Sweet Protein, and the T1R2–T1R3 Human Sweet Receptor. *Journal of Molecular Biology* **398**, 584–599 (2010).
135. Toniolo, C. & Temussi, P. Conformational flexibility of aspartame. *Biopolymers* **106**, 376–384 (2016).
136. Pin, J.-P. *et al.* The activation mechanism of class-C G-protein coupled receptors. *Biology of the cell* **96**, 335–42 (2004).
137. Xue, L. *et al.* Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. *Nature Chemical Biology* **11**, 134–140 (2014).
138. Rondard, P. & Pin, J.-P. Dynamics and modulation of metabotropic glutamate receptors. *Current Opinion in Pharmacology* **20**, 95–101 (2015).
139. Sack, J. S., Saper, M. A. & Quiocho, F. A. Periplasmic binding protein structure and function. *Journal of Molecular Biology* **206**, 171–191 (1989).

140. Brock, C. *et al.* Activation of a Dimeric Metabotropic Glutamate Receptor by Intersubunit Rearrangement. *Journal of Biological Chemistry* **282**, 33000–33008 (2007).
141. Kniazeff, J., Prézeau, L., Rondard, P., Pin, J.-P. & Goudet, C. Dimers and beyond : The functional puzzles of class C GPCRs. *Pharmacology & Therapeutics* **130**, 9–25 (2011).
142. Doumazane, E. *et al.* Illuminating the activation mechanisms and allosteric properties of metabotropic glutamate receptors. *Proceedings of the National Academy of Sciences* **110**, E1416–E1425 (2013).
143. Lakkaraju, S. K., Xue, F., Faden, A. I. & MacKerell, A. D. Estimation of Ligand Efficacies of Metabotropic Glutamate Receptors from Conformational Forces Obtained from Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **53**, 1337–1349 (2013).
144. Kim, S.-k., Chen, Y., Abrol, R., Goddard, W. A. & Guthrie, B. Activation mechanism of the G protein-coupled sweet receptor heterodimer with sweeteners and allosteric agonists. *Proceedings of the National Academy of Sciences* **114**, 2568–2573 (2017).
145. Audet, M. & Bouvier, M. Restructuring G-Protein- Coupled Receptor Activation. *Cell* **151**, 14–23 (2012).
146. Shonberg, J., Kling, R. C., Gmeiner, P. & Löber, S. GPCR crystal structures : Medicinal chemistry in the pocket. *Bioorganic & Medicinal Chemistry* **23**, 3880–3906 (2015).
147. Cooke, R. M., Brown, A. J., Marshall, F. H. & Mason, J. S. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. *Drug Discovery Today* **20**, 1355–1364 (2015).
148. Venkatakrisnan, A. J. *et al.* Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).
149. Dror, R. O. *et al.* Activation mechanism of the β 2-adrenergic receptor. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18684–9 (2011).
150. De March, C. A. *et al.* Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *Journal of the American Chemical Society* **137**, 8611–8616 (2015).
151. Deupi, X. & Kobilka, B. in *Advances in protein chemistry* 07, 137–166 (2007).
152. Venkatakrisnan, a. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–94 (2013).
153. Mühlemann, A. *et al.* Determination of key amino acids implicated in the actions of allosteric modulation by 3,3-difluorobenzaldazine on rat mGlu5 receptors. *European Journal of Pharmacology* **529**, 95–104 (2006).
154. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779–815 (1993).

155. Rowe, B. A. *et al.* Transposition of Three Amino Acids Transforms the Human Metabotropic Glutamate Receptor (mGluR)-3-Positive Allosteric Modulation Site to mGluR2, and Additional Characterization of the mGluR2-Positive Allosteric Modulation Site. *Journal of Pharmacology and Experimental Therapeutics* **326**, 240–251 (2008).
156. Toda, Y. *et al.* Two Distinct Determinants of Ligand Specificity in T1R1/T1R3 (the Umami Taste Receptor). *Journal of Biological Chemistry* **288**, 36863–36877 (2013).
157. Malherbe, P. *et al.* Identification of essential residues involved in the glutamate binding pocket of the group II metabotropic glutamate receptor. *Molecular pharmacology* **60**, 944–954 (2001).
158. Masuda, T. *et al.* Five amino acid residues in cysteine-rich domain of human T1R3 were involved in the response for sweet-tasting protein, thaumatin. *Biochimie* **95**, 1502–1505 (2013).
159. Schaffhauser, H. Pharmacological Characterization and Identification of Amino Acids Involved in the Positive Modulation of Metabotropic Glutamate Receptor Subtype 2. *Molecular Pharmacology* **64**, 798–810 (2003).
160. Yao, Y., Pattabiraman, N., Michne, W. F., Huang, X.-P. & Hampson, D. R. Molecular modeling and mutagenesis of the ligand-binding pocket of the mGlu3 subtype of metabotropic glutamate receptor. *Journal of Neurochemistry* **86**, 947–957 (2003).
161. Jiang, P. *et al.* The cysteine-rich region of T1R3 determines responses to intensely sweet proteins. *The Journal of biological chemistry* **279**, 45068–75 (2004).
162. Hemstapat, K. *et al.* A Novel Family of Potent Negative Allosteric Modulators of Group II Metabotropic Glutamate Receptors. *Journal of Pharmacology and Experimental Therapeutics* **322**, 254–264 (2007).
163. Kang, H. J. *et al.* Selectivity and Evolutionary Divergence of Metabotropic Glutamate Receptors for Endogenous Ligands and G Proteins Coupled to Phospholipase C or TRP Channels. *Journal of Biological Chemistry* **289**, 29961–29974 (2014).
164. Hampson, D. R. *et al.* Probing the Ligand-binding Domain of the mGluR4 Subtype of Metabotropic Glutamate Receptor. *Journal of Biological Chemistry* **274**, 33488–33495 (1999).
165. Molck, C. *et al.* Pharmacological Characterization and Modeling of the Binding Sites of Novel 1,3-Bis(pyridinylethynyl)benzenes as Metabotropic Glutamate Receptor 5-Selective Negative Allosteric Modulators. *Molecular Pharmacology* **82**, 929–937 (2012).
166. Pagano, A. *et al.* The Non-competitive Antagonists 2-Methyl-6-(phenylethynyl)pyridine and 7-Hydroxyiminocyclopropan[b]chromen-1a-carboxylic Acid Ethyl Ester Interact with Overlapping Binding Pockets in the Transmembrane Region of Group I Metabotropic Glutamate Receptors. *Journal of Biological Chemistry* **275**, 33750–33758 (2000).
167. Rosemond, E., Peltekova, V., Naples, M., Thogersen, H. & Hampson, D. R. Molecular Determinants of High Affinity Binding to Group III Metabotropic Glutamate Receptors. *Journal of Biological Chemistry* **277**, 7333–7340 (2002).

168. Hermit, M. B., Greenwood, J. R. & Brauner-Osborne, H. Mutation-induced Quisqualic Acid and Ibotenic Acid Affinity at the Metabotropic Glutamate Receptor Subtype 4 : LIGAND SELECTIVITY RESULTS FROM A SYNERGY OF SEVERAL AMINO ACID RESIDUES. *Journal of Biological Chemistry* **279**, 34811–34817 (2004).
169. Fukuda, J. *et al.* Identification of a novel transmembrane domain involved in the negative modulation of mGluR1 using a newly discovered allosteric mGluR1 antagonist, 3-cyclohexyl-5-fluoro-6-methyl-7-(2-morpholin-4-ylethoxy)-4H-chromen-4-one. *Neuropharmacology* **57**, 438–445 (2009).
170. Goudet, C. *et al.* A novel selective metabotropic glutamate receptor 4 agonist reveals new possibilities for developing subtype selective ligands with therapeutic potential. *The FASEB Journal* **26**, 1682–1693 (2012).
171. Galvez, T. *et al.* Mutagenesis and Modeling of the GABAB Receptor Extracellular Domain Support a Venus Flytrap Mechanism for Ligand Binding. *Journal of Biological Chemistry* **274**, 13362–13369 (1999).
172. Suzuki, G. *et al.* Pharmacological Characterization of a New, Orally Active and Potent Allosteric Metabotropic Glutamate Receptor 1 Antagonist, 4-[1-(2-Fluoropyridin-3-yl)-5-methyl-1H-1,2,3-triazol-4-yl]-N-isopropyl-N-methyl-3,6-dihydropyridine-1(2H)-carboxamide (FTIDC). *Journal of Pharmacology and Experimental Therapeutics* **321**, 1144–1153 (2007).
173. Sato, T., Shimada, Y., Nagasawa, N., Nakanishi, S. & Jingami, H. Amino acid mutagenesis of the ligand binding site and the dimer interface of the metabotropic glutamate receptor 1. Identification of crucial residues for setting the activated state. *The Journal of biological chemistry* **278**, 4314–21 (2003).
174. Mukherjee, R. S., McBride, E. W., Beinborn, M., Dunlap, K. & Kopin, A. S. Point Mutations in Either Subunit of the GABAB Receptor Confer Constitutive Activity to the Heterodimer. *Molecular Pharmacology* **70**, 1406–1413 (2006).
175. Malherbe, P. Mutational Analysis and Molecular Modeling of the Allosteric Binding Site of a Novel, Selective, Noncompetitive Antagonist of the Metabotropic Glutamate 1 Receptor. *Journal of Biological Chemistry* **278**, 8340–8347 (2003).
176. Knoflach, F. *et al.* Positive allosteric modulators of metabotropic glutamate 1 receptor : Characterization, mechanism of action, and binding site. *Proceedings of the National Academy of Sciences* **98**, 13402–13407 (2001).
177. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* **89**, 10915–10919 (1992).
178. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).
179. Wilbur, W. J. & Lipman, D. J. Rapid similarity searches of nucleic acid and protein data banks. *Proceedings of the National Academy of Sciences* **80**, 726–730 (1983).
180. Eddy, S. R. Profile hidden Markov models. *Bioinformatics (Oxford, England)* **14**, 755–63 (1998).

181. Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507–2524 (2006).
182. Trott, O. & Olson, A. J. AutoDock Vina : Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, NA–NA (2009).
183. Ciccotti, G., Ferrario, M. & Ryckaert, J.-P. Molecular dynamics of rigid systems in cartesian coordinates A general formulation. *Molecular Physics* **47**, 1253–1264 (1982).
184. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98–103 (1967).

Modélisation moléculaire de la perception de la saveur sucrée : approches structurales et dynamiques.

Résumé

La consommation excessive et chronique de sucre est un facteur de risque pour l'apparition de pathologies telles que le diabète de type II ou l'obésité. Une des solutions pour répondre à cet enjeu majeur de santé publique, tout en conservant le plaisir de la saveur sucrée, consiste en l'utilisation d'édulcorants en substitution du sucre. Actuellement, un certain nombre d'édulcorants sur le marché présentent arrière-goût amer ou sont sujets à débat quant à leurs effets sur la santé. Un des objectifs de ces travaux de thèse consiste à proposer de nouveaux édulcorants grâce à des approches rationnelles *in silico*. Un modèle statistique a été établi sur la base des structures chimiques et a permis d'identifier de nouveaux édulcorants d'origine naturelle. Ensuite, la reconstruction par homologie du récepteur à la saveur sucrée et l'étude des sites de liaison apportent des indices, à l'échelle atomique, qui permettront d'identifier ou même de concevoir de nouveaux édulcorants. L'étude dynamique d'un récepteur de la même famille (Récepteur Couplé aux Protéines G (RCPG) de classe C) a permis d'émettre une hypothèse sur le mécanisme d'activation, phénomène important pour la compréhension de la conversion du signal chimique en signal électrique.

RCPG, classe C, Édulcorants, Pouvoir Sucrant, QSAR, Modélisation Moléculaire.

Molecular modeling of the sweet taste perception: structural and dynamic approaches.

Abstract

Sugar overconsumption is a risk factor for pathologies such as type II diabetes or obesity. Sweeteners consumption is used to overcome this public health issue. Indeed, they have low caloric value but still preserve the pleasure of sweet taste. Currently, number of sweeteners are commercially available, but they present a bitter aftertaste or there is a debate about their safety. One aim of this work was to propose new intense sweeteners using computational modeling strategies. Through a statistical approach to predict the sweetness based on the chemical structure of already known sweeteners, new natural compounds have been identified. Furthermore, the structural study of the homology model of the sweet taste receptor provides some clues to design new sweeteners. The molecular dynamic study of a class C G-protein coupled receptor gives the first molecular hypothesis of the activation process.

GPCR, Class C, Sweeteners, Sweetness, QSAR, Molecular Modeling.