



**HAL**  
open science

# Étudier, modéliser, moduler le vivant : apport de la bioinformatique structurale

Stéphane Teletchea

► **To cite this version:**

Stéphane Teletchea. Étudier, modéliser, moduler le vivant : apport de la bioinformatique structurale. Sciences du Vivant [q-bio]. Nantes Université, 2023. tel-04381065

**HAL Id: tel-04381065**

**<https://hal.science/tel-04381065v1>**

Submitted on 8 Jan 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



# HABILITATION DIRIGER DES RECHERCHES HDR

NANTES UNIVERSITE

Spécialité : École doctorale n° 605, Biologie-Santé, Bioinformatique Structurale

Par

**Stéphane TELETCHÉA**

Etudier, modéliser, moduler le vivant : apport des approches de bioinformatique structurale

**Travaux présentés et soutenus à Nantes, le 15 décembre 2023**

**Unité de recherche :** Unité en Sciences Biologiques et Biotechnologies, Nantes Université, CNRS, UMR 6286, F-44000 Nantes

## Rapporteurs avant soutenance

**Anne-Claude CAMPROUX**  
**Florent BARBAULT**

Professeure, Université Paris Cité

Maître de conférences, Université Paris Cité

## Composition du jury :

Présidente : **Valérie TRICHET**  
Rapporteurs : **Catherine ETCHEBEST**  
**Gilles MARCOU**  
**Gwenaëlle ANDRE-LEROUX**  
Examineurs : **Jean-Yves LE QUESTEL**

Professeure, Nantes Université

Professeure, Université Paris cité

Maître de conférences, Université de Strasbourg

Directrice de Recherche, INRAE, Jouy-en-Josas

Professeur, Nantes Université



# Table des matières

<b>I</b>	<b>Préambule</b>	<b>1</b>
<b>II</b>	<b>Curriculum vitae</b>	<b>6</b>
<b>1</b>	<b>Curriculum Vitae</b>	<b>8</b>
1.1	État civil . . . . .	8
1.2	Formation . . . . .	9
1.3	Expériences professionnelles . . . . .	9
<b>2</b>	<b>Synthèse de l'activité d'enseignement</b>	<b>10</b>
<b>3</b>	<b>Synthèse de l'activité scientifique</b>	<b>12</b>
3.1	Modélisation moléculaire de macromolécules biologiques . . . . .	12
3.2	Modélisation moléculaire de protéines de la famille des RCPG et recherche d'inhibiteurs spécifiques par criblage virtuel . . . . .	14
3.3	Modélisation moléculaire de protéines, conception et validation d'inhibiteurs peptidiques . . . . .	16
3.4	Étude et modélisation des protéines membranaires du globule rouge, définition des partenaires protéiques et conception d'inhibiteurs spécifiques . . . . .	18
3.5	Étude des galectines et caractérisation d'inhibiteurs innovants . . . . .	19
3.6	Détermination de l'interface protéine-protéine entre RhoA et ArhGEF1 . . . . .	21
<b>4</b>	<b>Production scientifique</b>	<b>23</b>
4.1	Encadrement d'étudiants de M2 recherche . . . . .	25
4.2	Co-encadrement de thèses . . . . .	27
4.2.1	2016-2019, Ennys GHEYUCHE, co-encadrement à 40% . . . . .	27
4.2.2	2022-2025, Elena ALVAREZ-SANCHEZ, co-encadrement à 40% . . . . .	27
4.3	Participation à des jurys de thèse . . . . .	27
4.4	Comités de sélection et instances nationales . . . . .	28
4.5	Implication dans des programmes de recherche . . . . .	28
4.6	Financements en tant que partenaire, responsable ou porteur . . . . .	29
4.7	Collaborations nationales et internationales . . . . .	30
4.8	Vie universitaire . . . . .	31
<b>III</b>	<b>Projet de recherche</b>	<b>32</b>
<b>1</b>	<b>Prédiction de repliements protéiques non naturels</b>	<b>34</b>
<b>2</b>	<b>Intégration de données biologiques</b>	<b>37</b>





# Table des figures

3.1	Structure de l'adduit platiné G*G*G-Pt . . . . .	13
3.2	Page d'accueil du site de prédiction de la structure des RCPG . . . . .	15
3.3	Comparaison entre les prédictions réalisées à partir du modèle du récepteur ORIG1 et les résultats expérimentaux. . . . .	16
3.4	Définition de peptides inhibiteurs de la liaison RANK-RANKL se fixant sur RANK. . . . .	17
3.5	Page principale de l'interface de présentation des protéines présentes dans le globe rouge mature. . . . .	20
3.6	Organisation du site de liaison d'un résidu galactose 2NHAc lié à la galectine-1 (4XBL), à la galectine-3 (4XBN) et à la galectine-7 (4XBQ). . . . .	21
3.7	Comparaison des structures des complexes RhoA-ArhGEF-1 . . . . .	22
1.1	Analyse de la complexité structurale liée à la séquence du protéome humain. . . . .	35
1.2	Prédiction de structure d'une sélection de séquences non naturelles . . . . .	36





# Liste des tableaux

1	Encadrement d'étudiants depuis mon recrutement. . . . .	7
2	Financements ayant permis les gratifications et salaires des étudiants. . . .	7



# Première partie

## Préambule

*À mes parents, à ma famille*

On aime ce qui nous a émerveillé, et on protège ce que l'on aime.

---

*Jacques-Yves Cousteau*

La vie telle que nous la connaissons sur Terre a commencé à émerger il y a près de 3,8 milliards d'années [9]. À partir d'atomes tout d'abord assemblés en molécules élémentaires, puis en assemblages moléculaires complexes, les premiers éléments auto-réplicatifs ont commencé à émerger. En plus de la capacité à s'auto-répliquer, ces molécules primordiales ont aussi acquis la possibilité de conserver ces caractères d'une génération à l'autre, puis de transmettre une complexification toujours plus grande à leur descendance [31].

De nos jours, il est compliqué de valider quelles molécules sont apparues en premier, comment les premiers assemblages ont survécu pour donner des assemblages plus complexes, et comment cette complexité initiale a donné lieu aux organismes unicellulaires puis pluricellulaires. Les premières molécules présentant des activités auto-catalytiques répliquatrices sont pressenties comme étant basées sur des nucléotides, en particulier l'Acide Ribo-Nucléique (ARN), que l'on continue à étudier pour repousser encore notre connaissance sur les fonctions qu'il peut remplir [62].

L'hypothèse de la panspermie n'est pas non plus à exclure, avec la preuve encore récente que l'on retrouve des acides aminés ailleurs que sur terre, dans l'espace ou sur certaines météorites par exemple [8]. Lors d'explorations spatiales dans le futur, on trouvera peut-être aussi nombre de briques élémentaires du vivant ailleurs dans le système solaire : acides nucléiques, lipides, glucides.

Si l'origine de la vie et la détermination des étapes clés des éléments de sophistication du vivant sont difficiles, voire impossibles à dater précisément, nous avons cependant de très nombreuses traces biologiques, géologiques, et maintenant nucléiques de ce passé depuis au moins un milliard d'années [43].

Un travail considérable de classement et d'analyse a ainsi pu tracer les événements évolutifs qui ont mené au monde que nous connaissons. La recherche a permis de retracer le déroulé de l'apparition d'innovations évolutives, que l'on peut lister rapidement sans être exhaustif : apparition de la vie terrestre, apparition du vol, apparition des grands herbivores, apparition des modes de reproduction des animaux (ovipares, vivipares, ovovivipares), apparition de la respiration aérobie (et non plus anaérobie), apparition des mammifères, des primates et des humains. Chacune de ces grandes étapes a dû faire intervenir des changements continus ou brutaux dans les génomes des organismes. Nous conservons encore aujourd'hui dans nos génomes une trace des protéines impliquées dans les fonctions essentielles au fonctionnement de nos cellules [84].

Avec l'abandon progressif du religieux comme seul mode d'explication de notre origine, les naturalistes comme Linné [95] ou Buffon [93] ont effectué un travail colossal d'étude et de classement du vivant. Ces observations ont permis de comprendre qu'il existait des principes pour l'organisation de ces organismes, formulés par Darwin

[92] ou Lamarck [14] dans le domaine de l'évolution des espèces ou encore par Mendel pour l'explication de l'hérédité [85]. De la même manière, des progrès considérables ont été réalisés dans la compréhension de l'origine et de la transmission par des microbes, subdivisés encore ensuite en virus et bactéries, par des grands noms comme Louis Pasteur [60] ou Robert Koch [94].

Au XX<sup>ème</sup> siècle, la compréhension atomique de la matière par les méthodes de cristallographie par diffraction de rayons X a été utilisée pour comprendre l'organisation des macromolécules biologiques. Ce travail de longue haleine a vu chaque découverte fondamentale récompensée par un prix Nobel de Chimie, de Médecine ou de Physique [56]. Parmi quelques noms remarquables, on peut citer Summer, Northrop et Stanley pour leurs travaux sur les enzymes et les virus [1], Kendrew et Perutz pour leurs travaux sur la structure des protéines globulaires [4] ou encore Crick, Watson et Wilkins pour la découverte de la structure en double hélice anti-parallèle de l'ADN [3]. À partir des années 1970 une révolution a commencé, celle de l'utilisation du vivant pour des objectifs non dictés par les lois du vivant, mais pour des finalités biotechnologiques. C'est le début de l'utilisation des enzymes de restriction et des techniques de séquençage [2, 7]. Pour la première fois l'Homme a pris conscience des possibilités que ces technologies offraient, ce qui a amené aux lois de bioéthiques que l'on connaît maintenant et à une grande vigilance dans la manipulation du vivant [68]. Durant la même période, il est devenu aussi très clair que la structure des macromolécules individuelles permettait de comprendre la fonction de celles-ci, et qu'au-delà de la compréhension des mécanismes enzymatiques, un champ exploratoire immense s'ouvrait non plus pour la réutilisation du vivant, mais bien pour la conception de nouvelles molécules capables de cibler des constituants biologiques [42]. De nos jours, la recherche connaît une accélération telle qu'il devient impossible d'identifier une personne unique ou un événement isolé, des collections entières d'articles leurs sont maintenant consacrés [15].

Ces découvertes sont arrivées à un moment critique de notre développement technologique où l'informatique scientifique puis grand public a connu une accélération sans précédent. Ces développements informatiques ont permis à leur tour l'accélération des découvertes en biologie, et l'étude des macromolécules biologiques par des approches non plus expérimentales rassemblées dans des bases des données [78] mais *in silico* grâce à l'avènement de la modélisation moléculaire [55]. Nous sommes à nouveau à l'aube d'une nouvelle ère dans l'étude et la prédiction de la structure des macromolécules biologiques avec les développements récents de méthodes de prédiction de structure à très haut débit par des méthodes s'appuyant sur l'apprentissage profond (*deep learning*) s'appuyant sur les données expérimentales connues (AlphaFold2, RoseTTaFold, OmegaFold, etc) [79]. Les années à venir vont connaître des développements exceptionnels pour la prédiction de nouvelles protéines non naturelles à activité spécifique [59].

Ma recherche consiste à comprendre comment ces fonctions moléculaires jouent un rôle au niveau cellulaire, par l'intrication des protéines, des lipides, des glucides et des acides nucléiques, soit dans le but de déterminer leur mécanisme fondamental d'activation, soit dans le but de déterminer des inhibiteurs spécifiques de protéines impliquées dans des processus d'intérêt.

Je vais présenter dans ce mémoire quelques illustrations d'études de

mécanismes moléculaires par modélisation moléculaire. Pour chacun des sujets présentés, je vais introduire le contexte biologique, le questionnement lié à la stratégie de recherche, les méthodes et techniques mises en œuvre, ainsi que le rôle qu'ont joué les étudiants encadrés dans l'avancement du projet.

Les projets développés récemment ont été l'occasion de collaborations tripartites entre des biologistes experts des cibles et pathologies, des chimistes experts en synthèse avancée et de modélisateurs étudiant *in silico* la structure et la fonction des macromolécules. Ce partenariat exceptionnel a été soutenu par la Région Pays de La Loire via le projet PIRAMID (<http://pyramid-research.fr/>). Au-delà des avancées scientifiques importantes que le projet a connues, ces échanges ont posé les bases de collaborations solides entre chercheurs de la région, ont permis de développer nos expertises réciproques, ont été l'occasion d'encadrer ou de co-encadrer de nombreux étudiants de licence, de master et en thèse. Les travaux engagés durant ce projet auront une portée beaucoup plus grande que celle envisagée lors des réflexions engagées lors de la rédaction du projet initial.

Je conclurai ce mémoire par la présentation de quelques réalisations qui répondent aux missions de diffusion du savoir et de la connaissance auprès des pairs et du grand public. Hors de la culture scientifique inhérente à notre métier, il est en effet essentiel de permettre au plus grand nombre d'appréhender mieux le vivant, en fournissant un accès facilité à la connaissance via la conception et le déploiement de bases de données en ligne.

Deuxième partie

Curriculum vitae



J'ai choisi de faire une carrière scientifique pour analyser, étudier, modéliser puis moduler le vivant. À travers un parcours comprenant des recherches fondamentales ou finalisées, j'ai pu consolider ma compréhension des mécanismes moléculaires en faisant appel à la modélisation moléculaire et au criblage virtuel. Ces domaines sont maintenant connus sous les noms de bioinformatique structurale et de chémoinformatique mais ils recoupent des notions que l'on va retrouver en biochimie, en biologie, en chimie, en biophysique ou encore en statistique.

Avec l'expérience acquise dans ces domaines, je cherche maintenant à transmettre et partager aux étudiants comment la combinaison de ces approches multi-échelle permet de mieux décrypter la complexité biologique du vivant. Le Tableau 1 présente de manière synthétique le nombre d'étudiants encadrés, la présentation de leurs sujets est fournie plus loin dans le document.

Tableau 1 – Encadrement d'étudiants depuis mon recrutement.

Niveau d'études	Nombre d'étudiants	Durée
<b>Doctorat</b>	2 (co-encadrement)*	3 ans*
<b>Master 2</b>	12 (3 en co-encadrement)	5,5 mois
<b>Master 1</b>	22	2 mois
<b>Licence</b>	15	1 mois ou TER sur l'année

\* Première thèse soutenue 2020, seconde commencée en novembre 2022.

L'encadrement des étudiants de master 2 et des étudiants en thèse a été permis grâce au soutien du laboratoire permis par une politique de soutien des collaborations inter-équipes, par l'obtention de financements auprès de Nantes Université, de la région Pays de La Loire, d'associations comme La Ligue Contre Le Cancer, ou encore des soutiens par l'Agence Nationale pour la Recherche. Une liste détaillée de ces est indiquée dans le Tableau 2. Ces financements ont aussi servi à la gratification des stages de M1.

Tableau 2 – Financements ayant permis les gratifications et salaires des étudiants.

Niveau d'études	Nombre	Financement
Thèse	1	Projet régional PIRAMID
Thèse	1	ANRT et entreprise Affilogic
Master 2	4	Financement interne US2B
Master 2	8	Financements sur appels à projet
Master 1	4	Financements sur appels à projet

La suite du document présente le détail de mon activité d'enseignement et de recherche au format demandé par l'école doctorale Biologie Santé de Nantes.

# Chapitre 1

## Curriculum Vitae

### 1.1 État civil

Nom : Stéphane Téletchéa  
Date de naissance : 13 septembre 1975 à Saint-Pierre (975)  
Nationalité : Française  
Situation familiale : Marié, 3 enfants  
Adresse professionnelle : Unité des Sciences Biologiques et Biotechnologiques  
(US2B)  
Nantes Université, CNRS, UMR 6286  
2 chemin de la houssinière  
F-44000 Nantes  
☎ +33 251 125 636  
✉ stephane.teletchea@univ-nantes.fr

## 1.2 Formation

- 1993 : **Baccalauréat série D**, mention assez bien
- 1995 : **DUT de biologie appliquée**, Université de La Rochelle
- 1999 : **Maîtrise de biochimie**, Université de La Rochelle
- 2000 : **DEA d'Analyse de Génomes et de Modélisation Moléculaire**, Université de Paris VII Denis Diderot
- 2005 : **Doctorat des Aspects Moléculaires et Cellulaires de Biologie**. Université Paris V René Descartes, UMR CNRS 8601, Paris, sous la direction du Dr. J. Kozelka. *Étude par modélisation de dynamique moléculaire et spectroscopie RMN des déformations induites par la chélation du cisplatine sur l'ADN. Félicitations du jury.*

## 1.3 Expériences professionnelles

- **2005-2008, Post-doctorat** : Laboratoire Mathématique, Informatique et Génome, INRA, UR 1077, Jouy-en-Josas, laboratoire du Dr. J-F. Gibrat. *Prédiction et classement des affinités de liaison des ligands des récepteurs olfactifs.*
- **2008-2012, Post-doctorat** : Laboratoire de Physiopathologie de la Résorption Osseuse, INSERM U957 / EA 3822 Université de Nantes, Nantes, laboratoire du Pr. D. Heymann. *Modèles des partenaires protéiques intervenant dans l'ostéogenèse et définition d'inhibiteur peptidiques des interactions RANK-RANKL.*
- **2012-2013, Post-doctorat** : Laboratoire de Dynamique des Structures et Interactions des Macromolécules Biologiques, INSERM UMR\_S1134, Université de Paris 7, équipe du Pr. C. Etchebest. *Étude et modélisation moléculaire des protéines membranaires du globule rouge.*
- **2014-2022, Maître de Conférence classe normale** : Unité de Fonctionnalité et Ingénierie des Protéines (UFIP), Université de Nantes, CNRS, UMR6286.
- **2022-, Maître de Conférence hors classe** : Unité des Sciences Biologiques et Biotechnologies (US2B), Nantes Université, CNRS, UMR 6286/

# Chapitre 2

## Synthèse de l'activité d'enseignement

Je suis maître de conférences à l'université de Nantes depuis le 1<sup>er</sup> septembre 2014. Je me suis impliqué fortement dans la vie et les instances de l'université, en prenant en charge la formation des étudiants lors de mes enseignements mais aussi en encadrant des étudiants en stage au laboratoire. Mon implication principale en début de carrière a concerné le C2i, la Méthodologie du Travail Universitaire, l'Introduction à l'informatique pour la biologie, et le dispositif d'enseignant référent. J'ai ensuite stabilisé l'essentiel de mon enseignement dans plusieurs UE comme **responsable ou co-responsable : 1 UE de L2, 4 UE en M1 et 2 UE en M2.**

Depuis 2014, je suis responsable de l'UE de L2 « informatique pour les sciences de la vie » (180 étudiants en 2015, près de 250 actuellement). J'ai constitué une équipe pédagogique stable constituée de 4 permanents et de plusieurs contractuels (3-5 en fonction des années). Les cours sont remaniés régulièrement pour préciser les objectifs pédagogiques, l'espace numérique MADOC contient de nombreux documents pour favoriser la réussite des étudiants (annales, exercices d'auto-évaluation, contenu vidéo sur des points-clés, ...). Les documents de cours sont partagés avec l'association étudiante locale (SCINAPSE).

Depuis 2015 j'interviens plus spécifiquement dans l'enseignement de la bioinformatique, dès la licence (L3), puis en M1 et en M2. J'ai ainsi participé à la construction de la maquette du master de bioinformatique de Nantes en 2017 avec les responsables de formation Christine Sinoquet et Bernard Offmann. Je suis devenu **co-responsable avec le Pr. Bernard Offmann du master 2 pour le parcours d'ingénierie en bioinformatique réalisé en alternance.** Cette construction de maquette a compris des remaniements notables tant sur l'organisation des enseignements que sur les contenus (semaines regroupées pour l'initiation à la programmation, apparition des mégadonnées, ...). Dans ce master je suis aussi **co-responsable des UEs de bioinformatique structurale (3 UEs) avec Bernard OFFMANN et responsable de l'UE de M2 en biologie des systèmes.**

J'interviens dans d'autres masters dans des UEs à la frontière entre biologie et chimie ou biologie et informatique comme **co-responsable avec le Pr. Jean-Yves Le Questel** de l'UE de modélisation moléculaire (M1 Chimie-Biologie, Nantes), comme

responsable de l'UE « culture GNU/Linux » (M2 ACDI, Angers), ou ponctuellement en bioinformatique ailleurs (La Rochelle, Brest, Compiègne). Suite à l'obtention en 2021 de la PEDR, j'ai dû arrêter ces enseignements effectués dans des universités extérieures à Nantes.

Nous avons mis en place pour le parcours M2 « ingénierie bioinformatique » dont je suis co-responsable une pratique pédagogique innovante en lien avec le développement de l'alternance. Il s'agit en effet d'un **enseignement hybride, synchrone avec des étudiants du parcours en formation initiale** présents dans une salle tandis que **les étudiants en alternance suivent les cours à distance (en entreprise)**. Deux étudiants ont initialement choisi ce parcours en 2017-2018. Il y a cette année le maximum d'étudiants prévu dans la maquette (8). Le suivi des étudiants nous a offert un accès aux acteurs sociaux-économiques de toutes tailles, partout en France, ce qui facilite le recrutement de nos étudiants. J'assure une majeure partie du suivi régulier des étudiants en alternance, en effectuant en particulier le suivi longitudinal des étudiants avec leur encadrant en entreprise.

Pour accompagner ces transformations majeures de notre enseignement, j'ai suivi plusieurs formations organisées par le pôle de pédagogie de l'université, déposé et obtenu **4 appels à projets pour accompagner l'innovation pédagogique**. Cela a permis la production de contenus spécifiques et hors maquette pour nos étudiants, ainsi que l'achat d'un équipement informatique (serveur) pour la préparation de nos étudiants aux technologies émergentes en bioinformatique (cloud computing, vidéos d'introduction, etc.). Ces investissements profitent à tous les enseignants du master, participant ou non à l'appel à projet. J'ai donné plusieurs séminaires à l'université de Nantes pour partager l'expérience acquise lors de la mise en place de ces approches pédagogiques innovantes.

# Chapitre 3

## Synthèse de l'activité scientifique

Mon travail de thèse a tout d'abord porté sur l'analyse des adduits du cisplatine par des approches expérimentales (RMN) et de modélisation moléculaire. J'ai poursuivi ce travail par plusieurs post-doctorats portant sur (i) la modélisation moléculaire des récepteurs olfactifs, (ii) la recherche d'inhibiteurs innovants modulant la réparation osseuse, et (iii) sur l'étude structurale des protéines identifiées dans le globule rouge. Après mon recrutement en tant que maître de conférences à l'université de Nantes en 2014, j'ai continué l'analyse des macromolécules biologiques et la recherche d'inhibiteurs innovants en étudiant les galectines, la protéine RhoA et de la protéine Bcl-Xl. Je me suis ainsi spécialisé dans les interactions protéine-protéine et ces approches multi-échelles seront les éléments structurants de la suite de ma carrière. Ces travaux sont présentés brièvement ci-dessous par ordre chronologique

### 3.1 Modélisation moléculaire de macromolécules biologiques

J'ai étudié en DEA puis en thèse la structure de l'ADN et de l'ARN, les différents états de l'ADN (A, B, Z) et les possibilités d'appariements des acides nucléiques (Watson-Crick, Hoogsteen, Wooble). Lors de ma formation doctorale au sein de l'école doctorale du médicament, j'ai eu la possibilité de suivre un module de L3 en chimie inorganique pour mieux appréhender la modélisation moléculaire d'adduits du cisplatine. Pour confronter les résultats de modélisation moléculaire aux données expérimentales, j'ai été amené à synthétiser puis purifier des adduits du cisplatine, pour une analyse de ces adduits par RMN en collaboration avec une équipe norvégienne. Plus globalement, l'expérience acquise en DEA et en thèse m'a permis de mieux appréhender les mécanismes de reconnaissance entre une protéine et son ligand, en particulier quand le ligand est l'ADN.

La modélisation de la structure tridimensionnelle d'un ARN est un processus complexe. Au cours du stage de DEA au laboratoire de Biochimie Théorique (IBPC) sous la direction de Richard Lavery, le logiciel de modélisation des acides nucléiques JUMNA a été amélioré pour répondre à cette problématique. Il est désormais possible de construire un modèle tridimensionnel en épingle à cheveux de la structure d'un ARN à partir de sa

séquence primaire et de contraintes spécifiant quelles bases sont appariées et quelles bases sont empilées. Ce travail a été publié dans *Journal of Computational Chemistry* en 2003 [17] [REF 1].

Mon travail de thèse a consisté en l'étude par modélisation de dynamique moléculaire et par RMN des déformations que le cisplatine induit sur l'ADN, sa principale cible, et sur les interactions que les adduits du cisplatine sur l'ADN étaient susceptibles de former avec les partenaires protéiques disponibles dans la cellule. J'ai effectué ma thèse sous la direction de J. Kozelka au Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques (LCBT, UMR 8601, Dr. Mansuy, Paris V).

Le cisplatine (ou cis-diammine, dichloro-platine) est l'un des composés chimiques les plus utilisés en chimiothérapie anticancéreuse pour sa capacité à se lier à l'ADN et à perturber le cycle cellulaire [5, 6].

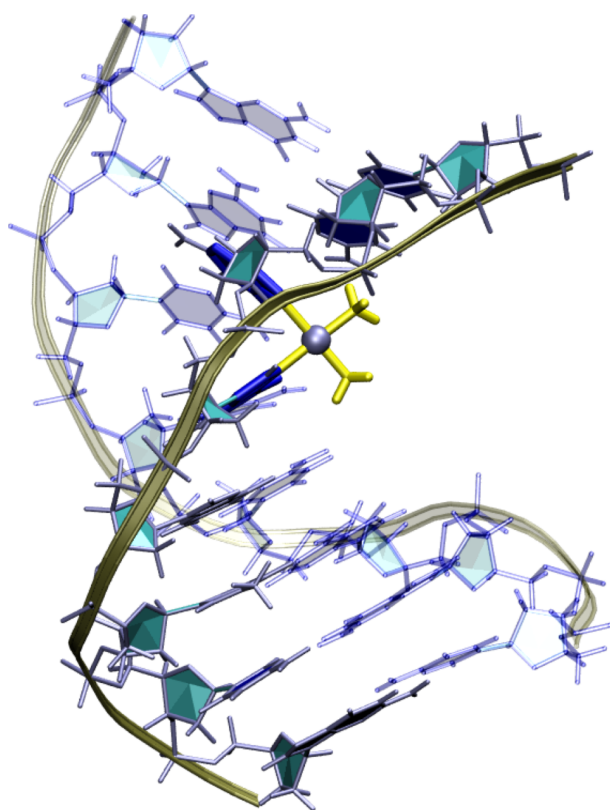


FIGURE 3.1 – Structure de l'adduit platiné G\*G\*G-Pt obtenue en couplant les analyses RMN et les calculs de dynamique moléculaire. G\* indique le site où le platine chélate la guanine dans le triplet de nucléotides. Le platine est indiqué par une sphère, les 4 coordinations sont représentées en bâtonnets jaunes.

Les simulations de dynamique moléculaire et les analyses de spectres RMN ont permis d'obtenir un accord important sur les conformations des sucres (populations relatives N et S) et les distances dérivées des signaux nOe : plus de 85 % des données expérimentales sont compatibles avec les calculs issus des simulations de dynamique moléculaire. Ce travail a été la première description d'un adduit sur un déca-nucléotide comportant trois guanines adjacentes (Figure 3.1). Ce travail a été publié dans *Chemistry – A European Journal* en 2009 [37]. [REF 6]

De la même manière, nous avons pu étudier les caractéristiques d'un autre adduit du platine, le pyrazolate bis-platine, et nous avons de nouveau pu obtenir un accord important entre la simulation de dynamique moléculaire et les données RMN. Ce travail a été publié dans *Chemistry – A European Journal* en 2006 [25]. [REF 4]

Nous avons enfin étudié le mode de reconnaissance possible de la déformation induite par le cisplatine sur l'ADN par le facteur de transcription LEF-I (Lymphoid Enhancer-binding Factor I) [11]. Cette reconnaissance ADN-protéine participe à la régulation de l'expression des récepteurs  $\alpha$  des lymphocytes T via la voie de signalisation Wnt [29]. La séquence de l'adduit platiné n'étant pas la séquence reconnue par le LEF-I, nous avons voulu vérifier si le cisplatine pouvait transformer cette liaison en une reconnaissance de structure puisque la déformation induite par le cisplatine a une structure très similaire à la déformation observée dans la structure du facteur LEF-I lié à l'ADN. La simulation de dynamique moléculaire du complexe ADN-protéine a mis en évidence que l'angle entre les bases reconnues par la protéine était compatible avec l'angle induit par le cisplatine sur les bases platinées, ce qui indique qu'une reconnaissance de l'ADN platiné par la LEF-I est plausible. Ce travail a été publié dans *JBSD* en 2005 [23]. [REF 2]

### 3.2 Modélisation moléculaire de protéines de la famille des RCPG et recherche d'inhibiteurs spécifiques par criblage virtuel

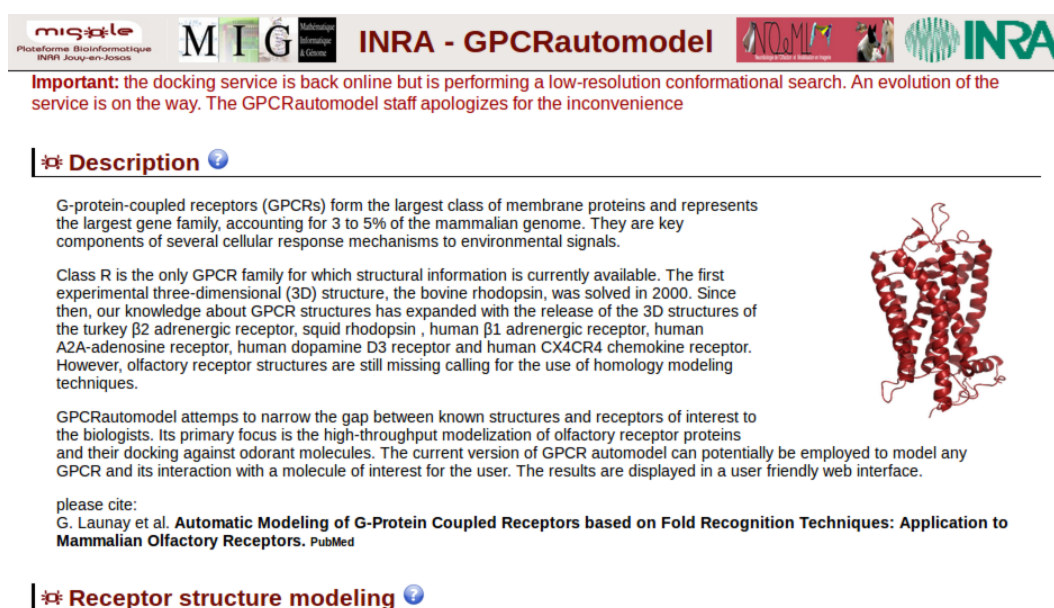
Lors de ce premier post-doctorat, je me suis familiarisé avec la structure des récepteurs couplés aux protéines G : leur agencement transmembranaire à 7 hélices et leur mode supposé d'activation qui conduit à la transduction du signal au niveau intracellulaire [27, 54, 65]. Ce travail était une collaboration entre deux équipes de l'INRA de Jouy-en-Josas, l'une étudiant les récepteurs olfactifs d'un point de vue expérimental, et l'autre par des approches de modélisation moléculaire. Pour faire le lien entre ces deux équipes, je me suis formé aux neurosciences, en suivant le cours dispensé par Roland Salesse sur le fonctionnement de la reconnaissance des odeurs, et j'ai réalisé une revue de la bibliographique sur les modes d'interaction des molécules odorantes sur les récepteurs olfactifs.

Aucune structure de récepteur olfactifs n'étant connue, j'ai développé un protocole de modélisation par comparaison en utilisant le logiciel MODELLER [61] à partir de la structure de la rhodopsine [13]. Afin de confronter les résultats expérimentaux de reconnaissance de ligand, une odothèque virtuelle a été créée puis criblée avec les logiciels AutoDock 4 [35] et Vina [38]. Pour étudier la flexibilité des récepteurs étudiés, j'ai utilisé la suite logicielle GROMACS [30], ce qui m'a permis de me former à la modélisation moléculaire de protéines membranaires. Cet environnement lipidique ajoute une complexité supplémentaire dans la représentation des macromolécules biologiques par un champ de force généraliste [19]. Afin d'étudier la phylogénie des différents récepteurs olfactifs, j'ai procédé à de nombreux alignements de séquence, à leur segmentation en hélices en fonction de leurs segments transmembranaires présumés, et utilisé ces informations pour enrichir la modélisation des récepteurs par comparaison. En parallèle, j'ai étudié les logiciels existants pour la prédiction de structure de protéines, en



sélectionnant les logiciels en fonction de leur performance dans le concours de Prédiction CASP (Critical Assessment of Structure Prediction, <http://predictioncenter.org/>).

Les récepteurs olfactifs appartiennent à la famille A des Récepteurs Couplés aux Protéines G (RCPG). Seul un petit nombre de récepteurs de cette famille a été caractérisé expérimentalement (rhodopsine, récepteurs adrénérgiques ou aux chimiokines). Pour modéliser la structure des récepteurs olfactifs, nous avons utilisé le logiciel de reconnaissance de repliements FROST afin de produire l'alignement optimal d'un récepteur olfactif donné par rapport à son support protéique [16, 32]. L'alignement structural ainsi produit a été utilisé pour fournir un modèle par comparaison qui a été utilisé pour effectuer un criblage virtuel sur les odorants les plus courants. L'énergie de liaison odorant-récepteur ainsi obtenue a été comparée avec les données d'activité disponibles pour un récepteur donné. Ce travail a été réalisé au sein de l'unité MIG de l'INRA de Jouy-en-Josas sous la supervision scientifique de Jean-François Gibrat.



**Important:** the docking service is back online but is performing a low-resolution conformational search. An evolution of the service is on the way. The GPCRautomodel staff apologizes for the inconvenience

### Description

G-protein-coupled receptors (GPCRs) form the largest class of membrane proteins and represents the largest gene family, accounting for 3 to 5% of the mammalian genome. They are key components of several cellular response mechanisms to environmental signals.

Class R is the only GPCR family for which structural information is currently available. The first experimental three-dimensional (3D) structure, the bovine rhodopsin, was solved in 2000. Since then, our knowledge about GPCR structures has expanded with the release of the 3D structures of the turkey  $\beta 2$  adrenergic receptor, squid rhodopsin, human  $\beta 1$  adrenergic receptor, human A2A-adenosine receptor, human dopamine D3 receptor and human CX4CR4 chemokine receptor. However, olfactory receptor structures are still missing calling for the use of homology modeling techniques.

GPCRautomodel attempts to narrow the gap between known structures and receptors of interest to the biologists. Its primary focus is the high-throughput modelization of olfactory receptor proteins and their docking against odorant molecules. The current version of GPCR automodel can potentially be employed to model any GPCR and its interaction with a molecule of interest for the user. The results are displayed in a user friendly web interface.

please cite:  
G. Launay et al. **Automatic Modeling of G-Protein Coupled Receptors based on Fold Recognition Techniques: Application to Mammalian Olfactory Receptors.** PubMed

### Receptor structure modeling

FIGURE 3.2 – Page d'accueil du site de prédiction de la structure des RCPG de la sous-famille des récepteurs olfactifs.

Pour résoudre la problématique posée par la modélisation à grande échelle de tous les récepteurs olfactifs un protocole de modélisation automatique a été développé. La prédiction est réalisée à travers un serveur en ligne, le modèle du récepteur ainsi que le résultat du criblage virtuel sont expédiés par courriel (<http://genome.jouy.inra.fr/GPCRautomodel>). L'utilisateur peut choisir la protéine support qui va être utilisée pour modéliser un récepteur olfactif donné, et décider s'il veut ou non procéder à un criblage virtuel du modèle obtenu avec la bibliothèque de ligands olfactifs. La validation expérimentale sur le récepteur OR1G1 des conséquences de mutations sur la reconnaissance de ligands a permis d'obtenir un accord qualitatif entre la prédiction fournie et le modèle présenté (Figure 3.3).

Ce travail a été publié dans PEDS en 2012. [50] [REF 9]

	Récepteur natif (% de cellules activées)	F104A mod	F104A exp	A112V mod	A112V exp	H176F mod	H176F exp	F200A mod	F200A exp
decanal	8,2 +/- 0,7	↗	↗					↗	↗
octanal	12,7 +/- 1,2			↗	↗				
nonanal	18,7 +/- 2,0	↗		↘		↗	↗		
nonanol	19,2 +/- 2,7					↘	↘		

FIGURE 3.3 – Comparaison entre les prédictions réalisées à partir du modèle du récepteur OR1G1 et les résultats expérimentaux.

F200A : mutation de la phénylalanine 200 en alanine. mod : prédiction issue de la modélisation, exp : résultat expérimental.

### 3.3 Modélisation moléculaire de protéines, conception et validation d'inhibiteurs peptidiques

Le Laboratoire de Physiopathologie de la Résorption Osseuse (LPRO) du Pr. D. Heymann étudie les mécanismes impliqués dans le dérèglement de la biologie de l'os et les pathologies associées. Deux types cellulaires principaux sont impliqués dans l'ostéogenèse : les ostéoclastes lysent la matrice osseuse (résorption) tandis que les ostéoblastes néo-forment l'os (apposition). Dans un contexte pathologique (cancer, ostéoporose), il existe un déséquilibre entre les activités de ces cellules [26]. Dans l'os sain, l'équilibre entre ces deux mécanismes est assuré par des communications inter-cellulaires à l'aide de deux cytokines sécrétées par les ostéoblastes : le ligand du récepteur activateur de la voie NF-κB (RANKL) et son récepteur leurre, l'ostéoprotégérine (OPG). L'activation des ostéoclastes est initiée par la liaison RANKL sur son récepteur RANK (récepteur activateur de la voie NF-κB) situé à la surface des ostéoclastes. L'état d'activation des ostéoclastes (quiescents ou en train de résorber la matrice osseuse) dépend de la biodisponibilité locale de RANKL, liée à la concentration locale de son récepteur leurre, l'OPG.

Durant cette période, j'ai pu découvrir la biologie du tissu osseux et participer à la caractérisation des événements moléculaires qui permettent à un os de grandir et de se réparer, dans un contexte physiologique ou pathologique. J'ai pu développer mes connaissances en immunologie et en cancérologie (cancers primaires de l'os), en particulier sur le rôle des cytokines dans la différenciation des monocytes en ostéoclastes dans un contexte inflammatoire ou cancéreux. [20, 22]

Le laboratoire a mis en évidence que la biodisponibilité de l'OPG est elle-même liée à la présence de protéoglycanes à la surface des cellules mais aussi à la présence et à l'état d'activation du facteur de von Willebrand dans la circulation générale. Nous avons pu caractériser par modélisation moléculaire la zone d'interaction entre le facteur de von Willebrand et l'OPG, ce travail a été publié dans Journal of Biological Chemistry en 2009. [33] [REF 8]

Pour moduler spécifiquement la liaison RANKL-RANK nous avons défini par modélisation moléculaire les zones d'interaction spécifiques de ces trimères de trimères. Ces modèles ont été confirmés en les comparant avec les structures tri-dimensionnelles

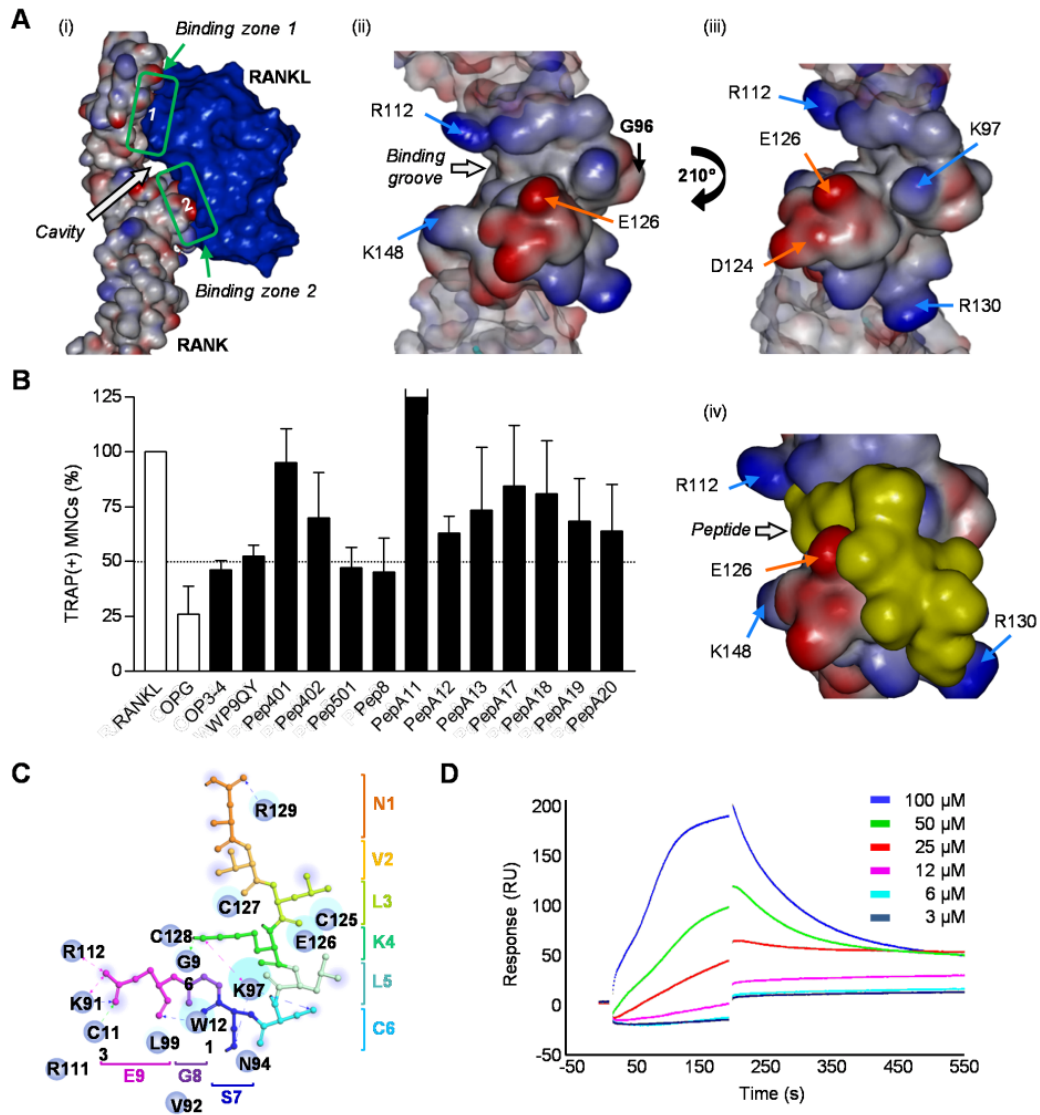


FIGURE 3.4 – Définition de peptides inhibiteurs de la liaison RANK-RANKL se fixant sur RANK.

**A** : la liaison entre RANKL et RANK s'effectue à deux endroits, ce qui permet de délimiter une cavité (i). A côté de cette cavité se trouve l'acide glutamique E126 indispensable à la liaison de la cytokine sur son récepteur, et la lysine K97 (ii et iii). Ces deux acides aminés servent à définir une charnière pour le changement d'orientation du peptide (iv), ce qui permet au peptide candidat de recouvrir au mieux le récepteur RANK, en possédant une zone spécifique de RANK non compétitive de RANKL (orientée vers R112) et une zone compétitive de la liaison RANKL-RANK (en direction de R130). **B** : validation d'une série de peptides publiés dans la littérature (OP3-4, WP9QY) et de peptides conçus durant ce travail (PepXXX) présentant une activité d'inhibition de l'ostéoclastogenèse comparable à celle du récepteur leurre de RANKL, OPG (concentrations des peptides : 50μM, OPG : 2 μM). **C** : interaction du meilleur peptide validé, pep8, avec les acides aminés de RANK. **D** : caractérisation de la liaison de pep8 à RANK par résonance plasmonique de surface (kD = 10 μM)

de RANK, RANKL et plus récemment de l'OPG. Nous avons développé des inhibiteurs se liant spécifiquement à RANK par criblage virtuel validé par des calculs d'énergie

d'interaction et une inspection visuelle (Figure 3.4). Une bibliothèque de peptides a été spécialement conçue pour ce projet, avec un tri spécifique pour que les peptides évalués présentent des conditions de solubilité optimales dans l'eau. Des candidats inhibiteurs ont été sélectionnés à partir de l'analyse *in silico*, et testés *in vitro* puis *in vivo*. Ce travail a été breveté [53] et publié [57] [REF 11].

La solubilité des peptides conçus *in silico* a parfois limité ou empêché la validation de l'activité de certains de ces peptides *in vitro* sur les ostéoclastes. Dans les cas où le peptide n'était pas soluble dans une solution compatible avec la culture cellulaire (PBS, milieu de culture, eau), il est possible de le reprendre dans une proportion variable de DMSO / eau (5-100 % de DMSO), ce qui permet de mesurer l'interaction sur le récepteur par des méthodes biophysiques (RMN, SPR) mais empêche l'étude de l'activité sur des ostéoclastes particulièrement sensibles à l'utilisation du DMSO, même en très faible quantité [46]. Comme aucune méthode publiée ne permettait de définir a priori la solubilité d'un peptide à partir de sa séquence, j'ai implémenté un protocole basé sur un arbre de décision pour écarter les peptides pouvant présenter potentiellement des problèmes de solubilité. Ces prédictions ont été validées sur notre collection de peptides dont la solubilité a été évaluée expérimentalement dans l'eau. Ce travail est en préparation, le serveur web permettant de tester la solubilité des peptides est néanmoins déjà disponible à l'adresse :

<http://mobylye.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::SolyPep>.

### 3.4 Étude et modélisation des protéines membranaires du globule rouge, définition des partenaires protéiques et conception d'inhibiteurs spécifiques

A partir de janvier 2013, j'ai effectué un post-doctorat au sein du laboratoire de Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), INSERM UMR\_S 1134, Université de Paris 7, dans l'équipe du Pr. C. Etchebest. Ce travail a pour objectif de rassembler en un endroit unique les données biologiques et structurales concernant initialement les protéines de la membrane du globule rouge. L'ensemble de ces informations va être utilisé pour prédire la structure des protéines insérées dans la membrane, pour valider les modèles produits grâce aux méthodes les plus performantes développées par d'autres équipes, et pour indiquer la position et le rôle de mutations ponctuelles identifiées ou non dans des pathologies associées au globule rouge.

Le globule rouge différencié ou érythrocyte est une cellule énuclée, dépourvue d'organelles, dans laquelle la synthèse protéique est absente [63]. Du fait de cette singularité cette cellule est un outil idéal pour l'étude des interactions métaboliques minimales nécessaire au maintien de l'intégrité de la cellule dans le corps humain [28, 39, 58]. De très nombreuses protéines membranaires jouent un rôle primordial dans la transfusion sanguines car elles peuvent être la cause de réaction de rejet immédiat ou différé de la transfusion du sang d'un patient à un autre. Compte-tenu de ces enjeux stratégiques, un effort important est fait pour caractériser la nature et la structure des protéines de l'érythrocyte. Il n'existe pas à ce jour d'endroit permettant de rassembler une information structurale sur les protéines d'importance, et compte-tenu de leur localisation membranaire, peu de protéines peuvent être modélisées facilement à partir des structures résolues expérimentalement [21]. Durant ce projet, j'ai mis en place une banque de données rassemblant les informations de différentes sources généralistes (Uniprot [91], OMIM [88], Gene Ontology [12], ENSEMBL [80], Entrez [71], BioGRID [77], PFAM [76], InterPro [89],

et diverses sources spécialisées comme dbRBC [47]. Ces informations accessibles librement (ou avec permission) sont enrichies de données structurales mises en place au sein du laboratoire à travers des outils de modélisation par comparaison (MODELLER, [61]), par enfilage (*threading* [45]), par modélisation *ab initio* [41], ou en combinant plusieurs approches [24, 49, 51].

La mise en place d'une base de données qui rassemble des informations de nature très diverse a nécessité l'implémentation à partir de zéro des protocoles de prédiction, de calcul, d'assemblage, et de mise à jour du recueil des informations. Afin de maîtriser l'évolution de la base de données au fur et à mesure de son développement, je me suis axé sur l'utilisation du framework web Django, qui, grâce à son ORM, m'a permis de générer les tables associées à l'organisation des données et de faire évoluer plusieurs fois le modèle de données sous-jacent. J'ai fait un usage intensif du langage python, sur lequel Django est aussi basé, pour récupérer les informations, les analyser, et les rassembler dans la base de données sous une forme synthétique par l'utilisateur final. Pour les analyses de séquence (conservation de motifs, détection d'acides aminés qui ont co-évolué, alignements multiples, ...), j'ai utilisé le langage perl et le module bioperl ou python et biopython quand cela s'avérait le plus utile [36]. La base de données est en ligne depuis plus de 5 ans avec près de 100 visiteurs uniques par mois (<http://www.dsimb.inserm.fr/respire>). Les évolutions de la base vont consister en l'incorporation de données de protéomique grâce à une collaboration avec Patrick Mayeux, DR CNRS, responsable de la plateforme de protéomique <https://3p5.recherche.parisdescartes.fr/Proteom'IC>, de la littérature scientifique, des pathologies spécifiques du globule rouge.

Nous avons mis en place un protocole de prédiction des structures tertiaires des protéines à haut débit pour produire des modèles spécifiques des protéines membranaires. Ces modèles peuvent être enrichis par l'ajout de contraintes structurales ou fonctionnelles basées sur les données de la littérature et sur les données expérimentales mises à la disposition par les partenaires du Laboratoire d'Excellence « Labex » GR-Ex (<http://www.labex-grex.com/>). L'objectif de ce portail web est de devenir le site de référence de portée mondiale pour les informations structurales concernant l'érythrocyte.

La base de données est disponible à l'adresse : <http://www.dsimb.inserm.fr/respire>, ce travail a été publié dans Plos One [67] [REF 16], une illustration de la page principale est présentée dans la Figure 3.5.

L'une des problématiques liées à la production de modèles de protéines est la détermination d'un indice de confiance dans le modèle produit. Ce problème est tout particulièrement critique dans le cadre des protéines membranaires où le nombre de structures disponibles est notoirement insuffisant pour servir de support pour la production de protéines homologues. Comme notre portail se propose de fournir un modèle de qualité pour les protéines présentées à l'utilisateur, nous avons fait appel à une fonction de score basée sur la méthodologie de la protéine, préalablement étalonnée au laboratoire. Cette méthodologie a été mise à la disposition de la communauté des modélisateurs via un site internet dédié ([http://www.dsimb.inserm.fr/dsimb\\_tools/hpmscore/](http://www.dsimb.inserm.fr/dsimb_tools/hpmscore/)). Ce travail a été publié récemment [90] [REF 28].

## 3.5 Etude des galectines et caractérisation d'inhibiteurs innovants

Les lectines sont des protéines reconnaissant un motif glycoconjugué composé de mono- ou poly- saccharides linéaires ou branchés. Cette superfamille de protéines est essentielle pour un grand nombre de processus cellulaires comme la migration cellulaire, l'inflammation, la réponse

## Welcome to RESPIRE

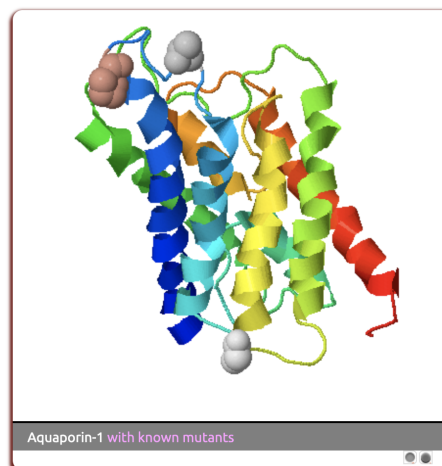
Welcome to the **Repository of Enhanced Structures of Proteins Involved in the Red blood cell Environment**. This database aims at providing a reference, ordered and protein-driven information on proteins available in the erythrocyte.

You will find on this database annotations for proteins detected in the latest differentiation stages of erythropoiesis. To help in finding useful content, reference external databases are queried regularly and then RESPIRE methods allow the integration of both this data and of additional processing into dedicated pages, one per protein ([See Documentation for details](#)).

With a combination of structural information, clinical information and links to reference external databases, it simplifies the access to RBC components for biologists, clinicians and structural biologists.

The protein structures presented in the carousel display models produced by the DSIMB laboratory for some protein of interest.

Start browsing the database by clicking an image in the carousel, by using pre-set searches in the **Search** dropdown menu, or by searching by protein name, gene name or uniprot identifier in the top search box. Many links are also provided in protein entries to propose alternative browsing facilities.



**Citation of the database:**

*Repository of Enriched Structures of Proteins Involved in the Red Blood Cell Environment (RESPIRE).*  
Téletchéa S, Santuz H, Léonard S, Etchebest C. (2019). *PLOS ONE* 14(2): e0211043. [CR](#)

FIGURE 3.5 – Page principale de l’interface de présentation des protéines présentes dans le globule rouge mature.

Le site permet de retrouver les protéines identifiées expérimentalement pour le stade de différenciation terminal des érythrocytes. Il permet le recensement rapide d’informations structurales associées à une protéine donnée et la visualisation interactive de la structure de ces protéines, en précisant si nécessaire la localisation des mutations connues.

immunitaire, l’autophagie et les interactions hôte pathogène [18, 48, 69]. Parmi cette superfamille 15 protéines différentes liant spécifiquement le galactose, sont connues chez l’homme. Ces 15 galectines sont prépondérantes dans les processus inflammatoires ou encore dans de nombreuses pathologies comme le cancer [66]. La reconnaissance spécifique du galactose est réalisée par une architecture très conservée en deux feuillets beta en sandwich dont la structure est rigidifiée par un réseau de phénylalanines interconnectées. Le site de reconnaissance du galactose comprend plusieurs résidus très conservés (Arg144, Asp148, His158, Glu165, Arg162, Asn174, Lys176, Glu184, Arg186) à la fois au niveau de leur séquence, mais aussi dans leur organisation structurale [34]. Ces résidus hydrophiles impliqués dans un réseau d’interactions hydrogènes avec les sucres sont complétés par une interaction hydrophobe très caractéristique apportée par le tryptophane 181. C’est cette balance entre la rigidité de la structure, le réseau de la liaison hydrogènes et la contribution hydrophobe apportée par le tryptophane qui permet une spécificité et sélectivité très importante des galectines pour les sucres contenant un galactose comme les lactosamines [82]. A ce jour il n’existe qu’un faible nombre d’inhibiteurs spécifiques des galectines-3 en phases cliniques comme le TD139, le GR-MD-02, le GB0139 ou le GB1211 [64, 87], il est donc indispensable de continuer à concevoir des inhibiteurs spécifiques de ces interactions.

Notre laboratoire possède une expertise spécifique dans la synthèse de lactosamine de type II ou la liaison entre le N-acétyl glucose et le galactose est réalisée entre les positions 1 et 3 par voie enzymatique alors que la liaison entre les lactosamine de type I se fait en position 1-4 (Gal  $\beta$  1,4 GlcNAc) [83].

Par criblage virtuel du mode de liaison entre des dérivés de lactosamine de type I et de type II

avec la galectine-3 chez l'humain, le rat et la souris, j'ai pu caractériser les éléments contribuant le plus significativement à la liaison avec des inhibiteurs synthétisés par l'équipe de glycobiochimie de notre laboratoire dirigée par Cyrille Grandjean. Deux articles ont ainsi été publiés sur le sujet, la Figure 3.6 présente le résultat du protocole de criblage virtuel que j'ai développé initialement puis consolidé lors du stage de 6 mois de master 2 par Charlotte Brunel-Cadic.

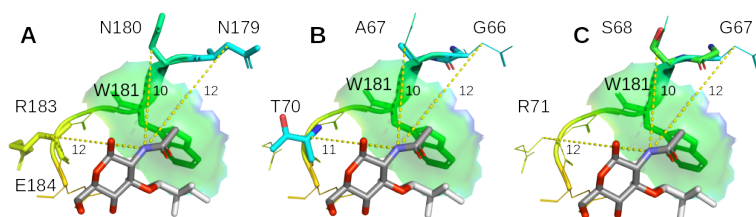


FIGURE 3.6 – Organisation du site de liaison d'un résidu galactose 2NHAc lié à la galectin-1 (4XBL), à la galectin-3 (4XBN) et à la galectin-7 (4XBQ).

Le tryptophane conservé dans les structures est indiqué en représentation de surface, seules les boucles proches du site de liaison sont affichées, les atomes du ligand sont colorés en gris pour les carbones, en rouge pour les oxygènes et en bleu pour les azotes. **A** Acides aminés impliqués dans la reconnaissance du galactose pour la galectine 3 avec les distances entre le ligand et la galectine en angstrom. **B** Les acides aminés retrouvés identiques à ces positions dans la galectine 1 ou **C** la galectine 7 sont indiqués en représentation batons, les acides aminés différents sont indiqués en représentation ligne.

### 3.6 Détermination de l'interface protéine-protéine entre RhoA et ArhGEF1

L'hypertension artérielle est caractérisée par une pression artérielle élevée en permanence dans les artères. Selon l'Organisation Mondiale de la Santé (OMS), l'hypertension artérielle est définie comme une pression artérielle systolique supérieure à 140 mmHg et une pression diastolique supérieure à 90 mmHg. Elle est considérée comme le principal facteur de risque de maladies cardiaques et d'accidents vasculaires cérébraux. Les causes de l'hypertension artérielle sont multiples, mais peuvent inclure des facteurs de style de vie tels que l'alimentation, l'exercice et le tabagisme, ainsi que des facteurs génétiques et métaboliques. Le traitement de l'hypertension artérielle peut inclure des changements de style de vie, des médicaments et, dans certains cas, des interventions chirurgicales. L'hypertension artérielle peut aussi apparaître avec l'âge, on parle alors d'hypertension essentielle.

L'équipe de Gervaise Loirand a démontré que deux protéines, RhoA et ArhGEF1, jouaient un rôle primordial dans la régulation de la tension artérielle [44]. RhoA est un membre de la super-famille des petites GTPases RAS [52]. RhoA est un régulateur important du remaniement du cytosquelette d'actine, participant ainsi à de multiples processus cellulaires tels que la contraction cellulaire, la migration, la prolifération et la transcription génique. RhoA est un interrupteur moléculaire qui relie les récepteurs de surface cellulaire aux voies effectrices intracellulaires en circulant entre un état inactif cytosolique lié à la guanosine 5'-diphosphate (GDP) et un état actif lié à la guanosine 5'-triphosphate (GTP) qui se transloque vers la membrane. L'activation de RhoA est médiée par les facteurs d'échange de nucléotide guanine Rho (*Guanine Exchange Factor*, GEF) qui favorisent l'échange du GDP pour du GTP, qui sont eux-mêmes activés par la liaison d'un effecteur sur son récepteur membranaire. ArhGEF1 est le Rho GEF responsable de l'activation de RhoA par l'angiotensine II à travers le récepteur de l'angiotensine II de type 1 dans les cellules musculaires lisses vasculaires [40]. Cette voie de

signalisation participe au contrôle physiologique de la tonicité vasculaire et de la pression artérielle mais joue aussi un rôle central dans la physiopathologie de l'hypertension.

J'ai étudié le complexe RhoA-ArhGEF1 dans le cadre du projet PIRAMID en supervisant Ennys Gheyouche. Durant ce travail, nous avons cherché à caractériser comment il était possible de construire un complexe macromoléculaire des deux protéines liées en s'appuyant sur les structures de RhoA en interaction avec d'autres GEF. Les deux approches de construction initiale du complexe, (i) soit par superposition, (ii) soit docking protéine-protéine ont permis d'identifier des conformations proches. Notre étude a montré que des simulations de dynamique moléculaire des complexes modélisés par chacune des approches ont permis d'améliorer grandement le positionnement relatif des deux protéines (Figure 3.7).

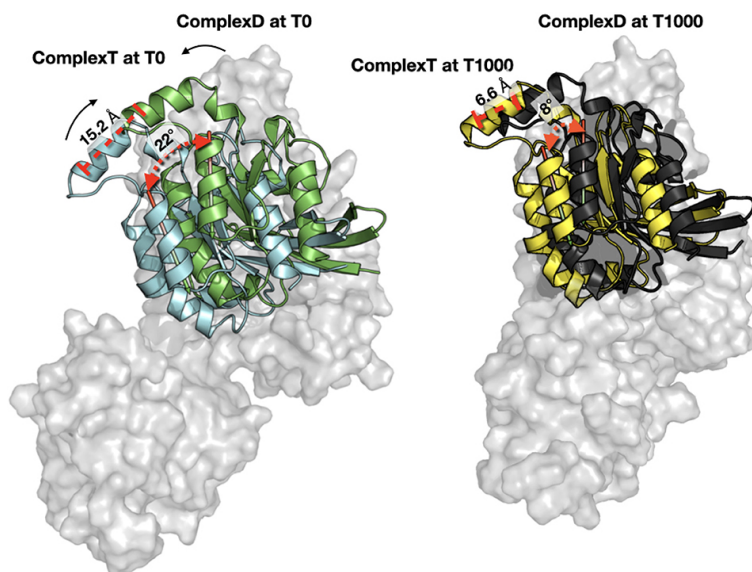


FIGURE 3.7 – Comparaison des structures des complexes RhoA-ArhGEF-1 obtenus par superposition (bleu, jaune) ou par amarrage moléculaire (*docking*) (vert, noir) au début des simulation (à gauche) et après dynamique moléculaire (droite). Le facteur d'échange ArhGEF1 est représenté en surface transparente grise.

Nous nous sommes appuyés sur l'analyse des séquences de chacune des protéines pour établir les profils de conservation des acides aminés de l'interface en déterminant quels acides aminés étaient spécifiques uniquement de RhoA et ArhGEF1 et quels acides aminés étaient retrouvés dans tous les complexes RhoX-ArhGEFX.

En couplant les analyses de séquence et de structure, nous avons ainsi fourni une méthodologie de modélisation des interactions protéine-protéine. Cette étude a été publiée dans la revue *Frontiers in Molecular Biosciences*. [73]. [REF 24]



# Chapitre 4

## Production scientifique

J'ai participé à 28 articles publiés, 1 article en pré-publication et 1 brevet. Je suis premier auteur de 6 articles et l'auteur de correspondance pour 2 articles. Plusieurs articles sont issus de collaborations auprès de collègues du laboratoire, de collègues de différents laboratoires nantais, ou d'autres laboratoires en France ou aux Etats-Unis. Ces dernières années, j'ai accordé une attention plus importante à la publication d'articles originaux en tant qu'investigateur principal sur des projets de recherche propres. Dans ces derniers articles, j'ai conçu l'étude, encadré un ou plusieurs étudiants de L3, M1, M2 ou en thèse, rédigé, puis soumis et apporté si nécessaire les corrections du manuscrit décrivant le travail. La liste des travaux publiés est présentée ci-dessous.

28. **Téletchéa S.**, Esque J., Urbain A., Etchebest C., de Brevern A.G. Evaluation of Transmembrane Protein Structural Models Using HPMScore. *BioMedInformatics* 2023, 3, 306-326. [lien](#)
27. Venturelli P.A., **Teletchea S.**, Bales A.M., Bartolozzi P., Bird A.T., Blevins T.K., Campaniello S.J., Caizergue M., Carlu L., Chancerelle G., Colletta B., Dauphin L., Doche B., Derolf P.M., De Wever T., Dewig E.M., Dixon L.M., Durand C., Eck M., Faatauiria T., Fisher S.M., Fix G., Fournier S., Gauthy A., Golitin C., Guyader S., Hachet F., Harnay P., Hawkins S.G., Kauffling A., Khan M., Kesterson W.J., Klein M.K., Lejeune C., Loiseau J., Loyau R., Luginbuhl S.B., Maeso J., Marc T., Martineau L., Meurillon T., Mesnieres E., Mohra R., Mccord A.O., Mcdonald Z.N., Mckay A.B., Miller T., Minhinnett S.R., Poujoulat R., Profit V., Psurny G.Q., Raymond G., Redinger R.R., Rech G., Rider A.L., Rodriguez L., Sanders S.S., Salou G., Saucier T.E., Schwer J.D., Seymour R.D., Sez nec C., Shook B.L., Soler J., Tettling L., Thornburg G.E., Tottoli T., Veber E., Verdier L., Verin R., Vigot M., Vigouroux E., Voss K.N., Weir J.L., Teletchea F. A student-based expansion of the strategies of reproduction in fish (STOREFISH) database to 288 North American freshwater and anadromous species for 14 egg and larval traits. *Cybium*, 2023, p. 1-10. [lien](#)
26. Poulet A, Rousselot E, **Téletchéa S**, Noirot C, Jacob Y, van Wolfswinkel J, Thiriet C, Duc C. The Histone Chaperone Network Is Highly Conserved in *Physarum polycephalum*. *Int J Mol Sci.* 2023 Jan 5 ;24(2) :1051. [lien](#)
25. Poulet A, Mishra LN, **Téletchéa S**, Hayes JJ, Jacob Y, Thiriet C, Duc C. Identification and characterization of histones in *Physarum polycephalum* evidence a phylogenetic vicinity of Mycetozoans to the animal kingdom. *NAR Genom Bioinform.* 2021 ; 3(4) :lqab107. [lien](#)

24. Gheyouché E, Bagueneau M, Loirand G, Offmann B, **Téletchéa S**. Structural Design and Analysis of the RHOA-ARHGEF1 Binding Mode : Challenges and Applications for Protein-Protein Interface Prediction. *Front Mol Biosci*. 2021 ; 8 :643728. [lien](#)
23. Floch A, **Téletchéa S**, Tournamille C, de Brevern AG, Pirenne F. A Review of the Literature Organized Into a New Database : RHeference. *Transfus Med Rev*. 2021 ; 35(2) :70-77. [lien](#)
22. Muñoz-García J, Cochonneau D, **Téletchéa S**, Moranton E, Lanoe D, Brion R, Lézot F, Heymann MF, Heymann D. The twin cytokines interleukin-34 and CSF-1 : masterful conductors of macrophage homeostasis. *Theranostics*. 2021 ; 11(4) :1568-1593. [lien](#)
21. **Téletchéa S**, Teletchea F. STOREFISH 2.0 : a database on the reproductive strategies of teleost fishes. *Database (Oxford)*. 2020 ; 2020 :baaa095. [lien](#)
20. Dussouy C, **Téletchéa S**, Lambert A, Charlier C, Botez I, De Ceuninck F, Grandjean C. Access to Galectin-3 Inhibitors from Chemoenzymatic Synthons. *J Org Chem*. 2020 ; 85(24) :16099-16114. [lien](#)
19. Le Caignec C, Ory B, Lamoureux F, O'Donohue MF, Orgebin E, Lindenbaum P, **Téletchéa S**, Saby M, Hurst A, Nelson K, Gilbert SR, Wilnai Y, Zeitlin L, Segev E, Tesfaye R, Nizon M, Cogne B, Bezieau S, Geoffroy L, Hamel A, Mayrargue E, de Courtivron B, Decock-Giraudaud A, Charrier C, Pichon O, Retière C, Redon R, Pepler A, McWalter K, Da Costa L, Toutain A, Gleizes PE, Baud'huin M, Isidor B. RPL13 Variants Cause Spondyloepimetaphyseal Dysplasia with Severe Short Stature. *Am J Hum Genet*. 2019 ; 105(5) :1040-1047. [lien](#)
18. Gheyouché E, Launay R, Lethiec J, Labeeuw A, Roze C, Amossé A, **Téletchéa S**. DockNmine, a Web Portal to Assemble and Analyse Virtual and Experimental Interaction Data. *Int J Mol Sci*. 2019 ; 20(20) :5062. [lien](#)
17. Tripathi N, Vetrivel I, **Téletchéa S**, Jean M, Legembre P, Laurent AD. Investigation of Phospholipase C $\gamma$ 1 Interaction with SLP76 Using Molecular Modeling Methods for Identifying Novel Inhibitors. *Int J Mol Sci*. 2019 ; 20(19) :4721. [lien](#)
16. **Téletchéa S**, Santuz H, Léonard S, Etchebest C. Repository of Enriched Structures of Proteins Involved in the Red Blood Cell Environment (RESPIRE). *PLoS One*. 2019 ; 14(2) :e0211043. [lien](#)
15. Dion J, Advedissian T, Storozhylova N, Dahbi S, Lambert A, Deshayes F, Viguier M, Tellier C, Poirier F, **Téletchéa S**, Dussouy C, Tateno H, Hirabayashi J, Grandjean C. Development of a Sensitive Microarray Platform for the Ranking of Galectin Inhibitors : Identification of a Selective Galectin-3 Inhibitor. *Chembiochem*. 2017 ; 18(24) :2428-2440. [lien](#)
14. Atmanene C, Ronin C, **Téletchéa S**, Gautier FM, Djedaïni-Pilard F, Ciesielski F, Vivat V, Grandjean C. Biophysical and structural characterization of mono/di- arylated lactosamine derivatives interaction with human galectin-3. *Biochem Biophys Res Commun*. 2017 ; 489(3) :281-286. [lien](#)
13. Ségaliny AI, Brion R, Brulin B, Maillason M, Charrier C, **Téletchéa S**, Heymann D. IL-34 and M-CSF form a novel heteromeric cytokine and regulate the M-CSF receptor activation and localization. *Cytokine*. 2015 ; 76(2) :170-181. [lien](#)
12. Craveur P, Joseph AP, Esque J, Narwani TJ, Noël F, Shinada N, Goguet M, Leonard S, Poulain P, Bertrand O, Faure G, Rebehmed J, Ghozlane A, Swapna LS, Bhaskara RM, Barnoud J, **Téletchéa S**, Jallu V, Cerny J, Schneider B, Etchebest C, Srinivasan N, Gelly JC, de Brevern AG. Protein flexibility in the light of structural alphabets. *Front Mol Biosci*. 2015 ; 2 :20. [lien](#)
11. **Téletchéa S**, Stresing V, Hervouet S, Baud'huin M, Heymann MF, Bertho G, Charrier C, Ando K, Heymann D. Novel RANK antagonists for the treatment of bone- resorptive

- disease : theoretical predictions and experimental validation. *J Bone Miner Res.* 2014 ; 29(6) :1466-77. [lien](#)
10. Baud'huin M, Duplomb L, **Teletchea S**, Lamoureux F, Ruiz-Velasco C, Maillason M, Redini F, Heymann MF, Heymann D. Osteoprotegerin : multiple partners for multiple functions. *Cytokine Growth Factor Rev.* 2013 ; 24(5) :401-9. [lien](#)
  9. Launay G, **Téletchéa S**, Wade F, Pajot-Augy E, Gibrat JF, Sanz G. Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Eng Des Sel.* 2012 ; 25(8) :377-86. [lien](#)
  8. Picarda G, Surget S, Guiho R, **Téletchéa S**, Berreur M, Tirode F, Pellat- Deceunynck C, Heymann D, Trichet V, Rédini F. A functional, new short isoform of death receptor 4 in Ewing's sarcoma cell lines may be involved in TRAIL sensitivity/resistance mechanisms. *Mol Cancer Res.* 2012 ; 10(3) :336-46. [lien](#)
  7. Picarda G, Trichet V, **Téletchéa S**, Heymann D, Rédini F. TRAIL receptor signaling and therapeutic option in bone tumors : the trap of the bone microenvironment. *Am J Cancer Res.* 2012 ; 2(1) :45-64. [lien](#)
  6. **Téletchéa S**, Skauge T, Sletten E, Kozelka J. Cisplatin adducts on a GGG sequence within a DNA duplex studied by NMR spectroscopy and molecular dynamics simulations. *Chemistry.* 2009 ; 15(45) :12320-37. [lien](#)
  5. Baud'huin M, Duplomb L, **Téletchéa S**, Charrier C, Maillason M, Fouassier M, Heymann D. Factor VIII-von Willebrand factor complex inhibits osteoclastogenesis and controls cell survival. *J Biol Chem.* 2009 ; 284(46) :31704-13. [lien](#)
  4. **Teletchea S**, Komeda S, Teuben JM, Elizondo-Riojas MA, Reedijk J, Kozelka J. A pyrazolato-bridged dinuclear platinum(II) complex induces only minor distortions upon DNA-binding. *Chemistry.* 2006 ; 12(14) :3741-53. [lien](#)
  3. Drumm M, **Teletchea S**, Kozelka J. Recognition complex between the HMG domain of LEF-1 and its cognate DNA studied by molecular dynamics simulations with explicit solvation. *J Biomol Struct Dyn.* 2005 ; 23(1) :1-11. [lien](#)
  2. **Teletchea S**, Hartmann B, Kozelka J. Discrimination between BI and BII conformational substates of B-DNA based on sugar-base interproton distances. *J Biomol Struct Dyn.* 2004 ; 21(4) :489-94. [lien](#)
  1. Harvey SC, Wang C, **Teletchea S**, Lavery R. Motifs in nucleic acids : molecular mechanics restraints for base pairing and base stacking. *J Comput Chem.* 2003 ; 24(1) :1-9. [lien](#)

## Brevet

- **Peptides targeting receptor activator of nuclear factor-kappa b (RANK) and their applications.** Heymann D, **Teletchea S**, Stresing V. 2012, WO2013093039A1.

## Article en pré-publication

- **Comparing two deep learning sequence-based models for protein-protein interaction prediction.** Richoux F, Servantie C, Borès C, **Teletchea S**. 2019, <https://doi.org/10.48550/arXiv.1901.06268>

## 4.1 Encadrement d'étudiants de M2 recherche

J'ai co-encadré 2 étudiants durant leur thèse, encadré ou co-encadré 12 étudiants en master 2, encadré 22 étudiants de M1. Plus de 15 étudiants de licence (L2, L3) ont fait des stages de

découverte de la bioinformatique, soit sous forme de TER, soit lors de stages volontaires en dehors des périodes d'enseignement.

Ces étudiants ont participé à l'élaboration ou à l'avancement de travaux en cours ou publiés, les projets sur lesquels ils ont travaillé ont parfois été réalisés par des étudiants qui ont effectué plusieurs stages sous ma supervision, souvent en L3 puis en M1. Les étudiants ayant apporté une contribution significative au travail publié sont indiqués comme auteurs dans les articles.

1. **Mars 2016 – Sept 2016 : Chaïmae Nachaoui – Université de Nantes**  
Caractérisation structurale de l'interaction RhoA/ArhGEF1
2. **Mars 2016 – Sept 2016 : Tristan Riolland – Université de Nantes**  
Caractérisation structurale de l'interaction Bcl-Xl – Bak et comparaison avec l'interaction BAK-PUMA
3. **Mars 2016 – Sept 2016 : Maxime Malenfant – Université de Aix-Marseille** Développement d'une interface Galaxy pour l'analyse de données en protéomique fonctionnelle
4. **Mars 2018 – Sept 2018 : Charlotte Brunel-Cadic – Université de Nantes**  
Caractérisation structurale et fonctionnelle des effecteurs des galectines
5. **Mars 2018 – Sept 2018 : Cynthia Borès – Université de Nantes** Étude de la relation structure-fonction appliqué à la caractérisation des interfaces protéine-protéine
6. **Mars 2019 – Sept 2019 : Maximilien Berne – Université de Nantes**  
Caractérisation structurale de l'entropie conformationnelle des protéines de la famille Bcl-2
7. **Mars 2020 – Sept 2020 : Stéphane Jedele – Université de Nice, Polytech**  
Caractérisation structurale de l'entropie conformationnelle des protéines de la famille Bcl-2 (poursuite travail M. Berne)
8. **Mars 2021 – Septembre 2021 : Anne-Emeline Huard – Université de Nantes** Caractérisation structurale de l'entropie conformationnelle des protéines de la famille Bcl-2 (poursuite travail S. Jedele)
9. **Décembre 2021 – Juin 2022 : Ellyn Rousselot – Université de Nantes**  
Étude de la composition en protéine du génome de *Physarum polycephalum* (co-encadrement avec Céline Duc, laboratoire US2B équipe 4)
10. **Mars 2022 - Septembre 2022 : Hamady Demba Ba – Université de Nantes** Libprotein : un explorateur C++ et Python rapide et polyvalent des modifications post-traductionnelles des protéines
11. **Mars 2023 - Septembre 2023 : Damien Garcia – Université de Nantes**  
Étude des éléments structuraux déterminant la force de l'interaction entre la protéine Spike du SARS-Cov-2 et le récepteur ACE2 de l'hôte
12. **Mars 2023 - Septembre 2023 : César Hunault – Université de Nantes**  
Étude de l'impact des modifications post-traductionnelles du segment intrinsèquement désordonné sur la structure de Bcl-xL dans un contexte membranaire

Parmi les étudiants encadrés quatre étudiants ont poursuivi en thèse : Gautier Peyrat, Romain Launay et Mallauray Vié (encadrement de M1) et Stéphane Jedele (encadrement en M2). Tous les autres étudiants ont été recrutés rapidement après l'obtention de leur diplôme de master 2.

## 4.2 Co-encadrement de thèses

J'ai eu l'occasion de co-encadrer une thèse et un deuxième co-encadrement a été initié en novembre 2022 grâce à l'obtention d'un financement de thèse CIFRE. Dans les deux cas j'ai assuré l'essentiel des encadrements en demandant ponctuellement un appui au Pr. Offmann, le co-encadrant principal des deux étudiants.

### 4.2.1 2016-2019, Ennys GHEYOUCHE, co-encadrement à 40%

Co-encadrement avec B. Offmann (50%) et G. Loirand (10%).

Le projet de thèse a pour titre **Développement de méthodologies innovantes pour la caractérisation d'interfaces protéine-protéine et la conception d'inhibiteurs spécifiques**.

Le support du financement de la thèse de Ennys Gheyouché a été apporté par le projet PIRAMID. Compte-tenu des attentes sur le projet, il a développé des approches innovantes pour la modélisation des interactions protéine-protéine et identifié des inhibiteurs spécifiques de l'interaction RhoA-ArhGEF1. Il a aussi participé à la mise en place d'un portail de synthèse des résultats pluridisciplinaires combinant biologie, chimie et modélisation. Ces travaux ont fait l'objet de deux publications [REF 18][REF 21] et ont été présentés par Ennys Gheyouché lors de 4 congrès nationaux et internationaux. Son doctorat a duré 3 ans.

Ennys Gheyouché a été recruté au sortir de sa thèse dans par l'entreprise Aqemia à Paris (<https://www.aqemia.com/>) puis poursuit actuellement sa carrière dans l'entreprise Iktos à Paris (<https://iktos.ai/>).

### 4.2.2 2022-2025, Elena ALVAREZ-SANCHEZ, co-encadrement à 40%

Co-direction avec B. Offmann (50%) et Simon Huet (10%)

Elena Alvarez-Sanchez effectue sa recherche sur la **Caractérisation structurale des déterminants de la spécificité des Nanofitines pour leurs cibles**. Ce travail de thèse est permis par l'obtention d'un financement CIFRE auprès de l'ANRT par la société Affilogic. Durant les six premiers mois de sa thèse l'étudiante a pu mettre un place un protocole robuste d'analyse des nanofitines qui sera présenté en mai 2023 lors du congrès international du GGMM à Toulouse (<https://ggmm2023.sciencesconf.org/>).

## 4.3 Participation à des jurys de thèse

1. **Thèse de Akhila Melarkode Vattekatte, examinateur, 25 mars 2019, Université de Paris** 3D structures of camelid antibodies : in-silico analyses, prediction and their dynamics.
2. **Thèse de François Bonardel, examinateur, 8 février 2021** Etude bioinformatique des lectines : nouvelles classification et prédiction dans les génomes.
3. **Thèse de Lindsay Rose, invité, 30 novembre 2022** Identification et caractérisation d'un nouvel inhibiteur de Rac1 à visée thérapeutique : implication en physiopathologies bronchiques et oncologie
4. **Thèse de Dylan Sérillon, examinateur, 12 avril 2023** Utilisation de méthodes computationnelles pour la prédiction d'énergie libre de Gibbs basée sur

des modèles d'apprentissage automatisés. thérapeutique. Implication en physiopathologies bronchiques et oncologie

## 4.4 Comités de sélection et instances nationales

J'ai participé à 4 comités de sélection de maître de conférences et à 1 évaluation d'autorisation de soutenance d'une HDR.

- **2018** Comité de sélection numéro 4364 pour le profil de **Bioinformatique Structurale** de l'université Paris Diderot.
- **2019** Examen de l'autorisation de soutenance d'une HDR, Université de Nice
- **2020** Comité de sélection numéro 4583 pour le profil de **Bio-informatique structurale et « Protein Design »**, Université de Lyon
- **2023** Comité de sélection numéro 328 pour le profil de **Chemoinformatique, Bioinformatique, Biostatistique** de l'université Paris Cité
- **2023** Comité de sélection numéro 4473 pour le profil de **Biostatisticien/Data scientist** de l'université de Angers

Depuis 2019 je suis membre suppléant du Conseil National des Universités en section 64. A ce titre j'ai participé à 3 comités de sélection : 1 comité d'avancement de grade, 1 comité d'examen de PEDR et 1 comité d'examen de demande RIPEC.

## 4.5 Implication dans des programmes de recherche

Depuis mon recrutement en tant que maître de conférences à Nantes Université en 2014, j'ai participé à de nombreux projets dont 11 ont été soutenus financièrement. Ces projets sont centrés autour de deux thématiques : (i) l'étude des interactions protéine-protéine, (ii) la conception d'inhibiteurs innovants sur une cible protéique donnée.

Les financements de ces projets ont été obtenus auprès de Nantes Université (2), d'associations (3), de la région Pays de La Loire (2), de l'agence nationale pour la Recherche (2) ou d'entreprises (2).

J'ai été **porteur de projet** pour la mise en place d'une stratégie de prédiction des interactions entre protéines par apprentissage profond (projet PREDICTI). Ce projet soutenu par Nantes Université a consisté à conjuguer les expertises en apprentissage profond (*deep learning*) d'un maître de conférences en informatique, Florian Richoux, avec mon expertise sur la modélisation moléculaire et l'analyse des complexes protéine-protéine. Nous avons pu explorer plusieurs pistes de prédiction avec deux étudiantes, Charlène Servantie (supervision par F. Richoux) et Cynthia Borès (sous ma supervision). Ce travail a été publié en tant que pré-soumission sur arxiv [PREPRINT 1]. Suite à l'obtention d'une disponibilité par F. Richoux pour un séjour au Japon, ce projet précurseur des avancées récentes dans le domaine a été suspendu.

Le projet le plus structurant pour lequel j'ai été impliqué est le projet PIRAMID (<http://pyramid-research.fr/>). Ce projet initié en 2015 a impliqué plus de 50 chercheurs de 11 laboratoires de la région Pays de La Loire. Il a été architecturé autour de 4 axes (+1 durant le projet) afin de mettre en oeuvre un tryptique scientifique biologie-chimie-modélisation. J'ai été membre du comité de pilotage du projet pour déterminer lors de chaque réunion l'état d'avancement des projets, les étapes de valorisation à mettre en oeuvre, et participer aux discussions avec les instances d'évaluation chaque année. J'ai été responsable de deux axes pour

la modélisation moléculaire des interfaces : (i) Axe 1 - RhoA, et (ii) l'Axe 2 - BCL-xL. Ce projet structurant à grande échelle a permis l'écriture de deux articles scientifiques, la mise en place de discussion sur la valorisation sous forme de brevet des molécules prometteuses identifiées, et l'obtention de 5 financements complémentaires.

Au niveau national je suis impliqué dans le groupe de travail MASIM du CNRS (<https://gt-masim.cnrs.fr/>) et j'ai participé au groupe de travail pour la mise en place de la fédération de recherche Thémosia (<https://www.chimie-theorique.cnrs.fr/spip.php?article338>). J'interviens aussi régulièrement dans les formations du Pôle Ouest du Réseau Français de Chimie Théorique (<https://www.chimie-theorique.cnrs.fr/spip.php?rubrique5>).

## 4.6 Financements en tant que partenaire, responsable ou porteur

1. **Projet Régional "Nouvelles équipes, nouvelles thématiques "ReMSIP".**  
Programme porté par B. Offmann et C. Grandjean, 185 k€ pour le laboratoire. Mon rôle a consisté à être partenaire scientifique du projet. J'ai mis en place et évalué les méthodologies de criblage virtuel des inhibiteurs de galectines.
2. **2013-2016 : Recherche d'inhibiteurs innovants de galectines.** Financement FRM Programme "Chimie pour la Médecine", porté par Cyrille GRANDJEAN), 220 k€ dont 144 k€ pour le laboratoire. Mon rôle a consisté à être partenaire scientifique du projet. J'ai participé à ce projet par le criblage de dérivés glycosidiques sur les galectines.
3. **2015-2017 : Recherche d'inhibiteurs innovants de galectines.**  
Contrat industriel avec le laboratoire Servier, budget global de 162 k€. Projet porté par C. Grandjean. Mon rôle a consisté à être partenaire scientifique du projet. J'ai apporté mon expertise en modélisation moléculaire et drug design. J'ai présenté les prédictions issues du criblage virtuel lors des réunions d'avancement du projet (à Nantes, en visio, ou dans les locaux du laboratoire Servier).
4. **2016-2021 : Protein interactions in rational approaches for medicinal innovative drugs (PIRAMID).**  
Financement par la région Pays de La Loire, budget global de 1,2 M€, dont 112 k€ pour mon partenariat. J'ai participé à la rédaction de l'appel à projets, tout particulièrement la partie concernant la valorisation des résultats, dans l'organisation et la présentation scientifique des travaux intermédiaires lors des réunions du projet et dans la rédaction des documents de synthèse-bilan pour chaque axe. Des résultats intermédiaires ont été publiés (deux articles) et les démarches de valorisation des molécules actives identifiées sont initiées depuis 2020.
5. **2017-2018 : Prédiction des interactions protéiques par apprentissage profond (PREDICTI).**  
Université de Nantes Phase 1, budget global de 15 k€ dont 7,5k€ sous ma responsabilité. J'ai été le porteur de projet de ce financement, ce qui a consisté à rédiger l'appel à projet, gérer le budget, rédiger le rapport de suivi, et co-rédiger de l'article décrivant le travail.
6. **2018 : Étude de dérivés glycosidiques pour inhiber la progression de cellules tumorales.** Financement La Ligue contre le Cancer Grand Ouest. Porteur de projet : Cyrille Grandjean, budget de 27 k€ dont 15 k€ pour le laboratoire. Mon rôle a consisté à être partenaire scientifique du projet sur la partie in silico.

7. **2019-2022 : Targeting Rho protein/Rho exchange factor interaction to disrupt Rho protein signaling (TROPIC), iSite Next.** Budget global de 140 k€ dont 35 k€ sous ma responsabilité. Je suis responsable du projet pour la partie in silico, j'ai participé à la rédaction de l'appel à projets, à la rédaction des documents de rendus intermédiaires et aux réunions de travail du projet.
8. **2021 : Analyse des complexes multiprotéiques de BCL-xL réfractaires à l'action inhibitrice des molécules BH3 mimétiques.** Porteurs de projet L. Maillet et S. Téletchéa, budget de 20 k€, dont 3,5 k€ sous ma responsabilité. J'ai participé à la rédaction de l'appel à projet, aux réunions de coordination et à la supervision d'une étudiante de master 2 (Anne-Emeline Huard).
9. **2022-2025 : Financement CIFRE avec la société Affilogic.** Ce financement a été octroyé par l'ANRT pour le financement d'une thèse pour l'étude du mécanisme de reconnaissance moléculaire des cibles des nanofitines. L'étudiante recrutée pour ce projet est Elena Alvarez-Sanchez.
10. **2022-2026 : Pre-clinical development of Rac1-dependent pathway inhibitors to fight severe asthma, projet ANR ORBIT.** Ce projet porté par Vincent SAUZEAU, avec un budget de 368 k€ dont 74 k€ sous ma responsabilité, a pour objectif de comprendre comment la petite GTPase Rac1 interagit avec ses facteurs d'échange de guanine afin de concevoir des molécules inhibitrices de cette interaction. Ma contribution au projet consistera en l'analyse in silico des complexes protéiques et au criblage virtuel des interfaces pour rechercher des molécules inhibitrices innovantes et sélectives. Ce projet sera réalisé à l'aide d'un post-doctorant pour un contrat d'un an à recruter en 2023.
11. **2022-2026 : Bcl-xL deamidation in platelets : mere age marker, or lifespan modulator ? projet ANR WRINXL** Projet porté par Muriel Priault, Bordeaux avec un budget total de 545 k€ dont 112 k€ sous ma responsabilité. Un post-doctorant de 18 mois sera recruté dès 2023 sur le projet pour déterminer les conséquences des modifications post-traductionnelles sur l'interaction entre Bcl-xL et BAK afin de déterminer des molécules protectrices de ces interactions.

## 4.7 Collaborations nationales et internationales

J'ai participé à de nombreux projets traitant des interactions protéine-protéine qui n'ont pas forcément été adossés à des financements spécifiques. Plusieurs concernent des travaux entre deux équipes de notre laboratoire, par exemple avec C. Grandjean, responsable de l'équipe de glycobiologie, B. Offmann sur les approches de bioinformatique structurale et Céline Duc pour les recherches de gènes d'intérêt chez *Physarum polycephalum*. Au niveau régional, j'ai aussi des collaborations de long terme avec des membres de l'Institut du Thorax, principalement Gervaise Loirand et Vincent Sauzeau (UMR\_S INSERM 1087/CNRS 6291, Nantes Université). A la suite du projet PIRAMID je continue à collaborer avec Laurent Maillet (CRCINA, UMR\_S1232 INSERM, Nantes Université) pour la caractérisation structurale et fonctionnelle des membres de la famille Bcl-2 en contexte membranaire. De par mon expertise dans le contexte osseux, je participe aussi à l'étude des partenaires impliqués protéiques impliqués dans le remodelage osseux avec Laurent Beck et Pierre Guihard (UMR\_S1229 INSERM, Nantes Université). J'ai aussi une collaboration de long terme sur la compréhension du rôle de la cytokine IL-34 dans le métabolisme osseux avec Dominique Heymann (US2B, CNRS UMR 6286, Nantes Université). Au niveau national je participe à la mise en place de bases de données spécialisées et de méthodes de prédiction pour la caractérisation des protéines impliquées dans la structure et la fonction du globule rouge avec Catherine Etchebest, Aline Floch et Alexandre de Brevern (BIGR, UMR\_S 1134 INSERM, Université Paris Cité).



J'ai aussi une collaboration de long terme avec F. Téletchéa à l'Université de Lorraine (UR AFPA, USC INRA 340) et à ce titre une collaboration internationale avec Paul Venturelli (Ball State University, Indiana State, USA). Dans ces travaux nous mettons en ligne des informations liées aux stratégies de reproduction des poissons (<https://www.storefish.org>) afin de rassembler des éléments connus dans la littérature scientifique, qui permettent l'analyse des facteurs importants pour la croissance des poissons. Ce bilan accessible sous forme de base de données va permettre de faire des analyses plus avancées comme celles liées aux problématiques de réchauffement climatique, comprendre les facteurs génétiques liés à ces taux de croissance des poissons ou encore de caractériser l'influence entropique sur la répartition des poissons dans le monde.

L'ensemble de ces collaborations, basées principalement sur des approches de bioinformatique et bioinformatique structurale, me permettent de trouver des synergies entre les différents projets. Pour la plupart des projets ces travaux ont donné lieu à des publications dans des journaux à comité de lecture, d'autres travaux sont en cours d'écriture, et ils permettent aussi aux étudiants du master de bioinformatique de Nantes d'avoir des sujets variés pour leurs stages de M1 et/ou M2.

## 4.8 Vie universitaire

J'ai été membre élu du conseil d'unité de l'UGIP pour le collège des C/EC de 2018 à 2022 et membre nommé du conseil de département biologie (depuis 2017).

Je suis membre élu titulaire, au titre du syndicat UNSA/SNPTEs du [Comité Social d'Administration \(CSA\) de Nantes Université](#) pour la période 2023-2026 après avoir été membre suppléant du CTE sur la précédente mandature (2019-2022). A ce titre j'ai participé à plusieurs groupes de travail dont celui sur les lignes directrices de gestion de la carrière des EC/C et celui concernant les modalités de gestion des contractuels.

Je suis membre suppléant nommé de la 64ème section du CNU depuis 2019. J'ai participé à trois concours de recrutement de MCU en section 64, j'ai évalué deux dossiers de promotion de collègues de l'université de Nantes, j'ai été membre du jury de quatre thèses.

J'ai co-organisé l'école thématique internationale DynaMoPPI, Nantes (<https://dynamoppi.sciencesconf.org/>) qui a accueilli 15 étudiants français et internationaux. J'ai pris en charge une majeure partie de l'organisation de cette école thématique (sous l'égide de la SFB et du GGMM) en gérant les inscriptions, l'interaction avec les tutelles, le budget, la venue des intervenants et leur hébergement et en coordonnant les demandes de co-financement.

J'ai été membre du bureau et trésorier de la SFBI de 2009 à 2015. Je suis membre de la SFB et de la SFBI depuis 2005.

Troisième partie

Projet de recherche

Les récents développements autour de la prédiction de structures de protéines comme AlphaFold [74], RoseTTAFold[72], ESMfold[81], OmegaFold [86], et encore d'autres [75], permettent d'étudier les structures du vivant avec des approches à haut débit inimaginables jusqu'alors. Les approches menées par Google (DeepMind/AlphaFold) ou encore Facebook (ESMfold) ont ainsi permis de faire un saut qualitatif sans précédent dans la prédiction de la structure tridimensionnelle des protéines monomériques. Beaucoup de recherches sont cependant nécessaires pour caractériser les changements conformationnels des protéines et les interactions entre plusieurs protéines, menant à la transduction du signal dans les cellules. Ce changement de paradigme permet aussi d'envisager l'exploration de l'immensité de l'espace des séquences et des structures des protéines, un domaine qui verra le 21<sup>ème</sup> siècle proposer des solutions biologiques à des problèmes chimiques créés au 20<sup>ème</sup> siècle : nous entrons de plain-pied dans l'ère du *de novo design* [59]. Mon projet de recherche s'articulera autour de deux axes. Le premier axe consistera en l'étude sans a priori de l'espace des protéines, pour identifier des îlots de stabilité tridimensionnelle de séquences non naturelles. Le second axe portera sur la mise en place et la maintenance de serveurs et base de données permettant de faire le lien entre les séquences, les structures et les fonctions des protéines.

# Chapitre 1

## Prédiction de repliements protéiques non naturels

La séquence primaire de chaque protéine est composée d'une suite d'acides aminés qui comprend à chaque position 20 possibilités. Pour une protéine de 100 acides aminés, cela représente déjà un nombre astronomique de combinaisons avec plus de  $10^{143}$  séquences, ce qui est de nos jours impossible à énumérer de manière exhaustive. Néanmoins, si chaque séquence était produite virtuellement, et s'il était possible de faire la prédiction de structure tridimensionnelle en une picoseconde, il faudrait pourtant encore  $10^{123}$  années pour prédire la structure de toutes ces protéines. Il est donc impossible et inutile de vouloir explorer la diversité des protéines par une approche de force brute. La Nature semble contourner cette complexité en conservant des protéines capables de se replier dans un temps très court, ce principe est connu sous le nom de paradoxe de Levinthal [10].

L'espace des protéines exploré par le vivant n'est cependant probablement pas le seul dans lequel il est possible d'identifier des repliements stables, les changements sélectionnés par l'évolution n'ont tout simplement pas eu assez de temps pour atteindre ces îlots de stabilité. Dans ce projet de recherche, je cherche donc à déterminer s'il est possible d'identifier des séquences de protéines non naturelles présentant un repliement stable. J'ai pour cela mis en place un calcul basé sur l'analyse des positions discrètes de l'espace occupé par les acides aminés dans les structures protéiques, appelées **rotamères**, et calculé le produit de ces conformations des chaînes latérales avec 4 états discrets de la chaîne principale des protéines : (i) hélice  $\alpha$ , (ii) brin  $\beta$ , (iii) tour, (iv) aléatoire (*coil*).

Ce calcul est ensuite normalisé par la longueur de la protéine pour obtenir une valeur de complexité intrinsèque de la protéine comprise entre 0 et 100% en fonction de la composition en acides aminés de la séquence. J'ai nommé ce calcul le *Structural Complexity Index*. L'analyse de la complexité du protéome humain est présentée dans la Figure 1.1.

L'analyse du protéome humain indique que la sélection naturelle a engendré des protéines à complexité structurale intrinsèque moyenne, on trouve ainsi que le score de la plupart des protéines ont un score proche de 50% avec une dispersion très limitée autour de cette moyenne. Ce résultat s'explique par la longueur moyenne d'une protéine humaine, autour de 400 acides aminés, ce qui permet d'avoir en moyenne une utilisation de chacun des acides aminés environ 20 fois. Cette distribution indique néanmoins qu'il est possible de trouver rapidement des séquences de protéines possédant un index structural identique avec notre méthode de calcul,

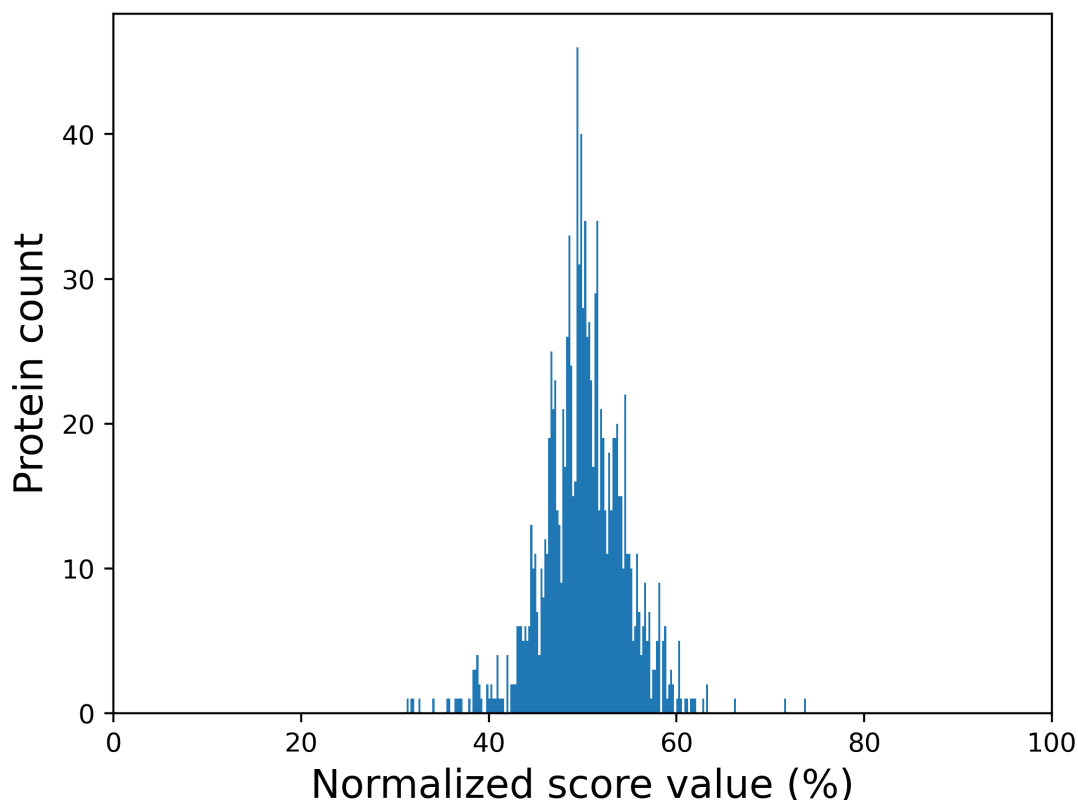


FIGURE 1.1 – Analyse de la complexité structurale liée à la séquence du protéome humain.

sans que cette séquence provienne d'un organisme vivant. J'ai donc simulé un ensemble de séquences présentant un score proche de 50% et étudié ce que les méthodes récentes de prédiction de structure proposeraient.

La Figure 1.2 présente les prédictions des structures de séquences non naturelles pour trois protéines potentielles. La coloration locale des modèles indique que de nombreuses sections des protéines sont modélisées avec peu de confiance (couleur rouge) comme pour la séquence 14806, avec une confiance moyenne à faible pour la séquence 14035 ou une confiance forte, proche d'un résultat issu d'analyse par cristallographie aux rayons X dans le cas de la séquence 1242.

Ces résultats préliminaires indiquent qu'il semble possible de déterminer une séquence non naturelle capable de présenter des formes régulières d'après les méthodes de prédiction de structures les plus avancées. Il reste à mettre en place une méthode consensus pour proposer de nouvelles séquences non naturelles présentant suffisamment d'éléments réguliers, bien modélisés, semblant fournir une protéine compacte, correctement repliée. Cette approche a fait l'objet d'une demande de financement en partenariat avec Jean-Christophe Gelly du laboratoire DSIMB à Paris, un expert des méthodes de prédiction de structure et de leur évaluation, et en partenariat avec Emilie Camberlein du laboratoire US2B, spécialiste de la production et caractérisation expérimentale de protéines recombinantes. Nous voulons prédire *in silico* la structure de 100 000 protéines non naturelles pour en sélectionner une centaine qui sera exprimée en cellule puis caractérisée par des approches biophysiques. Cette combinaison d'approches virtuelles et expérimentales permettra de continuer à développer des approches innovantes pour la prédiction de structure des protéines.

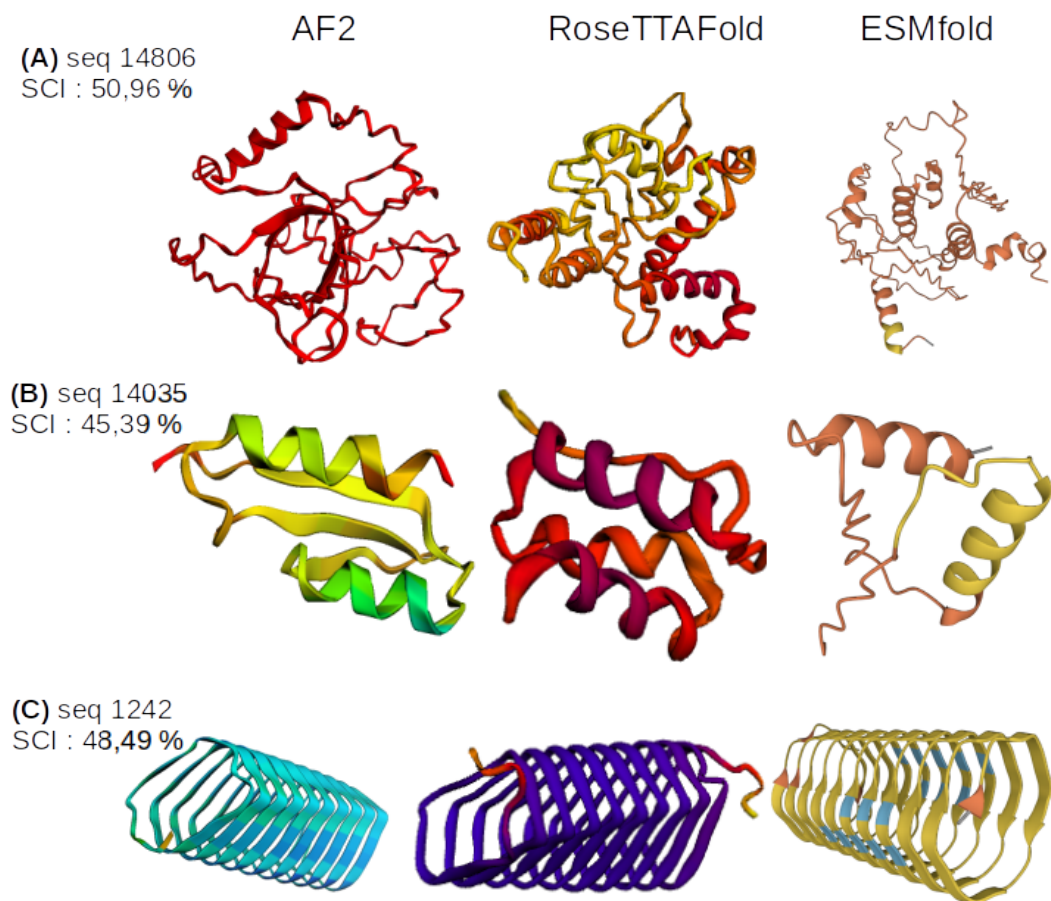


FIGURE 1.2 – Prédiction de structure d'une sélection de séquences non naturelles avec AlphaFold(2), RoseTTAFold et ESMFold. Les structures modélisées sont colorées selon la qualité locale estimée de la prédiction, du rouge (mal prédit) au bleu (très bonne prédiction).

# Chapitre 2

## Intégration de données biologiques

La littérature scientifique contient de très nombreuses données textuelles sur des thématiques très précises. A l'heure actuelle, et malgré les progrès récents dans les méthodes du traitement du langage naturel, il reste cependant très difficile d'extraire et d'organiser ces informations [70]. J'ai été amené à mettre en place plusieurs bases de données liées à des domaines très spécifiques et très variés. La première, RESPIRE, a consisté à faire un bilan des protéines identifiables dans le globule rouge en combinant des données de protéomique et des méthodes de prédiction de structure de protéines (Rosetta, I-TASSER). La seconde, RHeference, a consisté à organiser les données qu'une spécialiste du groupe RhD, Aline Floch, assemblait dans un fichier à plat. La troisième base de données, STOREFISH, contient des informations sur les caractères influant la croissance des poissons. La conception et la mise en place de ces bases de données accessibles sur internet m'a permis de déterminer les éléments importants de l'organisation des tables contenant les données, mais aussi de mettre en place des mécanismes de mise à jour. Je vais développer des interfaces de mises à jour contrôlées des données pour permettre à plusieurs utilisateurs de contribuer à l'évolution de celles-ci. À travers la définition de règles d'import de données et l'utilisation des méthodes avancées de traitement du langage naturel (*Natural Language Processing*, NLP), je vais continuer à procéder à l'intégration de données complexes pour la production de contenus scientifiques organisés.

# Chapitre 3

## Combiner structure et fonction

Je vais continuer à développer ces approches intégrées faisant appel à des méthodologies avancées pour les analyses de séquence, la prédiction de la structure de macromolécule, mais aussi pour procéder à l'extraction contrôlée d'informations disponibles dans la littérature scientifique. L'expertise acquise dans les travaux supervisés jusqu'à maintenant me permettra de développer des méthodologies originales utilisant des algorithmes avancés de prédiction de structure et des prédictions dérivées de méthodes de *machine learning* avancées.

Grâce à ces approches combinées, mon objectif est de **prédire des séquences de protéines à activité spécifique**. Comme nous l'avons vu plus haut, une première approche a consisté à caractériser à quel point les protéines naturelles faisaient usage des acides aminés. Nous avons vu que dans le cadre du protéome humain, cette diversité était globalement moyennée, et qu'il était ainsi possible de trouver rapidement de nouvelles séquences protéiques de complexité contrôlée. Après cette première étape, je vais mettre en place un protocole de prédiction de structure de protéines s'appuyant sur des annotations fonctionnelles disponibles dans la littérature scientifique, comme les modifications post-traductionnelles, les partenaires protéiques, les voies de signalisation dans lesquelles les protéines sont impliquées, ou encore les ligands ciblant ces protéines. Une première étape de cette stratégie a été réalisée en 2022 avec la création de la libprotein avec l'étudiant de master 2 Hamady. Mes recherches futures vont ainsi s'inscrire pleinement dans l'ère du *de novo protein design*, dans le but d'explorer le **foldiverse**.



# Bibliographie

1. *The Nobel Prize in Chemistry 1946* NobelPrize.org.  
<https://www.nobelprize.org/prizes/chemistry/1946/summary/> (2023).
2. *The Nobel Prize in Chemistry 1958* NobelPrize.org.  
<https://www.nobelprize.org/prizes/chemistry/1958/summary/> (2023).
3. *The Nobel Prize in Chemistry 1962* NobelPrize.org.  
<https://www.nobelprize.org/prizes/chemistry/1962/summary/> (2023).
4. *The Nobel Prize in Physiology or Medicine 1962* NobelPrize.org.  
<https://www.nobelprize.org/prizes/medicine/1962/summary/> (2023).
5. ROSENBERG, B. Platinum coordination complexes in cancer chemotherapy. *Die Naturwissenschaften* **60**, 399-406 (1973).
6. HILL, J. M. *et al.* Clinical studies of Platinum Coordination compounds in the treatment of various malignant diseases. *Cancer Chemotherapy Reports* **59**, 647-659 (1975).
7. *The Nobel Prize in Chemistry 1980* NobelPrize.org.  
<https://www.nobelprize.org/prizes/chemistry/1980/summary/> (2023).
8. EGLINTON, G., HENDERSON-SELLERS, A. & MOORBATH, S. Cosmochemistry and the origin of life. *Nature* **292**, 669-669 (1981).
9. CLOUD, P. How life began. *Nature* **296**, 198-199 (1982).
10. ZWANZIG, R., SZABO, A. & BAGCHI, B. Levinthal's paradox. *Proceedings of the National Academy of Sciences* **89**. Publisher : Proceedings of the National Academy of Sciences, 20-22 (1992).
11. LOVE, J. J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-795 (1995).
12. ASHBURNER, M. *et al.* Gene Ontology : tool for the unification of biology. *Nature Genetics* **25**, 25-29 (2000).
13. PALCZEWSKI, K. *et al.* Crystal structure of rhodopsin : A G protein-coupled receptor. *Science (New York, N.Y.)* **289**, 739-45 (2000).
14. BOWLER, P. in *International Encyclopedia of the Social & Behavioral Sciences* 4986-4992 (Elsevier, 2001).
15. CAMPBELL, I. D. The march of structural biology. *Nature Reviews Molecular Cell Biology* **3**. Number : 5 Publisher : Nature Publishing Group, 377-381 (2002).
16. MARIN, A., POTHIER, J., ZIMMERMANN, K. & GIBRAT, J.-F. FROST : A filter-based fold recognition method. *Proteins : Structure, Function, and Genetics* **49**, 493-509 (2002).
17. HARVEY, S. C., WANG, C., TELETSCHEA, S. & LAVERY, R. Motifs in nucleic acids : molecular mechanics restraints for base pairing and base stacking. *Journal of Computational Chemistry* **24**, 1-9 (2003).
18. HOUZELSTEIN, D. *et al.* Phylogenetic Analysis of the Vertebrate Galectin Family. *Molecular Biology and Evolution* **21**, 1177-1187 (2004).
19. OOSTENBRINK, C., VILLA, A., MARK, A. E. & VAN GUNSTEREN, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation : The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **25**, 1656-1676 (2004).
20. THEOLEYRE, S. *et al.* The molecular triad OPG/RANK/RANKL : involvement in the orchestration of pathophysiological bone remodeling. *Cytokine & Growth Factor Reviews* **15**, 457-475 (2004).
21. WHITE, S. H. The progress of membrane protein structure determination. *Protein Science* **13**, 1948-1949 (2004).

22. WITTRANT, Y. *et al.* RANKL/RANK/OPG : new therapeutic targets in bone tumours and associated osteolysis. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1704**, 49-57 (2004).
23. DRUMM, M., TELETCHÉA, S. & KOZELKA, J. Recognition complex between the HMG domain of LEF-1 and its cognate DNA studied by molecular dynamics simulations with explicit solvation. *Journal of Biomolecular Structure & Dynamics* **23**, 1-11 (août 2005).
24. SÖDING, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951-960 (2005).
25. TELETCHÉA, S. *et al.* A pyrazolato-bridged dinuclear platinum(II) complex induces only minor distortions upon DNA-binding. *Chemistry (Weinheim an Der Bergstrasse, Germany)* **12**, 3741-53 (2006).
26. BAUD’HUIN, M., LAMOUREUX, F., DUPLOMB, L., RÉDINI, F. & HEYMANN, D. RANKL, RANK, osteoprotegerin : key partners of osteoimmunology and vascular diseases. *Cellular and Molecular Life Sciences* **64**, 2334-2350 (2007).
27. CHEREZOV, V. *et al.* High Resolution Crystal Structure of an Engineered Human 2-Adrenergic G protein-Coupled Receptor. *Science (New York, N.Y.)* **318**, 1258-1265 (2007).
28. GOODMAN, S. R., KURDIA, A., AMMANN, L., KAKHNIASHVILI, D. & DAESCU, O. The Human Red Blood Cell Proteome and Interactome. *Experimental Biology and Medicine* **232**, 1391-1408 (2007).
29. CHVÁLOVÁ, K., SARI, M.-A., BOMBARD, S. & KOZELKA, J. LEF-1 recognition of platinated GG sequences within double-stranded DNA. Influence of flanking bases. *Journal of Inorganic Biochemistry* **102**, 242-250 (2008).
30. HESS, B., KUTZNER, C., van der SPOEL, D. & LINDAHL, E. *GROMACS 4 : Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation* (2009).
31. JAYARAMAN, K. Still clueless about origin of life. *Nature India* (2008).
32. TALY, J.-F., MARIN, A. & GIBRAT, J.-F. Can molecular dynamics simulations help in discriminating correct from erroneous protein 3D models? *BMC Bioinformatics* **9**, 6 (2008).
33. BAUDHUIN, M. *et al.* Factor viii/von willebrand factor complex inhibits rankl-induced osteoclastogenesis and controls cell survival. *Bone* **44**, S333-S334 (2009).
34. GAUTO, D. F., DI LELLA, S., GUARDIA, C. M. A., ESTRIN, D. A. & MARTÍ, M. A. Carbohydrate-Binding Proteins : Dissecting Ligand Structures through Solvent Environment Occupancy. *The Journal of Physical Chemistry B* **113**, 8717-8724 (2009).
35. MORRIS, G. M. *et al.* AutoDock4 and AutoDockTools4 : Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**, NA (2009).
36. RYU, T. Benchmarking of BioPerl, Perl, BioJava, Java, BioPython, and Python for Primitive Bioinformatics Tasks and Choosing a Suitable Language. *International Journal of Contents* **5**, 6-15 (2009).
37. TÉLETCHÉA, S., SKAUGE, T., SLETTEN, E. & KOZELKA, J. Cisplatin Adducts on a GGG Sequence within a DNA Duplex Studied by NMR Spectroscopy and Molecular Dynamics Simulations. *Chemistry - A European Journal* **15**, NA (2009).
38. TROTT, O. & OLSON, A. J. AutoDock Vina : Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, NA (2009).
39. D’ALESSANDRO, A., RIGHETTI, P. G. & ZOLLA, L. The Red Blood Cell Proteome and Interactome : An Update. *Journal of Proteome Research* **9**, 144-163 (2010).
40. GUILLUY, C. *et al.* The Rho exchange factor Arhgef1 mediates the effects of angiotensin II on vascular tone and blood pressure. *Nature Medicine* **16**, 183-190 (2010).
41. KAUFMANN, K. W., LEMMON, G. H., DELUCA, S. L., SHEEHAN, J. H. & MEILER, J. Practically Useful : What the Rosetta Protein Modeling Suite Can Do for You. *Biochemistry* **49**, 2987-2998 (2010).
42. KEISER, M. J., IRWIN, J. J. & SHOICHET, B. K. The Chemical Basis of Pharmacology. *Biochemistry* **49**, 10267-10276 (2010).
43. LANE, N., ALLEN, J. F. & MARTIN, W. How did LUCA make a living? Chemiosmosis in the origin of life. *BioEssays* **32**, 271-280 (2010).
44. LOIRAND, G. & PACAUD, P. The role of Rho protein signaling in hypertension. *Nature Reviews Cardiology* **7**. Number : 11 Publisher : Nature Publishing Group, 637-647 (2010).
45. ROY, A., KUCUKURAL, A. & ZHANG, Y. I-TASSER : a unified platform for automated protein structure and function prediction. *Nat. Protocols* **5**, 725-738 (2010).

46. LEMIEUX, J. M., WU, G., MORGAN, J. A. & KACENA, M. A. DMSO regulates osteoclast development in vitro. *In Vitro Cellular & Developmental Biology - Animal* (2011).
47. PATNAIK, S. K., HELMBERG, W. & BLUMENFELD, O. O. BGMUT : NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. *Nucleic Acids Research* **40**, D1023-D1029 (D1 2011).
48. ANATOLE A. KLYOSOV. in *Glycobiology and Drug Design 0 t.*, 3-22 (American Chemical Society, 2012).
49. JONES, D. T., BUCHAN, D. W. A., COZZETTO, D. & PONTIL, M. PSICOV : precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **28**, 184-190 (2012).
50. LAUNAY, G. *et al.* Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Engineering Design & Selection* **25**, 377-386 (2012).
51. REMMERT, M., BIEGERT, A., HAUSER, A. & SÖDING, J. HHblits : lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **9**, 173-175 (2012).
52. CHERFILS, J. & ZEGHOUF, M. Regulation of Small GTPases by GEFs, GAPs, and GDIs. *Physiological Reviews* **93**, 269-309 (2013).
53. HEYMANN, D., TELETCHÉA, S. & STRESING, V. *brev.* WO2013093039A1 (2013).
54. PALCZEWSKI, K. & ORBAN, T. From Atomic Structures to Neuronal Functions of G Protein-Coupled Receptors. *Annual Review of Neuroscience* **36**, 139-164 (2013).
55. *The Nobel Prize in Chemistry 2013* NobelPrize.org.  
<https://www.nobelprize.org/prizes/chemistry/2013/summary/> (2023).
56. JASKOLSKI, M., DAUTER, Z. & WLODAWER, A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *The FEBS Journal* **281**, 3985-4009 (2014).
57. TELETCHÉA, S. *et al.* Novel RANK Antagonists for the Treatment of Bone-Resorptive Disease : Theoretical Predictions and Experimental Validation. *Journal of Bone and Mineral Research* **29**, 1466-1477 (2014).
58. HEGEDŰS, T. *et al.* Inconsistencies in the red blood cell membrane proteome analysis : generation of a database for research and diagnostic applications. *Database* **2015**, bav056 (2015).
59. HUANG, P.-S., BOYKEN, S. E. & BAKER, D. The coming of age of de novo protein design. *Nature* **537**, 320-327 (2016).
60. *Notre Histoire* Institut Pasteur.  
<https://www.pasteur.fr/fr/institut-pasteur/notre-histoire> (2023).
61. WEBB, B. & SALI, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **54** (2016).
62. LANIER, K. A. & WILLIAMS, L. D. The Origin of Life : Models and Data. *Journal of Molecular Evolution* **84**, 85-92 (2017).
63. MORAS, M., LEFEVRE, S. D. & OSTUNI, M. A. From Erythroblasts to Mature Red Blood Cells : Organelle Clearance in Mammals. *Frontiers in Physiology* **8** (2017).
64. CHAN, Y.-C. *et al.* Dissecting the Structure-Activity Relationship of Galectin-Ligand Interactions. *International Journal of Molecular Sciences* **19**. Number : 2 Publisher : Multidisciplinary Digital Publishing Institute, 392 (2018).
65. HILGER, D., MASUREEL, M. & KOBILKA, B. K. Structure and dynamics of GPCR signaling complexes. *Nature Structural & Molecular Biology* **25**, 4-12 (2018).
66. JOHANNES, L., JACOB, R. & LEFFLER, H. Galectins at a glance. *Journal of Cell Science* **131**, jcs208884 (2018).
67. TÉLETCHÉA, S., SANTUZ, H., LÉONARD, S. & ETCHEBEST, C. Repository of Enriched Structures of Proteins Involved in the Red Blood Cell Environment (RESPIRE). *PLoS One* **14**, e0211043 (2019).
68. ANDORNO, R. *et al.* Geneva Statement on Heritable Human Genome Editing : The Need for Course Correction. *Trends in Biotechnology* **38**, 351-354 (2020).
69. BONNARDEL, F., PEREZ, S., LISACEK, F. & IMBERTY, A. in *Lectin Purification and Analysis* (éd. HIRABAYASHI, J.) 1-14 (Springer US, New York, NY, 2020).
70. LEE, J. *et al.* BioBERT : a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36** (éd. WREN, J.) 1234-1240 (2020).
71. SAYERS, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **48**, D9-D16 (D1 2020).

72. BAEK, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**. Publisher : American Association for the Advancement of Science, 871-876 (2021).
73. GHEYUCHE, E., BAGUENEAU, M., LOIRAND, G., OFFMANN, B. & TÉLETCHÉA, S. Structural Design and Analysis of the RHOA-ARHGEF1 Binding Mode : Challenges and Applications for Protein-Protein Interface Prediction. *Frontiers in Molecular Biosciences* **8** (2021).
74. JUMPER, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
75. LAINE, E., EISMANN, S., ELOFSSON, A. & GRUDININ, S. Protein sequence-to-structure learning : Is this the end(-to-end revolution)? *Proteins : Structure, Function, and Bioinformatics* **89**, 1770-1786 (2021).
76. MISTRY, J. *et al.* Pfam : The protein families database in 2021. *Nucleic Acids Research* **49**, D412-D419 (D1 2021).
77. OUGHTRED, R. *et al.* The BioGRID database : A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science* **30**, 187-200 (2021).
78. *PDB 50th Anniversary : celebrating the future of structural biology* Nature. <https://www.nature.com/collections/habhcahgig> (2023).
79. CALLAWAY, E. What's next for AlphaFold and the AI protein-folding revolution. *Nature* **604**, 234-238 (2022).
80. CUNNINGHAM, F. *et al.* Ensembl 2022. *Nucleic Acids Research* **50**, D988-D995 (D1 2022).
81. LIN, Z. *et al.* Language models of protein sequences at the scale of evolution enable accurate structure prediction Pages : 2022.07.20.500902 Section : New Results. 2022.
82. PEREZ, S. & MAKSHAKOVA, O. Multifaceted Computational Modeling in Glycoscience. *Chemical Reviews* **122**, 15914-15970 (2022).
83. PHANG, R. & LIN, C.-H. Synthesis of Type-I and Type-II LacNAc-Repeating Oligosaccharides as the Backbones of Tumor-Associated Lewis Antigens. *Frontiers in Immunology* **13**, 858894 (2022).
84. ROMEI, M. *et al.* Protein folds as synapomorphies of the tree of life. *Evolution* **76**, 1706-1719 (2022).
85. The true legacy of Gregor Mendel : careful, rigorous and humble science. *Nature* **607**, 421-422 (2022).
86. WU, R. *et al.* High-resolution de novo structure prediction from primary sequence Pages : 2022.07.21.500999 Section : New Results. 2022.
87. H, K. *et al.* Evaluating the affinity and kinetics of small molecule glycomimetics for human and mouse galectin-3 using surface plasmon resonance. *SLAS discovery : advancing life sciences R & D*. Publisher : SLAS Discov (2023).
88. MCKUSICK-NATHANS INSTITUTE OF GENETIC MEDICINE Johns Hopkins University (Baltimore, M. *Online Mendelian Inheritance in Man, OMIM®* <https://omim.org/> (2023).
89. PAYSAN-LAFOSSÉ, T. *et al.* InterPro in 2022. *Nucleic Acids Research* **51**, D418-D427 (D1 2023).
90. TÉLETCHÉA, S., ESQUE, J., URBAIN, A., ETCHEBEST, C. & DE BREVERN, A. G. Evaluation of Transmembrane Protein Structural Models Using HPMScore. *BioMedInformatics* **3**, 306-326 (2023).
91. THE UNIPROT CONSORTIUM *et al.* UniProt : the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523-D531 (D1 2023).
92. *Charles Darwin - Darwin, Huxley, and Evolutionary Naturalism | Britannica* <https://www.britannica.com/biography/Charles-Darwin/On-the-Origin-of-Species> (2023).
93. *Histoire naturelle | BNF ESSENTIELS* <https://gallica.bnf.fr/essentiels/buffon/histoire-naturelle> (2023).
94. *RKI - Robert Koch* [https://www.rki.de/EN/Content/Institute/History/rk\\_node\\_en.html](https://www.rki.de/EN/Content/Institute/History/rk_node_en.html) (2023).
95. SYSTEM. *Carl Linnaeus* <https://www.nrm.se/en/forskningochsamlingar/botanik/botaniskhistoria/carlvonlinne.480.html> (2023).



---

**Titre :** Étudier, modéliser, moduler le vivant : apport de la bioinformatique structurale

**Mots clés :** Bioinformatique Structurale, Modélisation Moléculaire, Criblage Virtuel, Récepteurs Olfactifs, Galectines, Django.

**Résumé :** Le développement de la vie a été un processus long, parsemé de longues évolutions et de transitions rapides, parfois façonnés par des changements climatiques majeurs. Cette histoire a mené à la complexité actuelle du vivant que nous commençons à comprendre de mieux en mieux. Ainsi, après des siècles d'étude minutieuse du vivant par les naturalistes, les avancées scientifiques et technologiques des deux derniers siècles ont permis un changement de paradigme dans notre compréhension du vivant à l'échelle moléculaire. Je présente dans ce document une présentation rapide de mes travaux de recherche principalement basés sur des approches *in silico* de modélisation

moléculaire. Ces travaux n'ont pu exister que par le travail de nombreux étudiants que j'ai eu le plaisir d'encadrer. J'ai cherché à leur montrer qu'il fallait dans un premier temps modéliser la structure de protéines isolées ou en complexe, puis dans un deuxième temps de rechercher et d'optimiser des petits ligands chimiques capables de bloquer spécifiquement un état d'une protéine. J'ai aussi mis en ligne avec ces étudiants, sous la forme de bases de données et de serveurs web, plusieurs services qui permettent d'accéder à des informations autrement difficiles à rassembler. Je conclus ce mémoire par une présentation de la suite envisagée pour mes recherches.

---

**Title :** Studying, modeling and modulating life : impact of structural bioinformatics

**Keywords :** Structural Bioinformatics, Molecular Modelling, Virtual Screening, Olfactory Receptors, Galectins, Django

**Abstract :** The development of life has been a long process, punctuated by long evolutions and rapid transitions, sometimes shaped by major climatic changes. This history has led to the current complexity of Life, which we are beginning to understand better and better. So, after centuries of meticulous study of living things by naturalists, the scientific and technological advances of the last two centuries have enabled a paradigm shift in our understanding of living things at the molecular level. In this document, I present a brief overview of my research work, based mainly on *in silico* approaches, mainly molecular modeling.

This work has only been possible thanks to the work of the many students I had the pleasure of supervising. I tried to show them that it was necessary firstly to model the structure of isolated proteins or complexes, and secondly to seek out and optimize small chemical ligands capable of specifically blocking a protein state. Together with these students, I have also set up a number of online services in the form of databases and web servers, providing access to data that would otherwise be difficult to gather. I conclude this dissertation with a presentation of the next steps in my research.