Simulations moléculaires des interactions ligand-récepteur : Bases moléculaires de la perception chimio-sensorielle

Sébastien Fiorucci

UNIVERSITÉ CÔTE D'AZUR

Habilitation à Diriger les Recherches

Spécialité Chimie

présentée à

l'Université Côte d'Azur
École Doctorale Sciences Fondamentales et Appliquées

par

Sébastien Fiorucci

Simulations moléculaires des interactions ligand-récepteur : Bases moléculaires de la perception chimio-sensorielle

Soutenue le 22 Juin 2017

JURY

Manuel Dauchez, Prof. Université de Reims Champagne Ardenne
Nicolas Ferré, Prof. Aix-Marseille Université
Olivier Taboureau, Prof. Université Paris Diderot – Paris 7
Loïc Briand, DR INRA, CSGA - Dijon
Laurent Chaloin, CR CNRS, CPBS - Montpellier
Serge Antonczak, Prof. Université Nice Sophia Antipolis

Université Côte d'Azur, Institut de Chimie de Nice, UMR 7272 CNRS
06108 Nice Cedex 2, France

Remerciements

Je tiens tout d'abord à remercier Serge Antonczak et jérôme Golebiowski, collègues de travail devenus amis, pour m'avoir fait confiance, pour leurs encouragements tout au long de ces années et pour m'avoir laissé le choix de mes activités de recherches.

Je n'oublie pas de remercier deux personnes importantes qui m'ont accompagné dans ce long processus qu'est de devenir « habilité à diriger des recherches » : mon directeur de thèse Daniel Cabrol Bass et mon chef de postdoc à Brême, Martin Zacharias.

Merci aux membres du jury, Manuel Dauchez, Nicolas Ferré, Olivier Taboureau, Loïc Briand et Laurent Chaloin qui m'ont fait l'honneur d'accepter de juger mon travail et que j'ai le plaisir d'inviter.

Ce travail n'aurait jamais pu se faire sans les forces vives du laboratoire : stagiaires, doctorants et post-doctorants, actuels ou passés. Je citerai dans un ordre plus ou moins chronologique: Landry Charlier, Juan Fernandez Carmona, Jérémie Topin, Julien Diharce, Felix Lohmann, Emmanuelle Bignon, Chloé Pontet, Maryline Viano, Claire de March, Gleb Novikov, Iurii Casciuc, Caroline Bushdid, Xiaojing Cong, Hubert Grunig, et une mention spéciale pour le futur docteur Jean Baptiste Chéron que j'encadre depuis bientôt 3 ans et sans qui mon activité de recherche actuelle ne pourrait avancer !

Et comme toute recherche ne se fait jamais seul, je tenais à remercier tous les collègues de l'Institut de Chimie de Nice avec qui j'ai l'occasion d'échanger et/ou de travailler. Ce serait long de tous les citer, je noterai dans le désordre les personnes de l'équipe APSM actuelle : Sylvain Antoniotti, Xavier Fernandez, Thomas Michel, Martine Adrian Scotto, Uwe Meierhenrich, Cornelia Meinert, Jean-Jacques Filippi, Nicolas Baldovini, Sophie Poulain, Sandra Olivero et Elisabeth Duñach.

Table des Matières

Chapitre 1 – Introduction p.5

1. Avant propos
    - Parcours
    - Production Scientifique
    - Responsabilités scientifiques et administratives
    - Enseignements et encadrement
    - Diffusion de l'information scientifique et technique
2. Activités de recherche & Méthodologie
3. Interdisciplinarité & Collaborations

Chapitre 2 – Curriculum Vitae détaillé p.10

1. Détails Personnels
2. Fonctions
3. Formation Universitaire
4. Activités de Recherche
5. Responsabilités Scientifiques et Administratives
6. Encadrements de Recherche
7. Expertises et Compétences
8. Activité d'Enseignement

Chapitre 3 – Activités de Recherche antérieures p.18

1. Thèse de doctorat
2. Attaché Temporaire d'Enseignement et de Recherche
3. Stage Postdoctoral

Chapitre 4 – Activités de Recherche actuelles et perspectives p.23

1. Thématique 1 : Prédiction des interactions protéine-protéine
    - contexte scientifique
    - collaborations et financements
    - résultats clés et perspectives
2. Thématique 2 : Biosynthèse des substances naturelles
    - contexte scientifique
    - collaborations et financements
    - résultats clés et perspectives
3. Thématique 3 (Principale) : Bases moléculaires des sens chimiques
    - contexte scientifique
    - collaborations et financements
    - résultats clés et perspectives

Chapitre 5 – Conclusion p.34

Annexes …… p.36

1. Bibliographie exhaustive des productions scientifiques
   - Publications
   - Logiciels et base de données
   - Conférences sur invitation
   - Communications orales
   - Séminaires sur invitations
   - Communications par affiche
   - Formations CNRS et école chercheur
   - Communications Grand Public

2. Articles représentatifs
   - The anatomy of mammalian sweet taste receptors *(Proteins, 2017)*
   - Sweetness prediction of natural compounds *(Food. Chem., 2017)*
   - Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channeling in the course of flavonoid biosynthesis. *(Phys. Chem. Chem. Phys., 2016)*
   - Isolation and functional characterization of a t-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*. (*Plant. Mol. Biol. 2014).*
   - Prediction/calculation of protein-protein binding affinities and mutation effect *(in Protein-protein complexes: Analysis, modeling and drug design, 2010).*

Chapitre 1

Introduction

Avant propos
Activités de recherche & Méthodologie
Interdisciplinarité & Collaborations

1. Avant propos

Le présent mémoire est rédigé en vue de l'obtention de l'habilitation à diriger des recherches et retrace mon parcours en détaillant les jalons de mes travaux de recherche. Ce premier chapitre donne un aperçu rapide sur mon parcours, ma productivité scientifique, mes responsabilités administratives et scientifiques, mes activités d'enseignement et d'encadrement, et la diffusion de l'information scientifique et technique. Le chapitre 2 détaille l'ensemble de ces points. Le chapitre 3 résume mes travaux de recherche antérieurs à ma titularisation en tant que Maître de conférence. Le chapitre 4 résume mes projets de recherche actuels et ma vision à moyen et plus long terme.

NB : les notes de bas de page de couleur bleu correspondent aux articles que je co-signe.

- **Parcours**

Physico-chimiste de formation, j'ai obtenu une maîtrise de Chimie à Nice et un DEA de Chimie Informatique et Théorique à Nancy. Pendant ma thèse à Nice, j'ai utilisé les méthodes de mécanique quantique pour étudier la réactivité de métabolites secondaires (flavonoïdes) et de modélisation moléculaire afin d'élucider les interactions entre ces composés et différentes enzymes (Quercetinase, Lipoxygenase). J'ai ensuite effectué un stage postdoctoral (ATER) à Nice sur la compréhension des mécanismes d'interaction entre une molécule odorante et une protéine de transport des odorants puis un deuxième stage postdoctoral à l'Université de Brême sur la compréhension des interactions antigène-anticorps à l'aide de simulations de *docking* (ou amarrage moléculaire) protéine-protéine, méthodes théoriques pour lesquelles j'ai développé un nouveau champ de force gros-grains.

J'ai été recruté en Septembre 2008 à l'Institut de Chimie de Nice (UMR 7272 CNRS) au sein de l'équipe Arôme Parfum Synthèse et Modélisation et plus précisément dans le groupe Modélisation Moléculaire récemment renommé *Chemosim* (acronyme regroupant les mots clés *Chemistry, Chemoinformatics, Molecular Modeling, Emotion & Simulation*). J'ai cette chance de connaître les membres de l'équipe, et plus particulièrement les professeurs Serge Antonczak et Jérôme Golebiowski, depuis de nombreuses années. Cela a grandement facilité mon intégration dans le groupe et dans les thématiques prioritaires (sens chimiques, métabolites secondaires) de l'équipe et du laboratoire.

- **Production Scientifique**

Mes travaux de recherche ont donné lieu à 30 publications dont 22 dans des revues à comité de lecture (rang A), 4 chapitres de livre et 4 actes de colloques. Ma thèse a donné lieu à 7 publications, et chacun de mes projets de recherche a été validé par plusieurs publications. Ma spécificité au laboratoire sur le thème *interactions protéine-protéine* puis plus récemment sur la compréhension des mécanismes moléculaires de la *perception des saveurs* m'a permis de gagner en indépendance et de signer désormais mes publications comme auteur principal. 8 de mes 30 publications ont été rédigé en tant qu'auteur correspondant ou co-correspondant.

- **Responsabilités scientifiques et administratives**

Depuis 2014, je suis membre du conseil de gestion de l'UFR Sciences de l'Université de Nice Sophia Antipolis. Depuis 2015, je suis membre du conseil scientifique de l'Institut de Chimie de Nice (UMR 7272 CNRS). Depuis la création du GDR 3713 O^3 en 2015 (Odorant – Odeur – Olfaction), je suis membre du conseil scientifique et co-animateur du thème « Odorat & Goût ».

Nos approches théoriques nécessitent des moyens de calculs importants. Au laboratoire je suis responsable de la maintenance de notre serveur de stockage (NAS) et de l'administration système de notre serveur de calcul. Je fais parti du panel d'experts sollicités pour l'évaluation des projets déposés auprès de l'infrastructure européenne de calcul intensif (PRACE). Je suis régulièrement sollicité pour la relecture d'articles en chimie théorique, chimie physique ou biophysique. Je suis intervenu plusieurs fois comme animateur (chairman) à des manifestations internationales.

Note sur les projets de recherche et leur financement:

J'ai participé à plusieurs projets de recherche financés (ANR, PEP's CNRS, PHC, ...) pour lesquels un rapport d'avancement ou de fin de projet a dû être rédigé et j'ai été porteur et coordinateur de 3 projets financés. Bien que partenaire de plusieurs projets ANR, je n'ai jamais été porteur d'un projet « majeur ». J'ai eu l'occasion de déposer en tant que coordinateur 4 projets à l'appel d'offre ANR JCJC ; en 2009 et 2010 sur les interactions protéine-protéine puis en 2014 et 2015 sur les bases moléculaires de la perception des saveurs. Le dernier en date a été très bien reçu par 3 des 4 rapporteurs (note de 21, 35, 40, et 41 sur 45) et je ne désespère pas de décrocher le sésame ANR (ou autre) avant mes 42 ans… Age charnière avancée par la revue Nature comme étant un jalon important dans la vie d'un chercheur en quête de son premier projet majeur financé. [1]

- **Enseignements et Encadrement**

J'ai bénéficié d'une décharge de service d'enseignement lors de ma première année en tant qu'enseignant-chercheur et je remercie l'Université d'avoir mis en place cette procédure. Cela m'a permis de développer mes activités de recherche (partenaire d'un projet ANR en 2009, rédaction d'un chapitre de livre en 2010, porteur d'un projet PHC en 2012) et de renforcer mes collaborations (avec le Prof. M. Zacharias notamment) pendant mes premières années en poste.
J'ai stabilisé mes enseignements, essentiellement autour de la chimie structurale et la modélisation moléculaire. Cela me permet de sensibiliser les étudiants sur la recherche réalisée au laboratoire, et plus particulièrement au travers de nouveaux travaux pratiques « machine » ; par exemple sur la prédiction de la structure de récepteurs chimiosensoriels, sur l'identification de pharmacophore de ligands de récepteurs transmembranaires ou encore sur la simulation par dynamique moléculaire de systèmes ligand-récepteur. Cette activité d'enseignement m'a permis avec mes collègues de développer de nouveaux outils pédagogiques. Dans ce cadre, plusieurs projets ont été été financés par Unisciel et sont disponibles en ligne.

L'une de nos missions d'enseignant-chercheur est de former les étudiants à la recherche par la recherche. J'attache une importance toute particulière à cette mission et je souhaite soutenir ma HDR pour encadrer officiellement un étudiant bien que cela se soit déjà produit et traduit par des co-publications ou plus récemment par des co-encadrements. Au total, j'ai encadré 12 étudiants lors de stages de recherche et 1 post-doctorant. Le premier doctorant que j'ai co-encadré a soutenu sa thèse en 2015. Une deuxième thèse sous notre co-direction avec le Prof. Serge Antonczak a démarré en octobre 2014 et devrait s'achever à la fin de cette année.

- **Diffusion de l'information scientifique et technique**

La dissémination de mes travaux de recherche a donné lieu à 28 communications par affiches, 27 présentations orales, dont 4 invitations à des conférences et 6 séminaires sur invitation, 1 logiciel librement distribué et la mise en place d'une base de données disponible sur le site web de notre groupe de recherche. J'ai participé à l'organisation de 6 manifestations nationales depuis 2011 et 1 manifestation internationale qui a eu lieu à Nice en 2013 pour laquelle je faisais

[1] *B. Maher. « Young scientists under pressure: what the data show ? », Nature* ***2016****, 538, 444*

également parti du comité scientifique. J'ai également participé à la création du GDR O^3 qui au-delà de la dimension recherche a pour but de sensibiliser et faciliter le transfert de connaissances scientifiques et techniques vers le grand public.

2. Activités de recherche & Méthodologie

La motivation première de mes travaux de recherche est de combiner des approches expérimentales et théoriques dans le domaine de la chimie physique pour atteindre une meilleure compréhension des phénomènes à l'échelle atomique. Mes travaux en cours traitent de systèmes d'intérêt biologique concernant les processus moléculaires impliqués dans les sens chimiques (olfaction et gustation), la biosynthèse de métabolites secondaires ou encore la formation d'assemblages macromoléculaires.

Dans les approches computationnelles, notre expertise consiste à déterminer le niveau de détail nécessaire pour répondre au problème posé et de faire la balance entre la précision de la méthode utilisée et la rapidité du calcul. Pour éviter les erreurs ou simplement pour éviter de perdre du temps en utilisant une méthode non adaptée au problème, il faut répondre à certaines questions. Tout d'abord, est-ce que la structure des entités mises en jeu est connue ? Répondre à cette question permettra de s'orienter vers les méthodes basées sur la connaissance du ligand (approches « *ligand-based »*) ou sur la connaissance du récepteur (approches « *receptor-based »)*. Quel est le niveau de résolution spatiale et temporelle nécessaire pour décrire/prédire le processus moléculaire ? Par exemple les méthodes de chimie quantique décrivent de façon pointue la structure électronique des molécules sur un temps relativement court en comparaison à la simulation de quelques millisecondes d'un système moléculaire de plusieurs centaines de milliers ou millions d'atomes avec une approche de mécanique classique.

Quelle que soit la méthode de modélisation moléculaire utilisée, un point commun aux différents projets sur lesquels j'ai pu travailler est la prédiction des interactions entre deux (ou plus) objets moléculaires et la quantification de différentes grandeurs thermodynamiques comme par exemple l'enthalpie libre qui régit les processus réactionnels. Qu'elle soit de complexation lors de l'association d'un ligand à un récepteur ou d'activation lors d'une réaction chimique entre deux composés, l'estimation de l'enthalpie libre reste un défi pour la chimie/biochimie théorique :

- La description des processus réactionnels est un enjeu majeur en chimie et en biochimie. Du point de vue microscopique, la chimie théorique permet de fournir les indices nécessaires à la compréhension de systèmes réactifs complexes. Les méthodes *ab initio*, ainsi que les méthodes basées sur la théorie de la fonctionnelle de la densité sont utilisées pour prendre en compte explicitement le rôle des électrons, et ainsi caractériser les intermédiaires réactionnels et les états de transition d'un mécanisme réactionnel. Les méthodes de thermodynamique statistiques telles que les simulations de dynamique moléculaire permettent d'échantillonner au cours du temps les changements conformationnels d'un système moléculaire et de faire le lien entre une grandeur macroscopique et les différents états microscopiques du système.

- La description et la prédiction des assemblages macromoléculaires est également un enjeu majeur en biologie puisque l'ensemble des processus biologiques résultent de l'interaction entre une ou plusieurs entités moléculaires. La modélisation moléculaire et la bioinformatique permettent notamment de reconstruire et d'étudier les macromolécules d'intérêt biologique et d'analyser leurs propriétés afin de mieux comprendre leurs fonctions. Les méthodes d'alignement de séquence et de reconstruction par homologie ont rendu accessible l'étude de biomolécules encore non-cristallisées. Enfin les méthodes de docking ont l'avantage de prédire un grand nombre de complexes ligand/récepteur pour un coût « processeur » relativement faible.

- En pharmacologie, lorsqu'une cible thérapeutique est inconnue, il est tout de même possible d'extraire des informations sur celle-ci en se basant sur la connaissance des ligands (agoniste, non-agoniste, agoniste inverse, ...). Dans ce contexte, la chémoinformatique s'appuie sur un ensemble de méthodes statistiques pour résoudre ces problèmes. Par exemple, les méthodes QSPR/QSAR, pour *Quantitative Structure Propriety/Activity Relationships*, permettent de relier la structure chimique de ces composés à une de leur propriété intrinsèque ou à une de leur activité, ici biologique.

3. Interdisciplinarité & Collaborations

L'étude des systèmes moléculaires a` l'aide de modèles numériques, est par nature transdisciplinaire. En sciences fondamentales, la notion de système existe aussi bien en chimie, en physique qu'en biologie. Lors de mes différents projets de recherche j'ai pu me rendre compte à quel point l'échange entre chercheurs de différentes disciplines scientifiques était parfois difficile. Il me semble important de promouvoir la pluridisciplinarité et l'interdisciplinarité à l'heure où la recherche est principalement financée pour répondre à des attentes sociétales, par nature globale et non disciplinaire.

Mes travaux ont un caractère interdisciplinaire à la frontière entre chimie physique, biophysique ou encore biologie structurale tout en m'appuyant sur les avancées en informatique appliquée et en calcul intensif. Mes projets de recherche sont le plus souvent enrichis par les collaborations fructueuses avec des chercheurs de différentes spécialités telles que la biologie moléculaire, les neurosciences ou encore l'informatique.

Ci-dessous une liste des principales collaborations mises en œuvre et attestées par des co-publications et/ou des projets financés ou en cours.

Sur le thème des interactions protéine-protéine :

- Prof. Martin Zacharias	TU Munich (Allemagne)
- Dr. Chantal Prévost	IBPC Paris
- Dr. Pierre Poulain	DSIMB Paris

Sur le thème de la biosynthèse de métabolites secondaires :

- Prof. Sylvie Baudino	LBVPAM St Etienne
- Dr. Frédéric Julien	LBVPAM St Etienne

Sur le thème des sens chimiques :

- Dr. Loïc Briand	CSGA Dijon
- Dr. Emmanuelle Jacquin Joly	iEES-Paris Versailles
- Prof. Hirohaki Matsunami	Duke Durham (USA)
- Prof. Minghong Ma	UPenn Philadelphie (USA)
- Dr. Peihua Jiang	Monell Philadelphie (USA)
- Dr. Alexander Bachmanov	Monell Philadelphie (USA)
- Dr. Jérôme Waldispühl	McGill Montréal (Canada)

Chapitre 2

Curriculum Vitae

Détails personnels
Fonctions
Formation universitaire
Activités de recherche
Responsabilités scientifiques et administratives
Encadrements de recherche
Expertises et compétences
Activités d'enseignement

Sébastien FIORUCCI
Maître de Conférence depuis le 1er Septembre 2008.
Cl. Normale 7ème échelon depuis 16/07/2016
Section CNU 31
Prime d'encadrement doctoral et de recherche (2015-2019)
Né le 3 Décembre 1979 à Cannes (38 ans)
Nationalité française
Marié, 2 enfants.

Université Côte d'Azur
Faculté des Sciences
Institut de Chimie de Nice UMR 7272 CNRS
Equipe APSM, groupe Chemosim
06108 Nice cedex 2
FRANCE

+33 (0)4 92 07 61 45

+33 (0)4 92 07 61 51

Sebastien.Fiorucci@unice.fr

http://chemosim.unice.fr

http://icn.unice.fr/fiorucci

Fonctions

2008 -	Maître de conférences	UCA - UNS
2007 - 2008	Stage Postdoctoral	Jacobs University Bremen
2006 - 2007	Attaché Temporaire d'Enseignement et de Recherche	UNS

Formation universitaire

2003 - 2006	Doctorat de Chimie	UNS
2002 - 2003	DEA Chimie informatique et Théorique	Uni. H. Poincaré, Nancy I
2001 - 2002	Maîtrise de Chimie	UNS
2000 - 2001	Licence de Chimie	UNS
1997 - 1999	DEUG - Sciences et Structure de la Matière	UNS
1996 - 1997	Baccalauréat (S)	Lycée A. de Tocqueville, Grasse (06)

Thèse, soutenue publiquement le 6 octobre 2006, intitulée « *Activités biologiques de composés de la famille des flavonoïdes : approches par des méthodes de chimie quantique et de dynamique moléculaire* » en présence du jury composé de :

Dr. Manuel Ruiz-López (Rapporteur), DR CNRS, Université Henri Poincaré,Nancy I
Dr. David Perahia (Rapporteur), DR CNRS, Uniuversité Orsay, Paris XI
Dr. Gilles Iacazio (Examinateur), MCF, Université Paul Cezanne, Aix-Marseille III
Dr. Patrick Trouillas (Examinateur), MCF, Université de Limoges
Pr. Daniel Cabrol Bass (Directeur de thèse), Pr. UNS
Dr. Serge Antonczak (Directeur de thèse), Pr. UNS.

obtenue avec mention « très honorable » et les félicitations orales du jury.

Activités de recherche

- **Thèmes de recherche en modélisation moléculaire:**

Thème prioritaire *80 % du temps recherche*	–	Description des bases moléculaires des sens chimiques (olfaction et gustation)
Thèmes secondaires *20 % du temps recherche*	– –	Catalyse enzymatique dans les voies de biosynthèse Prédiction des interactions protéine-protéine

- **Production scientifique:**

Résumé :
- 30 publications (dont 4 chapitres d'ouvrage et 4 *conference proceedings*)
- 4 conférences sur invitation
- 17 communications orales
- 6 séminaires sur invitation
- 28 communications par affiche.

La liste complète des publications et communications est fournie en annexe.

Bibliométrie* :
- H-factor : 12
- Nombre de citations totales : 461
- Moyenne de 15 citations par article et de 50 citations par an

**source croisée ISI web of knowledge / SCOPUS / Google Scholar en date du 23 mars 2017.*

Liste de 5 publications significatives :
*annexées à la fin de ce document, le sigle * signifie auteur correspondant* **.**
- The anatomy of mammalian sweet taste receptors. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Proteins.* ***2017****, in press.*
- Sweetness prediction of natural compounds. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Food. Chem.* ***2017****, 221, 1421-1425.*
- Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channeling in the course of flavonoid biosynthesis. J. Diharce, J. Golebiowski, S. Fiorucci*, S. Antonczak*. *Phys. Chem. Chem. Phys.,* ***2016*** *18, 10337-10345.*

- Isolation and functional characterization of a t-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*. F. Julien*, S. Moja, A. Bony, S. Legrand, C. Petit, T. Benabdelkader, K. Poirot, S. Fiorucci, Y. Guitton, F. Nicole, S. Baudino, J.L. Magnard. *Plant. Mol. Biol.* ***2014****, 84, 227-241.*
- Prediction/calculation of protein-protein binding affinities and mutation effect, S. Fiorucci*, S. Antonczak, J. Golebiowski. in *Protein-protein complexes: Analysis, modeling and drug design*, M. Zacharias, *World Scientific (****2010****) p.295-317.*

- **Autres productions :**

- Développement d'un logiciel de docking protéine-protéine: Ptools/Attract.
Citations: *A. Saladin et al. BMC struct. Biol (****2009****) + Schneider et al. Meth Mol Biol (****2012****)*
Téléchargeable sur Github ou sur http://chemosim.unice.fr

- Base de données de composés sucrés : SweetenersDB
Citation : *J.B. Chéron et al. Food Chem. (****2016****)*
Téléchargeable sur http://chemosim.unice.fr

Responsabilités scientifiques et administratives

- **Projets financés :**

2017-2020	**Partenaire** du projet **ANR** Demeter porté par le Dr. E. Jacquin Joly (UMR iEES 1392, Versailles) et en collaboration avec le Dr. D. Boujard (UMS 3387, Rennes). Budget : **591 k€** *Titre : "Bio-olfacticides : produire plus avec moins d'insecticides"*
2015-2018	**Partenaire** du projet **ANR-NSF** NeuroComp (Fr-US) en collaboration avec le Prof. H. Matsunami (Duke) et Prof. M. Ma (U. Penn). Budget : **148 k€** (Fr) & **300 k$** (US) *Titre : "Predicting odorant-dependent and independent olfactory neuron activation based on receptor dynamics"*
2014-2015	**Partenaire** du projet **PEP's CNRS** EXOMOD en collaboration avec le laboratoire BVpam (U. St Etienne). Budget: **30 k€** *Titre: "Utilisation de la biodiversité des roses botaniques pour l'étude du rôle d'une NUDIX hydrolase dans la biosynthèse du parfum"*
2014	**Porteur** du projet innovant de l'ICN (**U. Nice**) Budget : **3,5k€** en complément de la thèse de doctorat Jean-Baptiste Chéron déjà soutenue par le Giract (**3k€**) et la Gen Foundation (**3,5k€**) *Titre : "Les bases moléculaires de la perception sucrée"*
2014	**Porteur** du projet Crédit Scientifique Incitatif (**CSI**) de l'Université de Nice Sophia Antipolis en collaboration avec le laboratoire I3S (U Nice). Budget: **3,5 k€** *Titre: CLIO "Chemogenomics Links In Olfaction"*
2013-2014	**Partenaire** au projet d'échange international **Egide** (PHC AURORA n° 28890TJ) en collaboration avec N. Reuter, Professeur Titulaire d'une chaire de Biologie Computationnelle à Bergen (Norvège). Budget : **9 k€** *Titre : "Modeling of membrane bound metabolon"*

2012-2013 **Porteur** d'un projet d'échange international **Egide** (PHC PROCOPE n° 26435SG) en collaboration avec M. Zacharias, Professeur titulaire d'une chaire de biophysique théorique à TU Munich (Allemagne). **6 co-publications**. Budget: **7 k€**
Titre : "Multiresolution and Multicomponent Docking of large biomolecular assembly"

2009-2012 **Partenaire et PI** de la tâche 1 de l'**ANR** Blanche 2009-2012 (projet NADYN n°NT09_504609) sur le design d'un peptide ligand de la Dynéine. **1 publication**. Budget: **334 k€.**
Titre : "Nanoparticules d'ADN fonctionnalisées par des peptides ligands de la dynéine pour l'amélioration de leur trafic intracellulaire"

- **Responsabilités électives :**

2015 - **Conseil scientifique** de l'Institut de Chimie de Nice, **UMR 7272 CNRS**

2014 - **Conseil de gestion** de l'**UFR Sciences** de l'Université de Nice-Sophia Antipolis

2012 - 2014 **CPRH** du **département de Chimie** de l'Université de Nice-Sophia Antipolis (Comité Permanent des Ressources Humaines).

2009 - 2012 **Bureau** du **Département de Chimie** de l'Université de Nice-Sophia Antipolis, en charge de la commission Ressources Informatiques

2010 - 2011 **Comité de Direction** de l'Institut de Chimie de Nice, **UMR 7272 CNRS**

- **Réseaux & Groupements de Recherche :**

2014 - **Comité scientifique** du GDR 3713 Odorant-Odeur-Olfaction & **Coordonateur** du thème "Relations Odorat-Goût". Budget: **4 k€/an**. https://gdro3.wordpress.com/

2013 - 2014 **Participant** au projet de montage de dossier GDR Odorant-Odeur-Olfaction. Budget: **75 k€**. (45 k€ pour 2013 et 2014 + 30k€ pour 2 PEP's de site en 2014).

2014- **Participant** du COST GLISTEN: GPCR-Ligand Interactions, Structures, and Transmembrane Signalling: a European Research Network

2013- **Participant** du GDR Chemoinformatique (**organisateur** des journées SFCi en 2015)

Membre de la Société Chimique de France (**SCF**), de la Société Française de biophysique (**SFB**), du Groupe de Graphisme et Modélisation Moléculaire (**GGMM**), du Groupe d'Etude des Membranes (**GEM**), de la Société Française de Chémoinformatique (**SFCi**) et du Réseau Français de Chimie Théorique (**RFCT**).

- **Membre du comité d'organisation** *(ORG)* **/ scientifique** *(SCI)* **:**
  - *ORG+SCI :* 7èmes journées de la Société Française de Chemoinformatique, 8-9 octobre 2015, Nice, France. http://sfci2015.wordpress.com/
  - *ORG* : 2ème journées interdisciplinaires thématiques O3, 9-11 octobre 2013, La Colle sur Loup, France http://www.unice.fr/icn-s/spip.php?article253
  - *ORG* : 1ère journée interdisciplinaire thématique O3, 4 juillet 2013, Nice, France http://www.unice.fr/icn-s/spip.php?article253
  - *ORG+SCI* : 5ème Journée Arômes & Parfums, 7 juin 2013, Nice, France. http://www.unice.fr/icn/journee_ap/

- *ORG+SCI* : Integrative Approaches for Modeling Biomolecular Complexes 2013, 29-31 mai 2013, Nice, France. http://iambc2013.wordpress.com
- *ORG* : 13ème Rencontre des Chimistes Théoriciens Francophones, 1-5 juillet 2012, Marseille, France. http://rctf2012.wordpress.com/
- *ORG* : AROMAGRI, groupe de réflexion sur les sens chimiques, 13-14 octobre 2011, Grasse, France. http://aromagri2011.wordpress.com

- **Membre du comité éditorial :**
  - BioMed Research International, section Computational Biology (**IF 2016 = 2.1**) (http://www.hindawi.com/journals/bmri/editors/computational.biology/)
  - Frontiers in Molecular Biosciences, section Mathematics of Biomolecules (début en 2014, pas encore d'IF) http://www.frontiersin.org/Mathematics_of_Biomolecules/editorialboard
  - ISRN Thermodynamics (début en 2014, pas encore d'IF) http://www.isrn.com/journals/thermodynamics/editors/
  - Dataset Papers in Science, section Biophysics (début en 2013, pas encore d'IF) http://www.datasets.com/journals/biology/editors/biophysics/

- **Activité de « Peer-reviewing »** :
  Journal of Molecular Graphics and Modeling, International Journal of Quantum Chemistry, BMC structural biology, Journal of Physical Chemistry, BBA Proteins and Proteomics, Journal of Molecular Modeling, Molecular Simulation, PLoS One.

- **Membre du panel d'expert scientifique pour l'infrastructure européenne PRACE** (*Partnership for Advanced Computing in Europe*, http://www.prace-ri.eu/), plus d'une dizaine d'expertises réalisées.

- **Administration système, équipe APSM du laboratoire ICN (UMR 7272 CNRS)**
  - serveur de calcul haute performance (48 CPU)
  - serveur de sauvegarde (type NAS, stockage 25 To)

Encadrement de recherche

Postdoctorant :

- **Juan Fernandez Carmona** (2010) **: 1 publication commune** sur *les interactions tannins/protéines peu structurées* (*Mol Inf 2011*)

Doctorant :

- **Jean Baptiste Chéron** (2014-) : encadrement à **50%** sur la *modélisation des récepteurs membranaires impliqués dans les sens chimiques.* **3 publications** (*Food Chem, Proteins, L'Actu Chim 2017*) + **1 articles soumis** (*J Comp Chem*) et **3 en préparation**.
- **Julien Diharce** (2011-2014) : encadrement à **50%** sur la *formation de complexes protéine-protéine impliqués dans la biosynthèse de substances naturelles.* **3 publications** (*PCCP 2016, J. Phys Chem B* et *Structure 2014*)
- **Jérémie Topin** (2008-2012) **1 co-publication** sur la *prédiction de la cinétique de diffusion du dioxygène dans une hydrogénase.* (*J Phys Chem B 2014).*
- **Landry Charlier** (2006-2009) : **1 co-publication** sur la *prédiction d'enthalpie libre de complexation dans les systèmes hydrophobes* (*PCCP 2007).*

Master 2 recherche:

- **Hubert Grunig** (2016-2017) : encadrement à **100 %** sur *l'étude de nouveaux ligands de récepteurs olfactifs d'insectes par approches QSAR* (projet ANR Demeter 2017-2021).
- **Marilyne Viano** (2013-2014) : encadrement à **100%** sur *l'étude de protéines membranaires impliquées dans les sens chimiques par simulations de dynamique moléculaire.* **1 publication** (*Chem Senses 2016*)
- **Felix Lohmann** (2011-2012) : encadrement à **100%** sur *la simulation par dynamique moléculaire de systèmes ligand-récepteur*. **Bourse ERASMUS**

Master 1

- **Chloé Pontet** (2012-2013) : stage sur la *Description des premières étapes moléculaires de la gustation* (encadrement à **100%**)
- **Océane Dunand** (2011-2012) : mémoire bibliographique sur la *perception gustative* (encadrement à **100%**)
- **Sylvain Luciano** (2010-2011) : mémoire bibliographique sur les *calculs d'enthalpie libre par des méthodes de simulations de dynamique moléculaire*. (encadrement à **100%**)
- **Jérémie Godemert** (2009-2010) : mémoire bibliographique sur la *prédiction des interactions ligand-récepteur par des méthodes de simulations de dynamique moléculaire*. (encadrement à **100%**)

Autres :

- **Iurii Cascuic** (2014-2015) : stage sur la *Prédiction des interactions molécules sapides/récepteurs gustatifs* (encadrement à **100**%) **1 publication** (*Food Chem 2017*)

Expertises et compétences

- **Domaines d'expertise (et logiciels dédiés) :**

- Modélisation moléculaire (AMBER, Maestro, APBS, …)
- Chimie quantique (Gaussian, GAMESS, NBO)
- Bioinformatique structurale & Drug Design (Jalview, Clustal, serveurs Uniprot, PIR, Modeller, Autodock, Attract, ...)
- Visualisation (VMD, Chimera, Pymol, Rasmol, Molekel, Molden, ...)
- Analyse statistique (Knime, R)

- **Compétences informatiques :**

- Administration Système :
  Cluster de calcul haute performance (Serveur Dell C6100), Station de travail OS Linux, Station de travail Unix SGI : Octane et Indigo II (Irix), Station de travail Sun : Ultra60 et Ultra80 (Solaris 9).
- Programmation :
  Python, C-shell, C/C++, Delphi, Pascal. Quelques notions en HTML & Fortran.

Enseignements

2010-	Maître de Conférences	192 h. eq. TD
2009-2010	Maître de Conférences avec décharge « nouveaux entrants » de 36h eq. TD	156 h. eq. TD
2008-2009	Maître de Conférences stagiaire avec décharge « nouveaux entrants » de 48h eq. TD	144 h. eq. TD
2006-2007	½ ATER	96 h. eq. TD
2003-2006	Monitortat	64 h. eq. TD / an

- **Service d'enseignement actuel :**
  (réalisé en filière Physique-Chimie ou Chimie si non spécifié)
  - Modélisation Moléculaire (M1)
  - Modélisation Moléculaire (5e année ingénieur Polytech'Nice)
  - Drug Design (5e année ingénieur Polytech'Nice)
  - Thermodynamique Statistique (L3)
  - option Modélisation Moléculaire (L3)
  - Informatique Disciplinaire Chimie (L2)
  - option Structure des biomolécules (L1)
  - Chimie Structurale (L1)

- **Autres enseignements réalisés sur la période 2003-2016 :**
  - Bioinformatique Structurale (label national de Chimie théorique – RFCT), cours et TP dispensés à l'ENS Lyon
  - Informations Scientifiques et Techniques (M1)
  - Chimie quantique (M1)
  - Risque Cimique (4e année ingénieur Polytech'Nice)
  - Chimie des Solutions (Préparation Agreg PC)
  - Chimie atomistique (PCEM1)
  - TP de Chimie générale (L1), chimie des solutions (L1), chimie inorganique (L3) et option chimie en pratique (L2 SV)

- **Production de documents pédagogiques en ligne** (projets financés par Unisciel)
  - cours - TD - TP d'un module de *Modélisation moléculaire* (**niveau L3**)
    http://uel.unisciel.fr/chimie/modelisation/modelisation/co/modelisation.html
  - cours - TD - TP d'un module de *Chimie Quantique* (**niveau L3**)
    http://uel.unisciel.fr/chimie/modelisationII/modelisationII/co/modelisationII.html
  - Réorganisation du module *Structure Microscopique de la matière* (**niveau L1-L2**) + production de QCM http://socles3.unisciel.fr/
  - Cours - TD -TP d'un module de Biologie structurale (**niveau L1-L2**)
    http://sites.unice.fr/site/ffontaine/structurebio/co/structure.html
  - Cours - TD -TP d'un module de Drug Design (**niveau L3-M1**) *en cours d'édition.*

Chapitre 3

Activités de recherche antérieures

Thèse de doctorat
Activités post-doctorale

1. Thèse de doctorat (2003-2006 – Université de Nice Sophia Antipolis) :

« Activités biologiques de composés de la famille des flavonoïdes : Approches par des méthodes de chimie quantique et de dynamique moléculaire »

Les flavonoïdes font partie d'une des plus importantes familles de substances naturelles, les polyphénols. La quercétine est l'une des plus représentatives de cette famille de composés du fait de son activite´ dans la quasi-totalite´ des processus biochimiques pouvant impliquer un flavonoïde. La quercétine est un antioxydant et un composé antiradicalaire jouant un rôle dans le piégeage des espèces réactives de l'oxygène (ERO), mais sous certaines conditions elle peut être a` l'origine d'une activite´ pro-oxydante sur les acides nucléiques et aminés (notamment en présence de métaux de transition) et enfin avoir la propriété´ d'inhiber certaines enzymes. A l'aide des méthodes de chimie théorique, mon travail de thèse a consisté à décrire les propriétés intrinsèques de la quercétine, molécule modèle, et d'étudier divers mécanismes biologiques pour lesquels les flavonoïdes pouvaient être considérés comme substrats ou inhibiteurs. Voici un résumé des principaux résultats.

- **Propriétés réactionnelles des formes activées de la quercétine et étude de leur implication dans les processus biologiques.**

Une étude à l'aide de méthodes DFT nous a permis de caractériser les propriétés électroniques, énergétiques et structurales de la quercétine.[2] Le caractère aromatique des flavonoïdes est depuis longtemps connu et responsable de la stabilisation de leurs formes radicalaires. Il nous est pourtant apparu que la délocalisation était limitée pour certaines formes radicalaires de type semiquinone (*Fig. 3.1, gauche*). Ceci induit une spécificité des sites réactifs sur ces substances en fonction de la position des groupements hydroxyles ou méthoxyles. Une étude des co-réactifs (O_2, H_2O_2, $OH^{\cdot}$, ...) impliqués dans les réactions enzymatiques a permis d'établir des cycles thermodynamiques pouvant expliquer la faisabilité de telle ou telle réaction.
La quercétine est susceptible de métaboliser l'oxygène moléculaire au sein d'une enzyme spécifique: la Quercétine 2,3-Dioxygenase (2,3-QD). Le mécanisme réactionnel (*Fig. 3.1, droite*) a pu être élucide´ a` l'aide de méthodes de la fonctionnelle de la densité.[3] L'oxygénolyse enzymatique de la quercétine est précédée par une réaction de complexation avec l'ion cuivre(II) du site actif de l'enzyme et par une réaction de déprotonation du substrat. L'oxydation de la quercétine par le dioxygène implique l'ouverture du cycle aromatique « central » et une libération de monoxyde de carbone.

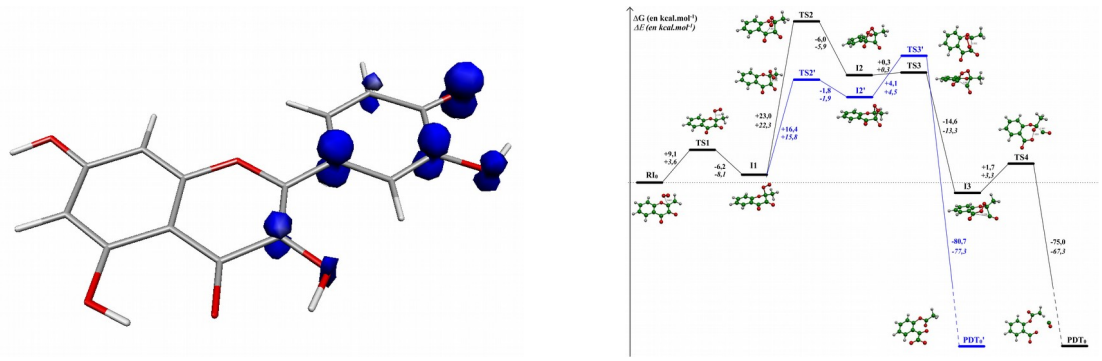


***Figure 3.1** : Gauche : Densité de spin (bleu) d'une forme semiquinone de la quercétine. Droite : Profil réactionnel d'un composé modèle permettant d'expliquer la réactivité de la quercétine vis-à-vis du dioxygène.*

[2] *Fiorucci et al., J. Agr. Food Chem.**, 2007,** 55, 903-91*
[3] *Fiorucci et al. Chemphyschem, **2004**, 5, 1726-1733*

- **Etude du caractère antioxydant de la quercétine - substrat de l'enzyme Quercétine 2,3-Dioxygénase.**

L'étude mécanistique (DFT) a été complétée en tenant compte du rôle de l'enzyme par une approche de simulations de dynamique moléculaire. Le cofacteur métallique (Cu^{2+}) entouré de trois résidus histidine et d'un résidu glutamate permet le transfert électronique et donc l'activation du substrat. La spécificité de l'enzyme pour la quercétine et le kaempférol a pu être explicitée en rationalisant le rôle de reconnaissance des acides aminés en fond de poche (*Fig. 3.2, gauche*). Cette étude a également permis de mettre en évidence l'existence d'un canal à dioxygène (*Fig. 3.2, droite*) qui mènerait directement une molécule de petite taille vers le site actif de l'enzyme.[4] Le passage du dioxygène à travers ce canal à été étudié à l'aide des méthodes de calcul du potentiel de force moyenne et d'échantillonnage avancé (LES).[5] La faible barrière énergétique ($\Delta G \sim 2$ kcal.mol^{-1}) montre que l'entrée du dioxygène via le canal est faisable.

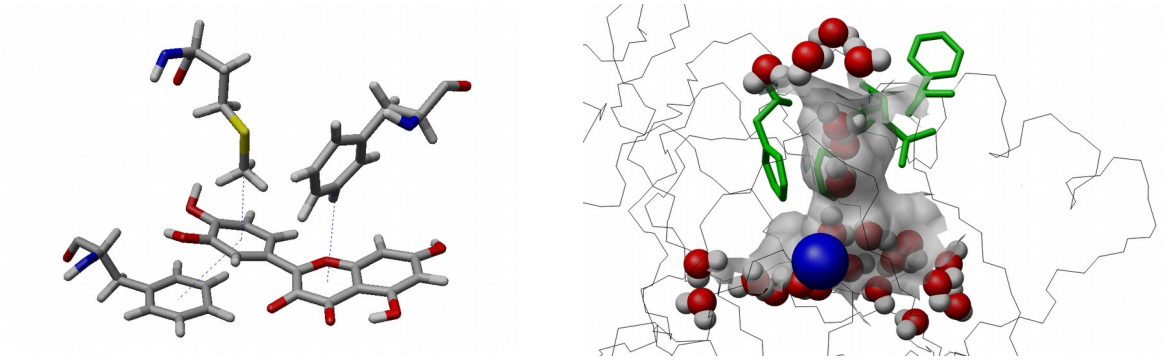


Figure 3.2 : *Gauche : Interactions hydrophobes permettant la reconnaissance du substrat (quercétine) par l'enzyme 2,3-QD. Droite : Canal à dioxygène (surface en gris) formé par quatre acides aminés (en vert). Le cuivre et les molécules d'eau du site actif sont respectivement en bleu et en gris/rouge.*

- **Etude du caractère inhibiteur de la quercétine sur la Lipoxygenase.**

Une étude a démontré que le produit de la dégradation de la quercétine par la lipoxygénase-3 (LOX-3) est un inhibiteur de cette enzyme.[6] Etant donné qu'aucune structure de la LOX-3 avec le substrat quercétine n'a pu être cristallisée à ce jour, nous nous sommes attachés à décrire le mode de complexation de la quercétine au sein de l'enzyme.

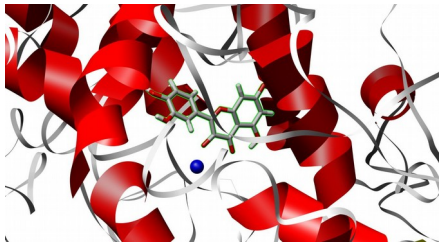


Figure 3.3 : *Site actif de la LOX-3 en interaction avec la quercétine*

A l'aide des méthodes de dynamique moléculaire, nous avons estimé l'enthalpie libre de complexation de la quercétine selon trois modes d'interaction.[7] Nos résultats montrent clairement que la quercétine se coordonne au centre métallique via sa fonction 3-hydroxychromone (*Fig. 3.3*). Une étude des interactions enzyme/substrat a permis d'identifier les résidus à l'origine de la reconnaissance et de l'activation de la quercétine.

[4] *Fiorucci et al., Proteins,* ***2007,*** *67, 961-970*
[5] *Fiorucci et al., Proteins,* ***2006****, 64, 845-850*
[6] *Borbulevych et al. Proteins,* ***2004****, 54, 13-9*
[7] *Fiorucci et al., Proteins,* ***2008****, 73, 290-298*

2. Attaché Temporaire d'Enseignement et de Recherche (2006-2007, UNS) :

« Les premières étapes moléculaires de l'olfaction : Simulation des interactions OBP - odorants»

La réception de composés odorants au sein de la muqueuse nasale implique différentes protéines du système olfactif. La transformation en signal électrique du signal chimique s'effectue par des récepteurs transmembranaires situés sur les cils olfactifs. Les protéines de liaison aux odorants (ou OBP pour *Odorant Binding Protein*), protéines localisées dans le mucus olfactif, joueraient un rôle de complexation, de solubilisation et de transport des odorants vers les récepteurs olfactifs.
A l'aide de simulations de dynamique moléculaire, nous avons caractérisé les interactions OBP/odorant intervenant lors du processus de complexation et de dé-complexation (*Fig. 3.4*).[8] Différentes approches (PMF, MMPBSA, TI) ont été utilisés pour prédire l'enthalpie libre d'association de ces systèmes très hydrophobes.[9] Nous avons pu comparer les résultats obtenus par ces différentes méthodes et certains sont en très bon accord avec ceux obtenus par titrage calorimétrique isotherme qui nous informe de la constante d'association, la stœchiométrie, l'enthalpie et l'entropie des complexes.

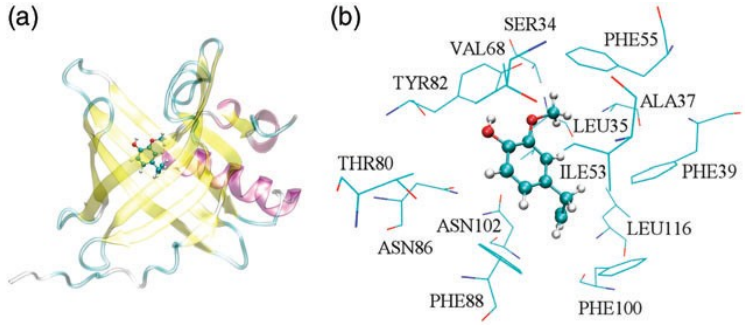


Figure 3.4* : *Structure d'une OBP (a) et détail des interactions OBP-odorant (b). *(Figure tirée de la référence [9])*

3. Stage Postdoctoral (2007-2008, Jacobs University Bremen, Allemagne) :

« Etude de complexes Antigène/Anticorps et de leur interaction avec le système du complément »

Cette étude entre dans le cadre d'un projet pluridisciplinaire européen « BacAbs ». L'objectif était d'étudier des candidats-vaccins contre la méningite (modèle : *Neisseria Meningitidis*), et de mettre au point des outils bioinformatiques pour prédire leurs propriétés structurales. Dans ce cadre, mon objectif a été de développer un algorithme de docking spécifique pour les complexes antigène/anticorps et de comprendre l'interaction de ces structures avec le système du complément (*Fig. 3.5, gauche*).

[8] *Golebiowski et al., Proteins, **2007,** 67, 448-458*
[9] *Charlier et al., Phys. Chem. Chem. Phys., **2007**, 9, 5761-5771*

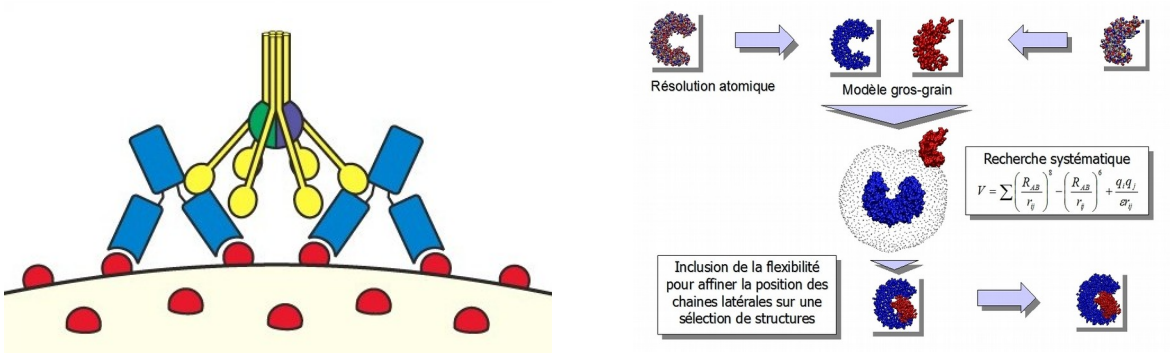


Figure 3.5 : *Gauche : Schéma d'une complexe formée par un antigène (en rouge), un anticorps (en bleu) et un des protagonistes du système immunitaire, le système du complément C1 (en jaune, vert et violet). Droite : Protocole de docking d'ATTRACT.*

La première partie de ce projet a porté sur le développement (en Python/C++) d'un programme de docking, Ptools/ATTRACT[10] (Fig. 3.5 droite), et d'optimiser la fonction de score gros-grains pour l'étude de complexes protéine-protéine[11]. La deuxième partie du projet[12] traite de l'analyse de la surface des protéines (forme de la protéine, accessibilité des résidus, profil électrostatique, caractère hydrophobe, etc.). Cela nous a permis d'identifier les possibles sites de complexation et de déterminer la région complémentaire sur un antigène (épitope) impliqué dans l'infection par la bactérie *chlamydia trachomatis*.[13]

[10] *Saladin et al., BMC Struct Biol,* ***2009****, 9, 27*
[11] *Fiorucci et al. Proteins,* ***2010****, 78, 3131-3139*
[12] *Fiorucci et al. Biophys J,* ***2010****, 98, 1921-1930*
[13] *Soriani et al. J. Biol. Chem.* ***2010****, 285, 30126-30138*

Chapitre 4

Activités de recherche actuelles et Perspectives

Thématique 1 : Prédiction des interactions protéine-protéine
Thématique 2 : Biosynthèse des substances naturelles
Thématique 3 (Principale) : Bases moléculaires des sens chimiques

1. Bases moléculaires des sens chimiques

- Thématique principale depuis 2014 -

- **Contexte scientifique**

Anatomie du système sensoriel. Notre cerveau est fait de milliards de neurones et est considéré comme l'organe le plus complexe de notre corps. Parmi nos cinq sens, l'odorat et le goût sont les deux sens chimiques permettant aux êtres vivants d'identifier et de réagir aux substances présentes dans leur environnement. Ces deux sens sont intimement liés. 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 sapides par la langue) à laquelle s'ajoutent des sensations olfactives (rétronasale) et somesthésiques (thermiques, tactiles et proprioceptives). L'espace chimique des molécules pouvant activer notre odorat ou notre goût est immense. L'homme peut par exemple distinguer plus de mille milliard de molécules odorantes.[14] L'extraordinaire capacité de discrimination de notre cerveau provient de l'organisation de notre système de détection. Au niveau physiologique, le premier niveau du traitement de l'information par le cerveau se situe lors de l'interaction entre une molécule odorante ou sapide et un récepteur de notre système sensoriel (Figure 4.1 et 4.2).

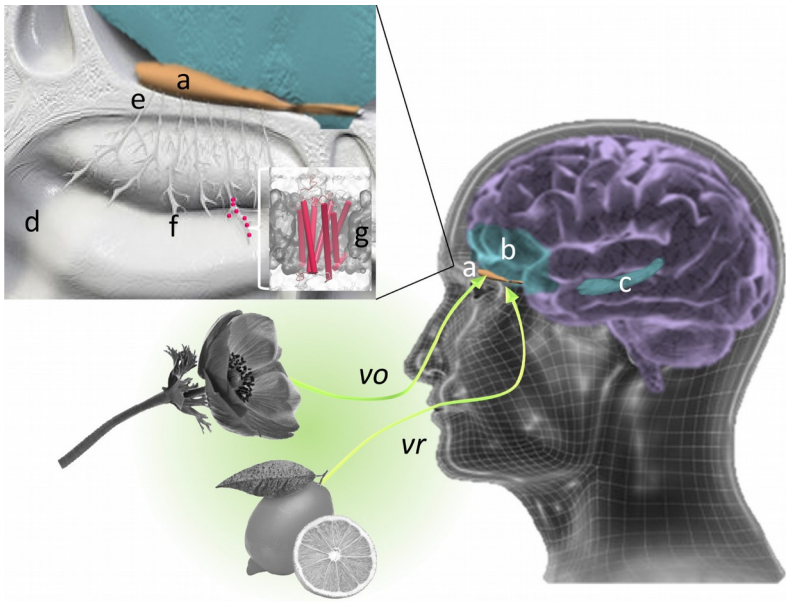


Figure 4.1*: Processus de chimio-réception de molécules odorantes. Lorsque nous inhalons (voie orthonasale, vo) ou ingérons (voie rétronasale, vr) des molécules odorantes, une partie de notre cerveau décode leur message moléculaire : il s'agit de notre bulbe olfactif (a). Le signal induit par ces molécules est alors transmis à différentes zones cérébrales (b,c) pour y être décodé. Au niveau moléculaire, les composés odorants sont inhalés via la cavité nasale (d), séparée de la boîte crânienne par la lame criblée de l'os ethmoïde (e) Les molécules entrent ensuite en interaction avec nos neurones olfactifs (f) qui expriment dans leur membrane leurs récepteurs olfactifs (g), à l'origine de l'influx neuronal. *(image issue de Bushdid et al. Actualité chimique 2016)*

[14] *Bushdid et al. Science, **2014**, 343, 1370-1372.*

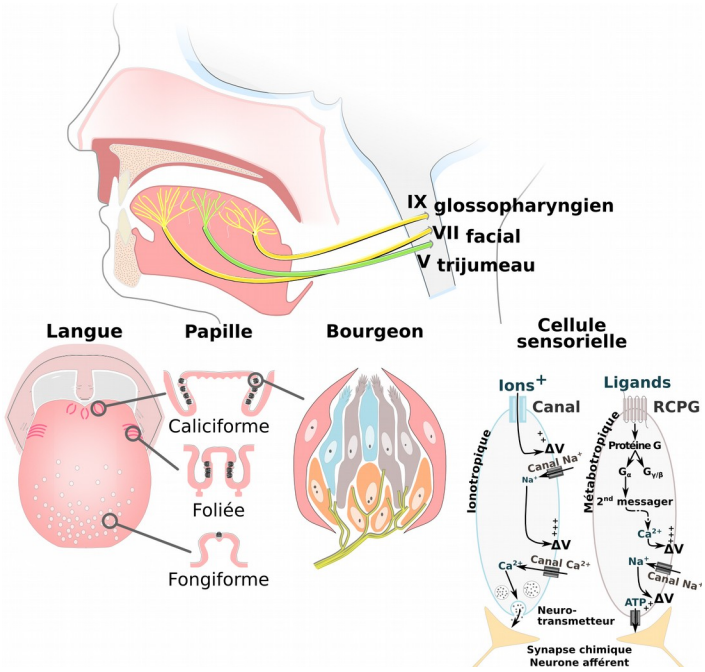


Figure 4.2* : *Anatomie du système gustatif. La surface de la langue possède des papilles gustatives qui varient en nombre, forme et localisation. Ces papilles contiennent des bourgeons gustatifs composés de plusieurs strates cellulaires en « forme d'oignons » (cellules basales, de soutien et sensorielles). La cellule sensorielle contient les récepteurs gustatifs de type canal ionique ou métabotropique. Selon la nature du récepteur, des voies de signalisation différentes sont sollicitées pour mener à l'activation du neurone afférent. Ces neurones 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 et analysée de manière consciente et inconsciente. Le nerf trijumeau est support de la sensation somesthésique. *(image issue de la thèse de Jean Baptiste Chéron)*

Codage de l'information chimique. Depuis les travaux de R. Axel et L. Buck (primés par le prix Nobel de physiologie et de médecine en 2004) sur la découverte de la famille de gènes des récepteurs olfactifs, on comprend mieux le traitement combinatoire du codage de l'information chimique. Les gènes codant pour l'ensemble des récepteurs olfactifs forment l'une des plus grandes familles du génome humain avec environ 900 représentants (dont 500 ne sont plus fonctionnels). Les mécanismes de chimio-réception impliqués dans le sens de la gustation sont similaires à ceux de l'olfaction. L'homme possède plus d'une trentaine de gènes pour distinguer les cinq saveurs primaires. L'identification de ces récepteurs dans les années 2000 a largement contribué à une meilleure compréhension des premières étapes moléculaires de la perception gustative. Sachant qu'une molécule (odorante ou sapide) peut activer plusieurs récepteurs sensoriels, qu'un récepteur peut répondre à plusieurs molécules différentes et que la perception d'une odeur ou d'une saveur peut provenir d'une seule molécule ou d'un mélange de molécules, la combinatoire est virtuellement infinie. Les récepteurs sensoriels du goût, canaux ioniques mis à part, et de l'odorat font partie de la famille des Récepteurs Couplés à une Protéine G (RCPG). D'un point de vue pharmacologique, une molécule à l'origine d'une odeur ou d'une saveur doit s'assembler avec un récepteur sensoriel sur le même principe que l'association clé-serrure. Leur mécanisme d'activation va ensuite dépendre fortement de la classe du récepteur en question :

Récepteur olfactifs. Les récepteurs olfactifs sont des RCPG de classe A dont le mécanisme d'activation est similaire à celui de la Rhodopsine, récepteur prototypique des RCPG de classe A. Ils agissent comme des « pinces » moléculaires qui peuvent s'ouvrir lorsqu'un ligand (ici une molécule odorante) s'associe avec eux. Si le ligand est un agoniste, le récepteur se lie à la protéine G enclenchant ainsi la cascade biochimique intracellulaire de transduction du signal chimique en signal électrique (dissociation de la sous unité alpha de la protéine G, libération du GDP, …). S'il s'agit d'un agoniste inverse (ou inhibiteur), la formation du complexe verrouille le

récepteur empêchant ainsi le couplage à la protéine G. Notre équipe a récemment démontré que la connaissance de la structure de ces récepteurs sensoriels permettait de comprendre leurs comportements dynamiques et le mécanisme d'ouverture du récepteur[15-16](Fig. 4.3).

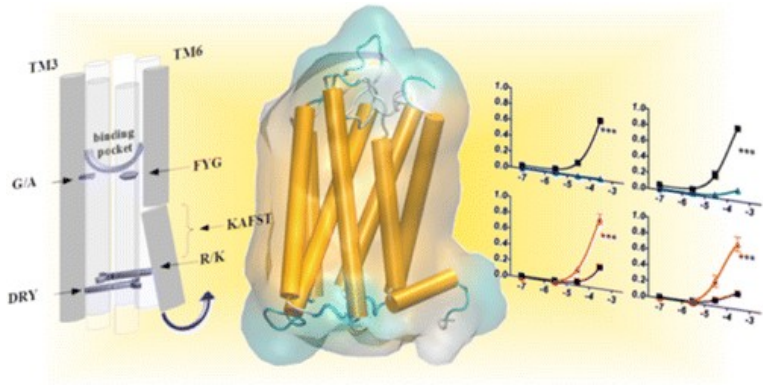


Figure 4.3* : *modèle 3D d'un récepteur olfactif (centre) et résidus importants validés expérimentalement (courbes dose-réponse à droite) impliqués dans la reconnaissance du ligand et dans le mécanisme d'activation (gauche). *(Image issue de la référence 15)*

Récepteurs Gustatifs. Les récepteurs gustatifs forment une famille hétérogène de récepteurs sensoriels (Fig. 4.4). Les canaux ioniques PKD2L1 et ENaC sont responsables de la perception du goût acide et salé et ne seront pas d'avantage détaillés dans ce document. Les récepteurs métabotropiques sont responsables de la perception des goûts sucré et umami pour les récepteurs gustatif de type 1 (ou T1Rs pour « Taste 1 Receptors») et du goût amer pour les récepteurs gustatifs de type 2 (ou T2Rs). On suppose que le mécanisme d'activation de ces derniers est proche des RCPG de classe A malgré une faible identité de séquence. Les récepteurs T1Rs se distinguent avant tout des autres récepteurs par leur structure. Les T1Rs, appartenant aux RCPG de classe C, possèdent un grand domaine extracellulaire et fonctionnent sous forme hétérodimérique (T1R1-T1R3 pour la perception du goût umami et T1R2-T1R3 pour la perception du goût sucré). Leur mécanisme d'activation reste encore flou, d'autant qu'ils possèdent de nombreux sites de liaison.

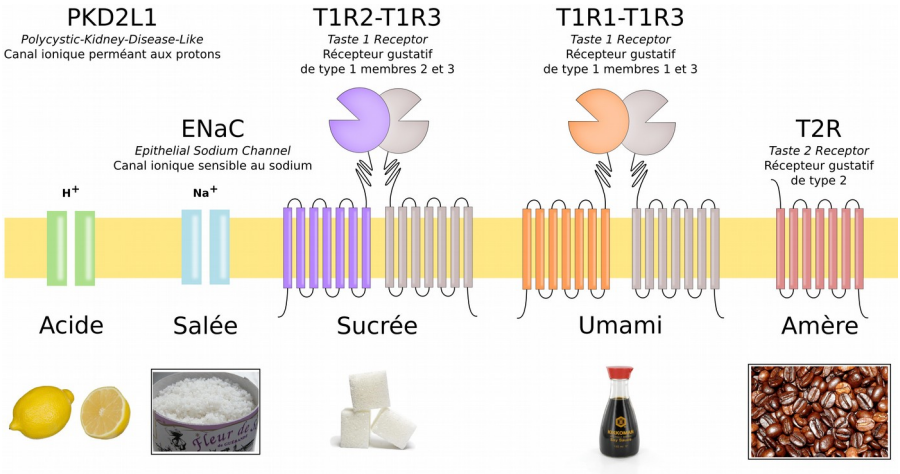


Figure 4.4 *: Schéma représentant les cinq familles de récepteurs gustatifs. Les saveurs acide et salée impliquent des canaux ioniques alors que les saveurs sucrée, umami et amère sont dépendant de l'activation de récepteurs couplés à une protéine G (RCPG).* [17]

[15] *de March et al. J. Am. Chem Soc.,* ***2015****, 137, 8611-8616*
[16] *Yu et al. Proc. Natl. Acad. Sci. USA,* ***2015****, 112, 14966-14971*
[17] *Chéron et al., Act. Chim.* ***2017****, 416, 11-18.*

Intérêt des approches théoriques. A ce jour, aucune structure expérimentale (par cristallographie RX ou spectroscopie RMN) d'un récepteur sensoriel (olfactif ou gustatif) n'a été obtenue. Cependant, plus d'une centaine de structures de RCPG ont été publiés et peuvent servir de support pour la prédiction de la structure des récepteurs sensoriels. Les méthodes de reconstruction par homologie de séquence permettent d'obtenir des modèles 3D en s'appuyant sur le paradigme : similarité de séquence = similarité de structure et donc de fonction. Il est alors possible d'avoir une vision atomique des interactions entre le ligand et le récepteur et d'étudier la dynamique des édifices moléculaires. Si le modèle est validé expérimentalement, il est possible de proposer de nouvelles hypothèses de travail à faire tester (par exemple par mutagenèse dirigée) pour compléter le modèle et pourquoi pas découvrir de nouveaux ligands.

• Collaborations et financements

- Olfaction :

Cette thématique est dirigée par le Prof. J. Golebiowski, notamment porteur du GDR 3713 O^3 (Odorant, Odeur, Olfaction). L'équipe bénéficie d'une solide expertise sur la modélisation des récepteurs olfactifs et a obtenu récemment plusieurs financements nationaux ou internationaux dans lesquels je suis impliqué. Un financement de coopération internationale Fr-USA (ANR-NSF, NeuroComp 2015-2018) pour une collaboration entre notre équipe et les équipes du Prof. H. Matsunami à Duke (Durham, USA) et du Prof. M. Ma à U.Penn (Philadelphie, USA) est en cours sur la compréhension du spectre de reconnaissance des récepteurs olfactifs.

Un autre projet vient de débuter (ANR Demeter 2017-2019) en collaboration avec le Dr. E. Jacquin Joly (INRA, iEEs – Paris Versailles) et le Prof. D. Boujard (UMS CRB Xenope, Rennes). Il porte sur l'identification de nouveaux ligands de récepteurs olfactifs d'insectes (bio-olfacticide) pour « *less pesticides & more food* ».

Enfin, une analyse bioinformatique sur la relation séquence-structure des récepteurs olfactifs est en cours avec le Dr. J. Waldispühl (McGill, Canada). Ce projet a bénéficié d'un « Crédit Scientifique Incitatif » de la part de l'UNS en 2014 puis d'un poste de Professeur Invité à Nice en 2016. McGill a également accordé au Dr. J. Waldispühl un congés sabbatique pour compléter sa venue à Nice en 2016.

- Gustation :

Dans l'équipe, cette thématique de recherche est la plus jeune et n'a bénéficié pour l'instant d'aucun financement majeur en dehors d'un soutien pour « projet innovant » de l'ICN et d'un financement de thèse MESR (Bourse de J.B. Chéron sur la période 2014-2017). En revanche, le doctorant J.B. Chéron a obtenu 4 subventions liées à son sujet de thèse : 2 bourses d'études pour des projets « food science » financées par les organisations Giract (2014) et The Gen Foundation (2016) et 2 bourses de voyage pour participer aux congrès COST Action GLISTEN (2015, Aschwill, CH) et 17th International symposium on Olfaction and Taste – ISOT (2016, Yokohama, JP).

Une étude sur l'identification de nouvelles molécules sucrées d'origine naturelle est en cours en collaboration avec le Dr. L. Briand (CSGA, INRA Dijon). Enfin, une collaboration sur l'étude des relations structure-fonction des récepteurs T1Rs est également en cours avec les Dr. P. Jiang et A. Bachmanov du Monell Chemical Senses Center (Philadelphie, USA).

- **Résultats clés et perspectives**

- Olfaction :

La thèse de Landry Charlier a été le point de départ des études portant sur la description des bases moléculaires de la perception olfactive. Nous avions commencé à décrire au niveau moléculaire les interactions entre odorant, OBP[8-9] (Cf. Chapitre 3) et récepteurs olfactifs[18]. Les récents progrès sur la structure et la dynamique des récepteurs olfactifs ont été réalisé sous la direction du Prof. J. Golebiowski. En voici un cours résumé même si je n'ai pas co-publié ces résultats :

« Une approche combinant biologie moléculaire et modélisation 3D d'un récepteur olfactif a permis de mettre en évidence un ensemble de résidus cruciaux pour la dynamique et la fonction du récepteur.[15-16] Il est également possible de prédire in silico si l'association d'un odorant avec un récepteur est capable de déclencher une réponse de ce dernier.[19] Enfin, une analyse bioinformatique a montré que les acides aminés impliqués dans la reconnaissance d'une molécule étaient relativement variables alors qu'il existait des invariants impliqués dans le mécanisme d'activation du récepteur.[20] »

Même si les résultats des différents projets en cours n'ont pas encore donné lieu à de nouvelles publications, ils ont été présentés à l'occasion de congrès internationaux spécialisés dans les sens chimiques (ECRO[21] et Achems[22]).

Dans un avenir proche, nous souhaitons étudier les relations entre l'espace chimique des molécules odorantes et le répertoire des gènes exprimant les récepteurs olfactifs chez les mammifères. La séquence de tous les récepteurs olfactifs étant par exemple connue chez l'homme, le rat ou la souris, il serait possible d'établir les liens chémo-génomiques de la perception chimio-sensorielle et tenter de répondre à une question cruciale en neuroscience : « *Peut-on prédire si une molécule active ou non un récepteur chimio-sensoriel ?* » A moyen terme, le modèle numérique permettrait d'identifier de nouvelles molécules candidates avant de les valider (ou non) *in vitro*. Dans un cercle vertueux, ces nouvelles données expérimentales permettront d'alimenter le modèle théorique et ainsi améliorer la prédiction. L'objectif à plus long terme est de décrypter le code combinatoire de la perception olfactive. L'impact sociétal et économique peut être très important. On peut, par exemple, citer la conception rationnelle de nouvelles molécules odorantes pour l'industrie des arômes & parfums.

- Gustation :

Les premiers travaux de l'équipe sur la perception gustative ont débuté par l'étude de l'interaction entre les tannins et les protéines peu structurés (financement ANR Protanin 2007) à l'origine de l'astringence des vins.[23] Depuis 2014, mes activités de recherche se focalisent principalement sur l'étude de la perception du goût sucré.

Nous avons commencé par étudier la diversité structurale des sucres naturels et de synthèse pour mettre au point un modèle (QSAR) de relation structure chimique – pouvoir sucrant (Fig. 4.5). Nous avons élargi l'espace chimique connus des molécules sucrés en réalisant un criblage virtuel de plus de 325 000 molécules naturelles.[24] La grande majorité des ~4000 nouvelles

[18] *Golebiowski et al. Chap. 96 in Flavour Sciences, V. Ferreira & R. Lopez, Academic Press (**2013**) p.221-232*
[19] *Topin et al. Chem. Eur J., **2014**, 20, 10227-10230*
[20] *de March et al. Prot. Sci. **2015**, 24(9), 1546-1548*
[21] *Viano et al. Chem. Senses, **2015**, 40, 242*
[22] *Golebiowski et al. Chem. Senses. **2016**, 41(7), E83*
[23] *Golebiowski et al., Mol. Inf., **2011**, 30, 410-414*
[24] *Chéron et al. Food. Chem. **2017**, 221, 1421-1425*

molécules identifiés font partie de la famille des di- ou tri-terpènes (respectivement basées sur le noyau stévioside ou saponine). Trois nouvelles molécules dont la structure est originale vont faire l'objet de tests *in vitro*.
Notre préférence innée pour le sucre et sa surconsommation dans les pays industrialisés sont à l'origine de nombreuses pathologies (diabète, obésité, ...). La découverte de nouveaux édulcorants a` faible pouvoir calorique ou encore de modulateurs du goût aiderait le développement de nouveaux produits alimentaires plus sains. Pour cela, l'identification des interactions ligand-récepteur est primordiale.

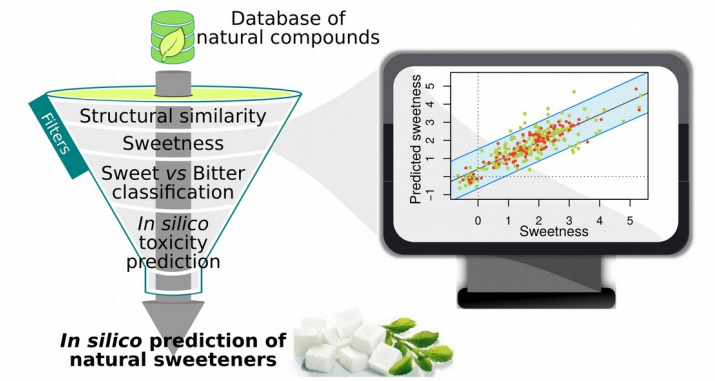


Figure 4.5 : *Modèle QSAR de relation structure – goût sucré (gauche) basé sur la prédiction du pouvoir sucrant (droite)*

La structure du récepteur au goût sucré étant encore inconnue, nous avons réalisé une étude bioinformatique[25-26] récapitulant les données expérimentales connues à ce jour (plus de 20 structures cristallographiques de différents domaines de RCPG classe C et plus de 600 données de mutagénèse dirigée). Cela a permis de fournir un nouveau modèle *in silico* du récepteur par homologie de séquence et d'identifier de nouveaux résidus clés pour l'activation du récepteur (Fig. 4.6). Certaines hypothèses sur le site de liaison de molécules impliqués dans la modulation de la réponse du récepteur sont en cours d'étude dans l'équipe du Dr. P. Jiang au Monell (Philadelphie, USA).

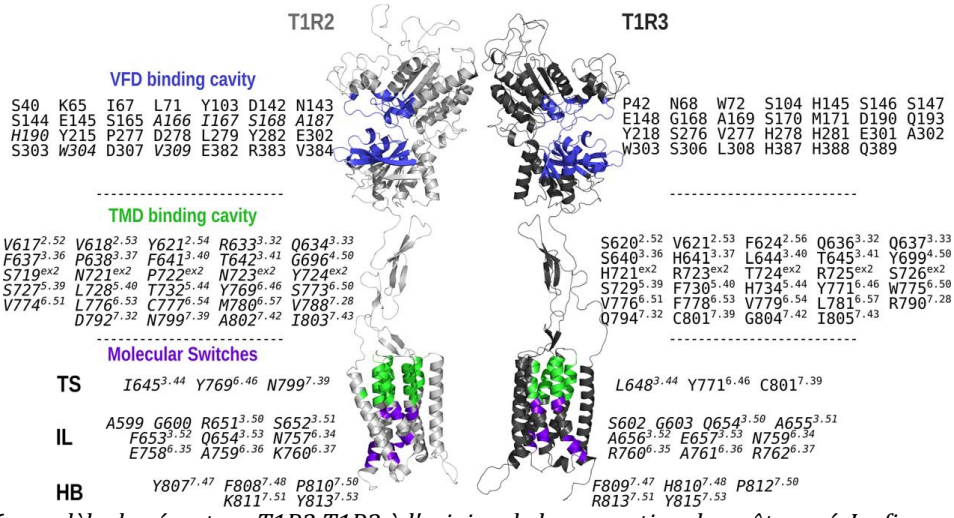


Figure 4.6 : *modèle du récepteur T1R2-T1R3 à l'origine de la perception du goût sucré. La figure récapitule les résidus clés impliqués dans la reconnaissance d'un ligand dans le site orthostérique (site VFD) et/ou allostérique (site TMD) ainsi que les résidus potentiellement impliqués dans le mécanisme d'activation du RCPG (molecular switches). Les résidus en italiques correspondent à des résidus identifiés sur la base de l'alignement de séquence des RCPG de classe C et de la structure du récepteur.*

[25] *Chéron et al. Proteins.* ***2017****, 85, 332-341*
[26] *Chéron et al. Chem. Senses,* ***2016****, 41(9), E179*

L'étude de la dynamique des interactions ligands-récepteurs dans les différents sites de liaison et du mécanisme d'activation du récepteur est actuellement en cours. Nous nous intéressons plus particulièrement aux résidus impliqués dans l'interface T1R2-T1R3 qui seraient responsables du transfert de l'information d'un monomère à l'autre.

Détailler le mécanisme d'activation d'un RCPG de classe C au niveau moléculaire aurait un impact direct sur la compréhension de la pharmacologie de toute cette famille de récepteur. Les RCPG de classe C regroupe notamment les récepteurs métabotropiques au glutamate (impliqués dans la régulation pré-et post-synaptique de l'activité neuronale) dont le dysfonctionnement est à l'origine de nombreuses pathologie du système nerveux central.

2. Biosynthèse des substances naturelles - *thématique secondaire* -

- **Contexte scientifique**

Il a été récemment proposé que dans une voie métabolique (succession de réactions chimiques catalysées par des enzymes au sein d'une même cellule permettant la transformation d'un métabolite), les enzymes n'agiraient pas de façon indépendante mais coordonnée à travers la formation de complexes multi-enzymatique appelé métabolon. Leur rôle premier est la régulation et l'optimisation des flux métaboliques, notamment par un phénomène appelé « substrate channeling ».[27] Dans une voie de biosynthèse, le produit d'une enzyme devient le réactif de la prochaine enzyme et la diffusion du substrat, dans un métabolon, est optimisée par la proximité des sites catalytiques. Cependant très peu d'études se focalisent sur le rôle exact de ces super-structures et très peu d'informations sont disponibles au niveau moléculaire. Toute nouvelle étude a donc un impact important dans la communauté scientifique.[28] Dans ce contexte les outils théoriques de modélisation moléculaire peuvent être d'une aide précieuse pour apporter de nouvelles hypothèses de travail tout en ayant une vision moléculaire du phénomène modélisé : Détailler un mécanisme réactionnel à l'aide de méthodes basées sur la mécanique quantique ; observer un processus dynamique à l'aide des méthodes de dynamique moléculaire ; prédire la formation de complexes biomoléculaires via les approches de docking ; etc.

- **Collaborations et financements**

Le laboratoire a une spécificité reconnue sur l'étude des matrices naturelles et plus spécifiquement sur les molécules odorantes (GDR O3[29]). Avec le Prof. Serge Antonczak, leader de la thématique « catalyse enzymatique » dans l'équipe, nous nous intéressons aux voies de biosynthèse des molécules odorantes dans les plantes. Nous collaborons notamment avec le Laboratoire BVPAM de l'Université de St Etienne et avons obtenu un financement PEP's CNRS Exomod en 2014 (renouvelé en 2015) pour l'étude de la spécificité d'une famille d'enzymes (Nudix hydrolase) chez la rose.

Nous nous intéressons également a une autre famille de métabolites secondaires, les flavonoïdes. L'un des objectifs de la thèse de Julien Diharce, que j'ai co-dirigé avec le Prof. S. Antonczak, et qui a été soutenue en 2015, était d'avoir une meilleure compréhension des phénomènes moléculaires impliqués lors d'un « substrate channeling ». Un des défis majeurs était la reconstruction *in silico* d'un modèle de métabolon. Un projet PHC (appel AURORA 2013) sur

[27] *Huang et al. Annu Rev Biochem.* ***2001****, 70, 149-180*
[28] *Laursen et al. Science,* ***2016****, 354, 890-893*
[29] *GDR « Odorant Odeur Olfaction » soutenu pour la période 2015-2018* https://gdro3.wordpress.com/

l'étude de la structure d'un métabolon fixé à une membrane a été déposé et soutenu en 2013 pour financer les échanges entre Nice et le laboratoire du Prof. Nathalie Reuter à Bergen (Norvège).

- **résultats clés et perspectives**

L'étude des voies de biosynthèse de composés odorants chez la rose est toujours en cours et espérons valoriser ce travail à plus ou moins court terme. Cependant la collaboration entre notre équipe et le LBVPAM a déjà donné lieu à deux publications communes. Un projet pluridisciplinaire faisant intervenir botanique, génie génétique, biologie cellulaire, chimie analytique et modélisation moléculaire a permis d'identifier et de caractériser deux nouvelles terpène synthase chez la lavande (*Lavandula angustifolia*): une cadinol synthase[30] et une bornyl diphosphate synthase[31]. Ce travail a fait la couverture de l'un des volumes de la revue scientifique Plant Molecular Biology en 2014 (Fig. 4.7, gauche). A plus long terme, on peut envisager de faire de l'ingénierie végétale pour optimiser la biosynthèse de telle ou telle molécule odorante par la plante, sachant que l'intérêt économique est potentiellement très important, notamment pour la rose.

Lors de la thèse de Julien Diharce, il a été simulé pour la première fois un phénomène de substrate channeling dans la voie de biosynthèse des flavonoïdes. L'étude théorique d'un complexe multi-enzymatique a permis d'observer, à résolution atomique, la diffusion d'un métabolite entre deux enzymes (Fig. 4.7, droite) et de mettre en évidence le rôle important de la microsolvatation lors du processus de « substrate channeling ».[32] Les outils théoriques développés au sein du laboratoire du Prof. Nathalie Reuter nous ont permis d'obtenir différents modèles d'un complexe multienzymatique enchassé dans une membrane cellulaire. Ces structures sont toujours en cours d'étude au laboratoire.

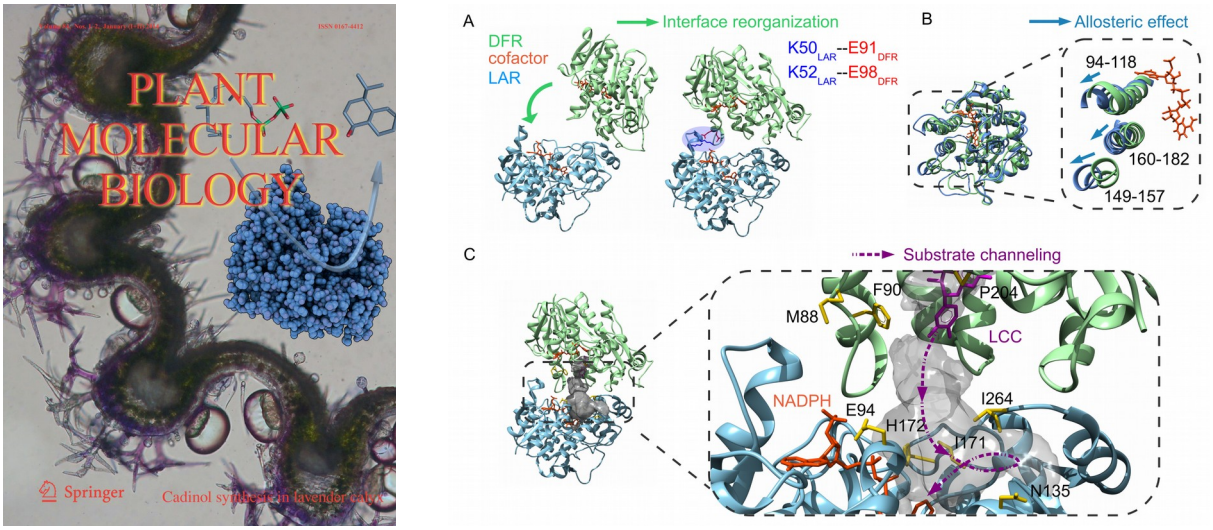


Figure 4.7 *: Gauche :couverture de Plant Molecular Biology illustrant les trichomes de la lavande sécrétant les molécules odorantes et la structure de l'enzyme τ-cadinol synthase. Droite : Formation d'un complexe enzyme-enzyme (A et B) permettant la diffusion d'un flavonoïde par substrate channeling (C).*

[30] *Julien et al. Plant Mol. Biol,* ***2014****, 84, 227-241*
[31] *Despinasse et al. Phytochemistry,* ***2017****, 137, 24-33*
[32] *Diharce et al. Phys. Chem Chem. Phys.* ***2016****, 18, 10337-10345*

3. Prédiction des interactions protéine-protéine *- thématique secondaire -*

- **Contexte scientifique**

Dans le domaine de la biologie moléculaire, la description des assemblages moléculaires est un enjeu majeur.[33] Actuellement il n'est toujours pas possible de déterminer expérimentalement toutes les structures protéine-protéine. La prédiction la plus réaliste possible de ces complexes est alors primordiale pour apprécier leur fonction biologique. Déterminer la structure native requiert alors l'échantillonnage le plus représentatif possible de l'ensemble des possibilités et l'estimation des propriétés thermodynamiques du système.[34] Un édifice moléculaire contenant plus de deux entités entraîne une explosion combinatoire lors de l'étude de toutes les configurations possibles et nécessite le développement de nouveaux outils théoriques. Dans ce contexte il est nécessaire de développer des méthodes basées sur des modèles simplifiés, comme les méthodes d'amarrage moléculaire (*docking*). Un des défis dans la mise en œuvre des méthodes de *docking* protéine-protéine est la prise en compte des changements conformationnels induits par la formation du complexe (flexibilité locale et globale) et la prédiction réaliste des énergies de complexation (fonction de score plus ou moins sophistiquée).

- **Collaborations et financements**

Je travaille en collaboration avec l'équipe de biophysique théorique du Prof. Martin Zacharias, détenteur d'une chaire en dynamique biomoléculaire à la Technische Universität München (TUM), le Dr. Chantal Prévost du Laboratoire de Biochimie Théorique (LBT, UPR 9080 CNRS, entité de l'IBPC à Paris) et le Dr. Pierre Poulain de l'équipe Dynamique des Systèmes et Interactions des Macromolécules Biologiques (DSIMB, UMR_S 1134 INSERM à Paris) sur le développement d'un logiciel de docking ATTRACT[10-12,35,36]. La particularité de ce code est l'utilisation d'un modèle gros-grains pour représenter la protéine et le développement d'un champ de force empirique basé sur la connaissance de structures déjà cristallisées (« knowledge based force-field »).

Dans ce cadre, un projet PHC (appel PROCOPE 2012) sur l'élucidation de la structure de complexes multiprotéiques a été déposé et soutenu en 2012 pour financer les échanges entre Nice et Munich. Un projet ANR (programme blanc ANR-09-BLAN-0113) a été obtenu en 2009 par le laboratoire en collaboration avec l'Institut de Biologie Valrose (IBV, UMR 7277-CNRS, UMR 1091 INSERM) sur le développement de nouveaux vecteurs de transfert de gènes. L'objectif était le design de peptides vecteurs pouvant se lier à la dynéine, protéine impliquée dans le trafic intracellulaire et ces nouveaux outils de docking ont été utilisé dans ce projet.[37]

- **Résultats clés et perspectives**

Un nouveau protocole de prédiction de structures a été testé avec succès ces dernières années lors du challenge international CAPRI, hébergé par l'Institut Européen de Bioinformatique (EBI)[38]. Par ailleurs, le design du code a été pensé en vue d'étendre la prédiction de complexes à

[33] *Protein-protein complexes: Analysis, modeling and drug design, Martin Zacharias, World Scientific (**2010**)*
[34] *Fiorucci et al. Chap. 11. in Protein-protein complexes: Analysis, modeling and drug design, M. Zacharias, World Scientific (**2010**) p.295-317*
[35] *Schneider et al., Chap.15 in Computational Drug Discovery and Design, Ricardo Baron, Springer Protocols (**2012**)*
[36] *Fiorucci et al. Chap. in Immunoinformatics, R.K. De & N. Tomar, Springer (**2014**) p.365-374*
[37] *Parassol et al. Plos One **2013**, 8, e82908*
[38] *Fiorucci et al. Proteins, **2010**, 78, 3131-3139*

plus de deux partenaires (*Fig. 4.8, gauche*). Une des limites aux approches de docking actuelles est qu'elles ne permettent pas d'avoir une estimation précise de l'affinite´ de liaison[39] contrairement aux méthodes théoriques les plus précises nécessitant des moyens de calculs très important. Nous avions rédige´ un article de revue traitant de ce sujet.[34] Nous développons actuellement un nouveau champs de force gros grains qui permettra d'estimer l'enthalpie libre de complexation protéine-protéine de façon extrêmement rapide. Les résultats que nous avons obtenus sont prometteurs.

Du point de vue applicatif, ces nouvelles approches de docking sont en support des deux thématiques du laboratoire présentées plus haut : i) la description des voies de biosynthèse de métabolites secondaires et ii) l'étude des phénomènes moléculaires impliqués dans les sens chimiques. Le premier projet a déjà été détaillé dans le paragraphe précédent (complexes multi-enzymatiques). Dans le deuxième projet, il est question d'étudier les interactions entre les protéines de liaison aux odorants (OBP) et les récepteurs olfactifs (OR). Le rôle de l'OBP en tant que transporteur n'est plus a` remettre en cause mais il a été´ montre´ que l'OBP pourrait également jouer un rôle de modulateur de l'interaction entre l'odorant et le récepteur olfactif.[40] En s'appuyant sur plusieurs indices expérimentaux, un modèle d'interaction entre le récepteur olfactif I7 du rat et l'OBP3 est en cours d'étude[41] (*Fig. 4.8, droite*). Il semblerait que la formation du complexe OBP-OR déstabiliserait le complexe odorant-OBP et entraînerait le relargage de l'odorant à proximité de l'OR, voire directement dans la cavité du récepteur. De nouveaux calculs et surtout de nouvelles données expérimentales sont encore nécessaires pour confirmer cette hypothèse.

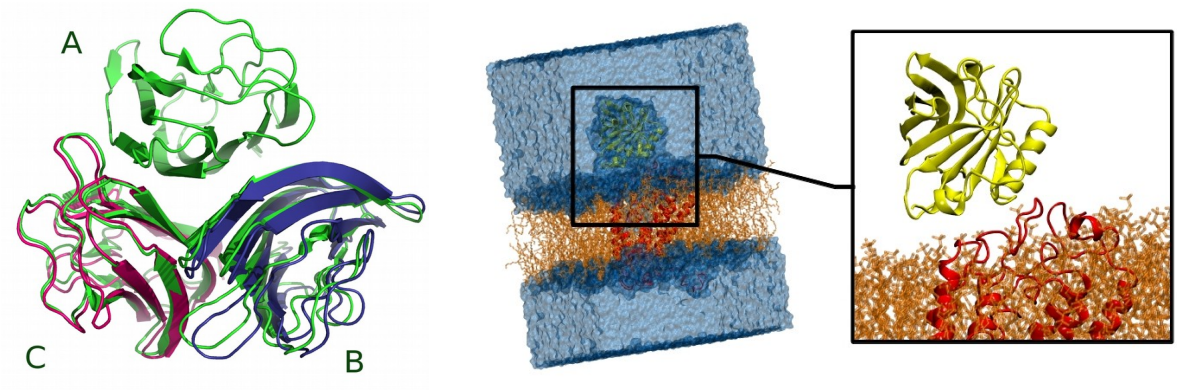


Figure 4.8 : *(Gauche) Résultats d'un docking multi-ligands. La structure expérimentale est en vert, la prédiction des protéines B et C respectivement en bleu et rouge. (Droite) Structure reconstruite du récepteur olfactif I7 (rouge) en interaction avec l'OBP3 (jaune) dans un environnement membranaire (orange) et de molécules de solvant (bleu) explicites.*

Ce thème de recherche n'a pas vocation à devenir prioritaire au sein du laboratoire. La spécificité nationale et internationale que nous avons sur le sujet fait que nous continuons à le supporter, essentiellement à travers les collaborations déjà établies et l'obtention de financement sur des projets plus appliqués reliés aux autres thèmes développés précédemment.

[39] *Kastritis et al. J. Proteome Res.*, ***2010***, *9(5), 2216-2225*
[40] *Vidic et al. Lab. Chip.*, ***2008***, *8, 678-688*
[41] *Viano et al. Chem. Senses,* ***2015****, 40, 242*

Chapitre 5

Conclusion générale

Conclusion générale

L'habilitation à diriger des recherches est une étape importante dans la vie d'un enseignant-chercheur validant son aptitude à élaborer une stratégie de recherche scientifique, sa capacité à collaborer avec d'autres chercheurs, souvent de disciplines différentes, à encadrer de jeunes chercheurs et à obtenir des financements permettant de faire avancer sa recherche. A travers ce mémoire détaillant mon parcours, mes activités de recherches passées et actuelles ainsi que les pistes de recherche que j'envisage pour la suite de ma carrière, je pense avoir démontré mon aptitude à diriger des recherches.

L'évaluation de la carrière d'un chercheur est (trop?) souvent lié à sa production scientifique. En résumant ma carrière en quelques chiffres, il semble que mes activités de recherche soient respectables pour mon âge (une trentaine d'articles, un facteur h de 12, ~460 citations au total, ~50 citations par an) mais il serait extrêmement réducteur de la limiter à la seule analyse bibliométrique. Cette vision chiffrée occulte nos autres activités mais également le parcours scientifique de tout chercheur, jalonné de réussites et d'échecs. Nos changements de postes au cours de notre carrière sont souvent responsables de changements thématiques à l'origine d'un ralentissement du rythme des publications scientifiques. Or la reconnaissance par les pairs passent souvent par cette évaluation chiffrée qui a tendance à encourager la surproduction et parfois à ses effets pervers.

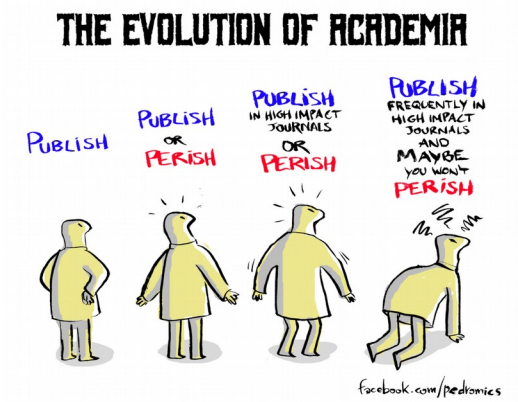


Figure 5.1 : *Evolution de la recherche académique ? (Illustration de Pedro Veliça)*

On peut regretter la liberté d'un chercheur sans contrainte, en quête de révolution scientifique, mais cette pression peut cependant être salutaire. Elle pousse à rationaliser nos efforts et à concentrer nos recherches sur des thématiques connectées au monde réel*. La recherche n'est pas une activité individualiste. Une grande partie de mes travaux a été possible grâce aux nombreux étudiants que j'ai pu encadrer et former et grâce aux échanges fructueux avec des collaborateurs de différentes disciplines. Pour ma part, j'ai eu la chance d'avoir une certaine continuité dans mes travaux de recherches. Cela s'est concrétisée par des publications scientifiques régulières sur l'ensemble des projets auxquels j'ai eu l'occasion de participer.

---

* *« J'ai toujours tâché de vivre dans une tour d'ivoire ; mais une marée de merde en bat les murs, à la faire crouler »* (Gustave Flaubert, *Correspondance*, 1872)

Annexes

1. Bibliographie exhaustive des productions scientifiques

Publications
Logiciel et base de données
Conférences sur invitation
Communications orales
Séminaires sur invitations
Communications par affiche
Formations CNRS et école chercheur
Communications Grand Public

2. Articles représentatifs

- The anatomy of mammalian sweet taste receptors. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Proteins. **2017**, 85, 332-341.*
- Sweetness prediction of natural compounds. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Food. Chem. **2017**, 221, 1421-1425.*
- Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channeling in the course of flavonoid biosynthesis. J. Diharce, J. Golebiowski, S. Fiorucci*, S. Antonczak*. *Phys. Chem. Chem. Phys., **2016** 18, 10337-10345.*
- Isolation and functional characterization of a t-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*. F. Julien*, S. Moja, A. Bony, S. Legrand, C. Petit, T. Benabdelkader, K. Poirot, S. Fiorucci, Y. Guitton, F. Nicole, S. Baudino, J.L. Magnard. *Plant. Mol. Biol. **2014**, 84, 227-241.*
- Prediction/calculation of protein-protein binding affinities and mutation effect, S. Fiorucci*, S. Antonczak, J. Golebiowski. in *Protein-protein complexes: Analysis, modeling and drug design*, M. Zacharias, *World Scientific (**2010**) p.295-317.*

Liste des Publications:

ART / OUV / PRO *précise s'il s'agit d'un* ***article*** *de recherche ou revue, d'un chapitre d'****ouvrage,*** *d'un «* ***conference proceedings »*** *ou actes de conférence (internationale).*
***** *signifie que je suis auteur (co-)correspondant*

30. ***ART*** *:* Ces molécules qui éveillent nos papilles. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *L'Act. Chim.* ***2017****, 416, 11-18.*
29. ***ART*** *:* Bornyldiphosphate synthase from *Lavandula angustifolia* : a major monoterpene synthase involved in essential oil quality. Y. Despinasse, S. Fiorucci, S. Antonczak, S. Moja, A. Bony, F. Nicolè, S. Baudino, J.L. Magnard, F. Julien. *Phytochemistry*, ***2017****, 137, 24-33.* ***(IF 2015 = 2.8)***
28. ***ART*** *:* The anatomy of mammalian sweet taste receptors. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Proteins.* ***2017****, 85, 332-341.* ***(IF 2015 = 2.5)***
27. ***ART*** *:* Sweetness prediction of natural compounds. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Food. Chem.* ***2017****, 221, 1421-1425.* ***(IF 2015 = 4.1)***
26. ***PRO*** *:* The anatomy of mammalian sweet taste receptors. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*. *Chem Senses.* ***2016****, E179-E179.* ***(IF 2015 = 2.5)***
25. ***PRO*** *:* Structure-function of Odorant and Sweet-taste Recept.ors Based on Molecular Modeling. J. Golebiowski, J.B. Chéron, C.A. De March, S. Fiorucci, S. Antonczak. *Chem. Senses.* ***2016****, 41(7),E83-E83.* ***(IF 2015 = 2.5)***
24. ***ART*** *:* Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channeling in the course of flavonoid biosynthesis. J. Diharce, J. Golebiowski, S. Fiorucci*, S. Antonczak. *Phys. Chem. Chem. Phys.,* ***2016*** *18, 10337-10345.* ***(IF 2015 = 4.4)***
23. ***PRO*** *:* Deciphering the odorant binding protein-olfactory receptor interactions. M. Viano, S.K. Kim, W.A. Goddard III, C.A. de March, J. Golebiowski, S. Fiorucci*. *Chem. Senses,* ***2015****, 40, 242.* ***(IF 2015 = 2.5)***
22. ***ART*** *:* Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. I. Kufareva, V. Katritch, Participants of GPCR Dock 2013[#],R.C. Stevens, R. Abagyan. *Structure.* ***2014****, 22, 1120-1139.* ***(IF 2014 = 5.6)***
    [#] *Among them: UNICE group involving C.A. De March, J. Diharce, S. Antonczak, J. Golebiowski, S. Fiorucci*
21. ***OUV*** *:* Computational antigenic epitope prediction by calculating electrostatic desolvation penalties of protein surfaces, S. Fiorucci*, M. Zacharias. in *Immunoinformatics,* R.K. De & N. Tomar*, Springer* (***2014***) *p.365-374.*
20. ***ART*** *:* Isolation and functional characterization of a t-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*. F. Julien, S. Moja, A. Bony, S. Legrand, C. Petit, T. Benabdelkader, K. Poirot, S. Fiorucci, Y. Guitton, F. Nicole, S. Baudino, J.L. Magnard. *Plant. Mol. Biol.* ***2014****, 84, 227-241.* ***(IF 2014 = 4.3)***
19. ***ART*** *:* O_2 Migration Rates in [NiFe] Hydrogenases. A Joint Approach Combining Free-Energy Calculations and Kinetic Modeling J. Topin, J. Diharce, S. Fiorucci, S. Antonczak, J. Golebiowski. *J. Phys. Chem. B,* ***2014****, 118, 676-681.* ***(IF 2014 = 3.3)***
18. ***ART*** *:* In vivo characterization of dynein-driven nanovectors using Drosophila oocytes.N. Parassol, C. Bienvenu, C. Boglio, S. Fiorucci, D. Cerezo, X.M. Yu, G. Godeau, J. Greiner, P. Vierling, S. Noselli, C. Di Giorgio, V. Van De Bor. *Plos One* ***2013****, 8, e82908.* ***(IF 2013 = 3.5)***
17. ***OUV*** *:* Molecular features underlying the chemoreception of odorant binding proteins and olfactory receptors. Insights from molecular modeling and biophysical data. J. Golebiowski, L. Charlier, J. Topin, S. Fiorucci, S. Antonczak. Chap. 96 in *Flavour Sciences,* V. Ferreira & R. Lopez*, Academic Press (****2013****) p.221-232.*
16. ***OUV*** *:* ATTRACT and PTools: Open sources programs for protein-protein docking, S. Schneider, A. Saladin, S. Fiorucci, C. Prévost, M. Zacharias. in *Computational Drug Discovery and Design* , Ricardo Baron, *Springer* (***2012***) *p.221-232.*

15. ***ART*** *:* Molecular features underlying perception of astringency as probed by molecular modeling. J. Golebiowski, S. Fiorucci, M. Adrian-Scotto, J. Fernandez-Carmona, S. Antonczak, *Mol. Inf.,* ***2011****, 30, 410-414.* ***(IF 2011 = 2.4)***
14. ***OUV*** *:* Prediction/calculation of protein-protein binding affinities and mutation effect, S. Fiorucci*, S. Antonczak, J. Golebiowski. in *Protein-protein complexes: Analysis, modeling and drug design*, M. Zacharias, *World Scientific (****2010****) p.295-317.*
13. ***ART*** *:* How to exploit antigenic diversity for vaccine design: the Chlamydia ArtJ paradigm.M. Soriani, P. Petit, R. Grifantini, R. Petracca, G. Gancitano, E. Frigimelica, F. Nardelli, C. Garcia, S. Spinelli, G. Scarabelli, S. Fiorucci, R. Affentranger, M. Ferrer-Navarro, M. Zacharias, G. Colombo, L. Vuillard, X. Daura and G. Grandi, *J. Biol. Chem.,* ***2010****, 285, 30126-20138.* ***(IF 2010 = 5.3)***
12. ***ART*** *:* Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT. S. Fiorucci, M. Zacharias, *Proteins*, ***2010****, 78(15), 3131-3139* ***(IF 2010 = 2.8)***
11. ***ART*** *:* Prediction of protein-protein interaction sites using electrostatic desolvation profile.
S. Fiorucci, M. Zacharias, *Biophys. J.,* ***2010****, 98(9), 1921-1930* ***(IF 2010 = 4.2)***
10. ***ART*** *:* PTools: an opensource molecular docking library. A. Saladin, S. Fiorucci, P. Poulain, C. Prevost, M. Zacharias. *BMC Struct. Biol.,* ***2009****, 9:27.* ***(IF 2009 = 2.8)***
9. ***ART*** *:* Theoretical Investigations of the Role Played by Quercetinase Enzymes upon Flavonoids Oxygenolysis Mechanism. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, *Phys Chem Chem Phys*, ***2009****, 11, 1491-1501.* ***(IF 2009 = 4.1)***
8. ***PRO*** *:* Enzymatic Mechanisms involving Flavonoïds : a Theoretical Point of View. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. Proceedings of the XXIV International Congress on Polyphenol*, Salamanques (Espagne),* 8-11 Juillet ***2008***.
7. ***ART*** *:* Molecular simulations enlighten the binding mode of quercetin to lipoxygenase-3. S. Fiorucci, J. Golebiowski, S. Antonczak, *Proteins*, ***2008****, 73, 290-298.* ***(IF 2008 = 3.4)***
6. ***ART*** *:* Binding free energy prediction in strongly hydrophobic biomolecular systems. L. Charlier, C. Nespoulous, S. Fiorucci, S. Antonczak, J. Golebiowski, *Phys. Chem. Chem. Phys.*, ***2007****, 9, 5761-5771.****(IF 2007 = 3.3)***
5. ***ART*** *:* Mechanistic events underlying Odorant Binding Protein chemoreception. J. Golebiowski, S. Antonczak, S. Fiorucci, D. Cabrol-Bass, *Proteins,* ***2007,*** *67, 448-458.* ***(IF 2007 = 3.4)***
4. ***ART*** *:* DFT Study of quercetin activated forms involved in antioxidant and prooxidant biological processes. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak, *J. Agr. Food Chem.****, 2007,*** *55(3), 903-911****. (IF 2007 = 2.5)***
3. ***ART*** *:* Molecular Simulations bring new insights into Flavonoid/Quercetinase Interaction Modes. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak, *Proteins,* ***2007,*** *67, 961-970.* ***(IF 2007 = 3.4)***
2. ***ART*** *:* Molecular simulations reveal a new entry site in Quercetin 2,3-Dioxygenase. A pathway for Dioxygen? S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak, *Proteins,* ***2006****, 64, 845-850****(IF 2006 = 3.7)***
1. ***ART*** *:* Oxygenolysis of flavonoid compounds. DFT description of the mechanism for the quercetin case. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak, *Chemphyschem,* ***2004****, 5, 1726-1733.****(IF 2006 = 3.6)***

Logiciel et base de données :

- Développement d'un logiciel de docking protéine-protéine: Ptools/Attract.
  Citations: *A. Saladin et al. BMC struct. Biol (****2009****) + Schneider et al. Meth Mol Biol (****2012****)*
  Téléchargeable sur Github ou sur http://chemosim.unice.fr
- Base de données de composés sucrés : SweetenersDB
  Citation : *J.B. Chéron et al. Food Chem. (****2016****)*
  Téléchargeable sur http://chemosim.unice.fr

Conférences sur invitation :

INT / NAT *précise la portée internationale ou nationale de la conférence ; l'intervenant est souligné.*

4. ***INT :*** Modelling of a metabolon involved in flavonoif biosynthesis : theoretical strategies. J. Diharce, S. Fiorucci, J. Golebiowski, S. Antonczak. 7èˢ Journées Franco-Italiennes de Chimie. 7Th JFIC-GCIC, Palazzo del Rettorato, 5-6 mai ***2014***.
3. ***NAT :*** Contribution de la modélisation moléculaire à la compréhension de la perception olfactive. S. Fiorucci. Journée interdisciplinaire Thématique Odorant-Odeurs-Olfaction, Université de Nice-Sophia Antipolis, 4 juillet ***2013***
2. ***INT :*** Multiresolution approaches for modeling biomolecular complexes. S. Fiorucci. Computational Challenges in Structural Biology, ESBS, Strasbourg, 14-15 novembre ***2012***
1. ***INT :*** Prediction of protein-protein interface using electrostatic profile. S. Fiorucci. Flexibility and Biological Recognition: from Biophysics to Data Models, INRIA Sophia Antipolis Méditerranée, 18-20 mars **2009**

Communications orales :

INT / NAT *précise la portée internationale ou nationale de la communication ; l'intervenant est souligné.*

17. ***INT*** *:* Sweet taste receptor: from sequence to structure. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. JFIC2016. *Avignon (France),* 25-26 avril ***2016***.
16. ***NAT*** *:* Étude de relations structure-activité du récepteur au goût sucré. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. SFCi2015. *Nice (France),* 8-9 octobre ***2015***. (**Prix de la meilleure communication orale**)
15. ***NAT*** *:* Description théorique d'un phénomène de “substrate channeling” dans la biosynthèse de flavonoïdes. J. Diharce, J. Golebiowski, S. Fiorucci, S. Antonczak. GGMM & GT Enzymes 2015. *Sète (France),* 25-28 mai ***2015***.
14. ***INT****:* Modelling of a metabolon involved in Flavonoid biosynthesis. J. Diharce, J. Golebiowski, S. Fiorucci, S. Antonczak. 27th International Conference on Polyphenols. *Nagoya* (Japon), 2-6 septembre ***2014***
13. ***NAT****:* Biosynthèse des flavonoïdes : une étude QM/MM MD de l'étape impliquant l'enzyme DFR. J. Diharce, J. Golebiowski, S. Fiorucci, S. Antonczak. 14ᵉ journée francophones des jeunes physico-chimistes. *Fréjus*, 14-18 octobre ***2013***.
12. ***INT****:* Modelling of a metabolon involved in flavonoid biosynthesis. J. Diharce, S. Fiorucci, S. Antonczak. 7ᵗʰ international Workshop on antocyanins. Porto *(Portugal)*, 9-11 septembre ***2013***.
11. ***INT****:* Theoretical strategies for the modelling of a metabolon involved in flavonoid biosynthesis. J. Diharce, S. Fiorucci, S. Antonczak. 6ᵗʰ Theoretical Biophysics Symposium (Theobio2013). *Gothenburg (Suède)*, 24-26 juin ***2013***.
10. ***NAT****:* Mécanismes enzymatiques de la production des anthocyanes: une étude par dynamique moléculaire QM/MM de l'étape impliquant DFR. J. Diharce, S. Fiorucci, I. Tunon, S. Antonczak. 18ᵉ congrès GGMM. *Saint Pierre d'Oléron*, 21-23 mai ***2013***.
9. ***NAT :*** Experimental and theoretical charaterization fo macromolecular assemblies at low resolution. S. Fiorucci. Discussion meeting CFCAM, *IBPC Paris*, 8-9 april ***2013***
8. ***INT :*** Modeling ligand-receptor interactions at the molecular level. Molecular dynamics used as a computational microscope to decipher the sense of smell. J. Golebiowski, L. Charlier, J. Topin, S. Fiorucci, S. Antonczak. 1ᵉʳ Congrès International ISI, *Meknès (Maroc),* 1-5 juin ***2011***
7. ***NAT :*** Enseigner la Chimie théorique en Licence ou comment profiter des ressources numériques et computationelles en la matière. J. Golebiowski, S. Fiorucci, S. Antonczak. 27ᵉ JIREC, 14ᵉ MIEC, *Paris (France)*, 24-27 mai ***2011***.

6. ***INT :*** Electrostatic profiles of targets NMB0033 and CT043. S. Fiorucci, M. Zacharias. Bacabs EU project meeting, *Milan (Italie)*, 10 et 11 Mars ***2008***.
5. ***INT :*** Antigen/Antibody docking improvment including hotspot and solvation energy. S. Fiorucci, M. Zacharias. Bacabs EU project meeting, *Barcelone(Espagne)*, 18 Janvier ***2008.***
4. ***NAT :*** Le mécanisme d'oxygénolyse de flavonoïdes : une synergie entre enzyme et substrat. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 10ème Rencontre des chimistes théoriciens Francophones, *Nancy (France),* 10-13 Juillet ***2006****.*
3. ***NAT :*** Un nouveau site d'entrée dans l'enzyme Quercetin 2,3-Dioxygénase révélé par simulations de dynamique moléculaire : un canal pour le dioxygène? S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. 3ème rencontre du Club Métalloprotéines et Modèles, *Carry-Le-Rouet (France),* 2-5 Octobre ***2005***.
2. ***INT :*** Oxygenolysis of flavonoid compounds. Theoretical mechanism description for the quercetin case. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. 5th European Conference on Computational Chemistry (EuCo-CC5), *La Londe les Maures (France),* 15-20 Juin ***2004***.
1. ***INT :*** *Oxygénolyse de flavonoïdes : Etudes DFT du cas de la quercétine. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 7ème journée Francophone des Jeunes Physico-Chimistes, Monastir (Tunisie)* 19-21 mars ***2004****.*

Séminaires (sur invitation) :

INT / NAT *précise la portée internationale ou nationale de l'invitation ; l'intervenant est souligné.*

6. ***NAT****:* Prédiction de la structure de complexes protéine-protéine. S. Fiorucci. Séminaire PIRAMID organisé par le CEISAM (UMR 6230 CNRS), Nantes, 05 juillet ***2016***.
5. ***INT*** *:* Molecular modeling of the sweet taste receptor. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. *UFIP UMR 6286 CNRS, Nantes,* 14-15 juin ***2016***
4. ***NAT*** *:* Apport de la modélisation moléculaire à l'étude des sens chimiques: le récepteur au goût sucré. S. Fiorucci. *CSGA (UMR INRA, CNRS & Univ. de Bourgogne), Dijon,* 28 janvier ***2016***
3. ***NAT :*** PTools: an opensource molecular docking library. S. Fiorucci. Séminaire du groupe modmol de l'Université de Nice-Sophia Antipolis, *IMPC (UMR 7275), Sophia Antipolis,* 20 octobre ***2009***
2. ***NAT :*** Mécanismes antioxydants et antiradicalaires impliquant des flavonoïdes. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. réseau INPL « Fonctionnalisation de molécules pour des applications biologiques », séminaires de modélisation moléculaire. *INPL, Nancy,* 22 mai ***2006***.
1. ***NAT*** *:* Etude DFT du Mécanisme d'Oxydation de la Quercétine par le Dioxygène. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass.. *ENSSPICAM (UMR CNRS 5516), Faculté des Sciences de St Jérôme, Marseille,* 25 septembre ***2003***.

Communications par affiche :

INT / NAT *précise la portée internationale ou nationale de la communication ; l'intervenant est souligné.*

28. ***INT****:* Structure of the human sweet taste receptor. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. 17th International Symposium on Olfaction and Taste (ISOT 2016), *Yokohama (Japon), 05-09 juin* ***2016****.*
27. ***INT****:* Structural anatomy of class C sweet taste receptor. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. CECAM meeting*, Juelich (Allemagne), 11 septembre* ***2015****.*
26. ***NAT****:* Molecular modelling of class C sweet taste receptor. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. GGMM*, Sète (France), 25-27 mai* ***2015****.*

25. ***NAT**:* Etude par modélisation moléculaire du récepteur impliqué dans la perception de la saveur sucré. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci., journée SCF PACA, Toulon (France)*, 22 avril **2015**.*
24. ***INT**:* Molecular study of sweet taste receptor. J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci. COST GLISTEN meeting*, Allschwill (Suisse), 01-02 avril **2015**.*
23. ***INT**:* Deciphering the odorant binding protein – olfactory receptor interactions M. Viano, S.K. Kim, W.A. Goddard III, C. de March, J. Golebiowski, S. Fiorucci, XXIVth International Conference of European Chemoreception Research Organization, *Dijon (France), 10-13 septembre **2014**.*
22. ***INT**:* How odorant binding proteins interact with olfactory receptors. M. Viano, K. Soo-Kyung , W.A.Goddard III, C. de March, J. Golebiowski, S. Fiorucci. 7Th JFIC-GIFC, *Turin (Italie) 5-6 mai **2014**.*
21. ***INT:*** Etude théorique d'enzymes impliqués dans la biosynthèse des flavonïdes. J. Diharce, S. Fiorucci, J. Golebiowski, S. Antonczak. 13ème rencontre des chimistes théoriciens francophones, *Marseille (France), 1er-5 juillet **2012**.*
20. ***INT:*** Theoretical Study of a keystep in the natural biosynthesis of Flavonoids. J. Diharce, J. Golebiowski, S. Fiorucci, S. Antonczak. 6ème journée franco-italiennes de chimie, *Marseille (France), 16-17 avril **2012**.*
19. ***NAT:*** Etude par modélisation moléculaire de la biosynthèse des flavonoïdes : l'exemple de la dihydroflavonol-4-reductase (DFR). J. Diharce, J. Golebiowski, S. Fiorucci, S. Antonczak. 5ème Colloque de la Chimie Azuréenne, *Nice(France), 21 octobre **2011**.*
18. ***NAT:*** Etude théorique de la biosynthèse des flavonoïdes : l'étape de transformation par la dihydroflavonol-reductase (DFR). J. Diharce, J. Topin, S. Fiorucci, J. Golebiowski, S. Antonczak. 22ème journée de la chimie de la Société Chimique de France PACA, *Toulon(France), 11 Mars **2011.***
17. ***NAT :*** Nouvelles approches de docking pour la compréhension des mécanismes de l'olfaction. S. Fiorucci, A. Saladin, P. Poulain, C. Prévost, M. Zacharias, N. Nafati, J. Golebiowski, S. Antonczak. 17e GGMM, *La Rochelle (France), 30 mai – 1er juin **2011.***
16. ***INT :*** Développement de nouvelles approches pour la prédiction des interactions protéine-protéine. S. Fiorucci, M. Zacharias. 12ème rencontre des chimistes théoriciens francophones, *Namur (Belgique), Juillet **2010**.*
15. ***INT :*** Ptools, an opensource molecular docking library, P. Poulain, A. Saladin, S. Fiorucci, M. Zacharias, C. Prévost. 9th European Conference on Computational Biology, *Ghent (Belgique), 26-29 Septembre **2010**.*
14. ***INT :*** Electrostatic profiles to predict protein-protein interaction sites. S. Fiorucci, M. Zacharias. CAPRI 2009 4th evaluation meeting, *Barcelone(Espagne), 9-11 décembre **2009**.*
13. ***NAT :*** Nouvelles approches de docking. S. Fiorucci, C. Prévost, M. Zacharias. 4ème colloque de la chimie azuréenne, *Nice(France), octobre **2009**.*
12. ***INT :*** Mécanisme enzymatique de l'oxygénolyse de flavonoïdes par la Qercetin 2,3-Dioxygenase. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 11ème Rencontre des chimistes théoriciens Francophones, *Dinard(France), juillet **2008**.*
11. ***INT :*** Inhibition de l'activité oxydante de la lipoxygénase par des flavonoïdes. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 10ème Rencontre des chimistes théoriciens Francophones, *Nancy (France), 10-13 juillet **2006**.*
10. ***INT :*** Le mécanisme d'oxygénolyse de flavonoïdes : une synergie entre enzyme et substrat. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 10ème Rencontre des chimistes théoriciens Francophones, *Nancy (France), 10-13 juillet **2006**.*
9. ***NAT :*** Un nouveau site d'entrée dans l'enzyme Quercétine 2,3-Dioxygenase: un canal à dioxygène? S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 3ème Colloque de la chimie azuréenne, *Nice(France) le 21 octobre **2005***.
8. ***NAT :*** Réaction d'Oxygénolyse de Flavonoïdes Catalysée par l'enzyme Quercetin 2,3-Dioxygénase. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. 3ème rencontre du Club Métalloprotéines et Modèles, *Carry-Le-Rouet (France), 2-5 octobre **2005***.

7. ***INT :*** Insights in the oxygenolysis mechanism of flavonoids by Quercetin 2,3-Dioxygenase. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 15th IUPAB, 5th EBSA International Biophysics Congress, *Montpellier 27 août – 1er septembre* ***2005***.
6. ***NAT :*** Oxygénolyse de la quercétine au sein de l'enzyme Quercétine 2,3-Dioxygénase. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. XVIIIème Journée de la Chimie, *SFC PACA, Toulon(France) le 7 avril* ***2005***.
5. ***INT :*** Aspects structuraux, électroniques et énergétiques des propriétés antioxydantes de la quercétine. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 9ème Rencontre des chimistes théoriciens Francophones, *Pau (France) 20-24 septembre* ***2004***.
4. ***INT :*** Comportement dynamique de la flavonoïde quercétine au sein de l'enzyme 2,3QD. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. 9ème Rencontre des chimistes théoriciens Francophones, *Pau (France) 20-24 septembre* ***2004***.
3. ***NAT :*** Etude théorique de la métabolisation du dioxygène par la quercétine : Rôle du cuivre et compétition entre deux mécanismes réactionnels. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. Colloque de la chimie azuréenne, *Nice (France) le 24 octobre* ***2003***.
2. ***NAT :*** Etude théorique de l'interaction d'une flavonoïde avec un modèle de site actif de la Quercetin 2,3-Dioxygenase. S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak. XVIème Journée de la Chimie, *SFC PACA, Marseille (France) le 3 avril* ***2003***.
1. ***INT :*** Theoretical study of the mechanism of molecular oxygen metabolisation by the quercetine flavonoid. S. Antonczak, S. Fiorucci, J. Golebiowski, D. Cabrol-Bass. Modelling chemical reactivity : from gas-phase to solution and enzymes. *Nancy (France) , France, July 16-18,* ***2003***.

Formations permanentes du CNRS :

L'ensemble de ces formations a eu lieu à la Fédération de Chimie moléculaire de Marseille

- VIIème journée du groupe de chimie théorique et modélisation moléculaire, 31 mars - 1er avril **2005**. Thème de la journée*:* «Macromolécule: quelques aspects théoriques».
- VIème journée du groupe de chimie théorique et modélisation moléculaire, les 25-26 mars **2004**. Thème de la journée*:* «Dynamique moléculaire».
- Vème journée du groupe de chimie théorique et modélisation moléculaire, les 26 et 27 mars **2003**. Thème de la journée: «Réactivité chimique : expérience et aspect théorique».

Ecoles Chercheurs de l'INRA :

- IIIème école Modmol organisée par le département CEPIA. *Bois d'Amont,* les 11-14 décembre **2006**. Thème de la journée *:* Modélisations moléculaire & supramoléculaire (méthode/applications/projets).

Communications Gd Public/Scolaire :

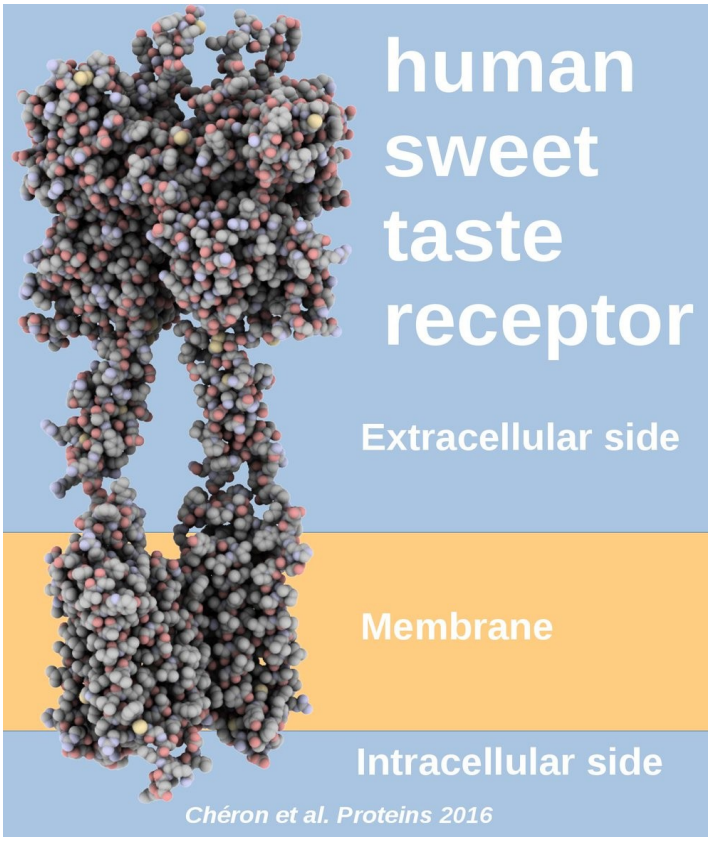
- Conférence Grand Public, « cinq saveurs, sept milliards de goûteurs », semaine du cerveau **2017**, Grasse.
- Conférence Grand Public, « de la molécule au goût », présenté par J.B. Chéron, semaine du cerveau **2016**, Grasse.
- Parrainage d'élèves 6e-5e, action Univercités sur un projet autour du « sens du goût » organisé par l'association « les petits débrouillards », **2014**, Nice.
- Conférence, Scolaire, « quand la chimie a du nez », fête de la science **2012**, Grasse.
- Rencontre Grand Public « les métiers de la chimie », fête de la science **2011**, Grasse.
- Film de vulgarisation « Au cœur de la matière », fête de la science **2011**, Nice.
- Conférence, Gd Public « Sherlock Holmes et la chimie », fête de la science **2011**, Grasse.

Article 1 :

The anatomy of mammalian sweet taste receptors.

J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*.

Proteins. ***2017****, 85, 332-341.*



PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The anatomy of mammalian sweet taste receptors

Jean-Baptiste Chéron,[1] Jérôme Golebiowski,[1,2] Serge Antonczak,[1] and Sébastien Fiorucci[1]*

[1] Université Côte d'azur, CNRS, Institut de Chimie de Nice UMR7272, 06108 Nice, France

[2] Department of Brain and Cognitive Science, DGIST (Daegu Gyeongbuk Institute of Science & Technology), Daegu, Korea

ABSTRACT

All sweet-tasting compounds are detected by a single G-protein coupled receptor (GPCR), the heterodimer T1R2-T1R3, for which no experimental structure is available. The sweet taste receptor is a class C GPCR, and the recently published crystallographic structures of metabotropic glutamate receptor (mGluR) 1 and 5 provide a significant step forward for understanding structure-function relationships within this family. In this article, we recapitulate more than 600 single point site-directed mutations and available structural data to obtain a critical alignment of the sweet taste receptor sequences with respect to other class C GPCRs. Using this alignment, a homology 3D-model of the human sweet taste receptor is built and analyzed to dissect out the role of key residues involved in ligand binding and those responsible for receptor activation.

Proteins 2016; 00:000–000.

Key words: chemical senses; sweet taste; GPCR; T1R; class C; structure–function relationships.

INTRODUCTION

Our innate preference for sweet foods acts as a major determinant for the over-consumption of sugar in industrial countries, which is considered a significant public health problem. Therefore, understanding the molecular basis underlying sweet taste perception is fundamental for identifying new low-calorie sweeteners or sweet taste modifiers to develop healthier food products with reduced amounts of sugar. We perceive our chemical environment via the activation of cellular receptors that trigger downstream signaling cascades. In vertebrates, G protein-coupled receptors (GPCRs), which mediate sweet, bitter, and umami tastes as well as vision and olfaction, constitute the largest and most versatile family of membrane-bound receptors.[1,2] In response to external stimuli, the receptor undergoes a series of structural rearrangements that results in signal transduction across the plasma membrane, with further propagation inside the cell and up to the brain.[3] The GPCR superfamily is divided into five classes based on sequence homology and functional similarity.[4] To date, a hundred X-ray structures of GPCRs have been solved, most of them being class A (Rhodopsin-like) GPCRs (regularly updated GPCR structures can be found in the GPCRDB[5] website at http://gpcrdb.org/structure/). As the human genome contains >800 functional GPCRs[4] and as GPCR expression is still a challenge,[6] a large number of GPCR structures remain unknown.

Sweet and umami taste receptors[7–9] (TAS1R or T1R) are expressed in oral but also in extraoral tissues[10] and belong to the class C GPCR family, which also includes eight metabotropic glutamate receptors (mGluR), two γ-aminobutyric acid type B receptors ($GABA_B$), the calcium sensing receptor (CasR), the class C-group 6-subtype A GPCR (GPRC6A) and seven so-called "orphan" receptors, for which no ligand has been identified.[11] Class C GPCRs share a similar structure comprised of a large extracellular domain called the Venus flytrap domain (VFD), which holds the orthosteric binding site, and by a seven-helix transmembrane domain (TMD), both of which are connected by a cysteine-rich domain (CRD).

The sweet taste receptor is a heterodimer comprised of T1R2 and T1R3 monomers (Fig. 1) that present different functional role.[12,13] Though only one receptor is devoted to sweet tastes, multiple ligand binding sites have been reported, including both T1R2 and T1R3 VFDs and the T1R3 TMD and CRD (Fig. 1). The sweet taste receptor has

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Sébastien Fiorucci; Université Côte d'azur, CNRS, Institut de Chimie de Nice UMR7272, 06108 Nice, France. E-mail: Sebastien.Fiorucci@unice.fr

Received 31 May 2016; Revised 9 September 2016; Accepted 30 November 2016
Published online 00 Month 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.25228

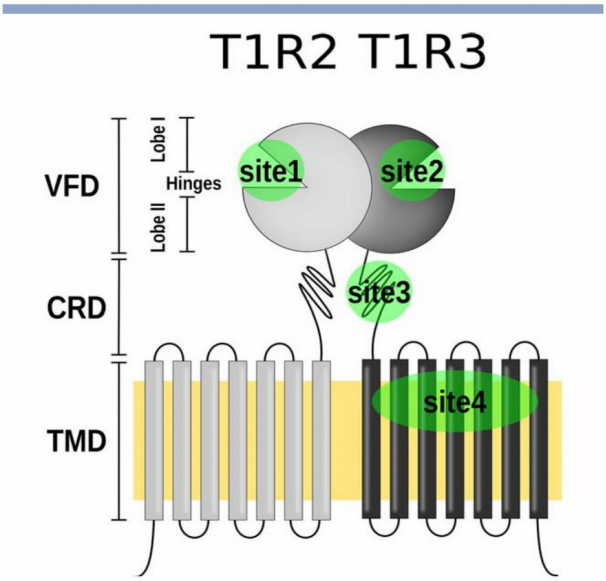


Figure 1

Schematic representation of the T1R2-T1R3 sweet taste receptor structure. The large N-terminal domain of each monomer, named the Venus flytrap domain (VFD), is connected to a seven-helix transmembrane domain (TMD) by a cysteine-rich domain (CRD). The orthosteric binding site is located in the VFDs (sites 1 and 2), while allosteric modulators bind to CRD and TMD (sites 3, 4, and 5). [Color figure can be viewed at wileyonlinelibrary.com]

never been crystallized, it then remains challenging to define the sweetener binding site on this receptor.

For GPCRs for which no experimental structure is available, computational modeling provided fruitful insights into their structure–function relationship using homology modeling. Typically, in the field of chemical senses, ligand–receptor interactions and dynamics were discussed on G protein-coupled odorant receptors.[14,15] More specifically, considering taste receptors, models of the sweet taste VFD and TMD have been proposed based on other class C VFD[16,17] and on class A GPCR templates,[18–20] respectively.

In the present study, a theoretical model of the human T1R2-T1R3 sweet taste receptor has been constructed based on an updated class C sequence alignment with constraints from available structural and site-directed mutagenesis data (Table I). The newly available experimental structure of mGluRs,[35,50] considered as prototypical class C GPCRs represents a significant breakthrough for predicting the structural features responsible for ligand recognition and receptor activation within this family. The molecular model accounts for the residues known to be involved in the orthosteric and allosteric binding sites. The structure of the sweet taste receptor sheds light on the molecular switches in the TMD. By comparing the inter-species variability of the sweet taste genes, we discuss the differences observed for sweet taste receptor activation in mammals.

MATERIALS AND METHODS

Sequence alignment

Class C GPCR sequences were aligned using the ClustalO algorithm.[51] The alignment was manually refined based on experimental constraints. The positions of residues in the VFD sequences were adjusted to satisfy class C GPCRs site-directed mutagenesis data and to conciliate the structural superposition of available crystallographic structures (Table I). The TMD sequence alignment was refined based on class C bundles site directed mutagenesis, transmembrane segment predictions[11,52] and crystallographic structures of mGluR1 and mGluR5 (Table I). Residue numbering of the class C GPCRs correspond to the Ballesteros–Weinstein definition proposed by Pin *et al.*[52]

Sweet taste receptor molecular modeling

The sweet taste receptor was constructed using the homology modeling software Modeller[53] by constraining GPCR canonical disulfide bridges and those corresponding to the VFD mGluR1 structure. All of the models were refined by energy minimization using the Amber molecular modeling package and the ff03 force field parameters.[54] X-ray structures of mGluR VFD (1EWT and 1EWK), open and closed receptor states, provide relevant templates to model various conformational states of the sweet taste monomers. The structure of the mGluR1 TMD (4OR2) was used as a template for the T1R2-T1R3 sweet taste receptor transmembrane bundle. The structure of Bovine Rhodopsin class A GPCR (1F88) has been chosen to build an alternate model of the sweet taste TMD as it has been proposed in previous studies[18–20] to discuss the influence of the template on homology modeling. The volume of the binding cavity was calculated with fpocket.[55]

RESULTS AND DISCUSSION

Extracellular domain

The venus flytrap domain (VFD) exhibits several binding pockets

From a structural point of view, the sweet taste VFD structure is a heterodimer made up of monomers containing approximately 250 amino-acids folded into two lobes and connected by three hinges[52] (Fig. 1). As revealed by the structural superposition of mGluR1, 2, 3, 5, and 7, $GABA_B1$ and $_B2$ VFDs experimental structures, a highly conserved topology is maintained in those dimers (Supporting Information Fig. S1).

Table I
Summary of Experimental Data Used to Align Class C GPCRs Sequences

		Crystallographic structures	References of site-directed mutagenesis experiments
VFD	T1R1	—	21,22
	T1R2	—	12,16,17,23
	mGluR1	1EWK, 1EWT, 1EWV, 1ISS, 1ISR, 3KS9	24–26
	mGluR2	5CNI, 5CNJ, 5CNK, 5CNM	27,28
	mGluR3	2E4U, 2E4V, 2E4W, 2E4X, 2E4Y, 2E4Z, and 3SM9	28
	mGluR4	—	26,29–32
	mGluR5	3LMK	
	mGluR7	3MQ4	
	mGluR8	—	32
	$GABA_{B\ (1\ \&\ 2)}$	4MQE, 4MQF, 4MR7, 4MR8, 4MR9, 4MRM, 4MS1, 4MS3, and 4MS4	33,34
TMD	T1R3	—	18–20
	mGluR1	4OR2	35–40
	mGluR2	—	41–43
	mGluR5	4OO9, 5CGC, and 5CGD	36,44
	mGluR8	—	45
	CasR	—	46–49

Accordingly, for these known structures, the profile of sequence alignment has been built under the constraint of their structural superposition. Then, the sequences of those receptors with unknown structures have been further added to the sequence alignment profile. A refined alignment was obtained by checking that the available site-directed mutagenesis *in vitro* data for all these receptors were fitting the model (Table I). One can note a minor difference with already published sequence alignments[16,17] for the residue K65 in T1R2 consistent with a putative role in the maintain of a closed structure of the VFD.[16,17,25] The alignment unveils a moderate degree of sequence identity within the class C sub-family (29 ± 1% identity between T1Rs and mGluRs and 39% identity between T1R2 and T1R3, cf. Supporting Information Table S1).

Above all, our sequence alignment recapitulates current knowledge determined from experimental data (Supporting Information Figs. S1 and S2) and contributes to understanding the ligand–receptor interactions involved in the recognition of sweeteners by T1R2 and T1R3. Similar to what is observed in X-Ray structures of mGluR VFDs,[56–59] the binding cavities for sweet compounds are located in between the T1R2 and T1R3 lobes (Fig. 2). We provide a list of the key residues that interact with sweeteners in Figure 3.

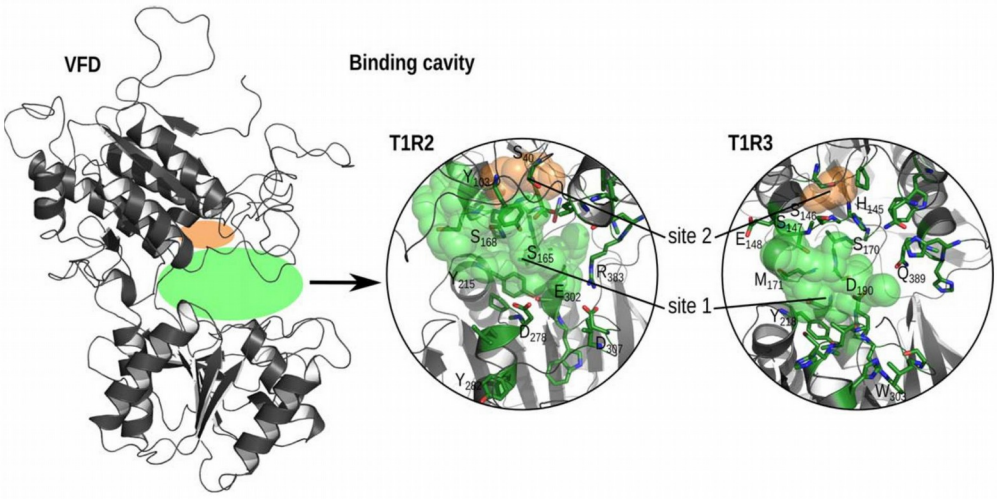


Figure 2
(Left) Structure of the T1R2 and T1R3 VFD homology model. (Right) The orthosteric binding site is in green and the extended pocket is in orange. The residues involved with the binding pocket are shown as sticks. [Color figure can be viewed at wileyonlinelibrary.com]

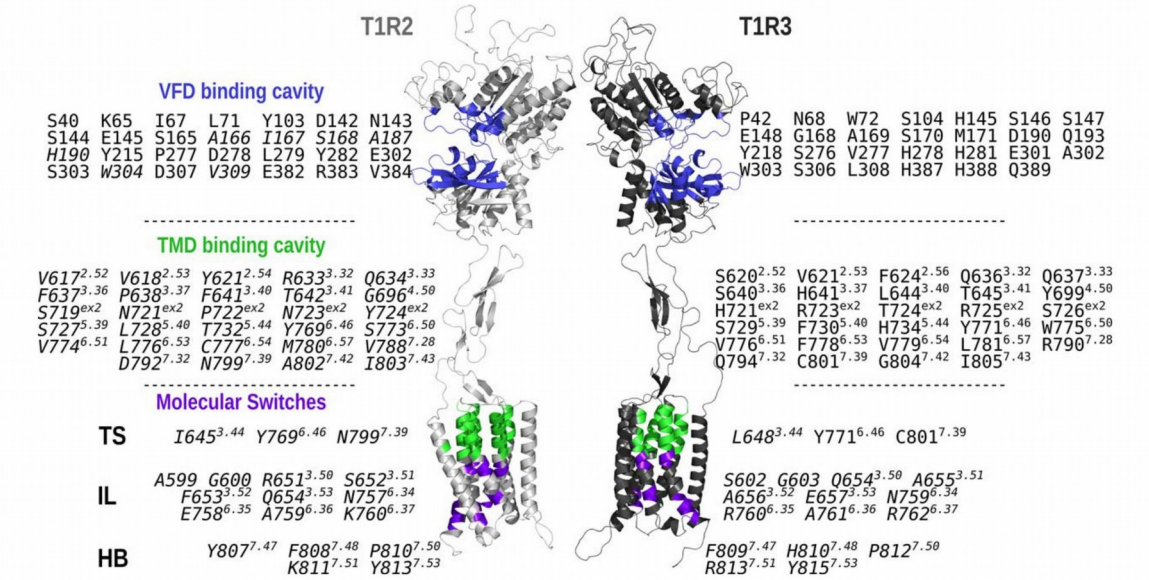


Figure 3

Structural overview of the sweet taste receptor model and a list of the key residues. Putative residues highlighted from the sequence alignments (Supporting Information Figs. S2 and S4) or based on the homology model are written in italic. Amino-acids involved in the VFD orthosteric binding site (in blue) correspond to the Wellendorph and Bräuner-Osborne's definition of the orthosteric binding site homology of class C GPCRs[11] and mutagenesis studies of Xu *et al.*, Proc Natl Acad Sci, 2004[12]; Zhang *et al.*, Proc Natl Acad Sci, 2010[23]; Masuda *et al.*, Plos One, 2012[16]; Maillet *et al.*, Chem Senses, 2015.[17] Residues of the allosteric binding sites (in green) come from experimental studies of T1R3 by Jiang *et al.*, J Biol Chem, 2005[18,19] and Winnig *et al.*, BMC Struct Biol, 2007[20] and then extrapolated to T1R2. The molecular signature of class C GPCRs (in purple) corresponds to the molecular switches: Transmission Switch (TS), Ionic Lock (IL) and Hydrophobic Barrier (HB). They are defined with the sequence alignment and the X-ray structures of mGluR1[35] (PDB identifier: 4OR2) and mGluR5[50] (4OO9, 5CGC, and 5CGD). [Color figure can be viewed at wileyonlinelibrary.com]

The sweetener chemical space contains hundreds of molecules that are mainly polar and soluble.[60,61] Structural analysis of our model shows that the T1R2 and T1R3 orthosteric binding pockets (site 1 in Fig. 2) have a volume of 4900 Å^3 in their open form and allows small as well as large sweeteners to bind to the receptor, in agreement with previously published homology models.[62] Consistent with the properties of sweeteners,[60,61] both the T1R2 and T1R3 cavities are hydrophilic, with 45 and 50% of their surface areas, respectively, being accessible to polar molecules. Inspecting the model, we also identified the presence of a secondary cavity (site 2 in Fig. 2) adjoining the main pocket that is similar to a pocket found on mGluR4.[63] In the model of T1R2, the residue S40 is located in this cavity. This finding is in line with the mutation S40A that decreases the response to aspartame.[17] As with mGluR, this second pocket offers an opportunity to design new potent sweeteners.

The cysteine rich domain (CRD) structure is constrained by disulfide bridges

As found in mGluRs, the VFD of the sweet taste receptor is connected to the TMD via a cysteine-rich domain (CRD).[52] The CRDs share nine highly conserved cysteine residues. The crystal structure of mGluR3 reveals that eight of them form four intra-domain disulfide bonds while the remaining one is bound to a cysteine residue in the VFD[57] It has been suggested that these rigidifying structural constraints favor the coupling between the VFD and the TMD upon protein activation.[11]

Our sequence alignment (Supporting Information Fig. S3) shows that the CRDs share modest sequence identities (Supporting Information Table S2) of 31 ± 2% between the T1R and mGluR protein families. The homology model of the sweet taste receptor is then highly dependent on the disulfide bound structural constraints. However, the ten latest residues of the CRD form an unstructured region, suggesting that they are flexible enough to enable a local deformation of the structure upon the VFD-TMD coupling. In our model, this flexible region is at roughly 10–15 Å away from the extracellular loop 2 of the TMD. One can suppose the formation of inter-domain interactions during the activation mechanism to stabilize the active state.

Model of T1R2/T1R3 transmembrane domain inspired from class C GPCRs Class C GPCRs share a 7-transmembrane helix domain that can be targeted by positive or negative allosteric ligands. It also can behave as class A GPCRs in the

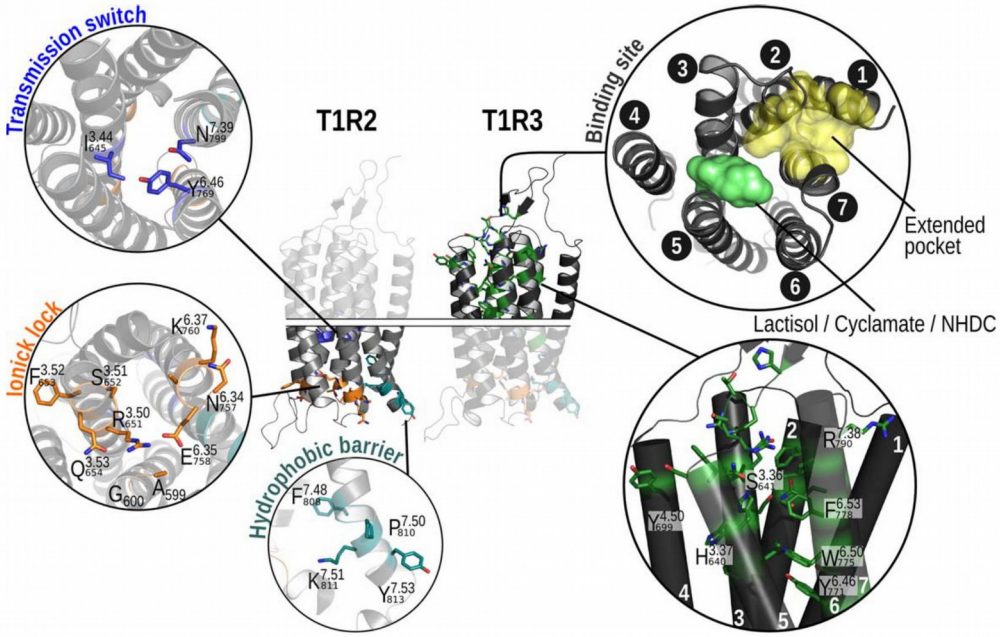


Figure 4

Structural details of the T1R2 and T1R3 TMD homology models. The molecular switches involved in GPCR activation are highlighted in the T1R2 TMD, which is the monomer that couples to the G-protein. Lactisol, cyclamate and neohesperidine dihydrochalcone (NHDC) ligands bind the T1R3 TMD binding pocket (residues colored in green), suggesting that the chemical information is transmitted to the T1R2 TMD through the protein–protein interface.

absence of the extracellular domain as demonstrated for mGluR5.[64] This suggests that class C TMDs contain a canonical GPCR superfamily binding site and the naming of the transmembrane pocket as either an allosteric or secondary orthosteric binding site can be debated.

An optimal alignment has been obtained by avoiding gaps in the TM helices. It is based on the conservation of key residues obtained from experimental data (Table I and Supporting Information Fig. S4). We propose for the first time homology models of T1R2 and T1R3 TMDs based on mGluR templates. Alignment of the sweet taste receptor sequences and those of the templates leads to sequence identities between 23 and 24% (Supporting Information Table S3). T1R2 and T1R3 TMD have a moderate sequence identity of 28% (Supporting Information Table S3) which suggests a different functional role. The tridimensional model of the sweet taste receptor proposed here is fully compatible with available experimental data as described in the following.

Structural details of the T1R2 and T1R3 binding pockets

T1R3, the common subunit of the sweet and umami taste receptor remains the only monomer for which mutation data and chimeric constructs have shown that the transmembrane portion of the protein binds ligands.[18–20] As shown by site-directed mutagenesis,[18,19] residues from TMs 3, 5, and 6 and extracellular loop 2 (denoted ex2) (Fig. 3) circumscribe the binding cavity of T1R3 for small ligands such as cyclamate and lactisole (Fig. 4). Additional site-directed mutations such as S620$^{2.57}$A, F624$^{2.51}$ L, Y699$^{4.50}$ L/F, and C801$^{7.39}$I suggest that the binding pocket may be extended to helices 2, 4 and 7 in the presence of larger ligands.[20] Consistently, our model, shown in Figure 4, reports that the above cited residues point toward the binding cavity.

T1R2 TMD also binds small size synthetic compounds such as perillartine, P4000 and S-819.[22,65] However, mutation data on T1R2 TMD binders are lacking to describe ligand–receptor interactions at the molecular level. The T1R2 TMD binding pocket was extrapolated from the sequence alignment and compared to T1R3. Variations in the TM3 and TM6 sequences between T1R2 and T1R3 lead to a structural modification in the TMD binding cavity. S640$^{3.36}$ in T1R3, belonging to the binding cavity as demonstrated by mutagenesis experiments[18,20] (S640A, Supporting Information Fig. S4), is replaced by F637$^{3.36}$ in T1R2, which is consistent with a steric hindrance for ligands in T1R2. H641$^{3.37}$, R723^{ex2}, and F778$^{6.53}$ in T1R3, highlighted by reports of single point mutations[18–20] (H641A, R723A, and F778A) constitute three anchor points for ligand recognition, via

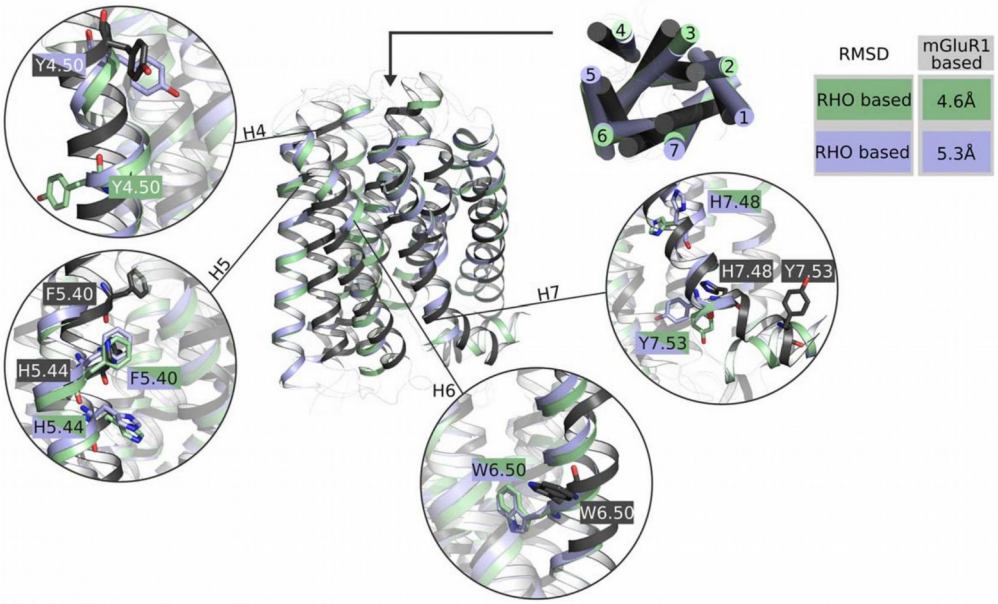


Figure 5

Comparison of the homology models based on the alignment proposed in Supporting Information Figure S4 and the mGluR5 crystal structure[50] as template (in dark grey, pdb identifier: 4or2) and two sequence alignments from Jiang *et al.* JBC, 2005[18] in green and Winnig *et al.* BMC, 2007[20] in blue with the rhodopsin crystal structure as template (pdb identifier: 1f88). The top view highlights the global shift of the TM helices. The side view reveals the local differences in the position of key residues involved in ligand binding or in TMD molecular switches. [Color figure can be viewed at wileyonlinelibrary.com]

electrostatic interactions and π-π stacking.[18] Based on the alignment, the corresponding T1R2 residues, P638$^{3.37}$, N721^{ex2}, and L776$^{6.53}$ differ from those in T1R3 (Supporting Information Fig. S4).

In the model of T1R3 TMD, we observe a principal binding pocket, which is shown in Figure 4, of 270 Å^3 in volume. 38% of its surface area is accessible to polar molecules. A smaller pocket with a volume of 210 Å^3 can be found in the T1R2 model. Ranging between 150 and 200 Å^3, the smallest ligands (lactisole and cyclamate) can fit into the T1R3 binding pocket but not the T1R2 binding pocket. These structural differences are in line with experimental data that show that lactisole, cyclamate and NHDC bind T1R3 but not T1R2. Nevertheless, a less buried 310 Å^3 T1R2 pocket surrounded by TMs 3, 5, and 6 is observed and may be important for designing new sweet taste receptor modulators.[22]

Functional differences in T1R2/T1R3 molecular switches

Molecular switches are important structural motifs involved in the transmission of the ligand-binding signal from the extracellular part or the receptor to the cytosolic side of the TMD where G proteins bind and have been extensively described for class A GPCRs.[66–69] With regard to class C GPCRs, molecular switches have been identified in recent X-ray structures of mGluR1[35] and mGluR5.[50] Sequence alignment shows a high conservation of the residues involved in these molecular switches in the mGluR subfamily (Supporting Information Fig. S4). Thus, the T1R molecular switches have been extrapolated from class C sequence alignment.

The class C transmission switch, acting as a ligand detecting sensor, involves residues 3.44, 6.46, and 7.39 according to Ballesteros–Weinstein numbering and is in line with their position in the X-ray structure and the effect of single point mutations Y659$^{3.44}$A and S809$^{7.39}$A/F in mGluR5 and T794$^{6.46}$A in mGluR1.[35,50] By comparing the mGluR and T1R sequences one can note the conservation of the physicochemical properties of the triad despite the inversion of the tyrosine residue (Y$^{3.44}$ in mGluRs and Y$^{6.46}$ in T1R2 and T1R3) from helices 3 to 6 (Fig. 3 and Supporting Information Fig. S5). N800$^{7.39}$ in T1R2 replaces a smaller cysteine and serine residue in T1R3 and mGluR, respectively. The mutations Y771$^{6.46}$A and C801$^{7.39}$I/L in the transmission switch of T1R3 abolish the response of the sweet taste receptor and highlight the functional role of these residues in the sweet taste receptor.

In mGluR1 and mGluR5, the K$^{3.50}$-E$^{6.35}$ ionic lock maintains the receptor in its inactive state.[35,50] As found in the crystal structures, strong electrostatic interactions

involving $K^{3.50}$, $R^{3.53}$ and two serine residues within ICL1 reinforce the TM3-TM6 lock. These residues in mGluR form two conserved KTNR and NEAK motifs within TM3 and TM6, respectively (Fig. 3 and Supporting Information Fig. S4). The sequence alignment and homology models allow us to highlight the corresponding critical residues involved in T1R activation. T1R1 and T1R2 have conserved functional amino acids, that is, a positively (arginine) and negatively (aspartate) charged residue at positions 3.50 and 6.35, respectively (Fig. 4). In T1R3, the position of polar residues $Q^{3.50}$ and $R^{6.35}$ is inverted. The model also suggests that residue $E^{3.53}$ can interact with $R^{6.35}$ then contributing to reinforce the T1R3 ionic lock. Even if the sequence alignment indicates this subtle variation, our homology model demonstrates a similar interaction scheme at the ionic lock.

In class C GPCRs, the [F/Y]xPKxY motif, which forms the hydrophobic barrier[50] preventing the flooding of the GPCR cytosolic side, is highly conserved (Supporting Information Fig. S4). Based on the sequence alignment, the residue $F/Y^{7.48}$ represents the hydrophobic barrier of T1R2 and is replaced by another aromatic residue ($H^{7.48}$) in T1R3. In the T1R models, one can notice that the hydrophobic barrier is reinforced by the contribution of another aromatic residue, $F^{7.47}$ (for T1R1 and T1R2) or $Y^{7.47}$ (for T1R3) (Fig. 4 and Supporting Information Fig. S4).

Influence of the sequence alignment and structure template

A critical comparison of our homology models based on class C templates and already published ones based on class A GPCR[18–20] has been carried out. Since no site-directed mutagenesis experiments on the TMD of T1R2 are available, the comparison has been performed on the T1R3 monomer. It reveals important structural differences on both the relative position of the helices and the orientation of the side chains as detailed in the following.

The use of class C GPCRs templates leads to a shift of the transmembrane segments with respect to the use of class A, principally for TM 2, 5, and 7. It also decreases the accessibility and the volume of the binding pocket (Fig. 5).[50] The shift of one helical turn in the structure of TM5 helix based on mGluR templates does not affect the list of residues involved in the binding cavity ($Y^{5.40}$ and $H^{5.44}$). Compared to the previous sequence alignment,[18] our model shifts the end of the TM4 segment by one and a half helical turn. Accordingly and consistent with site-directed mutagenesis, our model recovers that residue $Y^{4.50}$, is part of the cavity (Fig. 5).[20]

The orientation of $W^{6.50}$ also clearly differs between the models (Fig. 5). In previous studies,[18–20] $W^{6.50}$ in class C GPCRs was considered aligned with non-olfactory class A residue $W^{6.48}$.[66] However, the sequence similarity between class A and the sweet taste receptors is too low (for instance 14% with rhodopsin) to clearly address the questions regarding the key residues involved in receptor activation. In class C experimental structures, $W^{6.50}$ does not delimit the bottom of the binding cavity but point toward TM6 of the second unit. As the active state of the class C receptors involves TM6-TM6 interactions,[70] $W^{6.50}$ conservation and orientation suggest that its role in the activation process is different from that of class A GPCRs. The alignment reveals that the highly conserved residue $W^{6.50}$ is replaced by a serine in T1R2, suggesting a functional difference with respect to T1R3.

The class C hydrophobic barrier, formed by the motif [F/Y]xPKxY at the end of TM7, was aligned with the equivalent conserved motif $NPxxY^{66}$ in class A GPCRs. As found in the mGluR1 and mGluR5 X-ray structures, $F/Y^{7.48}$ (and not $Y^{7.53}$) would behave as $Y^{7.53}$ in class A. In light of the available class C template, it is very likely that our model is more accurate at these positions than previous models[18,20] (Fig. 5).

Interspecies variability of the sweet taste receptors

Based on the alignment of the eleven T1R2 and seven T1R3 mammalian reviewed sequences in the UniProt database, we observe at least 70% sequence identity for both T1R2 and T1R3 (Supporting Information Figs. S6 and S7). The VFD is less conserved than the TMD. Between humans and rodents, the corresponding percent identity is 70% and 80% for the VFDs and TMD, respectively. These high percentages of sequence identity suggest a similar structure and a conserved activation mechanism for mammalian sweet taste receptors. However, the mammalian sequence alignment also reveals subtle variations in key residues involved in the activation mechanism. These variations explain the differences observed in sweetener perception between the organisms.[8,9] The phylogenetic tree shows that the *Canis lupus familiaris, Mus musculus,* and *Rattus norvegicus* sequences diverge from other mammalian sequences. The chemical space of compounds that activate the sweet taste receptor in rodents is slightly different than that of humans. As reviewed by Bachmanov,[71] rodents do not perceive primate-specific sweeteners such as aspartame, neotame, cyclamate and NHDC. Aspartame and neotame are T1R2 VFD ligands,[12] while cyclamate and NHDC bind the T1R3 TMD pocket.[19,20] By comparing the residues that define the T1R2 VFD binding cavity in human and rodent sweet taste receptors, subtle variations in the side-chain lengths of the residues interacting with the ligands can be observed. S40, S168, and D278 in the human receptor correspond to T43, T171, and E282, respectively, in the mouse receptor. The alignment of mammalian sequences also reveals differences in the T1R3 TMD binding site (S640, T724, and R790 residues in human

are replaced by A645, [V/M]729 and Q795 in rodent sequences, respectively). In light of the provided T1R2-T1R3 model, such modifications of the ligand–receptor interactions provide a rational explanation for the loss in sweetener binding affinity[71] and will contribute to elaborating new hypotheses to understand interspecies variability.

CONCLUSIONS

The description of the T1R2-T1R3 structure is of crucial importance for deciphering the molecular basis underlying sweet taste perception and designing new non-caloric sweeteners or taste modifiers. The sweet taste receptor is also expressed in extraoral tissues where it plays a different role than its principal function.[10] It follows that a better understanding of sweet taste receptor structure and dynamics will also be beneficial for pharmaceutical research. To date, no structural experimental data for the sweet taste receptor is available and the mechanism of ligand recognition and receptor activation is still under debate. The publication of recent crystallographic structures of mGluR subfamily proteins provided a major leap for the modeling other class C GPCRs, like the sweet taste receptor. In the present article, a critical alignment of class C GPCRs sequences has been performed to provide a relevant three-dimensional structure of the sweet taste receptor that satisfies actual available structural and site-directed mutagenesis experimental constraints. These data provide an atomic point of view for T1R2-T1R3 and a robust starting point for initiating structural and mechanistic studies on the receptor.

REFERENCES

1. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature 2009;459:356–363.
2. Katritch V, Cherezov V, Stevens RC. Structure-function of the G protein-coupled receptor superfamily. Annu Rev Pharmacol Toxicol 2013;53:531–556.
3. Kobilka BK, Deupi X. Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci 2007;28:397–406.
4. Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 2003;63:1256–1272.
5. Isberg V, Mordalski S, Munk C, Rataj K, Harpsøe K, Hauser AS, Vroling B, Bojarski AJ, Vriend G, Gloriam DE. GPCRdb: an information system for G protein-coupled receptors. Nucleic Acids Res 2016;44:D356–D364.
6. Lv X, Liu J, Shi Q, Tan Q, Wu D, Skinner JJ, Walker AL, Zhao L, Gu X, Chen N, Xue L, Si P, Zhang L, Wang Z, Katritch V, Liu Z, Stevens RC. In vitro expression and analysis of the 826 human G protein-coupled receptors. Protein Cell 2016;7:325–337.
7. Zhao GQ, Zhang Y, Hoon MA, Chandrashekar J, Erlenbach I, Ryba NJP, Zuker CS. The receptors for mammalian sweet and umami taste. Cell 2003;115:255–266.
8. Nelson G, Chandrashekar J, Hoon MA, Feng L, Zhao G, Ryba NJP, Zuker CS. An amino-acid taste receptor. Nature 2002;416:199–202.
9. Nelson G, Hoon MA, Chandrashekar J, Zhang Y, Ryba NJP, Zuker CS. Mammalian sweet taste receptors. Cell 2001;106:381–390.
10. Laffitte A, Neiers F, Briand L. Functional roles of the sweet taste receptor in oral and extraoral tissues. Curr Opin Clin Nutr Metab Care 2014;17:379–385.
11. Wellendorph P, Bräuner-Osborne H. Molecular basis for amino acid sensing by family C G-protein-coupled receptors. Br J Pharmacol 2009;156:869–884.
12. Xu H, Staszewski L, Tang H, Adler E, Zoller M, Li X. Different functional roles of T1R subunits in the heteromeric taste receptors. Proc Natl Acad Sci USA 2004;101:14258–14263.
13. Li X, Staszewski L, Xu H, Durick K, Zoller M, Adler E. Human receptors for sweet and umami taste. Proc Natl Acad Sci USA 2002; 99:4692–4696.
14. March CA. d, Yu Y, Ni MJ, Adipietro KA, Matsunami H, Ma M, Golebiowski J. Conserved residues control activation of mammalian G protein-coupled odorant receptors. J Am Chem Soc 2015;137: 8611–8616.
15. Golebiowski J, Antonczak S, Fiorucci S, Cabrol-Bass D. Mechanistic events underlying odorant binding protein chemoreception. Proteins 2007;67:448–458.
16. Masuda K, Koizumi A, Nakajima K, Tanaka T, Abe K, Misaka T, Ishiguro M. Characterization of the modes of binding between human sweet taste receptor and low-molecular-weight sweet compounds. PLoS One 2012;7:e35380.
17. Maillet EL, Cui M, Jiang P, Mezei M, Hecht E, Quijada J, Margolskee RF, Osman R, Max M. Characterization of the binding site of aspartame in the human sweet taste receptor. Chem Senses 2015;40:577–586.
18. Jiang P, Cui M, Zhao B, Liu Z, Snyder LA, Benard LMJ, Osman R, Margolskee RF, Max M. Lactisole interacts with the transmembrane domains of human T1R3 to inhibit sweet taste. J Biol Chem 2005; 280:15238–15246.
19. Jiang P, Cui M, Zhao B, Snyder LA, Benard LMJ, Osman R, Max M, Margolskee RF. Identification of the cyclamate interaction site within the transmembrane domain of the human sweet taste receptor subunit T1R3. J Biol Chem 2005;280:34296–34305.
20. Winnig M, Bufe B, Kratochwil NA, Slack JP, Meyerhof W. The binding site for neohesperidin dihydrochalcone at the human sweet taste receptor. BMC Struct Biol 2007;7:66.
21. Toda Y, Nakagita T, Hayakawa T, Okada S, Narukawa M, Imai H, Ishimaru Y, Misaka T. Two distinct determinants of ligand specificity in T1R1/T1R3 (the umami taste receptor). J Biol Chem 2013; 288:36863–36877.
22. Zhang F, Klebansky B, Fine RM, Xu H, Pronin A, Liu H, Tachdjian C, Li X. Molecular mechanism for the umami taste synergism. Proc Natl Acad Sci USA 2008;105:20930–20934.
23. Zhang F, Klebansky B, Fine RM, Liu H, Xu H, Servant G, Zoller M, Tachdjian C, Li X. Molecular mechanism of the sweet taste enhancers. Proc Natl Acad Sci USA 2010;107:4752–4757.
24. 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. J Biol Chem 2003;278: 4314–4321.
25. Jiang Y, Huang Y, Wong H-C, Zhou Y, Wang X, Yang J, Hall RA, Brown EM, Yang JJ. Elucidation of a novel extracellular calcium-binding site on metabotropic glutamate receptor 1α (mGluR1α) that controls receptor activation. J Biol Chem 2010;285:33463–33474.
26. Kang HJ, Menlove K, Ma J, Wilkins A, Lichtarge O, Wensel TG. Selectivity and evolutionary divergence of metabotropic glutamate receptors for endogenous ligands and G proteins coupled to phospholipase C or TRP channels. J Biol Chem 2014;289:29961–29974.
27. Malherbe P, Knoflach F, Broger C, Ohresser S, Kratzeisen C, Adam G, Stadler H, Kemp J. a, Mutel V. Identification of essential residues involved in the glutamate binding pocket of the group II metabotropic glutamate receptor. Mol Pharmacol 2001;60:944–954.

28. Yao Y, Pattabiraman N, Michne WF, Huang X-P, Hampson DR. Molecular modeling and mutagenesis of the ligand-binding pocket of the mGlu3 subtype of metabotropic glutamate receptor. J Neurochem 2003;86:947–957.
29. Hampson DR, Huang XP, Pekhletski R, Peltekova V, Hornby G, Thomsen C, Thøgersen H. Probing the ligand-binding domain of the mGluR4 subtype of metabotropic glutamate receptor. J Biol Chem 1999;274:33488–33495.
30. Rosemond E, Peltekova V, Naples M, Thøgersen H, Hampson DR. Molecular determinants of high affinity binding to group III metabotropic glutamate receptors. J Biol Chem 2002;277:7333–7340.
31. Hermit MB, Greenwood JR, Bräuner-Osborne H. Mutation-induced quisqualic acid and ibotenic acid affinity at the metabotropic glutamate receptor subtype 4. J Biol Chem 2004;279:34811–34817.
32. Goudet C, Vilar B, Courtiol T, Deltheil T, Bessiron T, Brabet I, Oueslati N, Rigault D, Bertrand H-O, McLean H, Daniel H, Amalric M, Acher F, Pin J-P. A novel selective metabotropic glutamate receptor 4 agonist reveals new possibilities for developing subtype selective ligands with therapeutic potential. Faseb J 2012;26: 1682–1693.
33. Galvez T, Parmentier M-L, Joly C, Malitschek B, Kaupmann K, Kuhn R, Bittiger H, Froestl W, Bettler B, Pin J-P. Mutagenesis and modeling of the GABAB receptor extracellular domain support a venus flytrap mechanism for ligand binding. J Biol Chem 1999;274: 13362–13369.
34. Mukherjee RS, McBride EW, Beinborn M, Dunlap K, Kopin AS. Point mutations in either subunit of the GABAB receptor confer constitutive activity to the heterodimer. Mol Pharmacol 2006;70: 1406–1413.
35. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V, Conn PJ, Stevens RC. Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. Science 2014;344:58–64.
36. Pagano A, Ruegg D, Litschig S, Stoehr N, Stierlin C, Heinrich M, Floersheim P, Prezeau L, Carroll F, Pin J-P, Cambria A, Vranesic I, Flor PJ, Gasparini F, Kuhn R. 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. J Biol Chem 2000;275:33750–33758.
37. Knoflach F, Mutel V, Jolidon S, Kew JN, Malherbe P, Vieira E, Wichmann J, Kemp JA. Positive allosteric modulators of metabotropic glutamate 1 receptor: characterization, mechanism of action, and binding site. Proc Natl Acad Sci USA 2001;98:13402–13407.
38. Malherbe P, Kratochwil N, Knoflach F, Zenner M-T, Kew JNC, Kratzeisen C, Maerki HP, Adam G, Mutel V. Mutational analysis and molecular modeling of the allosteric binding site of a novel, selective, noncompetitive antagonist of the metabotropic glutamate 1 receptor. J Biol Chem 2003;278:8340–8347.
39. Suzuki G, Kimura T, Satow A, Kaneko N, Fukuda J, Hikichi H, Sakai N, Maehara S, Kawagoe-Takaki H, Hata M, Azuma T, Ito S, Kawamoto H, Ohta H. 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). J Pharmacol Exp Ther 2007;321:1144–1153.
40. Fukuda J, Suzuki G, Kimura T, Nagatomi Y, Ito S, Kawamoto H, Ozaki S, Ohta H. 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 2009;57:438–445.
41. Schaffhauser H. Pharmacological characterization and identification of amino acids involved in the positive modulation of metabotropic glutamate receptor subtype 2. Mol Pharmacol 2003;64:798–810.
42. Hemstapat K, Costa HD, Nong Y, Brady AE, Luo Q, Niswender CM, Tamagnan GD, Conn PJ. A novel family of potent negative allosteric modulators of group II metabotropic glutamate receptors. J Pharmacol Exp Ther 2007;322:254–264.
43. Rowe BA, Schaffhauser H, Morales S, Lubbers LS, Bonnefous C, Kamenecka TM, McQuiston J, Daggett LP. 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. J Pharmacol Exp Ther 2008;326:240–251.
44. Molck C, Harpsoe K, Gloriam DE, Clausen RP, Madsen U, Pedersen LO, Jimenez HN, Nielsen SM, Mathiesen JM, Brauner-Osborne H. Pharmacological characterization and modeling of the binding sites of novel 1,3-bis(pyridinylethynyl)benzenes as metabotropic glutamate receptor 5-selective negative allosteric modulators. Mol Pharmacol 2012;82:929–937.
45. Yamashita T, Terakita A, Kai T, Shichida Y. Conformational change of the transmembrane helices II and IV of metabotropic glutamate receptor involved in G protein activation. J Neurochem 2008;106:850–859.
46. Bai M, Quinn S, Trivedi S, Kifor O, Pearce SHS, Pollak MR, Krapcho K, Hebert SC, Brown EM. Expression and characterization of inactivating and activating mutations in the human Ca2+o-sensing receptor. J Biol Chem 1996;271:19537–19545.
47. Jensen AA, Spalding TA, Burstein ES, Sheppard PO, O'Hara PJ, Brann MR, Krogsgaard-Larsen P, Bräuner-Osborne H. Functional importance of the Ala(116)-Pro(136) region in the calcium-sensing receptor. Constitutive activity and inverse agonism in a family C G-protein-coupled receptor. J Biol Chem 2000;275:29547–29555.
48. Rus R, Haag C, Bumke-Vogt C, Bähr V, Mayr B, Möhlig M, Schulze E, Frank-Raue K, Raue F, Schöfl C. Novel inactivating mutations of the calcium-sensing receptor: the calcimimetic NPS R-568 improves signal transduction of mutant receptors. J Clin Endocrinol Metab 2008;93:4797–4803.
49. White E, McKenna J, Cavanaugh A, Breitwieser GE. Pharmacochaperone-mediated rescue of calcium-sensing receptor loss-of-function mutants. Mol Endocrinol 2009;23:1115–1123.
50. Doré AS, Okrasa K, Patel JC, Serrano-vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B, Weir M, Wiggin GR, Marshall FH. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature 2014;511:557–562.
51. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2014;7:539–539.
52. Pin J-P, Galvez T, Prézeau L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. Pharmacol Ther 2003;98:325–354.
53. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
54. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Götz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA. AMBER 12. San Francisco: University of California; 2012.
55. Guilloux VL, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics 2009;10:168.
56. Morikawa K, Kunishima N, Shimada Y, Tsuji Y, Sato T, Yamamoto M, Kumasaka T, Nakanishi S, Jingami H, Morikawa K. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. Nature 2000;407:971–977.
57. Muto T, Tsuchiya D, Morikawa K, Jingami H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. Proc Natl Acad Sci USA 2007;104:3759–3764.

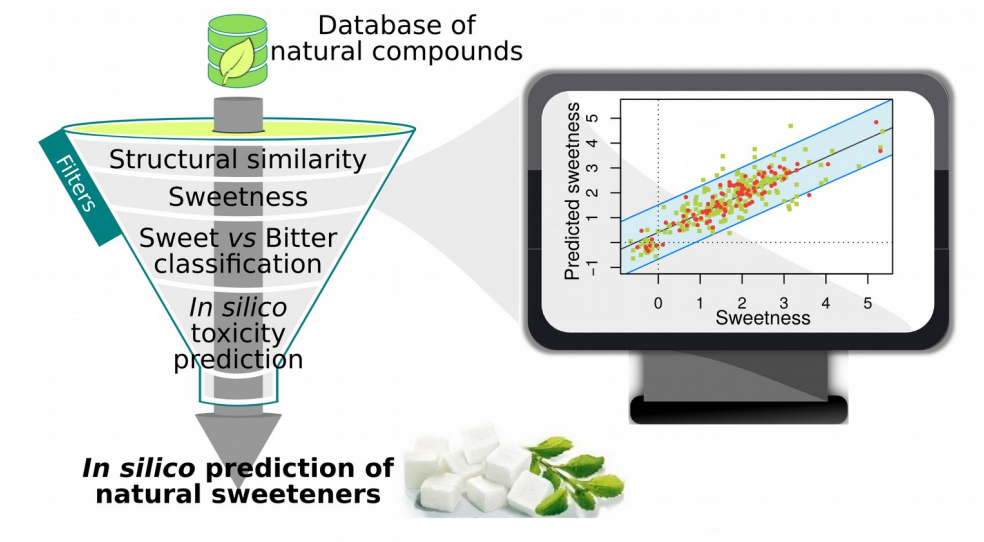
58. Tsuchiya D, Kunishima N, Kamiya N, Jingami H, Morikawa K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+. Proc Natl Acad Sci USA 2002;99:2660–2665.
59. Monn JA, Valli MJ, Massey SM, Hao J, Reinhard MR, Bures MG, Heinz BA, Wang X, Carter JH, Getman BG, Stephenson GA, Herin M, Catlow JT, Swanson S, Johnson BG, McKinzie DL, Henry SS. Synthesis and pharmacological characterization of 4-substituted-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylates: identification of new potent and selective metabotropic glutamate 2/3 receptor agonists. J Med Chem 2013;56:4442–4455.
60. DuBois GE, Prakash I. Non-Caloric Sweeteners, Sweetness Modulators, and Sweetener Enhancers. Annu Rev Food Sci Technol 2012;3:353–380.
61. Behrens M, Meyerhof W, Hellfritsch C, Hofmann T. Sweet and umami taste: natural products, their chemosensory targets, and beyond. Angew Chem Int Ed Engl 2011;50:2220–2242.
62. Temussi PA. New insights into the characteristics of sweet and bitter taste receptors. Inte Rev Cell Mol Biol 2011;291:191–226.
63. Acher FC, 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 2011;60:102–107.
64. Goudet C, Gaven F, Kniazeff J, Vol C, Liu J, Cohen-Gonsaud M, Acher F, Prézeau L, Pin JP. Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. Proc Natl Acad Sci USA 2004;101:378–383.
65. DuBois GE. Molecular mechanism of sweetness sensation. Physiol Behav 2016;164:453–463.
66. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S. Action of molecular switches in GPCRs–theoretical and experimental studies. Curr Med Chem 2012;19:1090–1109.
67. Deupi X, Standfuss J. Structural insights into agonist-induced activation of G-protein-coupled receptors. Curr Opin Struct Biol 2011;21:541–551.
68. Palczewski K. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. Science 2000;289:739–745.
69. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX. The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature 2011;471:656–660.
70. Xue L, Rovira X, Scholler P, Zhao H, Liu J, Pin J-P, Rondard P. Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. Nat Chem Biol 2015;11:134–140.
71. Bachmanov AA. Genetic Architecture of Sweet Taste. In: Weerasinghe DK, DuBois GE, editors. Sweetness and Sweeteners. Washington, DC: American Chemical Society; 2008. pp 18–47.

Article 2:

Sweetness prediction of natural compounds.

J.B. Chéron, J. Golebiowski, S. Antonczak, S. Fiorucci*.

Food. Chem. ***2017****, 221, 1421-1425.*



Food Chemistry 221 (2017) 1421–1425

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Short communication

Sweetness prediction of natural compounds

Jean-Baptiste Chéron, Iuri Casciuc, Jérôme Golebiowski, Serge Antonczak, Sébastien Fiorucci *

Université Côte d'azur, CNRS, Institut de Chimie de Nice UMR7272, 06108 Nice, France

ARTICLE INFO

Article history:
Received 30 June 2016
Received in revised form 11 October 2016
Accepted 25 October 2016
Available online 3 November 2016

Keywords:
Sweeteners
Structure-activity relationship
Chemical space
Natural compounds

ABSTRACT

Based on the most exhaustive database of sweeteners with known sweetness values, a new quantitative structure-activity relationship model for sweetness prediction has been set up. Analysis of the physico-chemical properties of sweeteners in the database indicates that the structure of most potent sweeteners combines a hydrophobic scaffold functionalized by a limited number of hydrogen bond sites (less than 4 hydrogen bond donors and 10 acceptors), with a moderate molecular weight ranging from 350 to 450 $g \cdot mol^{-1}$. Prediction of sweetness, bitterness and toxicity properties of the largest database of natural compounds have been performed. *In silico* screening reveals that the majority of the predicted natural intense sweeteners comprise saponin or stevioside scaffolds. The model highlights that their sweetness potency is comparable to known natural sweeteners. The identified compounds provide a rational basis to initiate the design and chemosensory analysis of new low-calorie sweeteners.

1. Introduction

Sugars, or saccharides, naturally present in food are nutrients associated with caloric intake. However in industrialized countries, overconsumption of high-calorie foods leads to major risks for metabolic disorders and pathologies such as obesity, cardiovascular diseases and type 2 diabetes (Lustig, Schmidt, & Brindis, 2012). To avoid these risks while preserving the sweetness perception, a large number of low-calorie sweeteners have been identified in natural extracts or chemically synthesized (Bassoli, Borgonovo, & Morini, 2011; Behrens, Meyerhof, Hellfritsch, & Hofmann, 2011; DuBois & Prakash, 2012). However, none of the currently available sweeteners presents the same chemosensory profile of sucrose. For instance, saccharin, the most consumed sugar substitute, is known to combine both a sweet and bitter taste (Kuhn, 2004). Moreover, some low-calorie sweeteners have been the subject of controversies about health and food safety (Suez et al., 2014). In this context, consumer interest in natural high potency sweeteners has grown spectacularly in recent years but only few of them are commonly used in the food industry. For instance, stevia is one of the fastest-growing sweeteners in the sugar substitute market. However, the identification of potent sweeteners in plants has always stemmed from an empirical knowledge of ethnobiological observations (Kinghorn & Soejarto, 1989) and a systematic study of natural compounds to detect new sweeteners has never been carried out. The search for novel sweeteners inspired by nature is still relevant.

Quantitative structure activity relationship (QSAR) modelling is a widely accepted computational method to predict biological activities from molecular structures (Cherkasov et al., 2014). Such ligand based approaches have been successfully applied to establish a link between the structure of a compound and its sweet taste. The first SAR and QSAR models date from the 60s and authors found that hydrophobicity is a crucial physico-chemical property for explaining the sweetness of a molecule. The first 3D approaches showed that two or three pharmacophoric structural features (hydrogen bonds, and a hydrophobic site) are shared by several families of sweeteners (Deutsch & Hansch, 1966; Kier, 1972; Shallenberger & Acree, 1967). A more comprehensive pharmacophore model highlighting eight interaction sites allowed a new class of sweet molecules derived from a guanidinium core to be discovered (Nofre & Tinti, 1996). More recently, several studies based on machine learning methods involved building statistical models to either predict sweetness (Yang, Chong, Yan, & Chen, 2011; Zhong, Chong, Nie, Yan, & Yuan, 2013) or to discriminate sweet from non-sweet molecules (Rojas et al., 2016). However, a systematic examination of natural compounds to detect new sweeteners has never been performed.

For the first time, virtual screening of a large database of natural compounds has been performed to identify new sweeteners. A QSAR model for sweetness prediction has been set up based on the most exhaustive database of compounds with available sweetness values. The model also incorporates additional filters to remove compounds with undesirable properties, such as bitter

* Corresponding author.
E-mail address: sebastien.fiorucci@unice.fr (S. Fiorucci).

http://dx.doi.org/10.1016/j.foodchem.2016.10.145
0308-8146/

off-taste and toxicity. The model highlights that the most potent natural sweeteners belong to the terpene family and predicts that their sweetness potency is comparable to known natural sweeteners. The identified sweeteners provide a good starting point for the rational design of an ideal low-calorie sweetener.

2. Materials and methods

2.1. Sweetener database

Three hundred and sixteen compounds belonging to seventeen chemical families (Fig. 1a) have been collected in a database of compounds with known sweetness values, ranging from 0.20 to 225,000 (Choi, Hussain, Pezzuto, Kingborn, & Morton, 1989; Zheng, Rao, & Jia, 2003; Zhong et al., 2013) (Table S1), which has been named SweetenersDB. The relative sweetness is defined as the concentration ratio between a given compound and the sucrose used as a reference that is perceived to have the same sweetness intensity. According to many biological processes (Koch, 1966) and due to the skewness of the sweetness value distribution, the sweetness values were converted into a base 10 logarithm, and denoted as log*S*. To set up a fast and accurate QSAR model, two strategies were employed to describe the structure of the molecules. Firstly, the compounds were stored in a 2D database using their smiles, and secondly, their 3D structures were generated by Marvin (*Marvin 14.9.29.0, 2014, Chemaxon*) and the three lowest energy conformers were selected. The protonation state was defined at pH 6.5 corresponding to a physiological salivary pH value (Aframian, Davidowitz, & Benoliel, 2006). Dragon descriptors (TALETE srl. Dragon, 2014) were computed for the 2D and 3D database. The descriptors of the 3D database were averaged by Boltzmann weighting. Each descriptor of the final matrix was normalized and redundant descriptors were removed (correlation greater than 0.9). The final matrices contain 244 and 265 descriptors from the 2D and 3D initial matrices, respectively.

2.2. QSAR model

Two supervised machine learning methods, Random Forest (Breiman, 2001) (RF) and Support Vector Regression (SVR) (Cortes & Vapnik, 1995), were assessed to select the fastest and the most accurate algorithm to predict the sweetness of a molecule. SweetenersDB was randomly split into a training set of 225 molecules and a test set of 91 molecules. Only molecules of the training set were considered for the optimization of the model, whereas those of the test set were not used to build the model, but rather to assess its performance. 2D and 3D QSAR models were built using the leave-one-out cross-validation methods. For the SVR and RF methods, a brute-force optimization was applied to assess the exhaustive parameter value combinations. The efficiency of the prediction of each approach was measured according to R^2, the squared of the correlation coefficient. It follows that the closer the value is to one, the better the prediction. The chemical space of SweetenersDB and SuperSweet (Ahmed et al., 2011), the largest database of sweet agents, was compared to assess the applicability domain of the QSAR model. The aim of determining the applicability domain is to estimate the uncertainty of the prediction and was used in this study to filter scaffolds that were too far from the chemical space of known sweet molecules. In a second phase, a large database of natural compounds, SuperNatural II (Banerjee et al., 2015), was virtually screened. Additional filters were used to exclude molecules with undesirable properties such as bitter off-taste or toxicity (Fig. 1b, see supporting information for more details).

3. Results

3.1. Chemical diversity of the SweetenersDB

The already known molecules with their associated sweetness values included in the SweetenersDB cover a large chemical space. They are composed of 8–89 heavy atoms including roughly a third of carbon atoms. Their molecular weights range from 120 to 1300 g·mol^{-1}. Despite the fact that the sweeteners should be hydrophilic to be soluble in saliva, the molecules belonging to the phenylpropanoid, polyphenol and terpene families also exhibit hydrophobic properties. As expected, the other chemical families (*e.g.* saccharides, polyols and amino acids) in the database share hydrophilic properties. The average octanol/water partition coefficient (log*P*) of 0.0 ± 2.1 across all of the compounds is in line with

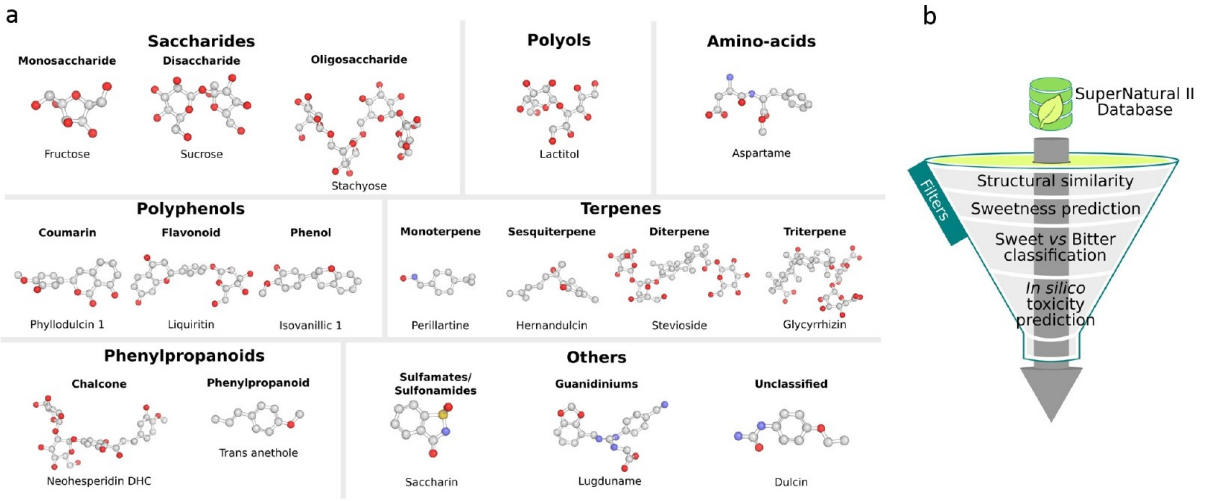


Fig. 1. (a) Molecular structure families of sweet compounds used in the QSAR model. Carbon, oxygen, nitrogen and sulfur atoms are coloured in grey, red, blue and yellow respectively. (b) Virtual screening workflow of the SuperNatural II database. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the broad diversity of the physico-chemical properties observed in the database. Fig. 2a and b shows that the number of hydrogen bond donors and acceptors of the highest potent sweeteners do not exceed 4 and 10 respectively. Analysis of the physico-chemical properties of the sweeteners reveals that the sweetness value ($\log S$) of a molecule is poorly correlated with most of the molecular descriptors except for the molecular weight (MW) and the partition coefficient (Fig. 2c and d). The correlation of $\log S$ and $\log P$ follows a linear tendency up to a maximal sweetness value corresponding to a $\log P$ value ranging from 3 to 3.5. The same observation can be reported considering MW up to a value ranging from 350 to 450 $g \cdot mol^{-1}$. This analysis reveals that the expected properties of an intense sweetener correspond to a moderate MW and a hydrophobic core with a limited number of hydrogen bond sites. Fig. 2 shows that natural sweeteners (highlighted in green in Fig. 2) are the heaviest compounds in the SweetenersDB and show the largest number of hydrogen bond donors and acceptors. Their physico-chemical properties exceed the limits defined above for the most intense sweeteners and are in line with their sweetness value which does not exceed 1000 ($\log S < 3$). This strengthens the idea that more potent sweeteners remain to be discovered.

3.2. Performance of QSAR models

To assess the predictability of the QSAR model the R^2 coefficient of determination has been calculated for each approach. Satisfying correlations between theoretical and experimental $\log S$ are observed for both SVR and RF (Fig. 3) meaning that the QSAR models, beyond being descriptive, are relevant to predict sweetness values. However, the coefficient of determination R^2 appears to be slightly better for SVR than RF when applying QSAR models to the molecules of the test set. RF models generally underestimate high $\log S$ values and the predictions reach a plateau for intense sweeteners. A similar performance ($R^2 > 0.8$) of the 2D and 3D SVR models is observed. The computation of the 3D conformers and descriptors is much more time consuming than for the 2D ones. As the strategy is to set up a fast and accurate QSAR model, the 2D SVR approach has been chosen to virtually screen a large database of natural compounds and identify new putative sweeteners (Table S2).

3.3. Comparison of the chemical spaces

Various chemical spaces have been compared to estimate the applicability domain of the QSAR model, i.e. to evaluate the uncertainty of the sweetness prediction. On one hand, the chemical space of SweetenersDB has been compared to the most exhaustive database of sweet agents, SuperSweet (Ahmed et al., 2011). As shown in Fig. 4 (right y axis, black dots), more than 99.5% of the compounds from the SuperSweet database are structurally close to a representative structure in the SweetenersDB. The chemical space of the SweetenersDB fully encompasses the structural diversity of SuperSweet. On the other hand, the chemical spaces of SuperNatural II and SweetenersDB have also been compared (Fig. 4, grey dots left y axis) to remove natural compounds sharing

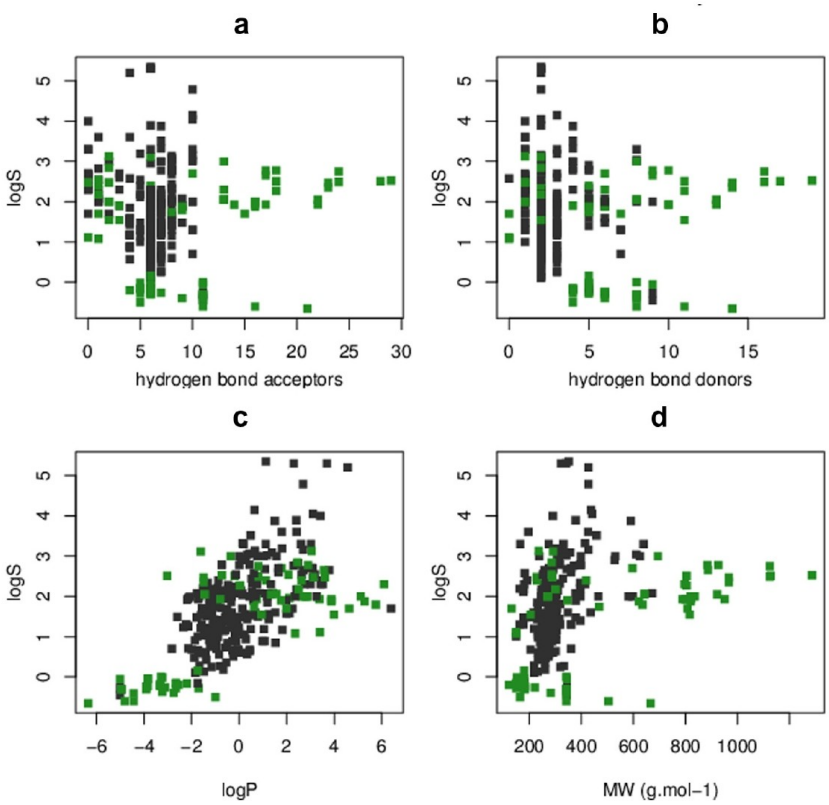


Fig. 2. Analysis of the physico-chemical properties of compounds in SweetenersDB: (a) number of hydrogen bond acceptors and (b) donors, (c) octanol/water partition coefficient ($\log P$), (d) molecular weight (MW). Natural compounds are highlighted in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

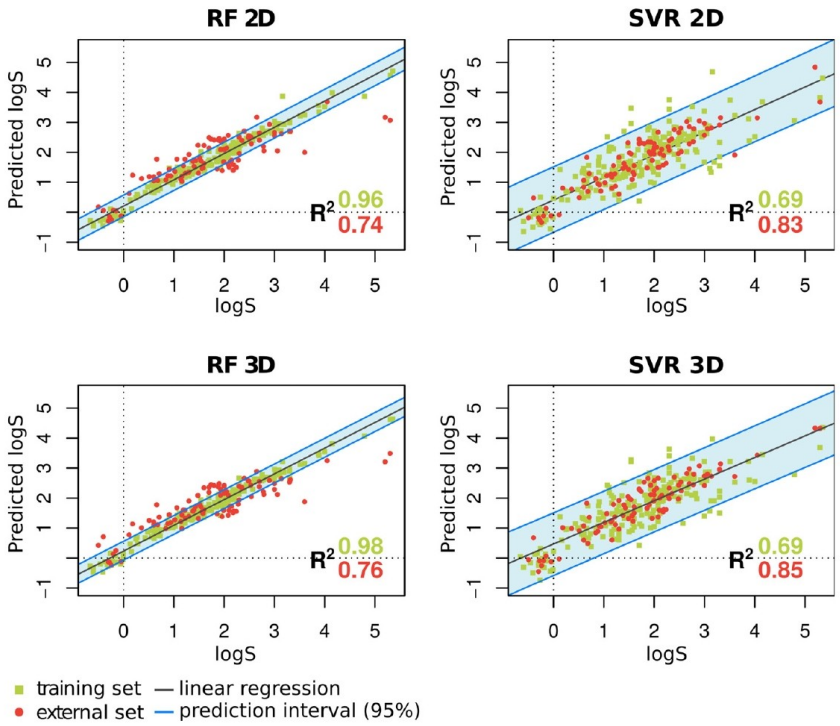


Fig. 3. Performance of sweetness prediction using Random Forest (RF) and Support Vector Regressions (SVR).

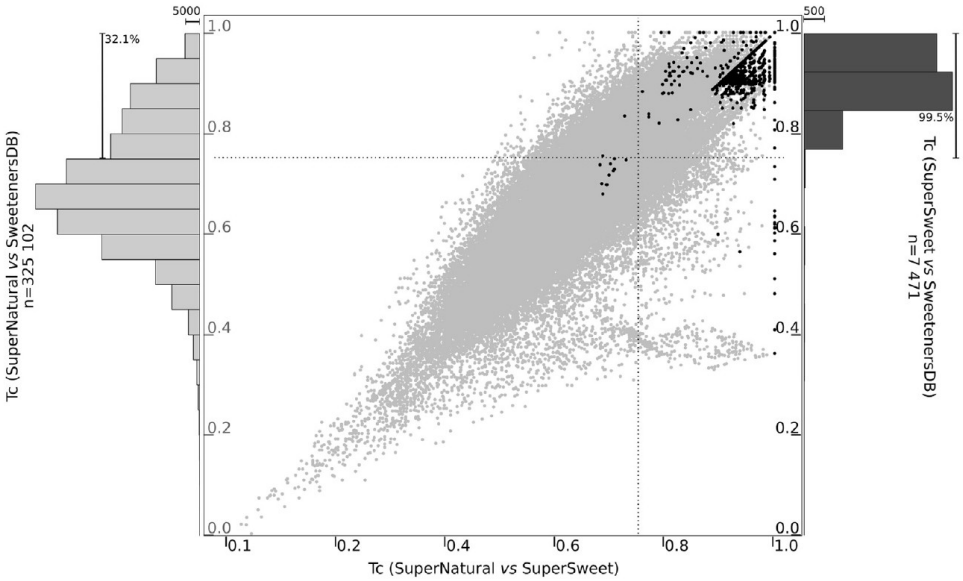


Fig. 4. Comparison of the chemical spaces contained in SuperSweet, SuperNatural II and SweetnersDB based on the Tanimoto coefficient (Tc) and pubchem fingerprints. Black and grey dots correspond to a molecular structure in the SuperSweet and SuperNatural II databases respectively. A Tc coefficient greater than 0.75 means that a compound is structurally close to a given chemical space.

low similarity with sweeteners used to build the QSAR model. As expected, SuperNatural II contains a very high diversity of structure with respect to the SweetenersDB. 105,620 of the 325,102 natural compounds belong to both chemical spaces, meaning that roughly 34% of SuperNatural II have been virtually screened to identify new putative sweeteners.

3.4. Prediction of natural sweeteners

More than 75,000 compounds within the applicability domain of the QSAR model applied to the SuperNatural II database exhibit a significantly high sweetness value ($\log S > 2$). Additional filters were applied to avoid undesirable properties. First, a classification

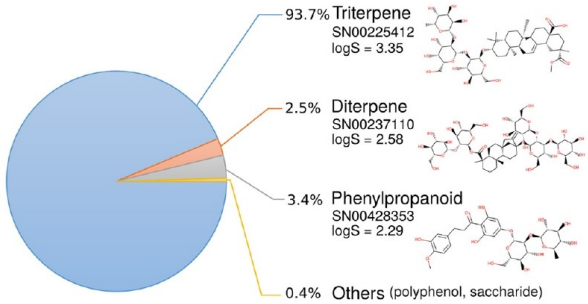


Fig. 5. Classification of natural compounds predicted as sweet by the QSAR model. A representative structure and its predicted sweetness value is shown for each of the most important chemical families.

model has been designed to discriminate compounds that share bitter and sweet tastes (Supporting information). Among the 75,000 natural molecules, 21% are unambiguously classified in the cluster of sweet taste and not in the bitter one. Secondly, molecules that contain undesirable scaffolds involved in toxicity problems have also been excluded. The final dataset, available online (http://chemosim.unice.fr/SweetenersDB), contains 4585 natural compounds (Table S2). Surprisingly, an overwhelming majority of the identified natural sweeteners belongs to the terpene family and less than 200 molecules are saccharides, polyphenols or phenylpropanoids (Fig. 5). It is noteworthy that the most potent natural sweeteners exclusively derive from the saponin and stevioside scaffolds. They present a relative sweetness 1000–10,000 times more intense than sucrose, in line with the natural sweeteners already characterized (Behrens, Meyerhof, Hellfritsch, & Hofmann, 2011; Kinghorn & Soejarto, 1989).

Their average MW of $950 \pm 158\ g{\cdot}mol^{-1}$ and the number of hydrogen bond donors and acceptors of 10 ± 3 and 19 ± 5, respectively, correspond to the upper bound of already known sweeteners. The physico-chemical properties of these glycosylated terpene cores go beyond the limits described above for the most intense sweeteners, i.e. a chemical structure that combines a hydrophobic scaffold functionalized by a limited number of hydrogen bond sites with a moderate MW. It follows that it might be difficult to discover a more intense natural sweetener than the existing ones. However, the sweeteners in the database provided can be rationally modified to enhance their sweetness potency by borrowing the hit-to-lead process applied in pharmaceutical research. Determination of the molecular structure of the sweet taste receptor will be a significant step forward to understand how a sweetener is perceived by the human body. Pending an answer to this question, the quest for an ideal low-calorie sweetener remains open and the newly identified sweeteners constitute a starting point for rational design and receptor-based screening experiments.

Acknowledgments

This work was supported by the French Ministry of Higher Education and Research [PhD Fellowship], by GIRACT (Geneva, Switzerland) [Promoting Flavor Research first year bursary award] and the Gen Foundation (Registered UK Charity No. 1071026) [a charitable trust which principally provides grants to students/researchers in natural sciences, in particular food sciences/technology].

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodchem.2016.10.145.

References

Aframian, D., Davidowitz, T., & Benoliel, R. (2006). The distribution of oral mucosal pH values in healthy saliva secretors. *Oral Diseases, 12*(4), 420–423. http://dx.doi.org/10.1111/j.1601-0825.2005.01217.x.

Ahmed, J., Preissner, S., Dunkel, M., Worth, C. L., Eckert, A., & Preissner, R. (2011). SuperSweet – a resource on natural and artificial sweetening agents. *Nucleic Acids Research, 39(Database)*, D377–D382. http://dx.doi.org/10.1093/nar/gkq917.

Banerjee, P., Erehman, J., Gohlke, B.-O., Wilhelm, T., Preissner, R., & Dunkel, M. (2015). Super Natural II – a database of natural products. *Nucleic Acids Research, 43*(D1), D935–D939. http://dx.doi.org/10.1093/nar/gku886.

Bassoli, A., Borgonovo, G., & Morini, G. (2011). Lost and found in sweeteners: Forgotten molecules and unsolved problems in the chemistry of sweet compounds. *Flavour and Fragrance Journal, 26*(4), 269–273. http://dx.doi.org/10.1002/ffj.2064.

Behrens, M., Meyerhof, W., Hellfritsch, C., & Hofmann, T. (2011). Sweet and umami taste: Natural products, their chemosensory targets, and beyond. *Angewandte Chemie International Edition, 50*(10), 2220–2242. http://dx.doi.org/10.1002/anie.201002094.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., ... Tropsha, A. (2014). QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry, 57*(12), 4977–5010. http://dx.doi.org/10.1021/jm4004285.

Choi, Y.-H., Hussain, R. A, Pezzuto, J. M., Kingborn, A. D., & Morton, J. F. (1989). Abrusosides A-D, four novel sweet-tasting triterpene glycosides from the leaves of Abrus precatorius. *Journal of Natural Products, 52*(5), 1118–1127. http://dx.doi.org/10.1021/np50065a032.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. http://dx.doi.org/10.1007/BF00994018.

Deutsch, E. W., & Hansch, C. (1966). Dependence of relative sweetness on hydrophobic bonding. *Nature, 211*(5044), 75. http://dx.doi.org/10.1038/211075a0.

DuBois, G. E., & Prakash, I. (2012). Non-caloric sweeteners, sweetness modulators, and sweetener enhancers. *Annual Review of Food Science and Technology, 3*(1), 353–380. http://dx.doi.org/10.1146/annurev-food-022811-101236.

Kier, L. B. (1972). A molecular theory of sweet taste. *Journal of Pharmaceutical Sciences, 61*(9), 1394–1397. http://dx.doi.org/10.1002/jps.2600610910.

Kinghorn, A. D., & Soejarto, D. D. (1989). Intensely sweet compounds of natural origin. *Medicinal Research Reviews, 9*(1), 91–115. http://dx.doi.org/10.1002/med.2610090105.

Koch, A. L. (1966). The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology, 12*(2), 276–290. http://dx.doi.org/10.1016/0022-5193(66)90119-6.

Kuhn, C. (2004). Bitter taste receptors for saccharin and acesulfame K. *Journal of Neuroscience, 24*(45), 10260–10265. http://dx.doi.org/10.1523/JNEUROSCI.1225-04.2004.

Lustig, R. H., Schmidt, L. A., & Brindis, C. D. (2012). Public health: The toxic truth about sugar. *Nature, 482*(7383), 27–29. http://dx.doi.org/10.1038/482027a.

Nofre, C., & Tinti, J.-M. (1996). Sweetness reception in man: The multipoint attachment theory. *Food Chemistry, 56*(3), 263–274. http://dx.doi.org/10.1016/0308-8146(96)00023-4.

Rojas, C., Ballabio, D., Consonni, V., Tripaldi, P., Mauri, A., & Todeschini, R. (2016). Quantitative structure–activity relationships to predict sweet and non-sweet tastes. *Theoretical Chemistry Accounts, 135*(3), 66. http://dx.doi.org/10.1007/s00214-016-1812-1.

Shallenberger, R. S., & Acree, T. E. (1967). Molecular theory of sweet taste. *Nature, 216*(5114), 480–482. http://dx.doi.org/10.1038/216480a0.

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C. A., Maza, O., ... Elinav, E. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature, 514*(7521), 181–186. http://dx.doi.org/10.1038/nature13793.

TALETE srl. Dragon, 2014. *Software for Molecular Descriptor Calculation*.Version 6.0.

Yang, X., Chong, Y., Yan, A., & Chen, J. (2011). In-silico prediction of sweetness of sugars and sweeteners. *Food Chemistry, 128*(3), 653–658. http://dx.doi.org/10.1016/j.foodchem.2011.03.081.

Zheng, J.-X., Rao, Z.-J., & Jia, C.-X. (2003). Study on the relationship between structure and sweetness of sucrose derivatives. *Food Science, 24*(5), 29–33. Retrieved from: <http://124.205.222.100/Jwk_spkx/EN/Y2003/V24/I5/29>.

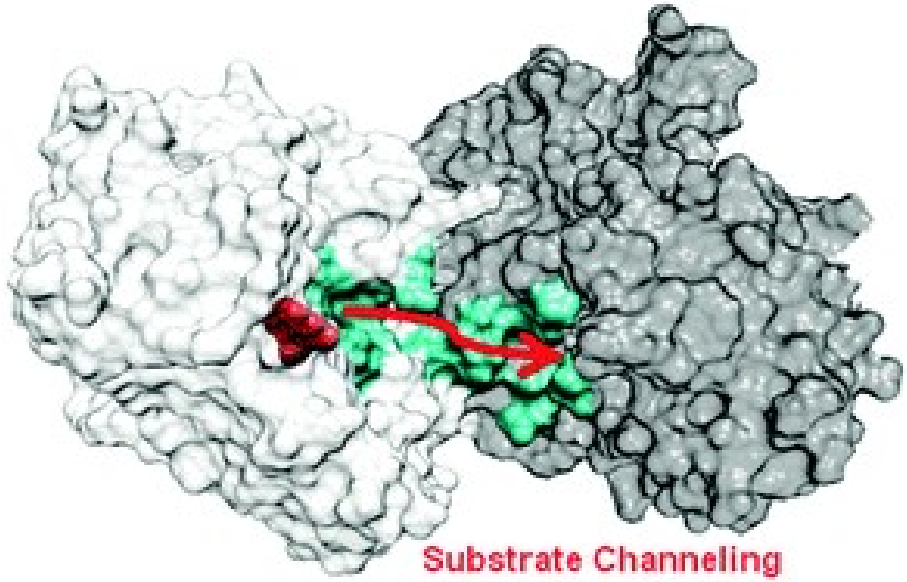
Zhong, M., Chong, Y., Nie, X., Yan, A., & Yuan, Q. (2013). Prediction of sweetness by multilinear regression analysis and support vector machine. *Journal of Food Science, 78*(9), S1445–S1450. http://dx.doi.org/10.1111/1750-3841.12199.

Article 3 :

Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channeling in the course of flavonoid biosynthesis.

J. Diharce, J. Golebiowski, S. Fiorucci*, S. Antonczak.

Phys. Chem. Chem. Phys., ***2016*** *18, 10337-10345.*



PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 10337

Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channelling in the course of flavonoid biosynthesis†

Julien Diharce, Jérôme Golebiowski, Sébastien Fiorucci* and Serge Antonczak*

Received 25th August 2015,
Accepted 10th March 2016

DOI: 10.1039/c5cp05059f

www.rsc.org/pccp

In the course of metabolite formation, some multienzymatic edifices, the so-called metabolon, are formed and lead to a more efficient production of these natural compounds. One of the major features of these enzyme complexes is the facilitation of direct transfer of the metabolite between enzyme active sites by substrate channelling. Biophysical insights into substrate channelling remain scarce because the transient nature of these macromolecular complexes prevents the observation of high resolution structures. Here, using molecular modelling, we describe the substrate channelling of a flavonoid compound between DFR (dihydroflavonol-4-reductase) and LAR (leucoanthocyanidin reductase). The simulation presents crucial details concerning the kinetic, thermodynamic, and structural aspects of this diffusion. The formation of the DFR–LAR complex leads to the opening of the DFR active site giving rise to a facilitated diffusion, in about 1 μs, of the DFR product towards LAR cavity. The theoretically observed substrate channelling is supported experimentally by the fact that this metabolite, *i.e.* the product of the DFR enzyme, is not stable in the media. Moreover, along this path, the influence of the solvent is crucial. The metabolite remains close to the surface of the complex avoiding full solvation. In addition, when the dynamic behaviour of the system leads to a loss of interaction between the metabolite and the enzymes, water molecules through bridging H-bonds prevent the former from escaping to the bulk.

1. Introduction

Currently, studies of metabolites produced in plants, including their characterization, localization, transport and production, have become the key point for better understanding their function. Beyond the investigations that focus on primary metabolic fluxes,[1,2] there has recently been a growing interest in the analyses of such data concerning secondary metabolites. The great molecular diversity of these compounds arises from the numerous enzymes involved in their biosynthesis. Recently, it has been proposed that these enzymes may not act independently but could associate to form active multienzymatic complexes that could more efficiently lead to the formation of metabolites. The initial proposal of the existence of such complexes, the so-called metabolons, emerged many years ago.[3,4] Two important features characterize these complexes: (i) the proteins associate transiently and (ii) when the complex is formed, the metabolites diffuse from one active site to the next active site through the so-called substrate channelling. Information regarding the nature, in terms of energy and structure, of such protein–protein complexes bound to a substrate, as well as details concerning the way the metabolite is transferred, remains scarce. Accessing such information will aid in deciphering the role of both the structure of the complex and the surrounding solvent molecules in the optimization of this intriguing molecular machinery.

Substrate channelling plays a central role in the regulation of metabolic flux[5,6] and presents numerous benefits such as the protection of unstable substrates and isolation of intermediates from competing reactions. Substrate channelling should also avoid long diffusing processes and lead to energy saving by circumventing metabolite solvation and desolvation events.[7–10] Numerous examples of this phenomenon have been reported, including tightly bound multienzymatic complexes[11,12] and assemblies of transient nature.[13,14] Two types of mechanisms have been proposed to describe such diffusion processes: a direct tunnelling of the metabolites confined within the protein matrix[15] or a channelling of the reaction intermediates at the surface of the protein complexes guided by electrostatic steering.[16,17] These studies focused primarily on the electrostatic properties of the channel or the proximity of the catalytic sites.[18]

Institut de Chimie de Nice, UMR-CNRS 7272, Faculté des Sciences, Université de Nice-Sophia Antipolis, 28 Avenue Valrose, 06108 Nice Cedex 2, France. E-mail: Serge.Antonczak@unice.fr, Sebastien.Fiorucci@unice.fr

† Electronic supplementary information (ESI) available: Structural analysis of DFR–LAR complexes and DFR–LAR docking solutions. Experimental free energy of binding *vs.* ATTRACT score; egress routes from DFR and LAR active sites; description of DFR and LAR active sites; electrostatic pattern during the diffusion process; and experimental procedures. See DOI: 10.1039/c5cp05059f

Finally, the gain in both kinetic[19] and thermodynamic aspects of this specific diffusion process still needs to be estimated.

Among these secondary metabolites, flavonoids are regarded as a prototypical case for metabolic studies due to their substantial molecular diversity.[20] Thus, numerous enzymes are involved in their biosynthesis, and their expression needs to be subtly regulated to manage the production of flavonoids within the cell. Each individual step constituting these enzymatic cascades appears to be currently well understood;[21] however, it has been proposed that these enzymes may not act independently but could associate to form a transient multienzymatic complex that more efficiently orchestrates the formation of the metabolites.[7,9,22–25] It is assumed that such a putative metabolon could be formed around a cytochrome P450 type enzyme anchored to the endoplasmic reticulum membrane, here the flavone 3′ hydroxylase enzyme (F3′H).[7,24] Many other enzymes could aggregate on F3′H, and at least three different types of multienzymatic complexes can be formed depending on the flavonoid sub-classes to be produced, *e.g.*, flavonols, anthocyanidins or proanthocyanidins. In these transient structures, the dihydroflavonol-4-reductase (DFR) enzyme is considered a hub for the biosynthesis of anthocyanin or procyanidin molecules.[26]

During the process of proanthocyanidin formation, the step following DFR action is catalyzed by the leucoanthocyanidin reductase enzyme (LAR).[27] Thus, in a scheme based on the existence of a metabolon, a dihydroquercetin type substrate (DHQ) is reduced in the DFR active site with the help of an NADPH cofactor, which is subsequently oxidized. The product of this enzymatic reaction, namely leucocyanidin (LCC), has to diffuse to the active site of the LAR enzyme where it acts as a substrate and undergoes reduction with the help of another NADPH cofactor (see Fig. 1).

Concerning protein–protein associations, some recent studies have demonstrated the existence and the importance of such interactions in the biosynthetic pathway of flavonoids[23,28] and, more generally speaking, has been the central topic of a series of review articles.[29] Despite recent insights on the existence of a metabolon in the course of flavonoid production, experimental evidence, at the molecular level, of a flavonoid compound shuttling between connected enzymes is lacking. However, LCC is known to be unstable in water. By protecting it from solvent effects during substrate channelling, the multi-enzymatic complex would optimize the metabolic flux.

To fully address the atomistic details underlying the molecular mechanisms of macromolecular complexes, high resolution techniques such as NMR spectroscopy are required. If high resolution approaches fail for such transient edifices (basically because single crystals of sufficient size cannot be obtained or if the size of the system is too large for X-ray and NMR experiments, respectively), molecular modelling methods are perfectly suited to act as a computational microscope. In the present work, performing protein–protein docking calculations and brute force molecular dynamics simulations reaching a micro-second time scale we present molecular and dynamic views of metabolite channelling between two enzymes involved in flavonoid biosynthesis, namely LAR and DFR.

2. Computational details

2.1. Initial coordinates and system setup

The initial structure of dihydroflavonol-4-reductase (DFR) was extracted from the crystal structure of a ternary complex with the co-factor NADP+ and the substrate of the enzymatic reaction, dihydroquercetin (PDB identifier 2C29).[26] The substrate was then substituted by the product and NADP+ was conserved. The stability of the newly formed complex has been improved through long molecular dynamics simulations. The initial structure of leucoanthocyanin reductase (LAR) was also extracted from the crystal structure of a ternary complex with both the co-factor and the enzyme product (NADPH and catechin, respectively; PDB identifier 3I52).[27] The position of the internal loop defined by residues 43–53 and both N- and C-terminal regions (respectively, defined by positions 1–10 and 318–347) of LAR are not present in the crystal structure.

Because too many residues were missing, a complete model of LAR was prepared. No specific secondary structure was predicted by the PSIPRED[30] server for both the N-terminal

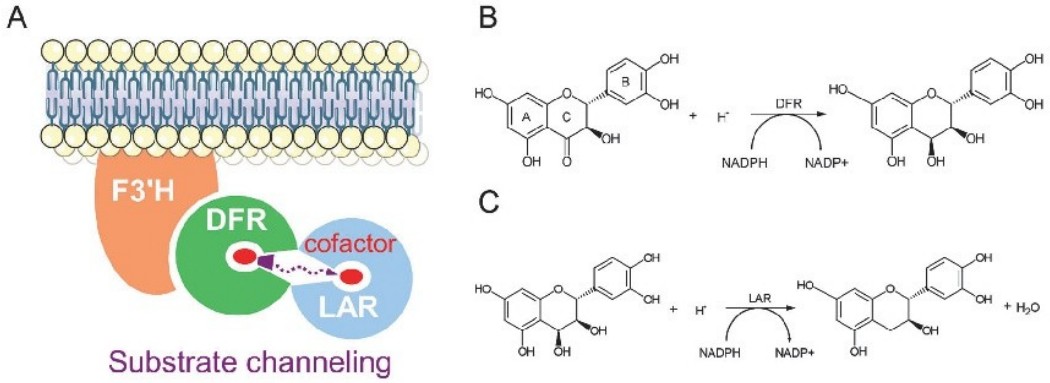


Fig. 1 Schematic view of a subset of the multi-enzymatic complex involved in flavonoid biosynthesis. The product of the DFR enzymatic reaction diffuses to the LAR active site where it will be converted to the following metabolite (A). Enzymatic reactions catalyzed by DFR and LAR, respectively, (B and C).

and the internal loop, but a short alpha helix was predicted at the C-terminal moiety (amino-acids 325–330). This helix was placed at the enzyme surface by minimizing the helix–enzyme interaction energy using a protein–protein docking protocol. Docking solutions were filtered following a distance constraint of 9 Å between the residue 325 and the latest crystallized residue (K317). The position of the helix 325–330 interacting with other helices (defined by residues 54–60 and 20–34) and associated with the lowest interaction energy were selected for the definitive LAR structure. Missing loops (positions 1–10, 43–53, 317–325 and 331–347) were built using Modeller.[31] The complete model of LAR was subjected to structural validation using PROCHECK[32] prior to protein–protein docking.

2.2. Protein–protein docking simulations

The ATTRACT docking protocol, which was detailed in Fiorucci *et al.*[33] and recognized in recent CAPRI rounds for the prediction of high quality models of protein–protein interactions, was applied to build the complex structure of DFR–LAR. Docking calculations take into account surface and electrostatic complementarity, as well as physico-chemical properties of interacting amino acids. A coarse-grained representation of the protein affords the energy minimization of thousands of starting configurations. A total of 436 initial positions were regularly distributed all around DFR, and 258 different orientations of LAR were generated for each starting point, leading to more than 110 000 docking solutions. Clustering analysis was performed to group similar docking poses (RMSD < 1 Å).

2.3. Force field parameters for NADPH/NADP+/LCC

NADPH and NADP+ parameters were taken from Holmberg *et al.*[34] Leucocyanidin (LCC) parameters were calculated using the Antechamber module of AMBER12 and following the RESP method for the atomic charges. Quantum-chemical calculations were performed using Gaussian03 at the RHF/6-31G* level of theory.

2.4. Molecular dynamics simulation protocol

Each simulation consists of a fully atomistic detailed, brute force MD simulation using amber ff03.r1 force field of the DFR–LAR complex previously determined by the ATTRACT docking protocol. The molecular systems were considered in the presence of one LCC molecule (located initially in the DFR active site) and the co-factors in both active sites, $NADP^+$ in DFR and NADPH in LAR cavities. The protonation state of titratable residues were determined at a physiological pH of 6.5 through the H++ server.[35] The system was neutralized by addition of 18 Na^+ ions located in the most negative region of space using the LEAP module of AMBER12.[36] Crystallographic water molecules were retained and the TIP3P solvent phase was extended to a distance of 8 Å from any solute atom leading to a box whose dimensions were $107 \times 95 \times 110$ or $115 \times 112 \times 88$ Å^3, depending on the considered DFR–LAR system (PP1 or PP2). MD calculations were performed using the PMEMD.cuda module of the AMBER12 program in the isothermal–isobaric thermodynamic ensemble at 300 K and 1 atm. The SHAKE algorithm was applied for bonds involving hydrogen atoms. A time step of 2 fs was chosen. An 8 Å cut-off was applied to non-bonded van der Waals interactions and the non-bonded pair list was updated every 15 steps. After the addition of the ions and the water molecules to the minimized complexes, 15 000 steps of minimization and 30 ps of molecular dynamics (MD) simulations using a restraint of 20 kcal mol^{-1} Å^{-2} on the solute atoms were performed, followed by four rounds of 15 000 steps minimization reducing the restraints by 5 kcal mol^{-1} Å^{-2} at each round, with 30 ps MD simulation. Particle mesh Ewald (PME) summation with GPU acceleration was used, and PME parameters were chosen to obtain a grid spacing close to 1 Å and a 9 Å direct space cut-off. The pressure and temperature coupling constants were fixed to 0.4 ps and 1 ps, respectively. Furthermore, the system was slowly heated from 100 to 300 K over a period of 50 ps. The equilibration was carried out for 5 ns after the 50 ps of heating. Production runs, with a timescale ranging from 150 ns to 1.2 μs, were performed in the *NPT* ensemble for each system. In total, we performed 8 simulations with a total duration of more than 5 μs. Structural analysis was performed using the ptraj module of AMBER12.

Accelerated molecular dynamics (aMD)[37] was used to enhance the sampling of the system in the vicinity of LAR cavity. To overcome conformational barriers, a bias was introduced to the dihedral angles and the total potential energy. A boost of 3.5 kcal mol^{-1} per residue was applied to the dihedral energy, and a threshold of 20% to the total energy potential.

2.5. Enzyme–substrate free energy calculations

The molecular mechanics generalized Born surface area (MM-GBSA) method[38] supplied with AMBER is based on the analysis of configurations obtained from equilibrated MD. The free energy calculation is a post-processing analysis of the trajectory of the complexes. The binding free energy is determined by the following equation:

$$\Delta G_{\text{bind}} = \langle G_{\text{complex}} \rangle - \langle G_{\text{receptor}} \rangle - \langle G_{\text{ligand}} \rangle$$

where $\langle G_x \rangle$ corresponds to the average of the total free energy of the component x over snapshots taken from the MD trajectory. An interesting feature of such a description of the energy lies in its capability to be decomposed on a per-residue basis because it is computed through additive potentials.

The total free energy of a system is expressed as the sum of several contributions:

$$G = E_{\text{MM}} + H_{\text{trans/rot}} + G_{\text{sol}} - \text{TS}$$

where E_{MM} is the average molecular mechanical energy, G_{sol} is the solvation free energy and $H_{\text{trans/rot}}$ is the energetic term corresponding to the six translational and rotational degrees of freedom ($H_{\text{trans/rot}} = 3kT \sim 1.8$ kcal mol^{-1} at 300 K). The solvation free energy is computed as a sum of polar and non-polar contributions. The non-polar contribution is due to cavity formation and van der Waals interactions between the solute and the solvent, which are typically calculated from the solvent-accessible surface area. The polar contribution of G_{sol} is obtained

following the generalized Born model (OBC model II[39]) available in AMBER.

In this article, the term $T\Delta S$ was estimated by a normal mode analysis. This method requires a system in equilibrium and the entropic contribution was exclusively calculated on initial and final states, for which it is roughly equal to -23 kcal mol^{-1}.

3. Results and discussion

3.1. Substrate channelling requires optimal conformation of the enzyme–enzyme complex

To predict the structure of the DFR–LAR complex, a virtual screening of protein–protein interfaces was performed using the ATTRACT docking program.[33] (Fig. S1 and Table S1, ESI†) The top 10 poses exhibit interaction energy typical for a weak protein–protein interaction (Fig. S2, ESI†). The broad spectrum of structures associated with different monomer orientations in conjunction with a narrow range of weak interaction energies was actually representative of a flat energy landscape, reflecting the transient nature of the complex.

Because the substrate has to diffuse from the DFR active site to the LAR one, only the protein–protein structures in which a possible pathway between cavities is present should be considered (Fig. S3, ESI†). Considering that the entrances should not be too far from one another, two poses were selected. These, noted PP1 and PP2, are ranked 4th and 7th and exhibit a similar interaction energy, corresponding to an estimated dissociation constant in the micromolar range. Analysis of the protein–protein interface highlights that both complexes (PP1 and PP2) satisfy the structural requirements of transient protein–protein complexes:[40] an interface ranging from 1000 to 3000 A^2 with a fraction of 50–60% of non-polar residues, approximately 50 residues are involved at the interface and approximately 10 hydrogen bonds.

Multiple non-constrained molecular dynamics (MD) simulations based on different initial velocities and all of which reached a hundred nanoseconds time scale were performed to sample the diffusion of the substrate (see Table 1). Interestingly, each protein structure undergoes weak structural modifications, as reflected by the relatively low RMSD values, while the complex is expected to bear important reorganizations. This point is not surprising since the refinement of the model, from coarse grain to full atom description, leads necessarily to a reorganisation of the amino-acids at the contact surface.

Considering the metabolite, LCC is initially located in the DFR cavity in contact with the reduced cofactor NADP$^+$. Molecular dynamics simulations on the PP1 complex show large reorganizations at the protein–protein interface that do not allow the substrate channelling phenomenon to occur. Indeed, these calculations lead to either a complete dissociation of the complex or substrate diffusion in the wrong direction. This suggests that an adequate structure of the complex is required for initiating substrate channelling.

Considering simulations applied on the PP2 complex, a diffusion of LCC from the active site of DFR towards that of LAR was observed in each of the four cases. However, the end of this diffusion is closely linked to the behaviour of the protein complex. In one of these cases, the dissociation of the complex occurs after 1.2 μs, while LCC has already started to diffuse from the protein–protein interface to the active site of LAR. In the three other cases, the complex remains stable due to the creation of two additional ionic interactions between $E91_{DFR}$ and $K50_{LAR}$ on one side and $E98_{DFR}$ and $K52_{LAR}$ on the other, creating a new contact area 25 Å away from the initial interface (see Fig. 2 and 3A). This induces a slight allosteric effect, which allows for the diffusion of LCC towards LAR (see Fig. 3B). This conformation is sampled only transiently at the beginning of the first simulation mentioned above, but the loss of these ionic interactions at the end of the trajectory could explain the dissociation of the complex. This structural refinement, observed in each PP2 simulations, is likely to be induced by the change in the resolution of the model, from coarse-grain (docking protocol) to full atom description.

This PP2 complex structure is stabilized by two areas where DFR and LAR interact, as shown in a representative structure extracted from the trajectory (see Fig. 2 and structure provided as ESI†). The interaction initially predicted by the ATTRACT protocol has slightly evolved but remains mainly hydrophobic (see Fig. 2B). It involves of two zones, $(A155\text{-}K156\text{-}K157\text{-}M158\text{-}T159)_{DFR}$ and $(N169\text{-}L177\text{-}T180\text{-}P265)_{LAR}$ on one side and $(N216\text{-}E217\text{-}A218\text{-}S221\text{-}I222)_{DFR}$ and $(E261\text{-}N262\text{-}I263\text{-}I264\text{-}P265)_{LAR}$ on the other. In this first contact area, some H-bond may be transiently formed but do not constitute the major part of the interactions. *A contrario*, in the second area, the interactions mainly involve charged amino acids leading to strong ionic stabilization (see Fig. 2A) as discussed previously.

In one trajectory out of the three others showing a diffusion of the metabolite at the entry of the cavity of LAR, LCC ultimately diffuses into the enzymatic active site, characterizing

Table 1 Summary of MD simulations and molecular events

MD	Simulation length (μs)	Average RMSD (Å) DFR	LAR	Complex	Molecular event
PP1 (1st)	0.055	1.9	1.7	6.9	[a]
PP1 (2rd)	0.16	2.2	2.0	4.0	[b]
PP1 (3rd)	0.49	2.5	1.8	3.7	[b]
PP1 (4th)	0.36	2.7	1.7	3.8	[c]
PP2 (1st)	1.20	3.0	2.0	7.5	[d]
PP2 (2rd)	0.94	2.2	1.9	6.2	Substrate channelling
PP2 (3rd)	0.57	2.6	2.0	8.0	[e]
PP2 (4th)	0.58	2.4	1.8	7.5	[e]

[a] Complete dissociation of the protein–protein complex. [b] Substrate diffusion along the green route shown in Fig. S3A (ESI) for bulk solvent. [c] The metabolite is rapidly trapped close to H137 and Q138 amino-acids located on the green route shown in Fig. S3A (ESI). Strong reorganization of the interface of the proteins prevents any substrate channelling. [d] Diffusion of the metabolite along the red route shown in Fig. S3A (ESI) until it reaches the protein–protein interface and dissociation of the complex after 1.2 μs due to the loss of the ionic interactions $E91_{DFR}$–$K50_{LAR}$ and $E98_{DFR}$–$K52_{LAR}$, see main text for discussion. [e] Creation of the $E91_{DFR}$–$K50_{LAR}$ and $E98_{DFR}$–$K52_{LAR}$ ionic interactions between the two enzymes and diffusion of the metabolite from the DFR active site to the vicinity of the LAR active site.

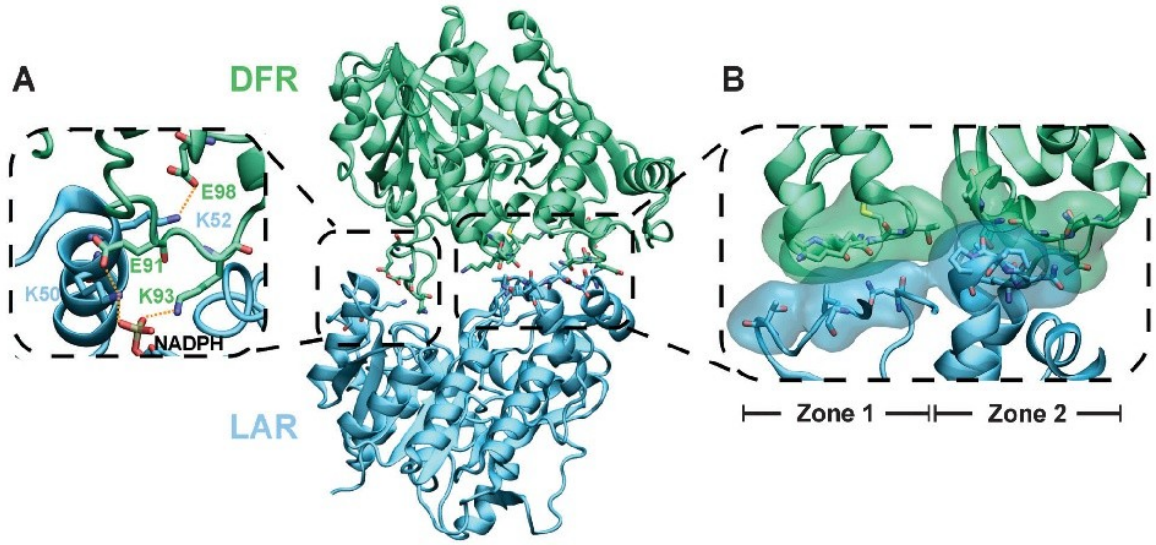


Fig. 2 Details of the interactions occurring at the protein–protein interface of the PP2 complex made up of DFR (green) and LAR (blue) after 155 ns of simulation. The interface is composed of two areas: (A) electrostatic interactions involving charged side chains and (B) hydrophobic contacts involving residues $(A155\text{-}K156\text{-}K157\text{-}M158\text{-}T159)_{DFR}$ and $(N169\text{-}L177\text{-}T180\text{-}P265)_{LAR}$ on one side (zone 1) and $(N216\text{-}E217\text{-}A218\text{-}S221\text{-}I222)_{DFR}$ and $(E261\text{-}N262\text{-}I263\text{-}I264\text{-}P265)_{LAR}$ on the other (zone 2) contribute to the stabilization of the complex.

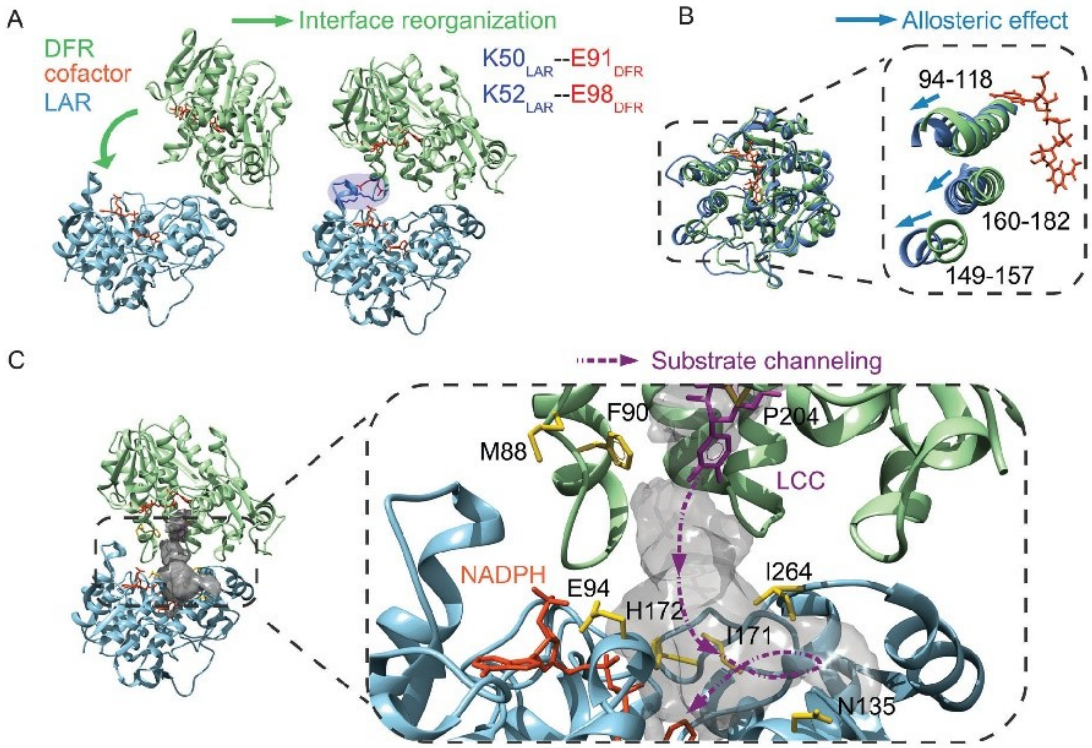


Fig. 3 Molecular events occurring during substrate channelling. The formation of two salt-bridges at the protein–protein (PP2) interface at the beginning of the molecular dynamics simulation (A) leads to the opening of the DFR cavity (B) and to the channelling of the substrate (C, in magenta). The enzymes DFR and LAR and the cofactors NADPH are colored green, blue and orange respectively. Key residues defining the gate of the enzyme cavities are represented in yellow.

a complete substrate channelling event between the active sites of the two enzymes (presented in Fig. 3C).

Finally, these eight simulations can be divided into two families. The four simulations based on the initial PP1 complex

do not lead to the appropriate diffusion of the metabolite. Considering PP2, which undergoes structural rearrangements, the four simulations reproduce the egress of LCC from the DFR active site and the diffusion towards LAR cavity. The difference between these simulations is detailed in the next point.

3.2. The allosteric effect induced by DFR–LAR association modifies the kinetics of the LCC release

Molecular dynamics simulations taking the PP2 complex as an initial structure show that DFR–LAR binding induces a tight conformational change of DFR helices defined by residues D94-K118, D149-K157 and A160-N182 (presented in Fig. 3B). This allosteric effect corresponds to an opening of the cavity, which gives rise to the diffusion of the metabolite out of the active site of DFR in roughly 150 ns. By comparison, considering a DFR–$NADP^+$–LCC system without any protein–protein interaction, this diffusion is observed after about 760 ns. Essential dynamics analysis (EDA),[41] which highlights the principal motions of a molecular system, allowed us to analyse this difference. Considering the protein–protein complex, the first principal motion is associated with the opening of the cavity of DFR, and corresponds to the shift of three helices made up of residues D94-K118, D149-K157 and A160-N182 (presented in Fig. 3B), known to be involved in the metabolite ingress and egress mechanisms (Fig. S3, ESI†). When considering only the DFR/$NADP^+$/LCC complex, EDA performed on long MD simulations revealed that the first principal motion does not involve these helices, although the release of LCC follows the same egress route. This allosteric effect has a key influence on the fluency with which the metabolite leaves the DFR active site and thus on the kinetics of this mechanism. In PP1 complexes, the inappropriate placement of LAR onto the surface of DFR hampers the opening phenomenon from occurring, revealing that an adequate interaction is necessary.

With the exception of the simulation leading to the complex dissociation, in the three other simulations performed on PP2, LCC diffuses close to the entry of the LAR active site, where it gets locked between helices 4 and 6 of LAR ($N135_{LAR}$ and $E94_{LAR}$) and interacts strongly with the NADPH co-factor through one or two H-bonds. The diffusion of LCC from DFR to this site takes approximately 300 ns. The calculation was continued for 1.7 μs, during which LCC remained in interaction with these two helices. This difficulty for the substrate to enter the active site of LAR has also been observed by considering the enzyme alone, *i.e.* not involved in a complex. Five molecular dynamics simulations of 500 ns each have been produced on a system composed of LAR alone and five LCC metabolites, free to diffuse around. None of these simulations led to the entrance of a LCC molecule in the active site, but in each of them, at least one LCC was interacting with the enzyme at the position described above. It can then be considered that the area described by NADPH and the helices 4 and 6 of LAR acts as a thermodynamic attractor for the substrates. Thus, to enhance the sampling of the potential energy surface associated with the final step of the diffusion process accelerated molecular dynamics (aMD) simulation was employed after 760 ns of simulation.[37] Finally, LCC enters the active site and interacts with residues which belong to the active site, namely Y137, K140, H122, and Q285 within 160 ns after the aMD simulation has begun (Fig. S5, ESI†). Fig. 3C shows the complete diffusion pathway followed by LCC metabolite transfer from the cavity of DFR to the active site of LAR.

It remains always difficult to extrapolate the simulation time using aMD to its equivalent when using classical MD. This depends strongly on the size and the rigidity of the system. Some estimates on small and large systems put forward ratios of roughly 2000 and 15 respectively.[42,43] Here, the 160 ns produced with the aMD procedure could correspond to a range of 2.5 μs to 25 μs in unconstrained molecular dynamics. This magnitude is consistent with the time when LCC remained locked into the thermodynamic attractor discussed above and the fact that no LCC entered the cavity of LAR when they were free to diffuse around the enzyme (total sampling time: 12.5 μs).

3.3. The substrate channelling is a multistep process

The diffusion process has been sampled without bias, indicating that it can easily overcome weak energy barriers at room temperature, such as sidechain reorientations. Thus, in order to obtain more details on how the structure of the complex could affect the diffusion phenomenon, we computed the interaction energy between the metabolite and the protein matrix throughout the trajectory following the MM-GBSA protocol. The aMD protocol, applied at the end of the diffusion process was an opportunity to accelerate sidechain movements without modifying the temperature of the system. MM-GBSA interaction energy is a good indicator of the evolution of the interaction between the metabolite and the protein complex though it lacks a complete description of the entropy term. The latter has been estimated on resting states, *i.e.* the initial and final interacting systems, to be about 23 kcal mol^{-1}.

On the basis of the estimated interaction energy, the substrate channelling phenomenon can be viewed as a three-step process, as shown in Fig. 4A. At the beginning of the MD simulation, LCC is located in the DFR active site and bears relatively high interaction energy with the complex, close to -20 kcal mol^{-1}. During the first phase, it leaves the DFR active site after $\sim$40 ns and interacts with residues M88, P204 and F90 defining the gate of DFR cavity. During the following 260 ns, the ligand diffuses on the protein–protein surface to finally reach the thermodynamic attractor site, mentioned above, close to the LAR active site entrance, where it interacts principally with residues E94, N135, I171, H172 and I264. In Fig. 4B the evolution of the interaction energy is plotted as a function of the distances between both enzymes and substrate. This analysis reveals a rapid diffusion of the metabolite. The interaction energy between the metabolite and the complex reaches a maximum close to zero when the metabolite is located at the junction between the enzymes. Nevertheless, the metabolite remains close to the surface in interaction with bridging water molecules.

During a second phase covering 50% of the simulation time, the LCC compound is locked between helices 4 and 6 of LAR.

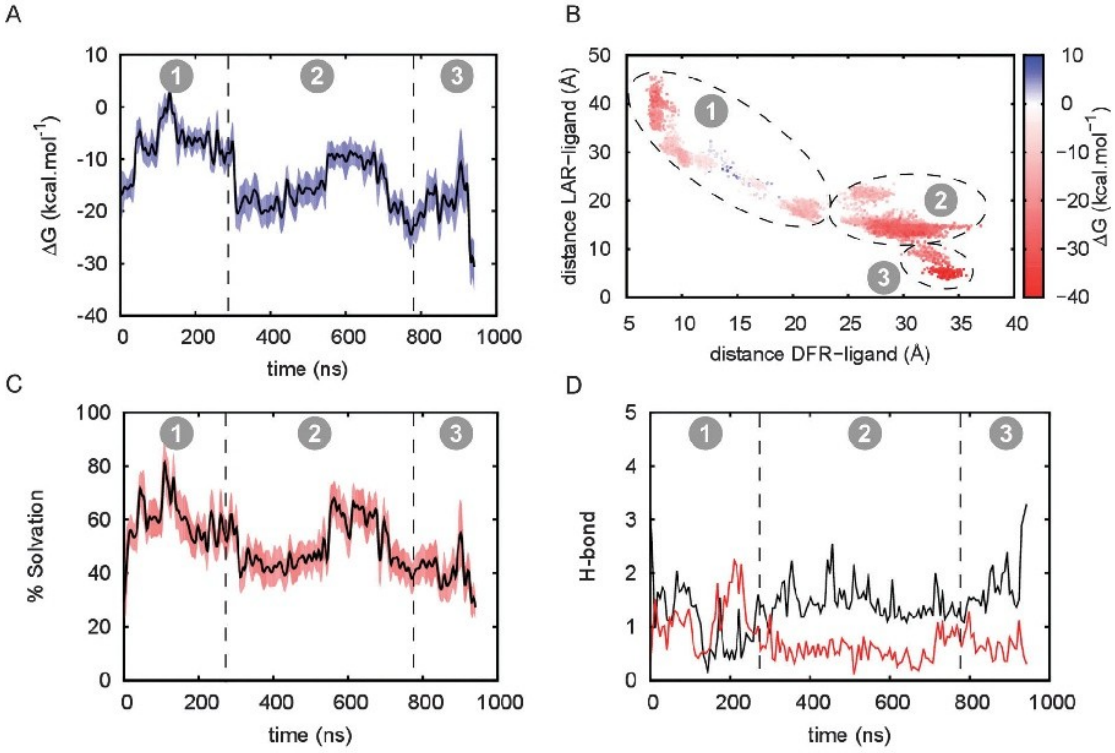


Fig. 4 Physico-chemical properties computed along LCC diffusion from DFR to LAR. The diffusion process is divided into three phases. (A) DFR–LAR/LCC binding free energy through the diffusion pathway calculated using the MMGBSA protocol. (B) 2D plot of interaction energy as a function of the two distances between substrate and enzyme centers of mass. (C) Percentage of solvation of LCC along substrate channelling. (D) Evolution of direct H-bond (black) and H-bond involving bridging water molecules (red) during the diffusion process. For each of these analyses, the average value and standard deviation were calculated from 156 successive subsets of 600 ps each, considering 1 frame per ps.

It is basically stabilized by strong interaction with the NADPH phosphate group and to a lesser extent by residues E94 and N135. The substrate sometimes loses its interactions with these residues of helices 4 and 6, allowing LCC to go back and forth, which keeps the anchor points with the phosphate groups of NADPH. These movements lead to a loss of interaction energy as reflected in Fig. 4A (period from 550 to 700 ns). This rocking movement is associated with a sampling of a narrow conformational attractor. At the end of this second phase, LCC flips and takes an orientation consistent with an entrance leading to a structure close to the one found in the LCC–LAR bound crystal structure.

In the final phase, the interaction with the NADPH moiety breaks, and LCC enters the active site of LAR in a position close to that found in the crystallographic structure (final ligand RMSD $\sim$5 Å). During this entrance, direct interactions between LCC and the cofactor are partially lost, explaining the decrease of interaction energy (see Fig. 4A). The major energy drops throughout the simulation occur at the exit of the DFR enzyme and after a final reorganization within the LAR target protein (Fig. S4, ESI†). The accurate prediction of interaction energy through the MMGBSA protocol on aMD trajectories has recently been addressed in an article devoted to the exploration of potential energy surfaces by such a method that confirmed the pertinence of this approach.[42]

Though the energies are slightly modified, the exploration of the surface is in good agreement with unbiased simulation and reproduces the general shape of the surface. On the basis of additional simulations, where the crystallographic structure LCC/NADPH/LAR system was considered as a starting point, the MM-GBSA energy of interaction between LCC and LAR is in a range of -31.5 to -34.5 kcal mol^{-1}, consistent with the energy value of -32 kcal mol^{-1} obtained at the end of the diffusion process. Thus, this last drop in energy is clearly associated with a stabilization of the substrate in the binding cavity of LAR. Moreover, since the entropy terms are similar for the initial and the final states, one can consider that the binding process is thermodynamically favoured by about -13.5 kcal mol^{-1} (see Fig. 4A).

3.4. Desolvation penalty is reduced with respect to a regular diffusion process

Electrostatic analysis performed using APBS software[44] suggested a negatively charged route that connected the two active sites (Fig. S5, ESI†). Thus, the driving force of the diffusion along this electrostatic pathway is the interaction of these residues with the different H-bond donor groups of flavonoid molecules. The molecular structure of LCC is of ambivalent nature. Its aromatic rings and hydroxyl groups confer both hydrophobic and hydrophilic physico-chemical properties. In this trajectory,

H-bonds appear between LCC and successive amino acids defining the diffusion pathway, keeping the metabolite close to the enzymatic surface. When most of the direct H-bonds are lost, bridging water molecules take the relay. Indeed, some water molecules get inserted between the metabolite and the protein surface and play the role of a bridge between the amino-acids and LCC, hampering the latter from escaping within the bulk. These water molecules cannot be considered as bulk solvent molecules and they are in interaction with both the surface of the proteins and the metabolite.

When LCC diffuses from DFR towards LAR, its solvation is only partial with respect to a metabolite fully hydrated in the bulk. The number of water molecules surrounding the metabolite in the first hydration shell, according to the post-processing analysis in AMBER, reaches 36. At the protein–protein interface, the solvation is reduced by at least 20% and up to 60% (see period 2 in Fig. 4C). The desolvation process of the ligand entering within the second enzyme (which can be considered as energetically penalizing) is then decreased accordingly. It reveals that a partially hydrated ligand still in interactions with the protein matrix during the substrate channelling process is thermodynamically more favourable than its complete rehydration.

This last point is particularly important because the present metabolite, beyond presenting H-bond possibilities, is primarily hydrophobic. LCC is known to be non-stable in water solution; this partial solvation protects it from possible degradation and could explain why the global biosynthesis process proceeds and produces the next metabolites.

4. Conclusions

It has been postulated for several years that enzymes of same biological cascade aggregate to form transient multienzymatic complexes that lead to a more efficient production of the metabolites. In these complexes, the metabolites can diffuse from an active site to the following active site through the so-called substrate channelling. For numerous reasons, primarily arising from the transient existence of these edifices, minimal experimental evidence of a metabolite diffusing between connected enzymes at the molecular level is available. Here, by performing molecular dynamics simulations, we bring a molecular point of view to depict such a substrate channelling event in the course of flavonoid biosynthesis. Considering two interacting successive enzymes of the biochemical cascade, namely DFR and LAR, we observed the diffusion of the product of DFR, namely LCC, from the active site to the cavity of LAR. This simulation allowed us to identify crucial events, at the molecular or atomic scale, which open the possibility of this diffusion.

First, the formation of the protein–protein (DFR–LAR) complex leads, in a relatively short time, to an opening of the DFR active site, due to the movement of three helices D94-K118, D149-K157 and A160-N182. This allosteric effect triggers a facilitated release of LCC from the enzymatic cavity with respect to what happens when considering a DFR/$NADP^+$/LCC system without interaction with LAR. The second important aspect of this diffusion lies in the role of solvent molecules and their interactions with the metabolite. Indeed, throughout the diffusion pathway, strong surface interactions between the ligand and the protein matrix are reported. Since the present metabolite exhibits hydrophobic tendencies, it does not diffuse naturally into the bulk. The evolution of the metabolite trajectory is then associated with forming and breaking of H-bonds involving amino acids at the protein surfaces. When the enzyme/metabolite interaction is lost, bridging water molecules get inserted between the two protagonists and prevent the metabolite from leaving the protein surface. As a consequence, in the present case, diffusion on the protein matrix minimizes the energy loss during desolvation since the LCC metabolite is on average only 50% solvated. The efficient diffusion of neutral but polarized intermediates is then controlled by a subtle balance between H-bond channelling and solvation/desolvation effects.

Despite the fact that the total diffusion process has been observed only once, this event can be considered as plausible since it has been sampled without any bias. Moreover, the leucocyanidin metabolite is not stable in the media and its concentration remains weak while a higher amount of catechin, (the enzymatic product of LAR) is detected. It can be then postulated that, through a substrate channelling event, the leucocyanidin could be protected from the media and thus reduced in the product of the enzymatic reaction catalysed by LAR. The conclusions reached here are specific to the DFR–LAR–LCC system, notably due to the particular properties of the metabolite (mainly hydrophobic but with the possibility to form numerous H-bonds due to its hydroxyl groups), but could be extended to similar systems involving other compounds of the flavonoid family.

References

1 I. G. L. Libourel and Y. Shachar-Hill, *Annu. Rev. Plant Biol.*, 2008, **59**, 625–650.
2 S. Schuster, D. A. Fell and T. Dandekar, *Nat. Biotechnol.*, 2000, **18**, 326–332.
3 H. A. Stafford, *Phytochemistry*, 1969, **8**, 743–752.
4 H. O. Spivey and J. M. Merz, *BioEssays*, 1989, **10**, 127–129.
5 E. W. Miles, S. Rhee and D. R. Davies, *J. Biol. Chem.*, 1999, **274**, 12193–12196.
6 X. Huang, H. M. Holden and F. M. Raushel, *Annu. Rev. Biochem.*, 2001, **70**, 149–180.
7 K. Jørgensen, A. V. Rasmussen, M. Morant, A. H. Nielsen, N. Bjarnholt, M. Zagrobelny, S. Bak and B. L. Møller, *Curr. Opin. Plant Biol.*, 2005, **8**, 280–291.
8 B. Winkel-Shirley, *Plant Physiol.*, 2001, **126**, 485–493.
9 B. S. J. Winkel, *Annu. Rev. Plant Biol.*, 2004, **55**, 85–107.
10 H. O. Spivey and J. Ovádi, *Methods*, 1999, **19**, 306–321.
11 I. B. Lomakin, Y. Xiong and T. A. Steitz, *Cell*, 2007, **129**, 319–332.
12 S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Hakansson,

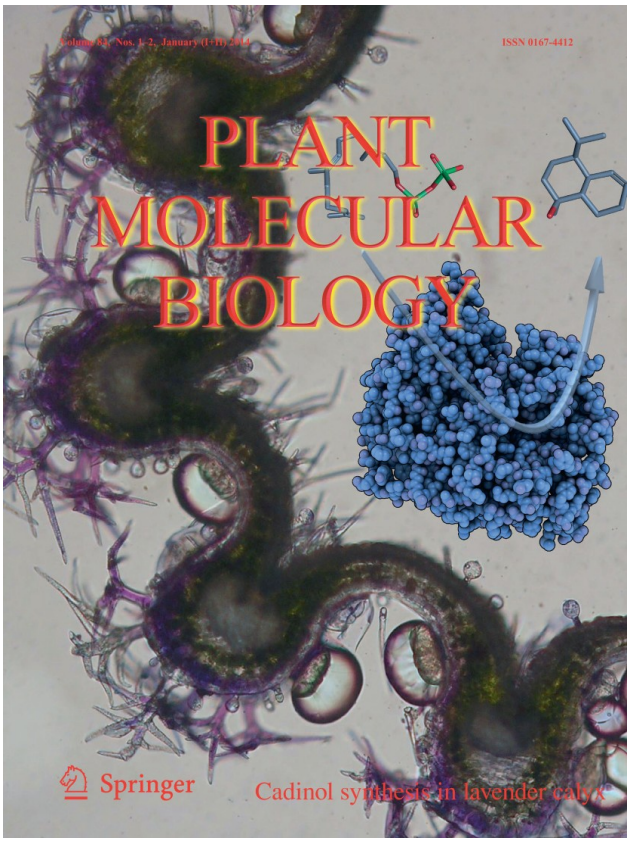
D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–517.
13 C. You, S. Myung and Y. H. P. Zhang, *Angew. Chem., Int. Ed.*, 2012, **51**, 8787–8790.
14 C. You and Y. H. P. Zhang, *ACS Synth. Biol.*, 2013, **2**, 102–110.
15 J. B. Thoden, H. M. Holden, G. Wesenberg, F. M. Raushel and I. Rayment, *Biochemistry*, 1997, **36**, 6305–6316.
16 R. M. Stroud, *Nat. Struct. Biol.*, 1994, **1**, 131–134.
17 A. H. Elcock and J. A. McCammon, *Biochemistry*, 1996, **35**, 12652–12658.
18 P. Bauler, G. Huber, T. Leyh and J. A. McCammon, *J. Phys. Chem. Lett.*, 2010, **1**, 1332–1335.
19 C. Eun, P. M. Kekenes-Huskey, V. T. Metzger and J. A. McCammon, *J. Chem. Phys.*, 2014, **140**, 105101.
20 O. M. Andersen and K. R. Markham, *Flavonoids: Chemistry, Biochemistry and Applications*, Taylor & Francis, 2005.
21 F. Ververidis, E. Trantas, C. Douglas, G. Vollmer, G. Kretzschmar and N. Panopoulos, *Biotechnol. J.*, 2007, **2**, 1235–1249.
22 H. A. Stafford and H. H. Lester, *Plant Physiol.*, 1982, **70**, 695–698.
23 I. E. Burbulis and B. Winkel-Shirley, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 12929–12934.
24 L. Ralston and O. Yu, *Phytochem. Rev.*, 2006, **5**, 459–472.
25 B. L. Møller, *Science*, 2010, **330**, 1328–1329.
26 P. Petit, T. Granier, B. L. d'Estaintot, C. Manigand, K. Bathany, J.-M. Schmitter, V. Lauvergeat, S. Hamdi and B. Gallois, *J. Mol. Biol.*, 2007, **368**, 1345–1357.
27 C. Maugé, T. Granier, B. L. d'Estaintot, M. Gargouri, C. Manigand, J.-M. Schmitter, J. Chaudière and B. Gallois, *J. Mol. Biol.*, 2010, **397**, 1079–1091.
28 K. C. Crosby, A. Pietraszewska-Bogiel, T. W. J. Gadella Jr and B. S. J. Winkel, *FEBS Lett.*, 2011, **585**, 2193–2198.
29 M. T. McManus, W. A. Laing and A. C. Allan, *Protein–protein interactions in plant biology*, Sheffield Academic Press, 2002.
30 L. J. McGuffin, K. Bryson and D. T. Jones, *Bioinformatics*, 2000, **16**, 404–405.
31 M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo and A. Šali, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 291–325.
32 R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
33 S. Fiorucci and M. Zacharias, *Proteins*, 2010, **78**, 3131–3139.
34 N. Holmberg, U. Ryde and L. Bulow, *Protein Eng.*, 1999, **12**, 851–856.
35 J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Res.*, 2005, **33**, W368–W371.
36 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER12*, 2012, DOI: citeulike-article-id:10779586.
37 D. Hamelberg, J. Mongan and J. A. McCammon, *J. Chem. Phys.*, 2004, **120**, 11919–11929.
38 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
39 A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383–394.
40 I. M. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
41 A. Bakan, L. M. Meireles and I. Bahar, *Bioinformatics*, 2011, **27**, 1575–1577.
42 Y. Miao, S. E. Nichols and J. A. McCammon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6398–6406.
43 L. C. T. Pierce, R. Salomon-Ferrer, C. Augusto F. de Oliveira, J. A. McCammon and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 2997–3002.
44 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10037–10041.

Article 4 :

Isolation and functional characterization of a t-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*.

F. Julien*, S. Moja, A. Bony, S. Legrand, C. Petit, T. Benabdelkader, K. Poirot, S. Fiorucci, Y. Guitton, F. Nicole, S. Baudino, J.L. Magnard.

Plant. Mol. Biol. ***2014****, 84, 227-241.*



Plant Mol Biol (2014) 84:227–241
DOI 10.1007/s11103-013-0131-3

Isolation and functional characterization of a τ-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*

Frédéric Jullien · Sandrine Moja · Aurélie Bony · Sylvain Legrand · Cécile Petit · Tarek Benabdelkader · Kévin Poirot · Sébastien Fiorucci · Yann Guitton · Florence Nicolè · Sylvie Baudino · Jean-Louis Magnard

Received: 5 February 2013 / Accepted: 15 September 2013 / Published online: 27 September 2013

Abstract In this paper we characterize three sTPSs: a germacrene D (LaGERDS), a (*E*)-β-caryophyllene (LaCARS) and a τ-cadinol synthase (LaCADS). τ-cadinol synthase is reported here for the first time and its activity was studied in several biological models including transiently or stably transformed tobacco species. Three dimensional structure models of LaCADS and *Ocimum basilicum* γ-cadinene synthase were built by homology modeling using the template structure of *Gossypium arboreum* δ-cadinene synthase. The depiction of their active site organization provides evidence of the global influence of the enzymes on the formation of τ-cadinol: instead of a unique amino-acid, the electrostatic properties and solvent accessibility of the whole active site in LaCADS may explain the stabilization of the cadinyl cation intermediate. Quantitative PCR performed from leaves and inflorescences showed two patterns of expression. *LaGERDS* and *LaCARS* were mainly expressed during early stages of flower development and, at these stages, transcript levels paralleled the accumulation of the corresponding terpene products (germacrene D and (*E*)-β-caryophyllene). By contrast, the expression level of *LaCADS* was constant in leaves and flowers. Phylogenetic analysis provided informative results on potential duplication process leading to sTPS diversification in lavender.

Keywords *Lavandula angustifolia* L. · 454 Pyrosequencing · Terpene synthase · Transcript regulation

Accession numbers: *LaCADS*: JX401282; *LaCARS*: JX401283; *LaGERDS*: JX401284.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-013-0131-3) contains supplementary material, which is available to authorized users.

F. Jullien (✉) · S. Moja · A. Bony · S. Legrand · C. Petit · T. Benabdelkader · K. Poirot · F. Nicolè · S. Baudino · J.-L. Magnard
Université de Lyon, 42023 Saint-Étienne, France
e-mail: jullien@univ-st-etienne.fr

F. Jullien · S. Moja · A. Bony · C. Petit · F. Nicolè · S. Baudino · J.-L. Magnard
Université de Saint-Etienne, Jean Monnet, 42000 Saint-Étienne, France

F. Jullien · S. Moja · A. Bony · C. Petit · F. Nicolè · S. Baudino · J.-L. Magnard
Laboratoire de Biotechnologies Végétales Appliquées aux Plantes Aromatiques et Médicinales, EA 3061, 23 rue du Dr Michelon, 42000 Saint-Étienne, France

S. Legrand
Stress Abiotique et Différenciation des végétaux Cultivés (SADV), UMR INRA 1281, Université Lille Nord de France, Université Lille 1, SN2, 59650 Villeneuve d'Ascq, France

T. Benabdelkader
Laboratoire d'Ecophysiologie Végétale, Ecole Normale Supérieure, Kouba, 16050 Alger, Algeria

S. Fiorucci
Institut de Chimie de Nice, UMR-CNRS 7272, Faculté des Sciences, Université de Nice-Sophia Antipolis, 06108 Nice Cedex 2, France

Y. Guitton
Laboratoire Stress Abiotique et Différenciation des végétaux Cultivés, UMR Lille1/INRA 1281, Université Lille 1, 59655 Villeneuve d'Ascq Cedex, France

Abbreviations

CADS	Cadinol synthase
CARS	β-Caryophyllene synthase
EO(s)	Essential oil(s)
EST(s)	Expressed sequence tag(s)
IPTG	Isopropyl-β-D-thiogalactopyranoside
FPP	Farnesyl diphosphate
GERDS	Germacrene D synthase
GPP	Geranyl diphosphate
LaCADS	*L. angustifolia* cadinol synthase
LaCARS	*L. angustifolia* β-caryophyllene synthase
LaGERDS	*L. angustifolia* germacrene D synthase
mTPS(s)	Monoterpene synthase(s)
Qpcr	Quantitative realtime polymerase chain reaction
sTPS(s)	Sesquiterpene synthase(s)
TPS(s)	Terpene synthase(s)

Introduction

Lavender, as many species of the *Lamiaceae* family, synthesizes and accumulates EOs in the secretory capitate and peltate oil glands located in abundance on the surface of the calyx and to a lesser extent on leaves (Online resource 1). Monoterpenes and sesquiterpenes are the main components of lavender EO and are derived from the condensation of isopentenyl diphosphate (IPP) and its allylic isomer, dimethylallyl diphosphate (DMAPP). These C5 isomers are produced through two different cytosolic and plastidial pathways (Fig. 1). The mevalonate pathway (MEV) is located in the cytosol and probably in its last steps in the peroxisome (Simkin et al. 2011). It begins with the condensation of two acetyl-CoA and leads through several steps to mevalonate diphosphate, then IPP, further isomerized in DMAPP. This pathway provides IPP/DMAPP precursors for sesquiterpene and triterpene syntheses. The alternative MEP pathway starts by the transfer of a C2 unit from pyruvate to glyceraldehyde-3-P and also leads to the synthesis of both IPP and DMAPP that are used as precursors for both monoterpene and diterpene syntheses. This MEP pathway was first characterized in bacteria (Rohmer et al. 1993) and then in the plastidial compartment of plants (Arigoni et al. 1997; Lichtenthaler et al. 1997). Although the MEV and MEP pathways are thought to be independently regulated they can both contribute to monoterpene and sesquiterpene syntheses through metabolite exchanges between cytosol and plastid. A specific membrane plastid transporter involved in the transport of IPP and/or GPP has been characterized indicating that prenylphosphate from the plastid could be used for the synthesis of sesquiterpenes in the cytosol (Bick and Lange 2003). Moreover, GPP could also be transported to cytosol leading to the production of monoterpene in the cytosol (Gutensohn et al. 2013).

Prenyltransferases catalyze the condensation of IPP and DMAPP to produce prenyl diphosphates. Condensation of one DMAPP and two IPP molecules catalyzed by farnesyl diphosphate synthase (FPPS) leads to the formation of FPP in the cytosol whereas the condensation of one DMAPP and one IPP catalyzed by geranyl diphosphate synthase (GPPS) leads to the formation of GPP in the plastid. Both GPP and FPP are substrates of terpene synthases for the synthesis of mono- and sesquiterpenes respectively. Five terpene synthases (Demissie et al. 2011, 2012; Landmann et al. 2007) have already been characterized from lavender (Fig. 1). Moreover, the expression of genes encoding linalool and limonene synthases showed that transcription paralleled terpene accumulation suggesting, at least in part, a transcriptional regulation of these genes (Guitton et al. 2010; Lane et al. 2010).

In this study, we characterized three terpene synthases responsible for the production of the major sesquiterpenes of the essential oil. A cadinol synthase is reported here for the first time in the plant kingdom. Correlations of terpene synthase transcript levels from qRT-PCR and essential oil accumulation at different developmental stages were performed to examine the regulation of these TPSs. 3D protein modeling of LaCADS and directed mutagenesis were performed to propose a mechanism explaining the biosynthesis of τ-cadinol.

Materials and methods

Plant material

Lavandula angustifolia plants cv 'Diva' were grown in an experimental field in Manosque (Alpes de Haute Provence—France). The whole plant was harvested to construct the 454 library whereas only leaves and flowers were used to quantify terpene synthase gene expression. In addition, plants cultivated in pots in the greenhouse have been used for leaf or flower RNA extraction and terpene synthase cloning.

Preparation of a normalized random-primed cDNA library, 454 pyrosequencing and reads analysis

Total RNAs were extracted from leaves, flowers and roots using the Tri reagent kit (Euromedex) according to the manufacturer's instructions. In total, 40 µg of an equimolar mixture of total RNA extracted from the different conditions was sent to Eurofins MWG GmbH. A normalized random-primed cDNA library was prepared from the RNA, an emulsion-based PCR was performed and one segment of a sequencing plate was sequenced on a GS FLX (454/Roche) to yield more than 600,000 reads delivered as assembled reads in FASTA format with quality scoring files of all clusters.

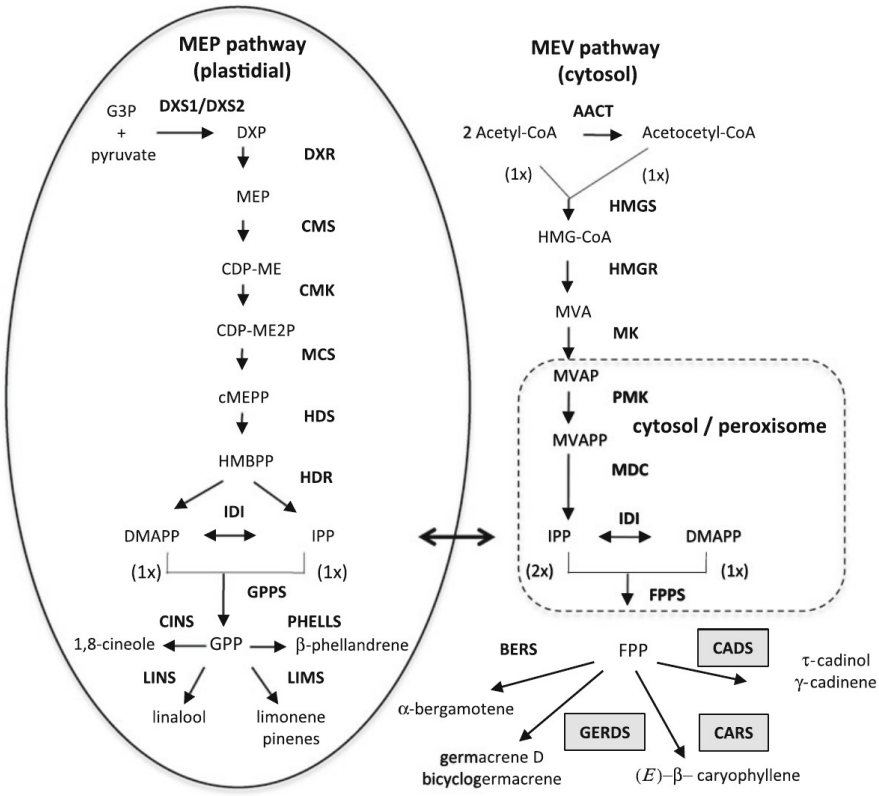


Fig. 1 Pathways of mono and sesquiterpene biosynthesis in *L. angustifolia*. Enzymatic steps in the *plain box* correspond to genes functionally characterized in the study. Enzymatic steps in the *dotted box* represent steps in which full length protein have been obtained from 454 library. AACT, acetoacetyl-coenzyme A thiolyase; BERS, α-bergamotene synthase; *CADS* cadinol synthase, *CARS* β-caryophyllene synthase, *CDP-ME* 4-(cytidine 5′-diphospho)-2-*C*-methyl-D-erythritol, *CDP-ME2P* CDP-ME-2-phosphate, *cMEPP* 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate, *CINS* cineole synthase, *CMK* 4-(cytidine 5′–diphospho) 2-*C*-methyl-D-erythritol kinase, *CMS* 2-*C*-methyl-D-erythritol 4-phosphate transferase, *DMAPP* dimethylallyl diphosphate, *DXP* 1-deoxy-D-xylulose 5-phosphate, *DXR* DXP reductoisomerase, *DXS* DXP synthase, *FPP* farnesyl diphosphate, *FPPS* FPP synthase, *GERDS* germacrene D synthase, *G3P* glyceraldehyde 3-phosphate, *GPP* geranyl diphosphate, *GPPS* GPP synthase, *HDS* 1-hydroxy-2-methyl-2(*E*)-butenyl 4-diphosphate synthase, *HDR* 1-hydroxy-2-methyl-2(*E*)-butenyl 4-diphosphate reductase, *HMBPP* 1-hydroxy-2-methyl-2(*E*)-butenyl 4-diphosphate, *HMGCoA* 3hydroxy-3-methylglutaryl-CoA, *HMGR* hydroxymethylglutaryl-CoA reductase, *HMGS* hydroxymethylglutaryl-CoA synthase, *IDI* isopentenyl diphosphate isomerase, *IPP* isopentenyl diphosphate, *LIMS* limonene synthase, *LINS* linalool synthase, *MDC* mevalonate pyrophosphate decarboxylase, *MDS* 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *MEP* 2-C-methyl-D-erythritol 4 phosphate, *MK* mevalonate kinase, *MVA* mevalonate, *MVAP* MVA-5-phosphate, *MVAPP* MVA-5-diphosphate, *PHLS* β-phellandrene synthase, *PMK* 5-phosphomevalonate kinase

RNA isolation and cloning of TPSs

mRNA isolation using the nucleospin® RNA plant kit (Macherey–Nagel) was followed by a reverse transcription. RACE assays were carried out to amplify 5′ and 3′ ends of the studied TPS using the Marathon® cDNA amplification kit (Clontech, San Diego, CA, USA), following the manufacturer's instructions. Sequences of primers used to clone LaCARS, LaGERDS or LaCADS are reported in online resource 2. DNA inserts cloned in the pGEMT®-Easy vector were sequenced by Eurofins MWG Operon (Germany, Ebersberg). For functional characterization the three sTPS were cloned in the pHGGWA vector (Busso et al. 2005) using the Gateway technology. The 5′ primer used to clone genes in the entry vector (pENTR/D-TOPO) (Invitrogen, USA) comprised an additional sequence for the hydrolysis of the GST-sTPS recombinant protein by thrombin.

LaCADS cloned into pENTR/D-TOPO was transferred by Gateway LR reactions into the binary destination vector pMDC32 (Curtis and Grossniklaus, 2003). The expression of pMDC32::CADS is under the control of a double CaMV 35S promoter.

Expression of recombinant LaCARS, LaGERDS and LaCADS in *E. coli*

Escherichia coli strain Rosetta (DE3) pLysS cells (Novagen, Darmstadt, Germany) were transformed with pHGGWA carrying the full length *LaCARS*, *LaGERDS* or *LaCADS* by heat shock. Production of the heterologous protein was performed during 14–16 h at 16 °C and 50 rpm in terrific broth supplemented with 0.5 % glycerol, 0.25 M D-sorbitol, 2.5 mM betaine after induction with 0.2 mM IPTG. The cells recovered by centrifugation were disrupted by incubation in native binding buffer (50 mM NaH_2PO_4, 0.5 M NaCl, 20 mM imidazole, 5 % glycerol, 5 mM DTT, pH 8) supplemented with 0.5 mg/ml lysozyme and sonication. After clarification of the lysate by centrifugation, the recombinant protein was purified by binding to the Talon® metal affinity resin (Clontech) according to the manufacturer's instruction. The resin-bound protein was incubated overnight at 4 °C in 200 µl of native binding buffer supplemented with 10 units of thrombin. The next day the TPS was recovered from the mixture by filtration. Protein concentration was measured using the BioRad reagent with bovine serum albumin as standard (Bradford 1976).

Transient and stable expression of LaCADS in tobacco

Four to six-week-old greenhouse-grown seedlings were co-infiltrated in triplicate with *Agrobacterium tumefaciens* C58pMP90 harboring the pMDC32::CADS binary vector and *A. tumefaciens* C58C1 carrying the viral suppressor p19 into pBIN61 (Voinnet et al. 2003) according to the method of Hellens et al. (2005). Freshly-grown bacteria cultures were centrifuged when OD_{600} nm raised 0.3–0.5 and re-suspended in infiltration media (50 mM MES buffer, 2 mM NaH_2PO_4, 0.2 mM acetosyringone, 0.5 % glucose) for 1 h at room temperature before injection into *N. benthamiana* leaves using a 1 ml syringe. After 7–10 days the treated leaves were detached and terpenes were extracted with hexane.

Transformation of leaf disks of *Nicotiana tabacum* cv. W38 by coculture with *A. tumefaciens* strain C58pMP90 harboring the pMDC32::CADS binary vector was performed according to Horsch et al. (1985).

Directed mutagenesis

Single base mutation were performed in order to replace the aspartate D_{532} or D_{533} by a glycine in the LaCADS sequence. LaCADS coding sequence cloned in pENTR/D-TOPO was used as template in a two step PCR amplification strategy. Mutation was introduced in reverse primers in PCR1 (PCR1R) coupled with the primer F full length (see online resource 2). In the same way, mutation was introduced in forward primer of PCR2 (PCR2F) coupled with the primer R full length (see online resource 2). PCR1 and PCR2 generate two products of respectively 1,614 and 89 pb. In a second step, PCR3 was performed using both products from PCR1 and PCR2 in equal molarity as template and primers F and R full length to amplify the full gene including the mutation on D_{532}. Mutation on D_{533} was performed using the same strategy with primers PCR1′R and PCR2′F (see online resource 2).

Enzyme assays

Enzymatic assays were performed in a final volume of 500 µl containing 15–50 µg purified recombinant protein, buffer (25 mM Tris–Cl, pH 7.5, 10 % glycerol, 1 mM DTT, 1 mg/ml BSA) and cofactors (10 mM $MgCl_2$, 1 mM $MnCl_2$). The reaction was started by addition of 50 µM geranyl or farnesyl diphosphate and the mixture was overlaid with 500 µl of hexane. After a 2 h incubation at 30 °C, the mixture was vigorously mixed and the upper hexane phase was collected, concentrated under nitrogen stream and analyzed by GC/MS. Negative controls were performed using the purified product from Rosetta (DE3) pLysS without expression vector.

Enzymatic assays were also performed from transgenic *N. tabacum*. Two grams of fresh leaves were crushed in liquid nitrogen supplemented with 20 mg PVPP. A pH 6.8 buffer with 20 % glycerol, 50 mM HEPES, 50 mM $Na_2S_2O_5$, 50 mM ascorbic acid, 10 mM $MgCl_2$, 5 mM DTT was added to the leaf powder for protein extraction. After centrifugation (2 min 12,000*g*), the supernatant was desalted on a PD10 column in the following buffer (10 % glycerol, 15 mM MOPS, 1 mM ascorbic acid, 5 mM sodium *meta*-vanadate, 2 mM DTT). For enzymatic reaction, the eluted proteins were supplemented with 20 mM $MgCl_2$, 0.5 mM $MnCl_2$ and 20 µM FPP and the solution was overlaid with 1 ml of pentane. Incubation was performed for 2 h at 30 °C, then the sample was vigorously vortexed, shortly centrifuged and the pentane fraction was analyzed by GC–MS.

GC–MS analysis

All GC–MS were performed on an Agilent GC 6850 gas chromatograph coupled with an Agilent 5973 ion trap mass detector. The instrument was equipped with a 30 m × 0.25 mm apolar capillary column DB5. Temperatures of injector and detector were 250 °C. Helium was used as the carrier gas at a flow rate of 1.0 ml/min. Oven temperature settings were: 4 min at 60 °C after injection followed by a 4 °C/min temperature ramp from 60 to 240 °C. Temperature was then kept on hold at 240 °C for 5 min. Injection volume was 2 µl in splitless mode. Molecule identification was performed using Wiley, NIST 05 and Adams mass spectra databases (Adams 2007). The GC products were *in fine* confirmed by comparing their retention time and mass spectrum to those of authentic

(*E*)-β-caryophyllene (Payan Bertrand France) and germacrene D samples (R.C. Treatt USA Inc.). An essential oil of basil (*Ocimum basilicum*) was used to characterize τ-cadinol (IFF France). Terpenes of *L. angustifolia* cv. Diva leaves and flowers were extracted with hexane as previously reported (Guitton et al. 2010) and analyzed as described above.

Transcript quantification by qPCR

LaCARS, *LaCADS* and *LaGERDS* expression levels were determined in leaves and flowers at different flower maturity stages. Total RNAs were extracted from leaves and from flowers of three *L. angustifolia* plants cv. 'Diva' using the Tri reagent kit (Molecular Research Center, Cincinnati, USA) according to the manufacturer's instructions and treated with the RQ1 RNAse-free DNase (Promega, Madison, USA). A total of 2 µg of RNA was reverse transcribed using SuperScript™ III Reverse Transcriptase (Invitrogen, Carlsbad, USA) at 50 °C for 1 h with an oligo-dT primer. Each RNA sample was reverse-transcribed in two independent reactions.

Quantitative PCR was performed with CFX96™ real-time detection system (Bio-Rad, Hercules, USA) using the SsoAdvanced™ SYBR Green Supermix (Bio-Rad, Hercules, USA). All reactions were carried out in a 20 µl volume using 2 µl of reverse transcribed cDNA as template and 500 nM of each of the primers according to the manufacturer's protocol. Gene primers (Online resource 2) were designed using the Primer3 software (Rozen and Skaletsky 2000). Each cDNA sample was amplified twice in two independent qPCR runs with each primer combination. As the RNA samples were reverse-transcribed in duplicate, this yielded four technical replicates per original sample. The following thermal profile was used for all PCR reactions: 95 °C for 30 s, followed by 40 quantification cycles [95 °C for 5 s, Tm (Online 2) for 30 s]. After 40 cycles, a melting-curve analysis (65–95 °C, one fluorescence read every 0.5 °C) was performed to check the specificity of the amplification.

Normalized expression values ($2^{-\Delta\Delta Cq}$ method, Livak and Schmittgen 2001) of *LaCARS*, *LaCADS* and *LaGERDS* were calculated by the CFX96TM data manager (Bio-Rad, Hercules, USA) using *β-ACTIN*, *TUBULIN* and *GAPDH* as reference genes (Online resource 2). The stability of expression of these reference genes was evaluated using Best-Keeper (Pfaffl et al. 2004), geNorm v. 3.5 (Vandesompele et al. 2002) and NormFinder (Andersen et al. 2004). We used REST 2009 (Pfaffl et al. 2002) to compare the expression level of a gene in a 'sample' group using a 'control' group as a reference by implementing a pairwise fixed reallocation randomization test (10,000 iterations). Differences in expression between 'sample' and 'control' cDNAs are considered significant for p values <0.05.

Model building and refinement

Blast and sequence alignment were performed using the UniProt webserver and MAFFT software. LaCADS and γ-cadinene synthase (UNIPROT entry: Q5SBP5) 3D modeling was performed from δ-cadinene synthase (PDB identifier 3G4F) and molecular models were generated with Modeller (Marti-Renom et al. 2000). Homology model of LaCADS with the highest DOPE score and no steric clash was then energy minimized using AMBER (Case et al. 2012) and the AMBER ff03 force-field Parameters. The solvation was treated implicitly with a generalized born model. The structural integrity of the homology model of LaCADS was verified with Procheck (Laskowski et al. 1993). Electrostatic potential was calculated with the APBS program and mapped onto the solvent accessible surface area. To obtain accurate electrostatic properties we use the two step focusing technique and a grid spacing lower than 0.5 Å in each space dimension. The molecular surface was generated using a water probe with a radius of 1.4 Å. The dielectric constant of the protein and the solvent was fixed to 2–80 respectively. pKa values of ionizable groups was predicted using PropKa (Rostkowski et al. 2011) and the amber forcefield (ff03.r1) was used to add atomic charges.

Results

Functional characterization of three sTPSs: LaCARS, LaGERDS and LaCADS

Partial sequences of three sesquiterpene synthases were obtained from 3 sets of ESTs obtained from a 454 cDNA library. Amplification of 5′ and 3′ ends was carried out by RACE assays. Open reading frames of putative *LaCARS, LaGERDS and LaCADS* coded for proteins with 548, 549, and 555 amino acids respectively. The molecular masses are predicted to be 63.6, 63.9, and 64.5 kDa respectively. Alignment of the three deduced proteins revealed alterations in characteristic conserved motifs of TPSs (Fig. 2). The arginine-rich N-terminal RR(x8)W motif required for cyclisation in mTPSs (Williams et al. 1998) was conserved only in LaGERDS but was altered in the other two synthases. The highly conserved aspartate-rich DDxxD motif required for Mg^{2+} or Mn^{2+} binding was found in both LaCARS and LaCADS whereas the third aspartate was substituted by a glutamate in LaGERDS. The other two characteristic motifs of TPSs LQLYEASFLL and (N,D)D(L,I,V)x(ST)xxxE are partially conserved (Degenhardt et al. 2009). Concerning the first motif, a sequence of four amino acids LYEA could be found in the three terpene synthases. The second motif was conserved in LaCADS but altered in both LaCARS and LaGERDS in which a glycine was found in place of a serine/threonine.

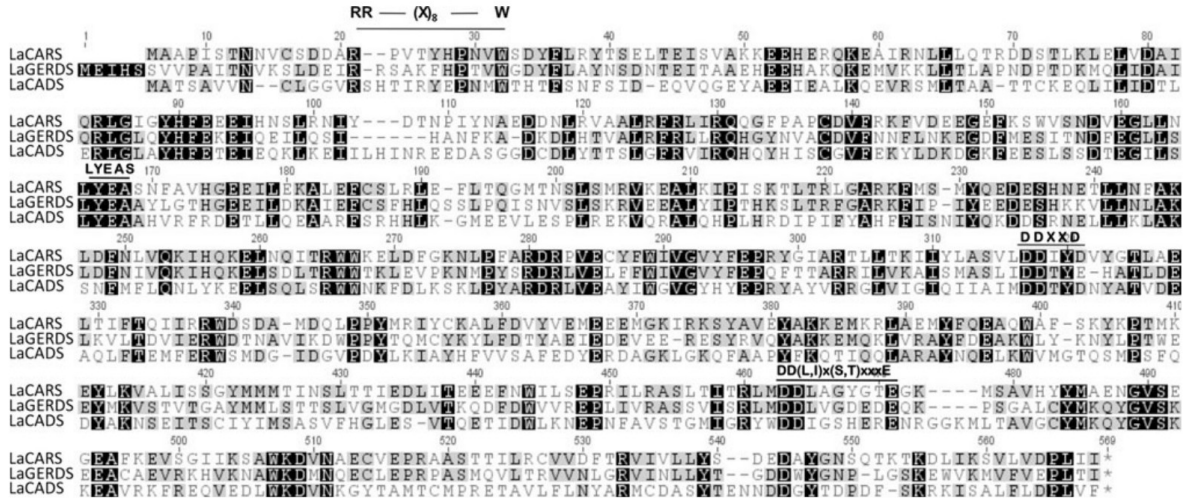


Fig. 2 Amino acid sequence alignment of the cloned sTPS from *L. angustifolia* LaCARS, (*E*)-β-caryophyllene synthase; *LaGERDS* germacrene D synthase, *LaCADS* τ-cadinol synthase. The four conserved regions of terpene synthases are *underlined*

Recombinant LaCARS, LaGERDS and LaCADS were produced in the *E. coli* Rosetta™ (DE3) pLysS strain using the pHGGWA expression vector (Busso et al. 2005) and incubated with FPP. The purified LaCARS produced a unique compound with retention time and mass spectrum identical to those of a (*E*)-β-caryophyllene analytical standard (Fig. 3a–c; online resource 3A-B). Purified LaGERDS incubated with FPP catalyzed the production of germacrene D and bicyclogermacrene as major and minor products respectively (Fig. 3b). Germacrene D could be identified by comparison of retention time with an analytical standard (Fig. 3c). Mass spectra of germacrene D from enzyme assay and analytical standard are available in supplementary data (online resource 3C-D). The purified LaCADS incubated with FPP produced τ-cadinol as major compound and γ-cadinene as minor product (Fig. 3d). The identification of τ-cadinol was performed using an essential oil (EO) of a τ-cadinol rich chemotype of *O. basilicum* as standard. τ-cadinol of *O. basilicum* had identical retention time and mass spectrum to those of the major product catalyzed by LaCADS activity (Fig. 3e, f). τ-cadinol was also found in lavender EO with similar mass spectrum and retention time to standard (data not shown). No monoterpene could be detected when LaCADS or LaGERDS were incubated with GPP. In contrast LaCARS was able to catalyze the production of several monoterpenes among which myrcene, limonene and camphene could be clearly identified (data not shown).

Because this is the first report of a τ-cadinol synthase activity, additional expression experiments were performed. *A. tumefaciens*-mediated transient expression (Hellens et al. 2005) in *Nicotiana benthamiana* leaves confirmed that LaCADS acted as a τ-cadinol synthase *in planta*: *N. benthamiana* leaves infiltrated with a double CaMV 35S-driven binary *LaCADS* vector produced significant amount of τ-cadinol (Fig. 4b). τ-cadinol was not produced in control *N. benthamiana* that had been infiltrated with P19 alone (Fig. 4a).

Moreover, stably transformed *N. tabacum* plants were obtained through transformation experiments with *A. tumefaciens* strain C58pMP90 harboring the pMDC32::CADS. No cadinol was found by GC–MS analysis after hexane extraction of transgenic leaves (data not shown). Proteins were then extracted from leaves of three tobacco plants coming from independent transformation events. Desalted proteins were overlaid with pentane and incubated with 20 μM FPP for 2 h. GC–MS analysis of concentrated pentane solution revealed the in vitro synthesis of τ-cadinol only from transformed plants (Fig. 4c, d).

Analysis of transcript levels for *LaCARS, LaGERDS and LaCADS* in leaf and during flower development

Quantitative RT-PCR was performed on leaf and inflorescence samples from three independent plants. Three flowering developmental stages were defined. In stage 1, the spike was short and green with flowers completely covered by bracts. In stage 2, the inflorescence was longer and a purple calyx appeared above bracts whereas in stage 3, the corolla was seen at the top of the calyx (Fig. 5). Two reverse transcription reactions were performed from each RNA extraction and each sample was analyzed in two repetitions leading therefore to 12 results per biological stage. Normalized expression relative to gene expression in leaf showed that *LaCADS* is poorly differentially expressed in flower samples relative to leaf (Fig. 6a). On the contrary, *LaCARS,* and *LaGERDS* exhibited higher transcript levels

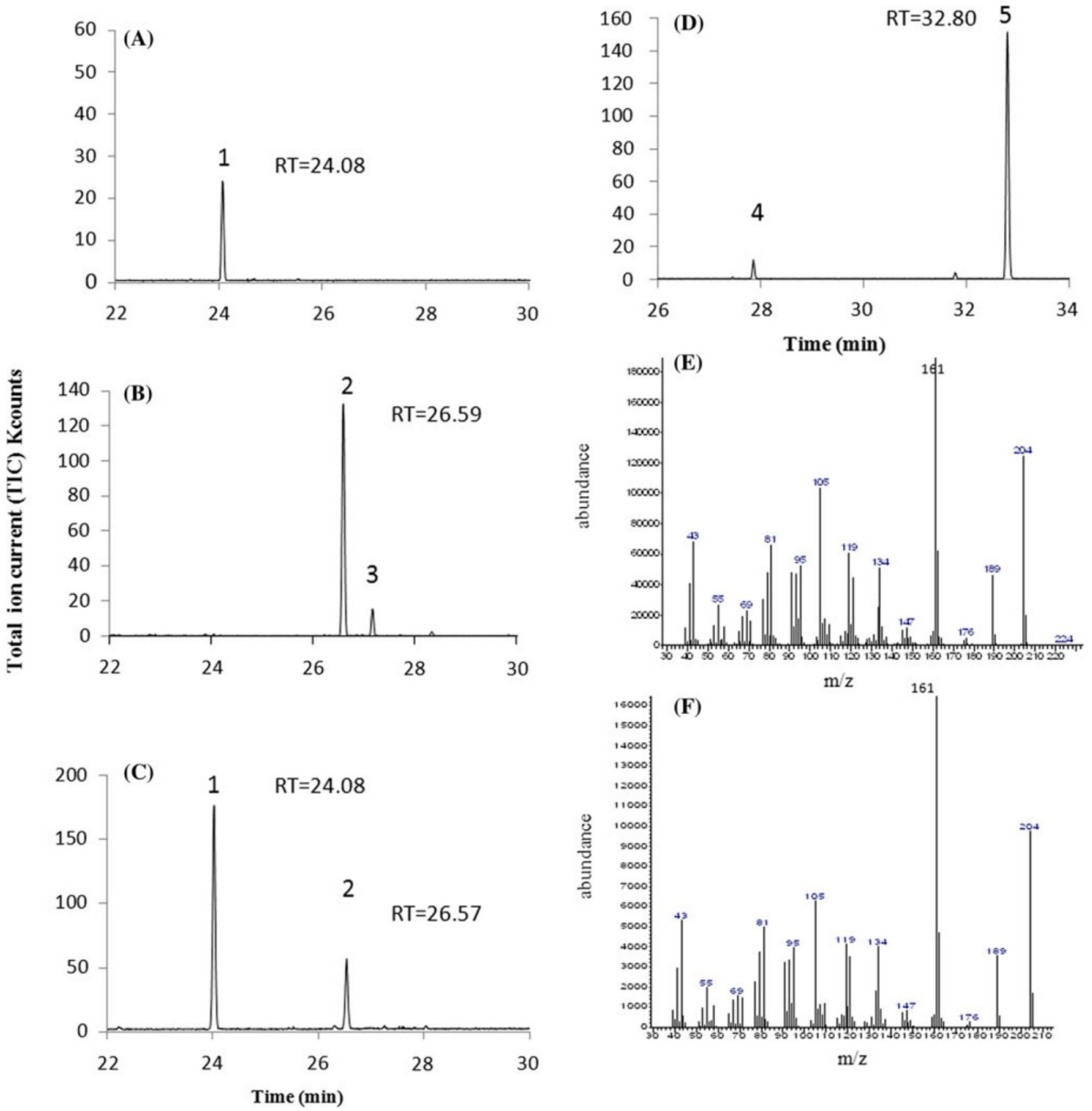


Fig. 3 Sesquiterpene products of LaCARS, LaGERDS and LaCADS measured in vitro and mass spectrum of τ-cadinol of basil and products of LaCADS. The enzyme were expressed in *E. coli*, extracted and incubated with 50 mM FPP and 10 mM $MgCl_2$ and 1 mM $MnCl_2$ ions. **a** Products of LaCARS, **b** products of LaGERDS, **c** standards of (*E*)-β-caryophyllene and germacrene D, **d** products of LaCADS. Mass spectrum of τ-cadinol: **e** in basil, **f** in products of LaCADS. Key to the terpene products was: *1* (*E*)-β-caryophyllene, *2* germacrene D, *3* bicyclogermacrene, *4* γ-cadinene, *5* τ-cadinol. *RT* retention time

compared to leaves during the first stage of flower development. The highest expression of both genes was noted in stage 2 inflorescences where transcript levels were over 30-fold higher than those of leaves.

Leaves and different flower stages were harvested on the same three plants. After hexane extraction, GC–MS analysis were performed in order to quantify τ-cadinol, (*E*)-β-caryophyllene and its related oxide and germacrene D (Fig. 6b). τ-cadinol was found in equal quantities in the different flower samples as expected from the corresponding transcript levels. Increased amounts of (*E*)-β-caryophyllene and its related oxide were noticed between stages S1 and S2 in accordance with transcript levels. However accumulation of these sesquiterpenes went forward further at stage 3

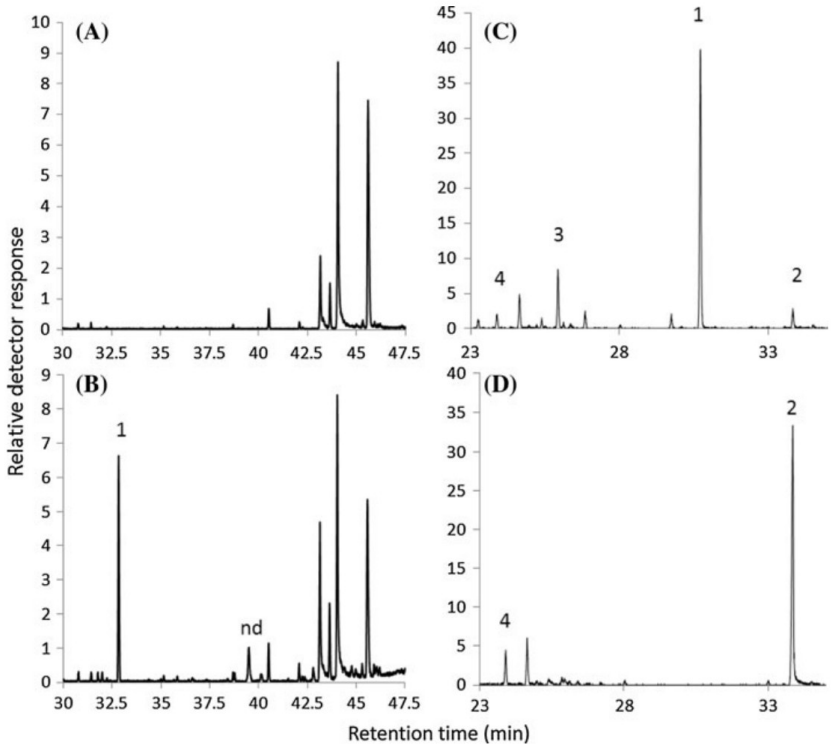


Fig. 4 Transient and stable expression of LaCADS in *Nicotiana*. (**a**, **b**) Transient expression in *N. benthamiana: A. tumefaciens* suspensions harbouring the pMDC32::CADS binary vector or the viral suppressor p19 into the pBIN61 were mixed and co-infiltrated into the adaxial side of young leaves (**b**). In control, infiltration was performed only with *Agrobacterium* carrying the viral suppressor p19 (**a**). Ten days after injection, volatiles were extracted by hexane from transformed leaves and analyzed by GC–MS. Stable expression in *N. tabacum* (**c**, **d**): leaf disks were cocultivated with *A. tumefaciens* strain C58pMP90 harboring pMDC32::CADS according to the method of Horsch et al. (1985). Transformed plants were in vitro regenerated on a selection medium with 50 mg/L hygromycin. Proteins were extracted from leaves (see section "Materials and methods") and after incubation with FPP volatiles were recovered in pentane and analyzed by GC–MS. **c** Transformed plants expressing LaCADS under the control of a double CaMV35S promoter. **d** Plants transformed with pBI121:GUS intron used as a control. Key to the terpene products was *1* τ-cadinol, *2* β-farnesol, *3* γ-cadinene, *4* β-farnesene

whereas the transcription of the related gene *LaCARS* was falling down. Correlations between transcript levels of *LaGERDS* and germacrene D synthesis during inflorescence development could also be observed but to a lesser extent. A relatively high amount of τ-cadinol and (*E*)-β-caryophyllene in leaves was an unexpected result if we consider the expression of their related genes.

Homology modeling of LaCADS and γ-cadinene synthase from 3D structures of δ-cadinene synthase

Blast analysis of the LaCADS sequence on the Uniprot webserver revealed that LaCADS clearly belongs to the terpene synthase family. γ-cadinene synthase is the best homologous protein with a sequence identity of roughly 63 % and a sequence similarity higher than 83 %. From the 250 best sequence homologs, redundant proteins were removed using a threshold of 90 % identity. Proteins with an unknown function were also removed. Finally a multiple sequence alignment using MAFFT (Katoh et al. 2002) was performed on the remaining sequences. Among them, 3D structures are known for *N. tabacum* 5-epi-aristolochene synthase (PDB id. 5EAT and 5EAU) and *Gossypium arboreum* δ-cadinene synthase (PDB id. 3G4D and 3G4F). δ-cadinene synthase (PDB identifier 3G4F and 3G4D) exhibits a relatively high sequence homology (E value $<10^{-100}$) with LaCADS and was used as template structure for homology modeling. Although the sequence identity is modest (34 %), the identical amino acids and those with similar physico-chemical properties reach 70 % of the sequence. Important residues within the δ-cadinene synthase active site are known; i.e. D307, R448, D451, E455, L526, Y527, Y535 and V536 (Gennadios et al. 2009). The coordinates of residues involved in substrate binding are well defined except for a short loop defined by residues 526-536 in δ-cadinene synthase. Online resource 4

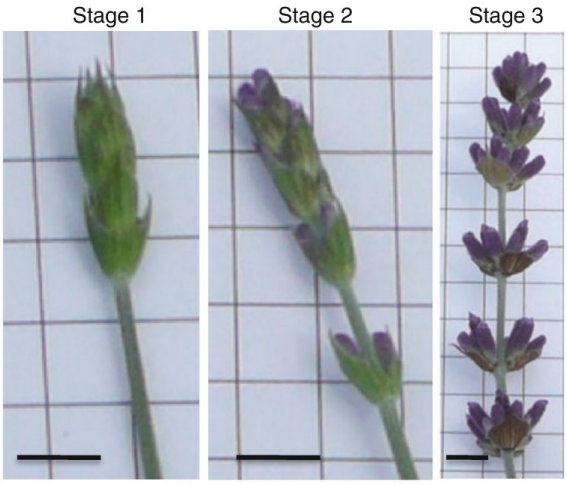


Fig. 5 Developmental stages of lavender inflorescence defined for qRT-PCR experiments and GC–MS analysis. *Scale bar* 5 mm

represents a focus of the multiple sequences' alignment on this loop. A tyrosine (numbered 527 in LaCADS) and a short sequence motif DxYT, 3–5 residues farther, are highly conserved in all aligned enzymes and are involved in substrate binding. Between these key residues, the loop region displays high sequence variability (Fig. 7). Three dimensional molecular models of LaCADS and γ-cadinene synthase of *O. basilicum* were generated with Modeller (Marti-Renom et al. 2000) using 3G4F as single template structure. The structural alignment of the LaCADS and γ-cadinene synthase models with the δ-cadinene synthase is shown together with an electrostatic potential map of the three enzymes in Fig. 8.

Directed mutagenesis on D532 of LaCADS do not change enzyme catalytic specificity

Modeling of active sites underlines the presence of an aspartate in position 532 in LaCADS that is substituted by a glycine in several cadinene synthases. Alignment of γ-cadinene synthase of *O basilicum* and LaCADS revealed a strong sequence homology for the active site. However the aspartate in position D532 in LaCADS is replaced by a glycine in the γ cadinene synthase (Fig. 7). The hydrophilic property of aspartate could help water to get inside the active site in LaCADS. In order to test this hypothesis, direct mutagenesis was performed on LaCADS leading to the substitution of aspartate 532 by a glycine. This mutation could be performed by a single A/G base substitution at position 1595 of *LaCADS* cDNA sequence. Alternatively D533, a conserved aspartate in most sTPSs, was also mutated and replaced by a glycine. The mutated sequences were transferred into the pHGGWA expression vector and

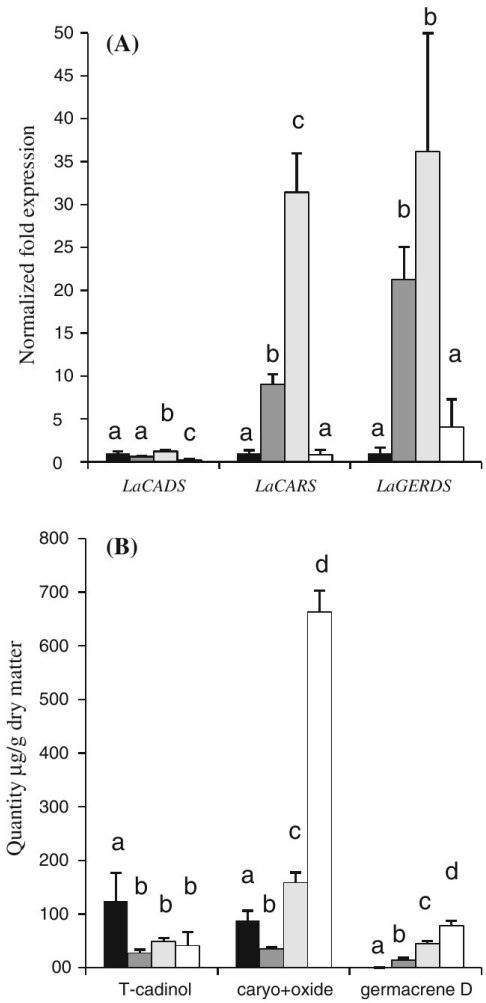


Fig. 6 Transcriptional activity of terpene synthases and terpenes accumulation in leaf and during inflorescences development in *L. angustifolia* cv Diva. **a** Transcript levels of *LaCADS*, *LaCARS* and *LaGERDS* normalized to β-actin, tubulin and GAPDH. **b** Sesquiterpenes accumulation corresponding to the activity of LaCADS, LaCARS and LaGERDS. Caryo + oxide: (*E*)-β-caryophyllene + caryophyllene oxide. *Bars colors*: leaf (*black*), S1 (*dark grey*), S2 (*light grey*) and S3 (*white*). Statistical analyses were performed with the software REST for qPCR data and a ANOVA followed by a test of Fisher for terpene analysis. Values with different letters indicated a difference with a p value <0.05. *Error bars* indicate standard deviation n = 12 in **a** and n = 3 in **b**

expressed using an identical protocol as the one described for LaCADS. Percentages of the different products of both mutated enzymes and LaCADS used as control were calculated from three replicate experiments (Table 1). Both mutations did not change product distribution and τ-cadinol rise to almost 90 % in all three assays. These observations were confirmed by a statistical Kruskal–Wallis test. The activity of mutated enzymes was also not different

```
                **          **
δ-Cadi  526 LYRE--GDGYTYV 536
γ-Cadi  511 SYNRNNGDGYTDP 523
LaCADS  526 SYTENNDDGYTDP 538
                **  *       *
```

Fig. 7 Part of the sequence alignment of LaCADS, γ-cadinene synthase of *O. basilicum* and δ-cadinene synthase from *Gossypium arboreum*. Residues involved in substrate binding are marked with a *star*. Highly conserved residues are *underlined* while the variable region is bordered by a *box*

compared to LaCADS. The amount of τ-cadinol produced from assays with 60 µg of LaCADS or mutated enzymes in D532G and D533G was evaluated respectively to 746, 649 and 814 pg. In independent repeated experiments a difference of 10–15 % of product amount could be observed between the three enzymes and the highest amount of products was never found with the same form of enzyme.

Discussion

Alterations in characteristic motifs of lavender sTPS sequences

Several structural motifs have been well characterized in the terpene synthase family among which the best known is an aspartate-rich region DDXXD involved in the binding of divalent metal ions which in turn interact with the diphosphate moiety of the substrate (Degenhardt et al. 2009). This DDXXD motif was conserved in both LaCADS and LaCARS whereas one aspartate was substituted by a glutamate in LaGERDS leading to a DDXXE amino-acid sequence. The same alteration has been reported only twice: in a bicyclogermacrene D synthase (OvBicyclGerDS) of oregano (Crocoll et al. 2010) and a 5-epi-aristolochene synthase of chili pepper (Zavala-Paramo et al. 2000). Considering the strong sequence identity (66 %) between LaGERDS and OvBicyclGerDS and the unusualness of such alteration we can hypothesize that this alteration might have occurred before the separation of genera *Lavandula* and *Origanum* in the *Lamiaceae* family.

A second motif designated NSE/DTE is also responsible for metal cofactor binding. This motif has apparently evolved from a second DDXXD aspartate-rich motif conserved in prenyl transferases to form a consensus sequence (N,D)D(L,I,V)x(S,T)xxxE (Christianson 2006). In LaCADS this consensus sequence was DDIxSxxxE whereas in LACARS and LaGERDS a DDLxGxxxE sequence was found underlying an alteration of the consensus where serine/threonine was substituted by a glycine. Such substitution has already been reported in a number of plant mono- and sesquiterpene synthases. Recent mutational investigations of the NSE/DTE motif led to the speculation that the presence of a water molecule close to the glycine could replace the missing hydroxyl side chain of serine/threonine without reducing the efficiency of this sequence to bind metal cofactors (Zhou and Peters 2009).

Several investigations with tritium-labeled GPP and non-cyclizable GPP analogs (Wheeler and Croteau 1987) led to conclude that a prerequisite to the formation of cyclic

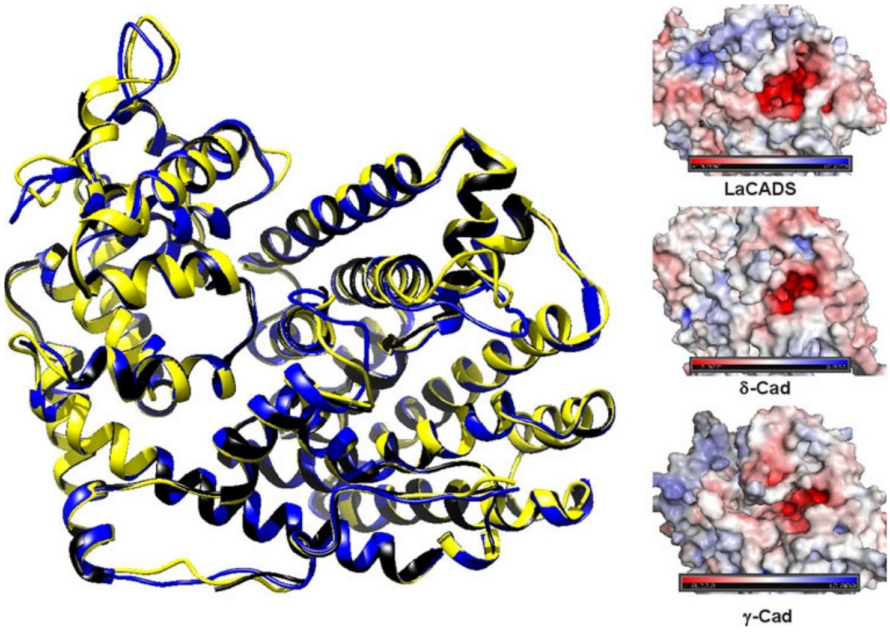


Fig. 8 Structural alignment (*left*) of homology models LaCADS (*yellow*) and γ-cadinene synthase from *O. basilicum* (*blue*) with the known 3D structure of δ-cadinene synthase from *Gossypium arboreum* (*black*). ±10 kbT/e electrostatic potential mapped onto the molecular surface of the proteins (*right*)

Table 1 Percentage distribution of the products biosynthesized by LaCADS and mutated enzymes after 2 h-incubation

Enzymes	γ-Cadinene	Unknown sesquiterpene	τ-Cadinol
LaCADS	8.6 ± 1.3	3.4 ± 1.0	87.9 ± 1.8
Mutation D532G	6.0 ± 0.6	4.5 ± 1.5	89.4 ± 1.8
Mutation D533G	6.3 ± 0.7	3.7 ± 0.3	89.9 ± 1.0

Experiments were repeated once with 15 µg and twice with 60 µg protein. SD, standard deviation. A kruskal–Wallis test was performed with the Excel-STAT software. The *p* value for γ-cadinene, unkown sesquiterpene and τ-cadinol data were respectively 0.066, 0.561 and 0.177 then all superior to 0.05 indicating that no statistical difference was observed for both products synthesized by LaCADS and both mutated enzymes

monoterpenes is the isomerization of the geranyl diphosphate cation into the linalyl diphosphate cation. A $RR(x)_8W$ motif close to the transit peptide may play a role in the initiation of the isomerization-cyclization reaction (Williams et al. 1998) or act to stabilize the protein (Hyatt et al. 2007). This motif is present in all mTPSs leading to the synthesis of cyclic monoterpenes but also conserved with variations in most sesquiterpene and diterpene synthases. This motif could therefore also be implicated in the isomerization reaction of the transoid farnesyl cation into the cisoid nerolidyl cation resulting in the synthesis of cyclic sesquiterpene. However the $RR(x)_8W$ motif was conserved in LaGERDS but one and two arginines were missing respectively in LaCARS and LaCADS. Considering that the biosynthesis of both τ-cadinol and germacrene D need, in a first step, the same isomerization of farnesyldiphosphate cation in nerolidyl phosphate cation, we can therefore assume that the $RR(x)_8W$ is not needed for the isomerization of FPP in nerolidyl diphosphate.

Sesquiterpene synthases are differentially expressed in *L. angustifolia*

Gene expression of the three characterized sesquiterpene synthases was studied by q-PCR. Two different patterns of expression could be distinguished. *LaCADS* showed a constitutive expression in the leaf as well as in early flowering stages. A high concentration of cadinol has been reported in leaves of several species in the genus *Teucrium* (Awadh-Ali et al. 2008; El-Shazly and Hussein 2004). In vitro experiments demonstrated that α-cadinol and τ-muurorol, two stereoisomers of τ-cadinol, have effective antifungic properties (Cheng et al. 2004). Therefore LaCADS could be expressed as a constitutive chemical defence in lavender. By contrast, *LaCARS* and *LaGERDS* showed an inducible response of transcription from stage 1 to stage 2 of floral development. Accumulation of (*E*)-β-caryophyllene and caryophyllene oxides as well as germacrene D was correlated in the first stages of flowering with increasing transcript levels of *LaCARS* and *LaGERDS* but in the latest stage transcription levels of the two genes dropped whereas sesquiterpenes production remained high. Similar observation was reported for nerolidol synthase expression during the anthesis of snapdragon (Nagegowda et al. 2008). *LaCARS* and *LaGERDS* are clearly regulated at the transcriptional level in early stages as already reported for several genes encoding terpene synthase (Tholl 2006; Nagegowda 2010) but stability of the enzymes (LaCADS and LaGERDS) and accumulation of products could explain that terpene production still increased in stage 3 whereas corresponding transcripts decreased. (*E*)-β-caryophyllene and germacrene D are common attractive sesquiterpenes and several pollinators such as flies have antennal receptors for these sesquiterpenes (Ibanez et al. 2010). Therefore increased biosynthesis of (*E*)-β-caryophyllene and germacrene D in the first stages of flowering could be related to pollination but additional studies are needed to evaluate by headspace the emission of these sesquiterpenes and their specific role in pollination.

Lavender sesquiterpene synthases come from different duplication processes

A phylogenetic tree was constructed using both the available terpene synthases of *L. angustifolia* and sesquiterpene synthases of *Lamiaceae* (Fig. 9). The three sTPS characterized in this study were clustered in the same TPS-a subfamily as expected from sesquiterpene synthases whereas α-bergamotene synthase (LaBERGS) fell in the subfamily TPS-b as reported by Landmann et al. 2007 with a zingiberene synthase of *O. basilicum*. These two sTPSs probably come from ancestral mTPSs that show convergent evolution through the lost of plastidial transit peptide and their neofunctionalization in sTPS. LaCARS and LaGERDS belong to the same clade and exhibit a high similarity compared to LaCADS. Gene duplication could have occurred before lavender and origano separation leading in both species to a (*E*)-β-caryophyllene synthase and a germacrene D/bicylcogermacrene D synthase. LaCADS clustered with the γ-cadinene synthase of *O. basilicum* sesquiterpene synthase in a separate clade from LaCARS and LaGERDS. Ancestral gene duplication could probably explain the emergence of these two divergent clades.

The activity of LaCADS needs the presence of a water molecule in the active site

Terpene alcohols can occur from the catalysis of hydroxylation by a P450 enzyme. Such reactions have been exemplified through the biosynthesis of menthol from limonene (Lupien et al. 1999) or by the oxidative transformation of epi-aristolochene into capsidiol (Ralston et al.

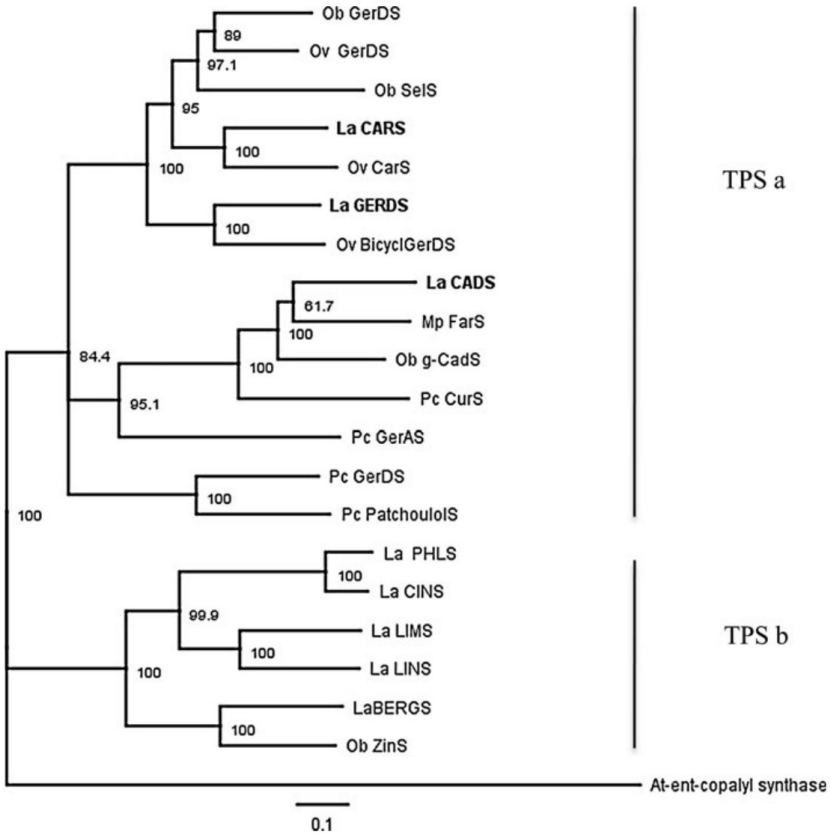


Fig. 9 Dendrogram analysis of monoterpene synthases of *L. angustifolia* and sesquiterpene synthases of *Lamiaceae* using the neighbour-joining method. At *Arabidopsis thaliana* ent-copalyl synthase (NM116512); *La Lavandula angustifolia* BERGS, (DQ263741); CADS, (JX401282); CARS, (JX401283); CINS, (JN701461); GERDS, (JX401284); LIMS, (DQ263740); LINS, (DQ263741); PHLS, (HQ404305) *Mp Mentha piperita* Fars, (O48935), *Ob Ocimum basilicum GerDS* germacrene D synthase (AY693644), *g-CadS* γ-cadinene synthase (AY693645); *ZinS* zingiberene synthase (AY693646), *SelS* selinene synthase (AY693643), *Ov Origanum vulgare* CarS (*E*)-β-caryophyllene synthase (GU385970); GerDS (GU385976), BicyclGerDS bicyclogermacrene D (GU385973); *Pc* Pogostemon cablin CurS, curcumene synthase (AY508726); *GerAS* germacrene A synthase (AY508728), GerDS (AY508727), patchoulol synthase (AY508730)

2001). Terpene alcohols can also be synthesized directly inside the active site of terpene synthases by deprotonation of a water molecule leading to the hydroxylation of the carbocation that recovers its stability. Kampranis et al. (2005) demonstrated that the ability for water capture of the active site was crucial to catalyze the formation of the monoterpene alcohol 1–8 cineole. We propose a similar reaction for the biosynthesis of τ-cadinol. After dephosphorylation, FPP could undergo several transformations in the active site of LaCADS to form the cadinyl carbocation. This molecule could recover its stability by hydroxylation from the deprotonation of a water molecule (Fig. 10). A similar reaction mechanism has been proposed for a cedrol synthase (Mercke et al. 1999) and a eudesmol synthase (Yu et al. 2008). To support this hypothesis we performed the homology modeling of LaCADS and *O. basilicum* γ-cadinene synthase from the template structure of *G. arboreum* δ-cadinene synthase. As shown in Fig. 8 the electrostatic properties within the active site depend on the surrounding amino-acids. The cavity is negatively charged for all the enzymes, in agreement with stabilization of a cadinyl cation. However, the active site of LaCADS is slightly more accessible to solvent molecules. The estimated volume of LaCADS, γ-cadinene and δ-cadinene synthase active site using the fpocket algorithm (Le Guilloux et al. 2009) is roughly 1,520, 1,460 and 1,430 $Å^3$ respectively. The alignment between *O. basilicum* γ-cadinene synthase and LaCADS showed strong homology of the active site except for the aspartate D532. Directed mutagenesis on LaCADS (D532G) did not change the catalytic properties of the enzyme underlying that the better facility for water to get inside the active site of LaCADS compared to the γ-cadinene synthase is rather depending on the global shape of the active site than to the ability of

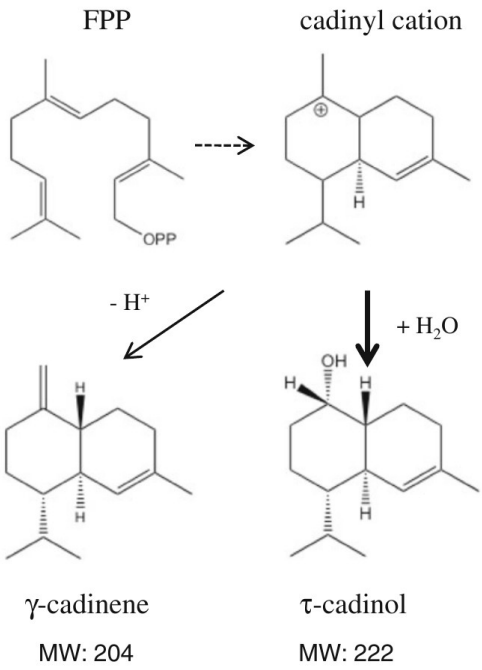


Fig. 10 Proposed biosynthetic pathway of τ-cadinol and γ-cadinene from FPP catalyzed by LaCADS. *MW* molecular weight

one hydrophilic amino acid D532 to trap a water molecule (Greenhagen et al. 2006).

We also address the involvement of D533 in LaCADS in catalysis since this residue is highly conserved in most of the sTPSs sequences (online resource 4). Despite this high conservation, mutagenesis on this amino acid (D533G) did not change the activity of the mutated protein and the main product was still τ-cadinol with γ-cadinene as a secondary product. This result raises questions on the actual function of this amino acid and the mechanism underlying its conservation. From a structural point of view, D533 belongs to a secondary pocket, containing the isoprenoid moiety of the substrate and adjacent to the main active site as it was suggested for the γ-cadinene synthase of *G. arboreum* (Gennadios et al. 2009). Despite its vicinity with the substrate, the loop region defined by residues 526–536 seems to be more flexible than the rest of the active site (i.e. undefined coordinates in the X-ray structure). One may think that D533 does not play a direct role in the reaction mechanism but may act as a modulator of the substrate recognition.

In conclusion, the characterization of these three sTPSs contributes to a better understanding of terpene biosynthetic pathways in lavender. Considering previous studies, four mTPSs and four sTPSs have been characterized in this species (Landmann et al. 2007; Demissie et al. 2011, 2012) and this plant is therefore one of the *Lamiaceae* species where the terpene biosynthetic pathway has been best depicted.

Acknowledgments This work was supported by a Grant of the Rhône-Alpes county. We thank Dr. Busso who kindly provided the pHGGWA vector. We also thank Florence Gros for her technical contribution to this work. The IFF company and CRIEPPAM are also thanked for providing respectively basil rich τ-cadinol and *L. angustifolia* cv'Diva' essential oils used as standards for τ-cadinol identification. We also thank Dr. Bendahmane (ENS Lyon) for critical reading and Frédéric Hache and Neil Finn (Department of Foreign Languages, UJM) for text revision).

References

Adams RP (2007) Identification of essential oil components by gas chromatography/mass spectrometry, 4th edn. Allured Publ Corporation, Carol Stream

Andersen L, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization applied to bladder and colon cancer data sets. Cancer Res 64:5245–5250

Arigoni D, Sagner S, Latzel C, Eisenreich W, Bacher A, Zenk MH (1997) Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeleton rearrangement. Proc Natl Acad Sci USA 94:10600–10605

Awadh-Ali NA, Wurster M, Arnold N, Lindequist U, Wessjohan L (2008) Chemical composition of the essential oil of *Teucrium yemense* Deflers. Rec Natl Prod 2:25–32

Bick JA, Lange BM (2003) Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. Arch Biochem Biophys 415:146–154

Bradford M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Busso D, Busso BD, Moras D (2005) Construction of a set gateway-based destination vectors for high throughput cloning and expression screening in *Escherichia coli*. Anal Biochem 343:313–321

Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Goetz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2012) AMBER 12. University of California, San Francisco

Cheng SS, Wu CL, Chang HT, Kao YT, Chang ST (2004) Antitermitic and antifungal activities of essential oil of *Calocedrus formosana* leaf and its composition. J Chem Ecol 30:1957–1967

Christianson DW (2006) Structural biology and chemistry of the terpenoid cyclases. Chem Rev 106:3412–3442

Crocoll C, Asbach J, Novak J, Gershenzon J, Degenhardt J (2010) Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. Plant Mol Biol 73:587–603

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol 133:462–469

Degenhardt J, Köllner TG, Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry 70:1621–1637

Demissie ZA, Sarker LS, Mahmoud SS (2011) Cloning and functional characterization of β-phellandrene synthase from *Lavandula angustifolia*. Planta 233:685–696

Demissie ZA, Cella MA, Sarker LS, Thompson TJ, Rheault MR, Mahmoud S (2012) Cloning, functional characterization and genomic organization of 1,8-cineole synthases from *Lavandula*. Plant Mol Biol 79:393–411

El-Shazly AM, Hussein KT (2004) chemical analysis and biological activities of the essential oil of *Teucrium leucocladum* Boiss. (Lamiaceae). Biochem Syst Ecol 32:665–674

Gennadios HA, Gonzalez V, Di Costanzo L, Li A, Yu F, Miller DJ, Allemann RK, Christianson DW (2009) Crystal structure of (+)-d-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. Biochemistry 48:6175–6183

Greenhagen BT, O'Maille PE, Noel JP, Chappell J (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthase. Proc Natl Acad Sci USA 103:9826–9831

Guitton Y, Nicolè F, Moja S, Valot N, Legrand S, Jullien F, Legendre L (2010) Differential accumulation of volatile terpene and terpene synthase mRNAs during lavender (*Lavandula angustifolia* and *L.* × *intermedia*) inflorescence development. Physiol Plant 138:150–163

Gutensohn M, Orlova I, Nguyen TTH, Davidovich-Rikanati R, Ferruzzi MG, Sitrit Y, Lewinsohn E, Pichersky E, Dudareva N (2013) Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. Plant J 75:351–363

Hellens R, Allan AC, Friel EN, Bolitho K, Grafton K, Templeton MD, Karunairetnam S, Gleave AP, Laing WA (2005) Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods 1:13. doi:10.1186/1746-4811-1-13

Horsch RB, Fry JE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT (1985) A simple and general method for transferring genes into plants. Science 227:1229–1231

Hyatt DC, Youn B, Zhao Y, Santhamma B, Coates RM, Croteau RB (2007) Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. Proc Natl Acad Sci USA 104:5360–5365

Ibanez S, Dötterl S, Anstett M-C, Baudino S, Caissard J-C, Gallet C, Deprés L (2010) The role of volatile organic compounds, morphology and pigments of globeflowers in the attraction of their specific pollinating flies. New Phytol 188:451–463

Kampranis SC, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Anssour S, Dunwell JM, Degenhardt J, Makris AM, Goodenough PW, Johnson CB (2005) Rational conversion of substrate and product specificity in a Salvia monoterpene synthase: structural insights into the evolution of terpene synthase function. Plant Cell 19:1994–2005

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

Landmann C, Fink B, Festner M, Dregus M, Engel KH, Schwab W (2007) Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). Arch Biochem Biophys 465:417–429

Lane A, Boecklemann A, Woronuk GN, Sarker L, Mahmoud SS (2010) A genomics resource for investigating regulation of essential oil production in *Lavandula angustifolia*. Planta 231:835–845

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics 10:168

Lichtenthaler HK, Rohmer M, Schwender J (1997) Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. Physiol Plant 101:643–652

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408

Lupien S, Karp F, Wildung M, Croteau R (1999) Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. Arch Biochem Biophys 368:181–192

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325

Mercke P, Crock J, Croteau R, Brodelius PE (1999) Cloning, expression and characterization of epi-cedrol synthase, a sesquiterpene cyclase from *Artemisia annua* L. Arch Biochem Biophys 369:213–222

Nagegowda DA (2010) Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. FEBS Lett 584:2965–2973

Nagegowda DA, Gutensohn M, Wilkerson CG, Dudareva N (2008) Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J 55:224–239

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in realtime PCR. Nucleic Acids Res 30:e36

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Bestkeeper-Excel-based tool using pair-wise correlations. Biotechnol Lett 26:509–515

Ralston L, Kwon ST, Schoenbeck M, Ralston J, Schenk DJ, Coates R, Chappell J (2001) Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1-3-dihydroxylase from tobacco (*Nicotiana tabacum*). Arch Biochem Biophys 393:222–235

Rohmer M, Knani M, Simonin P, Sutter B, Sahm H (1993) Isoprenoid biosynthesis in bacteria—a novel pathway for the early steps leading to isopentenyl diphosphate. Biochem J 295:517–524

Rostkowski M, Olsson MH, Sondergaard CR, Jensen JH (2011) Graphical analysis of ph-dependent properties of proteins predicted using PROPKA. BMC Struct Biol 11:6

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz SA, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Simkin AJ, Guirimand G, Papon N, Courdavault V, Thabet I, Ginis O, Bouzid S, Giglioli-Guivarc'h N, Clastre M (2011) Peroxisomal localisation of the final steps of the mevalonic acid pathway *in planta*. Planta 234:903–914

Tholl D (2006) Terpene synthases and the regulation, diversity and the biological roles of terpene metabolism. Curr Opin Plant Biol 9:1–8

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: research0034.1–0034.1

Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J 33:949–956

Wheeler CJ, Croteau R (1987) Direct demonstration of the isomerization component of the monoterpene cyclase reaction using a cyclopropylcarbinyl pyrophosphate substrate analog. Proc Natl Acad Sci USA 84:4856–4859

Williams DC, McGarvey DJ, Katahira EJ, Croteau R (1998) Truncation of limonene synthase preprotein provides a fully active pseudomature form of this monoterpene cyclase and

reveals the function of the amino-terminal arginine pair. Biochemistry 37:12213–12220

Yu F, Harada H, Yamasaki K, Okamoto S, Hirase S, Tanaka Y, Misawa N, Utsumi R (2008) Isolation and functional characterization of a b-eudesmol synthase, a new sesquiterpene synthase from Zingiber *zerumbet* Smith. FEBS Lett 582:565–572

Zavala-Paramo G, Chavez-Moctezuma MP, Garcia-Pineda E, Yin S, Chappell J, Lozoya-Gloria E (2000) Isolation of an elicitor-stimulated 5-epi-aristolochene synthase (gPEAS1) from chili pepper (*Capsicium annuum*). Physiol Plant 110:410–418

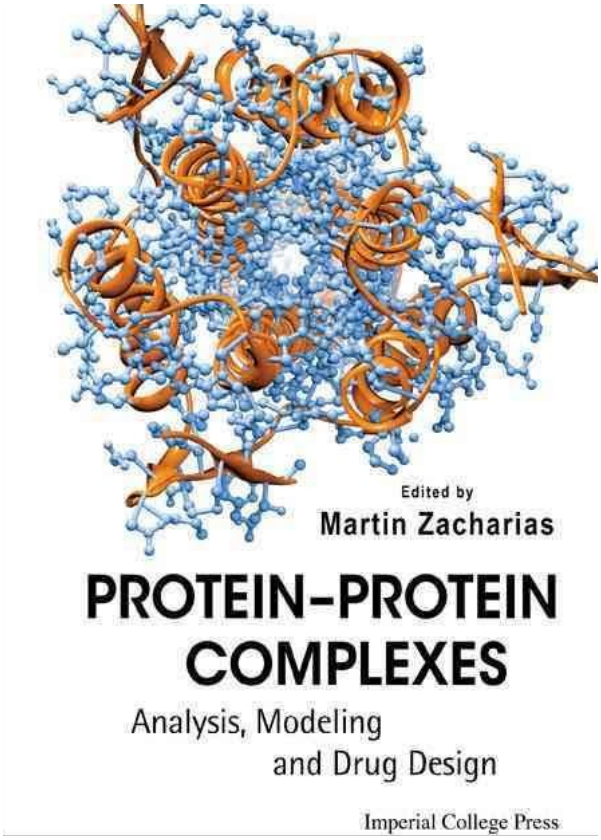
Zhou K, Peters RJ (2009) Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. Phytochemistry 70:366–369

Article 5:

Prediction/calculation of protein-protein binding affinities and mutation effect.

S. Fiorucci*, S. Antonczak, J. Golebiowski.

in *Protein-protein complexes: Analysis, modeling and drug design*, M. Zacharias, *World Scientific (**2010**) p.295-317.*



CHAPTER 11

PREDICTION AND CALCULATION OF PROTEIN–PROTEIN BINDING AFFINITIES AND MUTATION EFFECTS

Sébastien Fiorucci, Serge Antonczak, Jérôme Golebiowski

, Molecular Modelling Team, UFR Sciences,
Centre National de la Recherche Scientifique,
Université de Nice-Sophia Antipolis,
UMR6001 LCMBA, 28 Avenue Joseph Vallot,
06108 Nice Cedex2, France
Email: Sebastien.Fiorucci@unice.fr

This chapter reviews the most popular theoretical methods to compute changes in protein–protein binding affinities and relative free energies due to substitution of amino acid residues. It includes techniques for computing free energy changes associated with alchemical mutations, like free energy perturbation or thermodynamic integration, as well as more approximate methods such as the linear interaction energy method and approaches to combine molecular mechanics calculations and continuum descriptions of the surrounding solvents and ions. The applicability of the methods for calculating protein–protein interactions is also discussed.

11.1 Introduction

Molecular sciences have reached a point where manipulation either experimentally or theoretically of simple molecular systems is often no more a severe limitation to the exploration of their functionality. However, for more complex systems, several issues still need to be addressed in order to ensure that theoretical results are in agreement with experimental data. This is particularly true in the framework of

molecular modelling, where the size and complexity of studied structures are steadily increasing. The remarkable efforts and progress during the recent years have allowed us to gain detailed and relevant energetic descriptions associated with structural changes or interactions of multiple structures. Still, a lot needs to be done but many simulation approaches have matured to be applicable to a variety of biomolecular systems. In this chapter we will summarise the most common protocols that allow us to estimate protein–protein affinities and mutation effects.

Protein–protein association plays a crucial role in many biological processes including signal transduction, cell growth regulation, metabolism and adhesion, immune response and others.[1] Understanding how these macromolecular complexes are formed and what determines their specificity is not only fundamental for appreciating the underlying biological processes but also helpful in developing new therapeutic strategies. With the advent of the genomic and post-genomic era and the steady increase of computer capacity and speed, theoretical studies of highly complex systems become accessible. In combination with alanine scanning, single and multiple mutant cycles or saturation mutagenesis experiments, theoretical approaches allow the screening of a wide panel of amino acid sequences and provide energetic and structural information on the studied systems. In such context, computational protein design strategies have even been developed to engineer synthetic protein–protein interfaces.[2–5]

The tendency of molecular systems to react or to associate is represented by a thermodynamic quantity, the change in free energy. Predictions of ligand–receptor binding affinity, as well as mutation effects, remain a challenge for computational approaches, all the more so since two major difficulties hamper an accurate prediction: (a) despite spectacular progresses in both X-ray crystallography experiments and Nuclear Magnetic Resonance spectroscopy, it still remains unreasonable to hope for the experimental determination of the majority of protein structures in the near future. It is therefore necessary to build homology models[6–8] (if possible) for unknown protein structures; (b) Since many entries in the Protein Data Bank (PDB) do not describe macromolecular complexes but isolated proteins, theoretical approaches like docking methodologies[9,10] are needed to propose a binding mode for a protein–

protein interaction. In cases where the structure of a protein–protein complex is known and the mutation of a single amino acid does not induce large structural changes, computational predictions of binding free energy becomes feasible.

The accurate estimation of thermodynamic quantities can in principle be obtained by so-called 'first principle' approaches. However, the corresponding calculations are often time-consuming and not straightforward to set up. In parallel, the use of more approximate methods, based on empirical rules, becomes a possible alternative in case of high-throughput *in silico* screening.

In this chapter, we present the currently most accurate force field based methods such as Thermodynamic Integration (TI) and Free Energy Perturbation (FEP), based on a force field model to describe the physical interactions of the system. Then more approximate methods like Linear Interaction Energy (LIE) and Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) approaches will be described. Even simpler models based on empirical rules are also reviewed to complete the panel of computational approaches.

11.2 About Protein–Protein Interactions

The understanding of protein–protein interface organisation and composition contributes to identify the forces guiding the association of such macromolecular entities. The dissection of protein–protein binding sites has recently been the aim of many investigations in terms of geometry[11–14] and chemical nature of the interface (see Chapters 1 and 2 of this volume).[15–22] Considering different types of molecular assemblies such as heterodimers (protease-inhibitor, enzyme-inhibitor, antibody-antigen, etc.), homodimers or others, the distribution of amino acids at the interface differs from the rest of the surface exposed residues.[14,18] It is reflected by their interface propensities (defined as the ratio of the abundance of a given amino acid at the interface over its overall abundance on the surface) which are also quite different between the core and the rim of the interface.[16] It is commonly accepted that protein–protein interfaces are mainly composed of a buried hydrophobic core surrounded by a more hydrophilic ring partly exposed to the solvent.[12,23]

However, depending on the type of interface, even the inner part of the interface can contain some polar or charged residues. Interacting via hydrogen bonds or salt-bridges, these residues generally act as strong anchor points and maintain the structural integrity of the complex.[22,24] Although electrostatic complementarity[25] is strongly involved in protein–protein recognition, non-polar interactions play a major role in the binding affinity[21,24] since the interface is to a large extent densely packed.[15]

However, structural analysis alone cannot predict whether all of these interface contacts are important for binding. Alanine scanning experiments on human growth hormone and the extracellular domain of its first bound receptor showed for the first time that few specific residues contribute dominantly to the binding free energy.[26] Some authors have defined a 'hotspot' as a residue contributing a significant part of the binding free energy ($\Delta\Delta G > 2$ kcal·mol^{-1}) as measured by alanine substitution. Bogan and Thorn[15] also showed that hotspots are located within densely packed areas, i.e. at the inner part of the interface. Substitutions of a hotspot residue by an alternative residue may create holes or results in steric hindrance preventing a perfect fit and, thus explaining the critical loss of affinity. Clearly, the capacity to give a rationale to such mutations or to protein–protein affinities in general requires a realistic description of the associated changes of the free energy.

11.3 The Free Energy of Binding

The free energy is a thermodynamic function of state that encodes information about the equilibrium state of a system. When a system (in which the temperature, the number of particles and the volume are constant) is at equilibrium, the free energy (here the Helmholtz free energy, F) is at a minimum. Depending on the thermodynamic conditions, one can speak of either the Helmholtz free energy (F) if the number of particles (N), the volume (V) and the temperature (T) are kept constant or of the Gibbs free energy (G) if N, T and P (pressure) are constant, respectively. See Ref. 27 for detailed explanations. The accurate calculation of the free energy of a molecular system is,

however, difficult to perform. State-of-the-art approaches are rooted in a statistical thermodynamic treatment of the system. The free energy function (here called A) is directly connected to the partition function Q through the simple relation:

$$A = -kT \ln Q \tag{11.1}$$

Here, k corresponds to the Boltzmann constant and T to the temperature of the system. The partition function Q fills the gap between the macroscopic properties of a system and its microscopic representations. It can be simply described as a sum of Boltzmann factors corresponding to the partition of the particles constituting the system throughout accessible states. In a simple system of well defined localised and indistinguishable particles partitioned amongst quantified energy levels (e_i), the partition function reads:

$$Q = \frac{q^N}{N!}; q = \sum_i p_i e^{\frac{-e_i}{kT}} \tag{11.2}$$

In a more complex system, with interacting particles, the concept is identical but requires the calculation of the energy of the system through its Hamiltonian (H(r,p)), which is a function of both the positions (r) and the momenta (p) of the particles (phase space). H(r,p) is a continuous function and the summation becomes an integral.

$$Q = \frac{1}{N!}\frac{1}{h^3 N}\int \ldots \int e^{\frac{-H(r,p)}{kT}} drdp \tag{11.3}$$

In principle, this equation can give access to the partition function, provided that the integral can be computed. This is, however, out of reach for typical biomolecular systems, where the phase space (the space

of possible positions and momenta) is too large to be properly sampled with any kind of simulation to directly compute the partition function.

Fortunately, in the framework of molecular systems, one is often more interested in the difference of the free energies between two states than in the absolute free energy of a state. We indeed seek to estimate the free energy change throughout a given transformation that can be a chemical reaction, a protein folding process or the association of two protein molecules. In this case the free energy difference relies on a ratio of partition functions, which is in principle easier to estimate.

Figure 11.1 illustrates the typical thermodynamic cycle used to compute protein–protein binding affinity and the free energy change resulting from a mutation. The binding free energy ΔG_1 simply results from the association process between the two protagonists.

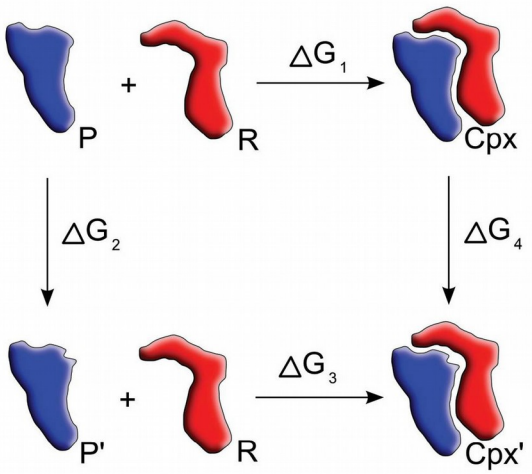


Fig. 11.1. Thermodynamic cycle used to estimate a protein–protein binding free energy and a free energy difference due to a mutation on one of the protein (P to P').

To compare the receptor (R) affinity between a wild-type protein and one of its mutant (called P') one can compute the free energy resulting from the mutation (ΔG_2), the free energy of binding of the mutant with the receptor (ΔG_3) and finally the free energy difference between the wild-type complex and the complex made up of the mutant and the receptor (ΔG_4).

The thermodynamic cycle indicates that $\Delta G_1 + \Delta G_4 - \Delta G_3 - \Delta G_2 = 0$. Consequently, the relative binding free energy is given by,

$$\Delta\Delta G_{\text{wt to mut}} = \Delta G_1 - \Delta G_3 = \Delta G_2 - \Delta G_4 \tag{11.4}$$

11.4 First Principle Methods and End-point Approaches

11.4.1 *Alchemical Mutations. Free Energy Perturbation (FEP) and Thermodynamic Integration (TI)*

The theory of free energy perturbation and thermodynamic integration approaches was initially developed in the 1950s while the first application to a biomolecular system was performed during the 1980s. For a detailed explanation and the historical development of the methods, see Ref. 28.

Let us consider a typical chemical equilibrium between two states of a given system:

$$A \leftrightarrow B \tag{11.5}$$

This equilibrium can be described by both the equilibrium constant and the corresponding free energy,

$$K = \frac{[B]}{[A]} \tag{11.6}$$

and

$$\Delta A = -kT \ln\left(\frac{Q_B}{Q_A}\right), \tag{11.7}$$

where Q_B and Q_A represent the partition functions of states B and A, respectively.

One can express the partition functions as shown above and the free energy expression reads

$$\Delta A = -kT \ln \frac{\int \ldots \int e^{\frac{-H_B}{kT}} drdp}{\int \ldots \int e^{\frac{-H_A}{kT}} drdp} \tag{11.8}$$

Considering that the difference between the partition function of states A and B is small, one can write the Hamiltonian of B as a perturbation of the Hamiltonian representing state A:

$$H_B = H_A + \Delta H \tag{11.9}$$

The change in free energy can be expressed as follows:

$$\Delta A = -kT\, ln \frac{\int \ldots \int e^{\frac{-H_A}{kT}} . e^{\frac{-\Delta H}{kT}} drdp}{\int \ldots \int e^{\frac{-H_A}{kT}} drdp} \tag{11.10}$$

This latter equation is just the ensemble average of the perturbation in the Hamiltonian, taken from a simulation obtained for the system in state A.

$$\Delta A = -kT\, ln \left\langle e^{\frac{-\Delta H}{kT}} \right\rangle_A \tag{11.11}$$

The Free Energy Perturbation (FEP) approach can be implemented either in Monte Carlo or Molecular Dynamics simulations. The approach is, in principle, exact (does not involve any approximations) and the accuracy of the calculated change in free energy depends on the sampling of configurations relevant for state B (under the control of the Hamiltonian representing state A).

From a more general point of view, even if the difference between states A and B is not significant, one can decompose the path going from A to B into several smaller steps ($A..A_1...A_n...A_z...B$) and then compute the sum of each associated small free energy change throughout the path from A to B.

The Thermodynamic Integration (TI) approach is also based on thermodynamic statistics of various simulations. In the TI scheme, one writes the free energy as a function of a coupling parameter, generally called λ. For the A to B transformation one has

$$\Delta A_{A\to B} = A(\lambda_A) - A(\lambda_B) = \int_{\lambda_A}^{\lambda_B} \frac{\partial A(\lambda)}{\partial \lambda} d\lambda \qquad (11.12)$$

It is possible to rewrite this free energy difference as a sum of averages of the Hamiltonian derivatives with respect to λ.

$$\Delta A_{A\to B} = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \qquad (11.13)$$

The free energy difference can then be obtained by a numerical integration of the ensemble average for the derivative of the Hamiltonian with respect to λ, obtained from various Monte Carlo runs or Molecular Dynamics simulations representing an alchemical transformation from A to B.

In both of these methods, the magnitude of the perturbation is critical for the accuracy of the calculated free energy difference. The free energy difference between two states might be calculated accurately, provided that state B can be considered as a small perturbation to state A. In this

case, the partition functions of states A and B overlap and the simulation sampling can be properly achieved within reasonable computing times.

11.4.2 *End-point Approaches: Molecular Mechanics Poisson–Boltzmann Surface Area (MMPBSA) and Linear Interaction Energy (LIE) Methods*

11.4.2.1 *MM-GB(PB)SA Approach*

Since the free energy is a state function, it is in principle sufficient to only evaluate the initial and the final states for computing the binding free energy.[29] The Molecular Mechanics Poisson–Boltzmann Surface Area MM-GB(PB)SA method is based on the analysis of configurations obtained from equilibrated MD simulations with explicit solvent or other approaches treating the solvent as a continuum. The total free energy of the system can be expressed as the sum of several contributions:

$$G = E_{MM} + H_{rot/trans} + G_{sol} - TS \tag{11.14}$$

Where E_{MM} is the molecular mechanics energy

$$E_{MM} = E_{bond} + E_{angle} + E_{dihedral} + E_{vdW} + E_{Coulomb} \tag{11.15}$$

G_{sol} is the solvation free energy and $H_{rot/trans}$ corresponds to the contribution due to putative changes in the translational and rotational degrees of freedom of the binding partners.

The solvation free energy is computed as a sum of polar and non-polar contributions. The non-polar contribution corresponds to cavity formation and van der Waals interactions between the solute and the solvent. This contribution is typically calculated from the solvent-accessible surface area of the molecule,

$$G_{SA} = \gamma SASA + \beta \tag{11.16}$$

where SASA is the solvent accessible surface area estimated by rolling the a solvent-sized probe over the solute surface, γ and β are constants which were extracted from a least-squares fit to a plot of experimental alkane transfer free energies versus accessible surface area.
In the MM-PBSA approach, the polar contribution of G_{sol} is obtained by a calculation of the electrostatic potential $\phi(r)$ from a solution of the Poisson (or Poisson–Boltzmann) equation,

$$\nabla\varepsilon(r)\nabla\phi(r)+4\pi\rho(r)=0 \qquad (11.17)$$

here $\varepsilon(r)$ is a position dependent dielectric constant, and $\rho(r)$ is the charge distribution of the solute.

Alternatively, in the MM-GBSA approach the electrostatic component is calculated using the Generalized Born equation. In the GB equation, the protein atoms are represented by spheres with a dielectric constant different from that of the exterior of the protein (solvent). The electrostatic energy can be calculated by,

$$G_{sol(polar)}=\left(1-\frac{1}{\varepsilon}\right)\sum\frac{q_i q_j}{\sqrt{r_{ij}^2+\alpha_i\alpha_j\exp-\left(r_{ij}^2/2\alpha_i\alpha_j\right)}} \qquad (11.18)$$

with q_i and α_i are, respectively, the charges and the effective Born radii of atoms i and j.

The conformational entropic term TS is often estimated by a normal mode analysis of the complex and the isolated protein partners.

In order to calculate free energy changes associated with the complex formation the free energies of isolated protein partners are subtracted from free energies calculated for conformations of the complex:

$$\Delta G_{bind}=\langle G_{complex}\rangle-\langle G_{protein1}\rangle-\langle G_{protein2}\rangle \qquad (11.19)$$

Where $\langle G_x\rangle$ corresponds to the average of the total free energy of the complex or isolated partner over snapshots taken from the MD trajectory.

An interesting feature of such a description of the free energy change is the possibility to decompose it on a per-residue basis.

Such a per-residue decomposition allows a semi-quantitative evaluation of alanine-scanning scoring by replacing a given residue side chain with an alanine side chain and performing the same free energy calculation. Often this is performed using the trajectory obtained for the wild type proteins by replacing in each snapshot a given side chain by alanine.[30,31]

11.4.2.2. *Linear Interaction Analysis*

The Linear Interaction Analysis (LIE) scheme is based on the idea that when a solute binds to a receptor, the change in free energy can be decomposed into polar and non-polar contributions.[32] The linear response theory is invoked to estimate the polar (electrostatic) component and the non-polar contribution is considered to scale proportional to the intermolecular van der Waals interaction energy, averaged over molecular dynamics simulations. The binding free energy can thus be written as:

$$\Delta G_{bind} = \alpha \Delta \left\langle V_{l-s}^{vdW} \right\rangle + \beta \Delta \left\langle V_{l-s}^{el} \right\rangle + \gamma \qquad (11.20)$$

Here, $\langle x \rangle$ denotes an average over a sampled trajectory (from MD or MC), for the van der Waals (vdW) or electrostatic (el) terms involving the binding partners (*l*-*s*). The Δ stands for the difference between the solute free in solution and bound to its partner. The parameters α, β and γ are fitted with respect to experimental results on a set of test cases with known binding free energy.

Several review articles on details and technical aspects of each approach to compute binding free energies in various systems have been published.[33,34] In this section the main aspects that have to be taken into account when setting up a free energy calculation will be discussed. These calculations can be applied to the prediction of both the total binding free energy or the effect of a residue mutation on binding free energy.

11.4.3 *Applications on Protein–Protein Complex Structures*

As for many computational approaches, free energy calculation techniques are subject to a trade-off between speed and accuracy. The methods are based on a statistical evaluation of various terms. The statistics are in general more accurately determined if a large number of configurations have been sampled. It is not surprising that in the framework of protein–protein association, the number of applications of the computationally of very expensive methods such as FEP and TI is still limited, since large numbers of configurations need to be evaluated.

Due to the large computational costs of alchemical transformations, applications typically involve single mutations often in model systems with known binding energies. TI was, for example, used to compute the relative binding free energy between a wild-type peptide and its Pro6 to Ala6 mutation on recognition by the T-Cell Receptor.[35] The TI calculation gave excellent agreement with experiment and allowed also the decomposition of the free energy change into various energetic contributions which helps to explain the driving forces for binding. More generally, the decomposition into various components and more particularly on a per residue basis is a powerful tool to predict hotspots at protein binding sites and potential mutations that affect peptide or protein binding. Such prediction approaches include 'computational alanine scanning' or 'virtual alanine scanning' and are frequently used to predict hotspots in proteins that have a dominant effect on affinity for a given receptor. Variants of the MMPBSA calculation methods have been developed to explore the effect of alanine substitutions.[30,31] The MMPBSA approach involves more approximations but is computationally less expensive compared to FEP or TI methods and has been applied successfully to identify hotspot residues in proteins.[36]

Looking more closely at protein–protein interactions, the balance of hydrophobic and hydrophilic regions at the interfaces has to be properly considered. Systems dominated by hydrophobic contacts may have characteristics that are more difficult to predict.[37] This phenomenon is due to the limited directionality and often weak nature of hydrophobic interactions and due to the indirect role played by solvent molecules. The LIE approach is particularly sensitive to the hydrophobic character of a

binding region and generally requires an additional term to produce accurate free energies for systems dominated by hydrophobic contributions.[33,38] In the LIE approach hydrophobic contacts are modelled by a scaled van der Waals interaction term. The decomposition of the interaction free energy between two proteins indeed emphasized the importance of the non-polar terms.[35] The electrostatic contribution can either play the role of directional constraints[39] or can even make a repulsive contribution to binding.[40]

The LIE approach can give results in good agreement with experiment if the mutation does not involve residues sensitive to the electrostatic environment, for example due to changes in protonation states or the presence of counter ions. Initially, LIE calculations involving aspartic or glutamic acids gave results in poor agreement with experimental data.[41] Appropriate treatment of possible changes in protonation states, however, can improve the results.[42,43] If the residue is not deeply buried within the interface it is also likely to form close contacts with water molecules around the interface.

In general, the binding free energy can dependent on buried water molecules present at the interface between the partner structures. During alchemical transformation in FEP or TI calculations, the removal of chemical groups or whole side chains can create large empty cavities at the interface which in reality are filled with water molecules.[44] In such cases the FEP/TI protocol may need to include the creation or annihilation of water molecules. In case of employing implicit solvent methods one should be aware that the structural role played by explicit water molecules may not be appropriately accounted for.

The effect of the substitution of a putative hotspot at the interface can be decomposed into several parts: (a) the direct interaction energy of the new residue with its neighbours, (b) the influence of this residue on the overall structure of both the mutated protein and the complex, (c) the organisation of the water molecules around this residue, also both in the free protein and potentially in the complex.

An accurate calculation of these contributions requires sufficient sampling of conformations of the wild type and mutated complex structures. The adequate sampling of relevant states is indeed of primary importance since it represents the cornerstone to achieve convergence on

the calculation of ensemble averages of a perturbation of free energy derivative. The large size of biomolecular systems implies that a thorough sampling is unfortunately not always possible. In case of screening a large number of possible substitutions it is often necessary to use much simpler empirically parameterized approaches.

11.5 Empirical Scoring Functions

The calculation of binding free energy changes using the methods discussed above is based on a statistical analysis of Monte Carlo or Molecular Dynamics simulations and usually requires the calculation of averages over many conformations. Specific force fields[45–48] have been developed to reduce the requirement of sampling many conformations. Simplifications of the energy function may, however, result in additional parameters to reproduce experimental data.

For instance the scoring of a predicted protein–protein complex usually involves the calculation of a scoring function for a single conformation of the complex (see the review of Halperin *et al.*[10] and references therein). Problems arising from docking simulations can be, in a general manner, split into two groups: (a) the generation of a pool of protein–protein complex conformations and (bi) the scoring of the predicted structures. Searching and scoring methods can simply be based on geometric rules that tended to maximise the packing of interface atoms. Protein–protein docking algorithms can, however, use more sophisticated scoring functions taking into account solvation effects and flexibility of the residue side chains and the backbone,[9,49] and may provide a qualitative estimate of the binding free energy.

The development of a fast and reliable simplified protein force field is a difficult task due to the subtle balance between different energy terms. In the following the theoretical basis of empirical force fields used to estimate free energy changes in the context of protein folding[47,48,50] and protein design[2–5] will be discussed.

11.5.1 *Empirical Force Fields*

Several computational approaches[45–47,51] have been employed to approximately calculate binding energies and effects due to substitution of residues at the binding site. In each of these methods, several empirically parameterized terms are introduced in the energy function. A set of training structures is used to fit those terms followed by validation through blind test experiments. Among others, the FOLD-X[46] and Robetta[52] software packages are examples of such approaches and are described below.

11.5.1.1 *FOLD-X*

FOLD-X, developed by Guerois *et al.*,[46] provides a fast and quantitative estimation of mutational effects on protein stability and protein–protein association. The free energy can be decomposed into a combination of terms demonstrated to be important for protein stability:

$$\Delta G = W_{vdW}\Delta G_{vdW} + W_{solvH}\Delta G_{solvH} + W_{solvP}\Delta G_{solvP} + \Delta G_{wb} + \Delta G_{hbond} + \Delta G_{el} + W_{mc}T\Delta S_{mc} + W_{sc}T\Delta S_{sc} \quad (11.21)$$

ΔG_{vdW} is the van der Waals interactions term. ΔG_{solvH} and ΔG_{solvP} represent the polar and non-polar solvation energy, respectively. ΔG_{wb} and ΔG_{hbond} stand for stabilising effect of water bridges and hydrogen bonds, respectively, and ΔG_{el} is the electrostatic contribution of charged groups. The entropic term, described using fitted parameters,[53,54] accounts for the cost to restrict the backbone (ΔS_{mc}) and side chain (ΔS_{sc}) mobility in the folded state.

The set of parameters and weights (W_{xx} in the equation) are fitted using experimental free energy differences ($\Delta\Delta G_{wt/mut}$) of 339 single-point mutants. The predictive performance of the method was examined on a blind test database of 667 protein mutations as well as 82 protein–protein complex mutations. A good correlation was obtained between $\Delta\Delta G_{\exp}$ and $\Delta\Delta G_{calc}$ with a standard deviation smaller than 0.85 kcal·mol^{-1}.

While hydrogen bonds, van der Waals and electrostatic interactions are explicitly evaluated, no computationally expensive simulations are needed by FOLD-X. The algorithm describes implicitly some specific properties of proteins (flexibility, the existence of an unfolded state) via the entropic terms and the fitted weights.

For issues like protein–protein interaction, where a critical point is to provide a fast and reasonably accurate estimation of the energetics of the system, the FOLD-X approach may be used to test the stability of a structural model.[55,56] For instance, Tur *et al.*[57] investigated the effect of mutations on protein–protein complexes (TRAIL-DR4 and TRAIL-DR5) involved in apoptosis using the FOLD-X method. A single amino acid mutation of TRAIL was predicted to have a favourable effect on binding affinity.

11.5.1.2 *Robetta*

While the FOLD-X approach is primarily used to predict binding energies, the Robetta algorithm[52] covers different aspects of protein design. On the one side, Rosetta is an *ab initio*[58] and comparative modelling[52] tool including NMR refinement[59] and side chain interface packing.[60] On the other side, the alanine-scanning module is of particular interest in the case of protein–protein interface analysis studies. Robetta approximates the binding free energy of protein–protein complexes accounting for shape complementarity at the interface as well as polar interactions and solvation effects.[45] As in the FOLD-X approach, the free energy function is a linear combination of various terms:

$$\begin{aligned}\Delta G = {} & W_{attr}E_{LJattr} + W_{rep}E_{LJrep} + W_{HB(sc-bb)}E_{HB(sc-bb)} \\ & + W_{HB(sc-sc)}E_{HB(sc-sc)} + W_{Coul}E_{Coul} + W_{sol}G_{sol} \\ & + W_{\phi/\psi}E_{\phi/\psi}(aa) + \sum_{aa=1}^{20} n_{aa}E_{aa}^{ref}\end{aligned} \tag{11.22}$$

The free energy is decomposed into the attractive and repulsive part of a Lennard–Jones potential (E_{LJattr} and E_{LJrep}). $E_{HB(sc-bb)}$ and $E_{HB(sc-sc)}$ stand for a side chain-backbone and a side chain-side chain hydrogen

bonding terms, respectively. E_{Coul} is a Coulomb potential and G_{sol} an implicit solvation term. Two additional terms are only used for the alanine-scanning calculation: an amino acid type-dependent backbone torsion angle propensity ($E_{\phi/\psi}(aa)$) and a reference value (approximating the interactions in the unfolded state) for each amino acid (E_{aa}^{ref}). The weights (W_x) of the energy terms are parameterized using thermodynamic measurements of mutation effects[61] on both monomeric proteins and on protein interfaces. The effect of residue mutation is estimated for both the bound and the unbound states for the wild-type and the corresponding mutation, leading to the estimation of the binding free energy difference:

$$\Delta\Delta G_{bind} = \Delta G_{bind}^{MUT} - \Delta G_{bind}^{WT} \tag{11.23}$$

For a test set of more than 2500 molecular systems, the unsigned error is roughly below 1.0 kcal·mol^{-1}.[45]

The influence of each term has been evaluated by removing its contribution to the free energy. The hydrogen bonding term (derived from an environment-dependent criterion) is a critical factor to realistically predict the existence of hotspots. As already discussed for MMPBSA, the lack of an explicit representation of the solvent is likely to result in an inaccurate description when water mediated hydrogen bonds are important for the protein-protein interactions. Applied to hotspot predictions, Robetta provided reasonable predictions of protein–protein binding free energies. Recent work reported the successful application of the model on the re-design of protein–protein interfaces with an experimentally verified significant improvement of the binding affinities.[62–64] For instance, Baker and co-workers[62] were able to increase the specificity of the colicin E7 DNase-Im7 immunity protein complex at least 300-fold.

11.5.2 *Knowledge-based Force Fields*

Knowledge-based (KB) potentials derive known proteins structure information to model interatomic interactions (see Refs 48, 50 and 65). The computational cost of such approaches allows them to estimate the

free energy of a large amount of structures. According to the Boltzmann statistics, a contact potential can be related to the population of a given structural feature through the following equation:

$$V\left(P_{ij}\right) = -kT \ln P_{ij} \cong G_{ij} \tag{11.24}$$

Where P_{ij} is the probability of finding a pair of atom within a cut-off distance.

A free energy change can then be defined as a linear combination of individual contributions. The free energy can even be refined by introducing interaction contributions of higher order between a third residue (or atom) k and the coupling between i and j ($\Delta\Delta G_{ijk}$):

$$\Delta G_i = \Delta G_i^{intrinsic} + \sum_j \Delta G_{ij} + \sum_j \sum_k \Delta\Delta G_{ijk} + \ldots \tag{11.25}$$

Here, $\Delta G_i^{intrinsic}$ represents the intrinsic change in free energy relative to the unfolded state.

Knowledge-based energy functions were used in protein folding or to evaluate the interaction energy in various receptor–ligand systems, including protein–protein,[51,66] protein–RNA[67] or protein–DNA[68] complexes.

11.6 About Computation Time

Computational demand is a critical issue for selecting an appropriate method to calculate or estimate binding free energy changes. Figure 11.2 gives a qualitative idea on the computational resources needed for the simulation approaches presented in the previous sections. The FEP and TI methods require an extensive and computationally demanding sampling of conformations to obtain accurate ensemble averages. Approaches based on few structures and using an empirically adjusted scoring function are largely devoid of sampling issues and are thus extremely useful for *in silico* high-throughput screening. In between these two extremes, the MMPBSA and LIE approaches usually require

shorter simulation times than the FEP/TI methods and for many purposes represent a reasonable trade-off between speed and accuracy.

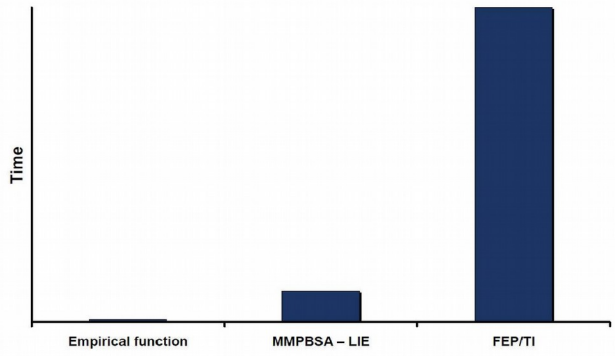


Fig. 11.2. Qualitative comparison of computation time needed to carry out a simulation using the different approaches discussed in this chapter.

11.7 Conclusions and Outlook

The ability to predict the binding mode of biomolecular complexes and to calculate the associated free energy change is a great challenge for computational chemists and biochemists. Several approaches exist that differ in computational demand and the necessary approximations to calculate free energy changes. The available approaches range from computationally very demanding free energy calculations including explicit solvent to empirically parameterized methods. The free energy simulation methods can in principle predict with a high accuracy the full structural and energetic features for biomolecular systems and are limited only by the accuracy of the force field and the sampling of relevant states. Due to the large computational demand, the application of free energy simulation approaches is often limited to the study of single mutations and its effect on protein–protein binding.

At the opposite end of the complexity scale, several approximate approaches are dedicated to produce qualitative results in very short time. These methods describe changes in the free energy of the system of interest using empirically derived terms and lack a detailed representation of the solvent and often also do not account for the flexibility of partner structures. The strength of these methods relies on a clever parameterisation and the capacity to deal with a large number of proteins structures in reasonable computer time.

References

1. Jones S., Thornton J.M. (1996). *Proc Natl Acad Sci USA* 93: 13–20.
2. Jiang L., Althoff E.A., Clemente F.R., Doyle L., Rothlisberger D., Zanghellini A., Gallaher J.L., Betker J.L., Tanaka F., Barbas C.F. 3rd, Hilvert D., Houk K.N., Stoddard B.L., Baker D. (2008). *Science* 319: 1,387–1,391.
3. Kaplan J., DeGrado W.F. (2004). *Proc Natl Acad Sci USA* 101: 11,566–11,570.
4. Kuhlman B., Dantas G., Ireton G.C., Varani G., Stoddard B.L., Baker D. (2003). *Science* 302: 1,364–1,368.
5. Looger L.L., Dwyer M.A., Smith J.J., Hellinga H.W. (2003). *Nature* 423: 185–190.
6. Forster M.J. (2002). *Micron* 33: 365–384.
7. Marti-Renom M.A., Stuart A.C., Fiser A., Sanchez R., Melo F., Sali A. (2000). *Annu Rev Biophys Biomol Struct* 29: 291–325.
8. Moult J. (2008). *Structure* 16: 14–16.
9. Gray J.J. (2006). *Curr Opin Struct Biol* 16: 183–193.
10. Halperin I., Ma B., Wolfson H., Nussinov R. (2002). *Proteins* 47: 409–543.
11. Bahadur R.P., Chakrabarti P., Rodier F., Janin J. (2004). *J Mol Biol* 336: 943–955.
12. Bahadur R.P., Zacharias M. (2008). *Cell Mol Life Sci* 65: 1,059–1,072.
13. Janin J., Rodier F., Chakrabarti P., Bahadur R.P. (2007). *Acta Crystallogr Sect D: Biol Crystallogr* 63: 1–8.
14. Jones S., Thornton J.M. (1997). *J Mol Biol* 272: 133–143.
15. Bogan A.A., Thorn K.S. (1998). *J Mol Biol* 280: 1–9.
16. Chakrabarti P., Janin J. (2002). *Proteins* 47: 334–343.
17. Keskin O., Ma B.Y., Nussinov R. (2005). *J Mol Biol* 345: 1,281–1,294.
18. Lo Conte L., Chothia C., Janin J. (1999). *J Mol Biol* 285: 2,177–2,198.
19. Reichmann D., Rahat O., Cohen M., Neuvirth H., Schreiber G. (2007). *Curr Opin Struct Biol* 17: 67–76.
20. Stites W.E. (1997). *Chem Rev* 97: 1,233–1,250.
21. Tsai C.J., Lin S.L., Wolfson H.J., Nussinov R. (1997). *Protein Sci* 6: 53–64.
22. Xu D., Lin S.L., Nussinov R. (1997). *J Mol Biol* 265: 68–84.

23. Rodier F., Bahadur R.P., Chakrabarti P., Janin J. (2005). *Proteins* 60: 36–45.
24. Xu D., Tsai C.J., Nussinov R. (1997). *Protein Eng* 10: 999–1,012.
25. McCoy A.J., Epa V.C., Colman P.M. (1997). *J Mol Biol* 268: 570–584.
26. Clackson T., Wells J.A. (1995). *Science* 267: 383–386.
27. Gasser R.P.H., Richards W.G. (1995). *An Introduction to Statistical Thermodynamics*. World Scientific.
28. Kollman P. (1993). *Chem Rev* 93: 2,395–2,417.
29. Srinivasan J., Cheatham T.E. 3rd, Cieplak P., Kollman P., Case D.A. (1998). *J Am Chem Soc* 120: 9,401–9,409.
30. Chong L.T., Duan Y., Wang L., Massova I., Kollman P.A. (1999). *Proc Natl Acad Sci USA* 96: 14,330–14,335.
31. Massova I., Kollman P. (1999). *J Am Chem Soc* 121: 8,133–8,143.
32. Aqvist J. (1994). *J Phys Chem* 94: 8,021–8,024.
33. Brandsdal B.O., Osterberg F., Almlof M., Feierberg I., Luzhkov V.B., Aqvist J. (2003). *Adv Protein Chem* 66: 123–158.
34. Wang W., Donini O., Reyes C.M., Kollman P.A. (2001). *Annu Rev Biophys Biomol Struct* 30: 211–243.
35. Michielin O., Karplus M. (2002). *J Mol Biol* 324: 547–569.
36. Taranta M., Bizzarri A.R., Cannistraro S. (2009). *J Mol Recognit* (published online: http://dx.doi.org/10.1002/jmr.1934).
37. Charlier L., Nespoulous C., Fiorucci S., Antonczak S., Golebiowski J. (2007). *PCCP* 9: 5,761–5,771.
38. Aqvist J., Marelius J. (2001). *Comb Chem High Throughput Screen* 4: 613–626.
39. Zoete V., Meuwly M., Karplus M. (2005). *Proteins* 61: 79–93.
40. Gohlke H., Kiel C., Case D.A. (2003). *J Mol Biol* 330: 891–913.
41. Brandsdal B.O., Aqvist J., Smalas A.O. (2001). *Protein Sci* 10: 1,584–1,595.
42. Almlof M., Aqvist J., Smalas A.O., Brandsdal B.O. (2006). *Biophys J* 90: 433–542.
43. Wang W., Kollman P.A. (2000). *J Mol Biol* 303: 567–582.
44. Brandsdal B.O., Smalas A.O. (2000). *Protein Eng* 13: 239–245.
45. Kortemme T., Baker D.(2002). *Proc Natl Acad Sci USA* 99: 14,116–14,121.
46. Guerois R., Nielsen J.E., Serrano L. (2002). *J Mol Biol* 320: 369–387.
47. Pokala N., Handel T.M. (2005). *J Mol Biol* 347: 203–227.
48. Vajda S., Sippl M., Novotny J. (1997). *Curr Opin Struct Biol* 7: 222–228.
49. Smith G.R., Sternberg M.J. (2002). *Curr Opin Struct Biol* 12: 28–35.
50. Poole A.M., Ranganathan R. (2006). *Curr Opin Struct Biol* 6: 508–513.
51. Jiang L., Gao Y., Mao F., Liu Z., Lai L. (2002). *Proteins* 46: 190–196.
52. Chivian D., Kim D.E., Malmstrom L., Bradley P., Robertson T., Murphy P., Strauss C.E., Bonneau R., Rohl C.A., Baker D. (2003). *Proteins* 53 Suppl 6: 524–533.
53. Guerois R., Serrano L. (2000). *J Mol Biol* 304: 967–982.
54. Hurley J.H., Mason D.A., Matthews B.W. (1992). *Biopolymers* 32: 1,443–1,446.
55. Tokuriki N., Stricher F., Serrano L., Tawfik D.S. (2008). *PLoS Comput Biol* 4: e1000002.

56. Kolsch V., Seher T., Fernandez-Ballester G.J., Serrano L., Leptin M. (2007). *Science* 315: 384–386.
57. Tur V., van der Sloot A.M., Reis C.R., Szegezdi E., Cool R.H., Samali A., Serrano L., Quax W.J. (2008). *J Biol Chem* 283: 20,560–20,568.
58. Simons K.T., Kooperberg C., Huang E., Baker D. (1997). *J Mol Biol* 268: 209–225.
59. Bowers P.M., Strauss C.E., Baker D. (2000). *J Biomol NMR* 18: 311–318.
60. Kuhlman B., Baker D. (2000). *Proc Natl Acad Sci USA* 97: 10,383–10,388.
61. Kumar M., Bava K., Gromiha M., Prabakaran P., Kitajima K., Uedaira H., Sarai A. (2006). *Nucleic Acids Res* 34: D204–206.
62. Joachimiak L.A., Kortemme T., Stoddard B.L., Baker D. (2006). *J Mol Biol* 361: 195–208.
63. Kortemme T., Joachimiak L.A., Bullock A.N., Schuler A.D., Stoddard B.L., Baker D. (2004). *Nat Struct Mol Biol* 11: 371–379.
64. Sammond D.W., Eletr Z.M., Purbeck C., Kimple R.J., Siderovski D.P., Kuhlman B. (2007). *J Mol Biol* 371: 1,392–1,404.
65. Pokala N., Handel T.M. (2001). *J Struct Biol* 134: 269–281.
66. Clark L.A., van Vlijmen H.W. (2008). *Proteins* 70: 1,540–1,550.
67. Chen Y., Kortemme T., Robertson T., Baker D., Varani G. (2004). *Nucleic Acids Res* 32: 5,147–5,162.
68. Zhang C., Liu S., Zhu Q., Zhou Y. (2005). *J Med Chem* 48: 2,325–2,335.