Simulations numériques de systèmes biologiques complexes : dynamique, structure et fonction de transporteurs, canaux et enzymes

Marc Baaden

UNIVERSITÉ PARIS VII - DENIS DIDEROT
U.F.R. SCIENCES DU VIVANT

MÉMOIRE

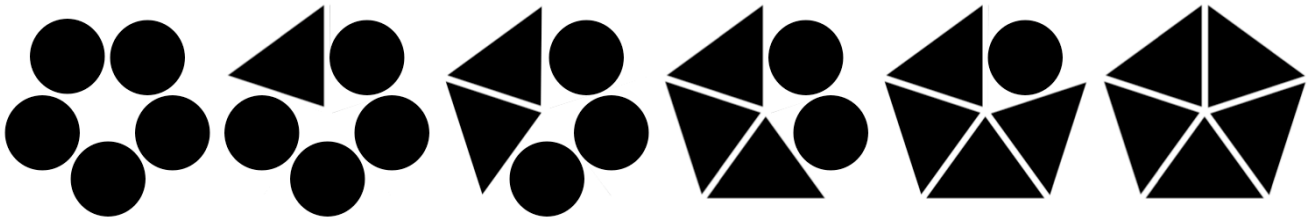
présenté pour obtenir

L'HABILITATION DE DIRIGER DES RECHERCHES DE L'UNIVERSITE DENIS DIDEROT (PARIS 7)

par

Marc BAADEN

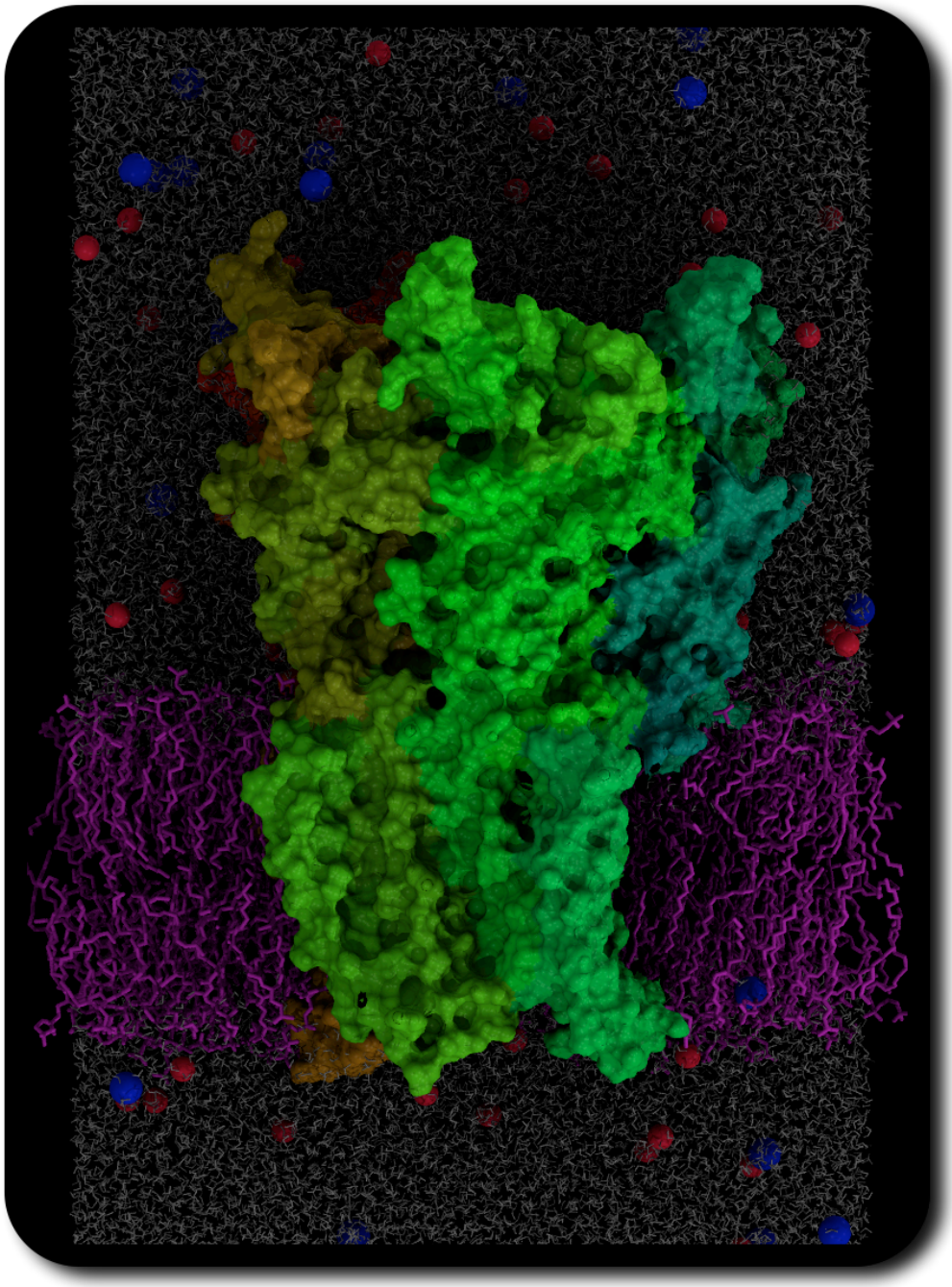
Simulations numériques de systèmes biologiques complexes : dynamique, structure et fonction de transporteurs, canaux et enzymes



Habilitation soutenue le 15 juin 2010

Président du Jury:	Prof. Frédéric DARDEL
Rapporteur:	Prof. Antoine FERREIRA
Rapportrice:	Dr. Anne IMBERTY
Examinateur:	Prof. Alexandre M.J.J. BONVIN
Examinatrice:	Prof. Catherine ETCHEBEST

SIMULATIONS NUMERIQUES DE SYSTEMES BIOLOGIQUES COMPLEXES : DYNAMIQUE, STRUCTURE ET FONCTION DE TRANSPORTEURS, CANAUX ET ENZYMES



Marc BAADEN

Novembre 2009

Illustration de couverture :

Simulation d'une forme ouverte du canal GLIC dans une bicouche lipidique. Il s'agit d'une simulation contrôle qui a permis de corroborer la stabilité de la structure cristallographique dans sa forme ouverte dans un environnement réaliste. Plus de détails sont donnés dans *Nature*, 2009, **457**, 111-114.

Pour ma "Cheche",

pour Mémé et pour mes parents

„ La science la plus profonde
N'est donnée à personne au monde
Par travail, argent, peine ou soins :
La connaissance universelle
En un instant se révèle
A ceux qui la cherchaient le moins.“

La Sorcière à Faust et Méphistophélès

dans Faust I, Johann Wolfgang von Goethe

Curriculum Vitae

Marc Baaden

Laboratoire de Biochimie Théorique, CNRS - UPR 9080
13, rue Pierre et Marie Curie
F-75005 Paris, France

Tél.: +33 1 58 41 51 76 *Mél.:* baaden@ibpc.fr
Fax: +33 1 58 41 50 26 *URL:* http://www.baaden.ibpc.fr

Détails personnels

Célibataire. Né le 23 Mai 1972 à Strasbourg, France.
Double nationalité franco-allemande.

Objectif Professionnel

Etude de phénomènes biologiques et chimiques par des approches théoriques émanants de la chimie physique permettant une compréhension au niveau atomique de ces phénomènes. Ceci tout en interagissant très intimement avec des expérimentateurs et faisant appel à des techniques novatrices telles que la réalité virtuelle.

Formation

Novembre 2000 — **Doctorat Européen en chimie physique / chimie théorique**, co-dirigé par le Prof. Georges Wipff et le Prof. Michel Burgard (Strasbourg) comprenant un séjour de 3 mois dans le groupe du Prof. Wilfred van Gunsteren (Zürich): 'Etudes de molécules extractantes en solution et aux interfaces liquide- liquide: aspects structuraux et mécanistiques des effets de synergie.'
Mention Très Honorable avec les Félicitations écrites du Jury.

1997 - 2000 — Université Louis Pasteur (ULP), Strasbourg, France et Ecole Polytechnique (ETH), Zürich, Suisse

Septembre 1997 — **DEA en chimie physique**, option spectroscopie, sous la direction du Prof. Pierre Granger et du Prof. Alain Strich: 'Calculs *ab initio* de tenseurs d'écran en RMN.'
Mention Bien.

1996 - 1997 — Université Louis Pasteur (ULP), Strasbourg, France

Juin 1997 — **Diplôme d'Ingénieur Chimiste EHICS**, Cursus tri-lingue (Français, Anglais, Allemand) en chimie, option physico-chimie des matériaux.
Major de la promotion.

1994 - 1997 — Ecole européenne des Hautes études de l'Industrie Chimique de Strasbourg (EHICS), Strasbourg, France

Avril 1994 — **'Diplom-Vorprüfung' en chimie**
Mention Bien.

1992 - 1994 — Université Technique de Karlsruhe, Allemagne

Juin 1991 — **'Abitur'** , suivant un cursus bilingue (Allemand/Français).
Mention Très Bien.

1982 - 1991 — Lycée Einstein de Kehl, Allemagne

Recherche

Depuis 2007 — **Chargé de Recherche de 1ère classe** • *chez le Prof. Ph. Derreumaux au CNRS Paris*
2003 - 2006 (4 ans) — **Chargé de Recherche de 2ème classe** • *chez le Dr R. Lavery au CNRS Paris*
Etudes de protéines membranaires, d'enzymes et de leur fonction biologique. Développements méthodologiques et applications dans le graphisme moléculaire interactif et en réalité virtuelle.
Résultats publiés dans Nature, Biophys. J., J. Comput. Chem., ChemPhysChem etc.

2001 - 2003 (2 ans) — **EC Research Fellow** • *chez le Prof. M.S.P. Sansom à l'Université d'Oxford (GB)*

	Simulations numériques de protéines membranaires bactériennes *publiées dans JMB, Biochemistry, Biophys. J., etc.*
1997 - 2000 (3 ans)	**Doctorant** • *chez le Prof. G. Wipff et le Prof. M. Burgard, France* Simulations de dynamique moléculaire de systèmes modèles (calixarènes, TBP, HNO_3) élucidant des aspects microscopiques de la reconnaissance moléculaire en solution, complexation et extraction liquide-liquide en milieu neutre *vs* acide. - *Résultats publiés dans JPCA, JPCB, Perkin. Trans., Inorg. Chem., J. Mol. Liquids, PCCP, etc.*
2000 (3 mois)	**Stage de recherche** • *chez le Prof. W. F. van Gunsteren, Suisse* Aspects énergétiques du transfert d'un soluté à travers l'interface eau/chloroforme. Calcul du potentiel de force moyenne par plusieurs approches.
1997 (6 mois)	**DEA** • *chez le Prof. P. Granger et le Prof. A. Strich, Strasbourg, France* Calcul *ab initio* de tenseurs d'écran en RMN (méthode GIAO) : étude de la rotation intramoléculaire dans une série de dérivés vinyliques. Détermination de la structure de fragments de la mélanine par analyse de déplacements chimiques calculés et comparaison avec l'expérience. - *Résultats publiés dans Mol. Phys.*
Publications	31 publications, 22 communications par affiches, 31 présentations orales, 2 logiciels distribués.
Encadrement	11 étudiants en stage de recherche, 3 post-doctorants, 2 doctorants, 2 ingénieurs.
Organisation	1 conférence internationale, 1 école d'été internationale, 1 journée porte ouverte
Relecture d'articles	JACS, Proteins, Biophys J, JPOC, JBSD, JMMG, PCCP et autres (environ 10 par an)
Bourses	Bourse ERASMUS de l'Union Européenne (1994) Allocation de Recherche du MNERT (26 kEuro, 1998-2000) Bourse Marie Curie de l'Union Européenne (114 kEuro, 2001-2003)
Contrats	*FonFlon*, 2006 à 2009, 153 kEuro, Fluctuations structurales dans les enzymes *FlowVRNano*, 2008 à 2010, 223 kEuro, Réalité virtuelle et simulations moléculaires

Contributions scientifiques majeures

Je suis reconnu pour mon expertise des protéines membranaires. J'ai contribué à l'élucidation d'une structure ouverte d'un canal ionique proche du récepteur nicotinique humain.

Enseignement

2003-2009	**Intervenant** • *principalement Université Denis Diderot, Paris 7 (~40 h p.a.)*
1998-99	**Vacataire puis ATER** • *Université Louis Pasteur, Strasbourg (196 h)*
1997/98	**Vacataire** • *Ecole de Chimie, Polymères et Matériaux, Strasbourg (104 h)*

Autres activités académiques

- Co-organisateur de l'école "BioImage", Paris, Juillet 2005, - http://www.ecole2005.ibpc.fr
- Co-organisateur de la conférence "Biomolecular Simulations", Bordeaux, Septembre 2005
- Secrétaire Scientifique de la section 13 du Comité National du CNRS
- Membre du comité utilisateurs du centre de calcul national IDRIS
- Co-auteur du livre "Itinéraires Bis" à paraître chez Connaissances et Savoirs début 2010

Distinctions et Affiliations

- Prix d'excellence de l'EHICS, Strasbourg, France en 1997
- Qualification aux fonctions de Maître de conférences (section 31) en 2001
- Prix de thèse du Conseil Scientifique de l'ULP, Strasbourg, France en 2001
- Prix jeune chercheur du Groupe de Graphisme et Modélisation Moléculaire en 2003
- Membre de la 'Gesellschaft deutscher Chemiker' et de la Société Française de Chimie
- Membre de la 'Biophysical Society'

Langues

Français, Allemand - **bilingue**.

Anglais - courant (**945 points au TOEIC, 23 ans de pratique**).
Japonais - connaissances de base (cours d'initiation de 4 mois **au Japon**).

Micro-Informatique

Administration système	Linux, Mac OSX, Unix, Windows.
Réseaux	TCP/IP (ftp, http, mail), Microsoft, Appletalk.
Programmation	Objective-C/C, Fortran, Perl, Python et autres
Modélisation Moléculaire	Gromacs, Yasara, NAMD, Amber, Gromos, Gaussian, Spartan, Macromodel.
Autre	What If, Modeller, VMD, VTK, R, XML, Docbook, LaTeX.
Bases de données	PDB, CCDB, Beilstein, bibliographiques (PubMed, ISI, CC, SCI, ...).
Internet	Création de sites de l'université (ULP), d'un club de sport et personnelles/professionnelles.

Renseignements complémentaires

Centre d'intérêt	Aïkido et Arts Martiaux (24 ans de pratique, ceinture noire).
	Musique (6 ans de piano).
Voyages	Europe (A, B, CH, D, E, F, GB, GR, H, I, NL, PL, TR), Afrique (MA), Asie (Japon) et états-unis.

Collaborations mises en œuvre

- Dr. David Bensimon	ENS Paris
- Dr. Peter Bond	Institut Max-Planck, Francfort
- Prof. Paolo Carloni	Sissa/Isas, Trieste, Italie
- Dr. Jean-Pierre Changeux	Collège de France Paris
- Dr. Pierre-Jean Corringer	Institut Pasteur Paris
- Dr. Marc Delarue	Institut Pasteur Paris
- Dr. Jonathan Essex	Université de Southampton
- Dr. Dirk Fasshauer	Institut Max-Planck, Goettingen
- Gilles Grasseau	IDRIS Orsay
- Dr. Brigitte Hartman	INTS Paris
- Dr. Ludger Johannes	Institut Curie Paris
- Prof. Ludovic Jullien	ENS Paris
- Prof. Anass Kettani	Faculté des Sciences Ben Msik, Casablanca, Maroc
- Dr. Syma Khalid	Université de Southampton
- Dr. Richard Lavery	IBCP Lyon
- Dr. Julie Ménetrey	Insitut Curie Paris
- Jean-Philippe Nominé	CEA Bruyères
- Dr. David Perahia	Université Paris 11, Orsay
- Dr. Chantal Prévost	IBPC Paris
- Dr. Bruno Raffin	INRIA Grenoble
- Dr. Sophie Robert	LIFO, Université d'Orléans
- Dr. Sophie Sacquin-Mora	IBPC Paris
- Prof. Mark S.P. Sansom	Université d'Oxford

Liste des Abbréviations

ABC *ATP-binding cassette*

ADN Acide désoxyribonucléique

ANR Agence Nationale de la Recherche

ATER Attaché temporaire d'enseignement et de recherche

ATP Adénosine triphosphate

BtuB *Outer membrane cobalamin transporter*

Cα Carbone alpha

CEA Commissariat à l'Énergie Atomique

DEA Diplôme d'études approfondies

DEISA Distributed European Infrastructure for Supercomputing Applications

DM Dynamique moléculaire

DNase Désoxyribonucléase

DSIMB Dynamique des Structures et Interactions des Macromolécules Biologiques

ELL Extraction liquide-liquide

FecA *Ferric Citrate Receptor FecA*

FepA *Ferric enterobactin receptor FepA*

FhuA *Ferric hydroxamate uptake receptor FhuA*

FonFlon Acronyme de projet ANR reliant fonction et fluctuations

FVNano Acronyme de projet ANR FlowVR Nano – un laboratoire virtuel

GGMM Groupe de Graphisme et Modélisation Moléculaire

GK Guanylate Kinase

GLIC *Gloeobacter violaceus ion channel*

GNM *Gaussian Network Model*

GPU *Graphics Processing Unit*

IBPC Institut de Biologie Physico-Chimique

IDRIS Institut du développement et des ressources en informatique scientifique

INRIA Institut national de recherche en informatique et automatique

INTS Institut National de la Transfusion Sanguine

LBT Laboratoire de Biochimie Théorique

LIFO Laboratoire d'Informatique Fondamentale d'Orléans

MDDriver *Molecular Dynamics Driver library*

MOAIS Multi-programmation et Ordonnancement sur ressources distribuées pour les Applications Interactives de Simulation

MsbA *MsbA ABC transporter*

OmpT *Outer membrane protease T*

OmplA *Outer membrane phospholipase A*

pH Potentiel hydrogène

PDB *Protein DataBank*

POPC Palmitoyloleoylphosphatidylcholine

POPS Palmitoyloleoylphosphatidylserine

rLDL *Low-Density Lipoprotein (LDL) Receptor*

RMN Résonance magnétique nucléaire

RPE Résonance paramagnétique électronique

RV Réalité virtuelle

RX Diffraction des Rayons X

SNARE *Soluble N-ethyl maleimide sensitive-factor Attachment protein REceptors*

SUR *Sulfonyl urea receptor*

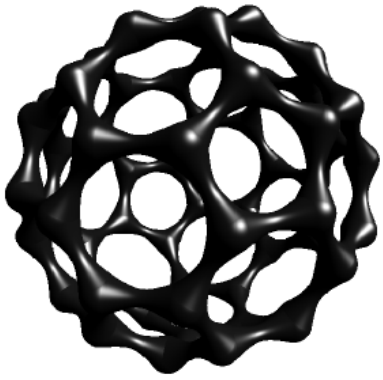
TBP tri-n-butyl phosphate

TMD *Targeted Molecular Dynamics*

VMD *Visual Molecular Dynamics*

VTK *Visualization Toolkit*

Table des matières



Remerciements · *Acknowledgements*

Je tiens à remercier ...

Le Professeur Philippe DERREUMAUX pour la confiance qu'il m'a témoignée, le soutien qu'il a accordé à mes projets de recherche au sein du laboratoire « LBT », et les agréables discussions que nous avons eues autour de boissons fermentées à base d'orge et de houblon.

Le Docteur Richard LAVERY pour m'avoir accueilli dans son laboratoire en 2003, pour la confiance et la liberté qu'il m'a accordée et pour la collaboration fructueuse que nous avons eue.

Le Professeur Cathérine ETCHEBEST pour ses nombreux conseils tout au long de ces dernières années, notamment en matière d'encadrement, et pour avoir participé au jury de soutenance de cette HDR.

Le Docteur Anne IMBERTY et le Professeur Antoine FERREIRA qui ont accepté de juger mon travail.

Les Professeurs Alexandre M. J. J. BONVIN et Frédéric DARDEL qui m'ont fait l'honneur de participer au jury de l'habilitation.

Thank you ...

Professor Mark S. P. SANSOM for introducing me to the wonderful world of membrane proteins, for your enthusiasm, support and encouragement, as well as your remarkable ability to share and impart your knowledge.

... et aussi ...

Mes collaborateurs proches de ces dernières années. Tout a commencé en 2003 avec les premières tentatives sur la réalité virtuelle, et dans un ordre à peu près chronologique je remercie Emeline GASSER, Jean-Philippe NOMINE et Christophe ARBEZ, Gilles GRASSEAU, Sophie ROBERT, Sébastien LIMET, Ahmed TURKI, Bruno RAFFIN, Antoine VANEL, Nicolas FEREY, Benoist LAURENT, Matthieu CHAVENT et Zhihan LU. Le deuxième projet démarré en 2004 concerne la fusion membranaire avec Marie-Pierre DURRIEU, Peter BOND, puis depuis 2009 Alex TEK et à l'autre bout du monde Sergio PANTANO et Leonardo DARRE. Les échanges avec Brigitte HARTMANN, Brahim HEDDI et Joséphine ABI-GHANEM ont accompagné de manière fort utile mes débuts d'encadrant. Et puis je remercie aussi Brigitte pour les nombreuses conversations plus ou moins scientifiques que nous avons eues et pour son franc parler. Le projet « FonFlon » m'a permis de collaborer avec David BENSIMON, Vincent CROQUETTE, Francesco MOSCONI, Olivier DELALANDE et Sophie SACQUIN-MORA. Rachid CHATER, Anass KETTANI, David PERAHIA, Marc GUEROULT, Pierre POULAIN, Christopher

AMOURDA, Adrien SALADIN, Chantal PREVOST, Christophe OGUEY, Jeremy ESQUE, Maria-Vittoria CUBELLIS, Dominique COSTA, Pierre-Alain GARRAIN, Pierre DEHLINGER, Daniel PICOT, Laurent CATOIRE et Danny PARTON ont éclairé d'autres collaborations fructueuses. Depuis 2008, un autre chapitre s'est ouvert autour des protéines bactériennes proches du récepteur nicotinique avec Marc DELARUE, Hugues NURY, Frédéric POITEVIN, Pierre-Jean CORRINGER et Jean-Pierre CHANGEUX. Depuis peu, les laboratoires Servier participent également à cette thématique. -- Je les remercie tous.

Toute l'équipe du laboratoire LBT, que ce soit dans le passé ou au présent. En particulier Dr. Emmanuel LEVY, Dr. Isabelle NAVIZET, Dr. Fabien CAILLEZ, Dr. Guillaume PAILLARD, Dr. Cyril DEREMBLE, Nicolas GUIOT, Dr. Jean-Christophe GELLY, Dr. Yasmine CHEBARO et Mohamed KHABZAOUI. Et enfin, Daniel, Alexey, Krystyna, Thérèse, Youri, Emmanuel G., Karine, Dragana, Peter, Cyril R., Antoine, Younès, David, Flore, Claire, Samuela, Charles, Mainak et les derniers arrivés.

Sans oublier tous ceux qui m'ont permis de participer à l'enseignement des étudiants: Prof. Sophie CRIBIER, Dr. Cécile BREYTON, Prof. Carlo ADAMO, Prof. Laurent JOUBERT, Dr. Ilaria CIOFINI et d'autres ...

Plus récemment, au sein de la section 13 du CoNRS, une autre vue sur la recherche scientifique française s'est offerte à moi, et je remercie beaucoup mes collègues pour leur investissement et l'exemple qu'ils donnent, en particulier Philippe et Odile.

Un paragraphe aussi pour tous les "Itis" qui ont contribué à l'aventure du livre "Itinéraires Bis", que ce soit en tant qu'auteur, participant à la formation ou organisateur de cette merveilleuse manifestation. La liste serait trop longue pour les nommer explicitement, mais je remercie tout un chacun pour avoir partagé des moments forts entre "jeunes chercheurs".

Et je n'oublie pas le Dr. Jean-Pierre HENRY et l'expérience BioImage, Germain TRUGNAN pour son écoute, Marc ROMERO et Laurent CHAMBON pour leur soutien en informatique, Michèle et ses croques, et surtout Isabelle LEPINE pour tout ce qui touche la vie de notre laboratoire et les besoins quotidiens de ses chercheurs.

Pour la préparation de cette habilitation, l'aide du Professeur Mireille VIGUIER, de Béatrice TREGUIER et de Didier PENELOPE à l'université Paris 7 Denis Diderot m'a été très précieuse.

Mes excuses pour ceux que j'aurai oublié de nommer, mais j'excelle dans le stéréotype du chercheur distrait, et je suis sûr que demain je m'en voudrai de ne pas avoir pensé à les remercier explicitement :)

Et le plus grand merci pour la fin: Merci Gisèle.

Chapitre 1

Avant-propos

"*Il est encore plus facile*
de juger de l'esprit d'un homme
par ses questions
que par ses réponses."

Duc de Lévis

1 Présentation

Le présent mémoire est rédigé en vue de l'obtention de l'habilitation à diriger des recherches et retrace mon parcours en détaillant les jalons de mes travaux de recherche. Ce premier chapitre donne des informations complémentaires sur mon parcours, ma productivité scientifique, mes activités administratives, les financements, l'encadrement, l'enseignement et la diffusion de l'information scientifique et technique.

2 Parcours

Physico-chimiste de formation, j'ai utilisé la modélisation moléculaire afin d'élucider des effets de synergie dans l'extraction liquide-liquide de cations métalliques pendant ma thèse. L'université de Strasbourg m'a décerné un prix de thèse pour ce travail. Un stage de trois mois dans le groupe de WILFRED VAN GUNSTEREN à l'ETH Zurich m'a permis d'élargir mes connaissances dans les méthodes de simulation. Mon prochain défi consistait dans la transition de la chimie physique traditionnelle aux systèmes biologiques. En tant que *EC*

research fellow avec le Prof. MARK SANSOM à l'université d'Oxford j'ai acquis une expertise dans la biophysique et étudié des complexes de protéines membranaires par simulation numérique. En 2003, le Groupe de Graphisme et Modélisation Moléculaire (GGMM) m'a attribué son prix jeune chercheur pour ce travail. Simultanément, j'ai été recruté comme chargé de recherche au CNRS dans le laboratoire de Biochimie Théorique (LBT) à l'Institut de Biologie Physico-Chimique (IBPC) à Paris. Au LBT, j'ai continué mes travaux sur les assemblages biomoléculaires complexes et persisté dans le développement de mes propres thématiques de recherche. Les fruits de cet effort sont une reconnaissance comme expert dans la simulation de protéines membranaires et une activité émergente dans l'utilisation d'approches de la réalité virtuelle dans de telles simulations.

3 Production scientifique

En somme, mes travaux de recherche ont donné lieu à 23 publications dans des revues à comité de lecture (rang A), 3 chapitres de livres et 5 actes de colloques. Ma première publication était une étape importante et formatrice pour moi. Elle concernait mon stage de DEA et j'étais entièrement en charge de la conception et de l'écriture du manuscrit. Pendant la thèse, j'ai publié 10 articles, dont 8 comme premier auteur. Après la thèse, j'ai continué à publier la majorité de mes travaux comme premier auteur. Avec la transition au CNRS, j'ai gagné en indépendance et je signe désormais mes publications principales comme dernier auteur.

4 Administration et évaluation de la recherche

Depuis 2004, je suis membre du comité utilisateur du centre de calcul national IDRIS. Je représente notre laboratoire auprès du Conseil de l'Institut (FRC 550 de l'IBPC). Au laboratoire, j'occupe la fonction de responsable scientifique informatique. En 2008, j'ai été élu au Comité National du CNRS en tant que secrétaire scientifique de la section 13. Je suis porteur et coordinateur de deux contrats ANR, le projet FonFlon de 2006 à 2010 et le projet FlowVRNano de 2008 à 2011. J'ai participé à 3 jurys de thèse. Je suis régulièrement sollicité pour la relecture d'articles en chimie physique, biophysique et réalité virtuelle. Je suis plusieurs fois intervenu comme animateur (*chairman*) à des manifestations internationales.

5 Encadrement et enseignement

J'ai encadré 11 étudiants lors de stages de recherche, 3 post-doctorants dans le cadre des contrats ANR FonFlon et FlowVRNano et 3 ingénieurs dans leur fonction au laboratoire LBT. Le premier doctorant que j'ai co-encadré a soutenu sa thèse en 2008. Une deuxième

thèse sous ma direction vient de démarrer. Tout au long de ma carrière, j'ai également saisi les opportunités pour participer à l'enseignement. D'abord comme vacataire, ensuite comme ATER et dernièrement comme intervenant à l'université Paris 7. J'ai notamment développé des travaux pratiques d'initiation à la simulation biomoléculaire qui ont connu un certain succès, même au-delà de mes enseignements.

6 Diffusion de l'information scientifique et technique

La dissémination de mes travaux de recherche a donné lieu à 22 communications par affiches, 31 présentations orales, dont 3 invitations au niveau international, et 2 logiciels distribués. J'ai participé à l'organisation de deux évènements internationaux, l'école « BioImage » qui a eu lieu à Paris en 2005 et la conférence « *Biomolecular Simulations* » à Bordeaux, en 2005 également. L'année prochaine, j'organiserai une session sur la réalité virtuelle lors de la conférence de biophysique à Shanghai. En 2011, j'organiserai les prochaines journées du GGMM.

Plusieurs autres opérations sont prévues dans le cadre du projet FlowVRNano qui vise à structurer une communauté autour des simulations interactives. Une retombée concrète est l'installation d'une salle dédiée (*showroom*) au sein de notre laboratoire. Plusieurs sites web que je maintiens viennent compléter le dispositif de dissémination.

Chapitre 2

Introduction et contexte

« Rien n'est indifférent,
rien n'est impuissant dans l'univers ;
un atome peut tout dissoudre,
un atome peut tout sauver ! »

Gérard de Nerval
Aurélia

1 Axes de recherche — grandes lignes

La motivation première de mes travaux de recherche est de combiner des approches expérimentales et théoriques dans le domaine de la chimie physique pour atteindre une meilleure compréhension des phénomènes à l'échelle atomique. Mes travaux en cours traitent de systèmes d'intérêt biologique concernant les processus membranaires et des phénomènes accessibles par des méthodes de nanomanipulation.

Les problèmes de la biophysique et biochimie sont au cœur de mes recherches. J'ai effectué des simulations complexes de protéines membranaires dans une bicouche lipidique qui se sont montrées tout à fait complémentaires et révélatrices par rapport aux études expérimentales de biologie structurale. Une récente collaboration exploitant cette complémentarité a donné lieu à une publication dans la revue *Nature* en début 2009. [1]

Les travaux récents visent à développer des approches combinant la réalité virtuelle avec les simulations moléculaires. [2] Les systèmes biologiques étudiés présentent à la fois un intérêt physico-chimique, biologique et médical et peuvent atteindre un grand nombre

d'atomes. En parallèle, je mène un travail de fond sur les méthodes de simulation et des approches novatrices.

2 L'interdisciplinarité au cœur de mon projet scientifique

Mes travaux ont un caractère interdisciplinaire allant de la chimie physique via la biochimie, la biophysique ou encore la biologie structurale jusqu'à l'informatique appliquée. L'approche est basée sur les principes de la chimie physique et s'applique aux systèmes d'intérêt biologique. Avec l'utilisation des techniques de la réalité virtuelle et le recours au calcul intensif parallèle, un lien organique avec l'informatique existe.

Ces fondamentaux sont présents dans l'analyse théorique et la construction de modèles, dans l'étude des assemblages moléculaires, de leurs structures, des propriétés dynamiques et thermodynamiques. Un travail de fond concernant les simulations numériques porte sur des aspects d'échantillonnage, le contrôle de qualité, l'importance de la structure de départ et le protocole de simulation. Ce travail est illustré avec des exemples de systèmes biologiques, mais est pertinent pour toute simulation de systèmes moléculaires.

3 Jalons de mes recherches au cours des quatre dernières années

La simulation du système SNARE inséré dans une double bicouche POPC/POPS chargée en détail atomique (340000 atomes, 40 ns) est une percée importante. Ce travail a récemment été présenté au niveau national et international, une première publication sur la mise en œuvre du calcul a vu le jour et plusieurs articles de vulgarisation sont parus. [3] Je suis le porteur de cette thématique au sein du laboratoire.

Une simulation d'une microseconde a pu être réalisée sur l'enzyme OmpT avec un modèle hybride gros-grain/tout atome. [4] D'autres travaux en cours laissent soupçonner un nouveau mécanisme catalytique pour ce système. L'obtention d'un financement de l'ANR dont je suis porteur m'a permis de pérenniser ces recherches.

Une collaboration intense et fructueuse avec deux équipes à l'institut Pasteur s'est engagée autour du récepteur GLIC. Il subsistait un doute sur le fait que la structure cristallographique de GLIC soit ouverte ou fermée. En combinant les simulations et l'expérience, nous avons pu confirmer et publier la première structure d'un tel récepteur dans un état supposé ouvert. [1] J'ai ensuite réalisé une simulation d'une microseconde représentant la fermeture (*gating*) du canal GLIC, provoqué par un changement de pH. Ce dernier travail est en cours de rédaction.

Dans l'axe touchant à la conception de logiciels, le projet de réalité virtuelle FVNano a vu le jour (voir figure ci-dessous). Une première retombée concrète est la librairie MDDriver permettant des simulations de dynamique moléculaire interactives. [2,5] Un deuxième financement de l'ANR dont je suis le porteur et une collaboration avec le centre de calcul national IDRIS ont ouvert la voie à des percées importantes. Je considère cet axe comme particulièrement innovant et déploie beaucoup d'efforts pour établir une position de première importance dans ce domaine.

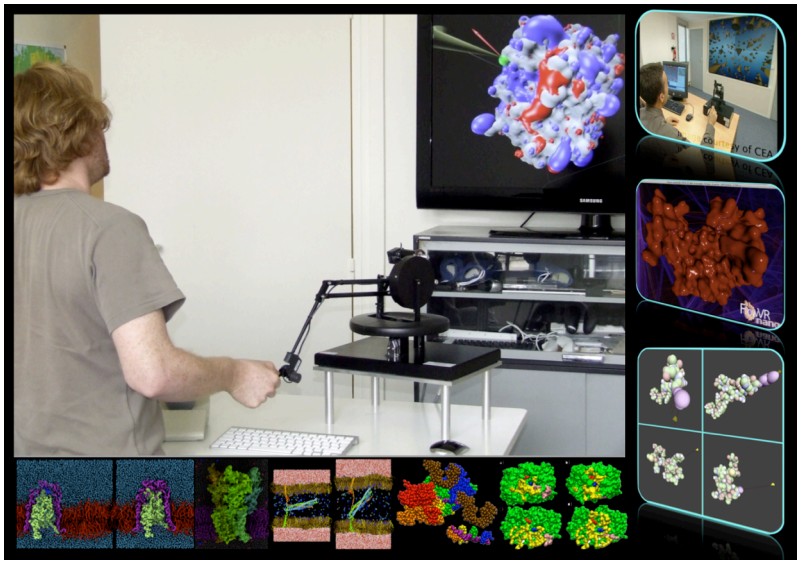


Figure 2.1: Le laboratoire « FVNano » en action. La partie centrale de l'image illustre une séance de travail typique. En bas de l'image quelques exemples de systèmes moléculaires étudiés. A droite plusieurs captures d'écran.

Références bibliographiques du chapitre 2

[1] N. BOCQUET, H. NURY, M. BAADEN, C. LE POUPON, J.P. CHANGEUX, M. DELARUE et P.J. CORRINGER: "*X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation*", 2009, **Nature**, 457, 111-114.

[2] O. DELALANDE, N. FEREY, G. GRASSEAU et M. BAADEN : "*Complex Molecular Assemblies at hand via Interactive Simulations*", 2009, **J. Comput. Chem.**, 30, 2009, 2375-2387.

[3] a) E. KRIEGER, L. LEGER, M.P. DURRIEU, N. TAIB, P. BOND, M. LAGUERRE, R. LAVERY, M.S.P. SANSOM et M. BAADEN : "*Atomistic modeling of the membrane-embedded synaptic fusion complex: a grand challenge project on the DEISA HPC infrastructure*", 2007, dans **ParCo 2007, Parallel Computing: Architectures, Algorithms and Applications**, édité par C.B.G.R. JOUBERT, F. PETERS, T. LIPPERT, M. BUECKER, B. GIBBON, et B. MOHR, Vol. 38, pp. 729-736, John von Neumann Institute for Computing, Juelich, Allemagne ; b) CSC News 3, 2007, 10-11; 12. Tietoyhteys 3, 2007, 10-11; 12-13. DEISA Newsletter 5, 2007. DEISA Newsletter 5, 2008. iSGTW 1/2008; iSGTW 10/2008. CORDIS News 11/2007. Supercomputing online 11/2007 etc.

[4] M. NERI, M. BAADEN, V. CARNEVALE, C. ANSELMI, A. MARITAN et P. CARLONI: "*Microseconds Dynamics Simulations of the Outer-Membrane Protease T*", 2008, **Biophys.J.** 94, 71-78.

[5] http://www.baaden.ibpc.fr/projects/mddriver/

Chapitre 3

Activité scientifique antérieure - doctorat et stage postdoctoral

"*La Science constate,*
mais n'explique pas :
c'est la fille aînée des chimères."

L'Isle-Adam

1 Grandes lignes

Mon tout premier travail de recherche fut réalisé en DEA menant à une publication sur le calcul *ab initio* de tenseurs d'écran en RMN. [6] La thèse qui suivit portait sur **l'extraction liquide-liquide**. Lors de cette thèse, dont l'orientation était majoritairement théorique, j'ai pu vérifier les résultats de mes simulations par une étude par RMN en solution et en cristallisant un complexe clé. Les résultats expérimentaux sont en parfait accord avec les structures et phénomènes prédits. [7] Les systèmes étudiés dans ce contexte émanent de la chimie supramoléculaire et se situent à l'interface de la chimie et de la biologie. Les « expériences sur ordinateur » permettent de mieux appréhender au niveau atomique les bases des phénomènes étudiés tels la reconnaissance et les assemblages moléculaires.

Le stage post-doctoral m'a permis d'étendre mes connaissances acquises en modélisation et de les appliquer aux problèmes de la biophysique et biochimie, des domaines qui sont au cœur de mes projets de recherche actuels. J'ai effectué des simulations complexes de **protéines membranaires** dans une bicouche lipidique qui se sont montrées

tout à fait complémentaires et révélatrices par rapport aux études expérimentales de biologie structurale.

2 Extraction liquide-liquide

2.1 Introduction

Ces travaux présentés dans la thèse de doctorat démontrent l'utilité des simulations de dynamique moléculaire (DM) pour l'étude de la reconnaissance moléculaire en solution, de la complexation et de l'extraction liquide-liquide (ELL). L'ELL et le transport des ions métalliques à travers une interface eau/« huile » sont des processus complexes et encore incompris au niveau microscopique. Rétrospectivement, il faut ajouter que de nombreuses ressemblances existent avec des systèmes biologiques, notamment les canaux ioniques que j'étudie actuellement.

2.2 Etudes de molécules extractantes en solution et aux interfaces liquide-liquide: aspects structuraux et mécanistiques des effets de synergie

Des études préliminaires concernant la représentation adéquate des cations trivalents terres rares La^{3+}, Eu^{3+} et Yb^{3+} dans les simulations de DM classiques ont révélé une mobilité des molécules de solvant inattendue. Dans la première sphère de solvatation en solution dans l'acétonitrile, le solvant s'échange plus rapidement avec Yb^{3+} qu'avec La^{3+}. [8] Ensuite, la complexation de ces cations par un calixarène développé récemment a été étudiée. Les calixarènes sont des molécules macrocycliques qui permettent de « cibler » la sélectivité par des effets d'adéquation entre la taille de l'ion métallique d'une part et la topologie, taille et basicité de la molécule extractante d'autre part comme c'est également le cas dans les filtres de sélectivité de certaines protéines transportant des ions. Ces investigations ont montré que les cations ne sont pas isolés du solvant, mais directement coordinés par quelques molécules d'eau ce qui implique que seulement 4 des 8 atomes donneurs potentiels du calixarène peuvent se coordiner au cation. [9]

La sélectivité peut également être modifiée et améliorée par la mise en jeu dans les systèmes d'extraction de deux (ou plusieurs) molécules extractantes. Ce phénomène appelé synergie est abordé dans un autre volet des travaux qui concerne des systèmes d'extraction liquide-liquide industriels impliquant le tri-n-butyl phosphate (TBP) comme co-solvant, extractant, surfactant et agent de synergie. J'ai examiné des effets de concentration, d'acidité de la phase aqueuse et des aspects de synergie dans des systèmes d'extraction mixtes TBP/calixarène. [10] Ces simulations ont apporté les premières vues microscopiques de tels

phénomènes et ont montré que le TBP possède certaines caractéristiques d'un lipide. Il a été observé la formation d'une troisième phase entre l'eau et l'huile, voire l'agrégation de micelles.

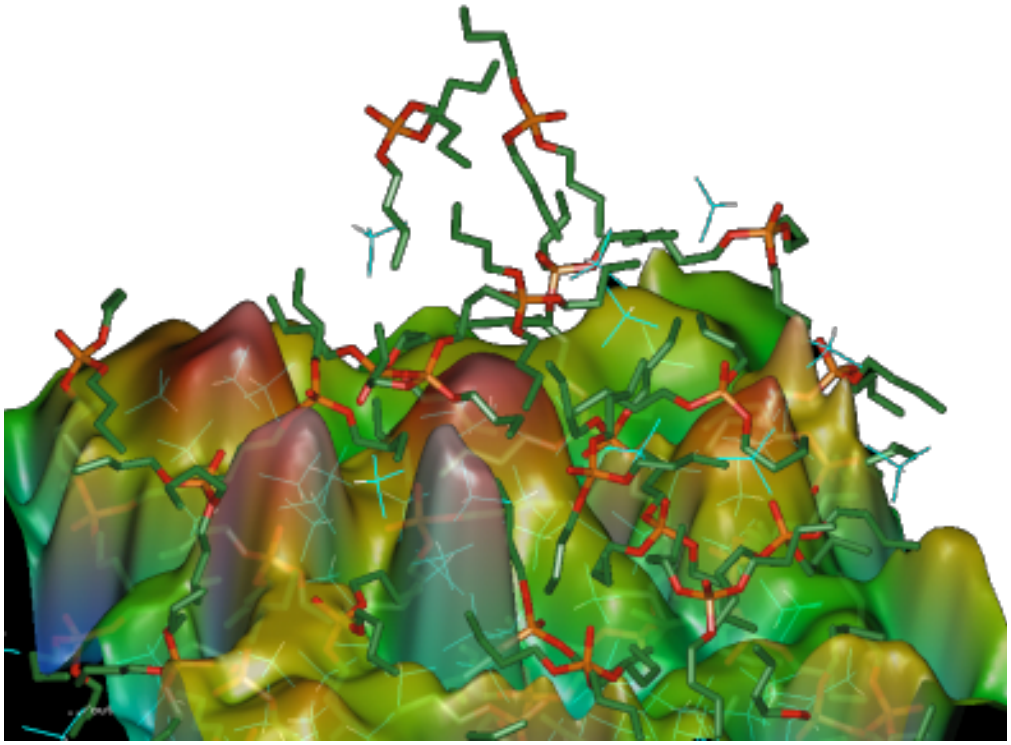


Figure 3.1. Représentation de l'interface eau/chloroforme en présence de TBP.

Enfin, j'ai simulé le transfert d'un soluté à travers l'interface eau/chloroforme. Le potentiel de force moyenne d'un tel processus a été calculé et par des méthodes standards dites « umbrella sampling » et par des approches nouvelles où j'ai exploité les mêmes données pour calculer la force nécessaire à retenir le système près d'un état donné. Une méthode similaire a permis d'élucider le mécanisme de transport du K+ dans le canal ionique KcsA. [11]

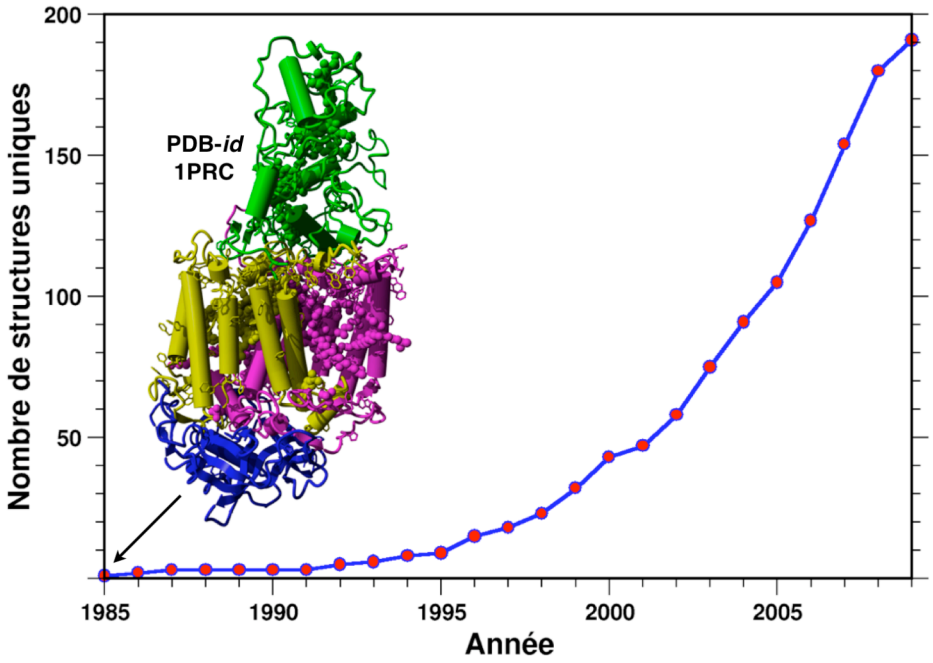


Figure 3.2. Evolution temporelle du nombre de structures uniques de protéines membranaires. La première protéine membranaire cristallisée en 1985 est montrée. [12] *Figure adaptée de* [13].

3 Protéines membranaires

3.1 Introduction

Les protéines membranaires sont de première importance en biologie cellulaire et interviennent par exemple comme canaux ioniques, récepteurs de médicaments et transporteurs de solutés. En effet, il a été estimé que 30 % des gènes encodent des protéines membranaires, et que celles-ci représenteront la moitié des cibles potentielles pour le développement de nouveaux médicaments. [14] Néanmoins, il n'existe que très peu d'information structurale sur ces protéines (voir Figure 3.2 ci-dessus). A ce jour on connait les structures atomiques d'environ 191 protéines membranaires alors qu'il en existe des milliers.

3.2 Protéines membranaires à repliement « tonneau β » - modélisation de l'enzyme OmplA

Un de mes axes de recherche concerne ces protéines membranaires sous forme de tonneaux β présents dans la membrane externe des bactéries gram négatives qui sont encore peu étudiés par rapport aux assemblages d'hélices α plus communément observés pour les autres protéines membranaires.

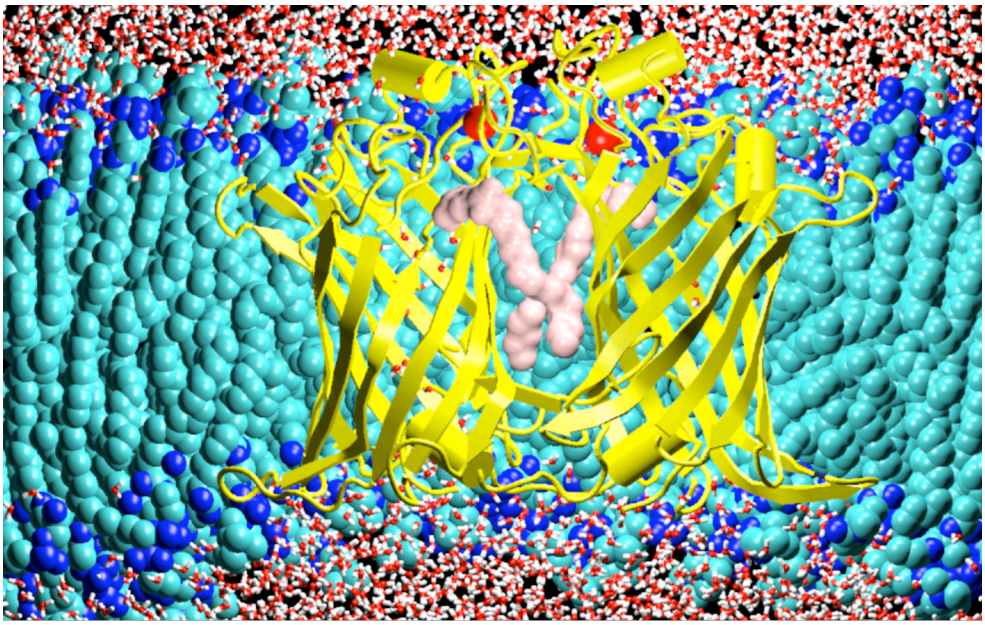


Figure 2.2. La protéine dimérique OMPLA dans une bicouche lipidique hydratée.

L'enzyme OmplA présente une telle architecture de tonneaux β et est d'une taille d'environ 33 kDa. Les fonctions exercées par cette protéine dans la membrane externe comprennent la stabilisation de la structure membranaire et l'activité enzymatique. Les structures de plusieurs formes fonctionnelles de cette enzyme ont été déterminées par cristallographie, mais n'ont pas permis d'élucider les relations entre structure et fonction. Pour obtenir des informations complémentaires, j'ai effectué des simulations de dynamique moléculaire dans une bicouche lipidique pour trois formes de cette enzyme: le monomère, inactif, le dimère sans substrat et le dimère en présence d'un inhibiteur. J'ai montré que la présence de molécules d'eau bien spécifiques, d'ions calcium et la formation d'un réseau de liaisons hydrogène autour de ceux-ci jouent un rôle important dans le cycle enzymatique. De même, la dynamique conformationnelle de la poche accueillant le substrat semble déterminante pour réguler l'activité de l'enzyme: en absence de substrat, la poche se ferme et empêche l'accès au site actif. Ce travail interdisciplinaire a été publié dans un journal de biologie moléculaire [15] et a été primé par le groupe de graphisme et modélisation moléculaire (GGMM) qui m'a attribué son prix de jeune chercheur lors de son congrès biannuel 2003 à Cabourg.

3.3 Modélisation et étude de transporteurs ABC: les espèces MsbA et SUR

Les protéines de la famille dite ABC (*ATP-Binding cassette*) se chargent du contrôle de l'osmorégulation et de la détoxication des xénobiotiques tels que les métaux lourds, les agents pathogènes ou certains médicaments. Leur importance clinique et biologique est liée à

leur capacité de conférer la résistance à divers traitements, notamment la chimiothérapie, ou au fait qu'une mutation peut entraîner des maladies telles que le diabète ou la mucoviscidose.

En 2001, la structure de l'homologue bactérien MsbA présentait un excellent point de départ pour une étude sur cette famille de protéines, car MsbA est plus proche des p-glycoprotéines des mammifères que les autres transporteurs ABC bactériens. Nous avons fait appel aux méthodes théoriques pour complémenter ces données de basse résolution.

La majeure partie de ce travail était effectuée par mon collègue J. CAMPBELL, à ce moment là doctorant à Oxford. Ce dernier m'a consulté en vue de profiter de mon expérience acquise lors de la modélisation d'autres protéines membranaires. Nous avons ainsi pu proposer un modèle détaillé du transporteur ABC MsbA qui pouvait servir de point de départ pour la modélisation par homologie d'autres protéines de la même famille qui sont des cibles thérapeutiques. [16] Lors de ce travail, il s'est avéré que le dimère résolu par cristallographie n'était pas viable et nous avons proposé un modèle alternatif. La structure correspondante a récemment été retirée de la PDB. [17] Une deuxième structure à l'état solide obtenue très récemment confirme en grande partie notre modèle proposé. [18]

Notre collaboration sur les transporteurs ABC s'est poursuivie dans l'étude du complexe SUR/Kir menée sur les fronts expérimentaux et théoriques simultanément. Vu la rapide évolution dans ce domaine et l'apparition de nouvelles structures, ces derniers travaux n'ont néanmoins pas été publiés.

Références bibliographiques du chapitre 3

[6] M. BAADEN, P. GRANGER et A. STRICH: "*Dependence of NMR isotropic shift averages and nuclear shielding tensors on the internal rotation of the functional group X about the C-X bond in seven simple vinylic derivatives H2C=CH-X*", 2000, **Mol.Phys.** 98, 329-342.

[7] M. BAADEN, G. WIPFF, M.R. YAFTIAN, M. BURGARD et D. MATT: "*Cation coordination by calix[4]arenes bearing amide and/or phosphine oxide pendant groups: how many arms are needed to bind Li+ vs. Na+? A combined NMR and molecular dynamics study*", 2000, **J.Chem.Soc. Perkin Trans. 2,** 1315-1321.

[8] M. BAADEN, F. BERNY, C. MADIC et G. WIPFF: "*M3+ lanthanide cation solvation by acetonitrile: The role of cation size, counterions, and polarization effects investigated by molecular dynamics and quantum mechanical simulations*", 2000, **J.Phys.Chem.A**. 104, 7659-7671.

[9] M. BAADEN, M. BURGARD, C. BOEHME et G. WIPFF: "*Lanthanide cation binding to a phosphoryl-calix[4]arene: the importance of solvent and counterions investigated by molecular dynamics and quantum mechanical simulations*", 2001, **Phys.Chem.Chem.Phys**. 3, 1317-1325.

[10] M. BAADEN, M. BURGARD et G. WIPFF: "*TBP at the water-oil interface: the effect of TBP concentration and water acidity investigated by molecular dynamics simulations*", 2001, **J.Phys.Chem.B**. 105, 11131-11141.

[11] S. BERNECHE et B. ROUX: "*Energetics of ion conduction through the K+ channel*", 2001, **Nature** 414, 73-7.

[12] a) J. DEISENHOFER, O. EPP, K. MIKI, R. HUBER et H. MICHEL : "*Structure of the protein subunits in the photosynthetic reaction centre of Rhodospeudomonas viridis at 3 Å resolution*", 1985, **Nature** 318, 618-624. ; b) J. DEISENHOFER, O. EPP, I. SINNING et H. MICHEL : "*Crystallographic refinement at 2.3 A resolution and refined model of the photosynthetic reaction centre from Rhodopseudomonas viridis*", 1995, **J.Mol.Biol.** 246, 429-57.

[13] http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html (Laboratoire de S. WHITE à UC Irvine).

[14] E. WALLIN et G. VON HEIJNE: "*Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*", 1998, **Protein Sci.** 7, 1029-38.

[15] M. BAADEN, C. MEIER et M.S.P. SANSOM: "*A molecular dynamics investigation of mono- and dimeric states of the outer membrane enzyme Ompla*", 2003, **J.Mol.Biol.** 331, 177-189.

[16] J.D. CAMPBELL, P.C. BIGGIN, M. BAADEN et M.S.P. SANSOM: "*Extending the structure of an ABC transporter to atomic resolution: modelling and simulation studies of Msba*", 2003, **Biochemistry** 42, 3666-3673.

[17] G. CHANG, C. B. ROTH, C. L. REYES, O. PORNILLOS, Y.-J. J. CHEN et A. P. CHEN: "*Retraction*", 2006, **Science** 314, 1875.

[18] A. WARD, C. L. REYES, J. YU, C. B. ROTH et G. CHANG : "*Flexibility in the ABC transporter MsbA: Alternating access with a twist*", 2007, **Proc. Natl. Acad. Sci. U.S.A.** 104, 19005-10

Chapitre 4

Modélisation de systèmes biologiques complexes à l'échelle atomique

"La biologie est la moins mathématisable des sciences parce que la plus lourde en contenu concret."

Jean Rostand

1 Structures expérimentales, modèles et simulations

La structure de départ utilisée pour simuler un système moléculaire provient souvent d'une détermination expérimentale ou pour les protéines parfois d'un modèle par homologie. J'ai montré une corrélation entre la qualité de la structure de départ et les caractéristiques d'une courte simulation de dynamique moléculaire (DM) d'environ deux nanosecondes (ns) sur un ensemble de protéines membranaires. [19] La qualité d'un modèle peut également être évaluée par DM et les modèles de qualité moyenne sont souvent améliorés par la simulation. Un deuxième travail examine l'impact de structures de départ différentes, soit déterminées par RMN soit par cristallographie, sur la simulation. [20] Dans le cas du récepteur GLIC se posait la question si la structure cristallographique ouverte était un artefact du à des molécules de lipides qui se sont logés au sein du pore de ce canal. Mes simulations ont pu confirmer la stabilité de la structure ouverte en l'absence de lipides sur une échelle de temps de 20 ns. [21] Ce travail continue, le but étant de cristalliser la structure

fermée et de la caractériser par DM (voir chapitre 5). Cela ouvre la voie à l'étude de la transition entre états ouvert et fermé, une première pour cette famille de récepteurs.

Les temps de simulation accessibles pour les systèmes biologiques - de l'ordre de dizaines de nanosecondes au début des années 2000 - sont souvent trop courts par rapport aux temps caractéristiques des phénomènes étudiés. De surcroît, il est difficile de quantifier l'échantillonnage et la convergence d'une simulation. Une comparaison quantitative de la convergence de 10 simulations de protéines membranaires a permis d'évaluer la qualité de l'échantillonnage pour chaque partie de la protéine étudiée. [22] Souvent, ce sont les parties très flexibles et non structurées qui sont mal échantillonnées.

Pour pallier les problèmes d'échelle de temps et de longueur imposés par la complexité biologique, des méthodes réduisant le détail de la représentation du système (*coarse-graining*) gagnent actuellement beaucoup de terrain. J'ai préparé une revue de ces techniques pour résumer les dernières avancées de ce domaine en effervescence et pour donner quelques conseils sur le choix de la méthode et de la représentation pour traiter un problème donné. [23] Plusieurs applications faisant appel à une représentation gros-grain sont en cours. [24],[25a]

Lors de ces travaux j'ai encadré ou co-encadré plusieurs étudiants qui ont permis aux projets d'aboutir: C. COX (co-encadrement du début de thèse), M.-P. DURRIEU (co-encadrement de la thèse), O. DELALANDE, N. FEREY et M. CHAVENT (encadrement des stages post-doctoraux).

2 Rôle des protéines SNARE dans la fusion membranaire

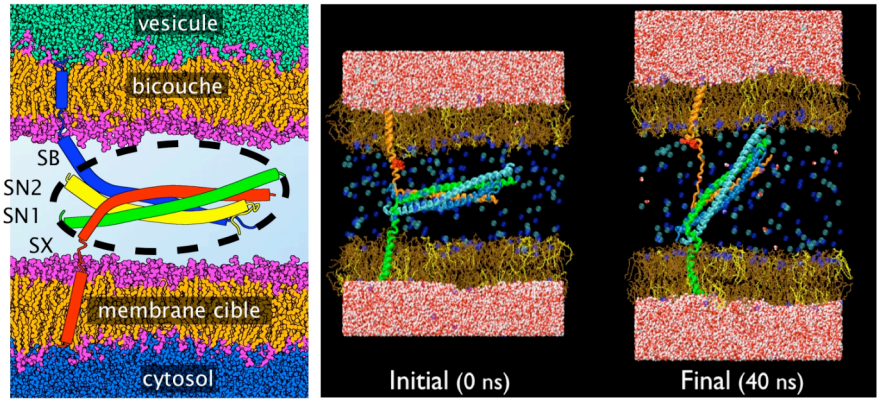


Figure 4.1: Le complexe SNARE inséré dans une double bicouche (340 000 atomes).

Le transport de matériel entre les différents compartiments de la cellule eucaryote ou vers le milieu extracellulaire implique en général la fusion des membranes des vésicules avec la membrane cible. Cette fusion membranaire est un processus de première importance pour la communication entre les cellules. Les protéines SNARE (*Soluble N-ethyl maleimide sensitive-*

factor Attachment protein REceptors) y semblent jouer un rôle majeur en formant un complexe sous forme de tresse composée de quatre hélices. Ce complexe est ancré dans les deux membranes pour permettre leur rapprochement. A l'aide de simulations moléculaires, j'étudie les propriétés de ce complexe, sa stabilité vis-à-vis de forces imposées par le processus de fusion, ainsi que son influence sur l'organisation des deux membranes. Mon objectif est de comprendre et de décrire en détail les forces auxquelles la tresse d'hélices est soumise et d'étudier son comportement dans un environnement réaliste.

Ce travail comporte trois volets. Premièrement, l'étude de la partie soluble du complexe, dont plusieurs simulations dans l'eau explicite ont été effectuées en partant de la structure cristallographique. Il en résulte des simulations et des conformations de référence du complexe à l'équilibre qui sont analysées en termes structuraux et énergétiques. Lors de la thèse de M.-P. DURRIEU, nous avons mis en évidence la structure en couches du complexe qui mène à des motifs de liaisons hydrogène et de ponts salins particuliers. [26] Une des hélices du complexe, la synaptobrévine, se distingue par sa plus grande rigidité.

Dans le cadre de la « *DEISA Extreme Computing Initiative* » 2005 j'ai entrepris une simulation très ambitieuse de l'intégralité du complexe SNARE inséré dans une double bicouche. Ce système de 340 000 atomes a été simulé pendant 40 ns et fait l'objet d'une collaboration étendue avec les groupes de M. LAGUERRE à Bordeaux, M.S.P. SANSOM à Oxford, E. KRIEGER en Autriche et R. LAVERY à Lyon. Les premiers résultats montrent une fragilité de la région juxtamembranaire de la synaptobrévine, le positionnement des tryptophanes 89 et 90 à l'extérieur de la membrane et un effet sur l'inclinaison des parties transmembranaires par la tension exercée sur le complexe. [27],[25a]

Un troisième volet concerne l'ancrage de la synaptobrévine dans la vésicule. Selon des études de RPE, la région juxtamembranaire de la synaptobrévine isolée est insérée dans la membrane, rendant indisponibles une dizaine de résidus pour la formation du complexe, et suggérant ainsi un mécanisme de régulation. Des simulations par DM du domaine transmembranaire inséré dans une bicouche lipidique ont été réalisées avec un modèle gros-grain. Ce modèle permet une meilleure exploration conformationnelle des protéines au sein des membranes qu'un modèle atomistique. Différentes conditions ont été explorées : le domaine trans- et juxtamembranaire seul, avec le domaine cytosolique de la synaptobrévine, ou inclus dans le complexe SNARE entier, dans le but de préciser leur influence sur le comportement de la synaptobrévine isolée ou complexée. [24] J'ai pu mettre en évidence des sous-états conformationnels de longue durée de vie. Les résultats sont en très bon accord avec l'expérience, mais diffèrent dans certains détails susceptibles de jouer un rôle dans la fonction biologique. La comparaison entre rat et levure a également révélé quelques différences.

L'ensemble de ces travaux a pu se faire avec la participation de M.-P. DURRIEU que j'ai co-encadrée en thèse avec R. LAVERY, et B. LAURENT qui a récemment effectué son stage

Master 1 avec moi. La suite est prévu pour la thèse de A. TEK qui a démarré en septembre 2009 et est décrite dans le chapitre 5.

3 Systèmes enzymatiques et relation fonction/fluctuation structurale

Un premier travail sur **l'enzyme bactérienne OmpT**, une protéine membranaire, a été publié. [28] J'y ai mis en évidence l'architecture du site actif et l'importance d'une molécule d'eau qui pourrait intervenir dans la réaction enzymatique. Un premier modèle approximatif d'un complexe enzyme-substrat issu de calculs d'amarrage moléculaire a été proposé et ce modèle a depuis été amélioré. Une récente étude avec un modèle hybride tout atome/gros-grain m'a permis d'explorer davantage le complexe OmpT-ARRA. Pour s'affranchir des problèmes d'échantillonnage, une simulation d'une microseconde a été effectuée. [29] Cette simulation a montré qu'il fallait à nouveau effectuer des simulations de détail atomique pour explorer des degrés de liberté approximés. Ce travail est en cours et pourrait bien indiquer un mécanisme enzymatique jusqu'alors ignoré. Dans le cadre du projet *BioSimGrid* j'ai contribué de façon importante à la comparaison d'un ensemble de simulations sur quatre enzymes de type hydrolase dont OmpT fait partie. [30] Les sites actifs de ces enzymes à priori distantes montrent une grande similitude dans les fluctuations. J'ai également utilisé des simulations interactives (décrites ci-dessous) pour étudier les propriétés de OmpT. [25a]

Le projet « FonFlon » dont je suis porteur vise à explorer la relation entre les fluctuations structurales d'une enzyme et son activité catalytique. Un nombre croissant d'expériences montre que la flexibilité structurale peut être essentielle au bon fonctionnement d'une protéine, soit en permettant les déformations nécessaires à son activité catalytique (dans le cas d'une enzyme) soit en lui permettant de s'associer avec son substrat par un mécanisme d'ajustement induit (*induced fit*). Comme il a été montré ailleurs, les fluctuations structurales sont directement liées aux propriétés mécaniques. [31]

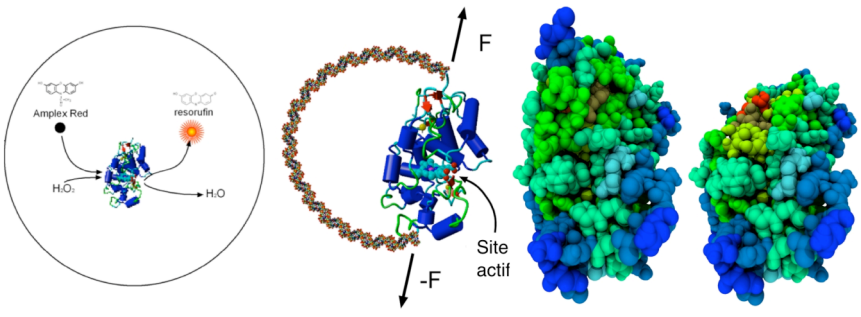


Figure 4.2: L'enzyme de la peroxidase du raifort comme modèle pour l'expérience (gauche, détection par fluorescence et application de contraintes par un ressort d'ADN) et le calcul (droite, visualisation du profil de rigidité calculé).

La peroxidase de raifort (*horseradish peroxidase*) et la guanylate kinase sont deux systèmes enzymatiques auxquels je m'intéresse dans le cadre de ce projet. En collaboration avec O. DELALANDE, S. SACQUIN-MORA, F. MOSCONI et D. BENSIMON, il s'agit d'effectuer des simulations qui permettraient de compléter les expériences de type molécule unique en cours visant à suivre l'activité enzymatique sous contrainte. Le financement de l'Agence Nationale de la Recherche nous permet de renforcer et de pérenniser cette collaboration. L'étude interactive par réalité virtuelle, faisant intervenir le projet FVNano, a également apporté des informations sur le mécanisme de cet enzyme. [25]

Dans le cadre d'une collaboration avec B. HARTMANN je m'intéresse également aux enzymes agissant sur l'ADN. La **DNase I** est une glycoprotéine hydrolysant les liens phosphodiéster de l'ADN en présence de cations divalents, Ca^{2+} et Mg^{2+}, et son activité dépend de la séquence d'ADN. Le système DNase I/ADN est un modèle plutôt simple et représentatif d'un complexe non-spécifique. Nous avons montré que les ions Ca^{2+} et Mg^{2+} sont essentiels pour optimiser la complémentarité électrostatique entre l'ADN et la protéine. Nous avons également identifié deux nouveaux sites de fixation de cations. [32] Il pourrait s'agir d'une facette encore peu étudiée des relations structure/fonction gouvernant les interactions non-spécifiques entre ADN et protéines. Lors de ce travail j'ai co-encadré M. GUEROULT (stage Master 2 recherche et début de thèse) et à moindre mesure J. ABI GHANEM.

4 Réalité virtuelle, graphisme interactif et manipulation d'objets nanoscopiques

Le projet FVNano a pour objectif le développement d'outils logiciels pour la simulation interactive haute performance couplant réalité virtuelle, visualisation scientifique et simulation parallèle. L'application principale au cœur du projet est la manipulation et l'exploration de simulations d'objets biologiques ou physiques complexes à l'échelle nanoscopique sur une plateforme de réalité virtuelle. En partant d'un prototype qui s'appuyait en grande partie sur des logiciels existants (VMD et « Ligand »), j'ai entamé une extension vers des outils plus adaptés (VTK et « FlowVR »). Le module graphique interactif impliquant le bras haptique permet de suivre visuellement les calculs effectués en temps réel en réponse à l'introduction d'une force.

Ce projet nécessite des compétences pointues dans plusieurs domaines proches de l'informatique et est réalisé en collaboration avec le laboratoire d'Informatique Fondamentale d'Orléans (LIFO), le CEA/DIF/DSSI à Bruyères-le-Chatel et le MOAIS de l'INRIA Rhône-Alpes, financé par l'ANR. Ces laboratoires partenaires sont des spécialistes de la visualisation interactive et de la réalité virtuelle et le projet s'appuie sur l'application de réalité virtuelle « FlowVR » développée par deux d'entre eux. La collaboration avec l'IDRIS

sur l'outil « MDDriver » pour la visualisation interactive de simulations par DM vient renforcer cet axe. [25a] Tous les développements découlant de ce projet sont mis à disposition de la communauté scientifique. [33]

Ce travail à l'interface avec l'informatique a été présenté à plusieurs manifestations internationales du domaine. [25b,c] L'outil MDDriver et les nouvelles possibilités qu'il offre pour l'étude d'assemblages biologiques complexes sont décrits dans une récente publication. [25a] Au sein du laboratoire et suite à mon initiative nous avons dédié une salle spécifique pour les manipulations interactives, et l'utilisation fréquente du dispositif faite par mes collègues témoigne de l'intérêt de cette nouvelle méthode. Dans le cadre de l'étude des propriétés mécaniques, j'espère également pouvoir tirer pleinement profit du retour haptique afin de rendre ces propriétés directement accessibles à l'utilisateur. De manière plus générale, l'outil FVNano devra faciliter la visualisation et l'analyse des résultats, et puis permettre une fouille efficace du grand nombre de données accumulées lors des simulations.

5 Autres projets

5.1 Modélisation du transporteur bactérien du fer FepA

Les structures tridimensionnelles de plusieurs transporteurs du Fer, FhuA, FepA, FecA, et la structure du transporteur de vitamine B12 BtuB ont été résolues au début des années 2000. Ces protéines d'environ 75 kDa se présentent sous forme de tonneaux β à 22 feuillets dont l'intérieur est « bouché » par le domaine N-terminal d'environ 150 acides aminés.

La dynamique moléculaire m'aide à comparer les structures *apo* et *holo* de ce système, et j'espère pouvoir en tirer des conclusions par rapport à l'éventuel détachement du « bouchon » N-terminal. J'ai commencé par la modélisation de la protéine FepA et de la molécule entérobactine qui forme un complexe avec le fer lors de son transport. De longues simulations de DM (75 ns) dans un environnement modèle d'une membrane biologique en détail atomique ont été entreprises. L'analyse détaillée de ces simulations est en cours. Il y a un volet méthodologique en collaboration avec C. OGUEY qui vise à utiliser les polyèdres de VORONOI/LAGUERRE pour caractériser les canaux d'eau à travers FepA. Dans ce contexte je co-encadre le doctorant J. ESQUE sur cette application. Ces travaux ont récemment été présentés lors d'un congrès national. [34] L'analyse interactive des structures de FepA fait partie des stages post-doctoraux de O. DELALANDE et M. CHAVENT.

Je dispose également d'une base de données de plusieurs simulations par dynamique moléculaire dans un environnement membranaire explicite d'homologues de FepA que j'analyse en collaboration avec S. KHALID et M.S.P. SANSOM à Oxford.

5.2 Caractérisation des protéines motrices de type kinésine

Les moteurs moléculaires de type kinésine possèdent la propriété de convertir l'énergie chimique libérée lors de l'hydrolyse de l'ATP en énergie mécanique, leur permettant de transporter un « cargo » le long du cytosquelette. Les expériences de type molécule unique ont permis de bien caractériser les propriétés mécaniques de ces protéines. L'interprétation des forces mesurées en termes de déformations structurales reste floue.

J'ai exploité une approche simple appelée GNM (*Gaussian Network Model*) où la molécule est représentée comme un ensemble de ressorts liant des paires d'atomes Cα. Ceci nous a permis d'effectuer une étude comparative de 17 structures de kinésines connues et de mettre en évidence des propriétés mécaniques remarquables de la tête motrice. Le modèle GNM donne accès à des informations relatives à la flexibilité et aux mouvements collectifs des résidus.

La méthode employée possède un intérêt plus général et nous préparons actuellement un serveur web pour l'automatiser (en cours d`élaboration [35]). Ce projet implique J.C. GELLY maintenant dans le laboratoire DSIMB à l'INTS et E. LEVY actuellement en stage post-doctoral à Montréal.

5.3 Changements conformationnels du récepteur des lipoprotéines de faible densité

L'épuration du cholestérol sanguin est assurée par l'activité des récepteurs de lipoprotéines, tel le récepteur des lipoprotéines de faible densité (rLDL). Un dysfonctionnement peut entraîner le développement de l'athérosclérose. Deux conditions sont d'un intérêt particulier pour le récepteur : son état à pH 7, correspondant au milieu extracellulaire au moment de la capture des lipoprotéines, et celui à pH 5, correspondant au milieu de l'endosome et au relargage des lipoprotéines.

La structure du récepteur rLDL déterminée par cristallographie nous a servi de point de départ pour une étude de modélisation. Ce travail s'insère dans une collaboration avec D. PERAHIA à Orsay et le laboratoire du Prof. A KETTANI au Maroc, et a donné lieu à plusieurs échanges, dont un stage de recherche de R. CHATER au sein du laboratoire. Nous avons comblé les lacunes de la structure expérimentale en ajoutant les résidus et chaînes latérales manquants. L'étude de la structure complexe du rLDL qui comporte 700 acides aminés a été abordée avec ce modèle au niveau atomique en tenant compte des effets de solvant. Ces simulations visent en particulier la forme extracellulaire du rLDL où le « bras » de la molécule se détache de son « cœur », ce qui nécessite un grand changement conformationnel.

Références bibliographiques du chapitre 4

[19] R.J. LAW, C. CAPENER, M. BAADEN, P.J. BOND, J. CAMPBELL, G. PATARGIAS, Y. ARINAMINPATHY et M.S.P. SANSOM: "*Membrane protein structure quality in molecular dynamics simulation*", 2005, **J.Mol.Graph.Model.** 24, 157-165.

[20] K. COX, P.J. BOND, A. GROTTESI, M. BAADEN et M.S.P. SANSOM: "*Outer Membrane Proteins: Comparing X-Ray and NMR Structures by MD Simulations in Lipid Bilayers*", 2008, **Eur. Biophys. J.** 37, 131-141.

[21] N. BOCQUET, H. NURY, M. BAADEN, C. LE POUPON, J.P. CHANGEUX, M. DELARUE et P.J. CORRINGER: "*X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation*", 2009, **Nature** 457, 111-114..

[22] J.D. FARALDO-GOMEZ, L.R. FORREST, M. BAADEN, P.J. BOND, C. DOMENE, G. PATARGIAS, J. CUTHBERTSON et M.S.P. SANSOM: "*Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations*", 2004, **Proteins** 57, 783-791.

[23] M. BAADEN et R. LAVERY : "*There's plenty of room in the middle: multi-scale modelling of biological systems*", 2007, dans **Recent Advances in Protein engineering**, Research signpost, India, édité par A.G. DE BREVERN, pp. 173-195, Research Signpost, Trivandrum, Kerala, Inde.

[24] M.P. DURRIEU, P.J. BOND, M.S.P. SANSOM, R. LAVERY et M. BAADEN: "*Coarse-grain simulations of the R-SNARE fusion protein in its membrane environment detect long-lived conformational sub-states*", 2009, **ChemPhysChem** 10, 1548-155.

[25] a) O. DELALANDE, N. FEREY, G. GRASSEAU et M. BAADEN : "*Complex Molecular Assemblies at hand via Interactive Simulations*", 2009, **J. Comput. Chem.** 30, 2375-2387 ; b) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*A VR framework for interacting with molecular simulations*", 2008, dans **Proceedings of the 2008 ACM symposium on Virtual reality software and technology**, édité par E. KRUIFF, pp. 91-94, ACM, Bordeaux, France ; c) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*From Interactive to Immersive Molecular Dynamics*", 2008, dans **Proceedings of the International Workshop on Virtual Reality and Physical Simulation (VRIPHYS'08)**, édité par F. FAURE et M. TESCHNER, pp. 89-96, Eurographics, Grenoble, France.

[26] M.P. DURRIEU, R. LAVERY et M. BAADEN: "*Interactions between neuronal fusion proteins explored by molecular dynamics*", 2008, **Biophys.J.** 94, 3436-3446.

[27] a) E. KRIEGER, L. LEGER, M.P. DURRIEU, N. TAIB, P. BOND, M. LAGUERRE, R. LAVERY, M.S.P. SANSOM et M. BAADEN : "*Atomistic modeling of the membrane-embedded synaptic fusion complex: a grand challenge project on the DEISA HPC infrastructure*", 2007, dans **ParCo 2007, Parallel Computing: Architectures, Algorithms and Applications**, édité par C.B.G.R. JOUBERT, F. PETERS, T. LIPPERT, M. BUECKER, B. GIBBON, et B. MOHR, Vol. 38, pp. 729-736, John von Neumann Institute for Computing, Juelich, Allemagne ; b) CSC News 3, 2007, 10-11; 12. Tietoyhteys 3, 2007, 10-11; 12-13. DEISA Newsletter 5, 2007. DEISA Newsletter 5, 2008. iSGTW 1/2008; iSGTW 10/2008. CORDIS News 11/2007. Supercomputing online 11/2007 etc.

[28] M. BAADEN et M.S.P. SANSOM: "*OmpT: molecular dynamics simulations of an outer membrane enzyme*", 2004, **Biophys.J.** 87, 2942-2953.

[29] M. NERI, M. BAADEN, V. CARNEVALE, C. ANSELMI, A. MARITAN et P. CARLONI: "*Microseconds Dynamics Simulations of the Outer-Membrane Protease T*", 2008, **Biophys.J.** 94, 71-78.

[30] K. TAI, M. BAADEN, S. MURDOCK, B. WU, M.H. NG, S. JOHNSTON, R. BOARDMAN, H. FANGOHR, K. COX, J.W. ESSEX et M.S.P. SANSOM: "*Three hydrolases and a transferase: comparative analysis of active-site dynamics via the BioSimGrid database*", 2007, **J.Mol.Graph.Model.** 25, 896-902.

[31] a) S. SACQUIN-MORA et R. LAVERY: "*Investigating the local flexibility of functionnnal residues in hemoproteins*", 2006, **Biophys. J.** 90, 2706-2717; b) S. SACQUIN-MORA, E. LAFORET et R. LAVERY: "*Locating the active site of enzymes using mechanical properties*", 2007, **Proteins** 67, 350-359.

[32] O. DELALANDE, N. FEREY, B. LAURENT, M. GUEROULT, B. HARTMANN et M. BAADEN: "*Multi-resolution approach for interactively locating funtionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device*", accepté à **Pacific Symposium for Biocomputing 2010**.

[33] Projet ANR Calcul Intensif et Simulation 2007: *FlowVRNano, un laboratoire virtuel pour modéliser les systèmes moléculaires nanoscopiques en biologie et dans les matériaux*
http://www.baaden.ibpc.fr/projects/fvnano/

[34] J. ESQUE, M. BAADEN et C. OGUEY: "*Analyse de protéines trans-membranaires par polyèdres de Voronoi/Laguerre - les canaux d'eau à travers FepA*", affiche, GGMM 2009, Mittelwihr.

[35] http://www.shaman.ibpc.fr/modo

Chapitre 5

Programme de recherche – processus moléculaires à l'interface des cellules neuronales

"The difficulty lies, not in the new ideas,
but in escaping the old ones, which ramify,
for those brought up as most of us have been,
into every corner of our minds."

John Maynard Keynes

1 SNAREs et processus membranaires : vers des modèles à l'échelle mésoscopique

La prochaine étape de ce travail introduit au chapite 4 sera réalisée dans le cadre de la thèse d'ALEXANDRE TEK qui a démarré en septembre 2009. Il s'agit de passer au complexe de fusion intégral, inséré dans deux membranes adjacentes et de mettre en évidence le côté mécanique sous-jacent en analysant les forces en jeu, mais aussi en séparant les hélices formant le complexe SNARE. Une combinaison innovante de la dynamique moléculaire, de méthodes de simulations mises au point au LBT et de la réalité virtuelle permettra d'effectuer ces calculs extrêmement ambitieux, moyennant l'accès aux ressources de calcul haute performance des centres nationaux tels que l'IDRIS.

Les simulations classiques avec le premier modèle tout-atome de ce système sont limitées par une échelle de temps de l'ordre de plusieurs dizaines de nanosecondes. Pour y pallier, nous entreprendrons des simulations à des résolutions plus faibles (« gros-grain ») qui permettent d'atteindre la microseconde. [33] Cette approche multiéchelle représente une

voie très prometteuse, mais doit être validée par une cohérence entre plusieurs méthodes. Je propose trois niveaux de représentation, le plus détaillé étant la représentation tout-atome. Un niveau intermédiaire peut être étudié grâce au modèle gros-grain de Bond et Sansom développé à l'université d'Oxford. [34] J'entretiens une collaboration de longue date avec ce groupe. Dans ce modèle, tous les composants protéiques, lipidiques et solubles sont présents, mais à une granularité plus faible que dans le modèle tout-atome. Une autre piste sera explorée lors de la visite de LEONARDO DARRE (1/10/2009 au 31/1/2010) qui est en thèse avec SERGIO PANTANO à l'institut Pasteur de Montevideo. Ce dernier a développé un modèle gros-grain hybride alliant détail atomique pour la partie protéique avec une faible granularité pour l'eau et les lipides et est également un spécialiste du système SNARE. [35] Le troisième niveau concerne des modèles plus simplifiés développés dans notre laboratoire. [36] Ces dernières représentations permettent une abstraction des molécules de solvant qui représentent une part importante dans le temps de calcul (81 % des particules à simuler).

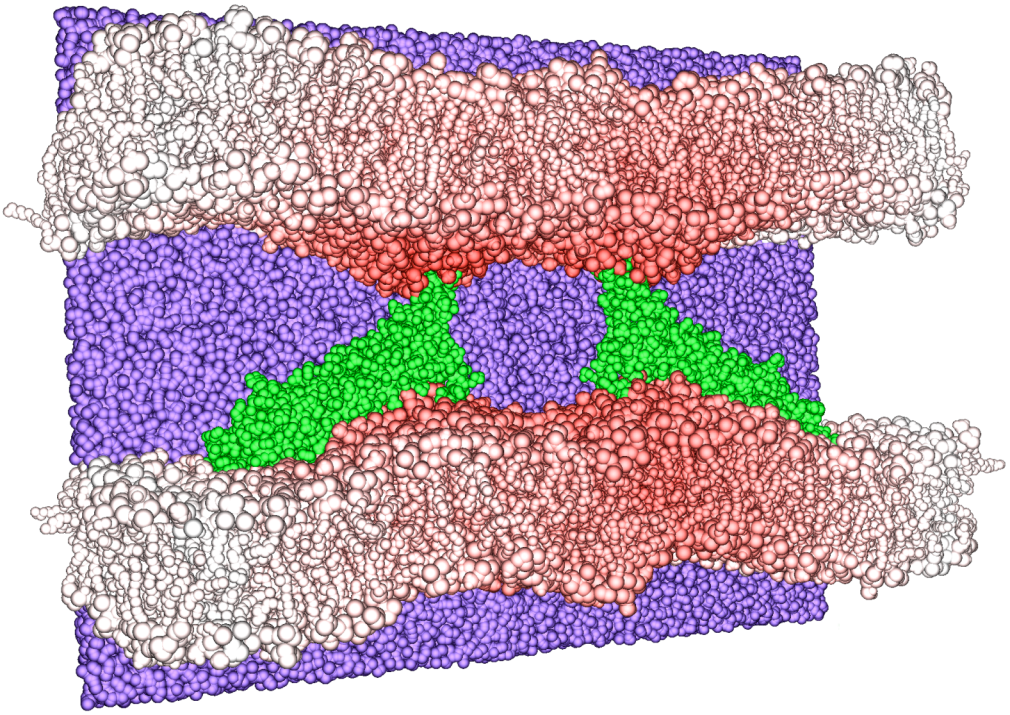


Figure 5.1: Vue d'esprit d'une dynamique augmentée d'un modèle de deux complexes SNARE (en vert). Une fine couche de molécules d'eau est montrée en pourpre. La bicouche lipidique est colorée selon l'intensité de la tension exercée par les deux tresses d'hélices du rouge (tension intense) au blanc. Image créée par Alex Tek.

Une plateforme logicielle de réalité virtuelle, exploitant un système graphique interactif équipé d'un bras haptique (à retour d'effort), FlowVRNano [37], est mise au point dans notre laboratoire (voir ci-dessous). Le cas du complexe SNARE représente un grand défi applicatif pour valider ce dispositif et pour le faire évoluer en fonction des besoins de l'utilisateur. Associé aux méthodes de simulation décrites ci-dessus, ce système permettra d'étudier les propriétés mécaniques de macromolécules biologiques et de leurs assemblages

en temps réel. [38] Dans le cadre de la présente thèse, il s'agira d'établir le lien entre les propriétés mécaniques du complexe SNARE et sa fonction biologique. Je souhaite pour cela développer une nouvelle méthode dite « dynamique augmentée » qui permettra de visualiser des propriétés de simulations en temps réel (voir figure ci-dessus).

2 Vers une compréhension moléculaire de l'activation du canal GLIC par le pH

Les complexes SNAREs décrits ci-dessus interviennent notamment dans la libération de neurotransmetteurs. Dans la transmission synaptique, il suit une étape de transduction du signal opérée par des canaux ioniques.

Les canaux ioniques déclenchés par un ligand forment une classe de protéines spécialisées dans la transduction d'un signal chimique/électrique à la surface de la membrane cellulaire. Parmi elles, les récepteurs dits « Cys-loop » forment une superfamille représentée dans toutes les formes de vie : eucaryotes, procaryotes et archaea. Cette classe majeure de récepteurs de neurotransmetteurs dans le système nerveux central humain est la cible de médicaments très importants. J'ai naguère démarré une collaboration intense et fructueuse avec les équipes de MARC DELARUE et PIERRE-JEAN CORRINGER à l'Institut Pasteur. Ces laboratoires ont récemment résolu la structure tridimensionnelle d'un analogue bactérien du récepteur nicotinique de l'acétylcholine, GLIC, déclenché par un changement de pH entre 7.0 (fermé) et 4.6 (ouvert). La structure a révélé un pentamère dans une conformation *a priori* ouverte. Il subsistait un doute sur le fait que la structure cristallographique de GLIC soit ouverte ou fermée. En combinant les simulations et l'expérience, nous avons pu confirmer et publier la première structure d'un tel récepteur dans un état supposé ouvert. [39] J'ai ensuite réalisé une simulation d'une microseconde représentant la fermeture (*gating*) du canal GLIC, provoquée par un changement de pH. Le canal se ferme et la protéine tend vers une structure stable, même si elle est hautement asymétrique. Ce travail est en cours de rédaction.

Par la suite, il s'agit d'aller plus loin dans l'étude des relations structure-fonction de ce canal, grâce à la combinaison entre simulations et expériences. Premièrement, l'état de protonation de la protéine à pH 4.6 reste à être clarifié ainsi que sa stabilité dans le temps, car je ne dispose pas à l'heure actuelle du bon jeu de charges partielles de cette protéine à pH acide. Deuxièmement, le groupe de P.J. CORRINGER a généré plusieurs mutants permettant d'induire un pont disulfure « non-naturel » dans la boucle M2-M3, à des endroits différents. La caractérisation fonctionnelle a été menée à bien et l'ouverture du canal au passage des ions (ou non) en fonction du pH a été mesurée. La structure de ces mutants a été cristallisée et résolue dans le groupe de M. DELARUE. On observe un changement conformationnel important, identique pour chacun des mutants, qui met en évidence une possible forme

« intermédiaire » du canal, forme qui se situe entre états ouvert et fermé. Les simulations permettront d'étudier cette forme dans un environnement non cristallin et proche de l'état naturel, une bicouche lipidique hydratée. Dernièrement, mon objectif serait de simuler la transition entre formes ouverte et fermée. Pour cela il est nécessaire d'affiner le modèle de la forme fermée en cours d'élaboration. Des simulations dirigées (*targeted molecular dynamics*) permettront ensuite, en partant de la forme ouverte, de cibler la forme fermée et ainsi de caractériser l'état de transition de ce processus. A terme, j'envisage d'examiner *in silico* l'effet de mutations ponctuelles sur la transition et la hauteur de la barrière énergétique séparant les deux états. Ces prédictions pourront être comparées à des mesures expérimentales qui seraient effectuées par le groupe de P.J. CORRINGER.

A nouveau, ce système représente un très bon cas d'étude et d'application pour les outils de réalité virtuelle que je développe. De surcroît, on peut s'attendre à des retombées importantes de ce travail pour la compréhension du fonctionnement des récepteurs de neurotransmetteurs, une cible majeure de médicaments très importants.

3 Valoriser les développements : vers un laboratoire virtuel

Le projet « FlowVRNano » a pour objectif le développement d'outils logiciels pour la simulation interactive haute performance couplant réalité virtuelle, visualisation scientifique et simulation parallèle. L'application principale au cœur du projet est la manipulation et l'exploration de simulations d'objets biologiques ou physiques complexes à l'échelle nanoscopique sur une plateforme de réalité virtuelle. Ces développements viendraient supporter les applications décrites ci-dessus.

Les avancées infographiques récentes permettent désormais de visualiser l'évolution dynamique de grandes structures biologiques à des temps qui rendent possible l'approche interactive. Les chercheurs peuvent ainsi agir de manière invasive au cours du calcul, en modulant des paramètres de simulation ou en manipulant le modèle moléculaire par application de forces. Celles-ci peuvent être rendues de manière tactile via une interface qui produit un retour de force (haptique). J'ai implémenté cette approche interactive avec succès et l'ai exploré en utilisant différents codes de simulation. [38,40] Cette démarche offre de nouvelles perspectives pour la compréhension de phénomènes biomoléculaires et pour la génération d'hypothèses qui pourront être vérifiées expérimentalement. Un contrôle-qualité est facilement accessible grâce à cette approche, la visualisation d'une simulation en cours permettant à l'utilisateur de détecter un problème de manière anticipée, la simulation de systèmes complexes étant parfois imprévisible. *Via* le dispositif d'interaction, l'utilisateur peut guider la simulation vers des états rares, ou bien qui ne seraient visités spontanément qu'après de longs calculs, ce qui permet de surmonter les limitations relatives aux échelles de

temps. Les simulations moléculaires interactives imposent toutefois quelques requis techniques. Le rendu graphique doit être efficace dans la mise à jour de l'affichage d'une molécule complexe et cet aspect est l'objet d'un travail en cours (figure 5.2). Le code de simulation doit permettre la réception et le traitement des données d'interaction produites par l'utilisateur. Simuler dans un temps interactif signifie que l'utilisateur doit voir et percevoir les changements occasionnés sur le système à son échelle de temps (entre 10 et 30 Hz).

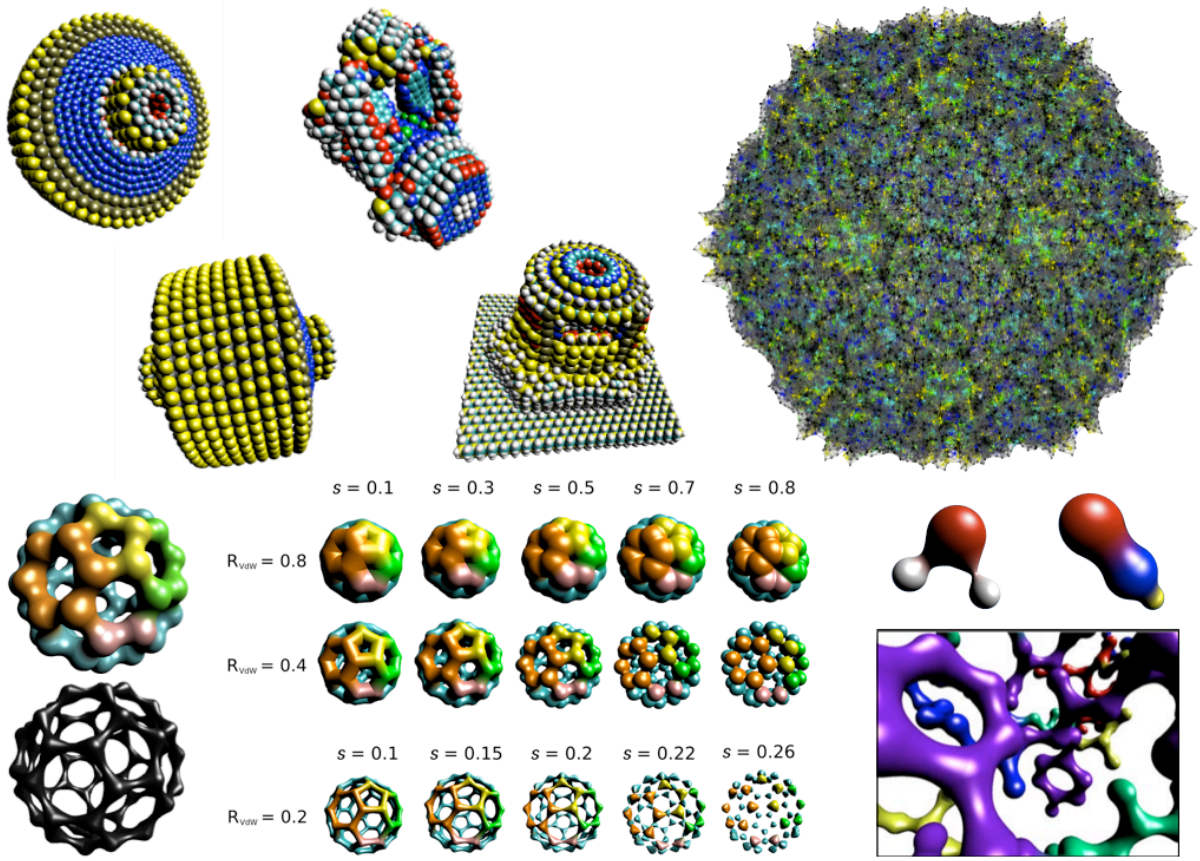


Figure 5.2: Rendu graphique développé dans le cadre du projet FlowVRNano lors du stage postdoctoral de Matthieu Chavent. Implémentation sur GPU utilisant la méthode de "Raycasting" et extension à une représentation par hypberboloïdes. Image adaptée à partir des résultats de M. Chavent.

Un autre aspect innovant concerne l'utilisation intensive de techniques de Réalité Virtuelle (RV). La stéréovision avancée augmente la perception 3D, ce qui est essentiel pour la visualisation de structures tridimensionnelles de grande taille. Les requis imposés par la simulation interactive en termes d'affichage nécessitent l'utilisation de technologies de rendus graphique de pointe, disponibles en RV. De plus, les périphériques d'entrée 3D ou 6D, couramment utilisés dans des environnements de RV sont sensiblement plus intuitifs que la souris (2D) pour l'interaction avec les objets moléculaires (3D). En proposant un retour tactile, les dispositifs haptiques s'avèrent tout à fait adaptés pour l'interaction avec des simulations — elles-mêmes souvent basées sur des forces. L'équipement étendu de RV que

j'ai mis en place dans notre laboratoire offre des conditions idéales pour la conception d'un laboratoire virtuel accessible au plus grand nombre de chercheurs. Cet outil permettra d'optimiser les échanges interdisciplinaires au sein d'une même équipe, ou entre chercheurs expérimentateurs et théoriciens. Le dispositif devrait également s'avérer extrêmement attractif pour l'enseignement et les étudiants, ou plus généralement, pour toute expérience pédagogique.

Outre les applications aux systèmes SNARE et GLIC, plusieurs applications génériques sont en cours de développement : identification interactive de sites de fixation d'ions, insertion de protéines membranaires dans une bicouche modèle et un dialogue avec des méthodes spectroscopiques, notamment la RMN et la RPE.

4 Perspectives

Fasciné par les assemblages moléculaires complexes du monde biologique et par les apports de la chimie physique combinée avec la réalité virtuelle pour comprendre leur fonctionnement, j'en fais le « *Leitmotiv* » de mes travaux futurs. J'ai également le souci de maintenir un équilibre entre l'application à des exemples concrets et le développement de nouvelles approches. L'aspect particulier que je compte explorer dans l'immédiat est le rôle des propriétés mécaniques moléculaires pour les processus biologiques.

Les applications prévues concernent en grande partie des études déjà entamées sur les protéines membranaires. Pour le complexe SNARE impliqué dans la fusion membranaire cela concerne son interaction avec les membranes. Pour le canal GLIC, il s'agit de mieux comprendre la transition entre états ouvert et fermé. L'axe de recherche sur les propriétés mécaniques implique également des travaux sur les outils disponibles et l'exploration de différentes méthodes pour en sélectionner les plus adaptées. Pour en donner un exemple, les simulations avec une représentation à gros-grain (« *coarse graining* ») ont été particulièrement avantageuses pour les expériences interactives.

L'originalité de ces recherches tient tout particulièrement à l'approche impliquant la réalité virtuelle que je compte approfondir. Un deuxième logiciel – à priori même accessible aux néophytes en modélisation – est en cours de finalisation.

Références bibliographiques du chapitre 5

[33] M. BAADEN et R. LAVERY : "*There's plenty of room in the middle: multi-scale modelling of biological systems*", 2007, dans **Recent Advances in Protein engineering**, Research signpost, India, édité par A.G. DE BREVERN, pp. 173-195, Research Signpost, Trivandrum, Kerala, Inde.

[34] M.P. DURRIEU, P.J. BOND, M.S.P. SANSOM, R. LAVERY et M. BAADEN: "*Coarse-grain simulations of the R-SNARE fusion protein in its membrane environment detect long-lived conformational sub-states*", 2009, **ChemPhysChem** 10, 1548-155.

[35] C. MONTECUCCO, G. SCHIAVO et S. PANTANO: "*SNARE complexes and neuroexocytosis: how many, how close?*", 2005, **Trends Biochem. Sci.** 30, 367-72.

[36] P. DERREUMAUX: "*Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations*", 2000, **Phys. Rev. Lett.** 85, 206-209.

[37] Projet ANR Calcul Intensif et Simulation 2007: *FlowVRNano, un laboratoire virtuel pour modéliser les systèmes moléculaires nanoscopiques en biologie et dans les matériaux*
http://www.baaden.ibpc.fr/projects/fvnano/

[38] O. DELALANDE, N. FEREY, G. GRASSEAU et M. BAADEN : "*Complex Molecular Assemblies at hand via Interactive Simulations*", 2009, **J. Comput. Chem.** 30, 2375-2387.

[39] N. BOCQUET, H. NURY, M. BAADEN, C. LE POUPON, J.P. CHANGEUX, M. DELARUE et P.J. CORRINGER: "*X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation*", 2009, **Nature**, 457, 111-114.

[40] a) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*A VR framework for interacting with molecular simulations*", 2008, dans **Proceedings of the 2008 ACM symposium on Virtual reality software and technology**, édité par E. KRUIFF, pp. 91-94, ACM, Bordeaux, France ; b) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*From Interactive to Immersive Molecular Dynamics*", 2008, dans **Proceedings of the International Workshop on Virtual Reality and Physical Simulation (VRIPHYS'08)**, édité par F. FAURE et M. TESCHNER, pp. 89-96, Eurographics, Grenoble, France ; c) O. DELALANDE, N. FEREY, B. LAURENT, M. GUEROULT, B. HARTMANN et M. BAADEN: "*Multi-resolution approach for interactively locating funtionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device*", accepté à **Pacific Symposium for Biocomputing 2010**.

Chapitre 6

Conclusion générale

" Very simple was my explanation,
and plausible enough -
as most wrong theories are!"

The Time Machine
H.G. Wells

1 Parcours et progression

En conclusion de ce mémoire, je tiens à souligner une forte continuité dans mes travaux avec une progression constante depuis la thèse. Cette progression a démarré avec la chimie physique et théorique en lien fort avec l'expérience dans le contexte de l'extraction liquide-liquide de cations. J'ai ensuite élargi le champs d'études aux systèmes biologiques, et aux protéines membranaires en particulier. Ces systèmes sont encore actuellement au cœur de mes recherches. Plus récemment, j'ai complété ces études par le développement de nouvelles méthodes de simulation, l'implémentation d'approches faisant appel à la réalité virtuelle et la mise à disposition de mes logiciels à la communauté scientifique.

2 Une vision au-delà des aspects scientifiques

Le raccourci souvent utilisé pour résumer (ou même évaluer) une carrière scientifique consiste à chiffrer la production en terme de publications. Même si ma « performance » en vue d'une telle analyse bibliométrique était tout à fait respectable (facteur h de 13, 40 citations par an en moyenne, 397 en tout), je suis d'avis que ces indicateurs donnent une vision trop individualiste et comptable du travail de recherche. Une grande partie de mes

travaux a été rendue possible grâce à l'encadrement de nombreux étudiants, doctorants et post-doctorants. Cette direction effective et productive m'a accessoirement permis de former des biochimistes à l'informatique ainsi que des informaticiens à la biochimie. Le mélange des disciplines et l'interdisciplinarité inhérente de mes sujets sont une source importante de nouvelles idées et d'approches en dehors des sentiers battus. L'axe portant sur la réalité virtuelle en est une manifestation très réussie et originale.

3 Utilité et légitimité de l'habilitation

Aujourd'hui je porte de manière autonome tous mes sujets de recherche. Je dirige depuis environ 2007 une petite équipe de taille variable. Dans le cadre de mon projet scientifique et professionnel, je serai amené à encadrer des doctorants. L'habilitation à diriger des recherches est une étape importante sanctionnant mon aptitude à maîtriser une stratégie de recherche scientifique et ma capacité à encadrer de jeunes chercheurs. Je pense avoir démontré par le présent mémoire mon aptitude à diriger des recherches.

Figure 6.1: Illustration par Jean Paul METAILIE, directeur de recherche au CNRS. A paraître dans le livre "Itinéraires bis - Mon parcours de jeune chercheur", Éditions Connaissances et Savoirs, 2009.

Annexe A

Titres et travaux

1 Publications dans des revues à comité de lecture

1.) M. BAADEN, P. GRANGER et A. STRICH: "*Dependence of NMR isotropic shift averages and nuclear shielding tensors on the internal rotation of the functional group X about the C-X bond in seven simple vinylic derivatives H2C=CH-X*", 2000, **Mol.Phys.** 98, 329-342.

2.) B. LAMBERT, V. JACQUES, A. SHIVANYUK, S.E. MATTHEWS, A. TUNAYAR, M. BAADEN, G. WIPFF, V. BOEHMER et J.F. DESREUX: "*Calix[4]arenes as selective extracting agents. An NMR dynamic and conformational investigation of the lanthanide(III) and thorium(IV) complexes*", 2000, **Inorg.Chem.** 39, 2033-2041.

3.) M. BAADEN, G. WIPFF, M.R. YAFTIAN, M. BURGARD et D. MATT: "*Cation coordination by calix[4]arenes bearing amide and/or phosphine oxide pendant groups: how many arms are needed to bind Li+ vs. Na+? A combined NMR and molecular dynamics study*", 2000, **J.Chem.Soc. Perkin Trans. 2**, 1315-1321.

4.) M. BAADEN, F. BERNY, C. MADIC et G. WIPFF: "*M3+ lanthanide cation solvation by acetonitrile: The role of cation size, counterions, and polarization effects investigated by molecular dynamics and quantum mechanical simulations*", 2000, **J.Phys.Chem.A**. 104, 7659-7671.

5.) M. BAADEN, M. BURGARD, C. BOEHME et G. WIPFF: "*Lanthanide cation binding to a phosphoryl-calix[4]arene: the importance of solvent and counterions investigated by molecular dynamics and quantum mechanical simulations*", 2001, **Phys.Chem.Chem.Phys**. 3, 1317-1325.

6.) M. BAADEN, M. BURGARD et G. WIPFF: "*TBP at the water-oil interface: the effect of TBP concentration and water acidity investigated by molecular dynamics simulations*", 2001, **J.Phys.Chem.B**. 105, 11131-11141.

7.) M. BAADEN, R. SCHURHAMMER et G. WIPFF: "*Molecular dynamics study of the uranyl extraction by tri-n-butylphosphate (TBP): Demixing of water/"oil"/TBP solutions with a comparison of supercritical CO2 and chloroform*", 2002, **J.Phys.Chem.B**. 106, 434-441.

8.) M. BAADEN, F. BERNY, R. SCHURHAMMER, C. MADIC et G. WIPFF: "*Theoretical studies on lanthanide cation extraction by picolinamides: ligand-cation interactions and interfacial behavior*", 2003, **Solv.Extr.Ion.Exch.** 21, 199-219.

9.) J.D. CAMPBELL, P.C. BIGGIN, M. BAADEN et M.S.P. SANSOM: "*Extending the structure of an ABC transporter to atomic resolution: modelling and simulation studies of Msba*", 2003, **Biochemistry** 42, 3666-3673.

10.) M. BAADEN, C. MEIER et M.S.P. SANSOM: "*A molecular dynamics investigation of mono- and dimeric states of the outer membrane enzyme Ompla*", 2003, **J.Mol.Biol.** 331, 177-189.

11.) J.D. FARALDO-GOMEZ, L.R. FORREST, M. BAADEN, P.J. BOND, C. DOMENE, G. PATARGIAS, J. CUTHBERTSON et M.S.P. SANSOM: "*Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations*", 2004, **Proteins** 57, 783-791.

12.) M. BAADEN et M.S.P. SANSOM: "*OmpT: molecular dynamics simulations of an outer membrane enzyme*", 2004, **Biophys.J.** 87, 2942-2953.

13.) R.J. LAW, C. CAPENER, M. BAADEN, P.J. BOND, J. CAMPBELL, G. PATARGIAS, Y. ARINAMINPATHY et M.S.P. SANSOM: "*Membrane protein structure quality in molecular dynamics simulation*", 2005, **J.Mol.Graph.Model.** 24, 157-165.

14.) K. TAI, M. BAADEN, S. MURDOCK, B. WU, M.H. NG, S. JOHNSTON, R. BOARDMAN, H. FANGOHR, K. COX, J.W. ESSEX et M.S.P. SANSOM: "*Three hydrolases and a transferase: comparative analysis of active-site dynamics via the BioSimGrid database*", 2007, **J.Mol.Graph.Model.** 25, 896-902.

15.) K. COX, P.J. BOND, A. GROTTESI, M. BAADEN et M.S.P. SANSOM: "*Outer Membrane Proteins: Comparing X-Ray and NMR Structures by MD Simulations in Lipid Bilayers*", 2008, **Eur. Biophys. J.**37, 131-141.

16.) M. NERI, M. BAADEN, V. CARNEVALE, C. ANSELMI, A. MARITAN et P. CARLONI: "*Microseconds Dynamics Simulations of the Outer-Membrane Protease T*", 2008,**Biophys.J.** 94, 71-78.

17.) M.P. DURRIEU, R. LAVERY et M. BAADEN: "*Interactions between neuronal fusion proteins explored by molecular dynamics*", 2008, **Biophys.J.** 94, 3436-3446.

18.) N. BOCQUET, H. NURY, M. BAADEN, C. LE POUPON, J.P. CHANGEUX, M. DELARUE et P.J. CORRINGER: "*X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation*", 2009, **Nature** 457, 111-114.

19.) O. DELALANDE, N. FEREY, G. GRASSEAU et M. BAADEN : "*Complex Molecular Assemblies at hand via Interactive Simulations*", 2009, **J. Comput. Chem.** 30, 2375-2387.

20.) M.P. DURRIEU, P.J. BOND, M.S.P. SANSOM, R. LAVERY et M. BAADEN: "*Coarse-grain simulations of the R-SNARE fusion protein in its membrane environment detect long-lived conformational sub-states*", 2009, **ChemPhysChem** 10, 1548-155.

2 Contributions invitées à des revues à comité de lecture

21.) M. BAADEN, F. BERNY, C. BOEHME, N. MUZET, R. SCHURHAMMER et G. WIPFF: "*Interaction of trivalent lanthanide cations with phosphoryl derivatives, amide, anisole, pyridine and triazine ligands: a quantum mechanics study*", 2000, **J.Alloys&Compounds.** 303, 104-111.

22.) L. TROXLER, M. BAADEN, V. BOHMER et G. WIPFF: "*Complexation of M3+ lanthanide cations by calix[4]arene-CMPO ligands: a molecular dynamics study in methanol solution and at a water/chloroform interface*", 2000, **Supramol.Chem.** 12, 27-51.

23.) M. BAADEN, F. BERNY et G. WIPFF: "*The chloroform TBP aqueous nitric acid interfacial system: a molecular dynamics investigation*", 2001, **J.Mol.Liq.** 90, 1-9.

3 Actes de conférence avec comité de lecture

24.) M. BAADEN, F. BERNY, N. MUZET, R. SCHURHAMMER et G. WIPFF : "*Separation of radioactive cations by liquid-liquid extraction: computer simulations of water / oil solutions of salts and ionophores*", 2000, dans **Proceedings of the Euradwaste 1999 conference**, édité par C. Davies, pp. 390-393, EC, Luxembourg.

25.) E. KRIEGER, L. LEGER, M.P. DURRIEU, N. TAIB, P. BOND, M. LAGUERRE, R. LAVERY, M.S.P. SANSOM et M. BAADEN : "*Atomistic modeling of the membrane-embedded synaptic fusion complex: a grand challenge project on the DEISA HPC infrastructure*", 2007, dans **ParCo 2007, Parallel Computing: Architectures, Algorithms and Applications**, édité par C.B.G.R. JOUBERT, F. PETERS, T. LIPPERT, M. BUECKER, B. GIBBON, et B. MOHR, Vol. 38, pp. 729-736, John von Neumann Institute for Computing, Juelich, Allemagne.

26.) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*A VR framework for interacting with molecular simulations*", 2008, dans **Proceedings of the 2008 ACM symposium on Virtual reality software and technology**, édité par E. KRUIFF, pp. 91-94, ACM, Bordeaux, France.

27.) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*From Interactive to Immersive Molecular Dynamics*", 2008, dans **Proceedings of the International Workshop on Virtual Reality and Physical Simulation (VRIPHYS'08)**, édité par F. FAURE et M. TESCHNER, pp. 89-96, Eurographics, Grenoble, France.

28.) O. DELALANDE, N. FEREY, B. LAURENT, M. GUEROULT, B. HARTMANN et M. BAADEN: "*Multi-resolution and multi-physics approach for interactively locating funtionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device*", accepté à **Pacific Symposium for Biocomputing 2010**.

4 Chapitres de livre avec comité de lecture

29.) M. BAADEN, F. BERNY, N. MUZET, L. TROXLER et G. WIPFF : "*Interfacial features of assisted liquid-liquid extraction of uranyl and cesium salts: a molecular dynamics investigation*", 2000, dans **Calixarenes for separations**, ACS Symposium Series 757, édité par G. Lumetta, R.D. Rogers, et A.S. Gopalan, pp. 71-85, Oxford University Press, New York.

30.) M. BAADEN et R. LAVERY : "*There's plenty of room in the middle: multi-scale modelling of biological systems*", 2007, dans **Recent Advances in Protein engineering**, Research signpost, India, édité par A.G. DE BREVERN, pp. 173-195, Research Signpost, Trivandrum, Kerala, Inde.

31.) S. KHALID et M. BAADEN : "*Molecular dynamics studies of outer membrane proteins : a story of barrels*", 2009, dans **Molecular Simulations and Biomembranes: From Biophysics to Function**, édité par P.C. Biggin and M.S.P. Sansom, Royal Society of Chemistry, Royaume Uni (sous presse).

5 Autres publications scientifiques

1.) M. BAADEN : "*Molecular Modeling with the ChemOffice Ultra 4.5 program suite.*", publié en ligne, 1999, CambridgeSoft Corporation.

2.) M. BAADEN : "*Etudes de molécules extractantes en solution et aux interfaces liquide-liquide: aspects structuraux et mécanistiques des effets de synergie*", **Thèse**, 2000, Université Louis Pasteur, Strasbourg (N°: 3630), 2 vol., 218/42 pages.

3.) General Discussion, *Faraday Discuss.* **144**, 2010, 203–222.

4.) General Discussion, *Faraday Discuss.* **144**, 2010, 445–466.

6 Logiciels

1.) MDDriver - http://mddriver.sourceforge.net et publication 19

2.) FVNano – http://www.baaden.ibpc.fr/projects/fvnano et publications 26, 27

3.) MonPote – voir publication 28

4.) BioSpring – version béta

Annexe B

Encadrement de stagiaires, doctorants et post-doctorants

"L'art de diriger consiste à savoir abandonner la baguette pour ne pas gêner l'orchestre."

Herbert von Karajan

1 Stagiaires

MEIER Christoph (co-direction M.S.P. SANSOM) (mars 2001 – juin 2001)
Part II-Student à l'université d'Oxford
Sujet: Simulations par dynamique moléculaire de la protéine membranaire OmplA

1 publication dans J. Molecular Biology.

CAMPBELL Jeff (co-direction M.S.P. SANSOM) (janvier 2002 – décembre 2002)
Doctorant à l'université d'Oxford
Sujet: Modélisation par homologie et simulations de dynamique moléculaire de transporteurs ABC

1 publication dans Biochemistry.

COX Catherine (co-direction M.S.P. SANSOM) (septembre 2002 – décembre 2002)
Doctorant à l'université d'Oxford
Sujet: Simulations de protéines membranaires à partir de structures expérimentales issues de la cristallographie et de la RMN

1 publication dans European Biophysical Journal.

GASSER Emeline (mars 2003 – mai 2003)
Stagiaire du DESS IAB (Informatique appliquée à la Biologie)
Sujet: Projet SHAMAN - Manipulation sensitive des macromolécules

1er prototype du logiciel SHAMAN.

GEARD Laurence (co-direction B. HARTMANN) (mars 2003 – mai 2003)
Stagiaire du DESS IAB (Informatique appliquée à la Biologie)
Sujet: Développement d'une application d'uniformisation des données issues de modélisation ou de la Protein Data Bank

1 prototype de logiciel de manipulation de fichiers PDB.

LEVY Emmanuel (co-direction R. LAVERY) (janvier 2004 – juin 2004)
Stagiaire du DEA AGM2 (Analyse de Génomes et Modélisation Moléculaire)
Sujet: Etude des propriétés mécaniques des moteurs moléculaires de type kinésine

1 prototype de serveur web déterminant des domaines "dynamiques".
1 publication en préparation.

CHATER Rachid (aout 2005)
Doctorant de l'université de Casablanca, Maroc
Sujet: Modélisation du récepteur des lipoprotéines de faible densité

MOSCONI Francesco (co-direction D. BENSIMON) (aout 2005 – novembre 2005)
Doctorant à l'ENS Paris
Sujet: Micromanipulations et simulations d'un processus enzymatique

ARBEZ Christophe (co-direction J.Ph. NOMINE) (avril 2006 – septembre 2006)
Thèse professionnelle à l'Institut Image, Arts et Métiers, ParisTech
Sujet: Développement d'un prototype du système FlowVRNano

1er prototype du logiciel FlowVRNano.

GUEROULT Marc (co-direction B. HARTMANN) (mars 2008 – aout 2008)
Stagiaire du Master 2 Pro d'ingénierie génomique fonctionnelle, Paris 7
Sujet: Modélisation « gros grain » des interactions entre protéines et ADN

1 publication dans Pacific Symposium for Biocomputing, une autre en préparation.

LAURENT Benoist (mars 2009 – aout 2009)
Stagiaire du Master 1 MBI (Biologie-Informatique), Paris 7
Sujets: Nanomécanique de la fusion membranaire: sur la trace des protéines SNARE ; conception d'un logiciel de simulation interactive basé sur Cocoa et VTK

1 publication dans Pacific Symposium for Biocomputing.
logiciel "BioSpring".

2 Doctorants

DURRIEU Marie-Pierre (co-direction R. LAVERY) (aout 2005 – avril 2008)
Chimie Analytique et Chimie Physique (ED388), Paris 6
Sujet: Etude par modélisation moléculaire de la stabilité et de la dynamique du complexe SNARE impliqué dans la fusion membranaire

2 publications dans Biophysical Journal et ChemPhysChem. 1 proceedings de conférence.

TEK Alex (septembre 2009 – présent)
iViv - Interdisciplinaire pour le vivant (ED387), Paris 6
Sujet: Nanomécanique de la fusion membranaire : simulations haute performance des protéines SNARE

3 Post-doctorants

DELALANDE Olivier (novembre 2007 – présent)
Thèse en Biochimie et Biologie Moléculaire, Paris 7
Sujet: Relation fonction/fluctuation chez les protéines (projet FonFlon)

1 publication dans J. Comput. Chem., 2 Short papers, 2 logiciels.

FEREY Nicolas (février 2008 – présent)
Thèse en Informatique (Bioinformatique), Paris-Sud XI Orsay
Sujet: FVnano : un laboratoire virtuel pour modéliser les systèmes moléculaires nanoscopiques en biologie et dans les matériaux

1 publication dans J. Comput. Chem., 2 Short papers, 2 logiciels.

CHAVENT Matthieu (aout 2009 – présent)
Thèse en Chimie Informatique et Théorique, LORIA, Nancy I
Sujet: FVnano : un laboratoire virtuel pour modéliser les systèmes moléculaires nanoscopiques en biologie et dans les matériaux

Participation à CAPRI round 19.

4 Ingénieurs

PAILLARD Guillaume (avril 2004 – septembre 2004)
Thèse en Modélisation Moléculaire, Paris 7
Tâche: Gestion informatique du laboratoire LBT

GUIOT Nicolas (mars 2005 – octobre 2007)
Ingénieur Informatique et Réseaux de télécommunication, ESAIP Angers
Tâche: Encadrement et formation initiale de l'Ingénieur d'Etudes responsable du parc informatique du laboratoire

KHABZAOUI Mohammed (juin 2008 – octobre 2009)
Thèse en Informatique, Lille 1
Tâche: Gestion informatique du laboratoire LBT

Annexe C

Résumés des publications scientifiques

Note

Les pages suivantes reproduisent les résumés de mes publications scientifiques. Ces résumés sont rédigés en langue anglaise.

1 Publications dans des revues à comité de lecture

1.) M. BAADEN, P. GRANGER et A. STRICH: "*Dependence of NMR isotropic shift averages and nuclear shielding tensors on the internal rotation of the functional group X about the C-X bond in seven simple vinylic derivatives H2C=CH-X*", 2000, **Mol.Phys.** 98, 329-342.

The 'Gauge Including Atomic Orbitals' (GIAO) approach is used to investigate the question of intramolecular rotation. Ab initio GIAO calculations of NMR chemical shielding tensors carried out with GAUSSIAN 94 within the SCF-Hartree-Fock approximation are described. In order to compare the calculated chemical shifts with experimental ones, it is important to use consistent nuclear shieldings for NMR reference compounds like TMS. The influence of rotating functional groups X=CH3, CHO, NO2, NH2, CONH2, COOH or C6H5 on the shielding tensors in seven vinylic derivatives H2C=CH-X is studied; the molecules are propene, acrolein, nitroethylene, ethyleneamine, acrylamide, acrylic acid and styrene. We observe a marked dependence of nuclear shielding and chemical shift on the torsional movement. Different Boltzmann averages over the conformational states are considered and compared for gas phase, liquid and solid state NMR. Their applicability to model cases for rigid or freely rotating molecules and for fixed molecules (e.g. polymers or proteins) with rapidly rotating groups is discussed and simple calculation models are presented. On the basis of this work it can be concluded that intramolecular rotation clearly affects the observed averages. Effects of up to 2 ppm have been observed for isotropic chemical shifts, and up to 17 ppm difference have been observed for individual tensor components, for example, of the carboxylic C-13 atom in acrylic acid. The variation of the shielding tensor on a nucleus in a fixed molecular backbone resulting from an attached rotating group furthermore leads to a new relaxation mechanism by chemical shift anisotropy.

2.) B. LAMBERT, V. JACQUES, A. SHIVANYUK, S.E. MATTHEWS, A. TUNAYAR, M. BAADEN, G. WIPFF, V. BOEHMER et J.F. DESREUX: "*Calix[4]arenes as selective extracting agents. An NMR dynamic and conformational investigation of the lanthanide(III) and thorium(IV) complexes*", 2000, **Inorg.Chem.** 39, 2033-2041.

The lanthanide and Th4+ complexes with calix[4]arene ligands substituted either on the narrow or at the wide rim by four coordinating groups behave totally differently as shown by an NMR investigation of the dia- and paramagnetic complexes. Solutions of complexes were prepared by reacting anhydrous metal perchlorate salts with the ligands in dry acetonitrile (CAUTION). Relaxation time T-1 titrations of acetonitrile solutions of Gd3+ by calixarenes indicate that ligands substituted on the narrow rim form stable 1:1 complexes whether they feature four amide groups (1) or four phosphine oxide functions. In contrast, a ligand substituted by four (carbamoylmethyl)diphenylphosphine oxide moieties on the wide rim (3) and its derivatives form polymeric species even at a 1:1 ligand/metal concentration ratio. Nuclear magnetic relaxation dispersion (NMRD) curves (relaxation rates 1/T-1 vs magnetic field strength) of Gd3+, Gd3+.1 and Gd3+.3 perchlorates in acetonitrile are analyzed by an extended version of the Solomon-Bloembergen-Morgan equations. A comparison of the calculated rotational correlation times tau(r) shows that ligand 3 forms oligomeric Gd3+ species. The chelates of ligand 1 are axially symmetric (C-4 symmetry), and the paramagnetic

shifts induced by the Yb3+ ion are accounted for quantitatively. The addition of water or of nitrate ions does not modify the geometry of the complex. The metal chelates of 3 and its derivatives adopt a C-2 symmetry, and the paramagnetic shifts are interpreted on a semiquantitative basis only. Water and NO3- ions completely labilize the complexes of the heavy lanthanides. The very high selectivity of ligand 3 through the lanthanide series stems from a complex interplay of factors.

3.) M. BAADEN, G. WIPFF, M.R. YAFTIAN, M. BURGARD et D. MATT: "*Cation coordination by calix[4]arenes bearing amide and/or phosphine oxide pendant groups: how many arms are needed to bind Li+ vs. Na+? A combined NMR and molecular dynamics study*", 2000, **J.Chem.Soc. Perkin Trans. 2**, 1315-1321.

Combined spectroscopic and theoretical studies have been performed on two recently developed calix[4]arenes in the cone conformation, L1 (bearing two -CH2C(O)NEt2 and two -CH2P(O)Ph-2 substituents occupying respectively distal phenolic positions) and L2 (with four -CH2P(O)Ph-2 substituents), in order to compare the Li+ vs. Na+ cation binding mode. Molecular dynamics simulations indicate that coordination of the Li+ cation involves three of the four substituents (the two phosphoryl groups and one of the two amide functions of L1; three phosphoryl arms of L2). A variable temperature NMR study carried out with L1 . Li+ confirms this fourfold coordination and reveals that in solution the lithium cation moves between the two adjacent OPOPOamide units. The weaker binding of the Na+ cation results in a more symmetrical coordination of the four phenolic oxygen atoms and two carbonyls of L1 or four phosphoryls of L2.

4.) M. BAADEN, F. BERNY, C. MADIC et G. WIPFF: "*M3+ lanthanide cation solvation by acetonitrile: The role of cation size, counterions, and polarization effects investigated by molecular dynamics and quantum mechanical simulations*", 2000, **J.Phys.Chem.A**. 104, 7659-7671.

We report a molecular dynamics (MD) study on M3+ lanthanide (La3+, Eu3+ and Yb3+) cations in dry acetonitrile solution and in M(MeCN)(n)(3+) clusters (n = 1-15) where two classical force-field representations of the cations are compared, in conjunction with the OPLS model of acetonitrile. It is shown that a set of van der Waals cation parameters (set2) fitted from free energies of hydration overestimates the cation coordination numbers (CNs). Another set of parameters (set1), where the size of cations is scaled down by 2(1/6) (using the sigma van der Waals value for R*) yields better results. Quantum mechanical calculations performed on M(MeCN)(n)(3+) aggregates (n = 1-9) demonstrate the importance of charge-transfer and polarization effects. They confirm the preferred coordination number of eight for Yb3+, the Yb(MeCN)(8+1)(3+) species with one MeCN molecule in the outer coordination sphere being somewhat more stable than Yb(MeCN)(9)(3+) D-3h. Adding a polarization term for the 1-6-12 OPLS acetonitrile to the force field (set2+pol) indeed markedly improves the calculated CNs. In all MD simulations, a remarkable dynamic feature is observed in the first solvation shell where the lifetime of acetonitrile molecules increases from Yb3+ to La3+, that is, inversely to the cation-solvent interaction energies and to the aqueous phase behavior. Rare-earth salts with ClO4- and F3CSO3- anions and the question of ion binding selectivity by L ligands (formation of ML33+ complexes, where L is a pyridine-dicarboxamide ligand) in acetonitrile solution are investigated by free-energy perturbation simulations, comparing the set1, set2, and set2+pol models. It is found that selectivities are markedly determined by the change in solvation-free energies of the uncomplexed cations, with pronounced

counterion effects. The two simplest models (set1 or set2 without polarization) predict the correct order of complexation (Yb3+ > Eu3+ > La3+), whereas addition of polarization contribution leads to the inverse order, because of overestimation of the cation-anion interactions in the salt solutions.

5.) M. BAADEN, M. BURGARD, C. BOEHME et G. WIPFF: "*Lanthanide cation binding to a phosphoryl-calix[4]arene: the importance of solvent and counterions investigated by molecular dynamics and quantum mechanical simulations*", 2001, **Phys.Chem.Chem.Phys**. 3, 1317-1325.

Molecular dynamics simulations on the 1:1 M3+ lanthanide (La3+, Eu3+ and Yb3+) inclusion? complex of a t-butyl-calix[4]arene L substituted at the narrow rim by four CH2-P(O)Ph2 arms demonstrate the role of hydration and counterions on the cation binding mode and shielding. In dry chloroform and in the absence of counterions, the cation is endo?, fully encapsulated within the pseudo-cavity delineated by the four phosphoryl arms and the four phenolic oxygens. This endo? bidentate binding mode is supported by full ab initio quantum mechanical optimization of the calixarene M3+ complexes. In biphasic solution, the complexes are shown to be surface active and to adsorb at an oil?/water interface with the cationic site pointing towards water and the hydrophobic t-butyl groups in oil?. The cation is not encapsulated, but adopts an exo ? position, coordinated to the four PO oxygens of L, to water molecules, and to counterions. This complex is too hydrophilic to be extracted from the interface to an organic phase. The unexpected binding mode has important implications concerning the mechanism of liquid-liquid ion extraction and the microscopic state of the extracted complex in the organic phase.

6.) M. BAADEN, M. BURGARD et G. WIPFF: "*TBP at the water-oil interface: the effect of TBP concentration and water acidity investigated by molecular dynamics simulations*", 2001, **J.Phys.Chem.B**. 105, 11131-11141.

Molecular dynamics simulations provide microscopic pictures of the behavior of TBP (tri-n-butyl phosphate) at the water-"oil" interface, and in water-"oil" mixtures where "oil" is modeled by chloroform. It is shown that, depending on the TBP concentration and water acidity, TBP behaves as a surfactant, an interface modifier, or a solute in oil. At low concentrations, TBP is surface active and forms an unsaturated monolayer at the "planar" interface between the pure water and oil phases, adopting an "amphiphilic orientation". Increasing the TBP concentration induces water-oil mixing at the interface which becomes very rough while TBP orientations at the phase boundary are more random and TBP molecules solubilize in oil. The effect of water acidity is investigated with three nitric acid models: neutral NO3H, ionic NO3- H3O+ and TBPH+ NO3-. The role of these species on the properties of the water-oil boundaries and on the outcome of water-oil demixing experiments is presented. The neutral NO3H form is highly surface active. Hydrogen bonding between TBP and NO3H, TBPH+, or H3O+ disrupts the first TBP layer and leads, at high TBP concentrations, to a mixed third phase or to a microemulsion. These results are important for our understanding of the microscopic solution state of liquid-liquid extraction systems.

7.) M. BAADEN, R. SCHURHAMMER et G. WIPFF: "*Molecular dynamics study of the uranyl extraction by tri-n-butylphosphate (TBP): Demixing of water/"oil"/TBP solutions with a comparison of supercritical CO2 and chloroform*", 2002, **J.Phys.Chem.B**. 106, 434-441.

We report molecular dynamics simulations on the phase separation of "perfectly mixed" water/oil/tri-n-butylphosphate (TBP) solutions containing 30 or 60 TBP molecules and 5 UO[2](NO[3])[2] complexes. The oil phase is mimicked by one of two liquids, either chloroform or supercritical CO[2] (SC-CO[2]). The simulations demonstrate the importance of TBP concentration. In the TBP[30] systems, the water and oil phases separate on the nanosecond time scale, leading to two interfaces onto which all TBPs adsord. Some of them spontaneously form 1:1 and 1:2 complexes with UO[2](NO[3])[2] at the interface. With the more concentrated TBP[60] systems, water and oil do not separate but form microemulsions containing neat water "droplets" surrounded by oil in which TBP is solubilized, sitting at the oil-water boundaries delimiting the droplets. All UO[2](NO[3])[2] salts are complexed by TBP at the liquid boundaries, sitting somewhat more in oil than in water. Following the Le Chatelier rule, the proportion of 1:2 complexes is larger in the TBP[60] system than in the TBP[30] system. Thus, TBP acts not only as a complexant but also as an "interface modifier". The simulations reveal strong analogies between chloroform and SC-CO[2] as organic phases. These novel results are crucial for our understanding of the state of heterogeneous solutions involved in uranyl extraction by TBP as well as assisted cation extraction by other extractants.

8.) M. BAADEN, F. BERNY, R. SCHURHAMMER, C. MADIC et G. WIPFF: "*Theoretical studies on lanthanide cation extraction by picolinamides: ligand-cation interactions and interfacial behavior*", 2003, **Solv.Extr.Ion.Exch.** 21, 199-219.

We report theoretical investigations on the complexation and liquid-liquid extraction of trivalent lanthanide cations by picolinamide ligands L. The relative contributions of pyridine vs. amide moieties of L and the effect of alkyl substituents are investigated in the gas phase by quantum mechanical calculations. The uncomplexed ligands L with different lipophilic substituents and Eu(NO3)(3)L-3 complexes are simulated at a water-"oil" interface, where "oil" is modeled by chloroform. The simulations point to the importance of interfacial phenomena in the liquid-liquid extraction of cations by picolinamide ligands and the difference between pyridine vs. amide substituted ligands.

9.) J.D. CAMPBELL, P.C. BIGGIN, M. BAADEN et M.S.P. SANSOM: "*Extending the structure of an ABC transporter to atomic resolution: modelling and simulation studies of Msba*", 2003, **Biochemistry** 42, 3666-3673.

Molecular modeling and simulation approaches have been use to generate a complete model of the prokaryotic ABC transporter MsbA from Escherichia coli, starting from the low-resolution structurebased CR trace (PDB code 1JSQ). MsbA is of some biomedical interest as it is homologous to mammalian transporters such as P-glycoprotein and TAP. The quality of the MsbA model is assessed using a combination of molecular dynamics simulations and static structural analysis. These results suggest that the approach adopted for MsbA may be of general utility for generating all atom models from lowresolution crystal structures of membrane proteins. Molecular dynamics simulations of the MsbA model inserted in a fully solvated octane slab (a membrane mimetic environment) reveal that while the monomer is relatively stable, the dimer is unstable and undergoes significant conformational drift on a nanosecond time scale. This suggests that the MsbA crystal dimer may not correspond to the MsbA dimer in vivo. An alternative model of the dimer is discussed in the context of available experimental data.

10.) M. BAADEN, C. MEIER et M.S.P. SANSOM: "*A molecular dynamics investigation of mono- and dimeric states of the outer membrane enzyme Ompla*", 2003, **J.Mol.Biol.** 331, 177-189.

OMPLA is a phospholipase found in the outer membranes of many Gram-negative bacteria. Enzyme activation requires calcium-induced dimerisation plus bilayer perturbation. As the conformation of OMPLA in the different crystal forms (monomer versus dimer; with/without bound Ca2+) is remarkably similar we have used multi-nanosecond molecular dynamics (MD) simulations to probe possible differences in conformational dynamics that may be related to enzyme activation. Simulations of calcium-free monomeric OMPLA, of the Ca2+-bound dimer, and of the Ca2+-bound dimer with a substrate analogue covalently linked to the active site serine have been performed, all with the protein embedded in a phospholipid (POPC) bilayer. All simulations were stable, but differences in the dynamic behaviour of the protein between the various states were observed. In particular, the stability of the active site and the hydrophobic substrate-binding cleft varied. Dimeric OMPLA is less flexible than monomeric OMPLA, especially around the active site. In the absence of bound substrate analogue, the hydrophobic substrate-binding cleft of dimeric OMPLA collapses. A model is proposed whereby the increased stability of the active site in dimeric OMPLA is a consequence of the local ordering of water around the nearby calcium ion. The observed collapse of the substrate-binding cleft may explain the experimentally observed occurrence of multiple dimer conformations of OMPLA, one of which is fully active while the other shows significantly reduced activity.

11.) J.D. FARALDO-GOMEZ, L.R. FORREST, M. BAADEN, P.J. BOND, C. DOMENE, G. PATARGIAS, J. CUTHBERTSON et M.S.P. SANSOM: "*Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations*", 2004, **Proteins** 57, 783-791.

In the current report, we provide a quantitative analysis of the convergence of the sampling of conformational space accomplished in molecular dynamics simulations of membrane proteins of duration in the order of 10 nanoseconds. A set of proteins of diverse size and topology is considered, ranging from helical pores such as gramicidin and small beta-barrels such as OmpT, to larger and more complex structures such as rhodopsin and FepA. Principal component analysis of the C(alpha)-atom trajectories was employed to assess the convergence of the conformational sampling in both the transmembrane domains and the whole proteins, while the time-dependence of the average structure was analyzed to obtain single-domain information. The membrane-embedded regions, particularly those of small or structurally simple proteins, were found to achieve reasonable convergence. By contrast, extra-membranous domains lacking secondary structure are often markedly under-sampled, exhibiting a continuous structural drift. This drift results in a significant imprecision in the calculated B-factors, which detracts from any quantitative comparison to experimental data. In view of such limitations, we suggest that similar analyses may be valuable in simulation studies of membrane protein dynamics, in order to attach a level of confidence to any biologically relevant observations.

12.) M. BAADEN et M.S.P. SANSOM: "*OmpT: molecular dynamics simulations of an outer membrane enzyme*", 2004, **Biophys.J.** 87, 2942-2953.

Five MD simulations (total duration >25 ns) have been performed on the E. coli outer membrane protease OmpT embedded in a DMPC lipid bilayer. Globally the

protein is conformationally stable. Some degree of tilt of the beta-barrel is observed relative to the bilayer plane. The greatest degree of conformational flexibility is seen in the extracellular loops. A complex network of fluctuating H-bonds is formed between the active site residues, such that the Asp210-His212 interaction is maintained throughout, whilst His212 and Asp83 are often bridged by a water molecule. This supports a catalytic mechanism whereby Asp83 and His212 bind a water molecule which attacks the peptide carbonyl. A configuration yielded by docking calculations of OmpT simulation snapshots and a model substrate peptide Ala-Arg-Arg-Ala was used as the starting point for an extended Huckel calculation on the docked peptide. These placed the LUMO mainly on the carbon atom of the central C=O in the scissile peptide bond, thus favouring attack on the central peptide by the water held by residues Asp83 and His212. The trajectories of water molecules reveal exchange of waters between the intracellular face of the membrane and the interior of the barrel but no exchange at the extracellular mouth. This suggests that the 'pore-like' region in the centre of OmpT may enable access of water to the active site 'from below'. The simulations appear to reveal the presence of specific lipid interaction sites on the surface of the OmpT barrel. This reveals the ability of extended MD simulations to provide meaningful information on protein/lipid interactions.

13.) R.J. LAW, C. CAPENER, M. BAADEN, P.J. BOND, J. CAMPBELL, G. PATARGIAS, Y. ARINAMINPATHY et M.S.P. SANSOM: "*Membrane protein structure quality in molecular dynamics simulation*", 2005, **J.Mol.Graph.Model.** 24, 157-165.

Our goal was to assess the relationship between membrane protein quality, output from protein quality checkers and output from molecular dynamics (MD) simulations. Membrane transport proteins are essential for a wide range of cellular processes. Structural features of integral membrane proteins are still under-explored due to experimental limitations in structure determination. Computational techniques can be used to exploit biochemical and medium resolution structural data, as well as sequence homology to known structures, and enable us to explore the structure?function relationships in several transmembrane proteins. The quality of the models produced is vitally important to obtain reliable predictions. An examination of the relationship between model stability in molecular dynamics (MD) simulations derived from RMSD (root mean squared deviation) and structure quality assessment from various protein quality checkers was undertaken. The results were compared to membrane protein structures, solved at various resolution, by either X-ray or electron diffraction techniques. The checking programs could predict the potential success of MD in making functional conclusions. MD stability was shown to be a good indicator for the quality of structures. The quality was also shown to be dependent on the resolution at which the structures were determined.

14.) K. TAI, M. BAADEN, S. MURDOCK, B. WU, M.H. NG, S. JOHNSTON, R. BOARDMAN, H. FANGOHR, K. COX, J.W. ESSEX et M.S.P. SANSOM: "*Three hydrolases and a transferase: comparative analysis of active-site dynamics via the BioSimGrid database*", 2007, **J.Mol.Graph.Model.** 25, 896-902.

Comparative molecular dynamics (MD) simulations enable us to explore the conformational dynamics of the active sites of distantly related enzymes. We have used the BioSimGrid (http://www.biosimgrid.org) database to facilitate such a comparison. Simulations of four enzymes were analyzed. These included three hydrolases and a transferase, namely acetylcholinesterase, outer-membrane

phospholipase A, outer-membrane protease T, and PagP (an outer-membrane enzyme which transfers a palmitate chain from a phospholipid to lipid A). A set of 17 simulations were analyzed corresponding to a total of ~0.1 ms simulation time. A simple metric for active-site integrity was used to demonstrate the existence of clusters of dynamic conformational behaviour of the active sites. Small (i.e. within a cluster) fluctuations appear to be related to the function of an enzymatically active site. Larger fluctuations (i.e. between clusters) correlate with transitions between catalytically active and inactive states. Overall, these results demonstrate the potential of a comparative MD approach to analysis of enzyme function. This approach could be extended to a wider range of enzymes using current high throughput MD simulation and database methods.

15.) K. COX, P.J. BOND, A. GROTTESI, M. BAADEN et M.S.P. SANSOM: "*Outer Membrane Proteins: Comparing X-Ray and NMR Structures by MD Simulations in Lipid Bilayers*", 2008, **Eur. Biophys. J.**37, 131-141.

The structures of three bacterial outer membrane proteins (OmpA, OmpX and PagP) have been determined by both X-ray diffraction and NMR. We have used multiple (7 × 15 ns) MD simulations to compare the conformational dynamics resulting from the X-ray versus the NMR structures, each protein being simulated in a lipid (DMPC) bilayer. Conformational drift was assessed via calculation of the root mean square deviation as a function of time. On this basis the 'quality' of the starting structure seems mainly to influence the simulation stability of the transmembrane β-barrel domain. Root mean square fluctuations were used to compare simulation mobility as a function of residue number. The resultant residue mobility profiles were qualitatively similar for the corresponding X-ray and NMR structure-based simulations. However, all three proteins were generally more mobile in the NMR-based than in the X-ray simulations. Principal components analysis was used to identify the dominant motions within each simulation. The first two eigenvectors (which account for >50% of the protein motion) reveal that such motions are concentrated in the extracellular loops and, in the case of PagP, in the N-terminal α-helix. Residue profiles of the magnitude of motions corresponding to the first two eigenvectors are similar for the corresponding X-ray and NMR simulations, but the directions of these motions correlate poorly reflecting incomplete sampling on a ~10 ns timescale.

16.) M. NERI, M. BAADEN, V. CARNEVALE, C. ANSELMI, A. MARITAN et P. CARLONI: "*Microseconds Dynamics Simulations of the Outer-Membrane Protease T*", 2008, **Biophys.J.** 94, 71-78.

Conformational fluctuations of enzymes may play an important role for substrate recognition and/or catalysis, as it has been suggested in the case of the protease enzymatic superfamily. Unfortunately, theoretically addressing this issue is a problem of formidable complexity, as the number of the involved degrees of freedom is enormous: indeed, the biological function of a protein depends, in principle, on all its atoms and on the surrounding water molecules. Here we investigated a membrane protease enzyme, the OmpT from Escherichia coli, by a hybrid molecular mechanics/coarse-grained approach, in which the active site is treated with the GROMOS force field, whereas the protein scaffold is described with a Go-model. The method has been previously tested against results obtained with all-atom simulations. Our results show that the large-scale motions and fluctuations of the electric field in the microsecond timescale may impact on the biological function and suggest that OmpT employs the same catalytic strategy as

aspartic proteases. Such a conclusion cannot be drawn within the 10- to 100-ns timescale typical of current molecular dynamics simulations. In addition, our studies provide a structural explanation for the drop in the catalytic activity of two known mutants (S99A and H212A), suggesting that the coarse-grained approach is a fast and reliable tool for providing structure/function relationships for both wild-type OmpT and mutants.

17.) M.P. DURRIEU, R. LAVERY et M. BAADEN: "*Interactions between neuronal fusion proteins explored by molecular dynamics*", 2008, **Biophys.J.** 94, 3436-3446.

In this report, we present features of the neuronal SNARE complex determined by atomistic molecular dynamics simulations. The results are robust for three models, varying force fields (AMBER and GROMOS) and solvent environment (explicit and implicit). An excellent agreement with experimental findings is observed. The SNARE core complex behaves like a stiff rod, with limited conformational dynamics. An accurate picture of the interactions within the complex emerges with a characteristic pattern of atomic contacts, hydrogen bonds, and salt bridges reinforcing the underlying layer structure. This supports the metaphor of a molecular Velcro strip that has been used by others to describe the neuronal fusion complex. No evidence for directionality in the formation of these interactions was found. Electrostatics largely dominates all interactions, with an acidic surface patch structuring the hydration layers surrounding the complex. The interactions within the four-helix bundle are asymmetric, with the synaptobrevin R-SNARE notably exhibiting an increased rigidity with respect to the three Q-SNARE helices. The interaction patterns we observe provide a new tool for interpreting the impact of mutations on the complex.

18.) N. BOCQUET, H. NURY, M. BAADEN, C. LE POUPON, J.P. CHANGEUX, M. DELARUE et P.J. CORRINGER: "*X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation*", 2009, **Nature**, 457, 111-114.

Pentameric ligand-gated ion channels from the Cys-loop family mediate fast chemo-electrical transduction1, 2, 3, but the mechanisms of ion permeation and gating of these membrane proteins remain elusive. Here we present the X-ray structure at 2.9 Å resolution of the bacterial Gloeobacter violaceus pentameric ligand-gated ion channel homologue4 (GLIC) at pH 4.6 in an apparently open conformation. This cationic channel is known to be permanently activated by protons5. The structure is arranged as a funnel-shaped transmembrane pore widely open on the outer side and lined by hydrophobic residues. On the inner side, a 5 Å constriction matches with rings of hydrophilic residues that are likely to contribute to the ionic selectivity6, 7, 8, 9. Structural comparison with ELIC, a bacterial homologue from Erwinia chrysanthemi solved in a presumed closed conformation10, shows a wider pore where the narrow hydrophobic constriction found in ELIC is removed. Comparative analysis of GLIC and ELIC reveals, in concert, a rotation of each extracellular -sandwich domain as a rigid body, interface rearrangements, and a reorganization of the transmembrane domain, involving a tilt of the M2 and M3 -helices away from the pore axis. These data are consistent with a model of pore opening based on both quaternary twist and tertiary deformation.

19.) O. DELALANDE, N. FEREY, G. GRASSEAU et M. BAADEN : "*Complex Molecular Assemblies at hand via Interactive Simulations*", 2009, **J. Comput. Chem.** 30, 2375-2387.

Studying complex molecular assemblies interactively is becoming an increasingly appealing approach to molecular modelling. Here we focus on interactive molecular dynamics (IMD) as a textbook example for interactive simulation methods. Such simulations can be useful in exploring and generating hypotheses about the structural and mechanical aspects of biomolecular interactions. For the first time, we carry out low-resolution coarse-grain IMD simulations. Such simplified modelling methods currently appear to be more suitable for interactive experiments and represent a well balanced compromise between an important gain in computational speed versus a moderate loss in modelling accuracy compared to higher resolution all-atom simulations. This is particularly useful for initial exploration and hypothesis development for rare molecular interaction events. We evaluate which applications are currently feasible using molecular assemblies from 1 900 to over 300 000 particles. Three biochemical systems are discussed: the guanylate kinase enzyme, the outer membrane protease T and the SNARE complex involved in membrane fusion. We induce large conformational changes, carry out interactive docking experiments, probe lipid-protein interactions and are able to sense the mechanical properties of a molecular model. Furthermore such interactive simulations facilitate exploration of modelling parameters for method improvement. For the purpose of these simulations, we have developed a freely available software library called MDDriver. It uses the IMD protocol from NAMD and facilitates the implementation and application of interactive simulations. With MDDriver it becomes very easy to render any particle-based molecular simulation engine interactive. Here we use its implementation in the Gromacs software as an example.

20.) M.P. DURRIEU, P.J. BOND, M.S.P. SANSOM, R. LAVERY et M. BAADEN: "*Coarse-grain simulations of the R-SNARE fusion protein in its membrane environment detect long-lived conformational sub-states*", 2009, **ChemPhysChem** 10, 1548-155.

Coarse-grain molecular dynamics are used to look at conformational and dynamic aspects of an R-SNARE peptide inserted in a lipid bilayer. This approach allows carrying out microsecond-scale simulations which bring to light long-lived conformational sub-states potentially interesting in the context of the membrane fusion mechanism mediated by the SNARE proteins. We show that these coarse-grain models are in agreement with most experimental data on the SNARE system, but differ in some details that may have a functional interest, most notably in the orientation of the soluble part of R-SNARE that does not appear to be spontaneously accessible for SNARE complex formation. We also compare rat and yeast sequences of R-SNARE and find some minor differences in their behavior.

2 Contributions invitées à des revues à comité de lecture

21.) M. BAADEN, F. BERNY, C. BOEHME, N. MUZET, R. SCHURHAMMER et G. WIPFF: "*Interaction of trivalent lanthanide cations with phosphoryl derivatives, amide, anisole, pyridine and triazine ligands: a quantum mechanics study*", 2000, **J.Alloys&Compounds.** 303, 104-111.

We report *ab initio* quantum mechanical calculations on charged LM3+ and neutral LMCl3 complexes formed by lanthanide M3+ cations (M = La, Eu, Yb) and

model ligands L, where L are phosphorous derivatives R3PO (R = alkyl/O-alkyl/phenyl), R3PS and R2PS2- (R = alkyl/phenyl), and amide, pyridine, triazine and anisole ligands. Among all neutral ligands studied, Ph3PO is intrinsically clearly the best. However, the comparison of LM3+ to LMCl3 complexes demonstrates that the concept of 'ligand basicity' is not sufficient to compare the efficiency of cation coordination. Counterions play an important role in the structures of the complexes and for the consequences of substitution in the Ligand. For instance, in the absence of competing interactions, phenyl substituted R3PS or R2PS2- ligands interact better than alkyl substituted ones, but the order is reversed in the presence of counterions. Counterions also amplify the alkyl vs. O-alkyl substituent effect in R3PO complexes. Bidentate anions or more bulky anions are expected to amplify the effects observed with chloride anions. Thus, multiple interactions between counterions and the other species in the first coordination sphere markedly contribute to the 'effectiveness' and stereochemistry of ligand-cation interactions.

22.) L. TROXLER, M. BAADEN, V. BOHMER et G. WIPFF: "*Complexation of M3+ lanthanide cations by calix[4]arene-CMPO ligands: a molecular dynamics study in methanol solution and at a water/chloroform interface*", 2000, **Supramol.Chem.** 12, 27-51.

We report a molecular dynamics study on the 1:1 M3+ lanthanide (La3+, Eu3+ and Yb3+) inclusion complexes of an important extractant molecule L: calix[4]arene-tetraalkyl ether substituted at the wide rim by four NH-C(O)-CH2-P(O)Ph-2 arms. The M(NO3)(3) and MCl3 complexes of L are compared in methanol solution and at a water / chloroform interface. In the different environments the coordination sphere of M3+ involves the four phosphoryl oxygens and three to four loosely bound carbonyl oxygens of the CMPO-like arms. Based on free energy simulations, we address the question of ion binding selectivity in pure liquid phases and at the liquid-liquid interface where L and the complexes are found to adsorb. According to the simulations, the enhancement of M3+ cation extraction in the presence of the calixarene platform, examined by comparing L to the (CMPO)(4) "ligand" at the interface, is related to the fact that (i) the (CMPO)(4)Eu(NO3)(3) complex is more hydrophilic than the LEu(NO3) one and (ii) the free CMPO ligands spread at the interface, and are therefore less organized for cation capture than L.

23.) M. BAADEN, F. BERNY et G. WIPFF: "*The chloroform TBP aqueous nitric acid interfacial system: a molecular dynamics investigation*", 2001, **J.Mol.Liq.** 90, 1-9.

We report MD studies on a chloroform / nitric acid water interface, either neat, or saturated by TBP molecules. Two extreme models are compared, where the acid is either neutral HNO3 or dissociated to NO3- and H3O+. The latter species are found to be "repelled" by the neat interface, while the neutral HNO3 molecules are surface active. When the neat interface is saturated by TBP molecules, the latter form highly disordered arrangements instead of a regular monolayer, and water is dragged to the organic phase as 1:1, 1:2 and 2:2 hydrates of TBP. Simulations with the neutral HNO3 model lead to extraction of acid to the organic phase, hydrogen bonded to the phosphoryl oxygen of TBP, forming HNO3:TBP adducts of 1:1 and 2:1 types. Simulations with the ionic model lead to H3O+:TBP adducts of 1:1, 1:2 and 1:3 types in the organic phase and significant mixing of the chloroform, TBP and water liquids in the interfacial region.

3 Actes de conférence avec comité de lecture

24.) M. BAADEN, F. BERNY, N. MUZET, R. SCHURHAMMER et G. WIPFF : "*Separation of radioactive cations by liquid-liquid extraction: computer simulations of water / oil solutions of salts and ionophores*", 2000, dans **Proceedings of the Euradwaste 1999 conference**, édité par C. Davies, pp. 390-393, EC, Luxembourg.

> Despite the importance of liquid-liquid extraction (LLE) separation techniques to separate ions from nuclear waste solutions, little is known about the detailed ion capture and selection mechanism and on the nature and solvation of the involved species. This led us to undertake related theoretical studies in two main directions. First, assess by QM (quantum mechanics) calculations the intrinsic characteristics of ion-ligand interactions (precise structures, interaction energies, electronic properties) in the gas phase. This is achieved on small model ligands (phosphoryl derivatives, amides, pyridines) with different stoichiometries, including counterions. Second, simulate by molecular dynamics (MD) the solution behaviour of ions and of ligands in their free or complexed states. The Cs+ / Na+ extraction by calix[4]crown molecules represents a successful simulation study, where the cation selectivity was investigated as a function of the solvent, the nature of counterions and the conformation of the ligand.

25.) E. KRIEGER, L. LEGER, M.P. DURRIEU, N. TAIB, P. BOND, M. LAGUERRE, R. LAVERY, M.S.P. SANSOM et M. BAADEN : "*Atomistic modeling of the membrane-embedded synaptic fusion complex: a grand challenge project on the DEISA HPC infrastructure*", 2007, dans **ParCo 2007, Parallel Computing: Architectures, Algorithms and Applications**, édité par C.B.G.R. Joubert, F. Peters, T. Lippert, M. Buecker, B. Gibbon, et B. Mohr, Vol. 38, pp. 729-736, John von Neumann Institute for Computing, Juelich, Allemagne.

> The SNARE protein complex is central to membrane fusion, a ubiquitous process in biology. Modelling this system in order to better understand its guiding principles is a challenging task. This is mainly due to the complexity of the environment: two adjacent membranes and a central bundle of four helices formed by vesicular and plasma membrane proteins. Not only the size of the actual system, but also the computing time required to equilibrate it render this a demanding task requiring exceptional computing resources. Within the DEISA Extreme Computing Initiative (DECI), we have performed 40 ns of atomistic molecular dynamics simulations with an average performance of 81.5 GFlops on 96 processors using 218 000 CPU hours. Here we describe the setup of the simulation system and the computational performance characteristics.

26.) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*A VR framework for interacting with molecular simulations*", 2008, dans **Proceedings of the 2008 ACM symposium on Virtual reality software and technology**, édité par E. Kruiff, pp. 91-94, ACM, Bordeaux, France.

> Molecular Dynamics is nowadays routinely used to complement experimental studies and overcome some of their limitations. In particular, current experimental techniques do not allow to directly observe the full dynamics of a macromolecule at atomic detail. Molecular simulation provides time-dependent atomic positions, velocities and system energies according to biophysical models. Many molecular simulation engines can now compute a molecular dynamics trajectory of interesting biological systems in interactive time. This progress has

lead to a new approach called interactive molecular dynamics. It allows to control and visualise a molecular simulation in progress. We have developed a generic library, called MDDriver, to facilitate the implementation of such interactive simulations. It allows to easily create a network between a molecular user interface and a physically-based simulation. We use this library in order to study an interesting biomolecular system, simulated by various interaction-enabled molecular engines and models. We use a classical molecular visualisation tool and a haptic device to control the dynamic behavior of the molecule. This approach provides encouraging results for interacting with a biomolecule and understanding its dynamics. Starting from this initial success, we decided to use VR functionalities more intensively, by designing a VR framework dedicated to immersive and interactive molecular simulations. This framework is based on MDDriver, on the visualisation toolkit VTK, and on the vtkVRPN library, which encapsulates the VRPN library into VTK.

27.) N. FEREY, O. DELALANDE, G. GRASSEAU et M. BAADEN : "*From Interactive to Immersive Molecular Dynamics*", 2008, dans **Proceedings of the International Workshop on Virtual Reality and Physical Simulation (VRIPHYS'08)**, édité par F. Faure et M. Teschner, pp. 89-96, Eurographics, Grenoble, France.

Molecular Dynamics simulations are nowadays routinely used to complement experimental studies and overcome some of their limitations. In particular, current experimental techniques do not allow to directly observe the full dynamics of a macromolecule at atomic detail. Molecular simulation engines provide time-dependent atomic positions, velocities and system energies according to biophysical models. Many molecular simulation engines can now compute a molecular dynamics trajectory of interesting biological systems in interactive time. This progress has lead to a new approach called Interactive Molecular Dynamics. It allows the user to control and visualise a molecular simulation in progress. We have developed a generic library, called MDDriver, in order to facilitate the implementation of such interactive simulations. It allows one to easily create a network connection between a molecular user interface and a physically-based simulation. We use this library in order to study a biomolecular system, simulated by various interaction-enabled molecular engines and models. We use a classical molecular visualisation tool and a haptic device to control the dynamic behavior of the molecule. This approach provides encouraging results for interacting with a biomolecule and understanding its dynamics. Starting from this initial success, we decided to use Virtual Reality (VR) functionalities more intensively, by designing a VR framework dedicated to immersive and interactive molecular simulations. This framework is based on MDDriver, on the visualisation toolkit VTK, and on the vtkVRPN library, which encapsulates the VRPN library into VTK.

28.) O. DELALANDE, N. FEREY, B. LAURENT, M. GUEROULT, B. HARTMANN et M. BAADEN: "*Multi-resolution and multi-physics approach for interactively locating funtionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device*", accepté à **Pacific Symposium for Biocomputing 2010**.

Metal ions drive important parts of biology, yet it remains experimentally challenging to locate their binding sites. Here we present an innovative computational approach. We use interactive steering of charged ions or small molecules in an electrostatic potential map in order to identify potential binding sites. The user interacts with a haptic device and experiences tactile feedback related to the strength of binding at a given site. The potential field is the first

level of resolution used in this model. Any type of potential field can be used, implicitly taking into account various conditions such as ionic strength, dielectric constants or the presence of a membrane. At a second level, we represent the accessibility of all binding sites by modelling the shape of the target macromolecule via non-bonded van der Waals interactions between its static atomic or coarse-grained structure and the probe molecule(s). The third independent level concerns the representation of the molecular probe itself. Ion selectivity can be assessed by using multiple interacting ions as probes. The resulting multi-level simulation combines multiple physics models and is intrinsically multi-resolution. This method was successfully applied to the DNase I enzyme, where we recently identified two new cation binding sites by computationally expensive extended molecular dynamics simulations.

4 Chapitres de livre avec comité de lecture

29.) M. BAADEN, F. BERNY, N. MUZET, L. TROXLER et G. WIPFF : "*Interfacial features of assisted liquid-liquid extraction of uranyl and cesium salts: a molecular dynamics investigation*", 2000, dans **Calixarenes for separations**, ACS Symposium Series 757, edited by G. Lumetta, R.D. Rogers, et A.S. Gopalan, pp. 71-85, Oxford University Press, New York.

We report molecular dynamics studies on the interfacial behavior of species involved in the assisted liquid-liquid extraction of cesium and uranyl cations. The distribution of uncomplexed Cs+Pic- and UO2(NO3)2 salts is first described at the water/chloroform interface, where they are found to remain adsorbed, contrary to expectations from extraction experiments. The question of synergistic effects is addressed by simulating these calixarenes at a TBP saturated interface (TBP = tributylphosphate). Finally, in relation with the extraction of uranyl by TBP, we report a computer demixing experiment of a "perfectly mixed" ternary water / chloroform / TBP mixture containing 5 UO2(NO3)2 molecules. The phase separation is found to be rapid, leading to the formation of a TBP layer between the aqueous and organic phases and to spontaneous complexation of the uranyl salts by TBP. The complexes formed are not extracted to chloroform, but remain close to the water/organic phse boundary. The simulations reveal the importance of interfacial phenomena in the ion extraction and recognition process.

30.) M. BAADEN et R. LAVERY : "*There's plenty of room in the middle: multi-scale modelling of biological systems*", 2007, dans **Recent Advances in Protein engineering**, Research signpost, India, edited by A.G. de Brevern, pp. 173-195, Research Signpost, Trivandrum, Kerala, Inde.

Understanding proteins well enough to rationally modify their biological function requires understanding how these biomolecules behave in their natural cellular, or extra-cellular, environments. This, in turn, implies understanding their interactions with a wide variety of other species within a dense and heterogeneous medium. Molecular modelling and simulation can certainly contribute to improving our understanding in this area, however the range of molecules and processes involved in the biological systems implies that a range of modelling techniques will have to be applied, balancing the requirements for accuracy and precision against the constraints imposed by the size, time and energy scales involved. This article attempts to summarize the various representations, methodologies and target functions presently available to molecular modellers, discusses how different combinations of these basic features

can be combined to attack different problems and then considers the role of hybrid methods, an area where there is still much scope for development.

31.) S. KHALID et M. BAADEN : "*Molecular dynamics studies of outer membrane proteins : a story of barrels*", 2009, dans **Molecular Simulations and Biomembranes: From Biophysics to Function**, edited by P.C. Biggin and M.S.P. Sansom, Royal Society of Chemistry, United Kingdom (in press).

The outer membrane proteins (OMPs) of Gram-negative bacteria play a key role in the function and structural integrity of the outer membrane. The OMPs cover a number of different functions, including membrane pores, passive and active transporters, recognition proteins, and membrane-bound enzymes. From a biomedical perspective, OMPs are of some interest as potential targets for novel antimicrobial drugs and vaccines. Currently only ~30 high-resolution structures of OMPs are known, revealing all but two of them to be based on a transmembrane barrel architecture, with sizes ranging from 8 to 22 strands in the barrel.

5 Autres publications scientifiques

1.) M. BAADEN : "*Molecular Modeling with the ChemOffice Ultra 4.5 program suite.*", publié en ligne, 1999, CambridgeSoft Corporation.

Marc Baaden is a chemical engineer at the European Higher Institute of Chemistry of Strasbourg. He is currently preparing a PhD thesis in the field of molecular dynamics investigating synergistic effects in liquid-liquid extraction of ions.

2.) M. BAADEN : "*Etudes de molécules extractantes en solution et aux interfaces liquide-liquide: aspects structuraux et mécanistiques des effets de synergie*", **Thèse**, 2000, Université Louis Pasteur, Strasbourg (N°: 3630), 2 vol., 218/42 pages.

Molecular dynamics simulations reported herein provide new important insights into cation recognition and complexation in solution as well as liquid-liquid extraction, with a particular focus on the microscopic events taking place at the interface between two immiscible liquids. Preliminary studies concerned the representation of the trivalent rare earth cations La3+, Eu3+ and Yb3+ in force field simulations, probing structural and energetic features on an experimentally characterized model system based on substituted pyridine dicarboxamide ligands. Complexation of such cations by a novel calixarene derivative was investigated showing unexpected features, such as the position of the cation in the complex. Independent experimental studies published subsequently support these findings. Another part of the work is related to industrial liquid-liquid extraction systems using tri-n-butyl phosphate (TBP) as co-solvent, extractant, surfactant and synergist. We investigate 1) concentration effects simulating up to 60 TBP at a water/chloroform interface, 2) acidity using a neutral and ionic model of HNO3 and 3) synergistic aspects of mixed TBP/calixarene extraction systems. These simulations provide the first microscopic insights into such issues. We finally addressed the topic of solute transfer across the water/chloroform interface. The potential of mean force for such a process has been calculated by both standard methods and novel approaches.

Annexe D

Publications représentatives

1 - Coarse-grain simulations of the R-SNARE fusion protein in its membrane environment detect long-lived conformational sub-states, *2009, **ChemPhysChem** 10, 1548-155*

2 - X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation, *2009, **Nature** 457, 111-114*

3 - Complex Molecular Assemblies at hand via Interactive Simulations, *2009, **J. Comput. Chem.** 30, 2375-2387*

4 - Multi-resolution and multi-physics approach for interactively locating funtionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device, *accepté à **Pacific Symposium for Biocomputing 2010***

CHEMPHYSCHEM

DOI: 10.1002/cphc.200900216

Coarse-Grain Simulations of the R-SNARE Fusion Protein in its Membrane Environment Detect Long-Lived Conformational Sub-States

Marie-Pierre Durrieu,[a] Peter J. Bond,[b] Mark S. P. Sansom,[c] Richard Lavery,[d] and Marc Baaden*[a]

Coarse-grain molecular dynamics are used to look at conformational and dynamic aspects of an R-SNARE peptide inserted in a lipid bilayer. This approach allows carrying out microsecond-scale simulations which bring to light long-lived conformational sub-states potentially interesting in the context of the membrane fusion mechanism mediated by the SNARE proteins. We show that these coarse-grain models are in agreement with most experimental data on the SNARE system, but differ in some details that may have a functional interest, most notably in the orientation of the soluble part of R-SNARE that does not appear to be spontaneously accessible for SNARE complex formation. We also compare rat and yeast sequences of R-SNARE and find some minor differences in their behavior.

1. Introduction

Membrane-associated processes are vital to cellular function, but involve both large and complex molecular systems and dynamics on timescales from nanoseconds to microseconds.[1] They are consequently difficult to simulate and many important processes are out of the range of current all-atom molecular dynamics simulations. Coarse-graining offers an attractive solution, by reducing the number of particles to be simulated, smoothing the energy hypersurface, and allowing much larger time steps.[2–4] Earlier work has also shown that these methods can give results comparable to all-atom models in the case of protein-membrane complexes,[5–7] and processes such as membrane protein insertion[8] and assembly.[9]

Herein, we present a study of the membrane-associated R-SNARE protein, which forms part of the biologically important SNARE (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors) membrane fusion complex.[10–12] Membrane fusion is necessary for the transport of molecules such as hormones and neurotransmitters from membrane vesicles to other cellular compartments or to the surrounding medium. Neuronal SNARE proteins form a tight assembly involving four α-helices,[13] synaptobrevin and syntaxin each contributing one helix and SNAP-25 contributing two. The SNARE four-helix bundle has a characteristic layer structure with 15 conserved hydrophobic layers and one conserved central ionic layer.[13–15] The residues making up the ionic layer provide a general classification of the helices as R-SNARE (e.g. synaptobrevin, contributing an arginine) and Qa-, Qb- and Qc-SNAREs (e.g. syntaxin and SNAP-25, which both contribute glutamines).[16] Membrane fusion involves formation of the SNARE complex, possibly via a "zippering" mechanism,[17] while the transmembrane domains of synaptobrevin and syntaxin are embedded in the vesicular and plasma membranes respectively.

In the present study, we focus on the conformational and dynamic properties of the membrane inserted R-SNARE, which is thought to play a pivotal role in the fusion process.[18] We also consider species-related changes in the R-SNARE amino acid sequence that have been proposed to be at the origin of changes in fusional control.[19] We pay special attention to the tryptophan residues which play a role in anchoring the C-terminal SNARE motif to the membrane and may also be involved in controlling its activity.[19,20]

Following earlier work on the soluble part of the SNARE complex[21] and all-atom simulations of the complex anchored in two lipid bilayers,[22] we now turn to coarse-grain models to investigate processes that are too slow to be simulated at the all-atom level. We remark that although there have been some pioneering all-atom simulations of membrane-bound protein systems with roughly 120 000 atoms which reach the microsecond scale,[23] the systems we study herein would, in all-atom

[a] *Dr. M.-P. Durrieu, Dr. M. Baaden*
Institut de Biologie Physico-Chimique
Laboratoire de Biochimie Théorique, CNRS UPR 9080
13, rue Pierre et Marie Curie, F-75005 Paris (France)
Fax: (+33) 15841-5026
E-mail: baaden@smplinux.de

[b] *Dr. P. J. Bond*
Theoretical Molecular Biophysics Group
Max Planck Institute of Biophysics
Max-von-Laue-Strasse 3, 60438 Frankfurt am Main (Germany)

[c] *Prof. M. S. P. Sansom*
Structural Bioinformatics & Computational Biochemistry Unit
Department of Biochemistry, University of Oxford
South Parks Road, Oxford OX1 3QU (United Kingdom)

[d] *Dr. R. Lavery*
Institut de Biologie et Chimie des Protéines
Département des Biostructures Moléculaires
CNRS UMR 5086/IFR 128/Université de Lyon
7, passage du Vercors, F-69367 Lyon (France)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cphc.200900216.

representations, require of the order of 360 000 atoms, and multiple microsecond simulations on this scale are currently not feasible.

Computational Methodology

Coarse-grain (CG) molecular dynamics simulations were performed with the Gromacs 3 program suite,[24] following the protocol described earlier.[5,6] CG parameters for lipid molecules (dipalmitoyl-phosphatidylcholine, DPPC), Cl^- ions, and water molecules were as in ref. [2]. CG parameters for peptides were as in ref. [6]. CG peptides were generated from the corresponding all-atom structures. The yeast analogue used in simulation C1 was built by mutating the rat R-SNARE peptide using YASARA.[25] Harmonic distance restraints were applied between backbone particles separated by 7 Å or less than to mimic secondary structure hydrogen bonds. These restraints had equilibrium lengths taken from the corresponding atomic structures and force constants of 10 $kJ\,mol^{-1}\,Å^{-2}$.[6] Harmonic restraints at the junctions between the transmembrane domain (TMD) and juxtamembrane domain of each peptide were adjusted to avoid differences due to small changes in the initial conformations. The restraints on the soluble part in simulation R3 (see below) were taken from a 10 ns simulation in implicit solvent using conditions similar to those of earlier studies.[21] In the case of simulations using pre-formed lipid bilayers, these were obtained from simulations of the lipids alone in water.

All systems were solvated with CG water particles (equivalent to four water molecules), and Cl^- counterions were added to preserve overall electrical neutrality. Each system was energy minimized using the steepest-descent method to relax any steric conflicts before beginning the simulations. Simulations were carried out with periodic boundary conditions. Coulombic electrostatic interactions used a dielectric constant of 20 and shift functions were applied so that the electrostatic energies and forces were damped to zero between 9 and 12 Å. Similar shift functions were applied to Lennard–Jones interactions. The non-bonded neighbour list was updated every 10 steps. All simulations were performed using an NPT ensemble. Separate Berendsen temperature baths[26] were used for the lipid and solvent particles (323 K with a 1 ps coupling constant). Pressure was anisotropically coupled for simulations starting from random mixtures and semi-isotropically coupled for the others (1 bar with a 1 ps coupling constant and a compressibility of 5×10^{-6} bar^{-1}). A 40 fs integration time step was used.

Graphics were prepared with VMD.[27] Standard conformational analysis was carried out using Gromacs tools and locally written code. Statistical and data analysis was performed using the R statistical software package[28] and Xmgrace.[29] Sequence alignments were performed with ClustalW.[30]

2. Results and Discussion

CG simulations were carried out for R-SNARE peptides from rat (*Rattus norvegicus*) with only the trans- and juxtamembrane domains (simulation R1 and R2) or with the adjacent soluble domain (simulation R3). Comparisons were made with the yeast (*Saccharomyces cerivisea*) analogue Snc2p (simulation Y1) and with the rat Qa-SNARE peptide (also termed syntaxin, simulation C1). The corresponding sequences are given in the Supporting Information.

As shown in Table 1, we considered two methods for obtaining the peptides studied within a DPPC lipid bilayer, either starting from a random mixture of peptide, lipids, ions and water, or inserting the peptide into a pre-formed bilayer, based on the conformations obtained in earlier all-atom simulations of the membrane-embedded SNARE complex[22] (see also the Supporting Information). Both approaches led to indistinguishable final states if peptide insertion occurred. During the random starting point simulations, bilayer formation was rapid and reproducible (around 50 ns), but correct insertion of the peptide TMD did not always occur within the simulation timescales we tested. Our overall success rate was 50% for ten simulations ranging in duration from 100 ns to 1000 ns. When insertion did occur, it could be rapid (around 30 ns after the bilayer formation).

Peptide insertion into the DPPC bilayer did not cause significant perturbations judging from the bilayer thickness which had an average value of 41 Å, compared to 40 Å for simulations of a pure bilayer of identical lipids[2] and was only slightly above 38.3 Å measured experimentally.[31]

All the peptides studied adopt a similar overall position within the bilayer (Figure 1). The juxtamembrane domains lie between the lipid headgroups (defined by the glycerol, phosphate and choline CG particles) roughly 18 Å from the bilayer center, with the exception of a slight upward shift (around 1 Å) when the soluble part of R-SNARE is added in simulation R3. The middle of the transmembrane domains lie around 1.5 Å from the bilayer center, with the exception of Snc2p which is longer by three residues and settles at 3.6 Å. In passing, we remark that positions measured from the center of the bilayer are generally more reliable than those involving the lipid headgroups due to the large fluctuations that the headgroups undergo.

The overall quality of the simulations in conformational terms can be judged by comparing various amino acid positions with EPR data.[18] As shown in Figure 2, simulations R1, R2 and R3 show excellent agreement with the EPR data in almost all cases. This agreement is independent of the starting point used for the simulations. The few residue depths that disagree with the experimental data can probably be attributed to perturbations in the transmembrane position due to the side chain mutations necessary to insert spin labels. This seems to be the case for L86; from experimental data, its position is higher in the bilayer compared to the theoretically predicted position, although the results for the flanking residues agree

Table 1. Summary of the CG molecular dynamics simulations carried out.

Simulation	Peptide	Residues	Starting point
R1	Rat R-SNARE	Q73–T118	Random
R2	Rat R-SNARE	Q73–T118	Bilayer inserted
R3	Rat R-SNARE with soluble part	N27–T118	Bilayer inserted
Y1	Yeast Snc2p (TM)	A67–S115	Bilayer inserted
C1	Rat Qa-SNARE Syntaxin	R262–A288	Random

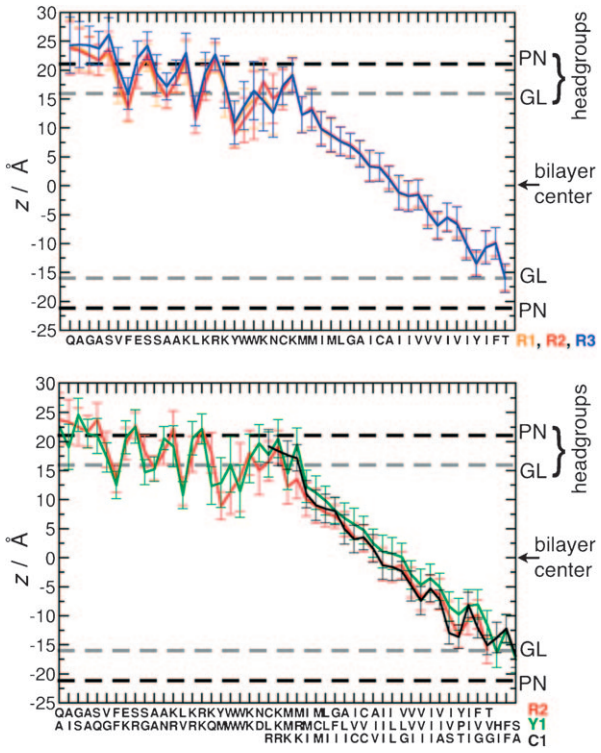


Figure 1. Top and bottom panels show the average insertion depths of R-SNARE juxta- and transmembrane domains with respect to bilayer center and headgroup regions for each residue of the peptide sequence. Error bars correspond to standard deviations. The top panel shows the results for the rat R-SNARE simulations R1, R2 and R3. Color coding is orange, red and blue for R1, R2 and R3, respectively. The bottom panel compares rat (R2) with yeast (Y1, green) and with a control simulation of a Qa-SNARE TMD (C1, black).

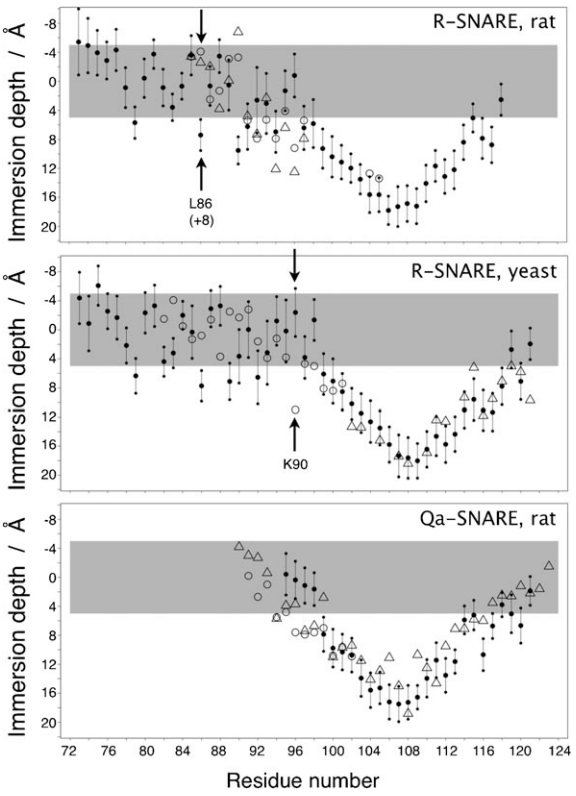


Figure 2. Simulated immersion depths with error bars corresponding to standard deviations for rat R-SNARE, yeast R-SNARE and rat Qa-SNARE from top to bottom, respectively. Experimental membrane immersion depths from the literature determined from EPR accessibility measurements are shown as open symbols. The head group region of the lipid bilayer is shown shaded. In the top panel, ○ correspond to data from ref. [20], △ to ref. [32]. In the central panel, ○ are from ref. [18], △ from ref. [33]. At the bottom, ○ are from ref. [34], △ ref. [35]. Two particular residues discussed in the text are highlighted.

with the simulations. A similar disagreement occurs for K90. This is probably an artifact, as the authors of the experiments point out, this charged residue is unlikely to be buried by 11 Å.

The tilt angle of the transmembrane domain (measured between residues at the ends of the α-helices and on the same face, M98-I114 for rat and R92-V108 for yeast) and the bilayer normal can also be compared to experiment. The simulations on the R-SNARE peptide yield an average value of 15°. This is in agreement with recent, but necessarily much shorter, all-atom simulations,[36] but smaller than several experimental measurements.[20,31] This may be due to the use of different lipid bilayers, but can also be linked to how the tilt angle is measured, as has recently been pointed out.[36,37]

In terms of sequence effects, there is little change in the overall position of the R-SNARE peptides from rat and yeast, and even the Qa-SNARE peptide occupies a similar place in the bilayer. While tilt angles are also similar for the R-SNARE peptides, those of the Qa-SNARE are clearly different and averages lie between 22° and 42°. In this case, the variability can be attributed to a glycine residue within the transmembrane domain that introduces a flexible kink into the α-helix. The only other sequence-dependent effect involves the angle between the trans- and juxtamembrane positions of the peptides. This angle is measured using the same definition of the transmembrane helix as given above and the V78-N94 vector for rat or the G72-D88 vector for yeast. This angle turns out to be noticeably larger for the rat sequence in simulations R1, R2 and R3 (average 102°–106°) than for the yeast sequence (average 90°). No calculations were made for the Qa-SNARE (simulation C1) where the juxtamembrane segment is too short.

We now turn to the analysis of movements of the transmembrane domains within the bilayer. These movements can be quantified by following the position of the C-terminal tryptophan residues of R-SNARE W91 and W92 (hereafter termed W and W′ respectively).

In the simulations starting from a pre-formed bilayer, these residues were initially placed just above the lipid headgroups. In all simulations, the tryptophans rapidly position themselves amongst the lipid headgroups. This occurs in less than 100 ns from the random starting point in simulation R1. The average positions adopted by both W and W′ agree well with both EPR[18] and fluorescence data.[20] However, if we look at the time evolution of the tryptophans, we observe slow rearrangements that are common to all simulations, irrespective of the peptide simulated.

We are able to identify three sub-states that we will term A, B and C, as illustrated in Figures 3 and 4. In sub-state A, W is above W′ (i.e. further from the center of the bilayer). In sub-state B, which dominates during our simulations, W′ is above W, with distances of 18 Å and 13 Å respectively. Figure 4 clearly shows that the opening angle between the juxtamembrane and transmembrane domains discussed above is largest for sub-state B, the A to B transition leading to a change from an acute to an obtuse angle. In sub-state C, which occurs towards the ends of simulations R2 and Y1, both tryptophans are deeply inserted in the membrane at distances around 9 Å, shifting the whole juxtamembrane domain downwards. In this conformation the opening angle decreases again, although not as much as in sub-state A, and also less markedly for the yeast sequence (see Figure 4). The average lifetimes of these sub-states appear to be several hundred nanoseconds, but they cannot be calculated accurately from our microsecond simulations. It is however noted that transitions between states A and C invariably pass through state B, which may therefore be a long-lived transition state, rather than a true sub-state.

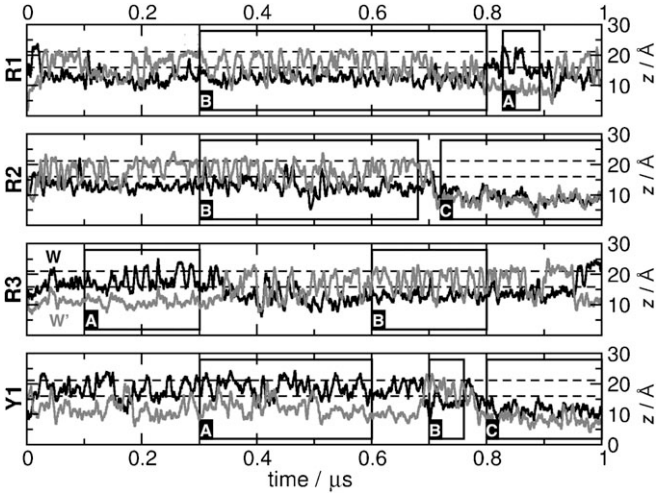


Figure 3. Tryptophan W (black curve) and W′ (grey curve) membrane insertion as a function of simulation time for systems R1, R2, R3 (W = W91, W′ = W92) and Y1 (W = W85, W′ = W86). PN and GL headgroups are indicated as dashed lines. Different intermediate states A, B and C can be identified and are shown as annotated boxes in the Figure. The distinction between these states is explained in the text.

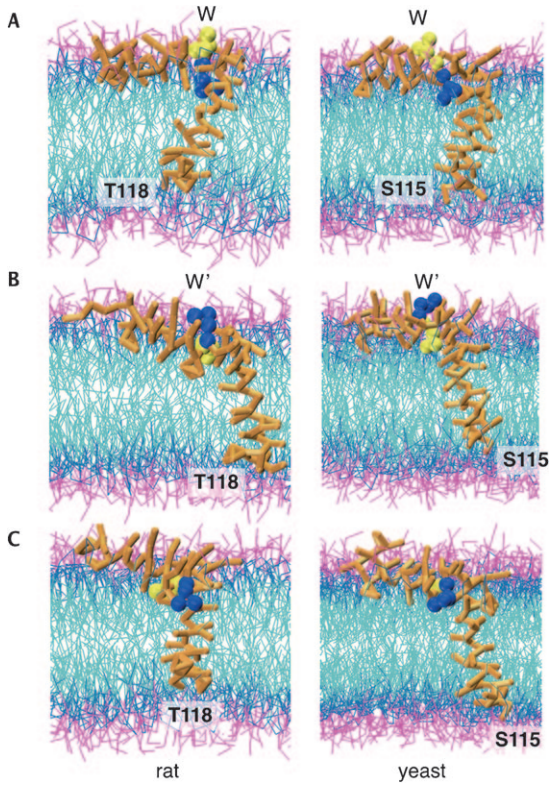


Figure 4. Snapshots of the rat and yeast simulations corresponding to the sub-states A (top), B (middle) and C (bottom), identified in Figure 3. The left column corresponds to simulations of the rat sequence. It shows selected snapshots of simulations R1, R1 and R2, from top to bottom, respectively. The right column corresponds to simulation Y1 of the yeast sequence. The peptide is shown as a licorice model and the tryptophans W and W′ are shown as ball-and-stick, colored in yellow and blue, respectively. The DPPC bilayer is shown as lines, water CG particles were omitted for clarity.

It should be noted that the lifetimes of the substates are long compared to the time for peptide insertion and this is true even taking into account that CG simulation times may have to be scaled down somewhat.[6,38] In our case this scaling does not seem to be large, since we observe peptide insertion and tilt angle equilibration on similar timescales to equivalent all-atom simulations.[37]

States A and B are compatible with fluorescence measurements on yeast R-SNARE, where a relatively exposed juxtamembrane segment is believed to be necessary to interact with t-SNAREs in forming the fusion complex. State C would fit with ideas on the neuronal R-SNARE, which is proposed to be deeply imbedded in the membrane and to require partner molecules to mediate its displacement prior to complex formation.[18]

In connection with forming the fusion complex, we note that the hydrophobic layer residues +6, +7 and +8 (i.e. residues F79, A83 and L86, positions termed *a* and *d* in the heptad repeat motif) are immersed in the bilayer below the lipid headgroups in all our simulations, even when the soluble domain of R-SNARE is present (simulation R3). This contrasts with earlier suggestions based on EPR studies[18] that these residues were accessible and available for complex formation. Our findings are however compatible with experiments on deletion mutants[19] which showed that removing the residues in question led to severely reduced R-SNARE mediated liposome fusion rates.

We lastly note that state C leads to a deeper insertion of the rat transmembrane domain than that of yeast. We attribute this effect to the aromatic residue (Y90), which precedes the first tyrosine in the rat sequence, as opposed to the polar methionine residue in the equivalent position of the yeast sequence. With respect to previous experimental studies, where a D88N mutant of yeast Snc2p was tested for increased membrane insertion,[18] we would thus suggest to examine a

double M84Y/D88N mutant. Preliminary simulations indicate that this double mutant clearly stabilizes a deep insertion of the two tryptophan residues in yeast (Supporting Information Figure 3). R92M might be another interesting mutation according to Supporting Information Figure 4.

3. Conclusions

Coarse-graining opens the door to simulations of slow processes in complex biological systems. The study of membrane embedded SNARE proteins carried out here is a good example of the possibilities of this approach. The results obtained in this first CG study of a key system for membrane fusion are in agreement with most available experimental studies, although there are some significant differences, notably concerning the accessibility of several of the hydrophobic-layer residues of R-SNARE (+6, +7 and +8) which, in our simulations, are firmly embedded in the lipid bilayer. Significantly, these simulations bring to light slow conformational transitions between substates, characterized by different dispositions of the C-terminal tryptophan residues, which are believed to play an important role in the control of fusion. Fully characterizing these substates would require simulations well beyond the microsecond scale studied here, emphasizing the multiple timescales involved in membrane-associated processes. However, on the basis of our study, we suggest one double and one triple mutant that should stabilize the deep insertion of the yeast R-SNARE. This is experimentally testable via tryptophan fluorescence measurements or EPR.

Comparisons between the rat and yeast R-SNARE sequences show what are apparently minor differences, although in the complex fusion process, it is difficult to conclude whether or not these observations (notably a change in the opening angle between the trans- and juxtamembrane domains and a change in the insertion depth of the transmembrane domain in the C sub-state) could have a functional impact.

Acknowledgements

We thank the French supercomputer centers IDRIS and CINES for providing computer resources (Projects No 041714, 051714 and LBT2411). M.P.D. acknowledges support from the French Ministry of Research. M.B. thanks ANR for support (Project ANR-06-PCVI-0025). P.J.B. is an EMBO fellow.

Keywords: biomembranes · lipids · molecular dynamics · peptides · vesicles

[1] J. H. Davis, M. Auger, R. S. Hodges, *Biophys. J.* **1995**, *69*, 1917–1932.
[2] S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B* **2004**, *108*, 750–760.
[3] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S.-J. Marrink, *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
[4] S. O. Nielsen, C. F. Lopez, G. Srinivas, M. L. Klein, *J. Phys. Condens. Matter* **2004**, *16*, R481–R512.
[5] P. J. Bond, J. Holyoake, A. Ivetac, S. Khalid, M. S. Sansom, *J. Struct. Biol.* **2007**, *157*, 593–605.
[6] P. J. Bond, M. S. Sansom, *J. Am. Chem. Soc.* **2006**, *128*, 2697–2704.
[7] P. J. Bond, M. S. Sansom, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2631–2636.
[8] K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, M. S. Sansom, *Structure* **2008**, *16*, 621–630.
[9] X. Periole, T. Huber, S. J. Marrink, T. P. Sakmar, *J. Am. Chem. Soc.* **2007**, *129*, 10126–10132.
[10] Y. A. Chen, R. H. Scheller, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 98–106.
[11] P. A. Harbury, *Structure* **1998**, *6*, 1487–1491.
[12] R. Jahn, T. Lang, T. C. Sudhof, *Cell* **2003**, *112*, 519–533.
[13] A. T. Brunger, R. B. Sutton, D. Fasshauer, R. Jahn, *Nature* **1998**, *395*, 347–353.
[14] J. A. Ernst, A. T. Brunger, *J. Biol. Chem.* **2003**, *278*, 8630–8636.
[15] W. Antonin, D. Fasshauer, S. Becker, R. Jahn, T. R. Schneider, *Nat. Struct. Biol.* **2002**, *9*, 107–111.
[16] D. Fasshauer, R. B. Sutton, A. T. Brunger, R. Jahn, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15781–15786.
[17] A. Brunger, K. Fiebig, L. Rice, E. Pollock, *Nat. Struct. Biol.* **1999**, *6*, 117–123.
[18] Y. Chen, Y. Xu, F. Zhang, Y. K. Shin, *EMBO J.* **2004**, *23*, 681–689.
[19] T. J. Siddiqui, O. Vites, A. Stein, R. Heintzmann, R. Jahn, D. Fasshauer, *Mol. Biol. Cell* **2007**, *18*, 2037–2046.
[20] D. Kweon, C. Kim, Y. Shin, *Nat. Struct. Biol.* **2003**, *10*, 440–447.
[21] M.-P. Durrieu, R. Lavery, M. Baaden, *Biophys. J.* **2008**, *94*, 3436–3446.
[22] E. Krieger, L. Leger, M.-P. Durrieu, N. Taib, P. Bond, M. Laguerre, R. Lavery, M. S. Sansom, M. Baaden in *ParCo 2007, Parallel Computing: Architectures, Algorithms and Applications*, Vol. 38 (Eds.: C. B. G. R. Joubert, F. Peters, T. Lippert, M. Buecker, B. Gibbon, B. Mohr), John von Neumann Institute for Computing, Jülich, **2007**, pp. 729–736.
[23] P. Bjelkmar, P. S. Niemela, I. Vattulainen, E. Lindahl, *PLoS Comput. Biol.* **2009**, *5*, e1000289.
[24] http://www.gromacs.org.
[25] E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, *Proteins* **2004**, *57*, 678–683.
[26] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
[27] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.
[28] R Development Core Team, *R: A Language and Environment for Statistical Computing*, **2007**.
[29] http://plasma-gate.weizmann.ac.il/Grace/.
[30] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
[31] J. F. Nagle, S. Tristram-Nagle, *Biochim. Biophys. Acta Rev. Biomembr.* **2000**, *1469*, 159–195.
[32] D. H. Kweon, C. S. Kim, Y. K. Shin, *J. Biol. Chem.* **2003**, *278*, 12367–12373.
[33] Y. Xu, F. Zhang, Z. Su, J. A. McNew, Y. K. Shin, *Nat. Struct. Mol. Biol.* **2005**, *12*, 417–422.
[34] D. H. Kweon, C. S. Kim, Y. K. Shin, *Biochemistry* **2002**, *41*, 9264–9268.
[35] Y. Zhang, Y. K. Shin, *Biochemistry* **2006**, *45*, 4173–4181.
[36] N. Taib, PhD thesis, Université de Bordeaux 1 (France), **2007**.
[37] S. Esteban-Martín, J. Salgado, *Biophys. J.* **2007**, *93*, 4278–4288.
[38] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

Received: March 19, 2009
Revised: April 30, 2009
Published online on May 28, 2009

LETTERS

X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation

Nicolas Bocquet[1]*, Hugues Nury[1,2]*, Marc Baaden[4], Chantal Le Poupon[1], Jean-Pierre Changeux[3], Marc Delarue[2] & Pierre-Jean Corringer[1]

Pentameric ligand-gated ion channels from the Cys-loop family mediate fast chemo-electrical transduction[1–3], but the mechanisms of ion permeation and gating of these membrane proteins remain elusive. Here we present the X-ray structure at 2.9 Å resolution of the bacterial *Gloeobacter violaceus* pentameric ligand-gated ion channel homologue[4] (GLIC) at pH 4.6 in an apparently open conformation. This cationic channel is known to be permanently activated by protons[5]. The structure is arranged as a funnel-shaped transmembrane pore widely open on the outer side and lined by hydrophobic residues. On the inner side, a 5 Å constriction matches with rings of hydrophilic residues that are likely to contribute to the ionic selectivity[6–9]. Structural comparison with ELIC, a bacterial homologue from *Erwinia chrysanthemi* solved in a presumed closed conformation[10], shows a wider pore where the narrow hydrophobic constriction found in ELIC is removed. Comparative analysis of GLIC and ELIC reveals, in concert, a rotation of each extracellular β-sandwich domain as a rigid body, interface rearrangements, and a reorganization of the transmembrane domain, involving a tilt of the M2 and M3 α-helices away from the pore axis. These data are consistent with a model of pore opening based on both quaternary twist and tertiary deformation.

Pentameric ligand-gated ion channels (pLGICs) are allosteric proteins regulating cellular excitability through the opening of an intrinsic transmembrane ion channel. Agonist binding to extracellular sites shifts a closed conformation into an open one, allowing ions to diffuse down their electrochemical gradient. Two agonist-free structures have been reported: the *Torpedo marmorata* nicotinic acetylcholine receptor (nAChR-*Tm*), solved by electron microscopy at 4 Å resolution[11,12], and the 3.3 Å ELIC X-ray structure[10]. X-ray structures of homologues of the extracellular domain (ECD) of nAChRs have also been described: the acetylcholine-binding proteins (AChBPs) co-crystallized with agonists and antagonists[13–15], and the ECD of α1-nAChR[16]. A key issue is now to understand how pLGICs open, select and translocate ions through the membrane. Most pLGICs undergo desensitization on prolonged exposure to agonist, complicating structural investigations of the transient open conformation. In contrast, GLIC is activated by protons but does not desensitize, even at proton concentrations eliciting the maximal electrophysiological response (pH 4.5)[5]. Here we present the first apparently open structure of this family, GLIC crystallized at pH 4.6.

The overall architecture of GLIC is similar to those of ELIC, the AChBPs, and nAChR-*Tm* (Fig. 1a). The five subunits are arranged in a barrel-like manner around a central symmetry axis that coincides with the ion permeation pathway (Fig. 1b). Subunits interact tightly through a 2,200 Å^2 interface containing charged residues and water molecules (Supplementary Fig. 8). The ECD of each subunit consists of a β-sandwich composed of five inner and three outer strands connected by loops (Fig. 1c). In eukaryotic pLGICs, the interface between ECDs holds the neurotransmitter-binding pocket[2,14]. The equivalent region in GLIC presents structural similarity despite a low sequence identity, and is well defined, notably the capping β9–β10 loop (loop C)[15]. The transmembrane domain (TMD) of each subunit consists of four helices (M1 to M4). M2 helices form the wall of the pore (Fig. 2a), bordered by rings of homologous residues, as previously established in nAChRs[6,17]: a charged E−2′, two polar T2′, S6′, and three hydrophobic I9′, A13′ and I16′/L17′ rings, that are close/homologous to α1-nAChR-*Tm* E−1′, T2′, S6′, L9′, V13′ and L16′/V17′ (the prime numbering starts at approximately the beginning of M2, at positions homologous to α1K242 of nAChR-*Tm*, Fig. 2c). M1 helices are kinked at P205 and form with M3 a second circle of helices interacting with M2. M4 helices are peripheral. Well-defined electron densities are observed in the grooves between M4 and both M1 and M3, close to residues labelled by hydrophobic probes on nAChR-*Tm*[18]. They were attributed to lipids that possibly contribute to holding M4 in its position (Fig. 1b, d, and Supplementary Fig. 3). The central ion permeation pathway consists of an extracellular hydrophilic vestibule more than 12 Å wide, followed by a funnel-shaped transmembrane pore (Fig. 2d). The M2 axes are tilted with respect to the pore axis, with outer hydrophobic side chains oriented towards the helix interfaces, and inner polar side chains oriented towards the pore.

A bundle of six detergent molecules (dodecyl-β-D-maltoside, DDM) obstructs the pore, with one detergent per monomer and one sitting on the five-fold axis, the sugar moiety being much less ordered than the aliphatic chains (Fig. 1b, d, Supplementary Fig. 3). The former interact extensively with the hydrophobic rings, shielding their side chains from the solvent with the polar heads pointing up, towards the vestibule, while the latter points its polar head down. To probe the influence of the detergent on the protein conformation, diffraction data have been collected with crystals grown in the presence of two DDM analogues (Supplementary Fig. 4). These detergents contain two bulky bromine atoms in their alkyl chain and could not bind all in the same position as described here (Fig. 1) without seriously disturbing the pore structure, for steric reasons. In both cases we observe that the protein conformation is unchanged (r.m.s. deviation <0.3 Å), whereas either the central detergent is not seen in the electron density (10,11-dibromoundecanoyl-β-maltoside) or the five central detergent-binding sites are only partly occupied (7,8-dibromododecyl-β-maltoside). Furthermore, we performed molecular dynamics simulations (20 ns) of the GLIC molecule in a lipid bilayer without DDM in the pore. We observe that its conformation is stable throughout the simulation (Supplementary Fig. 6).

[1]Pasteur Institute, G5 Group of Channel-Receptor, CNRS URA 2182. [2]Pasteur Institute, Unit of Structural Dynamics of Macromolecules, CNRS URA 2185. [3]Pasteur Institute, CNRS URA 2182, F75015, Paris, France. [4]Institut de Biologie Physico-Chimique, CNRS UPR 9080, 75005 Paris, France.
*These authors contributed equally to this work.

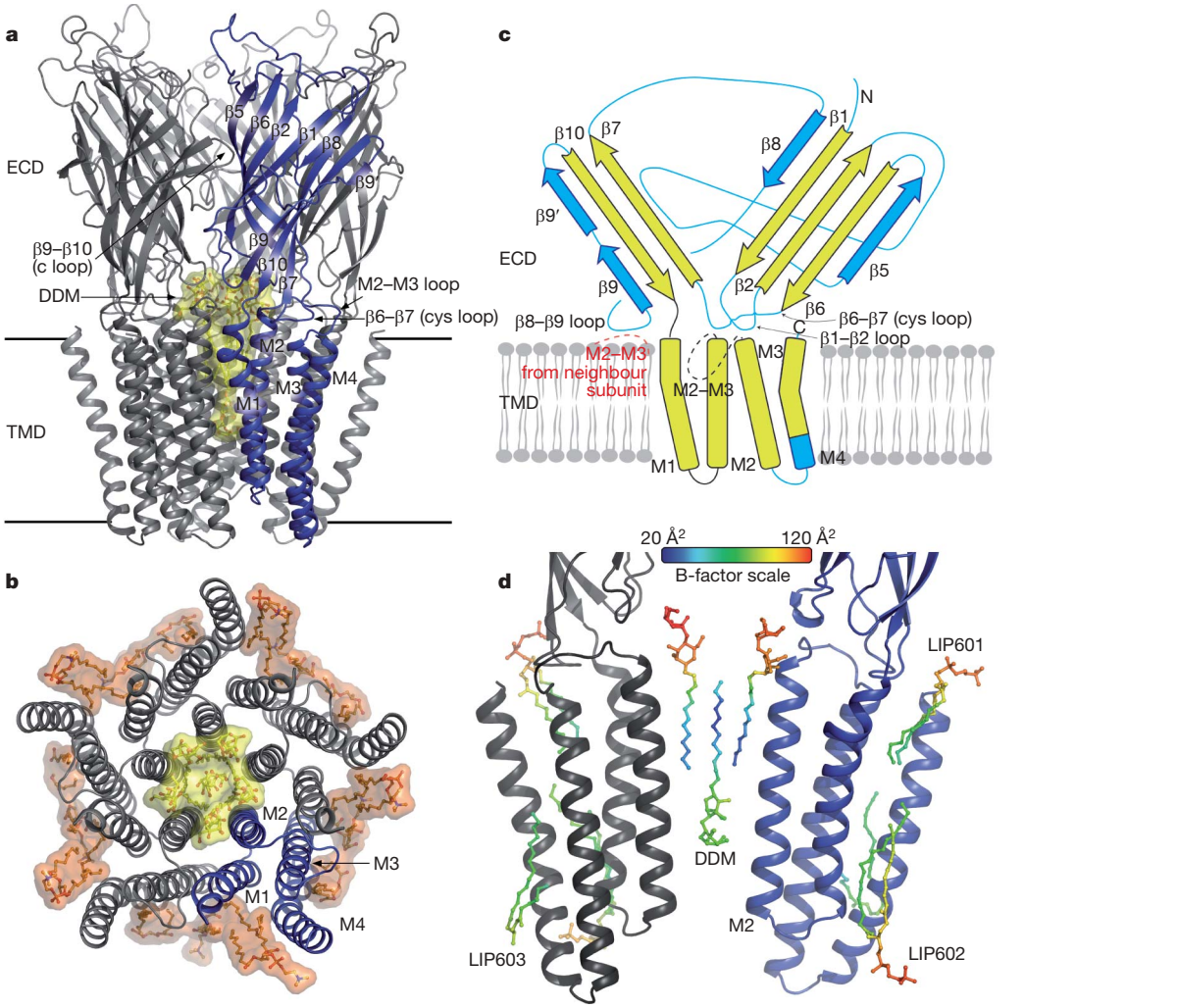


Figure 1 | GLIC structure. **a**, Ribbon representation of GLIC viewed from the plane of the membrane. DDM molecules bound in the channel are depicted as yellow sticks plus van der Waals surface. Horizontal lines represent the membrane limits. **b**, Transmembrane part of GLIC viewed from the extracellular side. The ECD is removed for clarity. Lipids are also depicted in orange. **c**, Topology of a GLIC subunit. The conserved core elements common to GLIC and ELIC are coloured in yellow. **d**, Close-up view of the TMD. Only two subunits are represented. The DDM molecules and the lipids (named LIP601/2/3) close to these subunits are coloured according to their atomic B-factor (colour scale at top).

The pore of GLIC resembles a cone with 12 Å outside and 5 Å inside diameters, sharply contrasting with that of ELIC, where outer residues L9′, A13′ and F16′ point their side chains towards the central axis, creating a 2 Å diameter hydrophobic barrier expected to prevent ion permeation (Fig. 2b). This wide opening of the pore is suggestive of an open structure. Three lines of evidence support this idea. First, GLIC was solved in the presence of a non-desensitizing agonist. Second, the GLIC pore structure is consistent with a wealth of biochemical data obtained on the open conformation of nAChRs that located the channel constriction between residues −2′ and 2′ (refs 19, 20), from measurement of cysteine accessibility[19] and channel conductance decreases following positively charged residue substitution[20]. Yet, permeability studies with organic cations suggest that this constriction is 7.5 Å wide in nAChRs[8], while it is only 5 Å wide in GLIC. Intrinsic flexibility of the side chains composing the constriction could compensate for this difference and transiently allow for accommodation of large cations. Third, molecular dynamics simulations based on the nAChR-*Tm* confirm that it corresponds to an impermeant structure, with rings 9′ and 13′ creating a hydrophobic barrier, but show that small increases (less than 2 Å) in diameter would produce a permeant channel[21,22]. The GLIC diameter is significantly wider than that of nAChR-*Tm* at these positions, further highlighting its compatibility with cation permeation.

Interestingly, mutagenesis has pointed to a key role of the −2′/2′ region in determining charge selectivity[6–9]; a ring of negatively charged residues at −1′ and an additional ring of prolines at −2′ respectively favour cationic and anionic selectivity of eukaryotic pLGICs. Thus, the −2′/2′ region was proposed to constitute the selectivity filter of the pore[7–9], an idea consistent with our structure, where the constriction is formed by the T2′ hydroxyl moieties, flanked on both sides by the S6′ hydroxyl and E−2′ carboxyl moieties. The 5 Å diameter at T2′ is too narrow to fit a fully hydrated ion, as indicated by the average distance between sodium or potassium and protein side-chain oxygen atoms (respectively 2.4 and 2.85 Å between atom centres[23]). Cation permeation would require the loss of equatorial water molecules; yet rings of polar side chains are located in the right position to transiently coordinate cations and could compensate for this loss. Regarding the pore conformation, data are thus consistent with the notion that GLIC and ELIC structures are in presumed open and closed conformations, respectively, while that of nAChR-*Tm* might be in a functionally closed but structurally nearly open conformation.

Assuming that GLIC and ELIC structures faithfully represent the open and closed forms of pLGICs embedded in a phospholipid bilayer, insight into the opening mechanism can be provided by analysing their rearrangements, even though they share only 18%

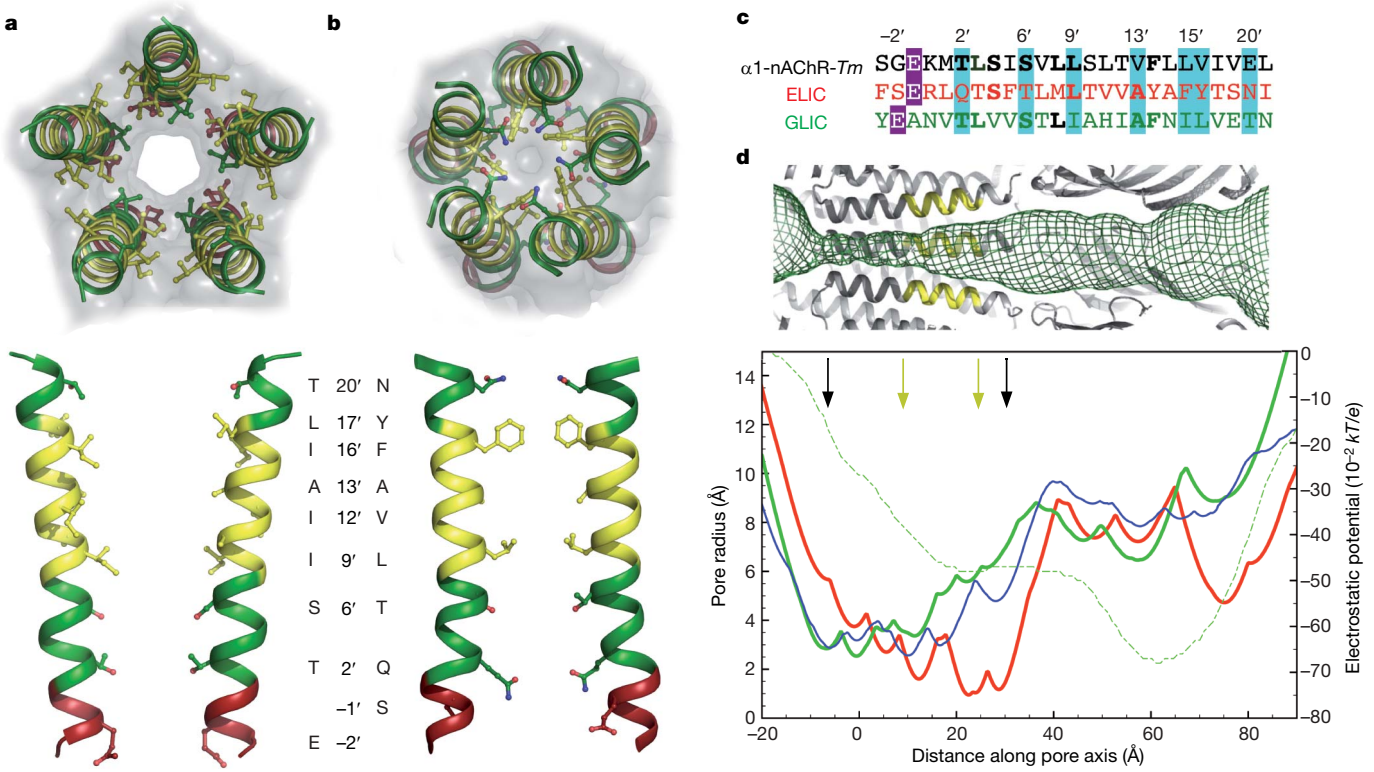


Figure 2 | GLIC and ELIC pores. **a**, Top and side view of GLIC M2 helices (top and bottom panels, respectively); helix backbones and side chains facing the pore are depicted. Hydrophobic, polar and negative residues are coloured yellow, green and red respectively. In the side view, only two subunits are shown. DDM molecules are removed for clarity. **b**, Top and side views of ELIC M2 helices. **c**, M2 sequence alignment. **d**, Pore radius for α1-nAChR-*Tm* (blue), ELIC (red) and GLIC (green) along its axis. The upper part is a mesh representation at the same scale of the GLIC channel. The electrostatic potential is represented as a dotted green line. Black and yellow arrows indicate the limits of the membrane and the hydrophobic rings, respectively.

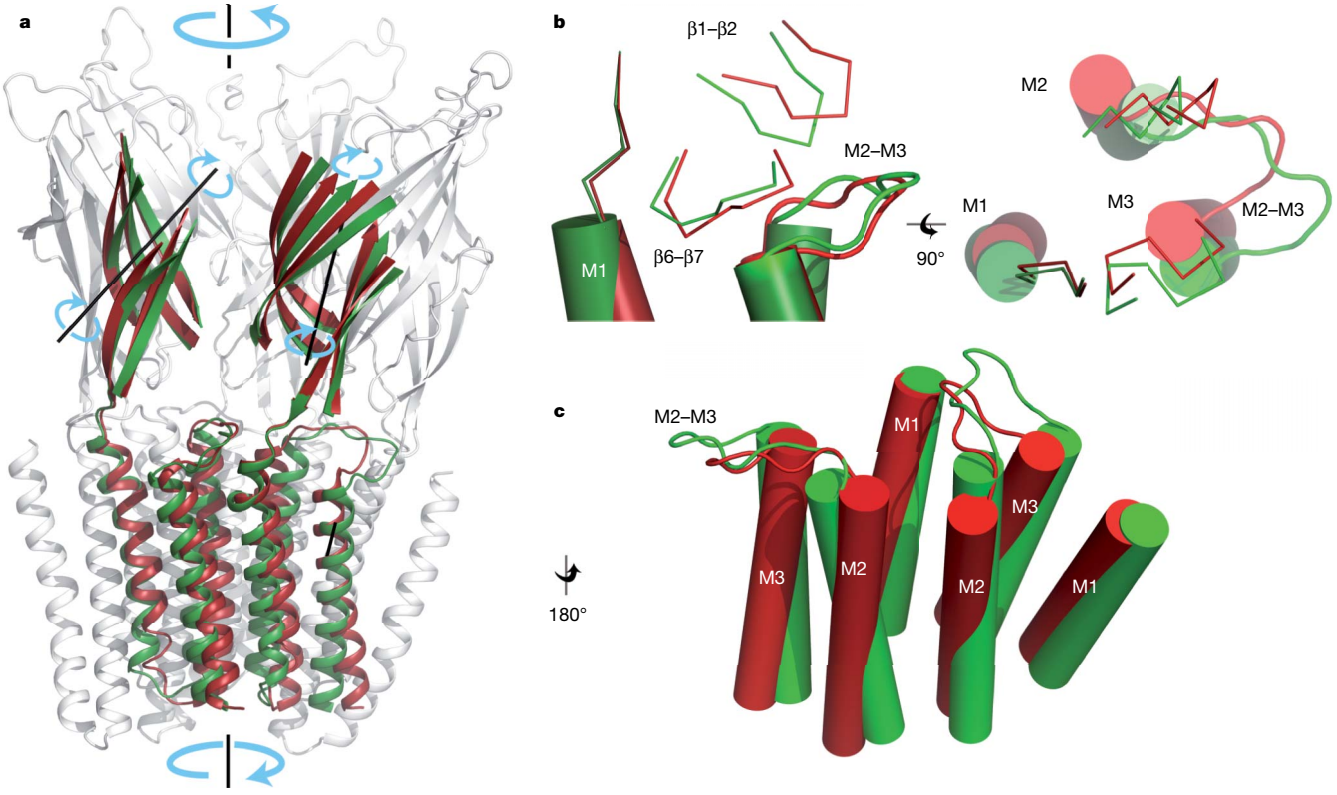


Figure 3 | Open GLIC and closed ELIC structure comparison. **a**, Side view of the structural superposition. For the two subunits in the foreground, only the common core is depicted, in green for GLIC, in red for ELIC. Other subunits are in grey. The ECD rotation axes and the twist axis are depicted. The M4 helix is omitted for clarity. **b**, Close-up view of the interface between the ECD and the TMD (side view in left panel and upper view in right panel). **c**, Close-up of transmembrane helices M1–M3 viewed from the channel.

sequence identity. Several elements can be unambiguously aligned with no gap or insertion, with identical secondary structure (Supplementary Fig. 2): helices M1, M2 and M3, and a large portion of the β-sandwich consisting of strands β1, β2, β6, β7 and β10, also well conserved in the AChBPs. These elements constitute the subunit 'common core'. Common core superimposition shows that the GLIC subunits display a quaternary twist compared to ELIC, with anti-clockwise (versus clockwise) rotation in the upper (versus lower) part of the pentamer, when viewed from the extracellular compartment (Fig. 3a). This is confirmed by normal mode analysis: the lowest frequency mode is precisely a twist mode and has by far the highest contribution (29%) to the transition (Supplementary Fig. 7a). However, we note that the first 100 lowest-frequency modes (usually the most collective ones) only explain about 50% of the transition. Other and more local movements occur: in the TMD, the outer ends of M2 and M3 of GLIC are tilted away radially from the channel axis, while the outer end of M1 is fixed. The inner ends of M1, M2 and M3 move tangentially towards the left, when viewed from the membrane (Fig. 3b). In the ECD, the core of the β-sandwich undergoes little deformation (Supplementary Table 2), but is rotated by 8° around an axis roughly perpendicular to the inner sheet of the β-sandwich (Fig. 3a), concomitant with a rearrangement of both the subunit–subunit and the ECD/TMD interfaces, regions known to contribute to neurotransmitter gating[24–27]. The latter contains the well-conserved β6–β7 and M2–M3 loops and the β1–β2 loop whose length is conserved in the pLGIC family. We observe a downward motion of the β1–β2 loop, concomitant with a displacement of the M2–M3 loop, M2 and M3 helices and β6–β7 loop towards the periphery of the molecule (Fig. 3c), thereby opening the pore.

This mechanism is different from a more local gating pathway suggested by the comparison of the conformations of the α and non-α subunits in nAChR-*Tm*[12]. It implies both a quaternary twist and tertiary deformations. Such twist to open motions, initially proposed from *ab initio* normal mode analysis of nAChRs[28], and observed for Kcsa[29], may plausibly be extended to eukaryotic pLGICs. The structural transition described here couples in an allosteric manner the opening–closing motion of the pore with distant binding sites—located at the ECD subunit interface for neurotransmitters, or within the TMD for allosteric effectors[30]—and may possibly serve as a general mechanism of signal transduction in pLGICs.

METHODS SUMMARY

GLIC was expressed as described previously[5], and solubilized/purified in 2%/0.02% DDM including an amylose affinity chromatography, two size exclusion chromatographies and thrombin cleavage of fused MBP. Crystals were grown at pH 4.6, with 15% PEG 4000, 100 mM NaOAc and 400 mM (NH_4)SCN in hanging drops. Data were collected at MX beamlines of the European Synchrotron Radiation Facility, and processed with XDS and CCP4 programs (Supplementary Table 1). The spacegroup is *C*2 with one pentamer in the asymmetric unit (Supplementary Fig. 1). Molecular replacement with Phaser using the ELIC structure gave the first density maps. Refinement was conducted with Coot and Refmac, using tight NCS restraints. The final model presents a good geometry and consists of residues 6–315 for the 5 subunits plus 6 DDM and 15 partial lipid molecules, and 115 water molecules (Supplementary Fig. 3).

Received 2 June 2008; accepted 29 September 2008.
Published online 5 November 2008.

1. Corringer, P. J. & Changeux, J. P. Nicotinic acetylcholine receptors. *Scholarpedia* **3,** 3468 (2008).
2. Lester, H. A., Dibas, M. I., Dahan, D. S., Leite, J. F. & Dougherty, D. A. Cys-loop receptors: New twists and turns. *Trends Neurosci.* **6,** 329–336 (2004).
3. Sine, S. M. & Engel, A. G. Recent advances in Cys-loop receptor structure and function. *Nature* **440,** 448–455 (2006).
4. Tasneem, A., Iyer, L. M., Jakobsson, E. & Aravind, L. Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. *Genome Biol.* **6,** R4 (2005).
5. Bocquet, N. *et al.* A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. *Nature* **445,** 116–119 (2007).
6. Imoto, K. *et al.* Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. *Nature* **335,** 645–648 (1988).
7. Imoto, K. *et al.* A ring of uncharged polar amino acids as a component of channel constriction in the nicotinic acetylcholine receptor. *FEBS Lett.* **289,** 193–200 (1991).
8. Keramidas, A., Moorhouse, A. J., Schofield, P. R. & Barry, P. H. Ligand-gated ion channels: Mechanisms underlying ion selectivity. *Prog. Biophys. Mol. Biol.* **86,** 161–204 (2004).
9. Corringer, P. J. *et al.* Mutational analysis of the charge selectivity filter of the α7 nicotinic acetylcholine receptor. *Neuron* **22,** 831–843 (1999).
10. Hilf, R. J. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452,** 375–379 (2008).
11. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423,** 949–955 (2003).
12. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346,** 967–989 (2005).
13. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411,** 269–276 (2001).
14. Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41,** 907–914 (2004).
15. Hansen, S. B. *et al.* Structures of *Aplysia* AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24,** 3635–3646 (2005).
16. Dellisanti, C. D., Yao, Y., Stroud, J. C., Wang, Z. Z. & Chen, L. Crystal structure of the extracellular domain of nAChR α1 bound to α-bungarotoxin at 1.94 Å resolution. *Nature Neurosci.* **10,** 953–962 (2007).
17. Giraudat, J., Dennis, M., Heidmann, T., Chang, J. Y. & Changeux, J. P. Structure of the high-affinity binding site for noncompetitive blockers of the acetylcholine receptor: Serine-262 of the delta subunit is labeled by [^{3}H]chlorpromazine. *Proc. Natl Acad. Sci. USA* **83,** 2719–2723 (1986).
18. Blanton, M. P. & Cohen, J. B. Identifying the lipid-protein interface of the *Torpedo* nicotinic acetylcholine receptor: Secondary structure implications. *Biochemistry* **33,** 2859–2872 (1994).
19. Wilson, G. G. & Karlin, A. The location of the gate in the acetylcholine receptor channel. *Neuron* **20,** 1269–1281 (1998).
20. Cymes, G. D., Ni, Y. & Grosman, C. Probing ion-channel pores one proton at a time. *Nature* **438,** 975–980 (2005).
21. Wang, H. L., Cheng, X., Taylor, P., McCammon, J. A. & Sine, S. M. Control of cation permeation through the nicotinic receptor channel. *PLOS Comput. Biol.* **4,** e41 (2008).
22. Beckstein, O. & Sansom, M. S. The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. *Phys. Biol.* **1,** 42–52 (2004).
23. Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: Sodium and potassium. *Acta Crystallogr. D* **58,** 872–874 (2002).
24. Jha, A., Cadugan, D. J., Purohit, P. & Auerbach, A. Acetylcholine receptor gating at extracellular transmembrane domain interface: The cys-loop and M2–M3 linker. *J. Gen. Physiol.* **130,** 547–558 (2007).
25. Grutter, T. *et al.* Molecular tuning of fast gating in pentameric ligand-gated ion channels. *Proc. Natl Acad. Sci. USA* **102,** 18207–18212 (2005).
26. Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438,** 243–247 (2005).
27. Lummis, S. C. *et al.* Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438,** 248–252 (2005).
28. Taly, A. *et al.* Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophys. J.* **88,** 3954–3965 (2005).
29. Shimizu, H. *et al.* Global twisting motion of single molecular KcsA potassium channel upon gating. *Cell* **132,** 67–78 (2008).
30. Li, G. D. *et al.* Identification of a GABAA receptor anesthetic binding site at subunit interfaces by photolabeling with an etomidate analog. *J. Neurosci.* **26,** 11599–11605 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Koehl for his program Aquasol and help with electrostatic calculations; P. Delepelaire and S. Edelstein for discussions; the staff of ESRF (Grenoble) ID14 and ID23 beamlines for data collection; facilities of the Pasteur Institute (A. Haouz for crystallogenesis, P. England for ultracentrifugation experiments, J. d'Alayer for mass spectroscopy controls and J. Bellalou for help in protein expression); and B. De Foresta (CEA, Orsay) for a gift of the two brominated DDM analogues. The latter diffraction data sets were collected at SLS and PSI (Villingen, Switzerland). We thank M. Fuchs for assistance during data collection; and the IDRIS supercomputer centre and its support staff for allocating CPU time at very short notice (project 082292). This work was supported by the Région Ile-de-France (N.B.), the Association Française contre les Myopathies, the Collège de France (C.L.P.), the Commission of the European Communities (Neurocypres project; H.N.) and the Network of European Neuroscience Institutes (ENI-NET).

Author Information Coordinates of GLIC have been deposited in the Protein Data Bank under accession number 3eam.Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.D. (marc.delarue@pasteur.fr) or P.-J.C. (pjcorrin@pasteur.fr).

Full methods

Protein expression and purification. GLIC protein was cloned as previously described[3] in a prokaryotic expression vector (Pet20b, Novagen) under an IPTG-inducible promoter. GLIC was fused with maltose binding protein (MBP) and its N-terminal signal peptide; a thrombin cleavage site was introduced between MBP and GLIC. GLIC protein was overexpressed in *Escherichia coli* C43 strain[31] cells, and expression was induced with 0.1 mM IPTG at $(OD)_{600}$=1 overnight at 20°C. Cells were harvested and lysed in a French press in buffer 1 (Tris 20mM pH 7.6, NaCl 300 mM, with proteases inhibitors from Roche); membranes were isolated by ultracentrifugation. The proteins were extracted from membranes with 2% DDM (Anatrace) under agitation at 4°C, and the solubilized fraction was cleared by ultracentrifugation. Solubilized proteins were first purified by affinity chromatography on an amylose resin. After twice 5 volume washes, in buffer 1 supplemented with 0.1% DDM and buffer 1 supplemented with 0,02% DDM (buffer 2) respectively, the MBP-GLIC fusion protein was eluted in buffer 2 containing 20 mM maltose. After concentration the protein was subjected to size exclusion chromatography on a Superose 6 10/300 GL column (GE Healthcare) equilibrated in buffer 2. Fractions of the peak corresponding to the pentameric MBP-GLIC were pooled and digested overnight at 4°C under gentle agitation, in the presence of 1 unit of thrombin (Calbiochem, bovin plasminogen-free) per 50 μg of protein and of 2 mM of CaCl2. The MBP fragment was first removed by incubation with amylose resin for 1 hour, then GLIC was finally submitted to a second size exclusion chromatography. Fractions of the peak corresponding to the pentamer were pooled and concentrated for crystallization experiments. These fractions were also subjected to analytical ultracentrifugation, which confirmed the protein is indeed a pentamer in solution (>98%).

Crystallization. GLIC was crystallized using the vapor diffusion method in hanging drops at 20°C. The concentrated (8-12 mg/ml) protein was mixed in a 1:1 ratio with reservoir solution containing 12-16% PEG 4000, 400mM (NH_4)SCN and 0.1M NaAc pH 4.6. The final pH was checked to be ~4.6 in

the drop. Crystals were obtained after 3-4 days, transferred in the mother solution supplemented with 20% glycerol for cryoprotection and flash-frozen in liquid nitrogen.

Structure determination. Several data sets of frozen single crystals were collected on beamlines ID14 and ID23 of the European Synchrotron Radiation Facility and processed with XDS[32] and CCP4[33] programs. The space group is C2 with one pentamer -as indicated by self-rotation functions- in the asymmetric unit. Using the best dataset, initial phases were obtained by molecular replacement with Phaser[34] and ELIC structure (2VL0) as a template. Clear deviations from the template transmembrane helices orientation were already apparent in the first density maps. We also took advantage of our independently obtained X-ray structure of the ECD at 2.3 Å, phased using SIRAS (Nury et al., in preparation), to validate the molecular replacement solution and place the ECD monomers into the first density maps. The model was built with Coot[35] and refined with Refmac[36] using strict NCS restraints except for residues 88, 139 to 147 and 222 for which loose side chain NCS restraints were applied, and for residues 53 to 65 (and 95 to 98 of chain B involved in crystal packing, see figure SI 1) for which no NCS restraints were applied. No NCS was imposed on DDM or lipids molecules, but it was used for waters molecules. The final model comprises 5 times 315 residues (6 are not visible either in the C- or N-termini), 6 DDM molecules, 15 partial lipid molecules and 115 water molecules. Molprobity[37] was used to validate the geometry: the overall score ranks our structure in the 86th percentile among structures of comparable resolution, the clash score is 17.4 (95th percentile) and 0.32% of residues are outliers in the Ramachandran plot.

Linear densities in 2Fo-Fc and Fourier difference Fo-Fc maps, located around the hydrophobic external surface of the protein, and perpendicular to the membrane plane, indicated the presence of bound lipids with partially ordered alkyl chains. The presence of more than 20 putative lipids can be inferred from electron density maps. Unambiguous determination of their chain lengths or polar heads chemical nature is not possible at this resolution. However, the functional importance of bound lipids is such that we chose to model the parts we could see. Therefore, 15 partial lipid

molecules were placed in their most probable conformations around the protein (see figure SI 4). Harmonic restraints were applied to them, so that they are included in the last refinement cycles but hardly move away from their initial positions.

Structural analysis. For electrostatic potential calculations, the pentamer was oriented with its symmetry axis aligned onto the (Oz) axis. Hydrogens were added and charges were assigned at pH 7 using the PDB2PQR software[38] and the Charmm22 set of partial charges, resulting in about 25,000 atoms. A program similar to APBS written by P. Koehl[39] was used to solve the linear Poisson-Boltzmann equation onto a 193x193x193 grid (the grid size is 0.7x0.7x0.9 Å). The permittivity ε was set to 2 inside the protein and 80 outside. No free ions were added. An horizontal slab of width 30 Å mimicking the membrane was added to the system, with ε=2 inside (but not through the protein channel, where ε=80). Similar profiles were observed with the sets of partial charges from Amber or Parse. Pore radii were calculated using the HOLE program[40]. Figures were prepared with PyMOL (www.pymol.org). The common core elements (Figure 1.c and Figure SI 2) for ELIC and GLIC structure comparison were established on the basis of structural and sequence homology, and then manually adjusted. They encompass sheets β1 (residue 17 to residue 28), β2 (36 to 47), β6 (100 to 111), β7 (123 to 132), β10, and the first three transmembrane helices – including the connecting loops- (183 to 278), and finally the second half of the M4 helix (296 to 313). This common core was used for the structural alignments shown in Figure 3, which was made within PyMOL using the 'super' command. Root mean square deviations (RMSDs) calculated between ELIC, GLIC and AChBP (apo-protein, pdb code 2BYN) common core elements are given in Table SI2. It is noteworthy that the superimposition result is only weakly dependent on the precise limits of the common core, and similar differences are observed when the common core is extended to GLIC loops that show reasonable conservation with ELIC, notably the β1-β2 and β6-β7 loops. This emphasizes that the differences in conformation illustrated in Figure 3 between GLIC and ELIC are robust. Normal Mode Analysis was performed using the NOMAD-Ref[41] web server

(lorentz.dynstr.pasteur.fr/index1.php). The cut-off for the elastic network model was 12 Angstroms and the models were restricted to pairs of equivalents Cα atoms, after structural alignment by SSM[42], resulting in 1210 atoms for each protein. Overlap coefficients[43] were also calculated using the NOMAD-Ref server. Similar results were obtained when calculating the normal modes either from the open or from the closed forms.

Probing the effect of the DDM binding in the pore.

As the structure shows the unusual presence of six DDM molecules inside the pore of the protein, the question arises whether these molecules influence the conformation seen in the crystal state or not. To answer this question, GLIC was extracted, purified and crystallized in the presence of two sorts of di-brominated DDM analogs kindly provided by Beatrice de Foresta from CEA, Saclay, France. These detergents are 7,8-dibromododecyl-β-maltoside (BrDM)[44] and 10,11-dibromoundecanoyl-β-maltoside (BrUM)[45] (Fig. S3a). Modeling studies predict that this modification would be sufficient to prevent full binding of the 6 detergents in the pore, for steric reasons (Fig. S4b). We collected 2 data sets at 3.1 and 3.5 Å resolution for BrDM and BrUM, respectively. Refinement (Table S1) showed that the protein model was not deformed, with an RMSD inferior to 0.3 Å with respect to the one obtained with DDM, while the DDM binding sites were only partly occupied and perhaps present some disorder as seen in the Fourier difference maps (Fig S4c,d,e). This is consistent with the idea that the pore conformation is independent on DDM binding.

Molecular Dynamics (MD) simulations. A full atomic model with all hydrogens present was built from the experimental GLIC structure. Subsequently, pKa calculations were carried out with the Yasara software[46] in order to determine the most likely protonation state of all ionizable residues at pH 4.6. Based on these calculations, we assigned the non-standard protonation states shown in table S3. The model was inserted in a fully hydrated palmitoyl-2-oleoyl-sn-glycerol-phosphatidylcholine (POPC) lipid bilayer (307 lipids, 43992 water molecules) leading to an initial system size of 130 Å × 130 Å × 150 Å dimension. The net charge of the system was neutralized with 54 Na^+ and 89 Cl^- counter-ions, achieving a salt

concentration of 100 mM. These steps were carried out within VMD[47], using the *membrane*, *solvate* and *autoionize* plug-ins. The simulations were carried out with NAMD[48] using the CHARMM27[49] force field. Two rounds of equilibration were carried out, first with the protein fixed (600 000 steps), then with restraints that were gradually released (2 500 000 steps). This was followed by 20 ns of production runs. Supplementary Figure S5 shows the model system at the end of the production run. Simulations were carried out at 310 K using Langevin dynamics. A Langevin piston algorithm was used to maintain the pressure at 1 atm. A short 10 Å cutoff was used for non-bonded interactions. Long-range electrostatic interactions were treated using the Particle Mesh Ewald method[50]. The r-RESPA multiple time step method[51] was employed with a 2 fs time step for bonded and for short-range non-bonded interactions, and a 4 fs step for long-range electrostatic forces. All bonds between hydrogens and heavy atoms were constrained with the SHAKE algorithm. All MD simulations were carried out at Babel, an IBM Blue Gene/P machine at the IDRIS Supercomputer Center in Orsay.

The GLIC structure remains in an open conformation during the whole production run (20 ns), exhibiting thermal fluctuations around an average structure in its transmembrane part. The extracellular part is more mobile. The conformational dynamics of the system are illustrated in a short movie provided as supplementary material. The conformation of the narrowest part of the pore closely matches the determined crystal structure and this similarity improves as the simulation progresses. Transiently, one can observe a constriction at the E-2' (E222) level which may result from two limitations of the simulation method used. First, the implicated E-2' residue is very mobile, and although the 20 ns simulation of this system is at the forefront of current computational possibilities, it only represents a very short timespan with respect to molecular reality. The sidechain conformations of mobile residues are thus insufficiently sampled (a more technical discussion can be found in Ref. 52). Second, the protonation state of this residue likely oscillates between protonated and deprotonated forms (predicted pKa of 3.9). However, the classical molecular dynamics method does not allow to alter the protonation state during the course of a simulation. In order to account and correct for these simulation artefacts, we

have also calculated pore profiles based on the protein backbone, assuming average volumes for each amino acid sidechain type. Both backbone- and all-atom-based pore profiles are shown in Supplementary Figure S6a. As expected, the backbone-based pore profile corrects for the limited sampling and approximate representation of dynamic protonation. It actually shows that the narrow part of the channel becomes wider in the course of the simulation. Transient increases of the pore diameter similar to those described by Wang *et al*.[22] could also be observed (Fig. S6b). On the basis of these simulation results, one can conclude that the crystallographic GLIC channel structure remains open when inserted in a biologically relevant membrane-mimetic bilayer environment. The DDM molecules that inserted upon crystallization could be removed without affecting the stability of the structure.

Table S1 - Data collection and refinement statistics

Data set	Native GLIC	BrDM GLIC	BrUM GLIC
Beamline	ESRF id23eh2	SLS PXI	ESRF id23eh1
Wavelength (Å)	0.873	1.00	0.919
Space group	C2	C2	C2
Unit cell (Å)	183.32 133.64 160.18 β=102.94°	182.67 133.69 160.12 β=102.24°	182.83 133.75 159.57 β=102.5°
Resolution (Å)	25-2.9 (3.06-2.9)	25-3.1 (3.27-3.1)	25-3.4 (3.58-3.4)
R_{merge} **(%) ***	10.1 (62.6) #	10.6 (45.0) #	12.9 (55.3)
Completeness (%)	99.5 (98.1) #	93.7 (95.4) #	98.1 (99.2)
<I/σ>	11.34 (2.05) #	7.0 (2.0) #	7.6 (2.1)
Observed / Unique reflections	321,255 (83260)	173,759 (63763)	161254 (50651)
Redundancy	3.9 (3.9) #	2.7 (2.7) #	3.2 (3.2)
Refinement statistics			
Resolution (Å)	25-2.9	25-3.1	25-3.4
Reflections working set (test set)	79080 (4186)	47413 (2506)	45361 (2415)
R_{cryst} (R_{free}) **(%)** &	21.0 (25.0)	20.8 (23.4)	21.6 (24.1)
Non-H protein (ligand) atoms	12630 (879)	12630 (0)	1260 (0)
<B-factors> (Å²)			
protein	50		
detergents	95		
lipids	92		
waters	42		
R.m.s deviations			
Bond lengths (Å)	0.012		
Bond angles (degrees)	1.12		
Molprobity scores			
Rotamer outliers (%)	7		
Ramachandran outliers (%)	0.32		
Ramachandran favored (%)	95.5		
Clashscore (percentile)	17.4 (95^{th})		
Overall quality score (percentile)	2.70 (86^{th})		

Numbers in parentheses correspond to the highest resolution shell.

* $R_{merge} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$

& $R_{cryst} = \Sigma_h | F_{obs}(h) - F_{calc}(h) | / \Sigma_h F_{obs}(h)$

Table S2 – RMSDs for structural superimpositions

	GLIC			ELIC		
	Subunit core	Pentamer core	Whole pentamer	Subunit core	Pentamer core	Whole pentamer
GLIC	-	-	-	2.02 (ECD+TMD) 0.99 (ECD)	2.72 (ECD+TMD) 1.62 (ECD)	3.20 (ECD+TMD) 2.3 (ECD)
AChBps						
2BYN	0.88	2.18	3.20	0.92	1.40	3.74
2BYP	0.81	2.15	2.88	0.92	1.39	4.04
2BYQ	0.94	2.25	3.22	0.84	1.56	3.52

Values are in Angstroms. RMSDs are calculated in PyMOL.

Table S3 – Protonation state changes adopted for the Molecular Dynamics simulation

Residue	Average pKa [#]
GLU 26	5,1
GLU 35	4,8
GLU 67	5,0
GLU 69	5,0
GLU 75	5,0
GLU 82	4,7
ASP 86	4,9
ASP 88	5,0
GLU 177	4,9
GLU 243	4,6
HIS 277	5,4

[#] Average predicted pKa for residues with more than 50% probability of being protonated by changing pH from neutral to pH=4,6. These predictions were made with the Yasara software (ref. 46) based on the GLIC crystal structure. Averages were calculated from equivalent symmetry-related residues. Several other residues are in a "twilight zone" with around 50 % probability of being protonated. These were kept in their standard protonation state in the current simulation.

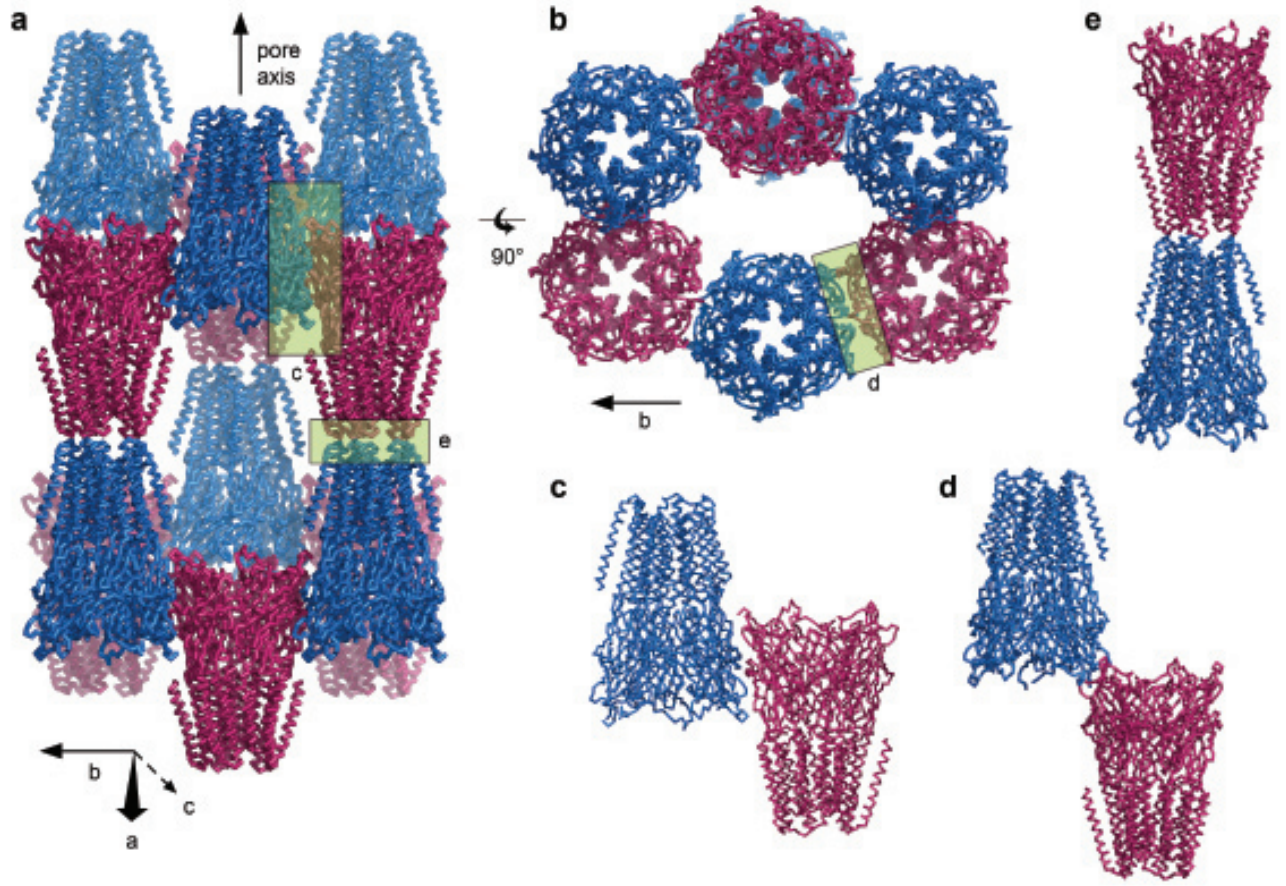


Figure S1. Crystal packing. a and **b,** Two orthogonal views of the crystal packing, with molecules depicted as magenta or blue ribbons depending on their orientation. The directions of cell axes are given. The green boxes highlight the three main pentamer-pentamer contacts, detailed in the corresponding c, d or e panels. A small molecule visible in the density but that cannot be assigned unambiguously participates in the e contact. D contact involves residues 52-56 and 95-98 of chains B of symmetry-relate pentamers and locally disrupts the non-crystallographic symmetry.

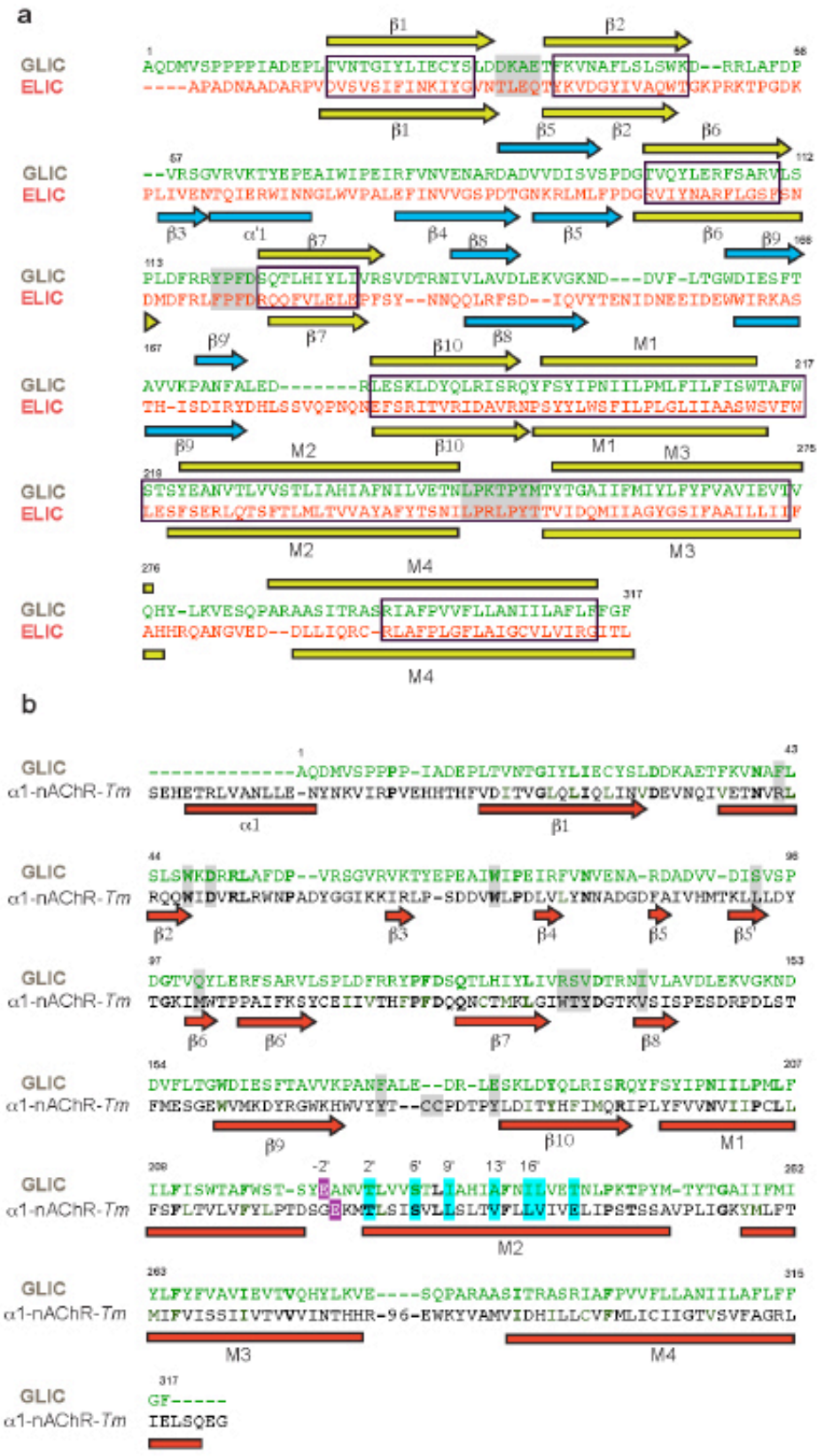


Figure S2. Complete sequence alignment. a, Structure-based alignment of GLIC (green) and ELIC (red, PDB code 2VL0). Numbering corresponds to the GLIC PDB file. Each secondary structure element is indicated, using an arrow for β-strands and a rectangle for α-helices. Common core elements are colored in yellow and delimited by black boxes while the remainder is colored in blue. The interface loops (β1-β2; β6-β7 or cys-loop; and M2-M3 loop) are shaded in grey. **b,** Sequence alignment between α subunit of the nicotinic acetylcholine receptor from *Torpedo marmorata* (black; α1- nAChR-*Tm*) and GLIC (green). Conserved residues are in bold letters. The secondary structure of α1-nAChR-*Tm*, inferred from reference 12 is in red. Residues known to contribute to the neurotransmitter binding are shaded in grey together with the homologous ones in the GLIC sequence. The important residues of the channel, belonging to M2, are colored in cyan. The cytoplasmic E(-2'), which is shifted by one position in GLIC sequence compared to its usual position, is colored in purple.

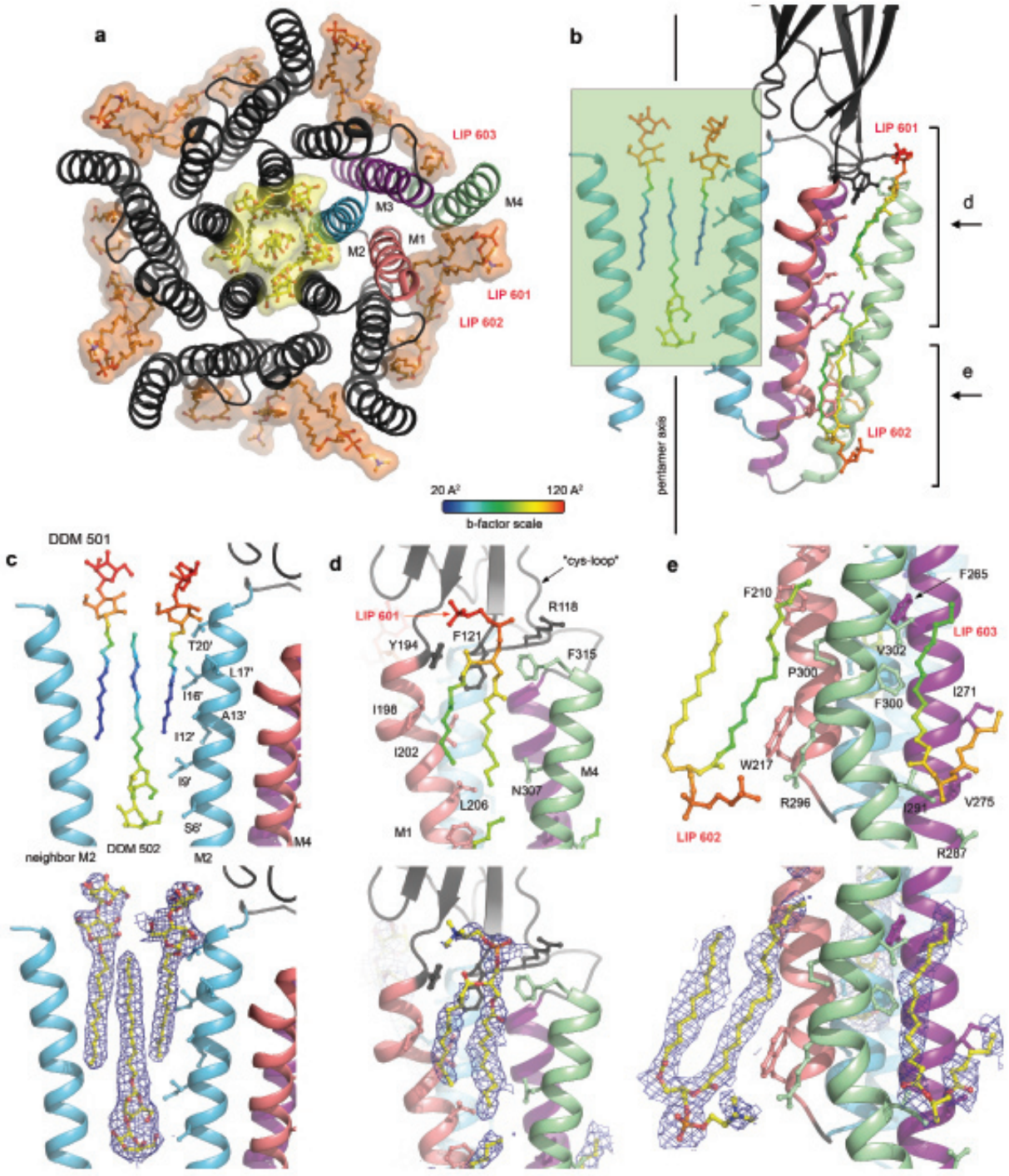


Figure S3. Detergents and lipids bound to GLIC. a, Ribbon representation of the transmembrane part of the pentamer, as viewed from the extracellular side. The orientation of the view is the same as in Figure 1a. The molecules are depicted in sticks plus transparent van der Waals surface, colored in orange and yellow for lipid (LIP 601, 602, 603) and DDM (501, 502) molecules, respectively. Only one subunit is colored with its transmembrane helices in pink (M1), cyan (M2), purple (M3) and green (M4). **b,** Side view of one subunit designed to help to locate the close-ups. The green box is represented in panel c, and the arrows indicate the directions for the view in panels d and e. **c,** Close-up views of the DDM molecules located in the channel. In the upper part the ligand colors indicate B-factors, from blue (low B-factor around 20 Å) to red (high B-factor around 110 Å²). In the lower part the 2Fo-Fc electron density map contoured at 1 σ is depicted as a blue mesh around the ligands. **d,** and **e,** Views as in panel c of the lipid molecules of the outer and inner leaflets of the membrane respectively. The protein is viewed from the side in the membrane, with M4 in the foreground.

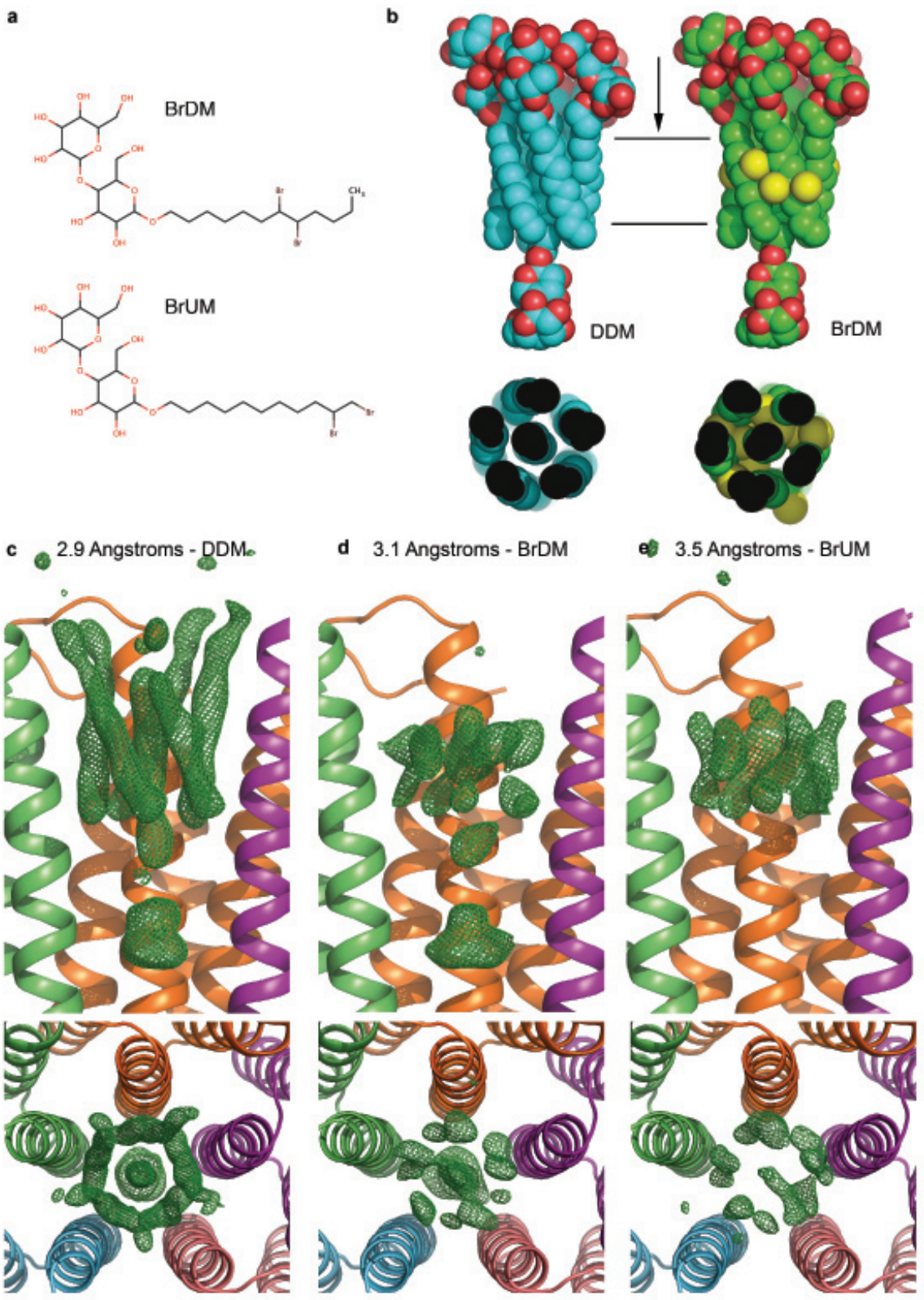


Figure S4. GLIC pore structure in the presence of DDM or brominated analogs. a, Chemical formulas of the two brominated DDMs, 7,8- dibromododecyl-β-maltoside (BrDM) and 10,11-dibromoundecanoyl β- maltoside (BrUM). **b.** Side view and cut top view (following the arrow direction) of the bundle of detergents as it is in the native structure (blue), in a van der Waals spheres representation; and of a model of what would look like the same bundle of BrDM detergents (green). The addition of the 12 bromide atoms is not possible without rearrangement of the molecules, because of steric clashes, as seen on the bottom view. **c, d** and **e.** Positive Fourier difference electron density maps, contoured at 3 σ, of the native dataset, of datasets recorded with BrDM and BrUM, in the pore region. The datasets resolution cutoffs are given. The discrepancy between the density maps supports a rearrangement of the detergent molecules within the pore, and particularly a partial occupation by less than 6 molecules combined with disorder. The difference between BrDM and BrUM might be due to the absence of the central detergent in the latter case. The density blob corresponding to the central molecule sugar's cycle was observed in all collected datasets, except in the presence of BrUM. All atoms RMSDs between protein models with brominated detergent and the native model are below 0.3 Å.

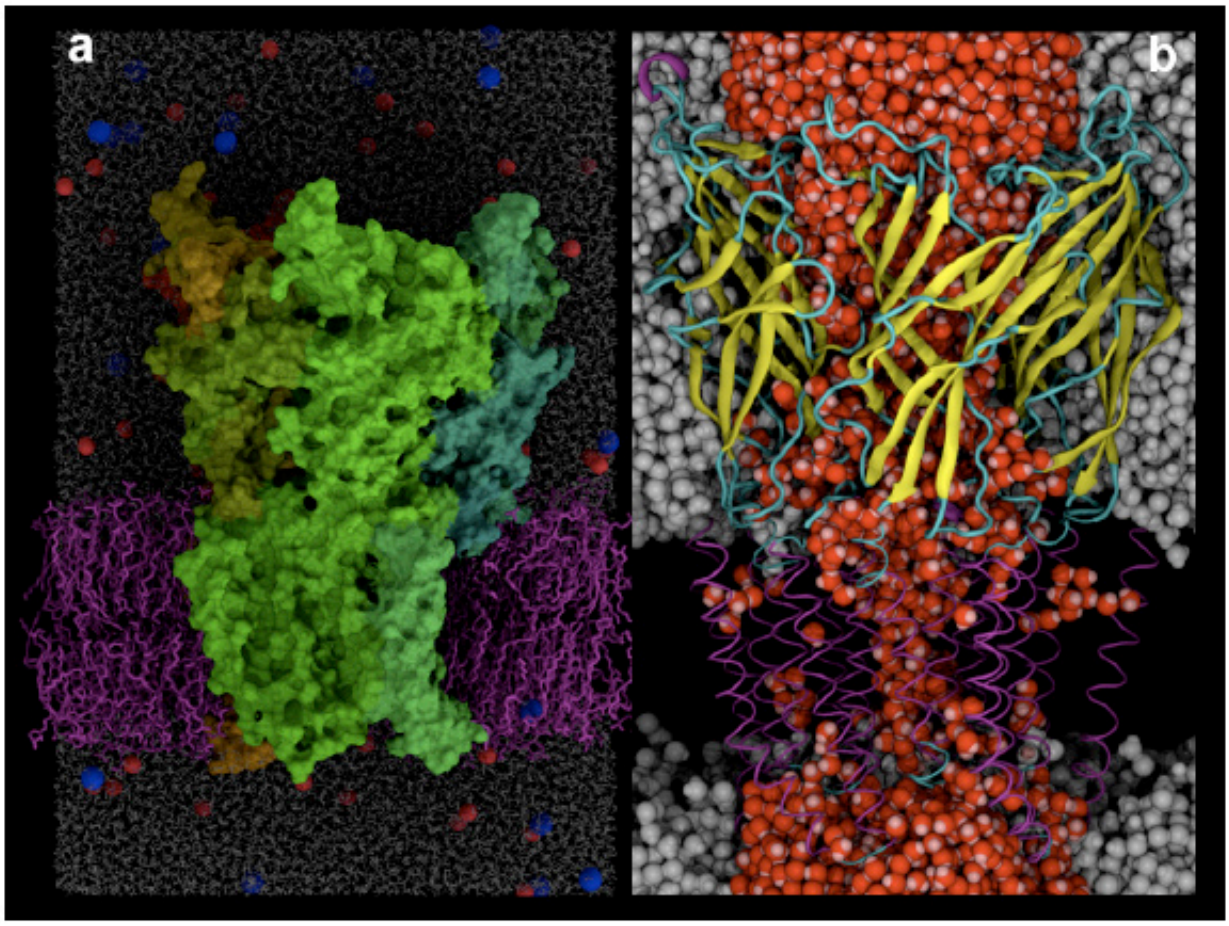


Figure S5. GLIC inserted in a fully hydrated lipid bilayer. a, membrane-inserted GLIC model at full atomic detail after 20 ns of molecular dynamics simulation. The surface of GLIC is shown colored from red to blue as a function of residue index. The POPC bilayer is colored purple, water molecules are silver, sodium cations blue and chloride anions red. Panel **b** zooms in on the continuously hydrated central channel. A cylindrical selection of water molecules is colored in red, other water molecules are white. The lipids were omitted for clarity. The protein transmembrane domain is shown as thin tubes and the extracellular part is drawn in cartoon representation. The protein is colored according to its secondary structure with helices in purple, sheets in yellow and everything else in cyan color.

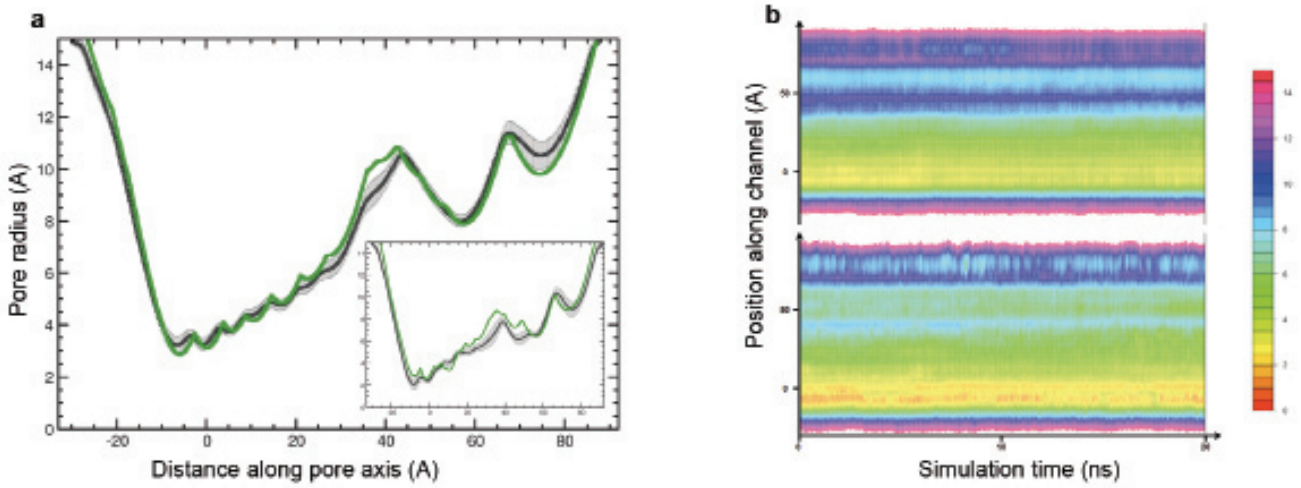
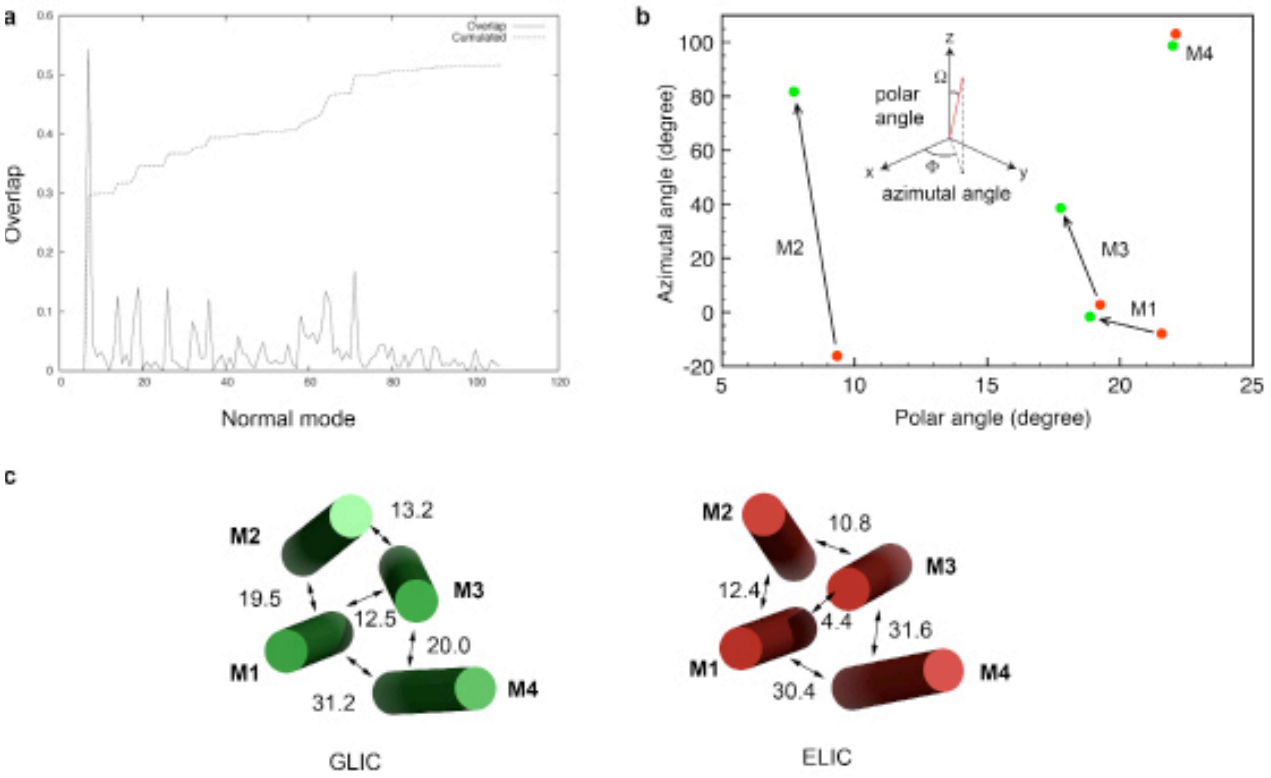


Figure S6. GLIC pore from simulation and from experiment. a, Pore radius for the GLIC crystal structure (green) and for the molecular dynamics simulation (black) along the channel axis. Origin is taken at the level of the GLIC T2' channel constriction. The molecular dynamics result is an average over the 20 ns production run with standard deviations shown as gray error bars. The main graph corresponds to backbone-derived data, where standard amino acid volumes were assumed in order to correct for simulation artefacts. The inset shows the deprecated all-atom graph. The backbone graphs were derived from the all-atom structures by preserving only backbone atoms and replacing the carbon alpha atoms by a sphere with the average volume of the amino acid. The amino acid volumes were adapted from ref. 53 and 54. **b,** time series of the effective GLIC pore radius from the simulation. Color-encoded plot of the effective pore radius along the channel axis calculated as in the previous pore profiles. Color code ranges from constricted (red) to wide open (purple). The color scale bar is shown on the right hand side of the graph. Origin of the channel axis is taken at the level of the GLIC T2' channel constriction. The molecular dynamics results over the full 20 ns production run are reproduced. The top panel corresponds to backbone-derived data, whereas the bottom panel shows the deprecated all-atom graph.



Figures S7. Structural differences between GLIC open and ELIC closed conformations. a,Overlap coefficient of the difference vectors between the open and closed forms of GLIC and ELIC (extended common core, 78% of the molecule Cα's) and the eigenvectors of the 100 lowest frequency normal modes calculated within the Elastic Network Model approximation. The first six modes are rotation-translation modes; the first non-trivial mode is mode #7, which can be visualized and seen to be a global twisting motion with the upper and lower domains of the protein rotating around the pore axis in opposite directions. **b,** Transmembrane helices polar and azimutal angles. The protein is oriented so that its symmetry axis matches the (Oz) axis. The orientation of helices axes is given in a polar vs. azimutal angles graph. The definition of these angles is included (inset). Arrows link the corresponding dots for ELIC (red) and GLIC (green).**c,** Inter-helical angle between adjacent pairs of helices in GLIC (green, left) and ELIC (red, right) structures. Values are in degrees.

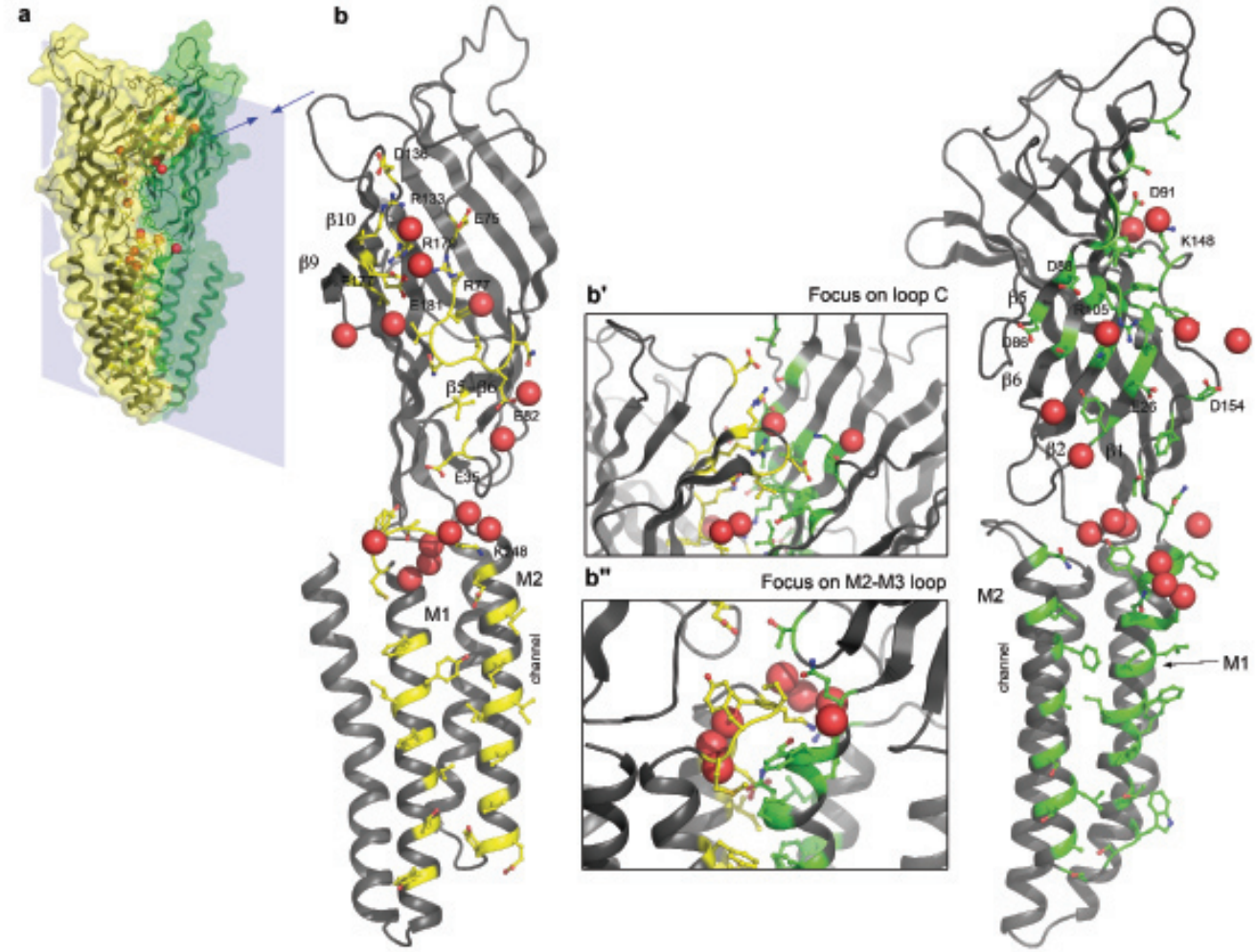


Figure S8. Water molecules and residues at subunit interfaces. a, Two neighboring subunits are represented as grey ribbons surrounded with van der Waals surface, colored in yellow for the principal subunit (the subunit containing the β9-β10 'c' loop), and green for the complementary subunit (the other one). Subunits are viewed from the membrane. The water molecules sitting near the interface are depicted as red spheres **b,** Detailed view of the side of each subunit that contribute to the interface between subunits. The orientation of the molecules is such that the interface is in the foreground, as indicated in panel a. Residues participating in the interface are depicted as green or yellow sticks. The two close-ups **b'** and **b''** represent i) the region corresponding to the neurotransmitter binding site in nAChRs, which includes one buried water molecule, and ii) the region near the M2-M3 loop (located at both subunit-subunit and ECD-TMD interfaces). Under this M2-M3 loop, there is a small cavity in which three ordered water molecules can be modeled.

References

31. Miroux, B. and Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289--298 (1996)

32. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Applied Cryst.* **26**, 795-800 (1993)

33. Collaborative Computationnal Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763 (1994)

34. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32-41 (2007)

35. Emsley, P. and Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126-2132 (2004)

36. Murshudov, G. N., Vagin A.A. and Dodson E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240-255 (1997)

37. Davis, I. W. Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B. 3rd, Snoeyink, J., Richardson, J.S., Richardson, D.C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, 375-83 (2007)

38. Dolinsky, T. J. Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N.A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, 665-7 (2004)

39. Azuara, C., Lindahl, E., Koehl, P., Orland, H., Delarue, M. PDB Hydro: incorporating dipolar solvents with variable density in the Poisson-Boltzmann treatment of macromolecule electrostatics. *Nucleic Acids Res* **34**, 38-42 (2006)

40. Smart, O. S., Neduvelil, J.G., Wang, X., Wallace, B.A., Sansom, M.S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph* **14**, 354-360 (1996)

41. Lindahl, E., Azuara, C., Koehl, P., Delarue, M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res*. **34**, 52-56 (2006)

42. Krissinel, E. and Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr*. *D* **60**, 2256-2268 (2004)

43. Delarue, M. and Sanejouand, Y-H. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J*. *Mol*. *Biol*. **320**, 1011-1024 (2002)

44. De Foresta, B., Legros, N., Plusquellec, D., Le Maire, M., Champeil, P. Brominated detergents as tools to study protein-detergent interactions. *Eur J Biochem*. **241**, 343-54 (1996)

45. De Foresta, B.,Gallay, J., Sopkova, J., Champeil, P. and Vincent, M. Tryptophan octyl ester in detergent micelles of dodecylmaltoside: fluorescence properties and quenching by brominated detergent analogs. *Biophys*. *J*. **77**, 3071-3084 (1999)

46. Krieger, E., Nielsen, J. E., Spronk, C. A. and Vriend, G. Fast empirical pKa prediction by Ewald summation. *J*. *Mol*. *Graph*. *Model*. **25**, 481-6

(2006)

47. Humphrey, W., Dalke, A. and Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-8, 27-8. (1996)

48. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-802 (2005)

49. MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616 (1998)

50. Darden, T., York, D. and Pedersen, L. Particle mesh Ewald - an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092 (1993)

51. Tuckerman, M., Berne, B. J. and Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990-2001 (1992)

52. Faraldo-Gomez, J. D., Forrest L.R., Baaden, M., Bond, P.J., Domene, C., Patargias, G., Cuthbertson, J., Sansom, M.S. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins* **57**, 783-91 (2004)

53. Counterman, A. E. and Clemmer, D. E. Volumes of Individual Amino Acid Residues in Gas-Phase Peptide Ions. *J. Am. Chem. Soc.* **121**, 4031-4039 (1999)

54. P.J. Linstrom and W.G. Mallard, Eds., NIST Chemistry WebBook,

NIST Standard Reference Database Number 69, June 2005, National Institute of Standards and Technology, Gaithersburg MD, 20899 (http://webbook.nist.gov)

Complex Molecular Assemblies at Hand via Interactive Simulations

OLIVIER DELALANDE,[1] NICOLAS FÉREY,[1] GILLES GRASSEAU,[2] MARC BAADEN[1]
[1]*Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique, CNRS UPR 9080, 13, rue Pierre et Marie Curie, Paris F-75005, France*
[2]*Institut du Développement et des Ressources en Informatique Scientifique, IDRIS-CNRS Bâtiment 506, Orsay cedex F-91403, France*

Received 2 October 2008; Revised 18 December 2008; Accepted 12 January 2009
DOI 10.1002/jcc.21235
Published online 7 April 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Studying complex molecular assemblies interactively is becoming an increasingly appealing approach to molecular modeling. Here we focus on interactive molecular dynamics (IMD) as a textbook example for interactive simulation methods. Such simulations can be useful in exploring and generating hypotheses about the structural and mechanical aspects of biomolecular interactions. For the first time, we carry out low-resolution coarse-grain IMD simulations. Such simplified modeling methods currently appear to be more suitable for interactive experiments and represent a well-balanced compromise between an important gain in computational speed versus a moderate loss in modeling accuracy compared to higher resolution all-atom simulations. This is particularly useful for initial exploration and hypothesis development for rare molecular interaction events. We evaluate which applications are currently feasible using molecular assemblies from 1900 to over 300,000 particles. Three biochemical systems are discussed: the guanylate kinase (GK) enzyme, the outer membrane protease T and the soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors complex involved in membrane fusion. We induce large conformational changes, carry out interactive docking experiments, probe lipid–protein interactions and are able to sense the mechanical properties of a molecular model. Furthermore, such interactive simulations facilitate exploration of modeling parameters for method improvement. For the purpose of these simulations, we have developed a freely available software library called MDDriver. It uses the IMD protocol from NAMD and facilitates the implementation and application of interactive simulations. With MDDriver it becomes very easy to render any particle-based molecular simulation engine interactive. Here we use its implementation in the Gromacs software as an example.

 J Comput Chem 30: 2375–2387, 2009

Key words: interactive molecular dynamics; virtual reality; guanylate kinase; SNARE complex; OmpT enzyme

Introduction

In 1994, the Sculpt program developed by Surles and collaborators permitted to change the shape of proteins interactively using continual energy minimization.[1] Since this pioneer work, the interactive molecular simulations field has been growing quietly but continuously. Initial interactive experiments using molecular mechanics techniques gave quickly rise to "guided" dynamics simulations[2,3] or steered molecular dynamics (SMD).[4,5] The interest in these methods increased with the enhancement of simulation accuracy. Thanks to technological and methodological progress, exciting new possibilities for dynamic structural exploration of very large and complex models of biological systems came within reach. In the emerging interactive molecular dynamics (IMD) approach, steering forces are applied to the molecular system in interactive time, with a chosen amplitude, direction and application point. This approach enables the user to explore the simulation system while receiving instant feedback information from real-time visualization or via haptic devices.[6,7]

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

Correspondence to: M. Baaden; e-mail: baaden@smplinux.de

Contract/grant sponsor: DEISA Consortium and EU; contract/grant numbers: FP6 508830/031513

Contract/grant sponsor: IDRIS (CNRS's National Supercomputer Center in Orsay)

Contract/grant sponsor: French Agency for Research; contract/grant numbers: ANR-06-PCVI-0025, ANR-07-CIS7-003-01

Recently, Schulten's group has carried out several convincing studies where IMD simulations were applied to macromolecular structures.[8,9] This effort led to the design of two innovative, efficient software tools greatly facilitating the process of setting up an IMD simulation: NAMD and visual molecular dynamics (VMD).[10,11] Similar projects proposing an interactive display for molecular simulations exist, such as the Java3D interface,[12,13] the Protein Interactive Theatre,[14] or others.[15,16] New autofabrication technology now even makes it possible to create tangible molecular models that can within certain limits be interfaced with molecular simulations.[17]

Recent Hardware Developments Open New Possibilities

Interactive simulations show a fast and promising evolution, owing to the wide spread of new computer hardware devices and increasing accessibility to powerful computational infrastructures. These techniques can be applied to the study of very large molecular systems comprising hundreds of thousands of atoms. Such applications are now in the reach of state-of-the-art desktop computing. This evolution was possible given the strong increase in raw computing power leading to faster and bigger processing units (multiprocessors, multicore architectures).[18] At the same time, important progress was made in simulation algorithms and methods. Currently ongoing technological developments such as graphics processing unit (GPU) or physics processing unit (PPU) computing[19–23] and the spread of parallelizable entertainment devices (PS3, Cell) with specific graphic and processing capabilities open exciting new opportunities for interactive calculations. These approaches could provide even more processing power for highly parallelizable computational problems - for instance by differentiating the parallelization of molecular calculations and graphical display functionalities. Given these developments, the range of computational methods and representations accessible to interactive simulations is bound to grow. It may soon be possible to extend the interactive approach to *ab initio* or quantum mechanics/Molecular mechanics (QM/MM) calculations. Such refined physical modeling methods increase the accuracy of the calculation. Alternatively, the precision of a given model can be improved by a better description of the molecular context and environment. This can be achieved via multiscale simulations,[24] which would largely benefit from interactivity for several reasons. First, the systems under study are often complex and the corresponding molecular models require fine-tuning of the interactions, which is difficult to achieve by noninteractive trial and error. Second, by using multiple scales it is possible to focus on the essential degrees of freedom of a complex biological system. This makes the user's interaction with the system more efficient. However, the raw increase in computer speed alone is not sufficient to grant a successful future evolution of these approaches. In addition, it is necessary to develop adapted software solutions, which are generally more efficient.[8]

Overview

In this article, we will focus on IMD as a textbook example for interactive simulation methods. We will point out how this approach can be used to explore new facets of molecular assemblies and gain an intuitive insight into the underlying physical properties of the model. Interactive simulations can thus be very useful in exploring and generating hypotheses about the structural and mechanical aspects of biomolecular interactions. This will be illustrated using a range of interesting biological applications including interactive docking. Possible applications further include method improvement and new developments. In this context, interactive simulations can facilitate exploration and adjustment of modeling parameters. Currently, remaining unsolved intrinsic problems with IMD will be discussed. In all related studies published up to date, all-atom molecular dynamics simulations were employed. In this study, we extend the application of the IMD approach to simplified modeling methods such as coarse-graining.[18,25–28] These methods allow for larger integration timesteps and thus span a much longer lapse of simulated time during the same real duration compared to classical simulations at atomic detail. Hereafter, we show that these techniques appear to be more suitable for interactive experiments. They represent a well-balanced compromise between an important gain in computational speed versus a moderate loss in modeling accuracy compared to higher resolution all-atom simulations. At the same time, coarse-grained IMD enables exciting applications thus helping to solve important biological and biochemical problems. This low resolution approach is particularly useful for initial exploration and hypothesis development with respect to rare molecular interaction events.

In the near future, the development of interactive simulations will surely allow to further increase the appeal of such an approach. One might imagine several lines of novel applications such as instantaneous data analysis, on-the-fly free energy computations or intuitive docking facilities guided by human interaction. Maybe it will even become possible to guide simulated chemical reactions interactively. In this perspective, we have tried to evaluate which applications are currently feasible. For this purpose, we have explored combinations of all-atom or coarse-grained modeling methods implemented in the Gromacs software package within IMD simulations driven via the VMD tool.[29] We have tested this approach on a large selection of complex molecular assemblies using a library prototype called MDDriver. The MDDriver software library was developed as a generic platform that can be used with any particle-based simulation code for rapid prototyping. It's purpose is to facilitate the implementation and application of interactive molecular simulations. Only minimal changes to existing software are required. Using MDDriver, it becomes easy to combine both a virtual reality-enabled visualization interface and a user's preferred calculation code, using a standardized and optimized treatment for the particle coordinate exchange. We hope that this freely available tool will spawn further development of the method and lead to a broad spread of interactive simulation methods among the computational chemistry and molecular modeling communities.

Materials and Methods

Technical Implementation

During the design of the MDDriver library, initial tests have been carried out with the VMD/NAMD ensemble. We characterized the data exchange occurring between both modules. The

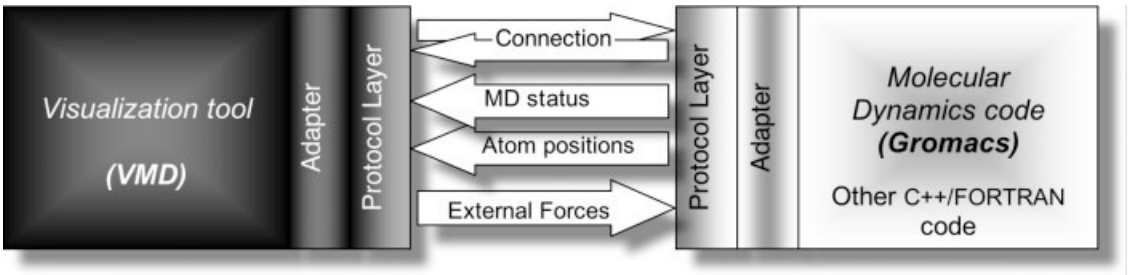


Figure 1. Data flow between a molecular visualization tool such as VMD and a generic simulation code, here molecular dynamics via Gromacs. The central arrows represent the main data flow, connected to a generic protocol layer. The subsequent adapter layer is in charge of adjusting the exchanged data for compatibility with a legacy MD application (or inversely to adapt the MD application data to the protocol layer).

communication mechanism is based on TCP/IP sockets and uses a specific "IMD protocol". Further analysis revealed that this approach is simple, robust and extensible. We have thus based the central communications within the MDDriver library on this previously developed protocol, preserving a complete compatibility with the NAMD-VMD tools. Figure 1 and Supporting Information Figure 1 illustrate how MDDriver couples a visualization and a calculation module. The IMD protocol was extracted from NAMD and encapsulated in an abstraction layer. The user part of this protocol layer consists of only five high-level subroutines shown in Supporting Information Table 1 and can potentially be called from any C/C++ or from FORTRAN code. This layer is mainly configured with the network port identifier and the coupling frequency (or data refreshment rate). As in the original IMD protocol, the MDDriver library includes integrated control functionalities such as pausing or stopping the simulation from the graphical user interface, and dynamically changing the data transfer rate. In addition, we have implemented a log message functionality that displays information on the standard output or writes it to a file using different verbosity levels. This functionality could easily be extended to record the user interaction for further analysis or replay. If the user wants to stop an interactive manipulation and pick it up at a later time, he can either pause the simulation or create a restart file at regular intervals. This functionality could be enhanced in a future version of the MDDriver software library.

An intermediate interface adapts the data from the protocol layer MDDriver library to the corresponding data structure in the application code, here Gromacs (see Supporting Information Fig. 1). This adapter is specific for a given simulation software. It is mainly in charge of transforming the physical values into adequate units, but it could also operate more complex transformations, such as changing data from a given discretization model to another one. The adapter is also in charge of redistributing and collecting the data in a parallel application. For implementing MDDriver in a given software package, the major part of the work consists of locating at which point in the simulation algorithm the data exchanges need to take place. The exchange with the visualization tool includes sending atom positions, receiving forces, and managing control events for the simulation.

Table 1. Characteristics of the Simulation Systems Used as Testbed for Validating the *MDDriver* Gromacs Implementation.

Nb.	Number of particles	PDB ID Code[a]	Resolution	Description
S1	1900	1S4Q	Coarse-grain	Guanylate kinase in water
S2	2741	1SRN	All-atom	S-peptide (residues 1 to 19) from semisynthetic ribonuclease in water
S3	6184	–[b]	Coarse-grain	Synaptobrevin-TMD in a DPPC lipid bilayer
S4	18,098	1S4Q	All-atom	Guanylate kinase in water
S5	29,965	1QD5	All-atom	OmplA monomer in an octane slab
S6	33,028	1SFC[c]	Coarse-grain	Full SNARE complex in a hydrated double DPPC lipid bilayer
S7	44,196	1QD5	All-atom	OmplA monomer in a POPC lipid bilayer
S8	55,607	1QD6	All-atom	OmplA dimer in a POPC lipid bilayer
S9	65,118	1I78	All-atom	OmpT enzyme in complex with ARRA peptide in a DMPC lipid bilayer
S10	72,466	1SFC	All-atom	SNARE core complex in water
S11	169,331	1N7D	All-atom	Low density lipoprotein receptor in water at pH 5
S12	304,564	1SFC[c]	All-atom	Full SNARE complex in a hydrated double lipid bilayer (mixed POPC/POPS)

[a]Protein Databank ID code of the starting structure used for building the model.
[b]The model was extracted from the model of Ref. 39.
[c]The full model was taken from the work described in Ref. 39.

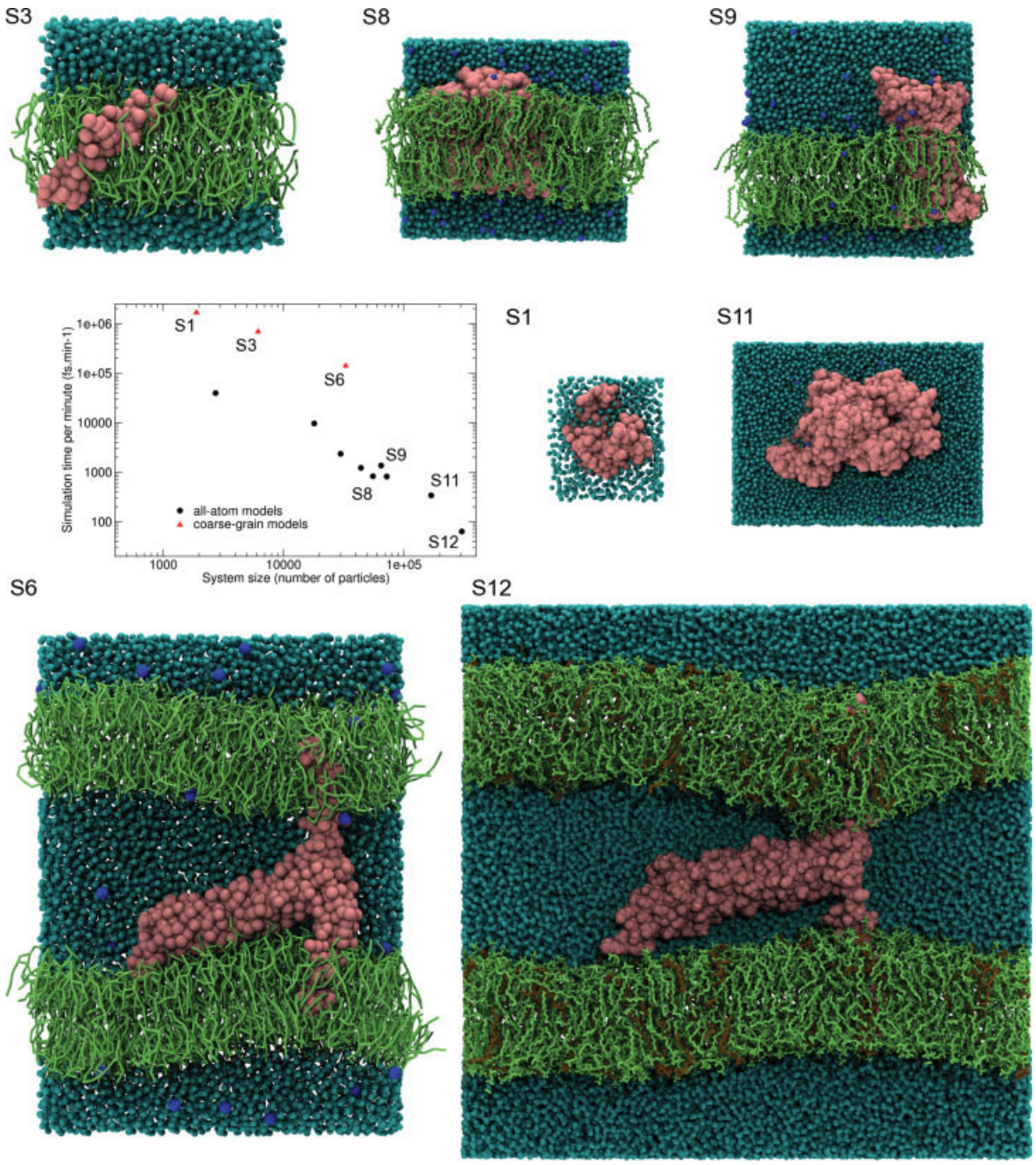


Figure 2. Illustration of some of the 12 test systems used for the validation of the MDDriver Gromacs implementation. The systems are described in detail in Table 1. The central graph indicates the distribution of system size versus the number of simulation steps per minute, a measure that describes how suitable the simulation speed is for interactive experiments. All-atom models are shown as black squares, coarse-grain models as red triangles. Selected snapshots are shown surrounding the central panel with proteins in pink, water in light blue, lipids in green, and other particles such as ions in blue. System 6 is a coarse-grained representation of system 12.

Such an adapter has been implemented for the Gromacs code and allows for keeping the changes to the initial MD code at a minimum. Only about 25 lines of the original Gromacs code needed to be changed (the Gromacs adapter which must redistribute and collect data, contains less than 500 lines). A detailed flow diagram illustrating the interactions between Gromacs and the MDDriver adapter layer is provided as Supporting Information Figure 2. The implementation of these changes in the

source code has been achieved in collaboration with the main Gromacs developers. The adapter implements control functionalities similar to VMD/NAMD. Three ways are provided to control a Gromacs simulation. The pause mode freezes the simulation until a new pause event occurs. The disconnection mode stops the interaction with the MD code while the simulation goes on—a follow-up connection is allowed. The kill mode properly stops the Gromacs code (output files are stored) and a subsequent simulation can be run with the adequate Gromacs input files. We have validated this approach in the light of prospective IMD simulations deployed on a large computing infrastructure (DEISA). These tests allowed us to point out some critical features related to performance loss due to parallelization and data exchange (the atom position transfer rate must be tuned), as well as to define the limits in the graphical rendering of a moving system in a supercomputing context. More technical details can be found in the MDDriver source code itself, in its accompanying documentation and tutorials.

The part of the MDDriver library presented in this article was developed for carrying out Gromacs IMD simulations using either all-atom or alternatively coarse-grained modeling methods. The final application is composed of a server part interfacing the simulation engine with the client part, the VMD visualization tool. The modularity of the MDDriver library permitted us to adapt the implementation template designed for the Gromacs application to several other simulation codes exchanging coordinates and forces. It is thus easy to extend any such simulation engine with an interactive interface. The use of the VMD program allowed us to benefit from an efficient interactive visualization tool with an evolved interaction interface. Furthermore, VMD includes bindings to the virtual reality peripheral network (VRPN) library[30] enabling an easy dialog with virtual reality (VR) devices. The haptic rendering of feedback forces in our coarse-grained IMD simulations relies on the same metaphors as initially implemented by the VMD developers.[11]

Validation on an Extended Benchmark

We have tested our MDDriver-based IMD approach using the implementation in the Gromacs code on a diverse selection of biological systems (see Fig. 2 and Table 1), ranging from very small to very large macromolecular assemblies. The corresponding models were constructed from the experimental structures given in Table 1; additional details are provided in the Supporting Information. The benchmark includes a small peptide in a water box, several larger systems composed of membrane protein complexes inserted into fully hydrated lipid bilayers, and a double-bilayer model of the membrane-inserted soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors (SNARE) complex with two mixed palmitoyloleoylphosphatidylcholine (POPC)/palmitoyloleoylphosphatidylserine (POPS) lipid membranes. The sizes for the simulated all-atom model systems range from 2741 to 304,564 atoms. For the coarse-grained systems, sizes range from 1900 to 33,028 particles (for further details see Table 1). In this report, we will focus specifically on three representative systems differing in their extent (from 1900 to more than 300,000 particles) and in the variety of applications that the IMD approach offers.

Large Conformational Changes

The guanylate kinase (GK) enzyme catalyzes the transfer of a phosphate group from an adenosine tri-phosphate (ATP) to a guanosine 5′mono-phosphate (GMP) nucleotide. This system has been extensively studied experimentally.[32–35] The protein shows an interesting structural behavior bearing two main conformations. An open form is observed before the docking of the GMP and ATP substrates or after the release of the GDP and ADP products. A closed form is essential for the enzymatic reaction. The mechanism of GK's closure and opening motions is still not clearly understood.[36] An interactive approach using IMD simulations does provide an innovative and dynamic structural exploration method shining light on these major conformational changes using computational models.

Interactive Docking

Previous molecular modeling studies (unpublished observation),[37,38] have thoroughly investigated the catalytic site of the OmpT membrane protein. This enzyme induces the hydrolysis of di-basic amino acid motifs such as in the ARRA peptide or in a potential autoproteolytic site located on one of its loops. Here we adopt the IMD method to examine the interactive docking of either a substrate molecule or of the L4 loop depicted as autoproteolytic site. Such an investigation does provide important complementary data to previous classical molecular dynamics (MD) simulations.

Protein Mechanical Properties and Lipid–Protein Interactions

The SNARE complex is a proteic assembly involved in the ubiquitous membrane fusion mechanism. This system seems particularly relevant for the study of protein mechanical properties and strong lipid–protein interactions.[31,39] Prior to fusion, the four-helical bundle constituting the SNARE complex is anchored to two opposing membranes that are to be fused. This is achieved via the insertion of two transmembrane domains (TMDs) belonging to two distinct helices of the bundle (syntaxin and synaptobrevin). A model of this state presents an interesting structural entity exhibiting crucial patterns specific to protein–protein and lipid–protein interactions. Using a force feedback device in IMD simulations of this system, we experienced a significantly enhanced intuitive perception of the mechanical behavior of our model of the SNARE complex via haptics.

Hardware and Simulation Protocols

All IMD simulations were performed on a 3GHz dual quad-core MacPro Apple computer using Gromacs (www.gromacs.org)[40] and either all-atom or coarse-grain models.[25–27] Additional performance benchmarks were carried out on the machine zahir, an IBM Regatta Power 4 computer at the IDRIS supercomputing centre in France. The molecular scene display was rendered using an NVIDIA Quadro FX 5600 graphics card. Stereo rendering was achieved via a Crystal Eyes stereovision device. The user interaction occurred via a Phantom Omni haptic device by Sensable Technologies.

Where available, the same simulation protocols as described in the literature were used for our interactive experi-

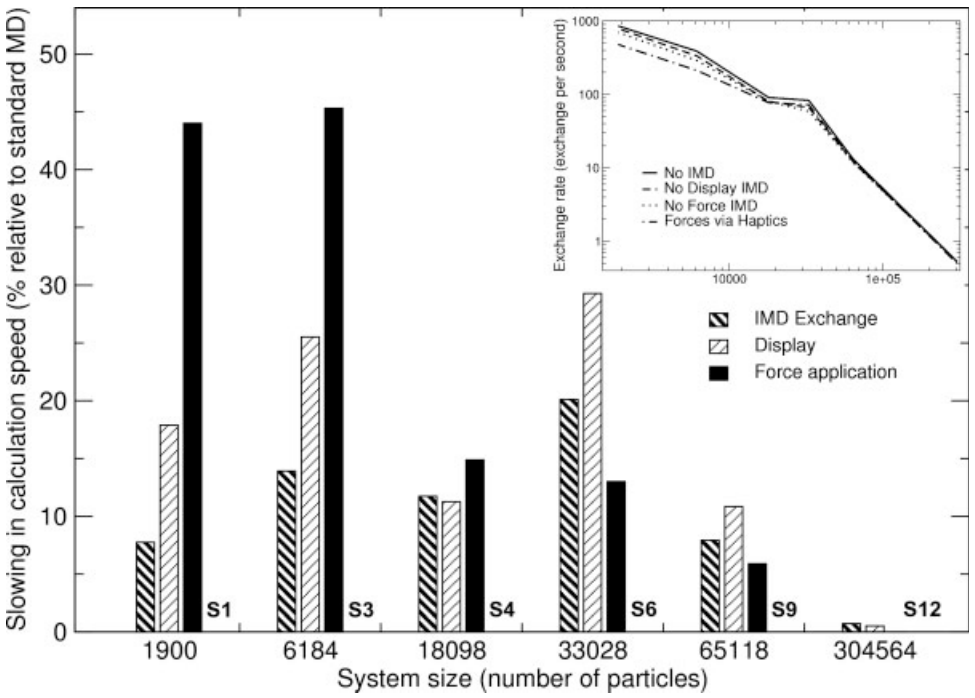


Figure 3. Inset: Benchmark results showing exchange rate in number of coordinate sets per second relative to system size (number of particles). Main plot: The slowdown in the exchange rate compared to classical noninteractive MD is characterized as percentage. Detailed results separated into effects induced by the connection to the MDDriver library (IMD Exchange), the execution of a visualization tool (Display) or the impact of force application. The effect of the user interaction on calculation speed decreases with increasing system size.

ments.[31,37,39,41,42] For unpublished simulations, such as those of the GK enzyme, similar MD protocols were used, based on commonly established standards. The most critical simulation parameters that can directly impact on the simulation performance are the simulation timestep, the neighbor list update frequency and the treatment of long-range electrostatics. The simulation timestep was 2 fs for all-atom and 40 fs for coarse-grain simulations. The neighbor list update frequency was set to every 10 simulation steps. In the coarse-grained simulations, electrostatic interactions were treated with the standard Coulombic potential with a dielectric constant of 20 and a 12 Å cutoff. Shift functions were applied so that the energies and forces vanish at the cutoff distance. For all-atom simulations, a particle mesh Ewald (PME) algorithm was used with a direct space cutoff between 10 and 18 Å. In general, the interactive simulations were run with the same parameters as previous noninteractive ones, without any tuning to enhance interactivity. Generation of the solvent or membrane models and construction of the coarse grain models is described in Supporting Information and Methods.

Results

Performance

We will briefly comment on the performance obtained with the MDDriver library implementation of the Gromacs code in the context of a desktop use. The data transfer rate for coordinates, simulation status and user-applied forces between calculation and visualization modules essentially depends on the size of the simulated system (see Fig. 3, inset). The time spent on handling user forces particularly affects IMD performance for small systems. It depends on the ratio of selected to total particles, which affects the amount of exchanged data. The display rendering step represents the major limitation in the case of larger systems, despite the fact that all measurements were carried out with the lightest representation of the scene appropriate for a given system. In most interactive simulations, we use a cartoon representation for displaying the global structure of the protein in combination with a space-filling display for active sites, binding pockets or substrate residues. With our desktop installation we have noticed two major inconveniences related to latency issues affecting user interaction. One issue originates from the pick and unpick events triggered by the haptic device when the user selects an atom. Another latency issue occurs even in the absence of interaction with the molecular scene. It is a direct effect of the MD calculation protocol, and corresponds to a slow-down every time the neighbor list used for the calculation of the pairwise interactions is recomputed by Gromacs. When the frequency of neighbor list calculation is set to every 10 simulation steps, a commonly used value, one observes a substantial decrease in the IMD exchange rate at precisely this interval (See Supporting Information Fig. 3). Obviously, these latencies in the

interactive molecular scene display do not exist in a classical postsimulation trajectory rendering. In an IMD context, they can have severe effects on the simulation, temporarily freezing the interactive session and perturbing the haptic rendering or more drastically threaten simulation stability.

Several test calculations were also deployed in a supercomputing context using a large computing infrastructure. Fundamentally, the Gromacs IMD simulations using MDDriver showed similar performance characteristics as observed in the context of a desktop computing use. As expected, it was possible to increase the size of the largest system that can still be treated in interactive time. This confirms the robustness of the MDDriver library when coupled to parallelized applications in a supercomputing environment. Generally, one can distinguish three situations (see Fig. 3). For very large systems (exceeding 60,000 particles), the scalability of the code performing the molecular simulation is the dominant performance bottleneck. Between 10,000 and 60,000 particles, network exchange, graphical rendering or user interaction slow down the interactive simulation equally. Below 10,000 particles, the interaction component becomes the biggest limitation, followed by the display component.

Applied Force via Springs

Interaction with the simulation was implemented as in the VMD tool,[11] allowing the user to impose forces on particles and experience a tactile feedback. The haptic device is used to control the direction of the forces applied to selected atoms and to adjust the amplitude of the force. This interaction method contains two stages. The first stage comprises the selection of a single atom or a set of atoms using a 3D tool attached to a haptic device and its buttons. In a second stage, the model described in[9] is used to compute the forces applied to the selected atoms. The main idea of this paradigm is to link the selected atoms and the 3D haptic tool with a spring. Instead of providing direct haptic rendering of forces as computed in the simulation, the force feedback only depends on the spring length, which in turn is influenced by the way the simulation reacts to the applied force. A strong point of this approach is that a low physical simulation framerate (5–10 Hz) does not cause instabilities and does not affect the quality of the haptic feedback. The reason is that the spring length is computed in haptic time, at 300 to 1000 Hz. A new aspect with respect to the haptic feedback concerns the user's perception of forces in a coarse-grain simulation.

Applied Force Parameters

Parameters related to the applied force were previously discussed by the VMD-NAMD developers. The feedback scale factor $\boldsymbol{S}_{\mathbf{fb}}$ and the applied force scale factor $\boldsymbol{S}_{\mathbf{f}}$ can be changed in interactive time. These parameters are the most affected ones upon lowering the resolution of the molecular representation from an all-atom to a coarse-grained level. The origin of this effect lies in the concomitant modification of the simulation timestep and of the potential energy surfaces.[26–28] The user can adjust the $\boldsymbol{S}_{\mathbf{fb}}$ and $\boldsymbol{S}_{\mathbf{f}}$ parameters. This is often required during an ongoing IMD simulation to adapt to the nature of the system under study. Tuning $\boldsymbol{S}_{\mathbf{f}}$ is a necessary but quite delicate task, ideally resulting in improved force efficiency. Yet in the case of an overestimation, this adjustment may induce strong structural distortions and threaten simulation stability. This point will be discussed in more detail below. Furthermore, proper force scaling is tied to a precise and judicious atom selection. This selection defines the application point of the force and is crucial to prevent irreversible or damageable losses of structural integrity. By performing low-resolution coarse-grain simulations rather than more detailed all-atom ones, while choosing well adapted $\boldsymbol{S}_{\mathbf{f}}$ and $\boldsymbol{S}_{\mathbf{fb}}$ parameters, an improved user perception of the haptic feedback is achieved.

Scientific Applications

We carried out a range of interactive molecular simulations using the MDDriver library to establish a dialogue between the Gromacs MD code and the VMD visualization tool. This approach allowed us to explore a large scope of biological and molecular interactions in atomistic and low-resolution models in an intuitive way. The biochemical events that have been examined interactively deal with a broad range of scenarios, including the study of protein–lipid as well as lipid–lipid interactions, enzyme-substrate recognition, macromolecular assembly and disassembly (see Fig. 4). Among all these applications, we will point out a few key findings, illustrating how the IMD approach can lead to the discovery of new and important structural and mechanical properties and thus generates new hypotheses for further validation. We will present an overview of potential applications to the field of structural biology and biochemistry. The interactive simulation approach provides access to a new dimension of complementary intuitive information about computer models of molecular assemblies. We will particularly insist on new features emerging from coarse-grain IMD simulations.

Exploring the OmpT Enzyme's Catalytic Site and Its Substrate Docking

One possible approach for characterizing the key residues for substrate docking in the OmpT enzyme is to observe the relaxation of its active site after a perturbation by the user. This experiment was carried out on the membrane-inserted enzyme using an all-atom model comprising 65,118 atoms. After a short (in user time) force application to guide the ARRA substrate inside the catalytic site of OmpT, the positively charged substrate is spontaneously recognized by surface residues that are part of the active site (D83, D85, and D97). No such recognition is observed deeper inside the enzyme, where the user has to precisely drive basic amino acids to create stable interactions with residues such as E27, D208, and D210. Substrate recognition and long-range interactions do not seem to involve the two basic amino acid residues on loop L4 which could play an autoproteolytic role. No specific conformational change of this loop is observed during docking of ARRA with a variety of user chosen interactive approaches. An alternative experiment is to start from a configuration where the substrate is already docked to the enzyme. The user then perturbs this structure by imposing small displacements. In this case, the user can observe a qualitative hierarchy among the contacts that are created between the

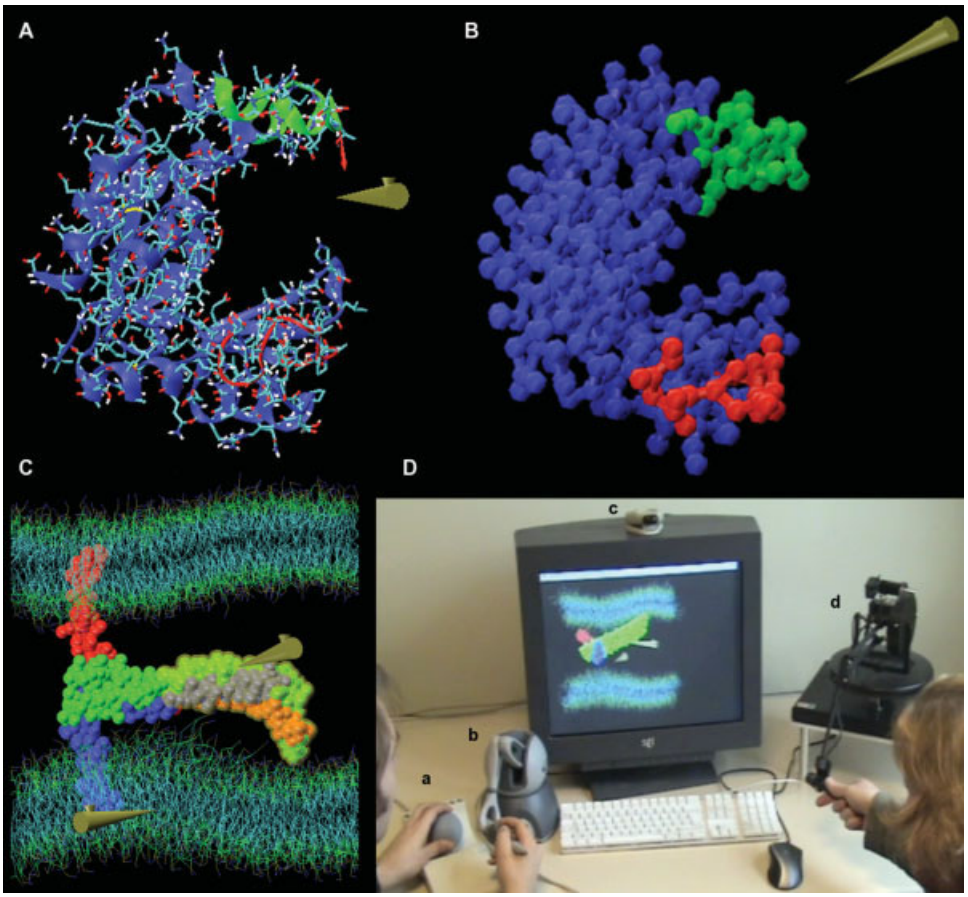


Figure 4. Scientific applications of IMD simulations to structural biology. The top row illustrates how one can induce large conformational changes by exploring the opening and closure mechanism of the U-shaped GK enzyme. IMD simulations were carried out using both an all-atom (Panel A) and a coarse-grain (Panel B) representation of GK. The enzyme is shown in blue with particularly interesting regions highlighted in green and in red color. In Panel A, a red arrow represents applied forces. The bottom row shows how MDDriver was used to probe protein mechanical properties and explore lipid–protein interactions in the SNARE system. Panel D is an example for *in situ* collaborative work where two scientists simultaneously interact with a given simulation. This Panel also provides an overview of the experimental setup used for IMD simulations with several VR devices such as a spaceball (a), a portable haptic device (b), a stereoview rendering equipment (c) and a second haptic stylus (d).

substrate and the different catalytic residues. This hierarchy can be further highlighted by an instant display of hydrogen bond formation and disruption. Strongly pulling the substrate out of the catalytic site leads to nonreversible changes. Concerning *apo* OmpT, we investigate several possibilities for the approach of the autoproteolytic loop L4 towards the catalytic site. It is possible to induce several conformations where the loop resides inside OmpT's active site after crossing a potential barrier. These experiments lead to a variety of new stable states. Instant visual feedback of favorable hydrogen bonds and system energies further helps the user to assess the most favorable hypothesis for such a conformational change. Steering the sidechain of arginine R218 close to the E27-D208-D210 triad seems sufficient to drive the whole loop into the site. This is not the case when a force is applied on lysine K217 towards the D83, D85, and D97 residues, or when a constraint is applied on the full loop backbone. Conducting such tests on a molecular system provides instant feedback. Based on the tactile haptic feedback, the user is intuitively able to assess whether the effect of the move that has been imposed is likely to lead to reversible or irreversible changes.

Inducing Closure and Opening of the Guanylate Kinase Enzyme

The closure mechanism of the U-shaped GK molecule consists of bringing its GMP and ATP binding sites closer together. It is an essential mechanism for the enzymatic reaction and has been investigated via IMD simulations based on both all-atom and coarse-grain methods. The all-atom model comprises 18,098 atoms and the coarse-grain representation consists of 1900

beads. Prospective experiments were carried out on the enzyme's *apo* form using coarse-grain simulations. This approach permitted an efficient and broad exploration of diverse ways to close the enzyme. As a guideline, we tried to reach a closed conformation similar to available homologous structures in the protein data bank. The crucial role of two structural elements in the initiation of GK's closure has thus been identified. Both the loop located between the third and fourth β-sheets (^{56}TRAPRPGEVDGVDY^{69} for *M. tuberculosis* GK) and the fourth α-helix play an essential role in this process. This has been confirmed in a second phase using more accurate all-atom IMD simulations. Hence, we gained an intuitive understanding of this model hinting at an early intermediate state, which seems to provide an impulse for the closure mechanism. These findings allow us to suggest an interesting new mechanistic hypothesis. GMP docking could actually initiate the displacement of the loop. The approach of the negatively charged GMP could drive the loop via long-range electrostatic interactions with two arginine residues located on this fragment. R57 and R60 link the substrate phosphate group to the enzyme and bridge it to the ATP's γ-phosphate at a later stage. This move would favor the global reorientation of the GMP-binding domain. It leads to the creation of new interactions between α-helix four and β-sheet five. Subsequently, specific contacts between the GMP-binding and LID domains are formed, thus achieving a global closure of the enzyme. In summary, a variety of induced conformational changes related to the GK enzyme, some reversible and some irreversible, have been efficiently explored via coarse-grain IMD simulations. The underlying conformational transitions were then refined at a higher atomic resolution.

Exploring the Membrane Anchoring of the SNARE Complex

Simulations of the SNARE complex were carried out with a reduced low-resolution model and with a refined and more intricate all-atom model. The coarse-grain approach comprises 72,466 particles, whereas the detailed representation requires modeling of 304,564 atoms. These simulations have provided twofold feedback. On the one hand, we gained theoretical data on the coarse-grain method that is still under development. Such feedback helps to continuously improve the method. On the other hand, biologically and functionally relevant structural information could be obtained for this system. The transmembrane insertion of the synaptobrevin helix into a lipid bilayer has been extensively studied via interactive simulations on a reduced SNARE system. This allowed us to explore the role of two consecutive tryptophan residues of the TMD. These tryptophans are considered key residues for helix anchoring.[39] Indeed, these residues adopt a firmly anchored stable position in the membrane and resist to a broad variety of perturbation-relaxation trials. Then, we essentially focused on the reorientation and repositioning of specific residues inside the membrane or at its interface with the aqueous solution. After a perturbation by the user it is possible to observe the relaxation of the simulated system and its return to equilibrium. A particularly informative experiment is to pull lipid molecules out of the membrane. Previously, non-interactive simulations were used to quantify the lipid adhesion forces,[44] which require complex postsimulation analysis. The haptic feedback provides intuitive clues about force-field parameterization and the balance of the various forces in the system, such as those between the lipids themselves and between the protein and the lipids. Comparatively, it is rather easy to extract a lipid from the membrane, whereas it is exceedingly difficult to detach the protein TMD.

Probing Mechanical Properties of the SNARE Assembly

Encouraged by the results on membrane insertion of SNARE transmembrane helices, we then examined the larger, extended model of the SNARE system. A pull on the N-terminal part of the four-helix bundle initiates a membrane perturbation. This hints at the possibility of a mechanical functional role of the bundle. It could lead to a perturbation of the bilayer and help catalyze the fusion event. At the same time, one also observes an important stretching of the global protein bundle. This indicates both a structural and a mechanical response of the whole helix bundle to an imposed force. Another observation is that the strength of the insertion of the syntaxin and synaptobrevin transmembrane helices into the two opposing lipid bilayers differs. The IMD approach allowed us to characterize a higher stability for the synaptobrevin transmembrane helix immersed in the vesicular membrane. The syntaxin protein is inserted more weakly into the target membrane. This finding might be a partial explanation as to why the syntaxin membrane anchoring is reinforced by the palmitoyl chains of synaptosome-associated protein of 25 kD (SNAP-25) in a biological context. Finally, the interactive simulations can be used to examine the structural integrity of the SNARE protein assembly. An interesting experiment consists of testing the self-association of the four SNARE helices driven by their hydrophobic complementarity. This can be done by first unzipping the complex, starting from the C-terminal part of the helix bundle, and then following the system while it returns to equilibrium. The synaptobrevin helix appears to be the weakest spot in the bundle, dissociating more readily than the other three helices.

Discussion

The following discussion exposes some of the advances that can be achieved via the IMD method. It is based on the case studies described in this report and on the observation of common features of interactive simulations, new perspectives and possible future enhancements. In particular, we focus on biological applications to the SNARE protein complex, and to both the GK and OmpT enzymes.

Interactive Simulations Provide an Intuitive and Enhanced Perception of Biochemical Systems

In an IMD simulation, the user guides the evolution of a given molecular system to overcome the short timescale of classical MD simulations. In this way, the user can explore rare events by enhancing their probability through his deliberate interactions. Particular benefits from the interactive approach are as follows. Firstly, the specification of a complex set of time varying conditions that drive the simulation is made easy. Second, the contin-

uous perceptual feedback produced during the course of the simulation is very valuable for the user. One can thus compel a nonspontaneous event to probe the related mechanical properties and structural deformations of a model of a given molecular system. Two potential issues concern the exhaustiveness and the reproducibility of the exploration. The approach is not by itself exhaustive; the pertinence of the exploration depends on the user's expertise to determine which explorations make sense. With MDDriver it is however possible to script some simpler deformations to achieve an automated procedure somewhat comparable to a grid search. Reproducibility can be quite delicate depending on how severely the system was perturbed. However, by saving the user interactions to a file for later replay one should be able to achieve a certain degree of reproducibility. This also enables the comparison of the directions and magnitudes of the forces in several separate simulations.

Membrane Anchoring, Interactive Docking and Novel Hypotheses on Membrane Fusion

Through this approach the user can easily explore the reversibility of structural events, induce large conformational changes or sense the mechanical properties of a molecular model. We have presented a first series of interactive investigations via a perturbation-relaxation approach. This technique has been applied to a lipid bilayer in the neighborhood of membrane proteins such as OmpT or synaptobrevin. The response of the TMDs of these proteins supplied new understanding about how these molecular models reproduce their insertion and anchoring into membranes. The positions of hydrophobic residues relative to the lipid chains and the membrane deformation after pulling on the synaptobrevin transmembrane helix have thus been efficiently studied. Such studies lead to new hypotheses that may be validated experimentally. Interactive and intuitive docking is another potential application as shown for the ARRA substrate interacting with OmpT. Multiple pathways for substrate approach were tested with this dynamic method and haptic rendering discussed below allowed for an enhanced perception of their feasibility. Interactive experiments were also informative for the study of enzymatic systems and allowed us to investigate the relationship between biological function and conformational fluctuations. Major conformational changes occurring during the closure of the GK enzyme could be investigated. In the case of the OmpT enzyme, we carried out a steered exploration of potentially correlated conformational transitions coupling the fluctuations of the substrate with those of nearby extracellular loops. In the same way, the reported autoproteolytic reaction of OmpT could be studied by driving the loop self-orientation via IMD simulations.

The study of the full-size model of the SNARE complex at coarse-grained resolution has provided valuable information about key features that might be essential for further understanding of the membrane fusion mechanism. Our experiments aimed at reproducing and exploring the mechanical action of the complex on the lipid bilayer. We realized this by pulling on the N-terminal hydrophilic and solvent-exposed part of the protein bundle. Our aim was to characterize the impact of this remote perturbation on the lipid bilayer. We have also observed a difference in the anchoring strength of the syntaxin and synaptobrevin TMDs in the membrane. This observation could possibly arise from an artefact in the coarse-grain parameterization, but it might as well hint at a functionally significant behavior. Higher resolution all-atom and noninteractive MD simulations will be necessary to confirm this feature. These first results highlight interesting perspectives for future investigations, such as the extension of the model system to multiple SNARE complexes.

Collaborative and Quantitative IMD Simulations

Another exciting application domain concerns cooperative IMD simulations. In particular for the SNARE system, this approach might provide further insight into the creation of a membrane pore. In that case, several users would simultaneously interact with the simulation system, each one pulling on one of the complexes. We did carry out first exploratory experiments on such *in situ* collaborative work using a setup with two haptic devices shown in Figure 4D.

In this report, we have mainly focused on intuitive and qualitative data. It has to be pointed out that reliable quantitative information can be reconstructed from IMD simulations as well. This is possible by recording simulation parameters and data such as trajectories, energy terms and imposed forces followed by postsimulation analysis. The user can for instance compute a potential of mean force from such simulations, as has been described for SMD simulations.[44,45]

Haptic Scene Rendering Provides Additional Benefits with Respect to Visual Feedback

The interactive experiments described in this article can be carried out without any specific equipment, just using a mouse and on-screen visual feedback. However, comparing these simulations with and without haptic feedback has shown that haptic devices provide several benefits. A quantitative evaluation of this question goes beyond the scope of this article, but some qualitative observations are discussed.

Generally speaking, immediate feedback from the simulation adds a new dimension of user perception. The instant visual feedback occurring after a user-induced perturbation is one of the most informative ways to use IMD for perceptive learning. Adding haptic feedback to this instant visualization further increases the understanding of a given system. Indeed, the force resistance provided by the haptic device using a mass-spring model, in addition to the visual feedback, allows the user to experience physical effects such as solvent viscosity or the mechanical resistance of a secondary structure element. This is intellectually easy to assimilate, even for structural biology or molecular modeling neophytes. We have successfully used this approach for teaching purposes, startling students with the high resistance forces preventing the extraction of an R-SNARE fragment from the membrane.

To interact with a simulation in progress, it is possible to use a mouse for adding force constraints on atoms or pseudo atoms. This provides 2 degrees of freedom for the interaction. However, using a haptic device is well adapted to this task, especially for selecting and catching moving atoms. Such a device with 3 instead of 2 degrees of freedom is more intuitive and efficient for interacting with a complex three-dimensional object. Further-

more, the immediate haptic feedback when an atom is actually picked significantly improves the user experience and greatly helps to immerse the user in the molecular scene. When using visual feedback only, the user often asks for additional explanations before getting started. With force feedback, this barrier is lifted, as the interactive simulation is comparable to dextrous manipulations such as those carried out in daily life.

It should also be pointed out that the hardware requirements are modest. In our experience, this approach is viable using a small and affordable haptic device, providing 3D positions and handling 3D directional force feedback. Such an entry-level solution designed for a desktop use is targeted at a large user community and is very easy to set up. The availability of convenient drivers that are compatible with software libraries such as VRPN[30] was a major criterium for our choice. We are also working on more advanced feedback, using six-dimensional output forces with additional torsion force feedback for experiencing more complex properties on biomolecules.

Using Interactivity to Enhance the Development of New Simulation Methods

Another, yet unexplored possibility for IMD simulation experiments concerns rapid and intuitive prototyping of novel modeling methods. This was for instance the case with the coarse-graining approach used in this work. The method is still under ongoing development. Several advantages could be gained from interactivity. After changing a forcefield or simulation parameter, one can visualize the evolution of related physico-chemical properties on the fly (lipid–lipid selfassembly, solvation effects). Feedback on the stability of a simulation with respect to a given conformation (integrity of a fold, heavy metal coordination sphere) is immediate. The impact of parameter or algorithm changes on calculation speed (frequency and algorithm of neighbor list calculation) is observed directly. For instance, we observed a slow-down every time the neighbor list used for the calculation of the pairwise interactions is recomputed by Gromacs (Supporting Information Fig. 3). These are just a few examples.

To the best of our knowledge this aspect has not been explored in preceding studies, focusing essentially on "steering" or "guidance" possibilities of interactive dynamics. Interactive simulations can for instance be used in an exploratory context. This implies less accuracy in the control of the induced force while providing intuitive and instantaneous tactile feedback via haptics. Such feedback illustrates how the program is simulating specific components of the forcefield, for instance independent terms of the potential energy. It can be used to improve a known forcefield for a given system or to test newly developed parameters for atypical residues. Using careful post-IMD analysis, one can furthermore obtain quantitative and reliable data enhancing a scientist's knowledge and perception of the underlying simulation method.

Comparative IMD Simulations at Different Resolutions

A comparative approach, where parameters are calibrated at different resolutions on the same reference system, can be used to improve a given simulation method. We have presented examples using both a detailed all-atom and an approximated coarse-grain representation. Such simulations help to point out discrepancies in the underlying modeling methods and highlight strengths and weaknesses. The IMD user can thus make an informed choice between the accuracy of a higher resolution representation and the flexibility of a low-resolution interactive simulation depending on the specific requirements for the study of a given molecular system. For instance, pulling a synaptobrevin transmembrane helix out of the membrane provided qualitative information about the anchoring strength of the protein fragment within the lipid bilayer. Comparison of the propensity of a single lipid molecule or of the TMD to return into the membrane with all-atom simulations indicated an excellent agreement with the coarse-grain model parameterization. This validated the more exhaustive and broader exploration carried out at the CG level. Simulating a system with an IMD based approach at different modeling scales also provides information about the accuracy, the pertinence or the implementation of methods in terms of reliable energies and physical or chemical properties.

Table 2. Simulation Time Explored During an Interactive Experiment of One Minute Duration Using IMD via *MDDriver*.

System	Number of particles	Resolution	Simulation time (ps)
Guanylate kinase	1900	Coarse-grain	17,000
	18,098	All-atom	10
SNARE complex	33,028	Coarse-grain	100
	304,564	All-atom	0.06

Two biological systems are examined using either a high-resolution (all-atom) or low-resolution (coarse-grain) representation.

Scripting Simulations via IMD

Finally, another aspect concerns the programmatic interaction with a molecular modeling algorithm such as a molecular dynamics simulation. The modularity of the MDDriver library permits an easy development of tools that interact with a calculation at each simulation step. This can for instance be used to compute certain parameters from the system coordinates and reinject external forces depending on these parameters into the simulation. With respect to the applications presented here, one could for example use this approach to impose a series of successive opening/closure events to the GK system.

Interactive Simulations at Coarse-Grained Resolution: A Well-Balanced Choice

Using a coarse-grain representation for interactive simulations provides two marked advantages. It permits bridging an expanded timescale by using large integration timesteps (see Table 2 and the central panel of Fig. 2) and leads to a smooth potential energy surface. These features facilitate imposing large conformational changes on a molecular structure. In comparison, at atomic resolution, one may need to apply exceedingly high forces over a short time interval to be able to observe a conformational change in interactive time. This principal inconvenience with high-resolution all-atom simulations is due both to

the rugged shape of the potential energy surface harboring many local minima and to the limited time-scale of these simulations. The high forces imply an elevated risk of loosing structural integrity. Furthermore, some large conformational changes cannot even be reached with all-atom IMD simulations because of their amplitude. Currently, the short integration timestep and the large number of particles that is required to model a complex biological system do in fact limit the application range of the IMD technique to all-atom simulations. To effectively observe large structural changes with all-atom simulations, noninteractive MD simulations are now performed for hundreds of nanoseconds or more. However, corresponding IMD simulations are intrinsically restricted to the time the user can stay in front of a computer. Typically the user's human timescale for interaction ranges from seconds to minutes. Implementing coarse-graining into IMD does significantly help to hedge these critical points and enables a better and quicker exploration of reversible events as the simulation system is more reactive. Furthermore, it enables the study of even larger systems. The two examples provided in Table 2 are particularly compelling. More than five orders of magnitude in the molecular time-scale can be bridged in a one-minute interactive experiment. For each system—the small GK enzyme and the four-helical SNARE bundle—the low resolution simulation spans a time-scale ~1700 times longer. The increase in integration time-step from 2 fs in the all-atom simulation to 40 fs represents a factor of 20. The roughly ten-fold decrease in the number of particles is responsible for the remaining speed-up.

CG IMD Simulations Do Not Require Excessive User Forces

A low-resolution approach significantly alleviates the problem related to high user-applied forces encountered for all-atom IMD simulations. At high atomic resolution, an effective event is produced with comparable speed (a few seconds at the user's timescale) after a force application with a scaling factor of around 30–50. This refers to the "Applied Force Scaling" tab in the "Tool controls" menu of the VMD software and corresponds to the $\boldsymbol{S}_{\mathbf{f}}$ parameter described above. This scaling factor is reduced to 1–5 in an equivalent CG simulation. In an all-atom IMD simulation, such an interactively imposed move has to be applied during more than 10 min (at a user's timescale) with a moderate force to preserve important secondary structure elements. This was for example the case with respect to the short beta sheets of the OmpT enzyme during the attempt to close its L4 loop.

Overall, a coarse-grained approach represents a well-balanced compromise. Some of the precision of the model is sacrificed to be able to manipulate the simulation at time-scales that are pertinent with respect to biological function. Hence, interactive CG simulations contribute to a better understanding of models of biological systems but also of the underlying modeling methods used for simulating a molecular system as already pointed out above.

Interactive Simulations, a First Step Toward Human Computation in Structural Biology?

In a particularly nice and enjoyable way, the "Fold It!" project (http://fold.it) has recently brought interactive simulations to the fore. "Fold It!" is a 3D-puzzle desktop game, in which the user's task is to fold proteins interactively and without any knowledge prerequisites. The puzzle is essentially based on an interactive-intuitive learning process. The protein folding work carried out by the "Fold It!" users can then be gathered and analyzed. This form of human computation[46] enables large-scale and intelligence-driven sampling of folded protein conformations which is yet beyond the reach of brute force computational approaches.

Conclusion

The MDDriver library introduced in this article offers a simple, modular and generic solution for combining any coordinates-based calculation code with various visualization programs. The interactive simulation method investigated in this work is a powerful tool for the exploration of biomolecular structures in large biological systems and greatly facilitates the generation of new hypotheses. It is now easily accessible, even via an affordable desktop computing environment.

The IMD simulation improvements presented here allow some biologically relevant questions to be addressed. More generally, if we consider our selection of systems, several types of applications can be suggested. They comprise protein/substrate interactions and the impact of long-range forces on steering and docking of substrates, but also local interactions such as those inside a catalytic pocket. It is possible to investigate stabilization, mechanical properties linked to functional-decisive conformational equilibria and the structural/mechanical behavior of domains in large macromolecular assemblies. Many other exciting applications are expected.

Simulations implementing a coarse-grain representation offer new perspectives and enable even more challenging interactive experiments. The low resolution allows one to increase the accessible time-scale for following the evolution of a given system. It becomes quicker and easier to manipulate structures during an IMD simulation. The relaxation process after a perturbation by the user provides instant information about the underlying calculation methods and force fields. An interesting evolution would be to design multiscale or multiphysics IMD methods offering the best compromise between accuracy and rapidity of the simulations.

In the context of methodological developments, IMD does also appear as a powerful tool to improve new methods under development or to fine-tune critical simulation parameters. We encourage experimenting with the MDDriver framework, which will soon be available at http://mddriver.sourceforge.net with technical documentation, movies and tutorials.

Acknowledgments

The authors thank Diana Doherty and Brigitte Hartmann for helpful suggestions and corrections for the revised manuscript. The authors thank the Gromacs developers, and in particular David van der Spoel and Erik Lindahl, for their kind assistance and helpful suggestions. The authors thank Peter J. Bond and Mark S. P. Sansom for access to their coarse-grain parameteriza-

tion. M.B. thanks Richard Lavery for getting him started with this area of research, both intellectually and by providing essential equipment.

References

1. Surles, M. C.; Richardson, J. S.; Richardson, D. C.; Brooks, F. P. J. Protein Sci 1994, 3, 198.
2. Wu, X.; Wang, S. J Chem Phys 1999, 110, 9401.
3. Wu, X.; Wang, S. J Am Chem Soc 2002, 124, 5282.
4. Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. J Mol Graph 2001, 19, 13.
5. Leech, J.; Prins, J. F.; Hermans, J. IEEE Comp Sci Eng 1996, 3, 38.
6. Rapaport, D. C. Physica A 1997, 240, 246.
7. Wollacott, A. M.; Merz, K. M. J Mol Graph Model 2007, 25, 801.
8. Grayson, P.; Tajkhorshid, E.; Schulten, K. Biophys J 2003, 85, 36.
9. Stone, J. E.; Gullingsrud, J.; Schulten, K. Proceedings of 3D Interactive Graphics, Research Triangle Park, NC; 2001, pp 191–194.
10. Nelson, M.; Humphrey, W.; Kufrin, R.; Gursoy, A.; Dalke, A.; Kale, L.; Skeel, R.; Schulten, K. Comp Phys Commun 1995, 91, 111.
11. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. J Comp Chem 2005, 26, 1781.
12. Knoll, P.; Mirzaei, S. Rev Sci Instrum 2003, 74, 2483.
13. Vormoor, O. Comp Sci Eng 2001, 3, 98.
14. Prins, J. F.; Hermans, J.; Mann, G.; Nyland, L. S.; Simons, M. Fut Gen Comp Sys 1999, 15, 485.
15. Ai, Z.; Frohlich, T. Comput Graph Forum 1998, 17, 267.
16. Cruz-Neira, C.; Langley, R.; Bash, P. A. SAR QSAR Environ Res 1998, 9, 39.
17. Gillet, A.; Sanner, M.; Stoffler, D.; Olson, A. Structure 2005, 13, 483.
18. Klein, M. L.; Shinoda, W. Science 2008, 321, 798.
19. Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. J Comp Chem 2007, 28, 2618.
20. Rodrigues, C.; Hardy, D.; Stone, J. E.; Schaeffer, F.; Hwu, W.-M. Proceedings of the ACM International Conference on Computing Frontiers, Ischia, Italy, 2008.
21. Meel, J. A. v.; Arnold, A.; Frenkel, D.; Zwart, S. F. P.; Belleman, R. G. Mol Simul 2008, 34, 259.
22. Liu, W.; Schmidt, B.; Voss, G.; Müller-Wittig, W. Comp Phys Commun 2008, 179, 634.
23. Anderson, J. A.; Lorenz, C. D.; Travesset, A. J Comp Phys 2008, 227, 5342.
24. Baaden, M.; Lavery, R. In Recent Adv in Structural Bioinformatics, Research Signpost, Trivandrum, Kerala, India, 2007; pp. 173–195.
25. Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. J Struct Biol 2007, 157, 593.
26. Bond, P. J.; Sansom, M. S. P. J Am Chem Soc 2006, 128, 2697.
27. Marrink, S. J.; de Vries, A. H.; Mark, A. E. J Phys Chem B 2004, 108, 750.
28. Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. J Phys Condens Matter 2004, 16, 481.
29. Humphrey, W.; Dalke, A.; Schulten, K. J Mol Graph 1996, 14, 33.
30. Taylor, R. M., II; Hudson, T. C.; Seeger, A.; Weber, H.; Juliano, J.; Helser, A. T. VRST '01: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Banff, USA, 2001.
31. Blaszczyk, J.; Li, Y.; Yan, H.; Ji, X. J Mol Biol 2001, 307, 247.
32. Hible, G.; Christova, P.; Renault, L.; Seclaman, E.; Thompson, A.; Girard, E.; Munier-Lehmann, H.; Cherfils, J. Proteins 2006, 62, 489.
33. Hible, G.; Renault, L.; Schaeffer, F.; Christova, P.; Radulescu, A. Z.; Evrin, C.; Gilles, A.-M.; Cherfils, J. J Mol Biol 2005, 352, 1044.
34. Vonrhein, C.; Schhlauderer, G. J.; Schulz, G. E. Structure 1995, 3, 483.
35. Choi, B.; Zocchi, G. Biophys J 2007, 92, 1651.
36. Baaden, M.; Sansom, M. S. P. Biophys J 2004, 87, 2942.
37. Bond, P. J.; Sansom, M. S. P. Mol Membr Biol 2004, 21, 151.
38. Neri, M.; Baaden, M.; Maritan, A.; Carloni, P. Biophys J 2008, 94, 1.
39. Krieger, E.; Leger, L.; Durrieu, M.-P.; Taib, N.; Bond, P. J.; Laguerre, M.; Lavery, R.; Sansom, M. S. P.; Baaden, M. Proceedings of ParCo 2007, Parallel Computing: Architectures, Algorithms and Applications, Juelich, Germany, 2007.
40. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. J Comp Chem 2005, 26, 1701.
41. Durrieu, M.-P.; Lavery, R.; Baaden, M. Biophys J 2008, 94, 3436.
42. Baaden, M.; Meier, C.; Sansom, M. S. J Mol Biol 2003, 331, 177.
43. Marrink, S. J.; Berger, O.; Tieleman, P.; Jähnig, F. Biophys J 1998, 74(2 Part 1), 931.
44. Balsera, M.; Stepaniants, S.; Izrailev, S.; Oono, Y.; Schulten, K. Biophys J 1997, 73, 1281.
45. Gullingsrud, J.; Braun, R.; Schulten, K. J Comp Phys 1999, 151, 190.
46. Ahn, L. v. Computer 2006, 39, 92.

To appear in *Pacific Symposium for Biocomputing 2010*

MULTI-RESOLUTION APPROACH FOR INTERACTIVELY LOCATING FUNCTIONALLY LINKED ION BINDING SITES BY STEERING SMALL MOLECULES INTO ELECTROSTATIC POTENTIAL MAPS USING A HAPTIC DEVICE

OLIVIER DELALANDE, NICOLAS FEREY, BENOIST LAURENT, MARC GUEROULT[§]
Laboratoire de Biochimie Théorique, CNRS UPR9080/IBPC, 13 rue Pierre et Marie Curie, F-75005, Paris, France

BRIGITTE HARTMANN
[§] *DSIMB team, INTS, 6 rue Alexandre Cabanel, F-75739 Paris Cedex 15, France*

MARC BAADEN
Laboratoire de Biochimie Théorique, CNRS UPR9080/IBPC, 13 rue Pierre et Marie Curie, F-75005, Paris, France

Metal ions drive important parts of biology, yet it remains experimentally challenging to locate their binding sites. Here we present an innovative computational approach. We use interactive steering of charged ions or small molecules in an electrostatic potential map in order to identify potential binding sites. The user interacts with a haptic device and experiences tactile feedback related to the strength of binding at a given site. The potential field is the first level of resolution used in this model. Any type of potential field can be used, implicitly taking into account various conditions such as ionic strength, dielectric constants or the presence of a membrane. At a second level, we represent the accessibility of all binding sites by modelling the shape of the target macromolecule *via* non-bonded van der Waals interactions between its static atomic or coarse-grained structure and the probe molecule(s). The third independent level concerns the representation of the molecular probe itself. Ion selectivity can be assessed by using multiple interacting ions as probes. This method was successfully applied to the DNase I enzyme, where we recently identified two new cation binding sites by computationally expensive extended molecular dynamics simulations.

1. Introduction

Metal ions drive important parts of biology, yet it remains experimentally challenging to locate their binding sites. A particular example may concern the mechanisms underlying the formation of non-specific protein-DNA complexes. In this context, we have recently studied the DNase I/DNA system as a representative and rather simple model of a non-specific complex. DNase I is a glycoprotein hydrolyzing DNA phosphodiester linkages in the presence of divalent cations, Ca^{2+} and Mg^{2+}. We demonstrated that Ca^{2+} and Mg^{2+} are crucial for optimizing the electrostatic fit between DNA and enzyme. In particular, extended molecular dynamics simulations at atomic detail allowed us to identify two new cation binding sites that are functionally important [1]. Such high-resolution atomistic methods are computationally expensive and require long simulation times. Could these ion binding sites also have been detected with simpler methods? Maybe this could be achieved at lower resolution and requiring less computational power? In the present manuscript we set out to describe a new method aimed at these goals, using the molecular dynamics results as a reference for assessing our findings. The purpose of this method is to provide a fast user-guided search in order to locate potential binding sites, prior to detailed investigations of these sites using more accurate but computationally expensive approaches.

Recently, we have described the exciting possibilities of interactive molecular simulations for studying biological macromolecules [2]. Here, we demonstrate that this approach can be used for locating ion binding sites. The concept of our method is to interactively explore electrostatic potential fields while being guided towards the binding sites by force feedback. This interactive haptic approach may be traced back to the work of Nagata *et al.* [3] who explored the concept in the context of protein-ligand docking. Their work was somewhat ahead of its time, as may be illustrated by the following observations: "Certain limitations remain; for example, only twenty protein atoms can be used to generate the electrostatic field. Furthermore, the system can only use globular probes, preventing drug molecules or small chemical groups from being simulated. These limitations are the result of our insufficient computer resources". Today, this situation has changed and we did not observe any significant computational limitations. Our approach has been tested on systems as big as a pentameric ligand-gated ion channel comprising 1 500 amino acid residues [4]. Of course, the raw increase in computer speed is not the only reason for such an improvement. Adapted software solutions and improved algorithms further render current approaches more efficient.

Using such an interactive approach for locating ion binding sites is a new idea. It will be illustrated with the DNase I example system. Previously, electrostatic steering has been described for a variety of cases such as

enzyme-ligand binding [5], tRNA binding [6], antibody-antigen association [7] and Cdc42 recognition [8]. Computational methods have been used in several of these studies, but interactive virtual reality (VR) approaches only marginally. Why are such VR approaches useful? VR efficiently introduces a human element in the process and benefits from the user's experience and insight. Enabling the user to sense the electrostatic potential field via tactile feedback is a major advantage. The electrostatic potentials of biomolecules are often complex volume and multi-level data, rendering the visual perception of individual binding sites difficult if not impossible. Adding haptic feedback to their interactive exploration significantly simplifies this task. Here, we use such an approach on a biomolecular system in order to explore potential variations at the DNase surface, to locate favourable binding sites, and to discover far-reaching electrostatic pathways guiding ions to these sites.

The representation used for the simulation system is innovative. It could be summarized as a multi-level simulation combining multiple physics models and is intrinsically multi-resolution. The potential field is the first level of resolution used in this model. At a second level, we represent the accessibility of all binding sites by modelling the shape of the target macromolecule at atomic or coarse-grained resolution. The third independent level concerns the representation of the molecular probe itself. By using multiple scales and resolutions it is possible to focus on the essential degrees of freedom of a complex biological system. This makes the user's interaction with the system more efficient.

The interactive VR-based approach renders the scientific task nice and enjoyable, hence encouraging the user to pursue his investigation. Another example of this kind is the ''Fold It!'' project (http://fold.it) that recently brought interactive simulations to the fore. ''Fold It!'' is a 3D-puzzle desktop game, in which the user's task is to fold proteins interactively and without any knowledge prerequisites. The puzzle is essentially based on an interactive-intuitive learning process. The similarities such applications bear with video games should not delude scientists to underestimate the scientific value of such approaches.

2. Materials and Methods

2.1. *Structural data and model preparation for the DNase I system*

Several experimental structures of the DNase I enzyme from different biological contexts were considered as reference in this study: the enzyme crystallized in its *apo* form (PDB code 3DNI [9]), in a DNase I:Actin complex (PDB code 1ATN [10]) or bound to an oligonucleotide duplex (PDB code 1DNK [11]). Calcium ions are found in some of these structures: *apo* DNase I (3DNI) bears two Ca^{2+} binding sites (described below as sites 1 and 2) and the DNase I:Actin complex reveals three Ca^{2+} binding sites (sites 1, 2 and 3).

In the interactive explorations, a model derived from previous molecular dynamics (MD) simulations is used to represent the enzyme. This model corresponds to the most representative conformer of the whole trajectory as determined by an exhaustive cluster analysis [1]. The detection of ionic binding pockets and their localization in the crystallographic structures have been based on this model. In the MD simulations, the *AMBER* parm99 force field has been used [12]. Consistently, we generated the parameters for the electrostatic potential calculation (*pqr* files) with the *pdb2pqr* program [13] using the *--ff=amber* option, including for ionic probes. The potential was determined with the *APBS* software [14] and standard input was obtained with the *--apbs-input* option. Ionic strength was introduced using a 0.15 M concentration of both +1 and -1 charged ions with a 2.0 Å radius value (*APBS* parameters).

The shape of the DNase I molecule was represented *via* van der Waals parameters from the *AMBER* parm99 force field (atomic resolution) or from the coarse-grained model by Zacharias [15] (low resolution).

Every interactive experiment consisted in at least four independent trials to detect binding pockets for each probe. We concluded that a binding site was found when the probe got stuck in a location where it was impossible to escape without increasing the user force range. Attractive pockets at the enzyme surface where the ion could easily be retrieved were not considered as binding sites.

2.2. *MyPal: an interactive simulation approach combining an electrostatic potential field and pairwise non-bonded interactions*

MyPal stands for **M**olecular scrutin**Y** of **P**otenti**AL**s. The application has a corresponding French name with the same meaning: *MonPote*, "**MO**lécules **N**aviguant sur un **POTE**ntiel". In this section we describe the underlying methodology developed in order to study the behaviour of ions or small molecular probes interacting with a static macromolecular target molecule. In this approach, based on a classical Newtonian simulation, ionic probes are immersed into and guided by an electrostatic potential field induced by the target molecule. In addition, the 3D shape of the target molecule is taken into account by representing its static 3D structure interacting with the probe molecules. The simulation deals with both electrostatic properties and steric constraints induced by the 3D shape of the target molecule. Moreover, in order to study the behaviour of several interacting ions or molecules,

non-bonded Coulomb and van der Waals interactions between all the probes are calculated. These inter-probe interactions are merged with the electrostatic potential from the map and the van der Waals interactions with the target molecule. The user can remain passive and observe the movements of the ionic probes from their initial positions during a simulation in progress. Alternatively, the user can act on the probes by selecting and applying external forces taken into account by the simulation, in order to explore the potential and identify binding sites.

A *MyPal* simulation requires two datasets as schematically shown in Figure 1. The first dataset corresponds to the structure and potential of the target macromolecule, including all static particles in the simulation (left). The second dataset describes a separate dynamic molecular structure including the ionic probes (right).

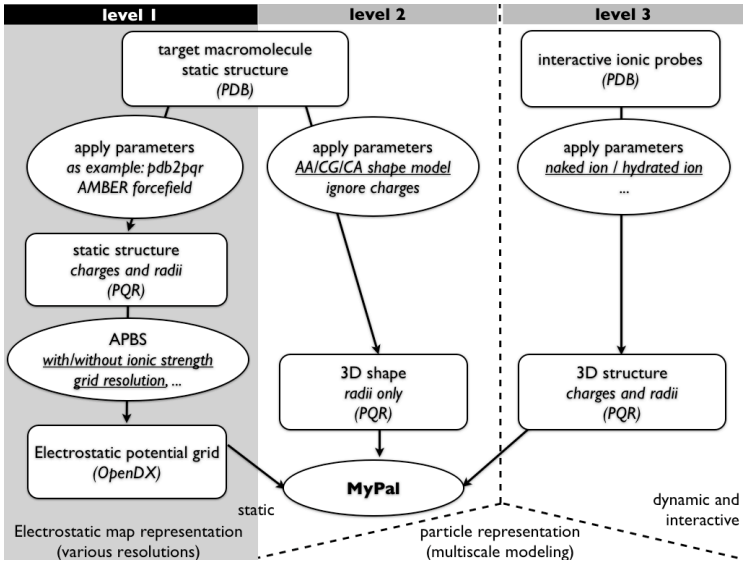


Figure 1: Schematic flow diagram for setting up a *MyPal* simulation. **Level 1** deals with the calculation of the electrostatic potential grid. Here we use the *pdb2pqr* tool in order to set up the radii and charges of the target macromolecule using the atomistic *AMBER* parm99 force field. Then we compute its electrostatic potential map using the *APBS* software [14]. The potential is saved as an *OpenDX* file and used as input for *MyPal*. At **level 2**, the 3D shape of the target molecule is modelled as a set of static spherical particles at the desired resolution (all-atom: AA, multi-bead coarse grained: CG or single-bead carbon alpha: CA). The ionic probes are modelled at **level 3**, using a set of radii and charges. The latter parameters may represent different resolutions such as naked *vs.* hydrated ions. Levels 1 and 2 are static, whereas the particles in level 3 are dynamic and can be guided by the user.

During the simulation, forces are computed and applied on the dynamic particle set ($P_{dynamic}$). We explicitly consider potential, van der Waals, Coulomb and external forces.

The potential forces F_{elect} act on the ionic probes and originate from the electrostatic potential map. They are defined by computing the gradient of the electrostatic potential. In Equation (1), we consider particle p belonging to the spatial cell $C_{i,j,k}$ of the electrostatic potential grid, and $E_{i,j,k}$ the value of the potential in this cell. We define the gradient as the mean of the difference between the $E_{i,j,k}$ potential and the potentials of the six adjacent cells, two for each axis. This method of computing the gradient reduces the bias related to the discretization of the grid.

$$\vec{F}_{elec}(p \in C_{i,j,k}) = \begin{bmatrix} \dfrac{\left(E_{i,j,k} - E_{i-1,j,k}\right) + \left(E_{i+1,j,k} - E_{i,j,k}\right)}{2.delta_x} q_p \\ \dfrac{\left(E_{i,j,k} - E_{i,j-1,k}\right) + \left(E_{i,j+1,k} - E_{i,j,k}\right)}{2.delta_y} q_p \\ \dfrac{\left(E_{i,j,k} - E_{i,j,k-1}\right) + \left(E_{i,j,k+1} - E_{i,j,k}\right)}{2.delta_z} q_p \end{bmatrix} \quad (1)$$

The van der Waals forces F_{vdw} applied to the ionic probes ($P_{dynamic}$) are computed taking into account all interactions between the ionic probes and the static particles defining the 3D shape of the target molecule, and between ionic probes themselves (P_{all}). As shown in Equation (2), these forces are computed using the following approximation for the van der Waals interactions.

$$\vec{F}_{vdw}(p \in P_{dynamic}) = \sum_{p' \in P_{all}} \vec{u}_{pp'} 4\varepsilon_{pp'} \left[\left(\frac{\sigma_{pp'}}{9d_{pp'}} \right)^9 - \left(\frac{\sigma_{pp'}}{7d_{pp'}} \right)^7 \right] \quad (2)$$

The Coulomb forces $F_{coulomb}$ applied to the ionic probes ($P_{dynamic}$) take into account all electrostatic interactions between the ionic probes according to Equation (3).

$$\vec{F}_{coulomb}(p \in P_{dynamic}) = \sum_{p' \in P_{dynamic}} -\frac{q_p q_{p'} \vec{u}_{pp'}}{4\Pi\varepsilon_0 d_{pp'}^2} \qquad (3)$$

Finally, these forces are summed with an external force provided by the user through the graphical interface during the simulation.

$$\vec{F}(p \in P_{dynamic}) = \vec{F}_{elec}(p) + \vec{F}_{vdw}(p) + \vec{F}_{coulomb}(p) + \vec{F}_{user}(p) \qquad (4)$$

This multi-resolution approach consists of three independent levels of modelling. The potential field is the first level of resolution used in this model (level 1 in Figure 1). In principle, any type of potential field could be used, not only electrostatic ones. Most of the scrutinized potentials in the present work are Poisson-Boltzmann type electrostatic potentials computed using the *APBS* software [14]. Such potentials are subject to several parameters (resolution, ionic strength, solvent and solute dielectric constants) and may thus implicitly take into account an averaging over selected solvent or solute degrees of freedom. More complex scenarios, for instance the presence of a biological membrane environment for a membrane protein target, could also be taken into account with tools such as AquaSol [16]. Deriving the potential field from an average over a molecular dynamics simulation could help to mimic the flexibility of the target molecule. These are only a few examples of how the resolution and underlying physics of the potential field can be varied. At a second level (level 2 in Figure 1), we model the accessibility of a binding site by representing the shape of the target macromolecule *via* non bonded van der Waals interactions between its static structure and the probe molecule(s). For this representation, a whole range of parameterizations commonly used in molecular mechanics is available. These comprise the detailed atomistic scale (AA), medium-resolution multi-bead coarse graining (CG) and single-bead carbon alpha models (CA) [17]. Level 3 in Figure 1 is the third, independent level of modelling and concerns the molecular probe itself. At this level, van der Waals and coulombic interactions are treated. Radii and charges can be chosen in order to fine-tune the desired properties of the ionic probes by considering effects such as solvation (increased radii) or charge screening (decreased charges).

It should be stressed that the *MyPal* approach is computationally cheap. The electrostatic map is computed offline prior to the interactive simulation and there are only a small number of dynamic ionic probes on which pairwise interactions are computed. Given the reduced number of degrees of freedom, namely the positions of the ionic probes, thorough sampling of phase space is easily achieved.

2.3. *Visualizing a simulation in progress*

During the design of the *MyPal* application, initial tests have been carried out with the VMD software [18]. This approach couples visualization and simulation with the *MDDriver* library [19], using the IMD protocol [20]. All screenshots in the present manuscript were obtained using VMD. VMD is however not optimally adapted to some of the specific tasks occurring during an interactive simulation. As an example, there is no visual feedback during the particle selection task, making it difficult to select the particle of interest with confidence. Multiple selections are not available using direct interactions with the mouse or a haptic device, but only using the Graphical User Interface (GUI). In order to surmount these shortcomings, we are now developing our own simple visualization tools specifically designed for interactive visualization. We use the visualization toolkit VTK [21] for visual rendering. VTK provides high-level features, such as isosurface rendering of electrostatic potential maps. In an enhanced version, our tool allows us to visualize up to several hundred thousand particles in interactive time using a GPU shader implementation of spherical representations. VTK was encapsulated into a Cocoa application, allowing us to quickly develop a GUI using the XCode and Interface Builder tools. As soon as this application will be finalized, it will be made freely available in order to simplify interactive potential exploration as described in this work [22].

2.4. *Interacting with the ionic probes during a simulation*

In order to interact with a *MyPal* simulation in progress, it is possible to use a mouse for adding force constraints on the probe particles. This approach provides two degrees of freedom, *e.g.* the x- and y-axes, for the interaction. However, using a haptic device is even better adapted to this task, in particular for selecting and moving particles in 3D space. Such a device with three instead of two degrees of freedom is more intuitive and efficient for interacting with a complex three-dimensional object. Furthermore, the immediate haptic feedback when a

particle is actually picked significantly improves the user experience and greatly helps to immerse the user in the molecular scene. When using visual feedback only, the user often asks for additional explanations before getting started. With force feedback, this barrier is lifted, as the interactive simulation becomes more intuitive and is comparable to dextrous manipulations such as those carried out in daily life. Hardware requirements are modest (details in section 2.6). In our experience, this approach is viable using a small and affordable haptic device, providing 3D positions and handling 3D directional force feedback. Such an entry-level solution designed for a desktop use is targeted at a large user community and is very easy to set up. The availability of convenient drivers that are compatible with software libraries such as *VRPN* [23] was a major criterion for our choice.

The interaction with the simulation was implemented as in VMD, allowing the user to impose forces on probe particles and experience a tactile feedback. The haptic device is used in order to control the direction of the forces applied to selected particles and to adjust the amplitude of these forces. This interaction method contains two stages. The first stage comprises the selection of a single probe particle or a set of particles that we will name $P_{selection}$, using a 3D tool attached to a haptic device and its buttons. In a second stage, the model described in Equation (5) is used in order to compute the forces $F_{simulation}$ applied to the selected particles and sent to the *MyPal* simulation as external force (see section 2.2). $F_{simulation}$ is proportional to the distance between the geometrical centre of the particle set and the tracker position P_T.

$$\vec{F}_{simulation}(p \in P_{selection}) = k_{simulation}(\vec{P}_T - \frac{1}{|P_{selection}|}\sum_{p' \in P_{selection}} \vec{P}_{p'}) \qquad (5)$$

The main idea of this approach is to link the selected atoms and the 3D haptic tool with a spring. Instead of providing direct haptic rendering of forces computed in the simulation, the force feedback $F_{feedback}$ only depends on the spring length according to Equation (6), which in turn is influenced by the way the simulation reacts to the applied force.

$$\vec{F}_{feedback} = -k_{feedback}(\vec{P}_T - \frac{1}{|P_{selection}|}\sum_{p \in P_{selection}} \vec{P}_p) \qquad (6)$$

We emphasize that the haptic loop computation frequency must be between at least 300 to 1000 Hz in order to provide a haptic rendering of good quality. A strong point of the approach described above is that a low physical simulation framerate does not cause instabilities and does not affect the quality of the haptic feedback. With this decoupled spring model, force feedback can be computed at a very high frequency required by the haptic device.

2.5. *Coupling simulation, visualization and interaction*

In this section, we briefly explain how simulation, interaction and visualisation modules exchange data. Figure 2 illustrates a schematic view of the data flow in the application presented in this work. We use the *MDDriver* library as central element in order to couple a visualization, interaction and *MyPal* simulation module. Previously, this library was successfully used to couple the *GROMACS* molecular dynamics engine [24] with *VMD* in order to carry out interactive studies of several biomolecular systems comprising enzymes and membrane proteins [2]. Here we use a similar approach in order to couple the *MyPal* program with our own visualization code based on *VTK* or with the visualization module *VMD*.

The *MDDriver* library provides a simplified and high-level API [19] for exchanging particle positions computed in a simulation module and custom user-generated forces from the visualization and interaction modules. The adapter layers shown in Figure 2 are specific for a given simulation or visualization software. They are mainly in charge of transforming the physical values into adequate units, but can also operate more complex transformations, such as changing data from a given discretization model to another one. The adapters are in charge of redistributing and collecting the data in a parallel application. For implementing *MDDriver* in a given software package, the major part of the work consists of locating the adequate routines in the simulation algorithm where the data exchange needs to take place. The exchange with the visualization tool includes sending atom positions, receiving forces and managing control events for the simulation.

Further coupling is needed by our interactive approach. It consists of coupling the peripherals management library, *e.g.* for working with a haptic device, and the visualization module. We use the *VRPN* library [23], which offers a device independent implementation, in order to read out the position and orientation of a tool used to select particles in the visualization module. *VRPN* is further used in order to generate force feedback along the lines described in section 2.4.

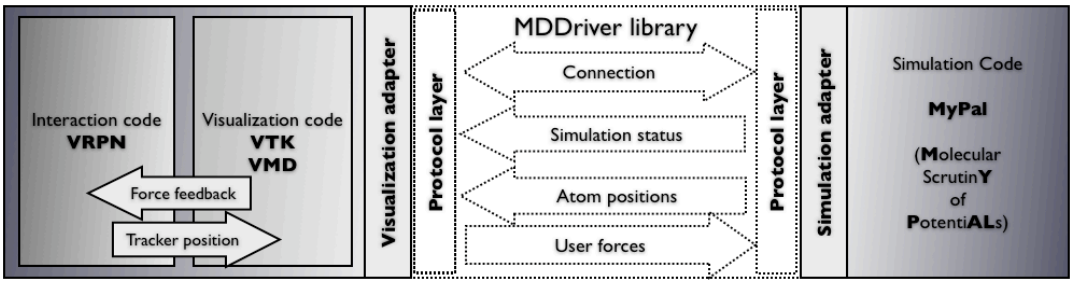


Figure 2: The *MDDriver* library couples a simulation code with a visualization code. *MDDriver* provides an API for exchanging particle positions computed in a simulation module and custom forces from the visualization and interaction modules [19]. Haptic device management is provided by the *VRPN* library [23] and force feedback is computed as described in section 2.4

2.6. *Hardware setup and typical configuration*

All interactive *MyPal* simulations were performed on a 3GHz dual quad-core MacPro Apple computer. Although the *MyPal* application itself is not parallelized, a multi-core workstation makes it possible to run all parts of the application on the same machine without performance issues. This includes visualization, haptic device server and calculation tasks. The molecular scene display was rendered using an NVIDIA Quadro FX 5600 graphics card. Stereo rendering was achieved via a Crystal Eyes stereovision device. Two devices were tested for user interaction and tactile feedback: a Phantom Omni haptic device and a higher precision Phantom Premium 1.5A, both by Sensable Technologies. Scene navigation was achieved with a Spaceball device providing six degrees-of-freedom.

The *MyPal* application can easily be run in a desktop context, with minimal spatial requirements, and does not require any sophisticated hardware setup. In our experiments, the typical configuration was to mount the haptic device, Spaceball, mouse and keyboard in front of a big screen. Figure 3 illustrates such a typical setup.

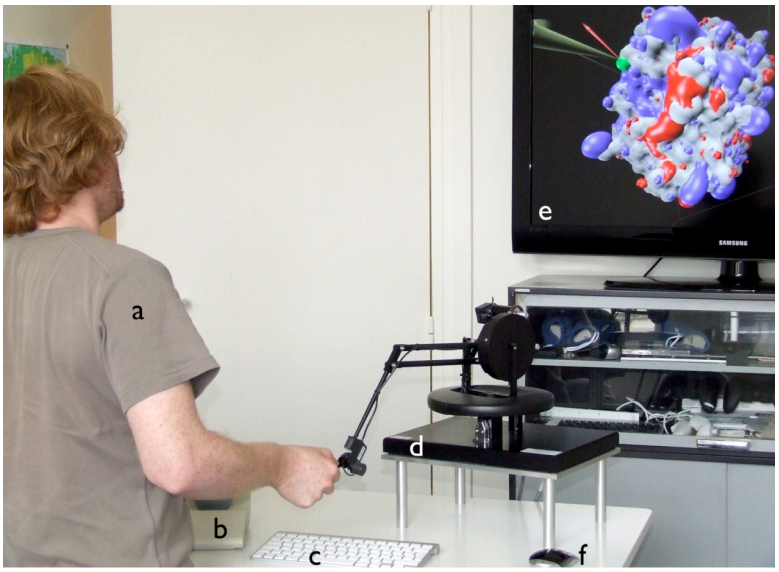


Figure 3: Typical hardware setup for an interactive exploration experiment based on the *MyPal* simulation tool. The user (a) can navigate the scene with a Spaceball device (b). Further control can be achieved *via* a keyboard (c) and a mouse (f). Interaction with the calculation and force feedback from *MyPal* occur *via* a haptic device (d). The scene is shown on a large screen (e) with an avatar representing the user-driven tool (cone in top-left corner of the screen), and a visual rendering indicating the user force (red arrow) acting on the ionic probe (green sphere).

3. Biological application – results and discussion

3.1. *DNase I cation binding sites revealed by atomistic molecular dynamics simulations*

We will use our previous results on cation binding sites in the DNase I enzyme as reference to compare our interactive simulations to. It has been shown elsewhere that molecular dynamics (MD) simulations are powerful – yet computationally expensive – tools for locating such sites [25]. Specific and non-specific protein-DNA interactions imply the formation of intermolecular interfaces requiring electrostatic and structural complementarity between the related partners. A precise functional role for metal ions in this process is certainly

an interesting hypothesis worthwhile exploring. The DNase I/DNA complex is a representative and rather simple model to study these non-specific interactions. We have recently carried out four 25 ns molecular dynamics simulations of the *apo* enzyme where the cation composition was varied between Na^{+}, Ca^{2+} and Mg^{2+} [1]. Detailed analysis of cation coordination in these simulations revealed four distinct sites. Two sites show a preference for calcium and two are selective for magnesium (Figure 4, left panel). The calcium binding sites were previously known [9], whereas the two magnesium sites were unexpected. We have shown that cations are essential for the biological function of DNase I, allowing DNA to bind to the enzyme. Indeed, the concurrent occupation of all four sites with the proper cation is required for an electrostatic fit between DNA, negatively charged, and the enzyme interface, also negatively charged in the absence of cations. This may imply that the information about the binding sites is encoded in the electrostatic potential map of the enzyme (Figure 4, right panel). In the following paragraphs, we reproduce key results of the MD study using the interactive *MyPal* application.

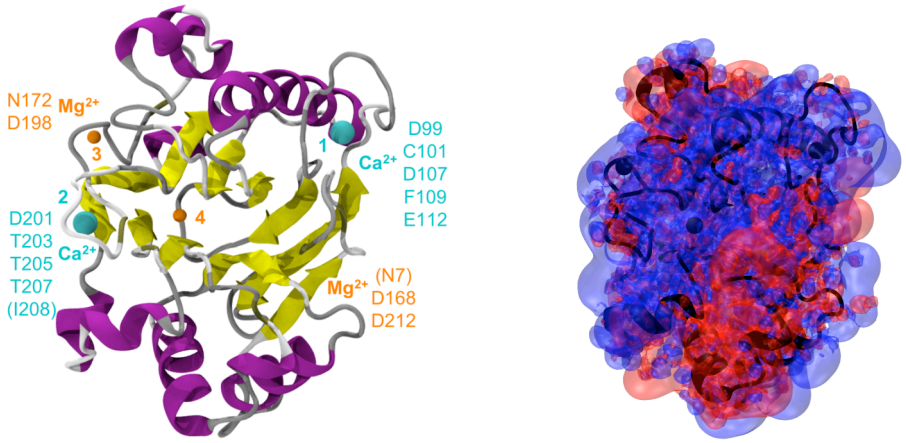


Figure 4: On the left, DNase I cation binding sites numbered from 1 to 4. Calcium binding sites are shown in cyan, magnesium binding sites in orange. The main amino acid residues constituting each site are indicated. On the right, electrostatic potential isosurfaces for potential values of -15.0(blue)/+15.0(red) kT/e are superposed onto the DNase I backbone structure and ions (in black).

3.2. *Locating the four previously identified ion binding sites with MyPal using interactive exploration at atomic resolution*

Figure 4 (right panel) illustrates the difficulty of visually identifying ion binding pockets due to the complexity of typical macromolecular electrostatic potentials. Even knowing their location (Figure 4, left panel), the binding sites cannot be distinguished. By using tactile feedback with our interactive *MyPal* application, it was however possible to locate all four binding sites with confidence and high precision (e.g. 0.9 Å positional deviation for the first site, mainly limited by the spacing of the potential grid). The interactive approach proved quick and intuitive.

Starting from a relaxed reference structure extracted from the MD trajectories, we interactively scanned the entire electrostatic potential surface of DNase I. An initial potential was calculated considering standard physiological ionic strength (0.15M NaCl). Na^{+}, Ca^{2+} and Mg^{2+} are potential ionic probes. For the identification of the binding sites we primarily used a hard sphere representation of the small divalent Mg^{2+} cation as it offers several advantages as a probe. Its double charge facilitates long-range electrostatic steering towards the binding pockets and its small size (0.79 Å) increases the accuracy for sensing the rough and detailed molecular surface at atomic resolution. The structural model of DNase I was extracted from MD simulations with binding sites occupied by ions that are mainly coordinated by amino acids and not water molecules. Consequently, it is important to consider a "naked" ion and not a hydrated one as probe. Otherwise the ion would not fit into the preformed coordination sphere at each site. More generally, considering hydrated ions using an implicit hydration shell remains feasible by increasing the probe radius and possibly varying the charge in order to take into account solvent shielding.

The parameters used for calculating the electrostatic potential affect the interactive exploration, in particular ionic strength. Taking into account ionic strength leads to locally more accurate potential maps. The downside of this for interactive experiments is that the detection of potential wells becomes harder as the range of electrostatic steering shortens. Without ionic strength, we achieved comparable precision, but did more easily detect several of the binding sites. The results for both series of experiments, with and without ionic strength, are presented in Figure 5. Despite the simplicity of the representations of protein surface and ions, all four ion binding sites identified by MD are retrieved by our approach. No other strong binding sites (false positive

locations) were detected, which can be a problem with other approaches [26]. We did observe several attractive interactions located at pockets on the enzyme surface. These locations correspond to shallow minima requiring little force to extract the ions and were not considered as binding sites.

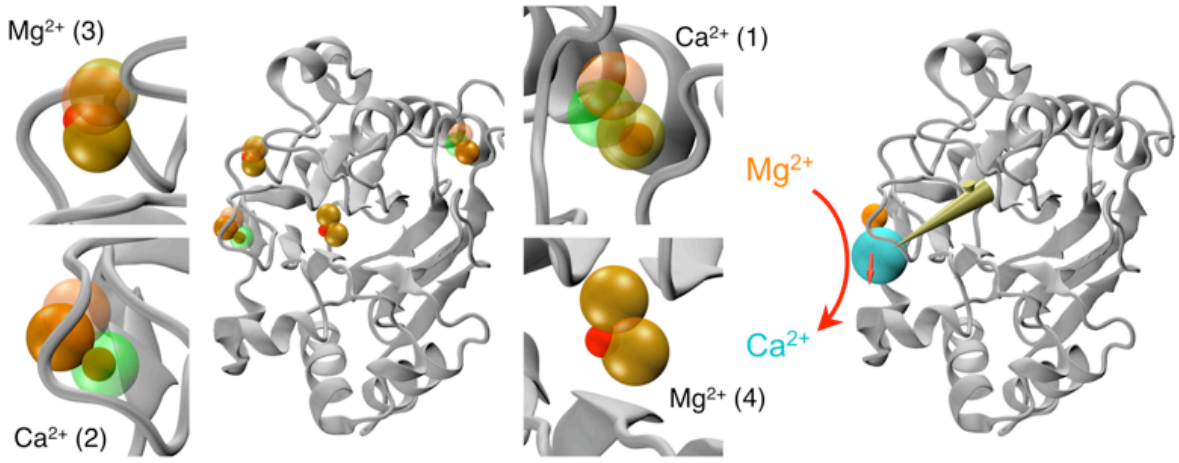


Figure 5: Visual summary of the interactive experiments. On the left, the results of a series of experiments for detecting *a priori* unknown ion binding sites are shown. The reference position of each binding pocket as determined by MD simulation is shown as a red sphere. Transparent spheres display *MyPal* predictions obtained for a potential map with (orange) or without (green) ionic strength. On the right, an ion substitution experiment ("molecular-billiard") at site 2 is depicted. Such an experiment probes the selectivity of a given ionic pocket for different ions.

3.3. *"Molecular-billiard" for probing site selectivity by ion substitution*

In the previous section, we successfully described how to locate the four DNase I cation binding sites. We did however not assess their selectivity. For this purpose, we have carried out a series of additional experiments, each starting with a different ionic probe at a given site. We then tried to interactively substitute the probe by another ion. We considered magnesium, calcium, sodium and chloride ions as probes. Figure 5 and Table 1 illustrate and summarize the results for these ion substitution "molecular-billiard" simulations.

As might be expected, chloride as an anion cannot be stabilized within any of the four cation binding pockets, nor can it displace a bound cation. Sites 1 and 2 are calcium selective, which is generally verified. It is however surprising that magnesium is able to substitute for calcium at site 1. This may be related to the simplicity of our model in which selectivity depends on the shape of the pocket itself and the pathway for accessing it. Generally speaking, buried and narrow sites are unreachable for large ions, whereas sites localized at the enzyme surface are readily subject to ion exchange. In the latter case, haptic feedback helps the user to distinguish between favourable and unfavourable substitutions. The favourable substitution of Na^+ with Ca^{2+} requires little user forces, whereas Na^+ can only displace Ca^{2+} using excessive force. Generally, Mg^{2+} can displace Ca^{2+} but the opposite remains impossible, except for the exposed calcium selective site 2. Sites 3 and 4 are selective towards the small divalent Mg^{2+} cation (r_{Mg2+}=0.79 Å). Site 4 is deeply buried and does not permit binding of the two bigger positive ions (r_{Ca2+}=1.71 Å ; r_{Na+}=1.87 Å). Site 3 can accommodate sodium and calcium in agreement with the 1ATN crystal structure [10]. Those ions are however easily displaced by magnesium. On the other hand, magnesium cannot be displaced by any of the other two cations, corroborating the magnesium selectivity of this site.

Table 1. Ion substitution simulation results. The table indicates whether exchange from X to Y is possible (→) or impossible (↛). For instance, Ca ↛ Cl means that Ca^{2+} cannot be displaced by Cl^-. A minus sign indicates that initial positioning of the chosen probe ion at the given binding pocket was not possible *via* our approach.

Probe (Site)	Ca^{2+} (1)	Ca^{2+} (2)	Mg^{2+} or Ca^{2+}(3)	Mg^{2+} (4)
Mg^{2+}/Ca^{2+}	Ca → Mg Mg ↛ Ca	Ca → Mg Mg → Ca	Mg ↛ Ca Ca → Mg	Mg ↛ Ca -
Na^+	Ca ↛ Na Na → Ca	Ca → Na Na → Ca	Mg ↛ Na Na → Mg	Mg ↛ Na -
Cl^-	Ca ↛ Cl -	Ca ↛ Cl -	Mg ↛ Cl -	Mg ↛ Cl -

The selectivity assessment remains a qualitative one, in particular due to the simplicity of the metal model. Recent work seems to indicate that induced polarization effects in the coordination sphere might be required to fully capture metal ion binding [27]. Such an improved potential function would presumably require further refinements, such as a completely polarizable forcefield, flexible sidechains and explicit water molecules [28, 29]. In our approach, we intentionally use approximations in order to speed up the search. More refined representations should be used *a posteriori* in order to refine the results of this initial haptic search.

The current implementation of *MyPal* was not designed in order to provide precise quantitative binding affinity estimates, but to be capable of distinguishing in real time between non-existing, weak and strong ion binding sites and assess the relative selectivity of significantly different ionic probes. The approximations made in the choice of the model representation limit the precision that could be obtained with the current implementation. Despite these limitations, it remains in principle possible to quantify the strength of binding. This aspect will be addressed in future work. One possible approach would be to calculate the work required by the user to extract an ion from its binding site.

These future extensions would also be beneficial for applications in docking and drug discovery. *MyPal* is already capable of handling rigid multi-atomic drug-like molecules. It is possible to displace a probe and assess whether it interacts favourably with cavities and pockets of a macromolecule. Currently the application does however lack tools to assess the ranking of different binding poses and the comparison between different probes. The implementation of such tools using the non-bonded interactions between target and probe molecules would be straightforward.

3.4. *Predicting the location of ionic binding pockets: multi-resolution exploration using a coarse grained protein shape*

Up to this point we have described experiments on a model structure with preformed ion binding sites. In this sub-section we validate our interactive approach using DNase I extracted from the DNase I/DNA 1DNK structure [11], crystallized without any divalent cation (ion-free DNase I). The expected locations of the binding sites were defined by superimposing the MD model onto the crystallographic structure.

We start with the same representation as previously, using a high-resolution atomistic model in order to describe the protein shape. All four binding sites can be located, albeit with reduced accuracy. The static conformation captured by the ion-free crystal structure biases the exploration by imposing an ill-suited local shape for the putative binding sites. For instance, in the case of the Mg^{2+} binding sites, the most solvent accessible and thus the least affected site deviates 2.5 Å from the reference location. Site 4, the most buried – *e.g.* potentially most biased – pocket can only be detected with significantly less precision and a deviation up to 3.9 Å from the expected site.

By lowering the resolution of the molecular shape representation, it should be possible to reduce these artefacts observed at atomic resolution. We used a coarse-grain representation with several beads per residue. In these simulations, the offset of the most buried magnesium site largely decreased from 3.9 to 2.2 Å. This decrease represents an important gain in precision. Localization of the more solvent-accessible pockets benefits to a lesser extent from coarse graining, with improvements ranging from 0.2 Å (site 3) to 0.4 Å (site 2).

Introducing flexibility in the model is an alternative to lowering the resolution of the representation. The protein or the complex could then be deformed in order to fit a binding interaction. The haptic device would help to render deformations fast and easy. Flexibility requires the introduction of bonded interactions in the model. Such a development is currently under way in our laboratory.

A positive conclusion can be drawn on the predictivity of our method. The multi-resolution approach combining an electrostatic potential grid computed from an all-atom structure and a coarse-grained molecular surface representation for steric repulsion is efficient for predicting ion binding sites in a quick and interactive manner. As an explanation, we may venture that the electrostatic contribution is a cooperative effect of several coordinating charged groups. A single misplaced sidechain is not critical and atomic resolution is appropriate even for distorted binding sites. A single atom sterically occluding the binding site is however a severe problem, calling for a low resolution representation of the molecular shape or alternatively for the introduction of flexibility in the model.

4. Conclusion and outlook

MyPal is a method for interactively locating ion binding sites by steering ionic probes into electrostatic potential maps using a haptic device. This multi-resolution approach, coupled with interactive devices is an important improvement of existing methods. This new strategy already emerges as potentially useful for other applications such as docking small ligands on proteins. Here we show how it facilitates the discovery of new relevant ion

binding sites by successfully retrieving the location of cation binding sites in DNase I and assessing their selectivity, combining atomic and coarse-grained resolutions.

The interactive experiments described in this article are particularly well adapted in order to acquire an intuitive understanding of the possible role of counterions in a biomolecular system. They can be carried out without any specific equipment, just using a mouse and on-screen visual feedback. However, we show that tactile feedback is essential for an efficient exploration of complex electrostatic potentials. Generally speaking, immediate feedback from the simulation adds a new dimension of user perception. Haptic feedback in particular is very intuitive and further increases the understanding of a given system. The force resistance allows the user to experience physical effects such as electrostatic steering or steric repulsion while extracting or inserting an ion in a binding pocket. A movie illustrating some interactive *MyPal* experiments on DNase I is available [22].

MyPal is complementary to existing methods such as deterministic or random computer generated searches for ionic binding sites. The human element based on the user's expertise is an essential feature of our approach. At the same time this human element makes it difficult to exhaustively test *MyPal* and compare it to other methods. Different users will use *MyPal* differently and a user's attention will wear off with time. Locating four ion binding sites such as in the case of DNase I remains feasible in a single work session. Such an exploration guided by the expertise of the user should be used prior to more computationally expensive approaches. For example, the 4th ion binding site in DNase I was quickly identified with *MyPal*, whereas it only appeared after extensive equilibration in an MD simulation. The MD simulation does however enable a very detailed and quantitative characterization of ion binding with extensive statistics.

The *MDDriver* library developed in order to couple molecular simulation engines with a visualization tool was used here in the context of a multi-resolution and multi-physics application. *MDDriver* is readily available on http://mddriver.sourceforge.net. The *MyPal* software implementing the simulation methodology presented in this manuscript, as well as the graphical tool designed for interactive simulations and simple visualization tasks based on Cocoa and VTK, will both be made available very soon. Important future improvements may concern visualization, *e.g.* the implementation of volume rendering, and ways to quantify, record and analyze interactive experiments more systematically.

Many future applications for *MyPal* can be imagined. Ions play a particularly central role inside ion channels and their selectivity, gating mechanisms and preferred ionic transport pathways could be investigated *via* interactive simulations. Closer to the example of the DNase I enzyme, the nucleosome still remains uncharted territory when it comes to locating its ion binding sites and assessing their functional role. Ion channels in membrane proteins and ions bound to the nucleosome are good examples of very large systems for which *MyPal* could provide a simple and straightforward method to capture molecular properties. Substantial time savings regarding simulation time and analysis are expected compared to classical approaches using MD. Furthermore, even at the present stage, *MyPal* is not limited to the docking of ions or small ligands, and should therefore be generally applicable in many biological domains. For instance, it opens the prospect to investigate larger molecular assemblies and their interactions. As an example, one could imagine moving the DNase I enzyme described in this manuscript along its DNA substrate to probe sequence specificity. This would be particularly challenging for a nucleosomal DNA substrate.

Acknowledgments

This interdisciplinary work was supported by the DEISA Consortium (co-funded by the EU, FP6 projects 508830/031513) and IDRIS (CNRS's National Supercomputer Center in Orsay). The project was funded by the French Agency for Research (grants ANR-06-PCVI-0025 and ANR-07-CIS7-003-01). We thank Christopher Amourda, Anthony Bocahut, Chantal Prévost and Sophie Sacquin-Mora for providing useful feedback and testing the method presented in this manuscript on a selection of biological systems. Their critical comments enabled us to continually improve the *MyPal* application.

References

1. M. Guéroult, J. Abi Ghanem, B. Heddi, C. Prévost, P. Poulain, M. Baaden and B. Hartmann, *in Albany 2009: Conversation 16* (2009).
2. O. Delalande, N. Férey, G. Grasseau and M. Baaden, *J. Comput. Chem*, in press, published online (2009).
3. H. Nagata, H. Mizushima and H. Tanaka, *Bioinformatics* **18**, 140 (2002).
4. N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J.P. Changeux, M. Delarue and P. J. Corringer, *Nature* **457**, 111 (2009).
5. R. C. Wade, R. R. Gabdoulline, S. K. Lüdemann and V. Lounnas, *Proc. Natl. Acad. Sci. USA* **95**, 5942 (1998).

6. D. Tworowski, A. V. Feldman and M. G. Safro, *J. Mol. Biol.* **350**, 866 (2005).
7. R.E. Kozack, M.J. d'Mello and S. Subramaniam, *Biophys. J,* **68**, 807 (1995).
8. L. Hemsath, R. Dvorsky, D. Fiegen, M.-F. F. Carlier and M.R. Ahmadian, *Mol Cell* **20**, 313 (2005).
9. C. Oefner and D. Suck, *J. Mol. Biol.* **192**, 605 (1986).
10. W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai and K. C. Holmes, *Nature* **347**, 37 (1990).
11. S. A. Weston, A. Lahm and D. Suck, *J. Mol. Biol.* **226**, 1237 (1992).
12. D. A. Case *et al.* (2002) AMBER 7 (University of California, San Francisco).
13. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucl. Acids Res.* **32**, W665 (2004).
14. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **98**, 10037 (2001).
15. M. Zacharias, *Protein Sci.* **12**, 1271 (2003).
16. C. Azuara, H. Orland, M. Bon, P. Koehl and M. Delarue, *Biophys. J.* **95**, 5587 (2008).
17. M. Baaden and R. Lavery, *in Recent Adv. in Protein Engineering*, 173 (2007).
18. W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics* **14**, 33 (1996).
19. N. Férey, O. Delalande, G. Grasseau and M. Baaden, *in Proceedings of the 15th ACM Symposium on Virtual Reality Software and Technology*, 91 (2008).
20. J. E. Stone, J. Gullingsrud, K. Schulten and P. Grayson, *in Proceedings of the 2001 ACM Symposium on Interactive 3D Graphics*, 191 (2001).
21. W. Schroeder, K. Martin and B. Lorensen, The *Visualization Toolkit An Object-Oriented Approach To 3D Graphics, 4th Edition*.
22. http://mddriver.sourceforge.net/
23. R. M. Taylor II, T.C. Hudson, A. Seeger, H. Weber, J. Juliano and A.T. Helser, *in Proceedings of the ACM Symposium on Virtual Reality Software & Technology,* 55 (2001).
24. http://www.gromacs.org
25. D. S. Glazer, R. J. Radmer and R. B. Altman, *Structure* **17**, 919 (2009).
26. C. Claperon, R. Rozenfeld, X. Iturrioz, N. Inguimbert, M. Okada, B. Roques, B. Maigret and C. Llorens-Cortes, *Biochem. J.* **416**, 37 (2008).
27. G. Kupparaj, M. Dudev and C. Lim, *J. Phys. Chem. B* **113**, 2952 (2009).
28. D. Jiao, C. King, A. Grossfield, T. A. Darden and P. Ren, *J. Phys. Chem. B* **110**, 18553 (2006).
29. F. Jalilehvand, D. Spångberg, P. Lindqvist-Reis, K. Hermansson, I. Persson and M. Sandström, *J. Am. Chem. Soc.* **123**, 431 (2001).

„It would appear that we have reached the limits

of what it is possible to achieve with computer technology,

although one should be careful with such statements,

as they tend to sound pretty silly in five years.“

John von Neumann, environ 1949