



HAL
open science

Dynamique d'assemblage de la capsidie des norovirus

Thibault Tubiana

► **To cite this version:**

Thibault Tubiana. Dynamique d'assemblage de la capsidie des norovirus. Biologie structurale [q-bio.BM]. Université Paris-Saclay, 2017. Français. NNT : 2017SACLS307 . tel-01773889

HAL Id: tel-01773889

<https://theses.hal.science/tel-01773889>

Submitted on 23 Apr 2018

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

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

Dynamique d'assemblage de la capside des norovirus

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°569 Innovation thérapeutique :
du fondamental à l'appliqué
Spécialité de doctorat : biochimie et biologie structurale

Thèse présentée et soutenue à Saclay, le 26 octobre 2017, par

Thibault Tubiana

Composition du Jury :

Mme. Sylvie NESSLER Professeure, Université Paris Sud – I2BC	Présidente du jury
M. Florent BARBAULT Maître de conférence, Université Paris 7 - ITODYS	Rapporteur
M. Patrice GOUET Professeur, Établissement – IBCP	Rapporteur
M. Marc BAADEN Directeur de recherche, CNRS – LBT	Examineur
M. Stéphane BRESSANELI Directeur de recherche, CNRS – I2BC	Directeur de thèse
M. Yves BOULARD Directeur de recherche, CEA	Co-Directeur de thèse

Dynamique d'assemblage de la capsid des Norovirus

Thibault Tubiana © 26 octobre 2017

Thèse de doctorat en Sciences de la Vie

Rapporteurs : Pr. Patrice Gouet et Dr. Florent Barbault

Directeur de thèse : Dr. Stéphane Bressanelli

Co-Directeur de thèse : Dr. Yves Boulard

Université Paris Saclay

École doctorale Innovation Thérapeutique : du fondamental à l'appliqué

Équipe IMAPP : Interactions et Mécanismes d'Assemblage des Protéines et des Peptides

I2BC - CNRS UMR9198

1 avenue de la terrasse

91190 Gif-Sur-Yvette

Remerciements

Je tiens tout d'abord à remercier le Pr Patrice Gouet et le Dr Florent Barbault pour m'avoir fait l'honneur d'être les rapporteurs de mes travaux de thèse. Merci également à la présidente du jury, le Pr Sylvie Nessler. Je remercie aussi le Dr Marc Baaden pour avoir accepté d'être jury dans mon comité de mi-thèse ainsi qu'à ma soutenance de thèse.

Je ne pourrais pas ne pas mentionner dans ces remerciements mes directeurs de thèse : les Drs Stéphane Bressanelli et Yves Boulard, qui ont été de grands mentors, des superviseurs géniaux, disponibles, à l'écoute et qui ont permis que cette thèse se déroule dans les meilleures conditions et dans la meilleure ambiance. Merci à eux de m'avoir donné cette chance incroyable.

Je remercie la région Île-de-France et la DIM Malinf dont les financements ont permis à cette thèse de se réaliser, qui plus est dans d'excellentes conditions.

Je remercie également ma mère et mon père, Sylvie Gallon et Bernard Tubiana, pour m'avoir soutenu durant toutes ces années, m'avoir laissé la liberté de choisir mes études et de m'avoir aidé sur le plan moral et financier durant toutes celles-ci. Merci également à ma grand-mère, Hélène Gallon, qui depuis que je suis tout petit me chouchoute et m'aide dans tous les aléas de la vie. Merci à ma belle-mère, Claude Cohen, dont les nombreux conseils m'ont permis de me préparer au mieux à certains entretiens et à la vie professionnelle en général. Également un grand merci à Sylvain Mounis, mon soutien de tous les jours, sans qui je n'aurais pas pu aborder la majorité de ma thèse sereinement et qui a pris le temps de corriger l'intégralité de ce manuscrit.

Merci également à Fabrice Allain pour ces nombreuses discussions scientifiques ou ludiques et ces après-midi de jeux de société qui nous permettaient

de décompresser lorsque nous en avons besoin. Je ne peux également pas m'empêcher de penser à toutes ces personnes qui m'ont permis d'arriver là où j'en suis aujourd'hui : je pense à mes amis Paule Labaune et Arnaud Recoquillay qui m'ont aidé là où j'avais des lacunes dans ces lointaines années lycéennes, Sonia Ziadia qui m'a incité à me tourner vers la licence de bio-informatique et Anne Badel qui, malgré mon dossier particulier, m'a donné ma chance dans ce parcours et qui, par la suite, m'a permis d'enseigner dans la Licence et Master Bioinformatique de Paris 7. Merci à Cathy Etchebest pour tous ses conseils prodigués durant mes années d'étude et au détour d'un congrès, qui m'ont permis d'avancer dans mes projets éducatifs et de recherche. Merci à Patrick Fuchs, Gautier Moroy et Pierre Poulain pour ces heures d'enseignement partagées au sein du Master Bioinformatique à Paris 7. Merci à eux et à tous ceux que j'aurais oubliés ici.

Je tiens aussi à remercier Nathalie Reuter qui m'a accueilli quelques mois dans son équipe à Bergen dans ce magnifique pays qu'est la Norvège, puis Marc Baaden qui a été le premier à me donner la chance de présenter dans un congrès international.

Je remercie également Jean-Charles Carvaillo, qui a été stagiaire de M2 sous ma supervision et qui a été essentiel à la réalisation de la dernière partie de cette thèse. Je voudrais bien entendu remercier tous les membres de l'équipe Interaction et Mécanisme d'Assemblage des Protéines et Peptides pour m'avoir supporté durant ces trois années, mais également pour la superbe ambiance qui y régnait tous les jours. Merci à Maité Paternostre, Sonia Fieulaine, Kaouther Ben Ouirane, Keinny François, Johann Habersetzer, Khac Minh Huy Nguyen, Maelenn Chevreuil, Christina Petcut, sans oublier mon voisin de bureau bien sûr : Pierre Chervy, alias Pierrot la Lune, qui a partagé mes délires et mes râleries pendant 3 ans. Merci également à Julien Maufront et Cyntia Taveneau qui m'ont permis de passer de bons moments quand ils étaient dans l'équipe.

Enfin, j'aimerais profiter de ces remerciements pour faire une mention toute personnelle. Lorsque j'avais 10 ans, mon oncle, Régis Gallon, m'a offert une cassette vidéo de la Cité des Sciences sur le système solaire dont les commentaires étaient dits par un homme passionné : André Brahic (que j'ai

eu la chance et la surprise d'avoir comme qu'enseignant durant quelques cours d'exobiologie à l'université Paris Diderot). Il a été, sans le savoir, parmi les personnes qui m'ont donné goût à la science, à la recherche et à la découverte et j'aime à penser qu'il a contribué à paver le chemin qui m'a mené jusqu'ici.



Thibault Tubiana

✉ tubiana.thibault@gmail.com • 🌐 www.tubiana.me
Née le 24/04/1990 • Permis B • <https://www.github.com/tubiana>

Éducation

Master de Bioinformatique - mention bien

Université Paris 7 - Diderot

2012-2014

Paris (France)

- Bioinformatique structurale (Modélisation et dynamique moléculaire, docking, drug design)
- Bioinformatique génomique et génomique fonctionnelle
- Bioinformatique intégrative systématique et protéomique
- Analyse de données, apprentissage des données
- Programmation avancée, algorithmique avancée

Licence de Bioinformatique

Université Paris 7 - Diderot

2009-2012

Paris (France)

- Structure et intégration des macromolécules
- Génétique et biologie moléculaire
- Programmation et algorithmique
- Base de données
- Statistique en biologique

Lycée Georges Sand

Baccalauréat Scientifique - option SVT - mention assez bien

2005-2009

Domont (France)

Expérience

Doctorat

I2BC - Institut de Biologie Intégrative de la cellule

Sous la direction du Dr. Stéphane Bressanelli et Dr. Yves Boulard

- Etude *in silico* de l'autoassemblage de la capsid des norovirus

Novembre 2014 – Actuellement

CEA Saclay (France)

Encadrement d'un stagiaire

Université Paris Diderot - Paris 7

Simulation gros grains du pentamère et hexamère de dimère de VP1 de Norovirus

Janvier – Juin 2017

CEA Saclay (France)

Monitorat

Université Paris Diderot - Paris 7

- Bioinformatique
- Biostatistique
- Programmation Python
- Programmes de visualisation scientifique

Septembre 2014 – Actuellement

Paris (France)

Stage de fin de master

Laboratoire de Biochimie Théorique (sous le tutorat de Dr. Marc Baaden)

- Développement de nouvelles visualisations standardisées des polysaccharides pour un logiciel de visualisation moléculaire basé sur un moteur de jeu 3D (Unity) : UnityMol.

Janvier – Juillet 2014

Paris (France)

Assistant de recherche

Computational Biology Unit (sous le tutorat du Pr. Nathalie Reuter)

- Etude de la présence de "cages aromatiques" sur un jeu de données de protéines périphériques, permettant la liaison de la protéine à des phosphatidylcholines sur une membrane cellulaire.

Juin – Juillet 2013

Bergen (Norvège)

Stage de master 1

Computational Biology Unit (sous le tutorat du Pr. Nathalie Reuter)

- Etude statistique de la distribution des acides aminés pour les protéines périphériques et transmembranaires.

Mars – Mai 2013

Bergen (Norvège)

Stage d'été

MTI (Molécules Thérapeutiques In Silico)

- Développement d'une plateforme coopérative de "serious gaming" pour les étudiants de première année

Juillet – Août 2012

Paris (France)

Stage de découverte de la recherche

Unité postulante de pathogenèse de *Helicobacter*, Institut Pasteur

- Dosage enzymatique de la quantité d'ammoniac produite par des *H. pylori* vivants, mutants ou non, selon le substrat qui est apporté.

1 semaine – Avril 2010

Paris (France)

Récompenses

Faraday Discussion 169
Prix Skinner - Meilleur poster étudiant

Nottingham (Angleterre)
2014

Autres projets

Tutoriel d'installation Pymol Windows/Mac/linux

Plus de 60000 téléchargement cumulés des archives pour l'installation sur Windows. 1000+ visites par mois
<http://tubiana.me/how-to-install-and-compile-pymol-windows-linux-mac/>

TTClust : A versatile molecular simulation trajectory clustering program

Programme Python pour le clustering hiérarchique de n'importe quelle trajectoire de dynamique moléculaire.
<https://github.com/tubiana/TTClust>

Compétences Informatiques

Langages Informatiques

Maîtrise: Python, R, C++

Bonne connaissance: Java, C, C#, Bash, L^AT_EX

Notions: HTML, JavaScript, Tcl/Tk

Logiciels

Maîtrise: Modeller, Pymol, VMD, Microsoft Office, Gromacs

Bonne connaissance: AMBER, NAMD, Unity3D, Maestro, Open Office

Notions: AutoDockTools, PostgreSQL

Systèmes d'exploitation

Maîtrise: Linux, Windows, Mac

Langue

Français : Langue maternelle

Anglais : courant

Centres d'intérêt

Piano (10 ans de cours)

Sciences aérospatiales

Sécurité informatique

Table des matières

Table des matières	ix
Table des figures	xiv
Liste des tableaux	xx
Abbréviations	xxi
Glossaire	xxii
A Communications scientifiques	0
1 Présentations	1
2 Posters	2
3 Articles	3
I Introduction générale	4
1 Les norovirus : Généralités	5
1.1 Historique	5
1.2 Taxonomie	6
1.3 Propagation du norovirus	7
1.3.1 Résistance du norovirus	7
1.3.2 Infectivité du norovirus	8
1.4 Impacts socio-économiques	10
1.4.1 Impact social	10

1.4.2	Impact économique	11
1.5	Piste thérapeutique	12
1.5.1	Utilisation de particules virales vides	12
1.5.2	Inhibition du norovirus	13
1.6	Organisation du Norovirus	13
1.6.1	Organisation génomique	13
1.6.2	Organisation des capsides icosaédriques	15
1.6.3	Organisation structurale de la capside	17
2	Théories d'assemblages	23
2.1	Théorie d'assemblage des petits virus à ARN positif simple brin	23
2.2	Modèles d'assemblage du norovirus	24
3	Virologie et simulation de dynamique moléculaire	26
3.1	Historique	26
3.2	Quelques simulations de capsides virales	27
3.3	Assemblage de capside virale <i>in silico</i>	29
3.4	Étude du norovirus <i>In silico</i>	32
II	Matériels et méthodes	36
1	Méthodes <i>in silico</i>	37
1.1	Modélisation par homologie	37
1.1.1	Principe	37
1.1.2	La modélisation comparative avec MODELLER	38
1.2	La simulation de dynamique moléculaire	41
1.2.1	Principe	41
1.2.2	Calcul de l'énergie potentielle	42
1.2.3	Conditions périodiques aux limites	48
1.2.4	Ensemble thermodynamique	49
1.3	Méthodes avancées	52
1.3.1	Simulation gros grains	52
1.3.2	Recuit simulé	54
1.3.3	Dynamique Moléculaire Dirigée (TMD)	56
1.3.4	Ajustement par dynamique moléculaire flexible	57
1.3.5	L'analyse par modes normaux	58

1.4	Paramètres utilisés dans nos simulations	59
1.4.1	Simulations de dynamique moléculaire classiques . . .	59
1.4.2	Simulations gros grains	60
1.4.3	Recuit simulé	63
1.4.4	Dynamique moléculaire dirigée	63
2	Méthodes expérimentales	64
2.1	S.A.X.S.	64
2.1.1	Principe	64
2.1.2	Analyse	66
3	Outils développés pendant la thèse	75
3.1	TTclust	75
3.1.1	Fonctionnement	76
3.1.2	Article	78
3.1.3	Étude d'un cas complémentaire à l'article	80
3.2	Comp. en masse de PDBs à des données SAXS	82
3.3	Comparaison d'angles entre deux domaines protéiques	83
3.4	Module Pymol : représentation de la distance entre 3 atomes par un triangle	84
III	Dynamique et asymétrie du dimère de la protéine de capsid du Norovirus	86
1	Contexte	87
2	Dynamics and asymmetry in the dimer of the norovirus major capsid protein	89
2.1	Article	89
2.2	Figures supplémentaires	111
3	Complément à l'article	116
3.1	Simulation d'un monomère	116
3.1.1	Contexte	116
3.1.2	Préparation	116
3.1.3	Étude de la stabilité d'un monomère C	117

3.1.4	Conclusion de la simulation du monomère C	119
3.2	Modélisation de la capsid de norovirus de GIII.2	121
3.2.1	Modélisation des domaines	122
3.2.2	Alignements dans la carte de densité	123
3.2.3	Ajustement avec MD-FLEXIBLE FITTING	124
3.2.4	Reconstruction atomique et améliorations possibles . .	127
3.3	Simulation de Dynamique moléculaire dirigée - GI.1	128
3.3.1	Contexte	129
3.3.2	Réalisation des deux modèles	130
3.3.3	Évolution du RMSD sur 1000 ps	131
3.3.4	Perspectives	133
4	Discussion complémentaire à l'article	134
IV	Étude d'intermédiaire d'assemblage de la capsid de Norovirus.	136
1	Contexte	137
2	Préparation des systèmes	139
2.1	Modélisation de l'unité asymétrique	140
2.2	Modélisation des capsomères	140
2.3	Précision de terminologie	141
3	Dynamiques du pentamère et de l'hexamère de dimère VP1	142
3.1	Simulations gros grains	142
3.1.1	Évolution du RMSD	142
3.1.2	Distance entre dimères	144
3.2	Analyse des modes normaux sur le pentamère et l'hexamère de dimères	147
3.3	Conclusion de l'étude de la dynamique des capsomères.	148
4	Docking d'un dimère interstitiel sur un pentamère de dimères	150
4.1	Clustering de la trajectoire	150
4.2	Docking avec ClusPro	152
4.2.1	Préparation	152
4.2.2	Résultats	152

4.2.3 Conclusion	154
5 Conclusions sur les intermédiaires d'assemblage de la capsid du norovirus.	155
V Discussion	157
VI Perspectives	166
Résumé	191

Table des figures

1.1	Première image du norovirus	6
1.2	Dendrogramme des 29 séquences représentatives de VP1	7
1.3	Voie de transmission du norovirus	9
1.4	Répartition par région des cas d'infections et de décès dus au norovirus	11
1.5	Répartition par tranche d'âge des cas d'infection et de décès dus au norovirus.	11
1.6	Liaison du citrate sur le domaine P de VP1	13
1.7	Représentation schématique de l'ARN génomique du norovirus. La succession des protéines codées dans les 3 ORF est indiquée. Les cadres ne sont pas à l'échelle.	14
1.8	Modélisation des NS1-2, NS3 et NS4 humaines et association putative à la membrane.	15
1.9	Exemple de calculs des paramètres h et k et du nombre T	16
	(a) Symétrie icosaédrique $T=3$	16
	(b) Symétrie icosaédrique $T=9$	16
	(c) Symétrie icosaédrique $T=27$	16
1.10	Description des axes de symétrie dans une capsidie icosaédrique.	17
	(a) Représentation schématique des axes de symétrie	17
	(b) Exemples sur la capsidie du norovirus	17
1.11	Structure de la protéine majeure de capsidie VP1.	18
1.12	VP1 : Site de liaisons d'un polysaccharide de groupe sanguin	19
1.13	Représentation de l'orientation du domaine S par rapport au domaine P pour les trois conformations.	20
1.14	Unité asymétrique et assemblages de dimères de la capsidie du norovirus	20

1.15	Capside du norovirus colorée par domaine et par chaînes	21
1.16	Région de VP1 en interaction avec VP2	22
1.17	Hypothèse d'assemblage de Devkota <i>et al.</i>	23
1.18	Charge nette des bras polypeptidiques vs. la taille du génome du virus	24
1.19	Modèle d'assemblage du norovirus de Prasad <i>et al.</i>	24
1.20	Modèle d'assemblage de Tresset <i>et al.</i> (2013)	25
	(a) Forme de l'intermédiaire d'assemblage du norovirus	25
	(b) Modèle d'assemblage de Tresset <i>et al.</i> (2013)	25
1.21	Les différentes échelles de simulation (taille/temps)	26
1.22	Exemple d'étude d'auto-assemblage par Chen <i>et al.</i>	30
	(a) Description du modèle gros grains utilisé	30
	(b) Résultats de la simulation de Monte-Carlo	30
1.23	Assemblage de la capside du VIH avec un modèle gros grains 3D	31
1.24	Assemblage de capside icosaédrique (T=1) par Johnston <i>et al.</i> . .	32
1.25	Exemple de nanoindentation sur la capside du norovirus.	33
1.26	Résultats des analyses par nanoindentation sur la capside du Norovirus	33
2.1	Nombre de structures déposées dans la PDB	37
2.2	Modélisation par homologie et identité de séquence : zone de confiance.	39
2.3	Distance entre deux atomes.	43
2.4	Exemple d'un potentiel harmonique : variation de l'énergie en fonction de la longueur de liaison	43
2.5	Illustration d'un angle entre trois atomes.	44
2.6	Représentation d'un angle dièdre entre 4 atomes.	44
2.7	Exemple de la variation de l'énergie en fonction de l'angle dièdre pour un groupement chimique.	44
2.8	Interactions électrostatiques	45
2.9	Approche de type PME sur un réseau bidimensionnel	46
2.10	Interactions de Van der Waals.	47
2.11	Potentiel de Lennard Jones.	47
2.12	Exemple d'une boîte de simulation	48

2.13	Périodicité de la boîte de simulation	49
	(a) Boite de simulation	49
	(b) Atomes visible avec un rayon de coupure.	49
2.14	Représentation gros grains de chaque acide aminé	53
2.15	Comparaison des réseaux élastiques	54
2.16	Schéma explicatif du recuit simulé	55
	(a) MD - exploration	55
	(b) Recuit simulé - Chauffage	55
	(c) Recuit simulé - Refroidissement	55
	(d) MD - exploration	55
2.17	Exemple d'utilisation de MDFF	58
2.18	Schéma du protocole de simulation. Réalisé par Jean-Charles Carvaillo.	61
2.19	Exemple d'application de l'algorithme domELNEDIN	62
	(a) Avant domELNEDIN	62
	(b) Après domELNEDIN	62
2.20	Description des cycles de recuit simulé utilisés dans notre protocole.	63
2.21	Schéma descriptif du fonctionnement du SAXS	64
2.22	Composition d'un échantillon de SAXS.	66
2.23	Exploitation d'une courbe SAXS	67
2.24	Relation entre la forme d'une particule observé en SAXS et la forme de sa courbe de diffusion	68
2.25	Représentation du fonctionnement de DAMMIN	69
2.26	Représentation simplifiée du fonctionnement de DADIMODO.	72
2.27	Représentation simplifiée du fonctionnement de TTCLUST.	77
2.28	Résultat du clustering pour un système gros grains (pentamère de dimères de la protéine majeure VP1 de la capsid du norovirus)	80
2.29	Illustration de la surface d'énergie potentielle en fonction d'un système tout atomes ou gros grains	81
2.30	Représentation simplifiée du fonctionnement des scripts de comparaison pour le SAXS.	83
2.31	Illustration des axes principaux des domaines S et P.	83

2.32	Représentation simplifiée du fonctionnement du script pour la comparaison d'angle et de hauteur des domaines.	84
2.33	Exemple de représentation "triangle" entre trois atomes	85
2.34	Représentation simplifiée du fonctionnement du plugin Pymol représentant la distance entre trois atomes tout au long de la trajectoire.	85
3.1	Cluster de résidus acides autour de l'axe quasi-6 de la VLP de GI.1	88
3.2	RMSD durant la dynamique du monomère C et du dimère C/C .	118
	(a) RMSD - monomère VP1 (conf. C)	118
	(b) RMSD - Dimère C/C VP1 (C1)	118
	(c) RMSD - Dimère C/C VP1 (C2)	118
3.3	Angle formé entre l'axe Z du monomère et l'axe principale du domaine S	119
	(a) Représentation de l'angle calculé dans la figure 3.3b	119
	(b) Angle formé entre l'axe Z et le domaine S durant la dynamique	119
3.4	ACP sur la dynamique du monomère C	120
3.5	Interface utilisateur de VEDA (Visual Environment for Dockin Algorithms)	123
3.6	Comparaison des unités asymétriques avant/après MDFF (complètement flexible)	125
3.7	Conservation des structures secondaires avant/après simulation par MDFF avec une flexibilité complète.	125
3.8	Idem que pour les figures 3.6 et 3.7 mais avec uniquement le domaine S flexible.	126
3.9	Parties flexibles (rouge) et rigides (vert) utilisées dans le protocole de MDFF.	127
3.10	Comparaison des unités asymétriques avant/après MDFF (boucle et coil du domaine S flexible)	128
	(a) Vue du dessous	128
	(b) Vue de côté	128
3.11	Modélisation de la capsid de GIII.2	129
3.12	Évolution du RMSD durant les simulations de TMD	132

4.1	Schéma récapitulatif des modèles d'assemblage	138
4.2	Schéma de la préparation des capsomères (POD et HOD). Réalisé par Jean-Charles Carvaillo.	139
4.3	Précision de terminologie sur le POD et le HOD	141
	(a) POD	141
	(b) HOD	141
4.4	Flexibilité du POD et du HOD (10)	142
4.5	Flexibilité pour chaque dimère du POD et du HOD (10 μ s) . . .	143
4.6	Distance entre chaque dimère du POD et du HOD (10 μ s) . . .	144
4.7	Distance entre chaque dimère du POD pour chaque réplique . .	145
4.8	Alternance des positions des éloignements et rapprochements entre dimères dans les simulations du POD	146
	(a) Simulation	146
	(b) Réplique	146
4.9	Distances entre le centre de masse des domaines S périphériques	147
4.10	Mode normaux sur le POD - mode 8	148
4.11	Modes normaux sur le POD - comparaison mode 8/9	149
	(a) Mode 8	149
	(b) Mode 9	149
4.12	Hypothèse d'extension à partir d'un POD.	149
4.13	Résultats du clustering des trajectoires de MD du POD	150
4.14	Résultats du clustering des trajectoires de MD du HOD	151
4.15	Résultat des <i>dockings</i> dimère CC avec un POD - position	153
4.16	Résultat des <i>dockings</i> x CC avec un POD - énergie	153
5.1	Schéma d'assemblage proposé du dimère à la capsid	162
5.2	Analyse de modes normaux sur l'intermédiaire d'assemblage (11 dimères)	163
5.3	Hypothèse d'interaction entre intermédiaires d'assemblage . . .	163
5.4	Hypothèse de dimérisation du domaine S	164
6.1	Protocole proposé pour l'étude de la formation d'un POD et d'un intermédiaire d'assemblage	168

6.2	Simulation proposée pour étudier les interactions entre POD et dimères	169
-----	--	-----

Liste des tableaux

3.1	Exemple de simulations réalisées sur des capsides de virus . . .	28
3.1	Référence cristallographique pour chaque région modélisée de VP1 GIII.2	122
3.2	Description des boîtes de simulation pour les systèmes sauvage et mutant	131

Abbréviations

APE Agence de Protection de l'Environnement

BPTI Bovine Pancreatic Trypsin Inhibitor

CDC Centres pour le contrôle et la prévention des maladies

CG Gros Grains (ou *Coarse Grain*)

HBGA Antigène de groupe sanguin (ou *histo-blood-group antigens*)

HOD Hexamère de dimère

MD Simulation de Dynamique Moléculaire

MDFF Ajustement par dynamique moléculaire flexible (ou *Molecular Dynamics Flexible Fitting*)

ORF Cadre de lecture ouvert (ou *Open Reading Frame*)

PDB Protein Data Bank

POD Pentamère de dimère

SAXS Small Angle X-ray Scattering

TMD Simulation de Dynamique Moléculaire Dirigée

VLP Virus Like Particule (Particule virale vide)

VP1 Protéine constitutive de la capsid du norovirus

Glossaire

Algorithme génétique

Un algorithme génétique a pour but d'obtenir une "solution approchée à un problème d'optimisation". Par analogie à la notion de sélection naturelle, une variation aléatoire de population est générée puis sélectionnée. Il faut donc une fonction d'évaluation de l'adaptation au problème donné. Plusieurs générations successives sont réalisées permettant ainsi d'affiner les résultats au fur et à mesure.

Analyse en Composante Principale

Analyse statistique consistant à rechercher des variables collectives au sein d'un système complexe (multidimensionnel). Une relation entre les variables est établie par une analyse en variance-covariance permettant de mettre en avant les corrélations qui peuvent exister et caractériser les mouvements collectifs au sein d'un système de [MD\[1\]](#)

ARN polycistronique

Un ARN polycistronique est un ARN messager codant plusieurs chaînes polypeptidiques. Il possède plusieurs codons d'initiation et codons-stop.

Espace canonique

Espace contenant toutes les conformations possibles d'un système.

Fonction objectif

Équation visant à être optimisée avec un certain nombre de contraintes qui doivent être minimisées ou maximisées.

Maturation protéolytique

La maturation d'une protéine est un ensemble de modifications post-traductionnelles. Chez les virus à ARN simple brin de polarité positive, la maturation de la polyprotéine de réplication se fait par clivages protéolytiques spécifiques.

Moment total (ou *Momentum*)

somme de la quantité de mouvement d'un objet en mouvement.

Monochromatique

Une onde monochromatique est une onde qui n'est composée que d'une seule longueur d'onde, une seule fréquence. Elle peut être décrite par une fonction sinusoidale au cours du temps.

ORF

Un **ORF** est ce qu'on appelle un cadre de lecture ouvert, une région sans codon stop susceptible d'être traduite en protéine.

Screening virtuel

Le *screening* virtuel est une méthode *in silico* visant à tester la possibilité de liaisons de centaines de molécules dans un site actif ou l'ensemble d'une protéine.

Unité asymétrique

Pour un objet symétrique (capside régulière, oligomère, cristal...) unité structurale minimum nécessaire pour reconstruire l'objet par symétrie. Dans le cas du norovirus, l'unité asymétrique de la capsidie icosaédrique est composée de 3 conformations différentes de la protéine structurale VP1.

A

Communications scientifiques

- **GDR 3696 - Protéines membranaires** **PRAZ-SUR-ARLY**
2015
- **Workshop de modélisation moléculaire** **SACLAY**
2015
- **Congrès des doctorants de la DSV (CEA)** **PORQUEROLLES**
2015
- **GDR 3696 - Protéines membranaires** **LA-LONDE-LES-**
MAURES
2016
- **Atelier de Modélisation des Molécules d'Intérêt** **CHÂTENAY**
Biologique de Paris-Saclay (AMMIB) **MALABRY**
2016
- **International Society of Quantum Biology and** **BERGEN**
pharmacology (ISQBP) **(NORVÈGE)**
2016
- **Structural Biology Meets Biophysics (SFB + GTBIO)** **OBERNAI**
2016
- **Atelier de Modélisation des Molécules d'Intérêt** **ORSAY**
Biologique de Paris-Saclay (AMMIB) *2017*
- **Groupe de Graphisme et Modélisation Moléculaire** **REIMS**
(GGMM) *2017*
- **GDR 3696 - Protéines membranaires** **PORQUEROLLES**
2017

- **Groupe de Graphisme et Modélisation Moléculaire (GGMM)** **SÈTE**
2015
- **Journée des doctorants** **CEA SACLAY**
2015
- **International Society of Quantum Biology and pharmacology (ISQBP)** **BERGEN**
(NORVÈGE)
2016
- **Journée des doctorants** **CEA SACLAY**
2016
- **Structural Biology Meets Biophysics (SFB + GTBIO)** **OBERNAI**
2016
- **Journées francophones de virologie (JFV)** **PARIS**
2017

- **Dynamics and asymmetry in the dimer of the norovirus major capsid protein.** Tubiana T., Boulard Y., Bressanelli S. (2017). PLoS ONE 12(7) : e0182056. <https://doi.org/10.1371/journal.pone.0182056>
- **TTClust : A versatile graphical molecular simulation trajectory clustering program.** Tubiana T., Boulard Y., Bressanelli S.. *Soumis*.
- **Membrane alterations induced by nonstructural proteins of human norovirus.** Doerflinger S. Y., Cortese M., Romero-Brey I., Menne Z., Tubiana T., Schenk C., White P. A., Bartenschlager R., Bressanelli S., Hansman G. S., Lohmann V.. *En révision*
- **A Second Phosphoinositide-Interacting Polybasic Domain in Septin 9 Controls Golgi Assembly and Apico-Basal Polarity.** Omrane, M; Taveneau, C; Peng, J; Tubiana, T; Yu, J; Guerois, R. ; Samuel, D. ; Goud, B. ; Poüs, C. ; Bressanelli, S. ; Gassama-Diagne, A. *En révision*.

I

Introduction générale

” *Winter is coming...*

— Ned Stark
Game of Thrones

Chaque année, particulièrement en période hivernale, la gastroentérite fait parler d'elle dans les médias. Elle peut être causée par deux facteurs principaux : par des bactéries (5% à 15% des cas) comme des salmonelles, staphylocoques dorés, *Escherichia coli* ; ou bien des virus comme le rotavirus ou le norovirus. Cependant, les norovirus sont la cause majeure des gastroentérites aiguës d'origine virale, particulièrement depuis le développement de vaccins contre le rotavirus depuis 2006 [2, 3].

Les norovirus : Généralités

Les norovirus sont la première cause de gastroentérite virale sporadique chez l'Homme, particulièrement dans les épidémies saisonnières associées aux aliments contaminés comme les coquillages. Ils ne touchent pas seulement les humains, mais également les bovins, les souris, les chiens... Chaque année, ils sont responsables de la mort de près de 220 000 personnes dans les pays en voie de développement. Aucun traitement antiviral n'est commercialisé à ce jour. Ils comptent parmi les virus les plus simples, constitués d'une coque de protéines encapsulant le code génétique du virus. Malgré cette simplicité, le mécanisme d'assemblage du norovirus ne reste qu'hypothétique. Cette thèse a pour but de comprendre davantage les mécanismes moléculaires impliqués dans l'assemblage de la capsid du norovirus.

1.1 Historique

En 1929, Zahorsky décrit une maladie qu'il appelle "Winter vomiting disease" [4] (Maladie du vomissement hivernal). Il observa durant des périodes s'étendant de novembre à février des épisodes de diarrhées, vomissements et de nausées, souvent au sein de la même famille. Durant les hivers 1941 et 1942, une épidémie de même nature sévit à Charleston en Caroline du Sud (États-Unis) touchant des milliers de personnes, principalement des enfants [5].

En 1968, une épidémie de gastroentérite similaire se déclara chez 50% des élèves et enseignants d'une école primaire à Norwalk dans l'Ohio (États-Unis), ainsi que 32% des proches de ces élèves [7]. L'agent responsable de cette maladie n'avait cependant pas été identifié. Il fut appelé alors "l'agent de Norwalk". Des expériences sur des prisonniers volontaires de la maison de correction du Maryland (Jessup, États-Unis) en 1971 ont permis d'écarter l'hypothèse de l'origine bactérienne de l'épidémie [8] et en 1972 les premières images d'une petite particule furent isolées par microscopie immunoélectronique [6] (cf. Fig. 1.1). Les efforts pour cultiver le virus de Norwalk *in*

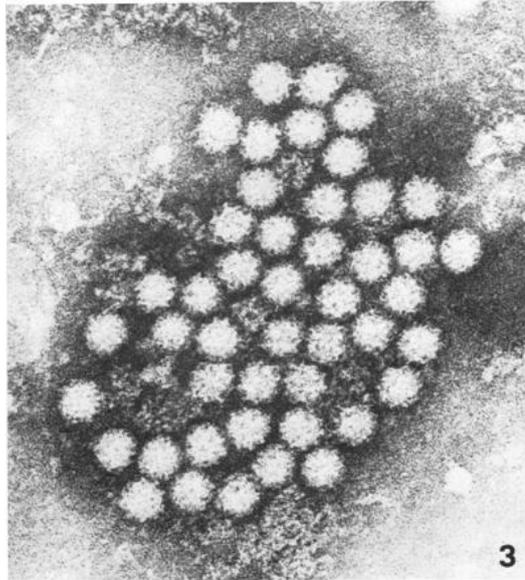


Figure 1.1: Première image du norovirus obtenue par microscopie électronique en 1972 par Kapikian *et al.* [6]

in vitro furent cependant infructueux et pendant longtemps, la littérature se concentra sur la description des symptômes provoqués par cet agent ainsi que les propriétés physiques de cette petite particule [9].

1.2 Taxonomie

Les observations faites sur cette particule montrèrent des ressemblances avec les membres de la famille des *Caliciviridae* : petit virus, non enveloppé, avec des dépressions circulaires à la surface de la capsidie nue, au génome ARN simple brin de polarité positive. Appelé jusque-là "virus de Norwalk", il fut alors renommé en "norovirus". Il a fallu attendre le séquençage complet du génome du norovirus en 1993 [11, 12] pour avoir une idée claire de la taxonomie du virus. Il rejoint alors la famille des *Caliciviridae* incluant les *Sapovirus*, *Lagovirus*, *Nebovirus*, *Vesivirus* et désormais, les *Norovirus* [13, 14]. Les norovirus, présentant une grande diversité génétique [15], sont subdivisés en 5 génogroupes majeurs, chacun pouvant contenir de nombreux génotypes. La classification se fait sur la base de l'identité de séquence entre protéines VP1, protéine majeure de la capsidie (cf. Fig. 1.2).

Les humains sont touchés par les génogroupes I et II (qui touchent également les porcs) tandis que le G.III touche les bovins [10]. Les norovirus de G.IV

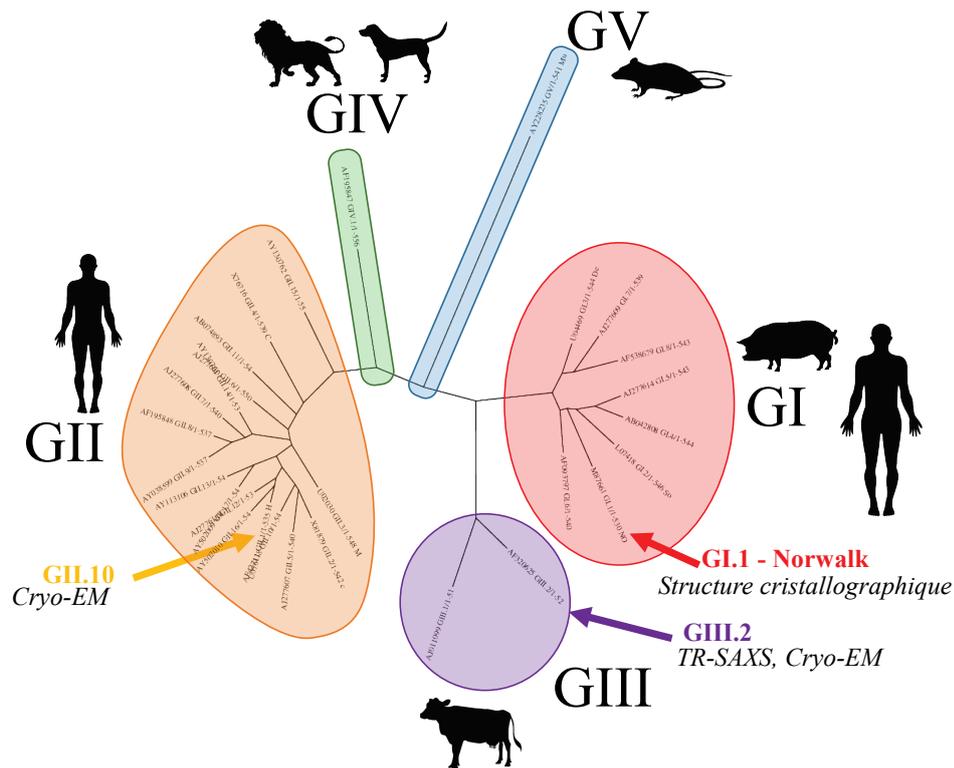


Figure 1.2: Dendrogramme des 29 séquences représentatives de VP1 [10]. Les capsides pour lesquelles des informations structurales sont disponibles (GI.1, GII.10 et GIII.2) sont indiquées par des flèches. Les séquences de chaque génogroupe sont indiquées dans la partie Matériel et méthode de l'Article page 89.

provoquent des gastroentérites aiguës chez le lion et le chien [16], et le G.V chez les souris [17].

1.3 Propagation du norovirus

Les norovirus se lient aux cellules épithéliales gastro-intestinales par des séquences polysaccharidiques spécifiques d'antigènes de groupe sanguin (HBGA)[18–21].

1.3.1 Résistance du norovirus

Une fois assemblé, le norovirus est extrêmement résistant : il reste stable sur des variations de température allant de -20°C à 60°C et des variations de pH de 3 à 7 [22]. Il reste également extrêmement stable en présence de chlore jusqu'à une concentration de 10 ppm [23] causant régulièrement

des problèmes sanitaires dans des piscines publiques où la concentration en chlore doit être inférieure à 1.5 ppm. Des études ont également montré que les norovirus pouvaient survivre une dizaine de jours sur un tapis contaminé [24, 25].

Il est donc nécessaire d'employer un désinfectant efficace. Selon le [CDC](#), un désinfectant avec une concentration de chlore de 1000 à 5000 ppm doit être utilisé. De récentes études ont également montré que l'acide citrique pouvait déstabiliser les capsides de norovirus [20]. L'[APE](#) a d'ailleurs validé un désinfectant contenant 5% d'acide citrique, moins toxique pour l'environnement que ceux utilisant du chlore. Il s'avère que l'acide citrique va venir se fixer au site de l'antigène de groupe sanguin, diminuant ainsi l'infectivité du norovirus [26].

1.3.2 Infectivité du norovirus

Le norovirus est transmis par voie feco-orale par l'eau ou la nourriture, ou bien de personne à personne par une contamination environnementale [27].

1.3.2.1 Contamination environnementale

Le virus est propagé dans l'environnement par les excréments de grandes quantités de virus ($>10^7$ particules/mL) dans les vomissures et les selles des personnes contaminées [29]. Une étude a montré, en reproduisant le vomissement d'un malade grâce à un système artificiel nommé "*Vomiting Larry*", que des particules virales pouvaient être projetées sur une distance supérieure à 3 mètres du malade, avec une diffusion latérale de 2.6 m [30]. Les auteurs préconisent donc une surface de désinfection de 8 m² quand une personne vient à vomir sur le sol.

Le norovirus est extrêmement infectieux : une particule virale de norovirus aurait une probabilité d'infection proche de 50% [31]. Une dizaine de particules suffirait donc pour déclencher une infection et développer une gastroentérite [31]. Cette haute résistance du norovirus sur les surfaces, combinée à sa haute infectivité fait donc de nombreux ravages.

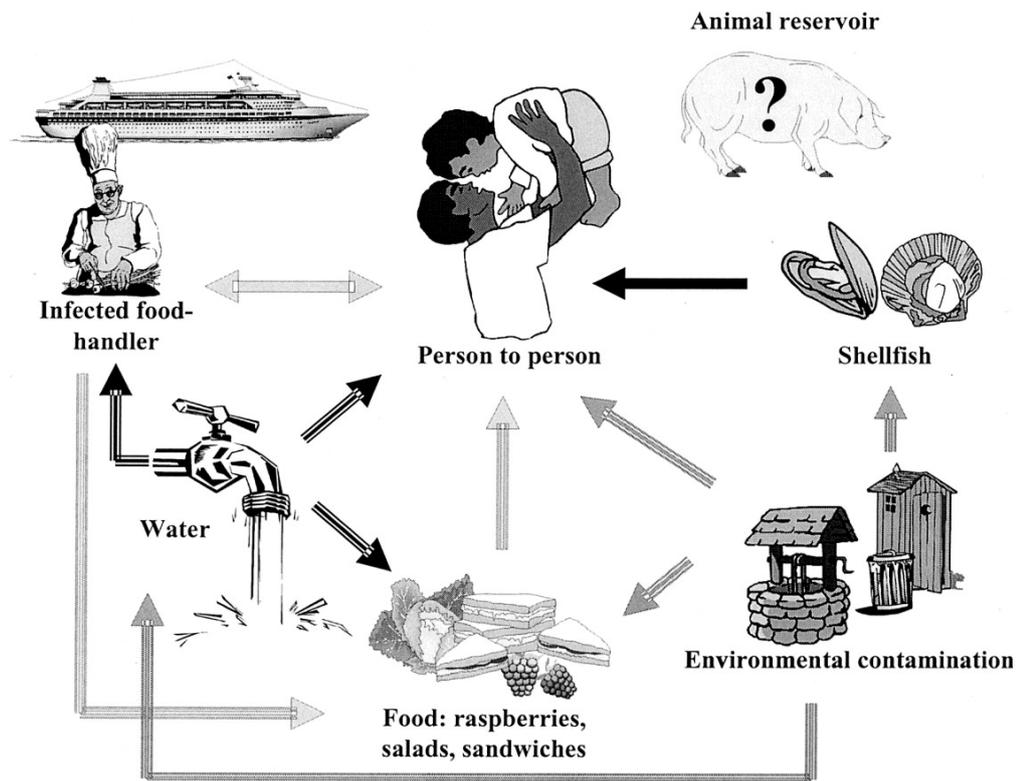


Figure 1.3: Voie de transmission du norovirus. Tiré de Moreno-espinosa *et al.* [28].

Ainsi, en 1999, 300 personnes, sur une période de 5 jours, ont été contaminées dans une salle de concert après qu’une personne malade eut vomi sur le sol [24]. Un mauvais désinfectant fut utilisé et le virus resta contagieux durant cette période. Le norovirus est également souvent responsable d’épidémies dans des contextes fermés, tels que des maisons de retraite [32], des hôpitaux ou bien des navires de croisière : il arrive également souvent qu’une personne infectée à bord finisse par contaminer un grand nombre de passagers au sein du bateau [33, 34].

1.3.2.2 Contamination par la nourriture

Les épidémies de norovirus dues aux aliments en Europe représentent 10 à 17% du nombre total d’épidémies [23]. Le norovirus se propage très facilement par l’eau. Des études sur des eaux de surface suisses ont montré sa présence dans la moitié des échantillons [23]. Il arrive également qu’une défaillance sur les réseaux d’évacuation des eaux usées conduise à une contamination de l’eau sanitaire.

Les particules virales excrétées par les individus infectés se retrouvent donc souvent rejetés dans les rivières, fleuves ou océans [35]. Les crustacés sont un formidable réservoir à norovirus [36] : ils filtrent l'eau de mer par la respiration ou l'alimentation et ils se développent dans les zones côtières (potentiellement infectées par les eaux usées). En filtrant l'eau, les particules virales vont s'y accumuler et si ces crustacés sont consommés crus, ils provoquent une contamination de l'individu. On parle alors d'une infection primaire par les denrées alimentaires [37]. Ainsi, il y a fréquemment des alertes sanitaires concernant la présence de norovirus dans les huîtres conduisant à la fermeture de certaines zones d'élevages. En 2003, une étude montra que près de 10% des huîtres importées en Suisse étaient contaminées par le norovirus [38].

La contamination d'un individu par les aliments ne s'effectue pas seulement par les crustacés, mais également par la nourriture issue de chaînes alimentaires. Un employé en tête de chaîne peut très facilement contaminer tout un lot conduisant à l'infection de centaines de personnes en bout de chaîne. C'est dans ce cas une contamination directe des denrées alimentaires [37]. Par exemple, en avril 2017 en France, des grands distributeurs furent contraints de rappeler des lots de framboises surgelées à la suite de détection de particules de norovirus dans certains lots.

Enfin, des fruits ou légumes peuvent également être contaminés par le lavage ou l'arrosage avec de l'eau contaminée. Il s'agit dans ce cas d'une infection secondaire [37]

1.4 Impacts socio-économiques

1.4.1 Impact social

Tous les ans, 700 millions de personnes environ sont touchées par la gastroentérite, qui est dans 220 000 cas mortelle [39]. Tous les continents sont concernés, l'Europe étant relativement moins touchée (Fig. 1.4). Si la répartition des infections du norovirus par région est similaire dans la plupart des continents (Fig. 1.4A), le nombre de décès est bien plus important dans les

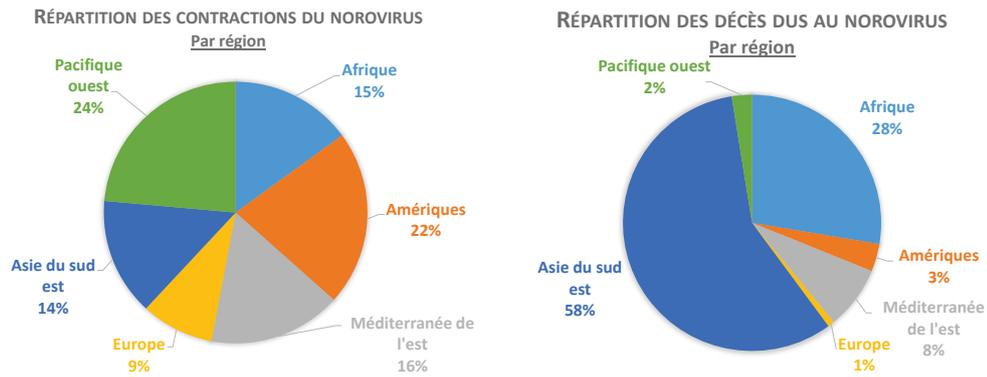


Figure 1.4: Répartition par région des cas d'infections et de décès dus au norovirus. Basée sur des données de Bartsch *et al.* [39].

pays en voie de développement (Fig. 1.4B). L'Asie du Sud-Est et l'Afrique sont bien plus touchées, notamment par un manque de prise en charge et d'infrastructures sanitaires suffisantes.

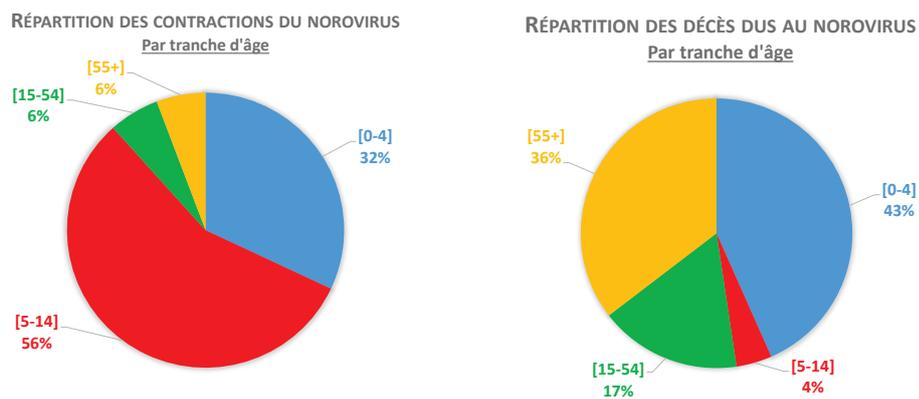


Figure 1.5: Répartition par tranche d'âge des cas d'infection et de décès dus au norovirus.

Les enfants sont plus sensibles au norovirus. Les individus de 0-4 ans et 5-14 ans représentent 88% des individus infectés par le norovirus^{1.5}. Cependant les infections chez les personnes "âgées" (55+ ans) et les enfants en bas âge (0 à 4 ans) sont plus mortelles (Fig. 1.5B).

1.4.2 Impact économique

Pendant longtemps, l'impact sur l'économie mondiale des conséquences d'une infection par le norovirus ne fut pas mesurée. En 2016, une étude a tenté d'approximer un coût annuel au norovirus en tenant compte des frais de

traitement, d'hospitalisation, mais également du manque de productivité associé à la gastroentérite entraînée par une infection due au norovirus [39]. Bartsch *et al.* [39] estime donc le coût annuel associé au norovirus à 64.5 milliards de dollars environ. 4.2 milliards concerneraient les systèmes de santé et 60.3 milliards de dollars à la perte de productivité. Les enfants de moins de 5 ans entraîneraient à eux seuls un coût estimé à 39.8 milliards de dollars chaque année.

Cette étude avait été faite dans le but d'apporter plus d'informations aux financeurs et décideurs afin d'améliorer la recherche sur le norovirus. En effet, selon les auteurs, l'attention accordée au norovirus n'est pas du tout suffisante par rapport à l'impact de ce virus dans le monde.

1.5 Piste thérapeutique

Depuis 2006, après la découverte de deux vaccins contre le rotavirus, le développement d'un vaccin contre le norovirus est devenu une priorité afin de prévenir les gastroentérites virales [3]. Cependant, les systèmes de cultures cellulaires actuels ne permettent pas de cultiver le norovirus, ce qui pose un réel problème au développement d'un vaccin [3, 40]. Plusieurs pistes sont donc explorées afin de se protéger du norovirus.

1.5.1 Utilisation de particules virales vides

En production dans un système baculovirus, VP1 s'auto-assemble en particule virale vide (VLP). Ces pseudo-particules virales sont antigéniquement et structurellement indistinguables des particules infectieuses collectées dans les selles de personnes contaminées par le virus de Norwalk [41]. Les sites de liaison aux récepteurs HBGA sont donc conservés. Il a également été montré que les VLPs pouvaient induire des réponses immunitaires sans développer de symptôme [42]. Ces virions sont donc une piste très sérieuse pour se défendre contre le norovirus.

En 1998, Ball *et al.* ont montré que des souris développaient une réponse immunitaire à la suite d'ingestion orale de VLPs de norovirus [43]. En 2001, Guerrero *et al.* ont démontré que la réponse immunitaire était plus élevée lors

d'une inoculation nasale [44]. La réponse immunitaire pouvait être encore plus élevée en utilisant certains adjuvants [44] ou des vecteurs viraux [45].

Des essais cliniques en phase I et II ont été réalisés et ont permis de prouver l'efficacité de VLPs sur les humains [46]. Le nombre de diarrhée et de vomissement chez les patients inoculés par des VLPs puis infectés par le norovirus a diminué, ainsi que la sévérité des symptômes de la gastroentérite [3].

1.5.2 Inhibition du norovirus

Suite à la découverte de l'efficacité de l'acide citrique sur la stabilité de la capsidie [20, 26], un traitement utilisant ce composé pourrait également être envisagé. L'acide citrique va en effet se fixer sur les sites de liaisons aux récepteurs HBGA (cf. Fig. 1.6) et ainsi diminuer l'infectivité du norovirus par inhibition compétitive.

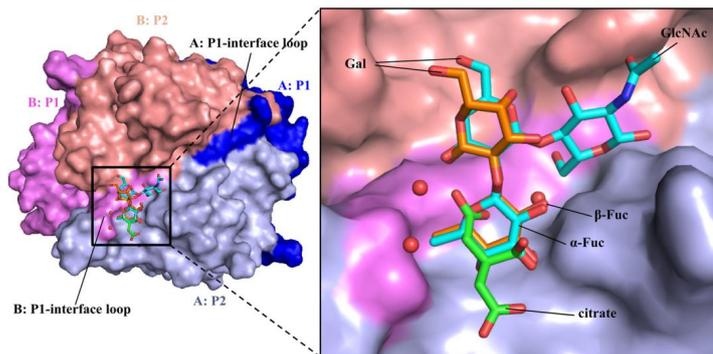


Figure 1.6: Liaison du citrate sur le domaine P de la protéine de capsidie virale (VP1). Le citrate est représenté avec un squelette vert, et l'HBGA en cyan. Tiré de Hansman *et al.* (2012)[20]

1.6 Organisation du Norovirus

1.6.1 Organisation génomique

Le norovirus est un virus dont le génome est un ARN positif simple brin de polarité positive[11]. C'est à dire que cet ARN se comporte comme un ARN messenger. Il est directement reconnu par les machineries de traduction de la cellule hôte.



Figure 1.7: Représentation schématique de l'ARN génomique du norovirus. La succession des protéines codées dans les 3 ORF est indiquée. Les cadres ne sont pas à l'échelle.

L'ARN d'environ 7500 bases est lié de manière covalente à une protéine virale génomique (VPg, ou NS5) en position 5', et polyadénylé en position 3' [47]. Trois cadres de lectures ouverts (**ORF***¹) sont présents. Le premier, l'ORF1, code 6 protéines non structurales (NSx). NS7 est une polymérase et participe à la synthèse de nouveaux génomes et de l'ARN sous-génomique, à partir duquel ORF2 et ORF3 sont traduits [48]. NS6 est une protéinase et participe à la **maturation protéolytique*** de la polyprotéine codée par l' ORF1 [49]. NS5, aussi appelé VPg, est liée de façon covalente à l'ARN génomique. Elle remplace la coiffe classique des ARN messagers et agit comme agent recruteur des facteurs de traduction de la cellule hôte [50]. Enfin, les fonctions de NS1-2, NS3 et NS4 sont encore peu claires mais liées aux réarrangements membranaires intracellulaires induits par la réplication des virus à ARN positif. Nos collaborateurs étudient ces questions et pour les aider, j'ai généré par modélisation par homologie, threading et *ab initio* des modèles pour préciser la topologie membranaire dans cette région (Fig. 1.8, Doerflinger *et al.*, en révision).

L'ORF2 et l'ORF3 codent les protéines structurales : respectivement VP1 et VP2. VP1 est la protéine majeure de la capside et c'est sur elle que ce travail de thèse est focalisé. VP2 est une seconde protéine structurale (protéine mineure de capside, présente dans le virion à quelques copies), de taille variable selon le génotype et dont le rôle n'est pas encore complètement défini à ce jour [51]. Cependant, la traduction de l'ORF3 est compliquée [51], une étude montre que VP2 est traduite à partir d'un ARN **bicistronique*** et que la région initiatrice de l'ORF3 se trouve en 3' de l'ORF2 [52].

1. Les mots et expressions suivis d'un astérisque sont définis dans le glossaire en début de document

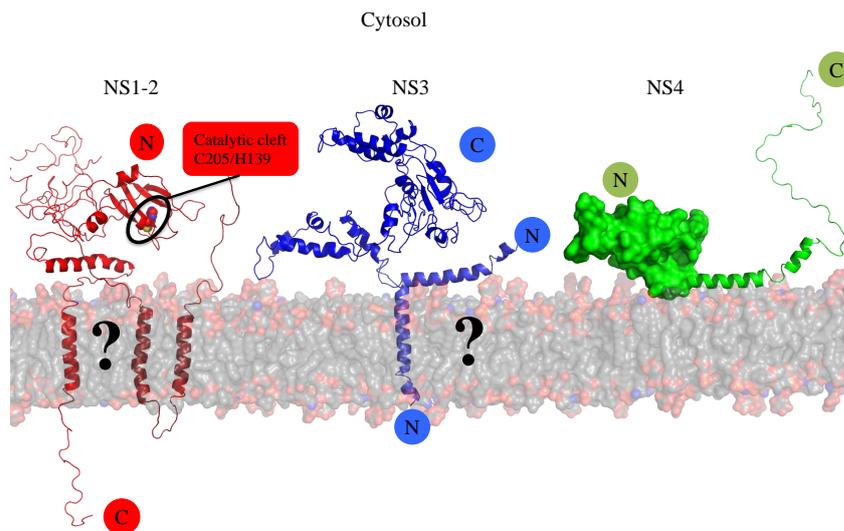


Figure 1.8: Modélisation des NS1-2, NS3 et NS4 humaine et association putative à la membrane. Figure réalisée pour Doerflinger *et al.*, en révision. Les deux principaux apports de cette modélisation ont été la définition des topologies membranaires possibles pour NS1-4 et l'identification d'une fonction enzymatique dans NS1-2.

1.6.2 Organisation des capsides icosaédriques

Avant de parler de l'organisation structurale de la capsidie du norovirus, il est important de préciser deux éléments sur l'organisation des capsides icosaédriques : ce qu'est la symétrie icosaédrique (les éléments de symétrie qui la composent) et la classification des capsides icosaédriques selon le nombre de triangulation T.

Un icosaèdre est composé de trois types d'axes de symétrie : axe 2, un axe 3 et un axe 5 (Fig. 1.10a). Une structure icosaédrique contient toujours le même nombre d'axes : six axes 5, dix axes 3 et quinze axes 2 [53]. Ces axes relient 60 **unités asymétriques*** identiques. On peut donc faire un icosaèdre avec 60 copies identiques d'une même protéine (ou d'un même complexe protéique). Mais au-delà de 60, ce n'est plus possible. Cependant, pour certains multiples de 60, les copies peuvent se positionner exactement sur un icosaèdre si elles sont dans des conformations légèrement différentes. Si le nombre de copies est $60T$, il y a T conformations. Les valeurs du nombre de triangulation T pour

lesquelles les conformations peuvent être presque identiques sont données par :

$$T = h^2 + h * k + k^2 \quad (1.1)$$

où h et k sont deux nombres entiers. C'est le principe de quasi-équivalence proposé en 1962 par Caspar et Klug [54]. h et k représentent les nombres d'hexamères entre deux pentamères via le plus court chemin. En pratique, en partant d'un pentamère, h représente le nombre d'hexamères en ligne droite (par exemple Fig. 1.9b) et k le nombre d'hexamères après avoir changé de direction (par exemple Fig. 1.9a et Fig. 1.9c).

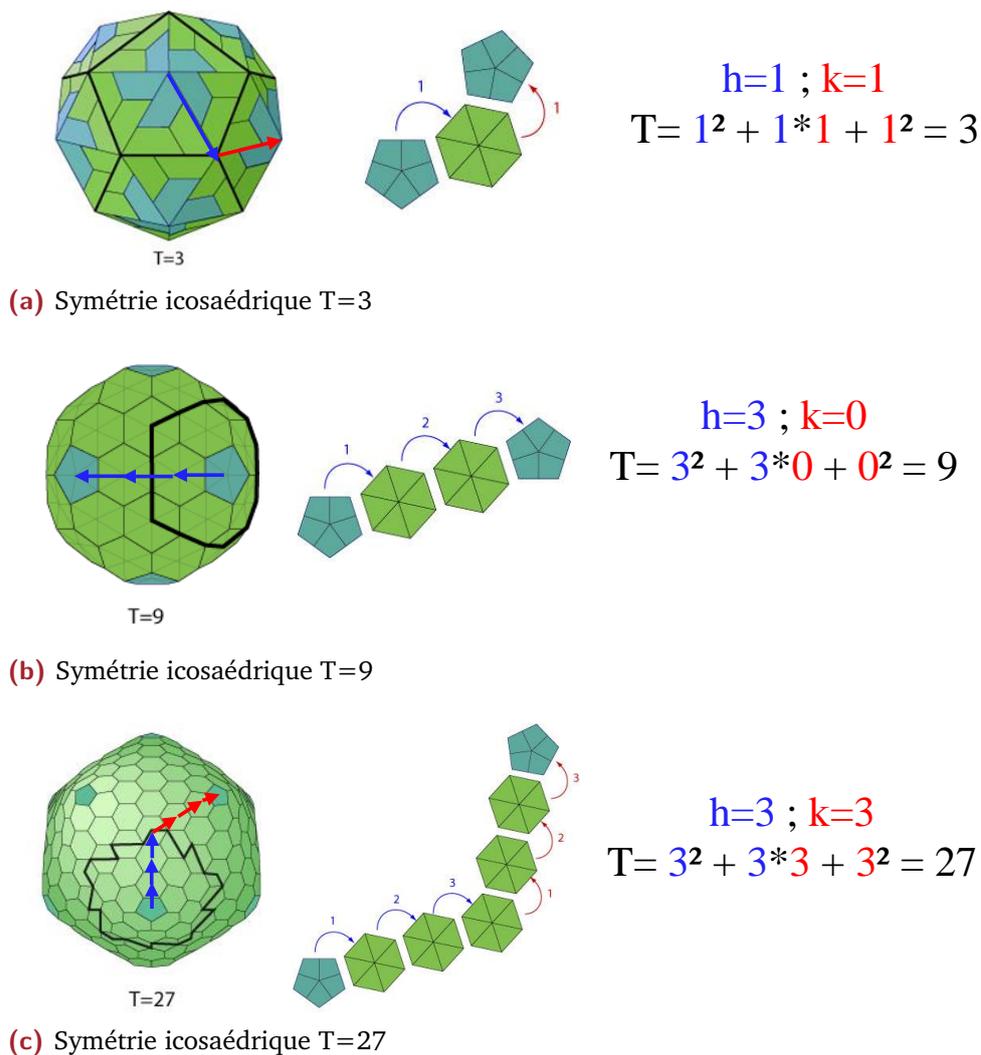
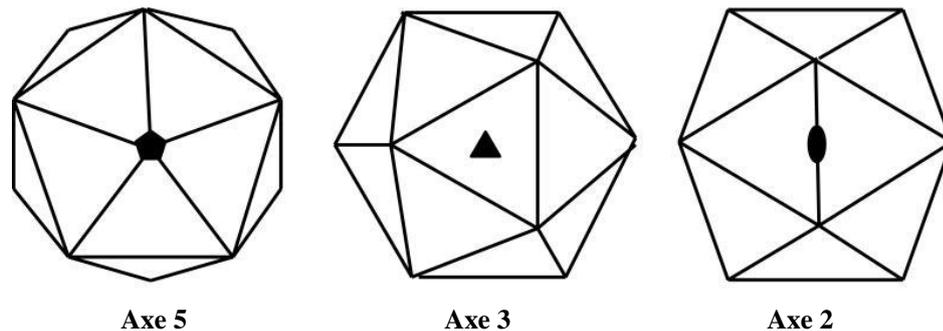
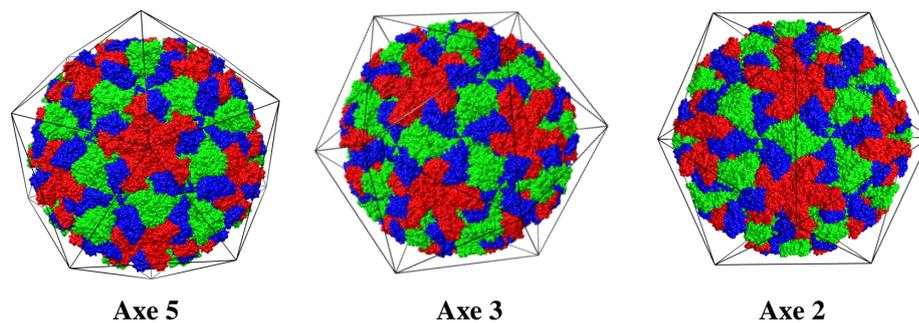


Figure 1.9: Exemple de calcul des paramètres h et k ainsi que du nombre T avec un exemple de symétrie icosaédrique de type T=3(1.9a), T=9 (1.9b) et T=27 (1.9c). Les flèches bleues montrent le parcours pour h , les flèches rouges pour k . Adapté du site ViralZone d'ExpASY (<http://viralzone.expasy.org/806>).

Ainsi pour un virus icosaédrique de type $T=3$ ($h=1, k=1$), comme le norovirus, la capside est composée de 180 copies d'une même protéine dans 3 conformations légèrement différentes appelées A, autour des axes 5, et B/C, autour des axes 3. L'axe 3 autour duquel B et C alternent est presque un axe 6 exact et représente une variation de l'axe 5 avec une copie supplémentaire (axe quasi-6). La forme refermée de la capside est obtenue grâce à l'alternance des hexamères quasi-6 qui sont plats et des pentamères qui sont plus courbés.



(a) Représentation schématique des axes de symétrie



(b) Exemples sur la capside du norovirus

Figure 1.10: Description des axes de symétrie dans une capside icosaédrique avec une représentation schématique (1.10a) tiré de Prasad & Schmid (2012)[55] et exemple sur le norovirus (1.10b)

1.6.3 Organisation structurale de la capside

1.6.3.1 VP1

VP1 est la protéine majeure de capside du norovirus. Elle est composée d'environ 530 acides aminés pour une taille de 58 à 60 kDa. Sa structure

dans le contexte de la capside a été déterminée par cristallographie aux rayons X en 1999 par Prasad *et al.* [56].

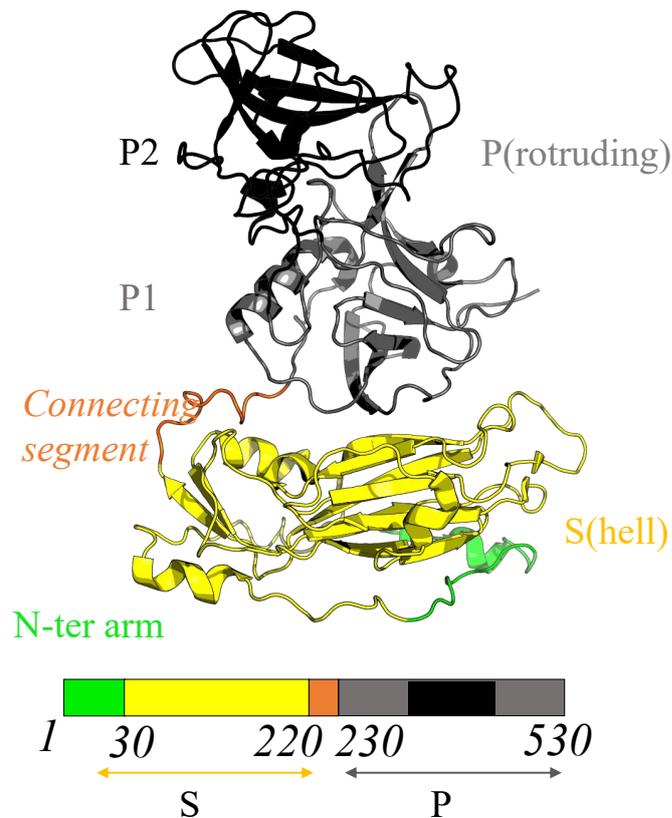


Figure 1.11: Structure de la protéine majeure de capside VP1. La partie N-terminale est représentée en vert, le domaine S en jaune et le domaine P en gris/noir. Le domaine P est composé de deux sous domaines : P1 en vert, et P2 en noir. Les domaines P et S sont séparés par un segment d'environ 10 acides aminés (orange).

La région N-terminale s'étend du premier résidu au résidu 30², la région C-terminale du résidu 521 au résidu 530. Ces deux régions ne sont pas structurées et absentes de la structure cristallographique, sauf le bras N-terminal de la conformation B, qui est ordonné depuis le résidu 10 et constitue le premier point de différence dans le cadre de la quasi-équivalence.

Le reste de la protéine est décomposé en plusieurs domaines (cf. Fig. 1.11) :

- le domaine S (résidus 46 à 220) composant la coque sphérique interne de la capside ("Shell") qui est la région la plus conservée [57].

2. La numérotation correspond à l'indexation des résidus pour le génogroupe I.1 (Virus de Norwalk)

- le domaine P (résidus 230 à 530) ("*Protruding*") formant les spicules et décorant la capsid. Ce domaine, est composé lui même de deux sous domaines :
 - Le sous-domaine P1 (résidus 230→278 + 406→520) qui est plutôt conservé
 - Le sous-domaine P2 (résidus 279 à 405) qui est le plus externe. Cette région est très variable [57, 58]. De nombreuses études ont montré qu'un site de liaison, présent en P2, était spécifique à certains HCGA via des séquences polysaccharidiques [18–21] (cf. Fig. 1.12).

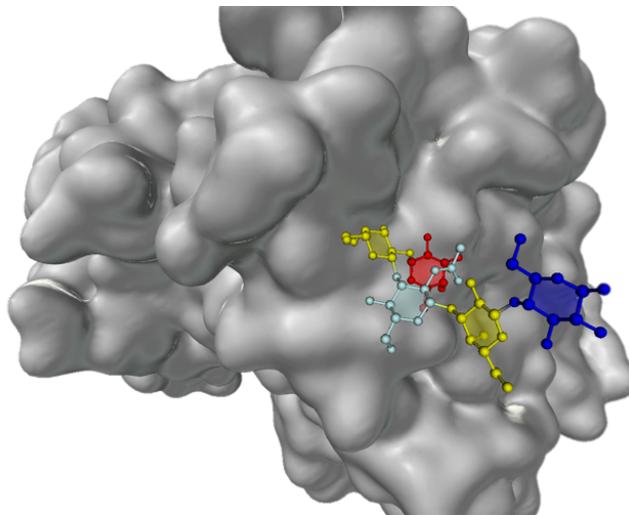


Figure 1.12: Antigène de groupe sanguin (pentasaccharide : aFuc12bGal13GlcNac13bGal14Glc) lié au site de liaisons HBGA du virus de Norwalk (PDBid : 2ZL6 [59]). Image rendue avec SweetUnityMol [60, 61]

Un segment d'environ 10 acides aminés connecte le domaine S au domaine P (220→230) et permet une flexibilité entre les deux domaines. Ainsi, trois conformations de VP1 existent où l'angle formé entre le domaine S et le domaine P est différent (cf. Fig. 1.13). C'est le deuxième point de différence dans le cadre de la quasi-équivalence, avec des angles de 87.4°, 88.1° et 94.8° respectivement pour les conformations A, B et C. Les différences principales sont donc cet angle beaucoup plus important dans la conformation C et la structuration du bras N-terminal dans la conformation B. Les trois conformations composent l'unité asymétrique permettant de reconstruire la capsid grâce aux 60 matrices de symétrie icosaédrique.

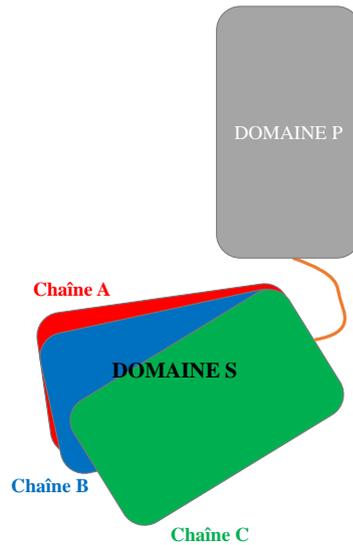


Figure 1.13: Représentation de l'orientation du domaine S par rapport au domaine P pour les trois conformations.

1.6.3.2 Capside

Le diamètre de la capside du norovirus est d'environ 39 nm [62]. Elle est constituée de 180 copies de VP1 arrangées avec une symétrie icosaédrique de type T=3 [54]. La surface de contact importante entre deux domaines P indique que la brique d'assemblage est un dimère de VP1 [56].

De ce fait, la description de cette capside en pentamères et hexamères n'est pas adaptée : les dimères A/B participent en effet à la fois aux pentamères de dimères (5 dimères A/B) et aux hexamères de dimères (6 dimères alternativement A/B et C/C) (cf. Fig. 1.14).

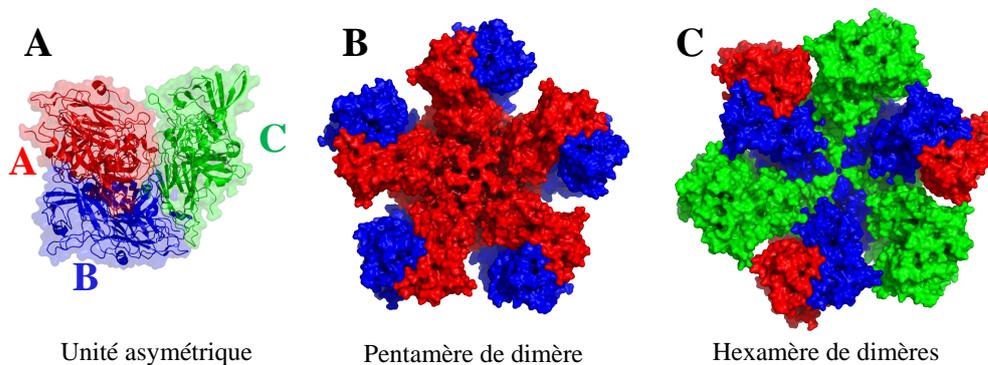


Figure 1.14: Unité asymétrique (A) et assemblages de dimères de la capside du norovirus : pentamère de dimères (B) et hexamère de dimères (C). Les chaînes A sont en rouge, les chaînes B en bleu et les C en vert.

Comme précisé dans la partie [VP1](#), le domaine S est responsable de la forme icosaédrique et sphérique de la capside tandis que le domaine P constitue des "pointes" à la surface de la coque interne [3, 63] (cf. Fig. [1.15](#)). Le domaine P2 est le domaine le plus externe, il contient les sites d'interaction avec les récepteurs HBGA [20, 21, 57, 59] et est en contact avec la réponse immunitaire de l'hôte. Il est intéressant de noter que le domaine S seul, peut former des pseudo particules icosaédriques [64], et que le domaine P seul peut former des particules octaédriques[65].

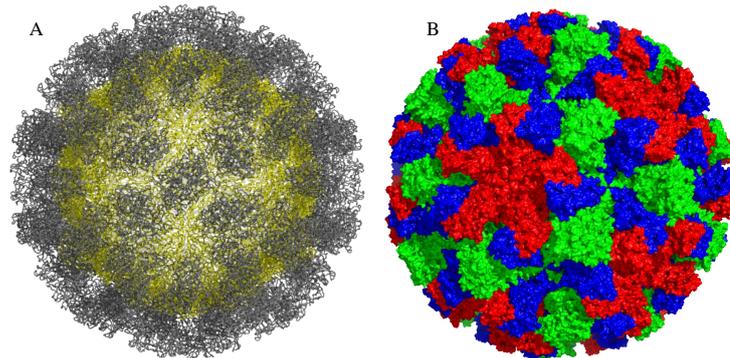


Figure 1.15: Capside du norovirus. (A) Représentation en ruban avec une coloration par domaine. Le domaine S est en jaune, le domaine P en gris. (B) Représentation en surface avec une coloration par chaîne (A→rouge, B→bleu, C→vert).

1.6.3.3 VP2

Une seconde protéine structurale est présente : VP2. Elle est plus petite (20 à 29 kDa) et présente une grande variabilité de taille selon les génogroupes [66]. Seules une ou deux copies seraient présentes dans la particule virale [67, 68]. Elle serait présente dans tous les calicivirus [69] et notamment chez le chat (FCV). VP2 stabilise la capside et est requise pour la production de particules virales infectieuses [68, 70]. En effet, son absence entraîne la formation d'un virion sans matériel génétique [71, 72].

VP2 est, contrairement à VP1, très basique [51]. Elle se fixe également à une région très conservée de VP1 [70], accessible à l'intérieur de la capside (entre deux domaines S) (cf. Fig. [1.16](#)).

Il a été émis l'hypothèse qu'elle ferait le lien entre l'ARN et la capside via des interactions protéine/protéine avec la protéine VPg [70], permettant

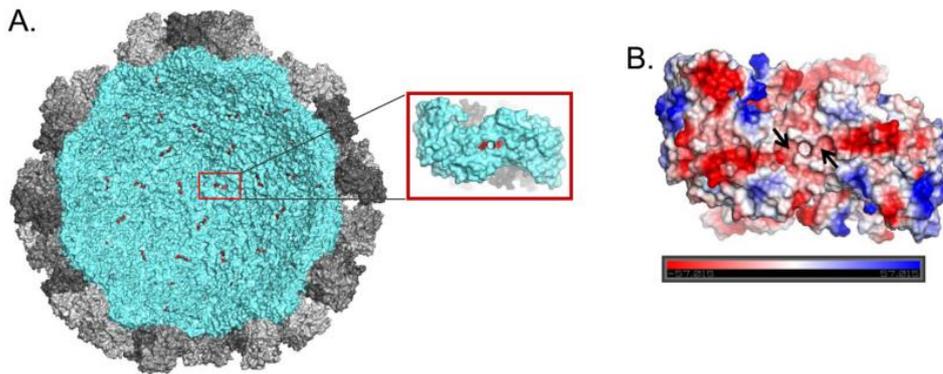


Figure 1.16: Région de VP1 en interaction avec VP2. (A) Représentation dans la capside. (B) Vue d'en dessous en représentation surface électrostatique. La région identifiée comme zone d'interaction avec VP2 se trouve en dessous de VP1, à l'intérieur de la capside et à l'interface dimérique. Tirée de Vongpunsawas *et al.* (2013) [70]

d'encapsuler l'ARN lors de la formation de la capside. Cependant, cette hypothèse n'a pas été vérifiée.

VP2 ne suivant pas la symétrie icosaédrique, il n'y a pas d'information structurale disponible [70].

2.1 Théorie d'assemblage des petits virus à ARN positif simple brin

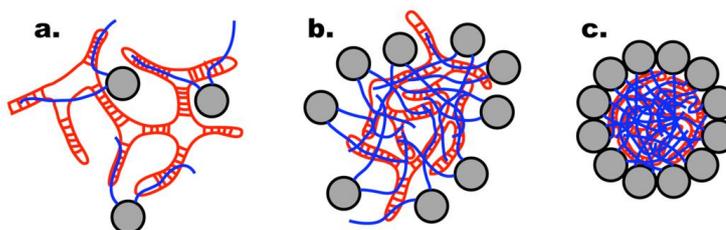


Figure 1.17: Hypothèse d'assemblage de Devkota *et al.* pour les petits virus à génome simple brin. Tiré de Devkota *et al.* (2009)[73]

En 2009, Devkota *et al.* ont proposé un mécanisme simple pour l'assemblage des nodavirus, un petit virus non enveloppé à ARN positif simple brin de symétrie de type T=3. Les bras N-terminaux des protéines de la capside étant chargés positivement, ils vont interagir avec l'ARN qui lui est chargé négativement [73] (Fig. 1.17a). Quand suffisamment d'ARN a été neutralisé (Fig. 1.17b), le complexe protéine/ARN se condense pour former la capside tout en encapsulant le génome [73](Fig. 1.17c).

En 2006, Belyi *et al.* ont étudié les structures de 19 virus (à ADN et ARN) ainsi que leurs génomes dont la taille variait entre 1 kb et 12 kb. Ils ont pu mettre en évidence que la charge nette du bras polypeptidique était proportionnelle à la taille du génome avec un facteur de 1.61 ± 0.03 [74] (cf. Fig. 1.18).

Devkota *et al.* ont donc étendu leur modèle d'assemblage du nodavirus à un modèle d'assemblage plus général pour les virus à génome simple brin [73]. Cependant, selon Belyi *et al.*, nous devrions trouver une charge nette d'environ 4658 sur les bras N-terminaux de VP1 (cf. Fig. 1.18). Or le bras N-terminal du norovirus (résidues 1→30) est négatif et à pH 7, cette charge nette rapportée à la capside entière vaut -360. Il apparaît donc que le modèle d'assemblage proposé par Devkota *et al.* n'est donc pas adapté au norovirus.

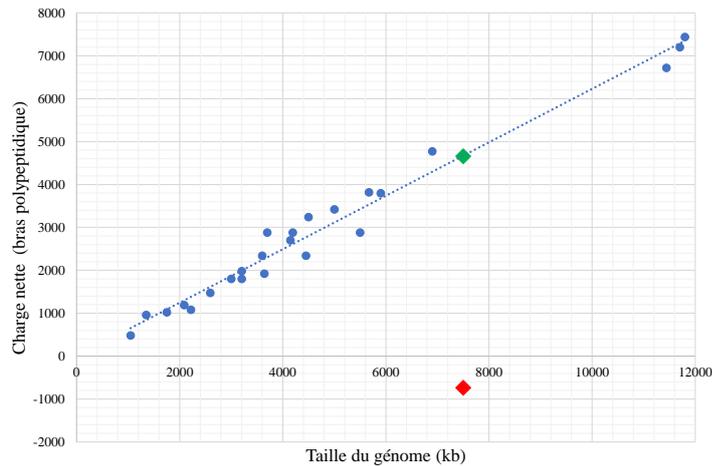


Figure 1.18: Charge nette des bras polypeptidiques en fonction de la taille du génome du virus. La charge nette correspond à la somme des charges sur tous les bras polypeptidiques en contact avec le génome. Il existe un rapport de 1.61 ± 0.03 entre la charge nette et la taille du génome. Les losanges représentent la charge nette des bras N-terminaux du norovirus théorique (orange) et réel à pH 7 (rouge). Adapté de Belyi *et al.* (2006) [74].

2.2 Modèles d'assemblage du norovirus

Lors de la publication de la structure cristallographique du Norovirus en 1999 [56], Prasad *et al.* ont remarqué une grande surface de contact (1800 \AA^2) entre paires de domaines S des chaînes A (autour de l'axe 5).

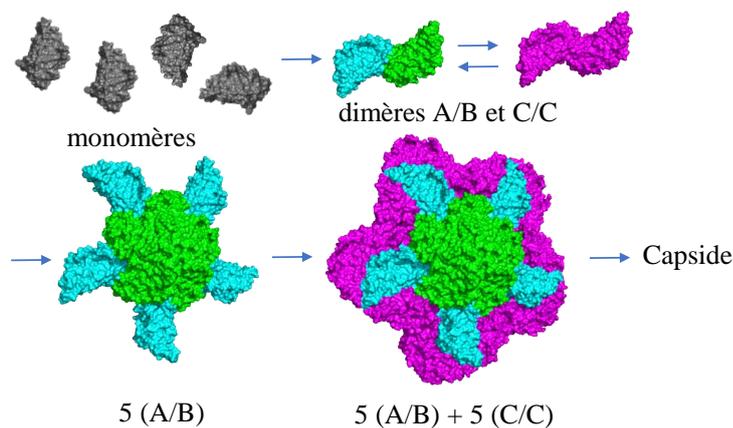
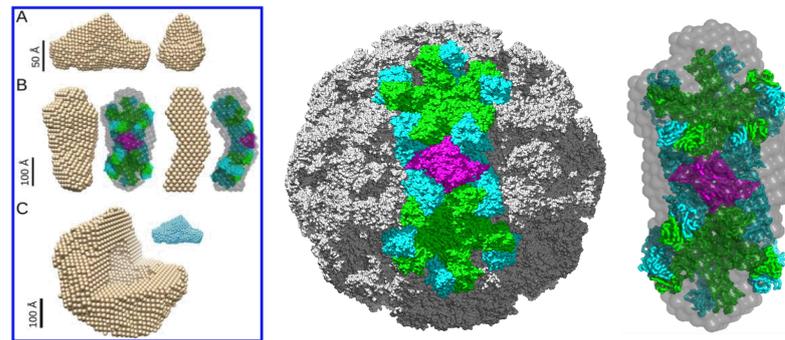


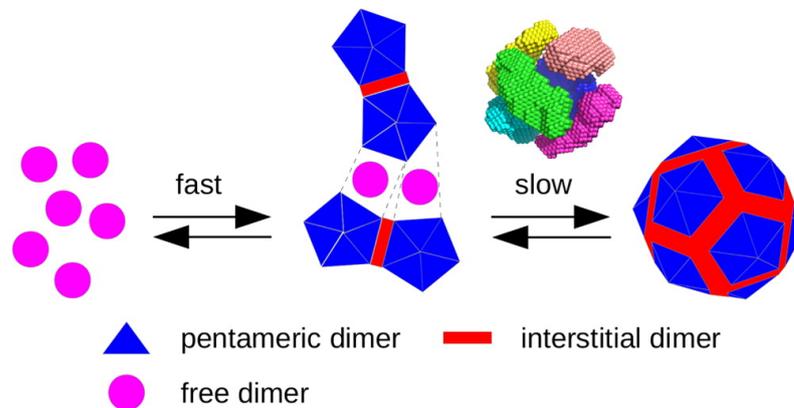
Figure 1.19: Modèle d'assemblage du norovirus de Prasad *et al.*. Adapté de Prasad *et al.* (1999) [56] par Jean-Charles Carvaillo. Dans cette figure, seuls les domaines S sont représentés.

Ils ont donc suggéré que le pentamère de dimères (POD) serait le premier intermédiaire dans le processus d'assemblage à partir de dimères [55, 56]

(Fig. 1.19). La suite de l'assemblage serait une croissance isotropique à partir du POD par l'insertion de dimères C/C [56].



(a) Forme de l'intermédiaire d'assemblage du norovirus



(b) Modèle d'assemblage de Tresset *et al.* (2013)

Figure 1.20: (1.20a) Forme de l'intermédiaire d'assemblage du norovirus. Adapté de Tresset *et al.* (2013)[75]. Cet intermédiaire semble correspondre à deux POD cimenté par un dimère interstitiel. (1.20b) Modèle d'assemblage de Tresset *et al.* en tenant compte de la forme de l'intermédiaire. Tirée de Tresset *et al.* [75].

En 2013, Tresset *et al.* ont montré par TR-SAXS que pour le norovirus bovin (GIII.2), l'assemblage passe par un intermédiaire allongé composé de 10 à 11 dimères [75]. L'enveloppe obtenue à partir des données SAXS est compatible avec une forme intermédiaire composée de deux POD reliés par un dimère interstitiel de type C/C (Fig. 1.20a). À partir de ce résultat, ils ont pu créer un modèle d'assemblage formé de 6 de ces intermédiaires et ponté par des dimères interstitiels (Fig. 1.20b).

Le but de cette thèse est de comprendre les mécanismes moléculaires permettant de passer du dimère à l'intermédiaire observé en TR-SAXS [75].

La simulation de dynamique moléculaire MD permet, grâce à des moyens informatiques, de simuler l'évolution de particules dans le temps. Cette méthode est très utilisée dans le domaine de la biologie/biochimie structurale pour l'étude conformationnelle de composés chimiques, de protéines, mais également dans l'interaction ligand/protéines, affinement moléculaire...[76]. La MD permet des observations à l'échelle atomique, dans un laps de temps de l'ordre de la nanoseconde (ns) (cf. Fig. 1.21). De nombreuses méthodes de simulation existent, permettant d'augmenter le temps de simulation ainsi que la taille du système. Finalement, les seules limites de ces méthodes sont notre imagination et les moyens de calcul à notre disposition.

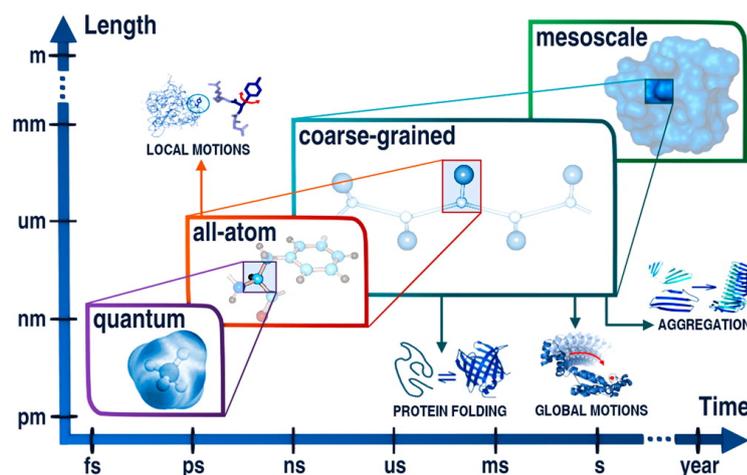


Figure 1.21: Les différentes échelles de simulation (taille/temps). Tirée de Kmiecik *et al.* (2016)[77].

3.1 Historique

En 1950, Metropolis *et al.* ont voulu étudier les propriétés d'un liquide. Pour ce faire, ils ont représenté le liquide par un modèle de disques à 2 dimensions [78]. Ce fut les premiers travaux de simulation moléculaire et cette méthode probabiliste fut appelée Monte-Carlo Metropolis. En 1957, Alder *et al.* ont développé une méthode déterministe afin de simuler des collisions entre des

sphères [79, 80]. C'est cette méthode qui a permis de poser les bases de la MD généraliste[81].

Il a ensuite fallu attendre la fin des années 70 pour voir arriver les premières simulations de MD avec les travaux de Martin Karplus [82]. La protéine étudiée fut la **BPTI** composée de 885 atomes avec un temps de simulation de 2 ps. Ce fut également durant cette période que les méthodes de simulations de mécanique moléculaire et quantique (QM/MM) furent mises au point par Michael Levitt et Arieh Warshel pour étudier l'hydrolyse d'un modèle de liaison peptidique par le lysozyme [83]. Ces travaux furent la base de la MD dans le domaine de la biologie structurale. Martin Karplus, Michael Levitt et Arieh Warshel furent récompensés par un prix Nobel de Chimie en 2013 pour "le développement de modèles multi-échelles pour des systèmes moléculaires complexes".

Depuis, les algorithmes ont évolué et la puissance de calcul n'a cessé d'augmenter tous les ans, permettant ainsi de simuler des systèmes bien plus importants. Ainsi, en 2013, Zhao *et al.* ont simulé la capsid complète du VIH pendant 100 ns (64 millions d'atomes)[84]. En 2015, Reddy *et al.* ont réalisé plusieurs simulations d'environ 5 μ s sur des systèmes gros grains de 31 millions d'atomes sur la capsid de virus de la grippe A [85]. Enfin, nous pouvons également mentionner une récente simulation de l'équivalent d'1/10 du volume du cytoplasme d'une bactérie (*Mycoplasma genitalium*) composé de 100 millions d'atomes pour une durée de 20 ns [86] qui, à notre connaissance, est l'une des plus grosses simulations réalisées à ce jour.

3.2 Quelques simulations de capsides virales

Bien que les capsides de virus soient de grosses structures, leur simulation n'est pas impossible. Elle demande cependant beaucoup de ressources et donc un accès à un super ordinateur. Quelques exemples de simulations de capsides entières sont présentés dans la table 3.1

Table 3.1: Exemple de simulations réalisées sur des capsides de virus en tout atome ou en gros grains.

Article	Système	N atomes	Type	Temps
Freddolino <i>et al.</i> (2006) [87]	Virus de la mosaïque du tabac	1 million	tout atome	5 x10 ns
Wang <i>et al.</i> (2013)[88]	Virus de la maladie hémorragique du lapin	~9.9 millions	tout atome	Optimisation MDFF
Zhao <i>et al.</i> (2013) [84]	Capside du VIH	64 millions	tout atome	100 ns
Goh <i>et al.</i> (2015) [89]	Multiples portions de capside de rétrovirus	4 x ~1.8 million	tout atome	100 ns
Reddy <i>et al.</i> (2015) [85]	Virus de la dengue	~912000	gros grains	5 μ s
Andoh <i>et al.</i> (2014) [90]	Capside du poliovirus	~6.4 millions	tout atome	25 μ s

La table 3.1 n'est, bien entendu, pas exhaustive. Cependant, il existe peu de simulations de capside de virus complète. Nous pouvons retrouver des simulations d'assemblage de capside, composée de quelques unités asymétriques ou capsomères [89], des simulations de capside entière ayant pour but d'améliorer un modèle cristallographique ou de Cryo-EM [88] (nous en avons nous-mêmes réalisé, voir [Modélisation de la capside de norovirus de GIII.2](#) page 121).

Nous pouvons retrouver quelques simulations de capsides en tout atome [84, 87, 88, 90] et notamment celle de Zhao *et al.* sur la capside du VIH comprenant plus de 64 millions d'atomes sur un temps de 100 ns. Une telle simulation a demandé des ressources considérables : 128 000 coeurs sur le super ordinateur de l'université de l'Illinois (Blue waters). Afin de pouvoir simuler sur des temps plus longs, certains décident de perdre en précision et passer à un système gros grains, ce qui permet de réduire le nombre d'atomes et donc le temps de calcul [85].

Enfin, nous pouvons également citer une étude par Arkhiupov *et al.* [91] qui ont utilisé un modèle dit "très gros grains" sur plusieurs capsides de virus

où chaque grain représente environ 200 atomes. Ainsi, un monomère est composé de seulement quelques billes gros grains. Ils ont pu ainsi simuler 8 systèmes sur des temps de 5 à 25 μs et étudier la stabilité de capsides vides. Ils ont observé que dans la moitié des cas, les capsides vides s'effondraient sur elles-mêmes ce qui avait déjà été rapporté dans une autre étude utilisant des modèles très gros grains sur des capsides virales [91].

3.3 Assemblage de capsides virales *in silico*

Il n'existe pas encore de simulation tout atome d'auto-assemblage(s) d'une capsides virale. Les principales raisons étant la taille importante du système ainsi que le temps nécessaire pour cet assemblage (pouvant dépasser la minute).

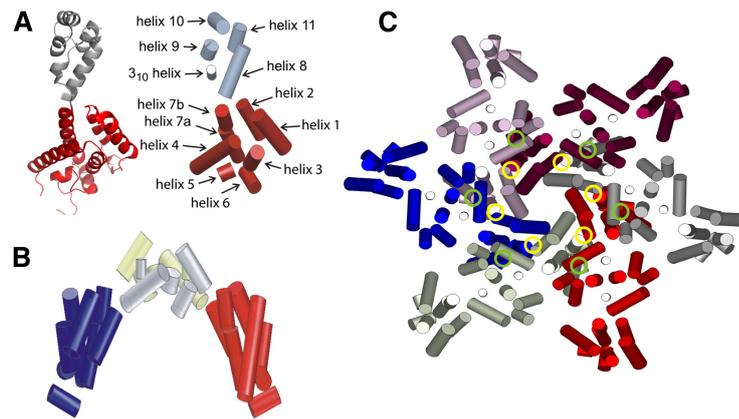
Diverses méthodes *in silico* peuvent cependant être utilisées pour étudier l'auto-assemblage des capsides virales. La plupart du temps, ces méthodes sont développées pour un problème en particulier et peuvent utiliser des méthodes "gros grains" (CG, voir partie [Simulation gros grains](#) page 52) à différentes échelles, ou encore des formes géométriques.

3.3.0.1 Méthodes gros grains

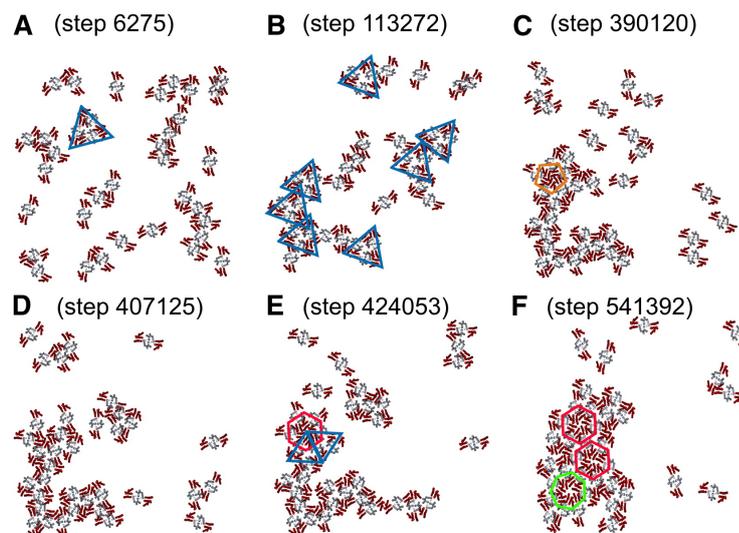
L'exemple de Chen *et al.*

Prenons par exemple le cas de l'étude de l'auto-assemblage de la capsides du VIH par Chen *et al.* (2011)[92]. Les auteurs ont utilisé une représentation gros grains imitant la structure du squelette de l'unité asymétrique ainsi que les interactions intermoléculaires observées expérimentalement (Fig. 1.22a).

Chen *et al.* ont réalisé ensuite une simulation de Monte-Carlo en deux dimensions et ils ont retrouvé la formation d'un réseau ordonné avec la symétrie hexagonale de la capsides (Fig. 1.22b). En modifiant la température de la



(a) Description du modèle gros grains utilisé



(b) Résultats de la simulation de Monte-Carlo

Figure 1.22: Étude des premières étapes de l'auto-assemblage de la capsid du VIH. (1.22a) Description du modèle gros grain 2D utilisé. (1.22b) Résultat d'une simulation de Monte-Carlo à partir d'un modèle gros grains. Tiré de Chen *et al.* (2011)[92].

simulation, ils ont alors pu établir un diagramme de phase approximatif et proposé un modèle d'assemblage en deux étapes principales :

1. Condensation des dimères (modèle CG) en groupes désordonnés
2. Nucléation du réseau par l'apparition d'un hexamère dans un groupe.

Ce type de simulation est particulièrement bien adapté pour la capsid du VIH de par sa grande taille (120x60 nm [93]) et sa géométrie : elle est composée majoritairement d'hexamères qui forment un plan, adaptée donc pour une simulation 2D. La courbure de la capsid du VIH est apportée par un petit nombre de pentamères [94].

L'exemple de Grime *et al.*

Nous pouvons également mentionner une seconde étude récente sur la capsid du VIH par Grime *et al.* (2016) [95]. Cette fois-ci, les trois dimensions sont utilisées afin de simuler, sous diverses conditions, la croissance de la capsid du VIH. Le modèle utilisé est un modèle gros grains basé sur les carbones alpha avec un réseau élastique entre les monomères de la protéine de capsid. Afin de mimer les interactions protéine/protéine requises pour l'auto-assemblage, une force attractive est rajoutée en positionnant des minimums locaux paramétrisés d'après les structures expérimentales [95].

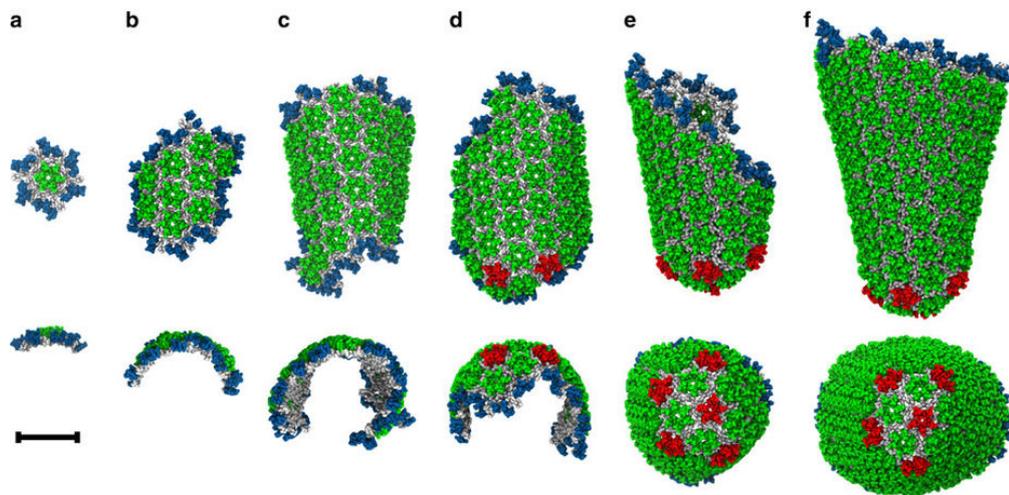


Figure 1.23: Assemblage de la capsid du VIH avec un modèle gros grains 3D. Tiré de Grime *et al.* (2016)[95].

Ils ont ainsi été en mesure de produire un auto-assemblage de dimères en une structure semblable à la capsid du VIH (cf. Fig. 1.23). En jouant sur les concentrations en dimères présents initialement dans le système, ils ont pu observer divers effets sur l'assemblage comme des phénomènes d'agrégation ou l'obtention d'assemblages incorrects.

L'exemple de Johnston *et al.*

En 2010, Johnston *et al.* ont utilisé un modèle beaucoup plus simple des capsomères de capsides icosaédriques. Les unités d'assemblage étaient simplifiées en des pyramides pentamériques (T=1 et T=3) et de plan hexamérique (T=3)[96].

En optimisant les paramètres de leur champ de force afin d'optimiser la distance et l'angle entre deux capsomères, ils ont pu étudier l'assemblage

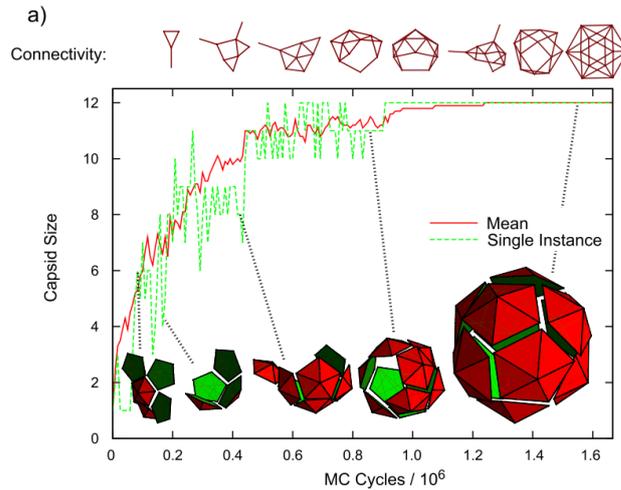


Figure 1.24: Assemblage de capsid icosaédrique ($T=1$) en utilisant des modèles "très" gros grains : les capsomères sont représentés ici pas des pyramides pentamériques. Tiré de Johnston *et al.* (2010)[96].

réversible de capsid de type $T=1$ (cf. Fig. 1.24) reproduisant ainsi les observations expérimentales d'une autre étude [96].

3.4 Étude du norovirus *In silico*

Le norovirus a été peu étudié par des méthodes *in silico*. À notre connaissance, une étude par nanoindentation a été réalisée sur la capsid du GI.1 [97] et deux études traitent de l'interaction entre des ligands et le site au récepteur HBGA sur le domaine P [98, 99].

3.4.0.1 Étude de la capsid par nanoindentation

L'étude de la stabilité d'un assemblage macromoléculaire par nanoindentation est une technique inspirée de la microscopie à force atomique. Dans un premier temps, une "sonde" *in silico* (généralement en forme d'hémisphère et composée de particules identiques d'une masse de 2kD) va être posée sur la surface d'une macromolécule. Dans un second temps, une force va être appliquée sur la sonde vers le centre de masse de la capsid. La sonde va alors rentrer dans la capsid en perturbant les interactions entre molécules et en rompant les contacts entre protéines (cf. Fig. 1.25). Ainsi, en modifiant l'endroit où la sonde est posée initialement (les axes de symétries par

exemple), et en jouant sur la force, il est possible de déterminer des zones où les interactions entre protéines sont plus faibles.

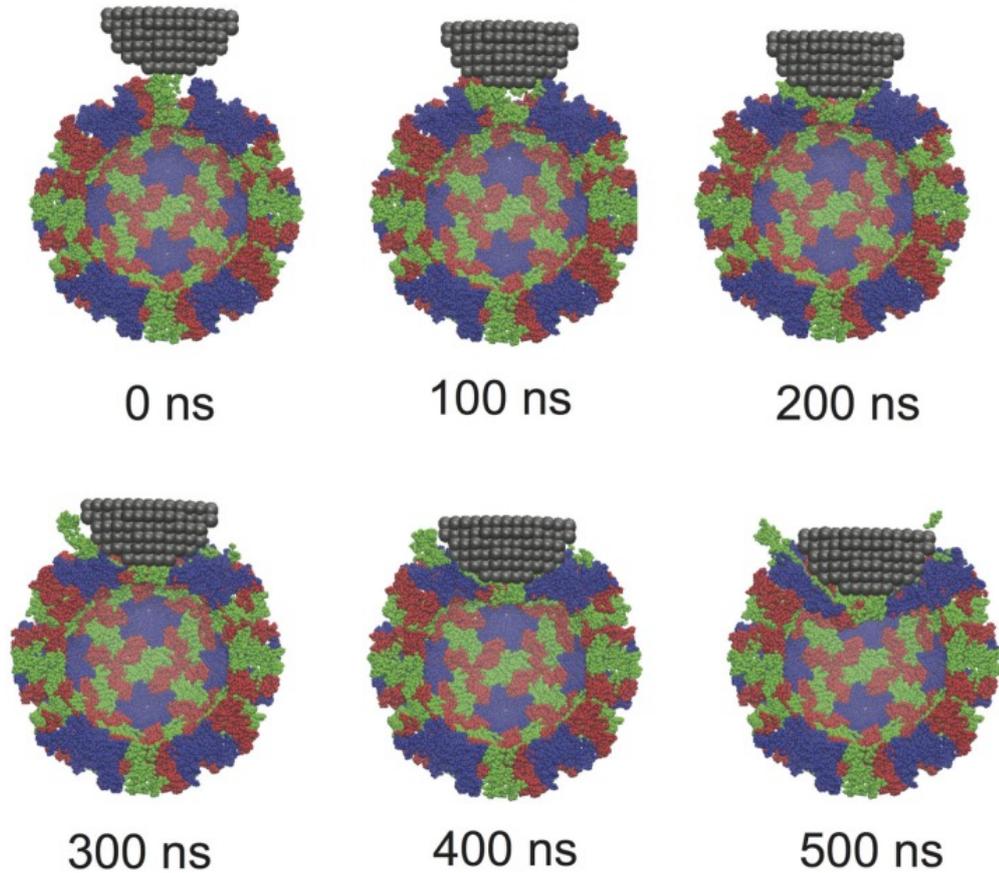


Figure 1.25: Exemple de nanoindentation sur la capside du norovirus. La sonde est représentée en hémisphère de particules grise, la capside est représentation "tronquée" colorée par chaîne (A :bleu, B :rouge, C :vert). Tiré de Boyd *et al.* (2015)[97].

Cette méthode est donc très utile pour étudier le désassemblage des capsides virales et a été utilisée en 2015 par Boyd *et al.* [97] sur la capside du GI.1 du norovirus en représentation gros grains.

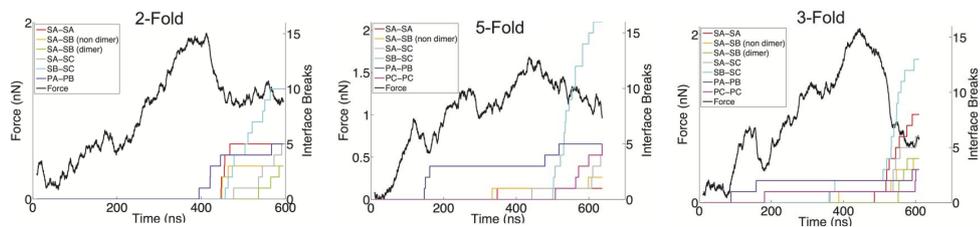


Figure 1.26: Résultats des analyses par nanoindentation sur la capside du Norovirus. "Les courbes colorées correspondent à la somme cumulée du nombre d'interface rompue". Tiré de Boyd *et al.* (2015)[97].

Les auteurs ont posé la sonde sur l'axe 2 (entre deux monomères C de VP1), sur l'axe 5 (pentamère de dimères) et sur l'axe 3 (hexamère de dimères). Ils ont ensuite suivi le nombre d'interfaces entre domaines S qui se rompaient sous la force de la sonde (cf. Fig. 1.26). Ils ont remarqué que le nombre d'interfaces rompues entre les domaines S de la chaîne B et C (Fig. 1.26 courbe cyan) est bien plus important dans les 3 cas, indiquant que cette interface dans la capsid est bien plus faible que les autres (axes 2 et 5).

3.4.0.2 Étude de liaisons de ligands

À notre connaissance, deux études ont été rapportées dans la littérature portant sur le dimère P de VP1 [98, 99]. Dans les deux cas, les structures utilisées sont un dimère de domaine P du GII.4 obtenu par Cao *et al.* avec des trisaccharides de groupes sanguins différents (A et B)[57].

Koppisetty *et al.* en 2010 ont remarqué, en simulant le dimère de P en complexe avec un trissaccharide de groupe sanguin de type A (tri-A) (2OBS [57]) que ce ligand comportait des erreurs géométriques, remettant en cause la justesse de cette structure. Ils ont procédé à un remplacement du tri-A par d'autres HBGA. Ils ont pu mettre en évidence que des fucoses liés en position $\alpha 1,2$ et des sucres avec des terminus en GalNAc α ou Gal $\alpha 1,3$ faisaient de fortes interactions avec la protéine, ouvrant la voie pour des pistes thérapeutiques.

Lundbord *et al.* ont donc utilisé la seconde structure de Cao *et al.* avec un trissaccharide de groupe sanguin B, que Koppisetty *et al.* avait jugé correct : 2OBT [57]. Ils ont réalisé du [screening virtuel*](#) suivi de courtes simulations de dynamique moléculaire afin de pouvoir calculer au mieux l'énergie libre de liaison de ces composés au site de liaison de la protéine. Ceci permettrait de prédire avec une meilleure confiance la liaison d'un ligand (ce qui est important dans la conception d'un potentiel inhibiteur). Ainsi, ils proposent 4 molécules avec une haute affinité comme inhibiteur du GII.4. Cependant, ils ont été évalués en 2016 comme étant d'une difficulté intermédiaire à synthétiser [100]. Aucune suite n'a été rapportée sur ces molécules.

L'étude de Koppisetty *et al.* a cependant permis de mettre la lumière sur un point particulier : le site de liaison à un HBGA est différent pour le GII.4 [100] et pour le GI.1 [59, 101] bien que ces deux souches touchent les humains.

Matériels et méthodes

” *Nous sommes tous des poussières d'étoiles*

— **André Brahic**

Enfants du soleil, histoire de nos origines.

Durant cette étude, plusieurs méthodologies ont été utilisées : des méthodes dites *in silico*, que nous pouvons traduire littéralement par "en silice", qui signifie que nous utilisons les ressources et la puissance computationnelle d'ordinateurs ou de super-ordinateurs. Ces méthodes peuvent être utilisées par exemple afin d'analyser des événements d'ordre atomique et expliquer ou corroborer certains résultats expérimentaux.

Nous avons également utilisé des méthodes biophysiques comme la diffusion de Rayons-X aux Petits Angles (ou SAXS pour *Small Angle X-Ray Scattering*) qui permet d'observer des événements à plus grande échelle (échelle nanométrique) et d'extraire des informations correspondant à la forme de l'objet observé.

Le but de cette partie n'est pas de faire un cours exhaustif du sujet, mais de donner au lecteur les bases nécessaires à la compréhension de ces méthodes.

Il existe de nombreuses méthodes *in silico*. Les plus utilisées pendant cette étude ont été la modélisation par homologie et la simulation de dynamique moléculaire (MD).

1.1 Modélisation par homologie

1.1.1 Principe

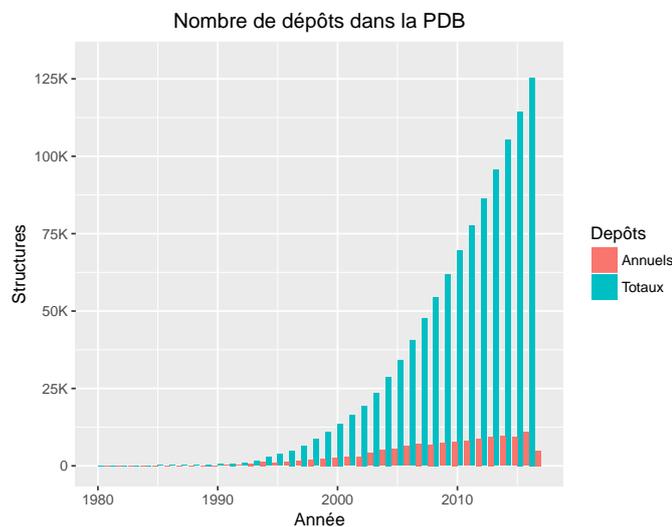


Figure 2.1: Nombre de structures déposées par année et en cumulé dans la Protein Data Bank depuis 1980.

Depuis la création de la Protein Data Bank [102] (PDB) en 1971 par le laboratoire national de Brookhaven (US), le nombre de dépôts de structures n'a cessé de croître. Avec les progrès technologiques, notamment dans le domaine de la cristallographie et de la RMN, le nombre de structures déposées a augmenté drastiquement depuis le début des années 90 (cf. Fig. 2.1) pour atteindre plus de 130 000 structures disponibles (mai 2017). Cependant, ce nombre est loin derrière les 84 millions de séquences de protéines déposées dans la base de données RefSeq [103] (mai 2017). Avec seulement 0.15% des protéines connues disposant d'informations tridimensionnelles, il était nécessaire de trouver des méthodes pouvant combler le manque de données

structurales. Parmi ces méthodes, nous pouvons compter la modélisation par homologie (ou modélisation comparative), qui se base sur le principe acquis que deux protéines de séquences proches auront des structures proches.

Cela permet, pour une protéine d'intérêt pour laquelle nous avons la séquence (séquence cible), d'obtenir un modèle structural basé sur un homologue dont la structure est connue, pour peu que les deux séquences soient suffisamment similaires (plus de 30% de similarité [104]).

De nombreux programmes de modélisation par homologie existent, les plus connus d'entre eux étant SWISS-MODEL [105], MODELLER ou I-TASSER [106]. I-TASSER et SWISS-MODEL peuvent s'utiliser via un serveur web. Bien que de nombreuses options pour affiner la modélisation existent, ces programmes sont assez simples d'utilisation (la séquence cible suffit à obtenir un modèle) et assez efficaces pour obtenir un résultat rapidement sans faire de recherche de structure référence et d'alignement de séquences au préalable (la recherche d'homologues structuraux et les alignements de séquences sont réalisés par les programmes). MODELLER nécessite une structure 3D d'une protéine de référence dont on connaît la séquence, la séquence cible ainsi qu'un alignement entre la référence et la cible.

1.1.2 La modélisation comparative avec MODELLER

Bien que les méthodes employées par les différents programmes de modélisation par homologie puissent être différentes, l'obtention d'un modèle structural passe par 4 étapes principales.

1.1.2.1 Identification de la structure référence

La première étape consiste à identifier une, ou plusieurs, structures de référence. Cette étape peut être effectuée à travers une recherche de séquences homologues pour les protéines dont une structure tridimensionnelle résolue est déposée à la PDB. De nombreux services en ligne existent pour réaliser cette étape. Nous pouvons citer par exemple PSI-BLAST [107], JACKHHMER

[108] ou bien HHPRED [109] qui, pour ce dernier, permet de produire un fichier d'alignement au format requis par MODELLER. Il est cependant très important d'avoir un regard critique sur la référence sélectionnée et de noter en premier lieu son identité de séquence (sa correspondance) avec la séquence cible.

Il est communément admis qu'il faut au moins 30% d'identité de séquence afin d'avoir un modèle par homologie pertinent [104, 110]. Les modèles basés sur une structure référence avec une identité de séquence entre 20% et 30% (cf. Figure 2.2) sont à manipuler avec précaution, il s'agit d'une "zone floue" (*twilight-zone*) [111]. Si aucune autre référence avec une meilleure identité n'est disponible, il faut faire attention à sa source (*ie.* ne pas modéliser une protéine membranaire virale à partir d'une protéine de pilus bactérien). En dessous de 20% d'identité, l'homologie de séquence n'est plus suffisante pour garantir une homologie structurale fiable, néanmoins selon la robustesse des programmes utilisés, l'obtention d'un modèle est toujours possible, mais principalement à but illustratif et non analytique (par exemple Fig. 1.8).

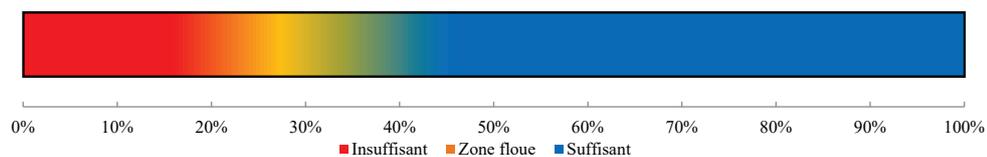


Figure 2.2: Modélisation par homologie et identité de séquence : zone de confiance.

Cette étape peut être réalisée par l'utilisateur, ou bien être réalisée automatiquement sur les serveurs comme I-TASSER ou SWISS-MODEL. Durant cette étude, la structure de référence étant déjà connue (pdb : 1IHM [56]) avec une identité de séquence d'environ 48%, cette étape n'a donc pas été nécessaire.

1.1.2.2 Alignement de séquences

Entre deux protéines homologues, il peut y avoir des résidus en plus (insertion), en moins (délétion, GAP) ou bien différents (mutation). Une fois la ou les structures références identifiées, leurs séquences doivent donc être

alignées avec la séquence cible. Un alignement de séquences va permettre d'obtenir une correspondance entre les deux séquences et produire une séquence en indiquant les résidus manquants par des GAP ("-") tout en faisant correspondre les résidus identiques ou similaires (*ie.* avec des propriétés physico-chimiques similaires). Ces informations sont stockées dans un fichier texte au format PIR qui permet d'inclure de nombreuses informations :

- les noms (et chemins) des structures références et cibles
- les acides aminés à modéliser, ainsi que l'identifiant de la ou des chaîne(s).
- l'origine de la protéine (bien que non nécessaire à la modélisation)
- la résolution de la structure cristallographique et son R-facteur (facteur de qualité de la structure).

1.1.2.3 Création du modèle

Basé sur l'alignement de séquence, le squelette (*backbone*) de la protéine cible ayant une correspondance structurale avec la protéine référence sera d'abord reconstruit [104]. Les chaînes latérales sont ensuite ajoutées suivant le même procédé.

Pour construire les séquences qui ne sont pas présentes dans la structure référence, un certain nombre de contraintes [110] vont être établies par MODELLER. Dans un premier temps, des contraintes de distances et d'angles dièdres seront créées, basées sur les informations structurales de référence (structures PDB) ainsi que l'alignement de séquences. Dans un second temps, des contraintes stéréochimiques (spatiales) vont être ajoutées [112] en se basant sur le champ de force CHARMM22 [113]. Tous ces termes sont ensuite introduits dans une **fonction objectif*** qui est ensuite optimisée en utilisant des méthodes de MD (gradient conjugué et recuit simulé [114], partie **Recuit simulé** page 54 pour plus de détails). Plusieurs modèles différents peuvent être produits qui doivent ensuite être évalués.

1.1.2.4 Évaluation des modèles

Il est généralement conseillé de réaliser plusieurs modèles (au minimum 5 [104]) que l'on évaluera ensuite. Les modèles sont évalués sur la valeur finale de la fonction objectif ainsi que sur le score DOPE [115]. Ce score représente une somme d'interactions entre des paires d'atomes. *In fine*, le meilleur modèle peut s'apparenter à celui avec la plus petite valeur de la fonction objectif [112]. Néanmoins, il faut garder un regard critique sur ces modèles et les valider par un moyen externe, comme une analyse de la carte de Ramachandran [116] pour vérifier qu'il n'y ait pas d'aberration dans le modèle obtenu.

1.2 La simulation de dynamique moléculaire

L'observation de détails atomiques est souvent indispensable à la compréhension des fonctions d'un objet biologique dans son ensemble. Si certaines méthodes, comme la cristallographie, la résonance magnétique nucléaire (RMN) et la cryo-EM, peuvent permettre l'obtention d'informations jusqu'à l'échelle de l'atome, elles ne produisent que des instantanés.

La simulation de dynamique moléculaire (MD) permet de s'affranchir des limites expérimentales à travers des simulations informatiques (*in silico*) et ainsi observer à une échelle atomique l'évolution d'un système dans des conditions proches de l'*in vivo* (solvant, force ionique...).

Afin de faciliter les calculs, la MD classique considère que chaque atome est une "bille" de masse et de charge définies, contrairement à la MD quantique qui tient compte de l'environnement électronique.

1.2.1 Principe

Durant les simulations, les atomes interagissent entre eux et leurs trajectoires sont obtenues en résolvant l'équation différentielle de la seconde loi de Newton ($F = ma$) :

$$\frac{d^2 x_i}{dt^2} = \frac{F_{r_i}}{m_i} \quad \text{avec} \quad F_{r_i} = \nabla E \quad (1.1) \quad (1.2)$$

Ainsi,

$$-\frac{\partial E}{\partial r_i} = m_i \frac{d^2 x_i}{dt^2} \quad (1.3)$$

L'équation 1.1 permet de décrire le mouvement d'une particule de masse m_i sur des coordonnées cartésiennes r_i avec une force F_{x_i} appliquée dans la direction du mouvement. La force exercée sur une particule dépend de ses interactions avec d'autres particules. Ces interactions sont définies selon certaines règles que l'on appelle champ de force. Ces champs de forces sont une somme de fonctions mathématiques permettant le calcul de l'énergie potentielle (E) d'un système (cf. equation 1.4).

1.2.2 Calcul de l'énergie potentielle

$$E_{pot} = \underbrace{\sum_{liaisons} \frac{k_l}{2} (l - l_0)^2 + \sum_{angles} \frac{k_\theta}{2} (\theta - \theta_0)^2 + \sum_{dièdres} \frac{V_n}{2} (1 + \cos(n\omega - \gamma))}_{\text{termes liés}} + \underbrace{\sum_{i=1}^{N_A} \sum_{j=1}^{N_{SEUIL}} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} + \sum_{i=1}^{N_A} \sum_{j=1}^{N_{SEUIL}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{\text{termes non liés}} \quad (1.4)$$

Certains termes (en rouge) sont fixés dans le champ de force. Ils doivent être calibrés et validés par des résultats expérimentaux. Les valeurs bleues, elles, sont variables et dépendent de l'état de la molécule à un instant t de la simulation.

Nous pouvons également séparer cette équation en deux parties : les termes liés et les termes non liés.

1.2.2.1 Les termes liés

Terme de liaison

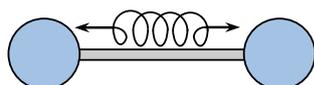


Figure 2.3: Distance entre deux atomes.

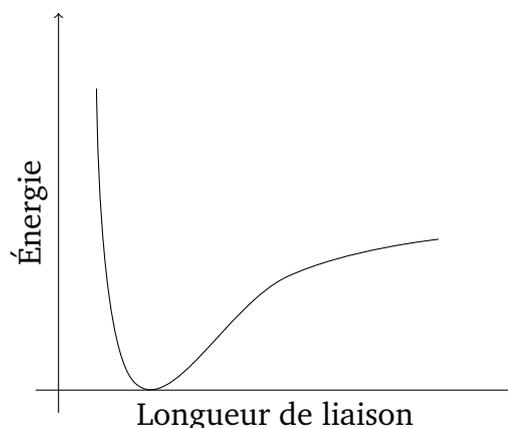


Figure 2.4: Exemple d'un potentiel harmonique : variation de l'énergie en fonction de la longueur de liaison

Le premier terme de l'équation de l'énergie potentielle (eq. 1.4) modélise l'interaction entre une paire d'atomes liés par une liaison covalente. Il s'agit d'un potentiel de Morse : un potentiel pseudo-harmonique permettant de mimer la liaison covalente entre deux atomes comme un ressort (cf. Fig. 2.3), mais en comprenant les effets de rupture de liaison. Ce potentiel va augmenter quand la longueur de liaison l dévie de sa valeur de référence l_0 (valeur équilibre intrinsèque au champ de force, spécifique au couple d'atomes) et induit une augmentation de l'énergie potentielle (cf. Fig. 2.4). Étant donné que la force entre deux atomes liés est très forte, une très grande quantité d'énergie est nécessaire pour faire dévier la longueur de liaison de sa valeur d'équilibre. Pour pallier ceci une constante de raideur, k_l , dépendante de chaque paire d'atomes, est rajoutée.

Angles

Le second terme de l'équation de l'énergie potentielle (eq. 1.4) est une somme sur tous les angles de valence (Fig. 2.5) dans la molécule. Là encore il s'agit d'un potentiel harmonique où l'angle θ oscille autour de la valeur référence θ_0 ,

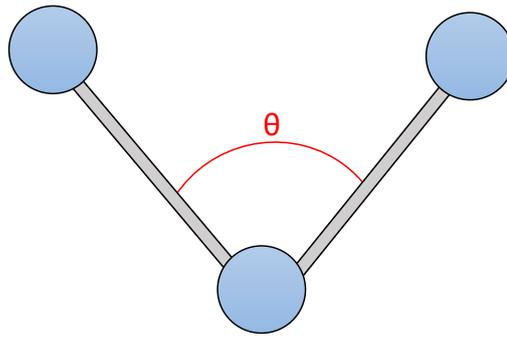


Figure 2.5: Illustration d'un angle entre trois atomes.

également intrinsèque au champ de force. θ_0 et l_0 est spécifique au trinôme d'atomes concerné avec une constante de force k_θ .

Angles dièdres

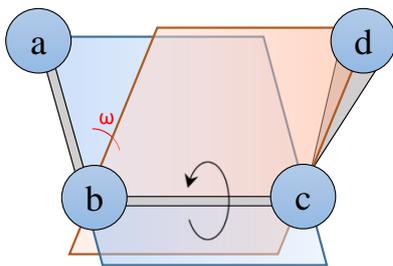


Figure 2.6: Représentation d'un angle dièdre entre 4 atomes.

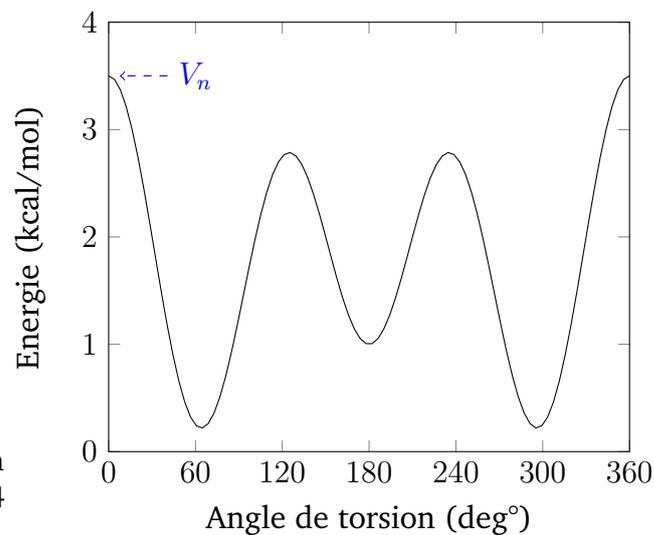


Figure 2.7: Exemple de la variation de l'énergie en fonction de l'angle dièdre pour un groupement chimique.

Le troisième terme, et le dernier des termes liés représente les angles dièdres : l'angle entre quatre atomes, liés de manière covalente. Il représente l'angle formé par les plans passant par les atomes a, b, c et les atomes d, c, b (cf. Fig. 2.6).

Dans l'équation 1.4 V_n représente la hauteur de la barrière énergétique (valeur maximum de l'énergie de torsion). ω représente l'angle dièdre entre

4 atomes, n la multiplicité, c'est-à-dire le nombre de points minimum dans la fonction sur 360° . Enfin, γ est l'angle de phase qui détermine où l'angle de torsion passe par sa valeur minimale.

La figure 2.7 est un exemple de l'énergie de torsion pour un groupement O-C-C-O avec le champ de force Amber [117]. Chaque quadrinôme d'atomes a ses constantes V_n , λ et n .

1.2.2.2 Les termes non liés

Interactions électrostatiques

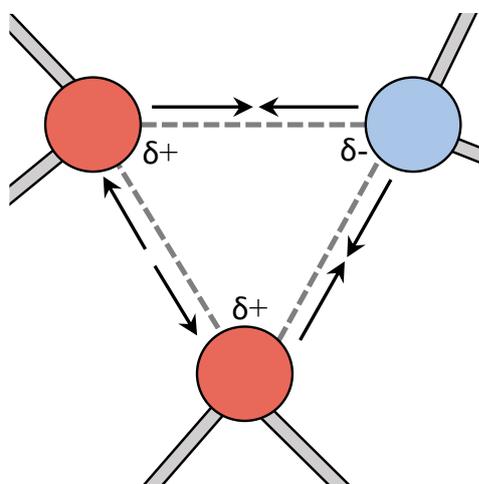


Figure 2.8: Interactions électrostatiques

Deux atomes peuvent interagir à distance sans être liés par une liaison covalente. Ces types d'interactions sont considérés dans deux termes dits "non liés" qui sont les interactions électrostatiques et les interactions de Van der Waals. Dans la nature, deux particules de même charge électrique, positive ou négative, vont s'éloigner. À l'inverse, deux particules de charges opposées vont s'attirer (cf. figure 2.8). Ce sont les interactions électrostatiques décrites par la loi de Coulomb.

Une des approches les plus utilisées pour le calcul de l'électrostatique est la sommation d'Ewald [118] permettant de calculer les interactions d'une molécule avec toutes les autres dans sa boîte de simulation, mais également dans les images périodiques. Elle permet d'obtenir la charge en tout point d'une molécule par le biais d'une somme de gaussiennes calculées dans

l'espace réciproque et dans l'espace réel à partir de charges ponctuelles (qui elles sont calculées dans l'espace réel).

La méthode de sommation d'Ewald est coûteuse, car elle nécessite de calculer les interactions pour toutes les paires d'atomes (algorithme en $O(N^2)$). Une méthode a été développée afin de réduire le temps de calcul : le PME (*Particle Mesh Ewald*). Basée sur la méthode d'Ewald, elle consiste à rajouter une grille virtuelle au système, puis pour chaque point de la grille, d'interpoler leur charge (cf. Fig. 2.9). Le potentiel électrostatique est ensuite calculé dans l'espace réciproque, grâce à une transformée de Fourier, et les forces interpolées en retour vers les atomes, dans l'espace réel, dont la position est mise à jour. En plus de cela, un *cutoff* de distance est utilisé afin de réduire le nombre de calculs à réaliser.

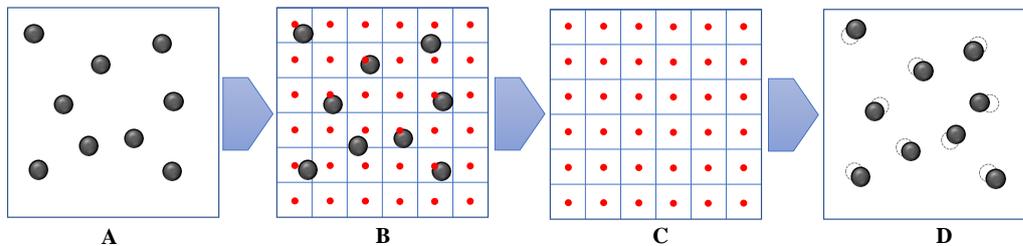


Figure 2.9: Approche de type PME sur un réseau bidimensionnel. (A) Système de particules chargées. (B) Les charges sont interpolées sur une grille bidimensionnelle. (C) Le potentiel et les forces sont évalués en chaque point de la grille dans l'espace réciproque grâce à une transformée de Fourier. (D) Les forces sont interpolées en retour (transformée de Fourier inverse) vers les particules, dont la position est ensuite mise à jour. Adapté de Chipot (2003)[119].

Dans l'équation 1.4 q_i et q_j sont respectivement les charges du point i et du point j . N_A est le nombre de points dans la grille électrostatique et N_{SEUIL} est le nombre de points se trouvant à une distance inférieure au *cutoff* de distance de l'atome i . r_{ij} est la distance entre l'atome i et j et ϵ_0 est la constante de la loi de Coulomb : la permittivité du vide (réponse du vide à un champ électrique appliqué). Cette constante est ajustée par la permittivité diélectrique relative du milieu considéré ϵ_r .

Interactions de Van der Waals

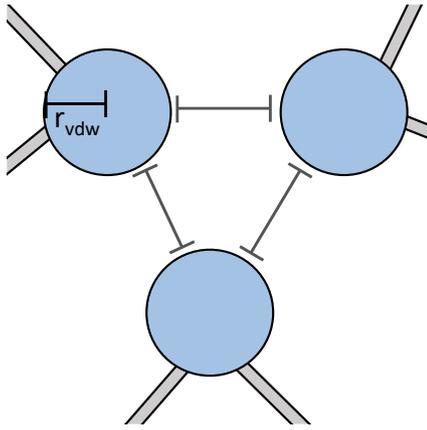


Figure 2.10: Interactions de Van der Waals.

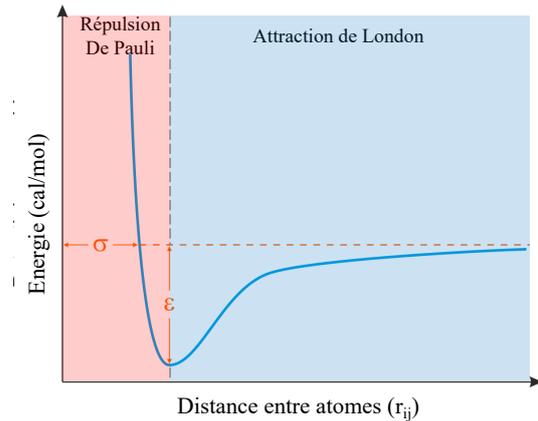


Figure 2.11: Potentiel de Lennard Jones.

Le dernier terme non lié, et de l'équation 1.4, représente l'attractivité et la répulsion entre deux atomes basées sur leur rayon de Van der Waals : les interactions de Van der Waals (cf. Fig. 2.10). Ces interactions sont prises en compte à travers un potentiel de Lennard-Jones (cf. Fig. 2.11) qui va être décomposé en deux termes :

- Un terme répulsif qui va repousser les atomes à courte distance (énergie très haute) grâce au principe d'exclusion de Pauli [120] : deux nuages électroniques ne peuvent pas s'interpénétrer car deux électrons ne peuvent pas occuper le même état quantique.
- Un terme attractif qui va attirer les atomes à longue ou moyenne distance selon le principe de London [121] : 2 atomes soumis à des interactions attractives, indépendamment de l'électrostatique, créent des dipôles instantané dus aux mouvements des électrons autour des noyaux.

Les interactions de Van der Waals n'agissent qu'à très courte distance contrairement à l'électrostatique.

La constante σ_{ij} , dépendante du champ de force et de la paire d'atomes i et j est le rayon de collision : la distance de séparation de deux atomes i, j où l'énergie vaut 0 (force répulsive et attractive identique). ϵ_{ij} correspond à un puits de potentiel, c'est-à-dire au minimum de la fonction de potentiel pour le couple d'atomes i, j . Enfin, r_{ij} la distance entre les atomes i et j .

Sur une longue distance, l'énergie vaut 0 et est donc négligée. Ce qui signifie que de même que pour les interactions électrostatiques, seuls les atomes étant à une distance inférieure à une valeur-seuil sont considérés pour le calcul de ce terme. Enfin, toujours dans le quatrième terme de l'équation 1.4, N_A correspond au nombre d'atomes et N_{seuil} aux nombres d'atomes proches ($< seuil$) de l'atome i .

1.2.3 Conditions périodiques aux limites

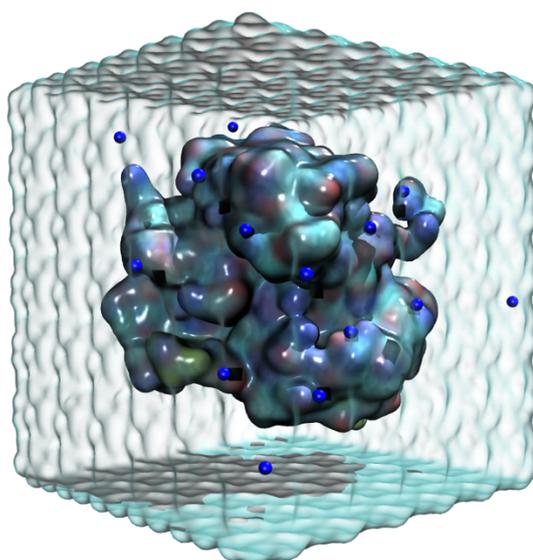


Figure 2.12: Exemple d'une boîte de simulation avec une protéine en représentation QuickSurf, les ions Na^+ en sphère de Van der Waals (bleu) et le solvant en QuickSurf (cyan semi-transparent). Réalisé avec le logiciel VMD [122].

Lors d'une simulation, la molécule considérée est plongée dans un solvant ; le plus souvent de l'eau. Afin de réduire le nombre de molécules d'eau, et ainsi le temps de calcul, une boîte de simulation de taille limitée est utilisée (cf. Fig. 2.12). Cependant, afin d'éviter des effets de bord (particules en contact avec le vide), la boîte de simulation est répliquée sur tous les côtés (cf. Fig. 2.13).

Ainsi, si une molécule sort d'un côté, elle reviendra dans la boîte par son côté opposé (Fig. 2.13a). Concernant le calcul des interactions longue distance

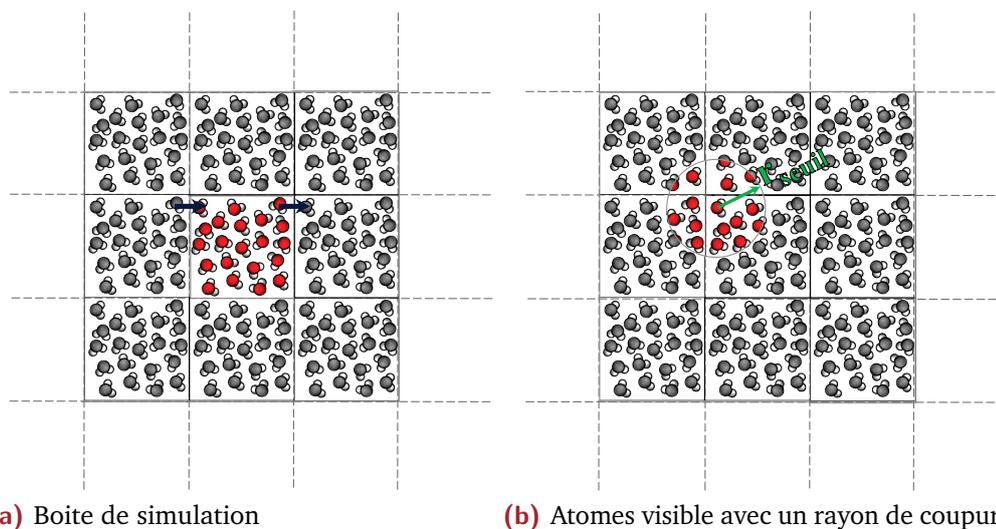


Figure 2.13: Représentation 2D des conditions périodiques avec du solvant (eau). (2.13a) La boîte de simulation est en couleur, les boîtes périodiques virtuelles en gris. (2.13b) Pour un atome d'oxygène d'une molécule, les interactions électrostatiques et de Van der Waals seront calculées avec les atomes rencontrés dans un rayon r_{seuil} (en couleur).

pour un atome i , il sera réalisé sur tous les atomes situés dans un rayon fixé (r_{seuil}) et dans les boîtes virtuelles (Fig. 2.13b).

Cependant, bien que cette méthode permette de simuler un milieu virtuellement infini, il faut veiller à ce que la boîte ne soit pas trop petite pour éviter que le système ne se voie lui-même avec la valeur-seuil pour le calcul des interactions non-liantes.

1.2.4 Ensemble thermodynamique

Durant une simulation de dynamique moléculaire classique, la pression et la température fluctuent autour d'une valeur référence, et restent stables. Le nombre d'atomes est quant à lui constant. C'est un ensemble thermodynamique isobare-isotherme, ou NPT (N pour le nombre d'atomes, P pour pression et T pour température). La pression et la température sont maintenues autour d'une valeur définie par le biais d'un thermostat et d'un barostat.

1.2.4.1 Température

Dans un ensemble NPT, la température est gardée constante autour d'une valeur d'équilibre par un thermostat pour, par exemple, éviter des changements de comportement du système durant la dynamique (comme le dépliement d'une protéine). Elle est directement liée à l'énergie cinétique (ainsi qu'à la vitesse des molécules) par l'équation :

$$E_{cin} = \sum_{i=1}^N \frac{|p_i|^2}{2m_i} = \frac{k_B T}{2} (3N - N_c) \quad (1.5)$$

Dans l'équation 1.5 p_i est le **moment total*** de la molécule (produit de la masse et de la vitesse soit $p = m_i v_i$) et k_B la constante de Boltzmann. N correspond au nombre de degrés de liberté, soit au nombre de particules, ici factorisé par 3, car chaque particule a 3 coordonnées cartésiennes. Enfin, N_c correspond au nombre de contraintes dans le système (conditions que le système est forcé de satisfaire).

Il existe plusieurs types de thermostats permettant de maintenir une température constante en modifiant la vitesse des particules :

- **BERENDSEN** [123] : corrige la température en multipliant les vitesses par un facteur λ à chaque pas de temps t afin d'amener le système à la température souhaitée. Cet algorithme est très efficace pour amener rapidement un système à la température souhaitée, mais il supprime la fluctuation de l'énergie cinétique, ce qui empêche le système de générer un **Ensemble canonique*** [124]. Il est très utilisé pendant les phases d'équilibration des systèmes afin de l'amener à son équilibre le plus rapidement possible (de manière exponentielle).
- **V-RESCALE** [125] : ce thermostat est très semblable au thermostat de Berendsen, mais avec un terme stochastique supplémentaire qui assure une distribution correcte de l'énergie cinétique et ainsi permet de générer un ensemble canonique [124]. C'est le thermostat V-RESCALE qui est le plus utilisé dans les dynamiques de production.

- **ANDERSEN** [126] : ce thermostat s'appuie sur une distribution de Maxwell-Boltzmann afin de retirer aléatoirement des vitesses pour les particules simulées. Il peut être utilisé afin de maintenir un ensemble thermostaté, mais il n'est pas entièrement parallélisable et compatible avec tous les algorithmes de simulations.
- **NOSÉ-HOOVER** [127, 128] : ce thermostat est bien plus compliqué que les autres. Il utilise un paramètre de friction avec son propre moment total et l'équation de mouvement qui entraîne des oscillations autour de la valeur d'équilibre. Il est également utilisé pendant les phases d'équilibration, mais contrairement au thermostat de Berendsen, l'état d'équilibre du système est beaucoup plus long à être atteint (4 à 5 fois plus long [124]).

1.2.4.2 Pression

Certains comportements de biomolécules ne peuvent se déclencher qu'à partir d'une certaine pression. Il peut se révéler utile dans certains cas de maintenir une pression constante dans le système, que ce soit pour simuler les conditions de pression à l'intérieur d'une capsid de bactériophage ou bien garder une pression similaire à celle du corps. Il est courant de garder une pression de 1 bar (1 atmosphère) maintenue grâce à un barostat. Un barostat est similaire à un thermostat, mais au lieu d'agir sur la vitesse des particules afin d'ajuster la température, le barostat va agir sur le volume du système en ajustant la position des particules. Trois types d'ajustement de volume peuvent être faits :

- **isotropique** : le volume est ajusté sans toucher à la boîte de simulation (principalement utilisé pour les systèmes avec une protéine dans l'eau)
- **semi-isotropique** : la boîte de simulation est modifiée sur les axes X et Y (adapté aux systèmes membranaires afin de permettre des fluctuations de l'aire de la membrane)
- **anisotropique** : La boîte de simulation est ajustée sur tous les axes (également adapté aux systèmes membranaires)

Deux algorithmes pour l'ajustement de la pression sont disponibles dans la suite GROMACS [129] :

- **BERENDSEN** : cet algorithme va ajuster les coordonnées des particules et la boîte de simulation (isotropique ou semi-isotropique) à chaque pas de simulation afin d'attendre la pression d'équilibre, et comme pour le thermostat de Berendsen, de manière exponentielle.
- **PARRINELLO-RAHMAN** : cet algorithme peut être comparé à celui du thermostat NOSÉ-HOOVER, car il permet une oscillation autour de la valeur d'équilibre, afin d'obtenir un vrai ensemble NPT [124]. Cependant, si le système est loin de l'équilibre, ce barostat peut induire de très grandes oscillations de la boîte et induire un *crash* du système.

1.3 Méthodes avancées

La MD classique n'est pas toujours adaptée à toutes les questions. Il existe de nombreuses méthodes de simulations avancées, afin de pouvoir, par exemple, simuler beaucoup plus longtemps ; franchir des barrières énergétiques ; étudier l'interaction spécifique d'un ligand à une protéine... Durant cette thèse, 4 méthodes dites "avancées" ont été utilisées : le gros grain, le recuit simulé, la dynamique moléculaire dirigée et les modes normaux.

1.3.1 Simulation gros grains

De manière générale, en simulation de dynamique moléculaire, la longueur des phases d'équilibration de température et de pression ainsi que la production de la dynamique augmentent avec la taille du système. De plus, le temps de simulation et le coût calculatoire augmentent également avec le nombre d'atomes contenu dans la boîte de simulation. La méthode du gros grains (ou *coarse grain*) est utilisée afin de simuler de très gros systèmes sur des durées de temps plus long.

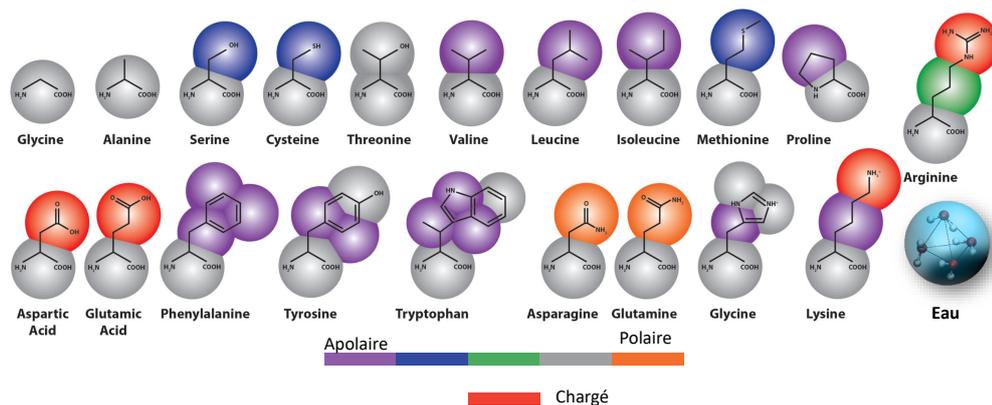


Figure 2.14: Représentation gros grains de chaque acide aminé. Les billes sont paramétrisées afin de reproduire les propriétés physico-chimiques du groupe d’atomes qu’elles remplacent. La nature (polaire ou apolaire) est ajustée ainsi que la charge. Tiré de Bradley *et al.* (2013) [130].

Principe

L’idée est de simuler des systèmes plus grands, plus longtemps, avec moins de ressources en réduisant le nombre d’atomes. Pour cela, certains atomes (avec des propriétés physico-chimiques similaires) des acides aminés vont être groupés pour former une plus grosse bille (cf. Fig. 2.14). L’eau, qui compose la grande majorité d’une boîte de simulation, est remplacée par des billes gros grains, qui chacune remplace 4 molécules d’eau. L’équation générale d’un champ de force (eq. 1.4) s’applique toujours, mais les termes spécifiques au champ de force (rouge) vont être modifiés afin de s’adapter à ces grains. Ainsi, la paramétrisation de chaque type de liaison, chaque type d’interaction, chaque angle est recalculé pour former le champ de force gros grains. La méthode gros grains la plus utilisée est MARTINI[131] qui a été validée expérimentalement. Un des avantages de ce champ de force, combiné au logiciel de simulation GROMACS, est que des molécules gros grains peuvent être combinées avec des molécules tout atomes.

Réseaux élastiques

Afin d’éviter d’avoir des effets de débobinement (déstructuration) des structures secondaires, un réseau élastique (potentiel harmonique) est inclus dans MARTINI en s’appuyant sur les interactions entre structures secondaires (feuilletts, hélices, boucles) préalablement calculées avec DSSP [132].

Deux types de réseaux élastiques existent dans le champ de force MARTINI. Le réseau élastique propre à MARTINI qui applique des contraintes uniquement sur le squelette de la protéine, et le réseau ELNEDYN qui est positionné en fonction de la distance entre les grains et qui contribue à la stabilité de la structure de la protéine au cours de la simulation (cf. Fig. 2.15). Bien entendu, il est possible de ne pas rajouter de réseau élastique afin de garder un maximum de flexibilité, mais dans ce cas, il faut bien connaître son système afin de ne pas voir la protéine se "débobiner". Ce système de réseaux est également totalement paramétrisable. Les unités de force du réseau en $\text{kJ.mol}^{-1}.\text{nm}^{-2}$ et du seuil de distance sont en nanomètre (nm).

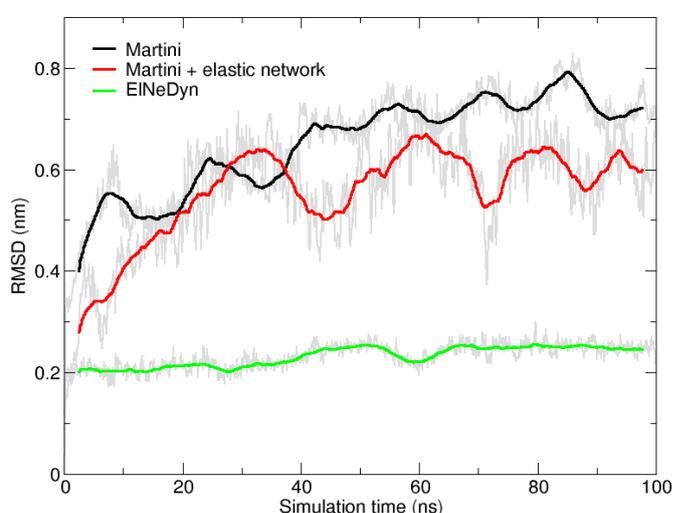


Figure 2.15: Comparaison du RMSD obtenu sur le squelette d'une protéase de HIV-1 avec les 3 différentes approches (Elnedyn, Martini et sans réseau élastique). Tiré du tutoriel Martini (<http://www.cgmartini.nl/index.php/tutorials-general-introduction/proteins>).

1.3.2 Recuit simulé

Dans une simulation de dynamique moléculaire classique, le système peut tomber dans un puits énergétique. Si ce puits de potentiel est trop profond, une dynamique moléculaire classique ne permettra pas au système de ressortir afin d'explorer d'autres puits énergétiques.

Le recuit simulé (ou *simulated annealing*) est inspiré par une technique de métallurgie qui consiste à chauffer un matériau puis le refroidir avec une

rampe de température contrôlée, afin de le rendre plus facile à manipuler, ou réduire ses défauts (par exemple avec des cristaux).

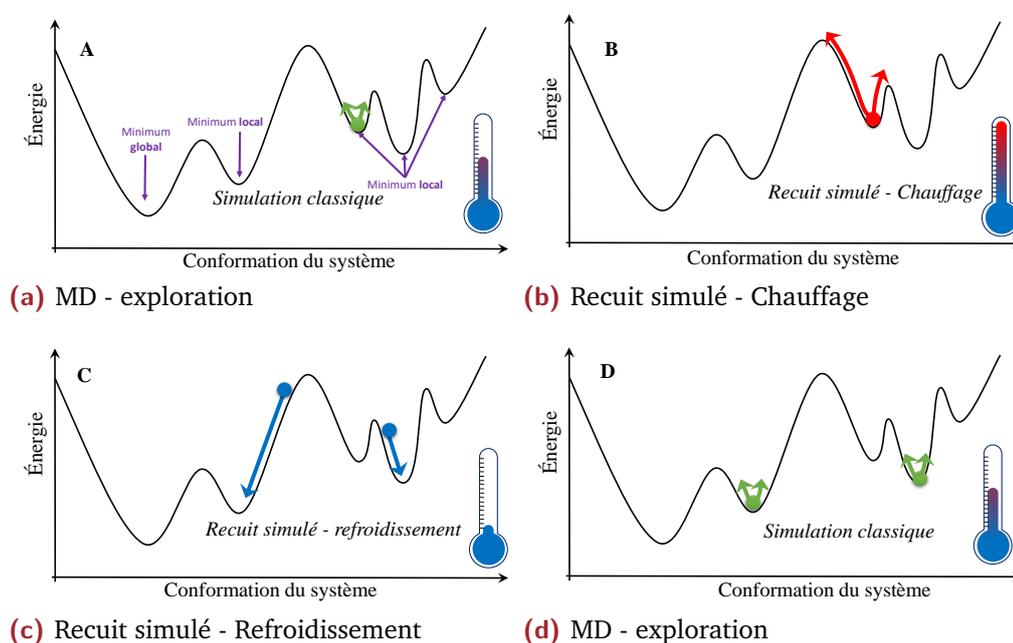


Figure 2.16: Représentation 2D d'une surface d'énergie potentielle d'un système. (2.16a) Exploration d'un puits de potentiel à travers une simulation de dynamique moléculaire. (2.16b) Phase de chauffage du recuit simulé : le système sort du puits de potentiel grâce à l'augmentation de la température. (2.16c) Phase de refroidissement du recuit simulé : le système va être piégé dans un autre puits de potentiel. (2.16d) Nous pouvons ensuite refaire une MD classique afin d'explorer de nouvelles conformations.

En simulation, le recuit simulé est utilisé pour explorer l'espace conformationnel. En effet, dans une MD classique, les fluctuations cinétiques ne sont pas assez prononcées pour permettre au système de sortir d'un puits énergétique (minimum local, ou global) (Fig. 2.16a). Nous avons vu dans l'équation 1.5 que la température est directement liée à l'énergie cinétique, donc en augmentant la température d'un système, nous allons apporter de l'énergie supplémentaire afin de permettre au système de pouvoir s'échapper d'un puits de potentiel pour en explorer d'autres (Fig. 2.16b). Une phase de refroidissement lente et contrôlée est ensuite appliquée pour que le système descende progressivement et reste bloqué dans une conformation correspondant à un minimum local d'énergie (cf. Fig. 2.16c).

Généralement, un protocole de recuit simulé est composé de nombreux cycles de chauffage/refroidissement, la température maximale dépendant du système et de la problématique étudiée. Dans le cas où la température est extrême, il convient d'adapter la simulation au logiciel et au champ de force utilisé afin d'éviter des phénomènes de fusion ou d'explosion. Il est par exemple possible de rajouter des forces de restrictions sur les atomes. Le recuit simulé est très utilisé dans la RMN, permettant ainsi l'obtention de structures à l'aide de contraintes d'angles et de distances issues des données RMN.

1.3.3 Dynamique Moléculaire Dirigée (TMD)

À ce jour, le temps simulé avec une dynamique moléculaire est encore très restreint. Cependant, certains événements biologiques, comme un changement conformationnel ou l'entrée d'un ligand dans une protéine, peuvent prendre plusieurs millisecondes, secondes ou plus. Ces temps sont très rarement accessibles, surtout pour des systèmes tout atomes, même avec les moyens de calcul actuels.

La dynamique moléculaire dirigée (TMD) va permettre de rajouter un terme énergétique supplémentaire dans la fonction d'énergie (eq. 1.4) basé sur la différence entre une structure de départ et une structure d'arrivée. Cette différence est appelée RMSD (*Root Mean Square Deviation*) et correspond à la racine carrée de la moyenne des carrés des distances entre paires d'atomes de deux structures pondérés par la masse de chaque atome.

$$RMSD = \sqrt{\frac{\sum_{i=1}^N m_i (r_i - r_i^0)^2}{\sum_{i=1}^N m_i}} \quad (1.6)$$

Avec r_i les coordonnées de l'atome i de la structure comparée et r_i^0 celui de la référence. m_i étant la masse de l'atome i et N le nombre d'atomes comparés.

L'ajout de ce terme énergétique aura pour effet de biaiser le système et de le diriger vers une conformation cible. Bien entendu, il faut que la cible ait le même nombre d'atomes et que la topologie soit identique avec la structure de

départ. Nous fixons un seuil de RMSD au système (par exemple, on voudrait que la différence moyenne entre notre structure de départ et notre structure cible soit de 2 Å), une force est exercée sur tous les atomes sélectionnés afin de provoquer les changements conformationnels nécessaires pour amener le système de départ à 2 Å du système d'arrivée.

La force exercée sur chaque atome est donnée par le potentiel :

$$E_{tmd_i} = \frac{1}{2}kN[RMSD_i - RMSD_i^*]^2 \quad (1.7)$$

$RMSD_i^*$ est la différence (distance) entre les coordonnées de l'atome i de départ et l'atome i d'arrivée. N est le nombre d'atomes sélectionnés et k une constante harmonique fixée par l'utilisateur. Ainsi, l'utilisateur peut faire varier le paramètre k afin d'augmenter ou de diminuer la force appliquée sur les atomes et ainsi déterminer une force minimale nécessaire pour passer d'une conformation A à une conformation B.

Il est important de noter qu'une seule topologie est utilisée pour les deux systèmes. Le système de la structure initiale et la structure cible doivent donc être identiques : même nombre d'atomes dans la protéine, même nombre d'ions et même nombre de molécules d'eau.

1.3.4 Ajustement par dynamique moléculaire flexible

La méthode d'ajustement par dynamique moléculaire flexible [133] ([MDFF](#)) est une méthode utilisée pour ajuster des structures atomiques dans des cartes de densité électronique. Cette méthode consiste à ajouter des forces extérieures, proportionnelles au gradient de la carte de densité dans la simulation de MD d'une structure atomique. Elle peut s'utiliser par exemple pour améliorer localement certaines régions de structure cristallographique [134], ou bien réorganiser des domaines protéiques sur une carte de cryo-microscopie électronique (cryo-EM) comme présenté dans la figure 2.17.

L'idée est d'intégrer la carte de densité électronique comme un potentiel qui va faire correspondre les endroits présentant la densité la plus forte (pic de

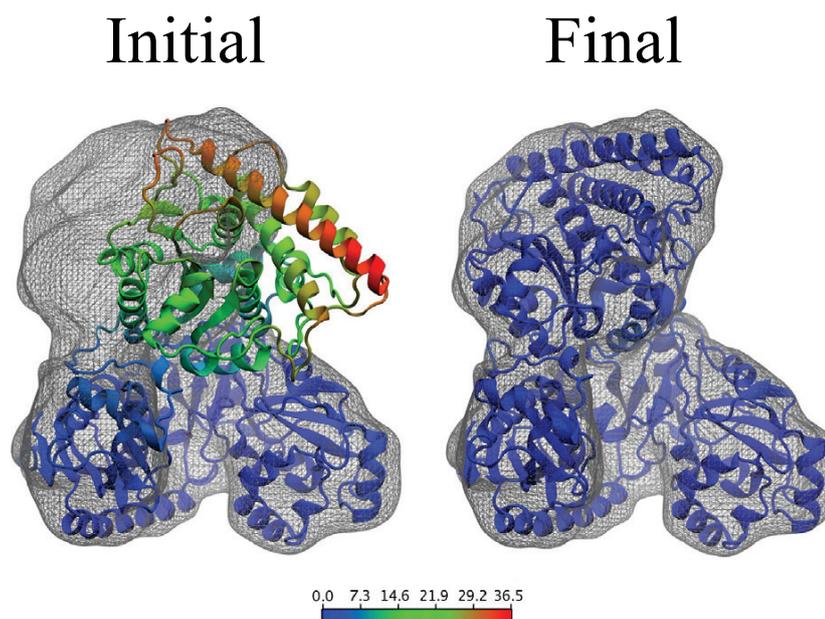


Figure 2.17: Exemple d'utilisation de la méthode MDFF sur une ACS/CODH (*Acethyl-CoA synthase/carbon monoxide dehydrogenase.*)[135]. La structure initiale et finale sont représentées en mode cartoon avec une coloration dépendante du RMSD entre la structure initiale et finale. Bleu : RMSD faible, rouge : RMSD élevé. La structure finale correspond à un ajustement dans une carte de densité simulé à 10 Å à partir de la structure cible. Tiré de Trabuco *et al.* (2008)[133].

densité) à des minimums énergétiques. Ainsi, les atomes dans la structure simulée seront soumis à des forces proportionnelles au gradient de la carte de densité.

MDFF peut également s'utiliser avec des restrictions de symétrie, ce qui est très utile pour les systèmes biologiques symétriques. L'utilisateur doit alors donner une structure parfaitement symétrique avec la même topologie. Durant la simulation de MDFF, la structure simulée peut dévier de sa symétrie à cause des fluctuations thermiques. Dans ce cas, un potentiel harmonique est rajouté et va minimiser le RMSD entre la structure simulée et la structure parfaitement symétrique donnée par l'utilisateur. Ce principe est similaire à la TMD (voir [Dynamique Moléculaire Dirigée \(TMD\)](#) page 56).

1.3.5 L'analyse par modes normaux

Comme précisé dans la partie [Dynamique Moléculaire Dirigée \(TMD\)](#), certains mouvements ne sont pas observables avec une simulation de dynamique

moléculaire dirigée. Cependant, nous n'avons pas toujours la possibilité d'avoir une structure cible pour faire une dynamique moléculaire dirigée ; soit parce qu'il n'existe pas d'autre conformation disponible, soit parce que nous ne savons pas quelle autre conformation une protéine pourrait adopter.

L'analyse par modes normaux est une méthode qui se base sur les coordonnées cartésiennes des atomes d'une protéine afin d'en extraire les mouvements de basse fréquence à partir d'une approximation harmonique.

1.4 Paramètres utilisés dans nos simulations

1.4.1 Simulations de dynamique moléculaire classiques

Les détails de l'obtention et de la préparation de chaque structure sont précisés dans la [Dynamique Moléculaire Dirigée \(TMD\)](#) page 56. Sauf indication contraire, les MD classiques ont été faites avec les mêmes paramètres et le logiciel de simulation GROMACS version 5.0 [129].

Les structures ont été neutralisées avec des ions Na^+ et solvatées avec le modèle d'eau TIP3P [136] à l'intérieur d'une boîte périodique rectangulaire avec au minimum 1.3 nm entre la structure et les bords de la boîte. Les structures obtenues après l'étape de minimisation d'énergie ont été utilisées pour initier les MD. Un pas d'intégration de 2 fs a été utilisé pendant les calculs, la liste des atomes voisins mis à jour tous les 25^{èmes} pas. Les interactions non liées de courte portée (Coulombiennes et Van der Waals avec un potentiel de Lennard-Jones) ont été calculées avec un rayon *cut-off* de 1 nm. Les interactions de plus grande portée ont été calculées avec la méthode PME [137] avec un maillage de 0.16 nm.

Afin d'éviter que les liaisons ne s'étirent trop, les longueurs de liaisons peuvent être figées grâce à un algorithme nommé LINCS. Sa version parallélisable (P-LINCS [138]) a été utilisée sur les liaisons covalentes. De la même manière,

l'algorithme SETTLE [139] qui a été utilisé, s'applique aux molécules d'eau en les considérant comme des corps rigides.

Durant les étapes d'équilibration, il est important de rajouter des restrictions de position sur les atomes afin d'éviter qu'ils partent à vive allure pendant les phases de chauffage. Il s'agit d'une force centrée sur l'atome fixée dans notre cas à $1000 \text{ kcal.mol}^{-1}$ induisant une pénalité énergétique lorsque l'atome dévie de sa position initiale.

Deux équilibrations de 500 ps en condition NVT ont été faites, l'une à 150°K suivie par une seconde à 300°K . Le solvant et la protéine ont été couplés séparément avec un bain de température externe à 300°K avec le thermostat velocity-rescaling [125] en utilisant une constante de temps de 0.1 ps. Nous avons ensuite réalisé une équilibration en condition NPT. La pression à l'intérieur de la boîte périodique est maintenue à une moyenne de 1 bar en utilisant le barostat isotropique Parrinello-Rahman [140] avec une constante de temps de 2 ps et une compressibilité de $4.5.10^{-5} \text{ bar}^{-1}$.

Après l'équilibration, les restrictions de position sont retirées afin de commencer la "vraie" simulation (production). Pendant la production, tous les autres paramètres de simulation sont identiques à ceux de l'équilibration. Nous rajoutons cependant une étape de pré-production afin laisser le temps au système de s'équilibrer. Suivant les résultats, cette région a été incluse ou non.

1.4.2 Simulations gros grains

Nos simulations gros grains ont toutes été réalisées par Jean-Charles Carvaillo sous ma supervision, avec Gromacs 5.1.4 [129] sur le supercalculateur OCCIGEN (CINES) en utilisant le champ de force MARTINI couplé à un réseau élastique ELNEDYN basé sur les structures secondaires. Le seuil de distance maximal du réseau est de 0.9 nm entre paires de grains et la constante harmonique est de $500 \text{ kJ.mol}^{-1}.\text{nm}^{-2}$. Afin de garder une flexibilité entre les domaines S et P, les liaisons élastiques entre ces deux domaines sont retirées grâce un algorithme appelé domELNEDIN [141]. domELNEDIN est un script tcl qui va parcourir un fichier de topologie gromacs et supprimer les liaisons

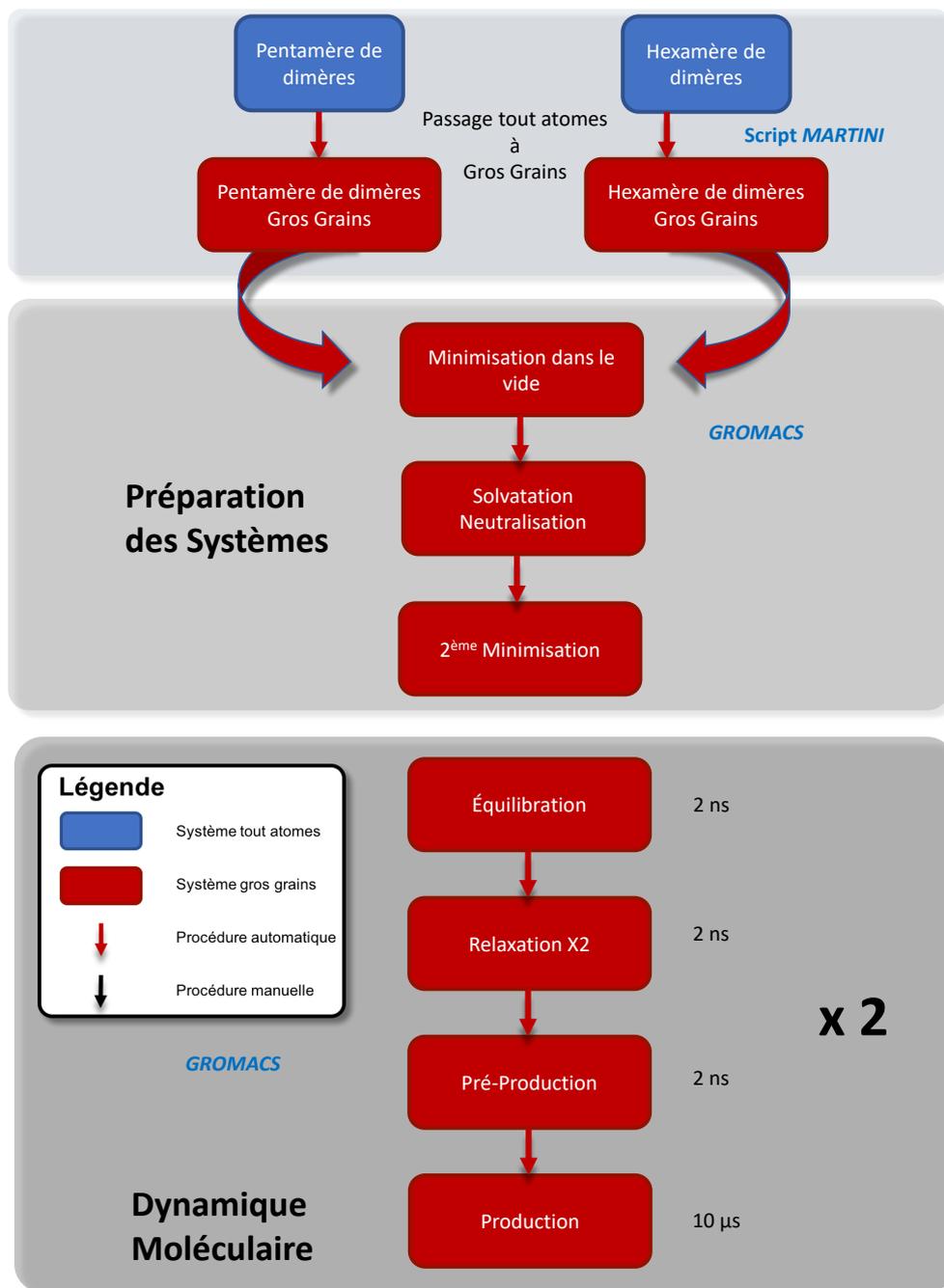


Figure 2.18: Schéma du protocole de simulation. Réalisé par Jean-Charles Carvaillo.

"HARMONICS" présentes entre deux régions d'une protéine spécifiés par l'utilisateur (Fig. 2.19).

Tel que décrit dans la figure 2.18, le système a d'abord été minimisé dans le vide avec 3 000 étapes de steepest-descent puis solvate avec des molécules d'eau standards gros grains MARTINI et neutralisé avec des ions Na⁺ dans une boîte périodique non cubique avec au minimum 1.3 nm entre n'importe

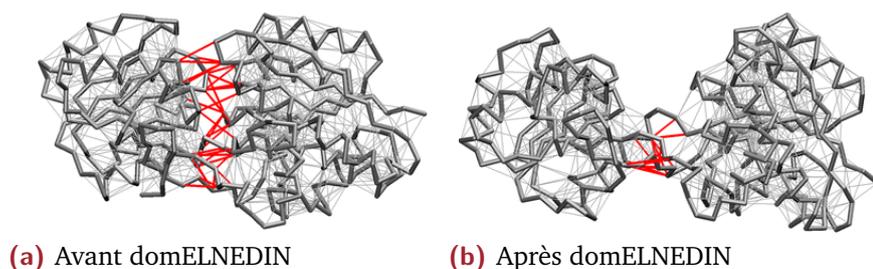


Figure 2.19: Exemple d'application de l'algorithme domELNEDIN sur une forme "ouverte" de la protéine de liaison à la leucine. Les liaisons grises fines représentent le réseau élastique maintenant la structure tertiaire dans les simulations gros grains. Les lignes rouges montrent le réseau élastique présent dans le modèle gros grains ELNEDIN (2.19a), mais supprimés dans le modèle gros grains domELNEDIN pour permettre une meilleure flexibilité des domaines protéines (2.19a)". Tiré de Siuda *et al.* (2013)[141]

quel grain du système et les bords de la boîte. Une deuxième minimisation de 20 000 étapes de *steepest-descent* est réalisée, suivie d'une étape d'équilibration de 2 ns où la température est amenée à 300°K avec l'algorithme de couplage *v-rescale* et la pression à 1 bar avec l'algorithme de couplage Berendsen avec un pas d'intégration de 10 fs. Deux étapes de relaxation de 2 ns chacune sont ensuite réalisées. L'algorithme de couplage de la pression est changé par Parinello-Rahman lors de la première relaxation, puis, lors de la seconde, le pas d'intégration est augmenté à 20 fs et les grains de la protéine cessent d'être restreints à leur position initiale.

Durant les simulations gros grains, les interactions non covalentes sont traitées par la méthode *switch* jusqu'à 1.2 nm pour les interactions de Coulomb et de 0.9 à 1.2 nm pour les interactions de Lennard-Jones, conformément au protocole standard de MARTINI [131, 142]. L'algorithme LINCIS [138] a été utilisé pour contraindre les longueurs de liaisons. Une constante diélectrique de 15 est choisie afin de reproduire les effets électrostatiques et les interactions de Van der Waals avec le modèle d'eau de MARTINI. La force de contrainte lors des phases de minimisation, équilibration et relaxation est de 1000 kJ.mol⁻¹.nm⁻² sur tous les atomes de la protéine.

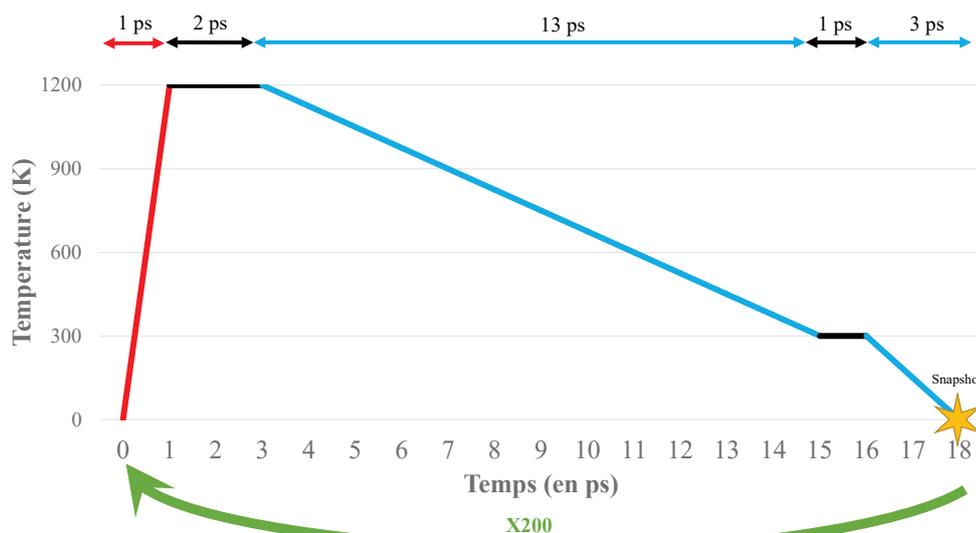


Figure 2.20: Description des cycles de recuit simulé utilisés dans notre protocole.

1.4.3 Recuit simulé

Amber 9 [143–145] a été utilisé pour le protocole de simulation de dynamique moléculaire par recuit simulé. Cette simulation a été réalisée en solvant implicite. La structure de départ était le modèle par homologie du GIII.2 minimisé. Les résidus 1 à 30 ont été gardés flexibles tandis que le reste de la protéine (résidus 31 à 522) a été fixée. Le système a été chauffé à 1200°K pendant la première ps puis stabilisé pendant 2 ps. Durant les 13 ps suivantes, le système a été refroidi progressivement jusqu'à 300°K puis le système a été stabilisé pendant 1 ps. Enfin, la température du système diminue jusqu'à 0°K pendant 3 ps et les coordonnées sont sauvegardées à la fin de chaque cycle. Cette boucle de 18 ps est réitérée 199 fois pour explorer l'espace conformation du bras N-terminal 2.20.

1.4.4 Dynamique moléculaire dirigée

Amber 14 [143–145] a été utilisé pour la TMD. Ces simulations ont été réalisées en solvant explicite avec de l'eau TIP3P [136] et neutralisés avec 28 ions Na⁺ pour la forme *WildType* du dimère et 32 pour la forme mutée du dimère (voir [Simulation de Dynamique moléculaire dirigée - GI.1 page 128](#)).

2.1 S.A.X.S.

Les rayons X sont très utilisés pour comprendre et obtenir des informations structurales sur la matière. Ainsi, suivant la diffusion ou la diffraction, nous pouvons obtenir de nombreuses informations comme l'organisation de la matière, sa composition ou bien encore la position des atomes la composant.

Durant cette thèse, en plus des méthodes *in silico* que j'ai utilisées, j'ai également réalisé quelques expériences de SAXS (Diffusion des rayons X aux Petits Angles ou *Small Angle X-ray Scattering*) afin de comprendre comment se comportait le dimère de VP1 en solution.

2.1.1 Principe

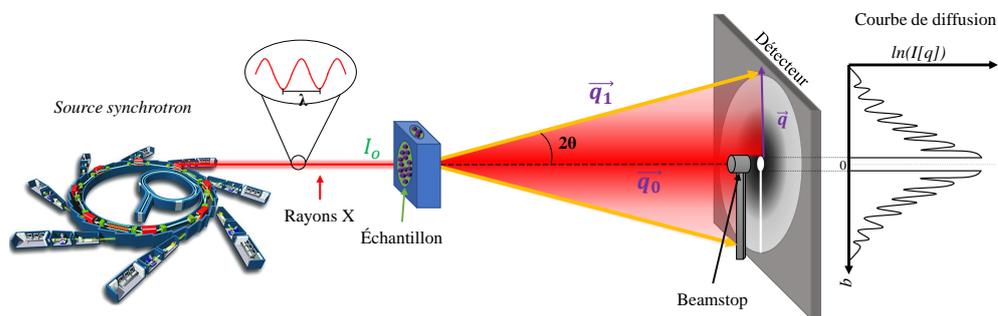


Figure 2.21: Description du SAXS. Une source synchrotron produit des rayons X qui iront ensuite traverser un échantillon. Une partie est diffusée par l'échantillon et enregistrée par un détecteur. La courbe de l'intensité diffusée en fonction de l'angle de diffusion 2θ (exprimé par le nombre d'onde q) est ainsi obtenue par intégration radiale. Le faisceau de rayons X direct sera arrêté par le *beam-stop*.

Le SAXS est une méthode très utilisée en biologie structurale. Contrairement à la cristallographie par rayons X, elle ne nécessite pas d'avoir de cristaux et est donc beaucoup plus facile à mettre en place. Seul un échantillon biologique en solution est nécessaire pour obtenir des données (à condition que l'échantillon soit de bonne qualité). Cependant, elle ne permet pas d'obtenir une structure atomique 3D, mais une enveloppe ou un facteur de forme de l'objet observé

(en fonction de l'espace dans lequel on se trouve) grâce à des analyses *ab initio*. L'espace réciproque permet d'obtenir le facteur de forme et l'espace réel, l'enveloppe. Les objets observés sont souvent de l'ordre de $1 \mu\text{m}$ à 1nm . C'est une technique dite de "basse résolution structurale", mais de haute précision quant à l'information de la taille et de la forme.

Le SAXS utilise des rayons X produits, la plupart du temps par une source synchrotron. Selon l'équation de Maxwell, "toute particule chargée se déplaçant de façon non uniforme émet un rayonnement électromagnétique". Ainsi, pour produire les rayons X nécessaires pour réaliser une expérience SAXS, des électrons sont accélérés dans un long tunnel circulaire et à chaque fois qu'ils sont courbés pour adapter leur trajectoire à ce tunnel, la perte d'énergie due à la courbure entraîne la production de rayons X.

Les rayons X sont ensuite filtrés pour qu'ils soient **monochromatiques*** et acheminés jusqu'à l'échantillon. Lorsque les rayons X traversent l'échantillon, ils diffusent dans toutes les directions de l'espace (diffusion *Thomson*). Cependant, une partie des rayons n'interagit pas avec la matière et sera interceptée par le *Beamstop* (l'intensité du rayonnement risquerait d'endommager le détecteur).

L'intensité mesurée sur le détecteur est ensuite intégrée radialement et exprimée en fonction du nombre d'onde q lié à l'angle de diffusion 2θ par la relation :

$$q = \frac{4\pi \sin \theta}{\lambda} \quad (2.1)$$

avec λ la longueur d'onde des rayons X. Ainsi, une courbe de diffusion représente le logarithme de l'intensité en fonction du nombre d'onde q : $\ln(I[q])$.

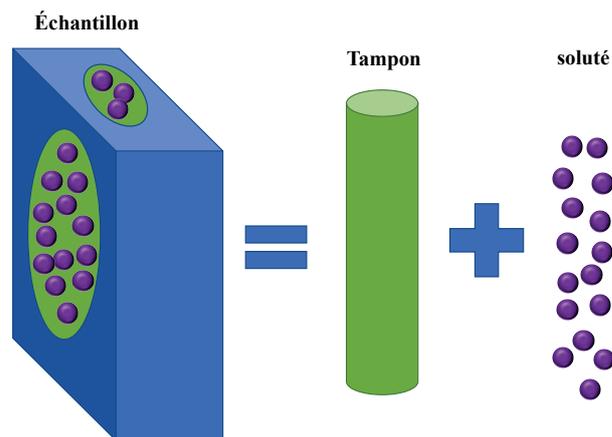


Figure 2.22: Composition d'un échantillon de SAXS.

2.1.2 Analyse

2.1.2.1 Traitement de la courbe de diffusion

La courbe de diffusion obtenue après une expérience SAXS n'est pas tout de suite exploitable. L'échantillon biologique est en solution dans un tampon, qui diffuse également les rayons X (cf. Fig. 2.22). Afin d'obtenir uniquement le spectre de notre soluté, nous devons d'abord prendre une mesure du tampon sans protéine. Nous obtenons ensuite un spectre que nous soustrayons à toutes acquisitions faites avec des protéines pour avoir uniquement le spectre de la protéine (cf. Fig. 2.23A).

Plusieurs informations peuvent être obtenues d'une courbe de diffusion des rayons X (cf. Fig. 2.24) :

- La forme générale de l'objet observé (sphère, tube, anneaux, haltère). En effet, la forme de la courbe est liée à la forme de l'objet observé (cf. Fig. 2.24).
- La région des petits q permet d'obtenir le rayon de giration de la particule.
- L'extrapolation de I_0 , l'intensité à $q = 0$. Avec cette valeur nous pouvons extraire le poids moléculaire, l'état d'oligomérisation et les possibles interactions intermoléculaires en solution.

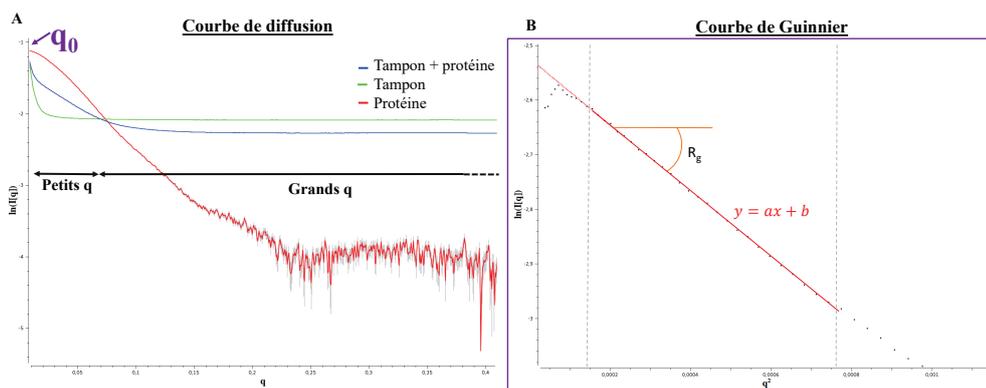


Figure 2.23: (A) Exemple de spectre de diffusion obtenu sur un échantillon de dimères de VP1. Le vecteur de diffusion q est calculé en fonction du logarithme de l'intensité. La courbe de diffusion de la protéine seule (rouge) est obtenue en soustrayant le tampon (vert) et le spectre du tampon et de la protéine. (B) Courbe selon la loi de Guinier (à l'échelle de q_0 , aux très petits q) pour une particule globulaire.

Rayon de giration

Le rayon de giration permet de mesurer la compacité d'une protéine et est défini par la formule suivante :

$$R_g = \sqrt{\frac{\sum_i m_i (r_i - r_0)^2}{\sum_i m_i}} \quad (2.2)$$

où r_i est la position de l'atome i , r_0 la position du centre de masse et m_i la masse de l'atome i . En SAXS, le rayon de giration R_g pour une particule globulaire peut être obtenu expérimentalement par la loi de Guinier :

$$I(q) = I(0) \exp\left(-\frac{R_g^2 q^2}{3}\right) \quad (2.3)$$

où $I(0)$ est calculé à partir de la courbe de tendance¹ d'équation $y = ax + b$ quand $x = 0$. Pour une protéine globulaire, nous pouvons simplifier le calcul du R_g par $R_g = \sqrt{-3a}$ (cf. Fig. 2.23B).

Évaluation du D_{max}

Le D_{max} est la distance maximale qui sépare les deux atomes les plus éloignés d'une protéine. Si la particule observée était une sphère parfaite, elle correspondrait à son diamètre. Afin d'estimer le D_{max} , la transformée de Fourier est

1. aux très petits q^2 (les premiers points sont ignorés)

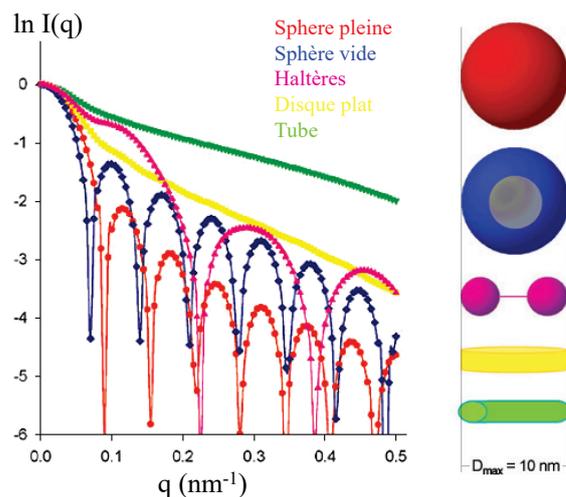


Figure 2.24: Relation entre la forme d'une particule observée en SAXS et la forme de sa courbe de diffusion. Tiré de Svergun *et al.* [146].

appliquée sur la courbe de diffusion pour obtenir la fonction de distribution de paires définie par :

$$P(r) = \int I(q) \exp(-2\pi iqr) dq \quad (2.4)$$

Cette fonction est une distribution des distances entre chaque paire d'atomes diffusant dans la particule. Elle est dépendante de la forme et de la compacité de la protéine. Le D_{max} est ainsi la valeur pour laquelle $P(r) = 0$ et est exprimé en *nm*. Cependant, cette valeur peut être surestimée ou sous-estimée en fonction de la gamme de la courbe de densité utilisée dans la transformée de Fourier.

2.1.2.2 Obtention d'une enveloppe

Grâce à la courbe de diffusion, il est possible de retrouver la forme de l'objet observé sous forme d'enveloppe. Nous avons utilisé le logiciel DAMMIN [147] afin de produire des enveloppes pouvant correspondre à nos données de SAXS. L'idée derrière ce programme est de trouver un objet tridimensionnel dont la courbe de diffusion théorique ressemble aux données expérimentales. Pour cela, le programme va utiliser des pseudo-atomes (ou *dummy atoms*) composés de deux états : atomes appartenant à la protéine, ou appartenant au solvant (cf. Fig. 2.25).

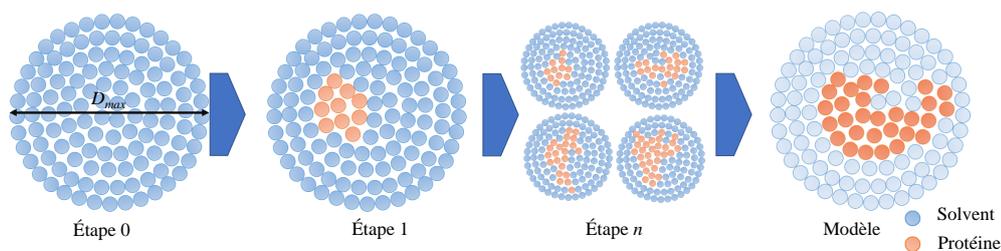


Figure 2.25: Représentation du fonctionnement de DAMMIN. Les sphères bleues représentent des pseudo-atomes appartenant au solvant, les sphères rouges ceux se trouvant dans la protéine. À chaque étape la courbe de diffusion théorique est calculée et comparée à la courbe expérimentale.

DAMMIN fonctionne à partir de la fonction de distribution de paires, que nous avons calculée à partir de la courbe expérimentale par le programme GNOM [148]. La sphère initiale (Fig. 2.25 - étape 0) est créée à partir du D_{max} expérimental. Puis, plusieurs étapes successives de recuit simulé vont être réalisées, permettant de changer le statut des pseudo-atomes afin de minimiser la fonction :

$$f(x) = \chi^2[I(q_i), I_{calc}(q_i)] + \alpha P(x) \quad (2.5)$$

où χ^2 représente la divergence entre la courbe calculée $I_{calc}(q_i)$ et la courbe expérimentale, $I(q_i)$ (voir partie [Calcul du \$\chi^2\$](#) page 70). Afin de ne pas s'éloigner d'un modèle cohérent au fur et à mesure des étapes, une fonction de pénalité $P(x)$ est rajoutée avec un poids α .

Une dernière chose est à prendre en considération lorsqu'on génère une enveloppe à partir de données SAXS : une structure tridimensionnelle (enveloppe en pseudo-atomes ou bien structure PDB) va avoir une seule courbe de diffusion théorique associée. En revanche, une courbe de diffusion expérimentale va correspondre à une multitude de modèles différents. C'est pourquoi il est important de générer plusieurs modèles afin d'obtenir un modèle moyen (moyennes de tous les atomes factices protéiques) et un modèle minimal (atomes factices protéiques en commun avec tous les modèles générés).

2.1.2.3 Calcul d'une courbe théorique

Afin de calculer une courbe de diffusion théorique à partir d'une structure tridimensionnelle, nous utilisons le programme CRY SOL[149]. La théorie physique du fonctionnement de ce programme est assez complexe, mais il faut savoir que pour réaliser une courbe de diffusion, CRY SOL va tenir compte du volume occupé par la protéine ainsi que d'une couche d'hydratation pour générer une courbe de diffusion.

Lorsque CRY SOL est utilisé avec une structure tridimensionnelle seule, une courbe de diffusion théorique sera seulement produite. Utilisé avec des données expérimentales de référence, CRY SOL va ajuster la courbe expérimentale avec la courbe théorique en minimisant le χ^2 2.7.

2.1.2.4 Calcul des résiduels

Les résiduels sont également un très bon indicateur d'un bon ou d'un mauvais ajustement. Ils se calculent pour chaque point de la courbe de diffusion et permettent de rendre compte des déviations locales quant à l'ajustement de la courbe théorique sur la courbe expérimentale.

$$Res_i = \frac{I(q_i) - I_{calc}(q_i)}{\sigma_i} \quad (2.6)$$

où σ_i l'écart-type (barre d'erreur) au i -ème point.

2.1.2.5 Calcul du χ^2

Le χ^2 est un terme calculé permettant de rendre compte de la différence globale entre une courbe de diffusion théorique d'un modèle (calculé par exemple avec CRY SOL) et une courbe expérimentale tel que :

$$\chi^2 = \frac{1}{N-1} \sum_i [Res_i]^2 \quad (2.7)$$

où N est le nombre de points sur la courbe expérimentale et Res_i le résiduel au i -ème point (cf. eq. 2.6).

Des valeurs faibles de χ^2 indiqueront que la correspondance entre deux courbes de diffusion, expérimentale et calculée, est bonne. Cependant il convient de faire attention au fait que le χ^2 est très dépendant de l'erreur expérimentale. Si l'erreur expérimentale est sous-estimée, la valeur de χ^2 sera surestimée même pour un modèle correct. Il est donc nécessaire de tester suffisamment de modèles et de les comparer entre eux. Ceci permet de diminuer l'impact d'un faible écart type.

2.1.2.6 Réalisation d'un modèle structural atomique

Si les données expérimentales sont suffisamment de bonne qualité et si des structures atomiques de domaines et/ou de conformations différentes sont disponibles, il est possible de créer un modèle structural correspondant. Plusieurs programmes existent modifiant la conformation d'une structure atomique existante. Tous les programmes cités dans cette section nécessitent une structure initiale (qui sera par la suite modifiée) ainsi que des régions flexibles et fixes. En effet, le coût calculatoire serait trop élevé pour tester toutes les combinaisons possibles, l'utilisateur spécifie donc des régions fixes (qui ne seront pas modifiées) et des régions flexibles (qui subiront des modifications afin de permettre la création de nouveaux modèles).

Dadimodo

DADIMODO [150] a été développé en association avec la ligne SWING au synchrotron SOLEIL pour permettre aux utilisateurs d'obtenir des modèles structuraux représentatifs de leurs données. Il s'appuie sur un [algorithme génétique*](#) :

1. À chaque génération, une population de nouveaux modèles structuraux est créée en modifiant l'angle dièdre aléatoirement dans la région flexible puis en minimisant ces modèles.
2. Chaque structure est ensuite évaluée grâce à une fonction de score interne. La qualité du *fit* de la courbe théorique de chaque structure, avec les données SAXS est calculée avec le programme CRY SOL.

3. La sous-population est ensuite sélectionnée avec la méthode du tournoi qui consiste à sélectionner un groupe de r candidats (tiré aléatoirement) parmi la population. Les meilleures structures enfants sont ensuite gardées et deviennent structures parents pour la génération suivante et l'algorithme revient au point 1.

Ces étapes décrites visuellement dans la figure 2.26, sont exécutées un certain nombre n de fois, fixé par l'utilisateur. L'étape de sélection devient de plus en plus stricte et à la fin DADIMODO renvoie une seule structure (représentative d'un groupe de structures en général presque identiques).

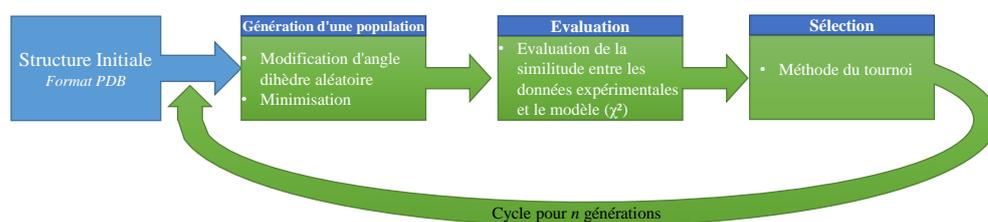


Figure 2.26: Représentation simplifiée du fonctionnement de DADIMODO.

Bunch

De nombreux autres logiciels existent comme BUNCH [151] de la suite ATSAS [152], qui est très utilisée la génération de modèles quasi-atomiques à partir de données SAXS pour la biologie. Bunch est bien adapté pour une seule protéine multi domaines séparés par des *linkers*. Il va générer plusieurs conformations grâce à une région flexible représentée par des pseudo-atomes. Un des avantages et défauts de ce logiciel est qu'il peut générer des homo-oligomères, mais uniquement symétriques. Ainsi, si l'oligomère n'est pas symétrique, les modèles ne seront pas de bonne qualité. Les complexes multi chaînes (hétéro-oligomères) sont difficilement exploitables par BUNCH également.

Coral

Une alternative à BUNCH permettant de s'affranchir de ces deux limitations est CORAL, toujours dans la suite ATSAS. Il est conçu pour la modélisation de complexes protéiques multi-domaines et peut être utilisé sur plusieurs courbes de diffusions expérimentales. Cependant, les *linkers* entre domaines ne sont pas générés *ab initio* par des pseudo-atomes comme Bunch, mais

depuis des bibliothèques de conformation de chaînes polypeptidiques. Coral peut dans ce cas être limité dans les conformations générées des modèles.

EOM

Un dernier logiciel a été utilisé dans l'analyse des données SAXS recueillies : EOM [153]. L'idée derrière EOM est de générer aléatoirement plusieurs milliers de modèles (physiquement possibles ou non) et de prendre des combinaisons de ces modèles pour générer une courbe de diffusion. Ceci permet de tester si une seule ou plusieurs populations sont présentes dans l'échantillon. Si la correspondance des courbes de diffusion est bien meilleur avec plusieurs modèles qui ont une forte différence entre eux, cela indique qu'il y a plusieurs conformations de la protéine en solution.

De même que pour DADIMODO, BUNCH et CORAL, EOM s'utilise en spécifiant des régions rigides et flexibles. Il peut s'utiliser avec ou sans contraintes de symétrie. Les deux ont été utilisés dans cette étude afin de pouvoir déterminer l'impact de l'asymétrie dans la modélisation de nos modèles.

2.1.2.7 Nos conditions expérimentales

Production de VLP

La production de VLP a été réalisée par Anna D. Koromyslova dans le groupe de Grant S. Hansmann à l'Université de Heidelberg. Le gène de la capsid du norovirus GI.1 (Virus de Norwalk) a été cloné dans un système d'expression en baculovirus précédemment décrit [154].

Acquisition des données et analyses

Après dialyse dans un tampon approprié (pour le tampon de dissociation, CHES 0.01 M, pH 9), les données ont été obtenues sur la ligne SWING du Synchrotron Soleil (Saint-Aubin, France) et traitées grâce à leur logiciel FOX-TROT. Des dilutions en séries de la protéine ont été mesurées (concentration initiale, dilution 2 puis dilution 4). Le tampon de dialyse a été mesuré avant et après chaque dilution. Toutes les mesures de tampon sont superposables et ont été moyennées et soustraites des mesures de solutions de protéine. Les analyses suivantes ont été faites en utilisant divers outils de la suite

ATSAS version 2.8.3 [152] comme indiqué. Trois dilutions ont été analysées dans PRIMUS [155] et trouvées superposables après mise à l'échelle. Nous avons donc analysé les données issues de l'échantillon non dilué. Le rayon de giration expérimental et le Dmax ont été calculés à partir de la fonction de distribution de paires $p(r)$ avec GNOM [148] à partir de la routine AutoRg dans PRIMUS. Nous avons également généré une enveloppe dans l'espace réel en utilisant un script Python qui rationalise et standardise les procédures classiques de moyenne de multi enveloppes. 30 enveloppes ont été générées avec DAMMIN [147] et alignées avec SUPCOMB [156]. Nous avons finalement l'enveloppe moyenne ainsi que l'enveloppe correspondant à l'enveloppe avec tous les points en commun entre toutes les enveloppes.

Modélisation corps rigide

Nous avons utilisé EOM [153] pour la modélisation avec plusieurs conformations et DADIMODO [150] pour la modélisation avec une seule conformation. Pour permettre d'avoir une flexibilité suffisante, nous avons choisi une version légèrement plus grande des régions terminales et de la région interstitielle. Ainsi, dans les deux cas, les trois corps rigides ont été définis comme étant les résidus 232 à 520 pour les chaînes A et B (dimère de P), les résidus 29 à 217 pour la chaîne A et 29 à 217 pour la chaîne B. Pour DADIMODO, qui utilise un modèle complet tout atomes, les régions N-terminales et C-terminales ont été ajoutées au dimère A-B cristallographique GI.1 avec MODELLER9.14. L'algorithme génétique de DADIMODO a été utilisé avec les paramètres suivants : à chaque pas, 20 enfants sont générés en changeant un angle dièdre aléatoirement dans la chaîne principale des acides aminés situés dans la région flexible avec un angle maximum de 30°. Les structures obtenues sont ensuite minimisées. Une population de 10 conformations est ensuite sélectionnée selon leur concordance aux données SAXS calculée avec CRY SOL [149], en augmentant la pression de sélection sur un maximum de 300 générations.

Durant ma thèse, j'ai été amené à développer plusieurs programmes ainsi qu'un certain nombre de scripts d'analyse et de pipeline afin d'automatiser certaines procédures. Que ce soit pour permettre l'identification de conformations différentes dans une trajectoire de dynamique moléculaire, rendre compte visuellement de la distance entre 3 acides aminés ou encore comparer plusieurs dizaines de milliers de modèles structuraux à une courbe de diffusion SAXS, vous trouverez dans cette partie un descriptif de certains programmes que j'ai réalisés durant ma thèse.

3.1 TTclust

Les trajectoires de MD peuvent produire une très grande quantité de données (termes énergétiques, vitesses et coordonnées des atomes) et générer un grand nombre d'images du système simulé. Par exemple dans nos simulations, nous avons en moyenne pour un système de 260 000 atomes et une durée de 100 ns, une trajectoire entière d'environ 50 Go et 10000 images du système. Bien entendu la taille de la trajectoire peut être compressée et réduite en retirant les molécules d'eau, mais nous aurons toujours 10000 images à traiter. Le clustering de trajectoire de MD permet de différencier le nombre de structures potentiellement différentes dans une simulation. Cette méthode n'est pas nouvelle, elle est déjà implémentée dans un certain nombre de programmes de simulation :

- `GMX CLUSTER` pour Gromacs. Complètement paramétrable, mais difficile d'utilisation : afin de trouver les bons paramètres de clustering, il convient de bien se renseigner sur quels paramètres affecter. Cet outil est également uniquement compatible avec les trajectoires générées par gromacs ou au format PDB.

- CPPTRAJ [157] pour AMBER. Également très modulable, mais difficile d'utilisation. Cependant CPPTRAJ est compatible avec plusieurs types de formats de trajectoire, incluant le format de Gromacs (bien que ce ne soit que depuis la dernière version d'AMBERTOOLS (2017) [158]).
- ou encore un plugin très facile d'utilisation intégré au logiciel de simulation Chimera [159]. Cependant, la lecture de trajectoire (en particulier provenant de Gromacs) dans Chimera est très laborieuse rendant l'utilisation très lente. De plus l'algorithme utilisé est NMRclust [160] qui est un algorithme de clusterisation hiérarchique automatisé et qui dans son implémentation actuelle ne permet pas de changer la valeur-seuil de clustering menant à un certain nombre de petits clusters composés de deux ou trois structures.

Dans notre équipe, nous réalisons la majorité des MD avec Gromacs et Amber, ainsi que NAMD dans de rares occasions. Afin de pouvoir clusteriser des trajectoires issue de différents programmes, facilement, avec la même méthode, j'ai donc réalisé un programme de clustering, nommé TTCLUST (pour *Trusty Trajectory Clustering*) combinant des modules Python pour la lecture de la trajectoire de dynamique moléculaire (MDTRAJ[161]) et pour le clustering hiérarchique (SCIPY [162]). Une note décrivant ce programme a été rédigée et soumise.

3.1.1 Fonctionnement

TTCLUST est écrit en Python et compatible avec les versions 2.7+ et 3.4+. Le code source est disponible sur github : <https://www.github.com/tubiana/ttclust>. Le fonctionnement détaillé est décrit dans la note associée (cf. partie 2.1) et la figure 2.27 décrit de manière simplifiée les cinq grandes étapes du programme :

1. L'utilisation du module MDTRAJ permet une compatibilité avec la plupart des formats de trajectoire de MD (*trr*, *xtc*, *dcd*, *crd*...) à condition d'avoir le bon fichier de topologie associé. Le format PDB permet de garder certaines informations comme le numéro de chaîne.

2. Une matrice de distances euclidiennes, représentant le RMSD entre chaque paire de frames, est calculée avec la fonction de calcul de RMSD de MDTRAJ. Cette matrice peut être calculée uniquement sur une certaine partie du système grâce au langage de sélection de MDTRAJ.
3. À partir de la matrice de distances euclidiennes, une matrice de distances hiérarchiques est calculée avec le module Scipy. Cette matrice représente une distance (en unité arbitraire) entre chaque paire de frames.
4. Le *clustering* s'effectue sur la base de la distance hiérarchique. Les clusters sont créés grâce à un seuil de distance. Tous les frames ayant une distance hiérarchique inférieure à la valeur seuil seront dans le même cluster. Cette valeur seuil peut être choisie par deux moyens : (a) en donnant explicitement le nombre N de clusters désiré et l'algorithme *maxclust* va sélectionner le cutoff pour avoir N clusters, (b) soit un dendrogramme s'affiche (cf. Fig. 2.28, étape 4) afin de permettre à l'utilisateur de choisir un cutoff adapté à son système.
5. Des graphiques sont ensuite générés afin de mieux analyser les clusters (proximité de certains clusters, tailles, largeur (ou *spread*)¹, ainsi que le changement de conformation s'opérant tout au long de la dynamique.

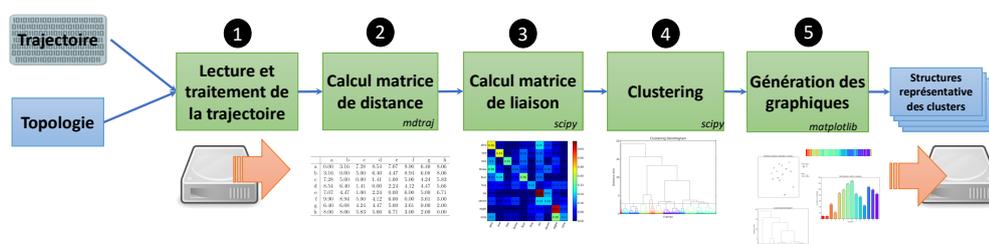


Figure 2.27: Représentation simplifiée du fonctionnement de TTCLUST.

1. Le *Spread* d'un cluster est le RMSD moyen entre toutes les structures appartenant au même cluster)

Structural bioinformatics

TTClust: A versatile graphical molecular simulation trajectory clustering program

Thibault Tubiana*¹, Yves Boulard¹ and Stephane Bressanelli*¹

¹Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris Sud, Université Paris-Saclay, 91198, Gif sur Yvette cedex, France.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXXX; revised on XXXXXX; accepted on XXXXXX

Abstract

Motivation: It is extremely helpful to be able to partition the thousands of frames produced in molecular dynamics simulations into a limited number of most dissimilar conformations. While robust clustering algorithms are already available to do so, there is a distinct need for an easy-to-use clustering program with complete user control, taking as input a trajectory from any MD package and outputting an informative overview of results.

Results: We present *TTClust* (for Trusty Trajectory Clustering), a python program based on the MDTraj package that fills this need. Its salient features are: (i) Ease of use: a simple command-line call with trajectory and topology files as arguments is enough in most cases; (ii) User control: the user can easily adjust major parameters such as clustering level and selection of atoms on which to perform alignment and clustering; (iii) Intuitive display of results, with plots allowing at-a-glance analysis of the trajectory.

Availability: The *TTClust* open source python program is freely available at <https://github.com/tubiana/TTClust> under the Gnu General Public Licence (GPL)

Contact: tubiana.thibault@gmail.com

Supplementary information: Fig. S1 (second case study), Text S1 (logfile produce by *TTClust*)

1 Introduction

Molecular dynamics simulations (MD) can produce a very large amount of data and routinely yield several thousand snapshots of large biomolecular systems. However differences between these snapshots may be minimal, especially in a short timescale. It is almost always desirable to ascertain which differences are only fluctuations of the system and which are different conformations with regard to the question at hand. Hierarchical clustering methods can be applied to group similar conformations and thus allow partitioning of the studied system into its different states during MD.

Only a few ready-to-use clustering tools are available. The USCF Chimera Trajectory Cluster tool (Pettersen *et al.*, 2004) can read many MD formats, but the user cannot control the clustering degree, which can lead to the creation of very small clusters. The user has more control with the Gromacs tool (Abraham *et al.*, 2015), but it is compatible only with the Gromacs or PDB formats.

We describe a new user friendly, fast and informative clustering tool derived from a methodology described in

the MDTraj package (McGibbon *et al.*, 2015). It directly takes as input most of the simulation data formats and has capabilities for adjusting the clustering degree and quickly visualising key properties of clusters, such as their sizes, spreads, separation and timeline of appearance.

2 Description

TTClust is a command-line python program (compatible with Python 2.7 or 3.4) and it can be used in the three major operating systems (Linux/Mac/Windows).

Thanks to the MDTraj package (McGibbon *et al.*, 2015) *TTClust* is able to deal with the trajectories of major MD programs like Gromacs (Abraham *et al.*, 2015), AMBER (Pearlman *et al.*, 1995), CHARMM (Brooks *et al.*, 2009), NAMD (Phillips *et al.*, 2005) and any trajectories in the PDB format. The trajectory is first aligned on the protein backbone (default) or any user-defined part. The RMSD between all pairs of frames is then calculated and used to generate a linkage matrix by the hierarchical cluster linkage function of the *SciPy* package (Jones *et al.*, 2001), by default with the ward algorithm (Ward, 1963) that minimises the variance within clusters.

T. Tubiana et al.

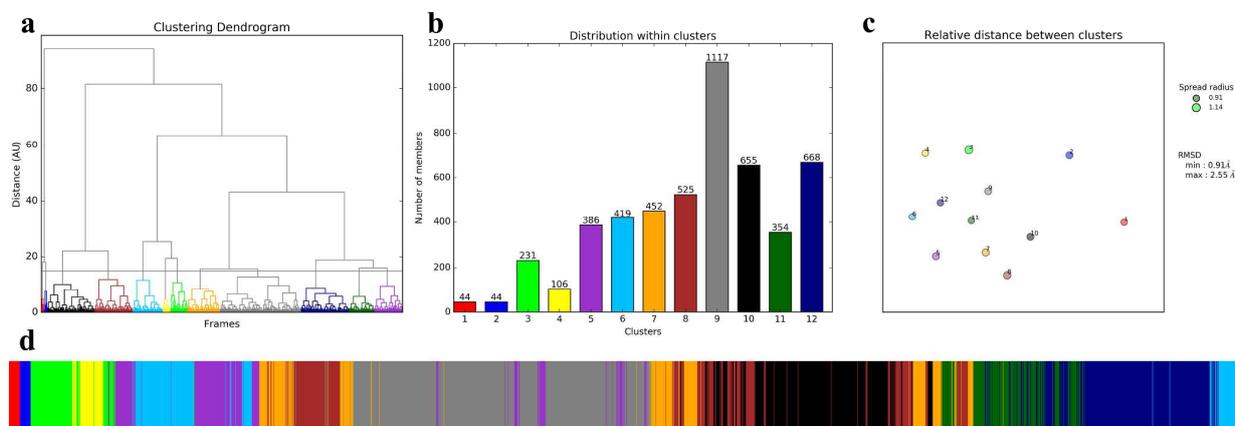


Fig. 1. *TTClust* applied to an all-atom 100-ns MD simulation. (a) Dendrogram of the RMSD-based linkage matrix between frames. The horizontal line is the clusterisation level chosen by point and click (here for 12 clusters). (b) Barplot numbers of frames within clusters. (c) 2D projection plot of the relative distances between clusters based on the RMSD between representative frames. The minimum and maximum distances are indicated, as well as the minimum and maximum spreads (average RMSD within clusters). (d) Timeline barplot along the simulation. The cluster color code is the same in dendrogram, histogram, 2D plot and timeline.

A dendrogram is then plotted with the Matplotlib package (Hunter, 2007) to have a graphical representation of the linkage matrix, see relations between all frames and select an appropriate cutoff value (Fig. 1a). Alternately the number of clusters can be defined directly on the command line, and in this case the *scipy* *maxclust* algorithm is used on the linkage matrix to segregate all frames at best.

A histogram of the cluster sizes in order of appearance in the simulation (Fig. 1b) and a timeline barplot of the trajectory (Fig. 1d) are generated. For all clusters, the spread is calculated and a representative frame (frame with the lowest average RMSD to all other frames in the cluster) is output as aligned PDB files for ready visualisation. Distances between clusters are computed as RMSD between pairs of representative frames. These two crucial pieces of information are displayed in a 2D projection graph (Fig. 1c) obtained with the multidimensional scaling algorithm (Kruskal, 1964) within the *scikit-learn* Python module (Pedregosa et al., 2011). All cluster data (members, representative frame, spread, RMSD matrix) are written into a logfile with all parameters for the run (cut-off, algorithm, input files, program version).

3 Case study

We applied this tool with default parameters on two MD simulations, one all-atom from PDB ID 4AEP (Scrima et al., 2012) (Fig. 1) and the other a large coarse grained system (Fig. S1). These case studies illustrate how major properties of the simulations are quickly highlighted thanks to *TTClust's* graphical summaries. Thus in the first case it can be seen at once that the first 4 clusters appear only at the beginning (Fig. 1d), are most sparsely populated (Fig. 1b) and lead to a conformational space segregated from the initial cluster (Fig. 1c). Thus the first few nanoseconds can be readily interpreted as a relaxation period, after which the protein explores a restricted conformational space. Indeed the final cluster 12 is close to cluster 6

(Fig. 1c), explaining how the simulation can come back to cluster 6 at the end. The logfile (Text S1) gives a more precise and complete breakdown of this analysis.

Funding

This work was supported by a grant from Région Ile-de-France to SB including doctoral funding for TT.

Conflict of Interest: none declared.

References

- Abraham, M.J. et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
- Brooks, B.R. et al. (2009) CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, **30**, 1545–1614.
- Hunter, J.D. (2007) Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, **9**, 90–95.
- Jones, E. et al. (2001) SciPy: Open source scientific tools for Python.
- Kruskal, J.B. (1964) Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115–129.
- McGibbon, R.T. et al. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.*, **109**, 1528–1532.
- Pearlman, D.A. et al. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, **91**, 1–41.
- Pedregosa, F. et al. (2011) Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Petersen, E.F. et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
- Phillips, J.C. et al. (2005) Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
- Scrima, N. et al. (2012) Two Crucial Early Steps in RNA Synthesis by the Hepatitis C Virus Polymerase Involve a Dual Role of Residue 405. *J. Virol.*, **86**, 7107–7117.
- Ward, J.H., Jr. (1963) Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.*, **58**, 236.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

3.1.3 Étude d'un cas complémentaire à l'article

Dans l'étude de cas présenté dans l'article (cf. section 2.1), TTCLUST est utilisé sur une dynamique tout atomes de 100 ns (simulation de NS5B [163]). Les premières nanosecondes de simulation sont une période de relaxation (conformation jamais retrouvée dans la suite de la simulation) puis la protéine explore un espace conformationnel restreint (le cluster 12 est proche du cluster 6). Cet exemple est représentatif d'un système qui est tombé dans un minimum local puis l'explore.

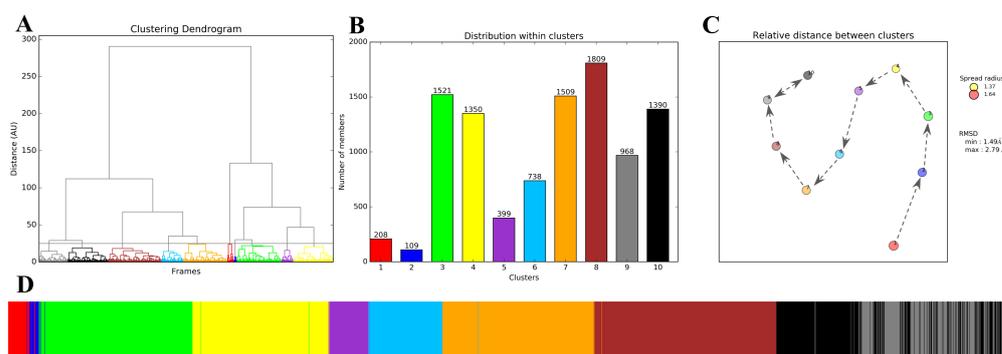


Figure 2.28: Clustering d'une trajectoire de MD de 10 μ s sur un système gros grains (pentamère de dimères). (A) Dendrogramme représentant graphiquement la distance hiérarchique entre les frames ainsi que les clusters grâce au code couleur. (B) histogramme de la taille de chaque cluster. (C) Projection 2D des distances entre chaque cluster, la taille du point étant liée à son *spread*. (D) Barplot des clusters tout au long de la simulation. Chaque cluster est représenté par une couleur qui est conservée entre tous les graphiques.

Nous pouvons comparer ce résultat à une deuxième étude de cas, celui d'une simulation gros grains d'un pentamère de dimères de protéines de capsid du norovirus pendant 10 μ s (cf. Fig 2.28). Le barplot temporel (cf. Fig. 2.28D) montre qu'il y a une direction tout au long de la simulation : chaque conformation observée n'est pas retrouvée dans la suite de la simulation, jusqu'à environ 8 μ s où on alterne entre deux conformations. Bien qu'il faille garder en perspective que la projection 2D des distances est basée sur un algorithme multiscalaire et que l'orientation du graphique est aléatoire, nous pouvons observer l'évolution de la dynamique d'un cluster à un autre jusqu'à l'alternance entre les clusters 9 et 10 sur la figure 2.28C. Si on regarde le dendrogramme (Fig. 2.28A), les clusters gris et noir sont très proches. Ceci est confirmé par la projection 2D des distances (Fig. 2.28C).

Ce type d'évolution directionnelle est représentatif d'un système qui va explorer l'espace conformationnel en tombant successivement dans des puits énergétiques et les explorant. Il est utile de préciser que dans le cas d'une simulation gros grains, la surface d'énergie potentielle est "lissée" ce qui permet une transition plus facile entre les puits d'énergie potentielle (cf. Fig. 2.29).

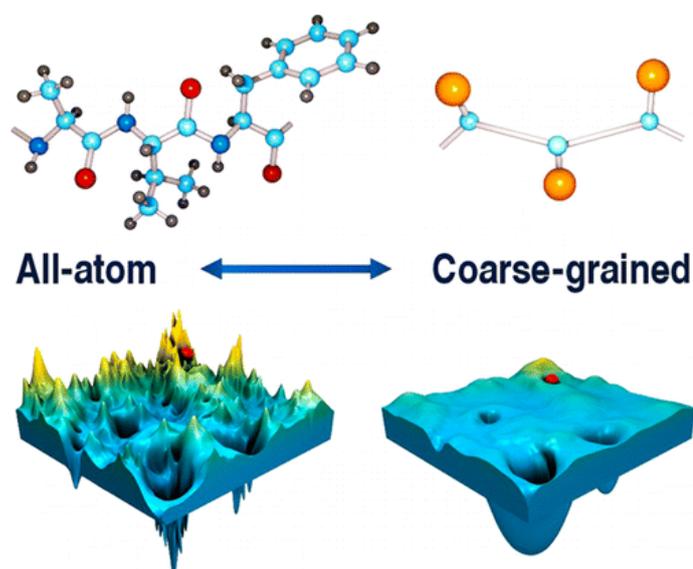


Figure 2.29: Illustration de la surface d'énergie potentielle en fonction d'un système tout atomes ou gros grains. Tiré de Kmiecik *et al.* (2016) [77]

Ainsi, TTCLUST permet de faire rapidement et facilement la différence entre deux types de MD :

- un système qui, après une relaxation, revisite des conformations déjà explorées durant la simulation
- un système avec une évolution directionnelle sans retourner dans des conformations déjà visitées.

Dans tous les cas, TTCLUST permet de lire des propriétés majeures des résultats du clustering grâce aux résumés graphiques.

3.2 Comparaison en masse de modèles structuraux à des données SAXS

Lorsque nous avons quelques modèles structuraux à comparer à des données SAXS, il convient d'utiliser CRY SOL à la main. Mais si nous sommes en présence de plusieurs dizaines, centaines ou milliers de modèles (comme ceux produits par EOM), il devient vite impossible d'exécuter manuellement CRY SOL. J'ai donc réalisé un programme, écrit en Python et compatible avec les versions 2.7+ et 3.4+ permettant de générer les courbes de diffusions théoriques d'un modèle structural et de le comparer à des données expérimentales grâce à CRY SOL.

CRY SOL est difficilement utilisable en ligne de commande. Il a été prévu pour un usage interactif. Ainsi pour chaque fichier PDB CRY SOL doit être exécuté deux fois : une première fois pour calculer la courbe de diffusion du modèle, une seconde fois pour fitter la courbe théorique avec la courbe expérimentale. Un fichier texte doit être généré à chaque exécution de Crysol contenant toutes les "réponses" aux questions que pose CRY SOL afin de générer les résultats (localisation du fichier PDB, rayons des atomes, nombre d'harmoniques...).

Les résiduels (voir [Calcul des résiduels](#) page 70) sont calculés et bien que CRY SOL calcule un χ^2 , si celui-ci est trop élevé (au-dessus de 100) il sera affiché par "*****" rendant ainsi certaines analyses comparatives impossibles. Un χ^2 est donc recalculé basé sur l'équation 2.7.

Nous avons également utilisé le programme FOXS [164] permettant d'avoir la valeur d'un χ (qui est ensuite mise au carré) même pour les modèles très divergents.

Finalement, les résultats entre FOXS et CRY SOL sont collectés et enregistrés dans un fichier csv pour de futures analyses.

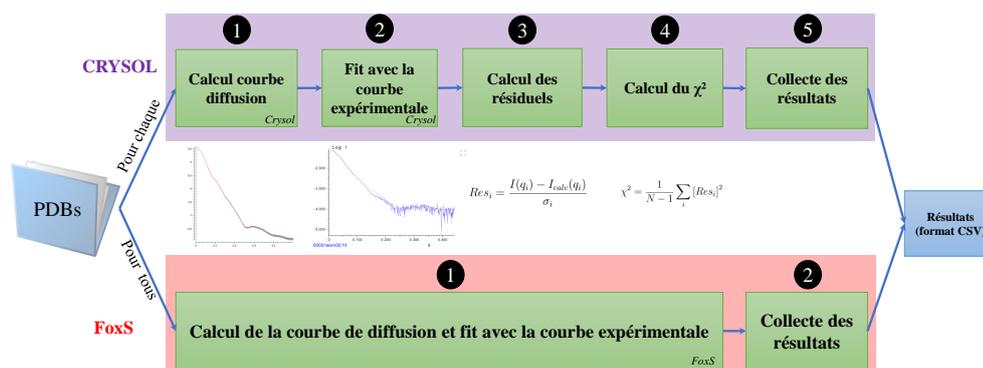


Figure 2.30: Représentation simplifiée du fonctionnement des scripts de comparaison pour le SAXS.

3.3 Comparaison d'angles entre deux domaines protéiques

Nous avons vu dans la partie sur VP1 (page 17) que les trois conformations de VP1 étaient dues principalement à un changement d'angle entre le domaine S et le domaine P (cf. Fig. 1.13). Ainsi nous avons voulu suivre la valeur de cet angle au cours de certaines de nos simulations. Le calcul d'un angle entre deux domaines protéiques n'est pas forcément trivial.

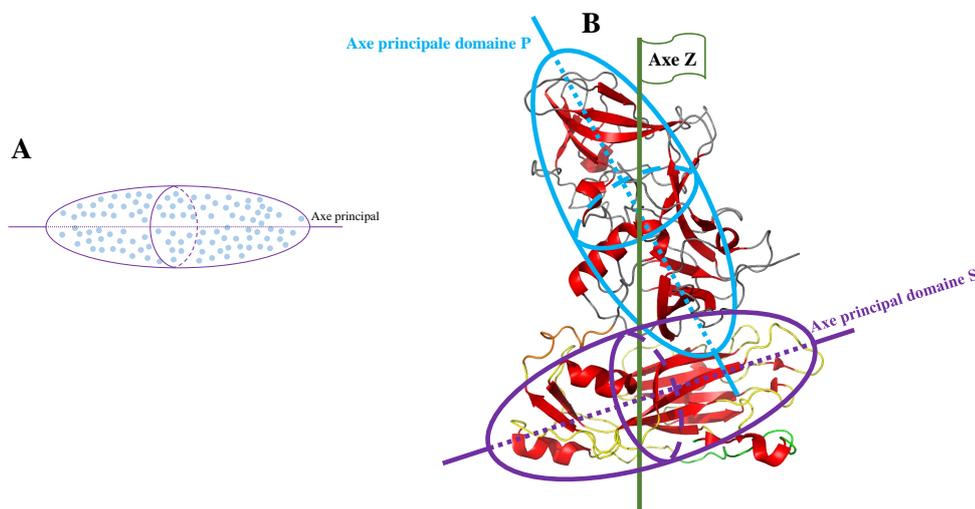


Figure 2.31: Illustration des axes principaux des domaines S et P. (A) Représentation d'un ellipsoïde englobant tous les points ainsi que l'axe principal passant par cet ellipsoïde. (B) Axes principaux des domaines S et P par rapport à l'axe Z. Le domaine S est en jaune, le domaine P en gris et les structures secondaires hélices et brins en rouge.

La stratégie que j'ai choisie est de calculer l'angle formé par l'axe principal du domaine et un axe directionnel (X,Y ou Z). Cette valeur sera suivie tout au long de la dynamique et comparée à la valeur référence qui est la valeur mesurée dans la structure. Afin de calculer l'axe principal d'un domaine, nous utilisons les coordonnées 3D de la protéine puis nous faisons une régression linéaire 3D à l'aide du module `SCIPY`. Afin de minimiser les fluctuations dues aux vibrations thermiques, nous pouvons nous restreindre au squelette de la protéine ainsi qu'aux régions ayant une structure secondaire hélice ou brin (cf. Fig. 2.31).

Enfin, la hauteur des domaines est également calculée avec pour référence la structure cristallographique : le centre de masse de chaque domaine est recalculé pour chaque frame puis comparé à une valeur référence. Cette valeur est calculée à partir d'une structure PDB donnée en plus au programme.

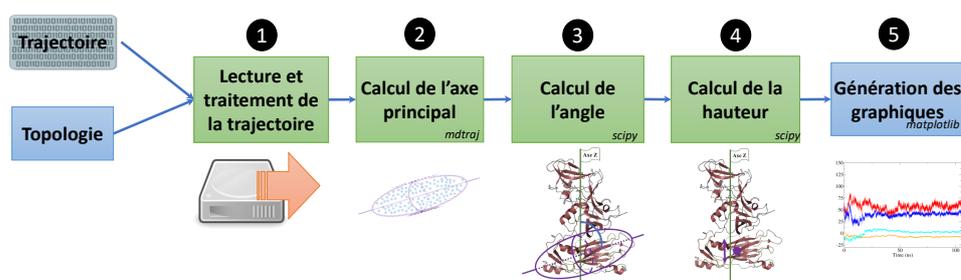


Figure 2.32: Représentation simplifiée du fonctionnement du script pour la comparaison d'angle et de hauteur des domaines.

3.4 Module Pymol : représentation de la distance entre 3 atomes par un triangle

Afin de comparer la forme d'un groupement de résidus acides durant une dynamique, j'ai développé un plugin PYMOL [165] permettant de représenter les distances entre trois résidus par un triangle.

Chaque sommet (ou *vertex*) du triangle est défini par un atome, ou bien le centre de masse de l'ensemble des atomes sélectionnés. Le triangle est un CGO (*Compiled Graphics Objects*, type d'objet graphique simple propre

à PYMOL) où chaque vertex doit être défini par un point de coordonnées [X,Y,Z] ainsi qu'une couleur). La couleur peut être fixée par l'utilisateur, sinon ce sera la couleur de l'atome qui sera choisie (cf. Fig. 2.33).

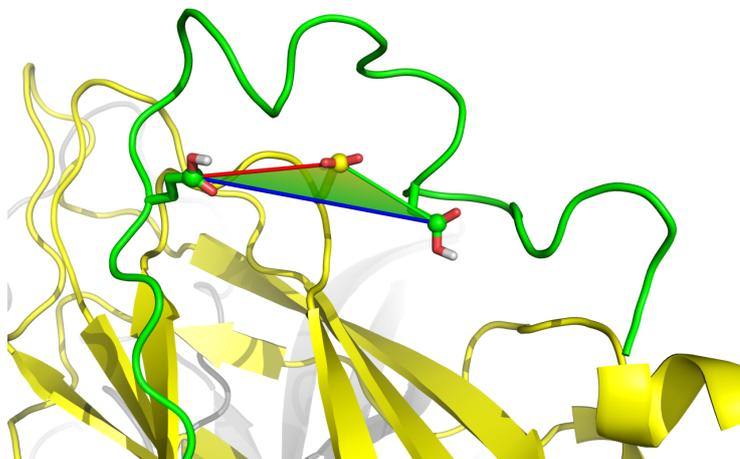


Figure 2.33: Exemple de la représentation "triangle" entre trois atomes. Image extrait d'une dynamique de 5000 instantanés. Le triangle est coloré avec un dégradé de couleur en fonction de la couleur de l'atome composant chaque sommet.

Afin que la lumière se reflète correctement sur la face du triangle, il est important de calculer la normale (axe perpendiculaire au plan du triangle). Ceci est calculé pour chaque frame de la trajectoire et le nouveau triangle est stocké dans un objet contenant tous les états précédents du triangle. Ainsi, une fois les triangles calculés, il sont lus en même temps que la trajectoire sans coût calculatoire supplémentaire.

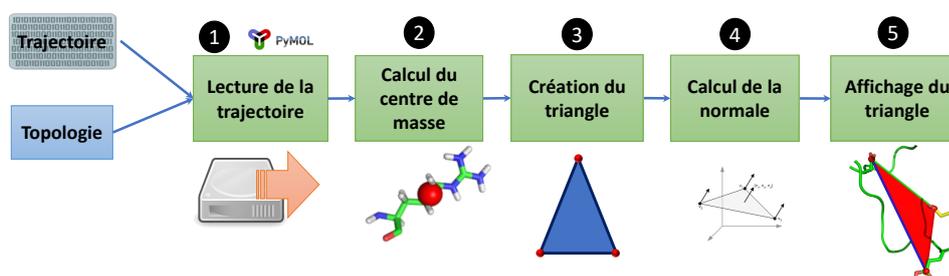


Figure 2.34: Représentation simplifiée du fonctionnement du plugin Pymol représentant la distance entre trois atomes tout au long de la trajectoire.

Dynamique et asymétrie du dimère de la protéine de capsid du Norovirus

” Interestingly, when he (ndlr : Albert Einstein) gave his Nobel Lecture, it was on relativity, even though he knew that he was supposed to talk about the photoelectric effect. Correspondingly, I traced the history of molecular dynamics simulations and their development in my lecture and did not emphasize the development of potential functions for simulations, the focus of the Chemistry Nobel Prize citation.

— **Martin Karplus**
Discours de prix Nobel 2013

Dans la plupart des petits virus icosaédriques quasiéquivalents de type T=3, un bras N-terminal sous le corps de la capsid (domaine S chez le norovirus) participe au mécanisme de formation de la quasiéquivalence [166].

La VLP du norovirus peut être désassemblée artificiellement à haut pH et faible salinité et réassemblée en abaissant le pH (VLP→dimères à pH9, dimères →VLP à pH6-8) [22]. Grâce à cette propriété, Stéphane Bressanelli et Guillaume Tresset en 2013 ont étudié l'assemblage et le désassemblage d'une VLP de norovirus analogue à la souche de Norwalk (GI.1), la souche NewBury2 (GIII.2) touchant les espèces bovines [167].

Durant cette étude, ils ont remarqué que dans la structure cristallographique du GI.1 [56], un groupe de résidus acides, conservés entre le GI.1 et le GIII.2, était présent sur le bras N-terminal et le domaine S autour de l'axe quasi-6 (cf. Fig. 3.1). Ces résidus ne sont pas neutralisés par d'autres résidus basiques à proximité. Ils peuvent donc être protonés en fonction du pH du milieu dans lequel la protéine se trouve, et ainsi participer à la relocalisation du bras N-terminal au sein du dimère ou de la capsid. Nous avons donc commencé à étudier le dimère de VP1 du GIII.2 afin de comprendre l'impact que pouvait avoir un changement de protonation sur la dynamique du bras N-terminal. Pour ce faire, nous avons utilisé des techniques de modélisation par homologie, recuit simulé et dynamique moléculaire classique.

Tresset *et al.* ont également montré dans la même étude que le désassemblage de cette capsid était similaire au GI.1, mais également que le dimère de VP1 du GIII.2 montrait des propriétés d'auto-assemblage étonnantes. Ils pouvaient former, en plus d'une capsid icosaédrique de type T=3, des particules beaucoup plus grandes (jusqu'à 76 nm dans des conditions expérimentales non physiologiques). En étudiant la dynamique d'assemblage de la VLP de GIII.2 en TR-SAXS, il ressort également que la taille du dimère est plus grande que le dimère cristallographique (GI.1)[75]. Nous avons donc réalisé des expériences en SAXS sur des VP1 de GI.1 (produit par Grant S. Hansman et Anna D. Koromyslova) afin de vérifier si cette propriété du dimère pouvait

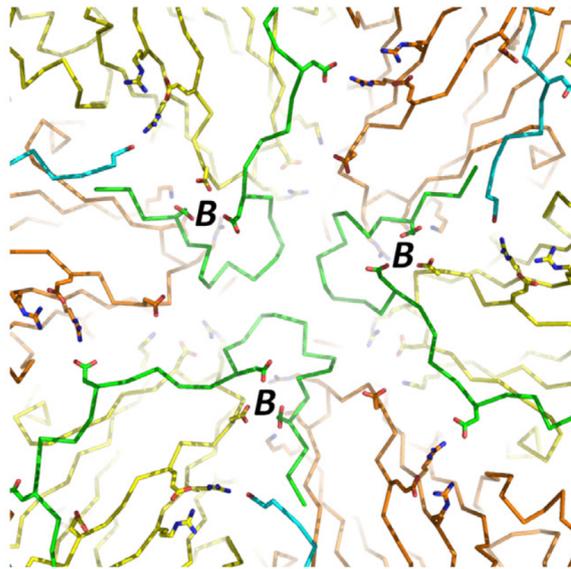


Figure 3.1: Cluster de résidus acides autour de l'axe quasi-6 de la VLP de GI.1. "La chaîne principale des domaines S de la chaîne B et C est visible. Les sous-unités C sont colorées en cyan (bras N-terminal) et orange. Les acides aminés acides (ASP et GLU) et basiques (LYS et ARG) sont représentés en mode sticks. Tous ces résidus sont neutralisés par des ponts aïlins sauf pour les acides aminés impliqués dans le cluster acide (marqué par "B" dans la sous-unité B." Tirée de Tresset *et al.* [167].

également être observée dans la souche de la structure cristallographique. Enfin, nous avons produit un modèle atomique du dimère de GI.1 correspondant aux données SAXS acquises sur la ligne SWING au synchrotron Soleil.

Les résultats obtenus ont permis de mettre en évidence la présence d'une asymétrie dans le dimère et ils ont fait l'objet d'une publication en juillet 2017 dans Plos One : "Dynamics and asymmetry in the dimer of the norovirus major capsid protein" [168]. Depuis que l'article a été soumis, nous avons également réalisé des travaux complémentaires : l'étude d'un monomère seul dans sa conformation C et une simulation de dynamique moléculaire dirigée (TMD) pour étudier le passage d'une conformation compacte à une conformation étendue et inversement. De plus, nous avons pu avoir une carte de cryo-EM de la VLP de GIII.2 grâce Julien Maufront et Jean Lepault. Nous avons donc réalisé un modèle atomique de la capsid du GIII.2 basé sur cette carte. Ces résultats sont présentés dans la partie complémentaire à l'article (voir [Complément à l'article](#) page 116).

Dynamics and asymmetry in the dimer of the norovirus major capsid protein

2.1 Article

RESEARCH ARTICLE

Dynamics and asymmetry in the dimer of the norovirus major capsid protein

Thibault Tubiana, Yves Boulard*, Stéphane Bressanelli*

Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris Sud, Université Paris-Saclay, Gif sur Yvette cedex, France

* yves.boulard@cea.fr (YB); stephane.bressanelli@i2bc.paris-saclay.fr (SB)



Abstract

Noroviruses are the major cause of non-bacterial acute gastroenteritis in humans and livestock worldwide, despite being physically among the simplest animal viruses. The icosahedral capsid encasing the norovirus RNA genome is made of 90 dimers of a single *ca* 60-kDa polypeptide chain, VP1, arranged with $T = 3$ icosahedral symmetry. Here we study the conformational dynamics of this main building block of the norovirus capsid. We use molecular modeling and all-atom molecular dynamics simulations of the VP1 dimer for two genogroups with 50% sequence identity. We focus on the two points of flexibility in VP1 known from the crystal structure of the genogroup I (GI, human) capsid and from subsequent cryo-electron microscopy work on the GII capsid (also human). First, with a homology model of the GIII (bovine) VP1 dimer subjected to simulated annealing then classical molecular dynamics simulations, we show that the N-terminal arm conformation seen in the GI crystal structure is also favored in GIII VP1 but depends on the protonation state of critical residues. Second, simulations of the GI dimer show that the VP1 spike domain will not keep the position found in the GII electron microscopy work. Our main finding is a consistent propensity of the VP1 dimer to assume prominently asymmetric conformations. In order to probe this result, we obtain new SAXS data on GI VP1 dimers. These data are not interpretable as a population of symmetric dimers, but readily modeled by a highly asymmetric dimer. We go on to discuss possible implications of spontaneously asymmetric conformations in the successive steps of norovirus capsid assembly. Our work brings new lights on the surprising conformational range encoded in the norovirus major capsid protein.

OPEN ACCESS

Citation: Tubiana T, Boulard Y, Bressanelli S (2017) Dynamics and asymmetry in the dimer of the norovirus major capsid protein. PLoS ONE 12 (7): e0182056. <https://doi.org/10.1371/journal.pone.0182056>

Editor: Heinrich Sticht, Friedrich-Alexander-Universität Erlangen-Nürnberg, GERMANY

Received: April 23, 2017

Accepted: July 11, 2017

Published: July 27, 2017

Copyright: © 2017 Tubiana et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by a grant from Région Ile-de-France to S.B. including predoctoral funding for T.T. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Introduction

Noroviruses are major causes of acute gastroenteritis worldwide [1]. They are extremely resilient and biodisponible, explaining in part their high infectivity by the fecal-oral route. The norovirus virion's major protein component comprises 180 copies of a single polypeptide chain (viral protein 1, or VP1) that together encase the *ca* 7500-base single-stranded RNA genome and a few copies of viral protein 2 (VP2). Upon production in the baculovirus system, VP1 self-assembles into empty virus-like particles (VLP) that are antigenically and structurally indistinguishable from infectious particles collected in infected individuals' stools [2]. This has

allowed early biochemical studies despite the difficulty in growing noroviruses in cell culture and is still the method of choice to study VP1. Indeed, the only atomic model of VP1 has been obtained for such a VLP with the crystal structure of the Norwalk virus capsid [3]. This work showed that in the particle the 180 copies of VP1 are arranged with classical $T = 3$ icosahedral symmetry [4]. There are thus 60 identical asymmetric units, each containing three VP1 molecules A, B and C with quasi-equivalent contacts to their neighbors. VP1 has a two-domain structure (Fig 1A), with an N-terminal shell domain (S) making up the spherical protein capsule and a C-terminal protruding domain (P) making up 90 dimeric spikes on the exterior surface of the shell. Indeed the large contact surface between pairs of P indicates that the capsid building block is a VP1 dimer. The segment connecting S to P and an arm N-terminal to S allow the small conformational adjustments that differentiate quasi-equivalent dimers A-B and C-C. Contacts between the two P domains provide most of the dimeric interaction surface but S domains also interact across the two-fold axis in both types of dimers, as remarked by Prasad et al. [3].

As with most RNA viruses there is a high diversity of noroviruses. They are classified in at least five genogroups, each genogroup harboring several genotypes [5]. Typically pairwise sequence identity between VP1 from distinct genogroups is about 50% (Fig 1B).

Norwalk virus, for which the VLP crystal structure is known, belongs to genotype 1 of genogroup I (GI.1, Fig 1C). Structural data for other genogroups includes cryo-electron microscopy (cryo-EM) for GV infectious virion [7] and GII VLP [8]. Those works showed a distinct conformation compared to the GI crystal structure of the 90 P dimers, that are raised off the shell and rotated, forming a cage-like structure around the S shell. Finally, our own work by time-resolved small-angle X-ray scattering (TR-SAXS) on a GIII VLP showed that after *in vitro* dissociation of the VLP into its constituting dimers, reassembly proceeds through long-lived, elongated intermediates made of 10–11 dimers [9]. Interestingly, we found then that the dissociated GIII VP1 dimer has a much larger radius of gyration than would be expected from the crystal structure of the GI dimers in the VLP. In the present work, we further explore the conformations and dynamics of the norovirus VP1 dimer by molecular modeling and molecular dynamics simulations and with additional SAXS data on the GI dimer.

Materials and methods

Homology modeling

A homology model of NB2 (GIII.2) VP1 was built based on the Norwalk virus (GI.1, Fig 1B and 1C) capsid protein crystallographic structure (PDB: 1IHM)[3]. Chain B of 1IHM was chosen as a structural template for the target GIII.2 dimer sequence as it is the more complete, comprising residues 10–520 of the 530-residue GI.1 VP1. The Modeller 9.14 program [10,11] was used to generate 100 structural models of the GIII.2 monomer that were further evaluated based on their pseudo-energy ("DOPE") value [12] and the number of residues outlier on the Ramachandran plot [13] with the RAMPAGE webserver (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). We selected the structural model with lowest DOPE value (8 Ramachandran plot outliers) and generated the dimer by aligning two copies on both A and B chains of the GI.1 structure with Pymol [14]. The model was then neutralized with 34 Na⁺ ions, hydrated with TIP3P water molecules and minimized afterward with the steepest-descent method during 20000 steps while restraining the protein atoms and ions with harmonic potentials with a constant force of 1000 KJ/mol. We compared this GIII.2 dimer model to the GI.1 A-B dimer (both dimers with missing N- and C-termini modeled *ab initio*) by performing a 100 ns molecular dynamics simulation for both (see *Preparatory simulations for the N-terminal arms study* for details of the simulation protocols).

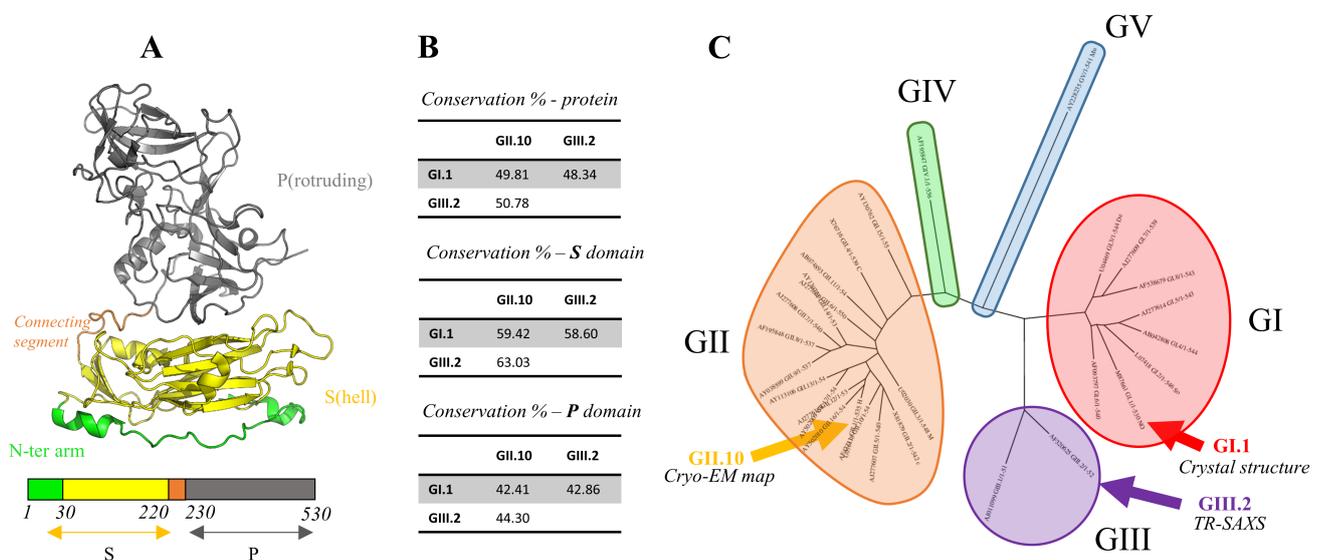


Fig 1. Crystal structure and sequence diversity of the norovirus VP1 capsid protein. (A) Structure of the norovirus major capsid protein (VP1). The N-terminal S domain is in yellow and the C-terminal P domain in grey. They are separated by a connecting segment of approximately 10 amino acids (orange). The N-terminal arm of the S domain is colored in green. Boundaries of domains are indicated between the primary structure diagram. (B) Pairwise sequence identities between VP1 for GI.1 (crystal structure), GII.10 (Cryo-EM Map) and GIII.2 (TR-SAXS) and between their P and S domains. (C) Dendrogram of 29 representative VP1 sequences taken from Zheng et al. [6]. The 29 genotypes' accession numbers are given in the "Materials and Methods" section. The genotypes are grouped in the five major norovirus genogroups GI to GV (colored circles). Capsids for which structural data are available (GI.1, GII.10 and GIII.2) are indicated by arrows.

<https://doi.org/10.1371/journal.pone.0182056.g001>

Simulations

Molecular dynamics simulations were performed with the GROMACS simulation package version 5.0 [15] with the Amber 99SB-ILDN force-field [16] except for simulated annealing molecular dynamics, for which we used AMBER 9 [17–20]. Protocols are detailed in each subsection.

Simulated annealing molecular dynamics

Simulated annealing molecular dynamics were performed in implicit solvent. The starting model was the GIII.2 dimer after homology modeling and minimization (see above). Residues 1 to 30 were set as flexible while residues 31 to 522 remained fixed. Under the protocol used, the system was heated at 1200°K for 1ps and stabilized for 2ps. During the 13 next ps we reached 300°K and stabilized the system at this temperature for 1ps. Finally we reached 0°K during 3ps and the coordinates of the trapped structure were written at the end of each cycle. This basic loop of 18 ps was iterated 199 times to explore the conformational space of feasible solutions for the N-terminal arm.

Clustering

We developed in python a software to clusterize our trajectories [21]. We used the MDTRAJ package [22] for the trajectory reading and RMSD matrix calculation. Numpy 1.11 [23] was used for the hierarchical clustering. The Ward method [10] was chosen and its graphical classification of the trajectory was used to select the clustering level. The representative frame is defined by the structure with the lowest RMSD against all the frames which belong to the same cluster. The source code is available at <https://github.com/tubiana/TTClust>.

Preparatory simulations for the N-terminal arms study

The starting structure was chosen as the representative structure of the main cluster for chain A obtained from simulated annealing molecular dynamics. We generated the 8 possible sets of protonation states for the 3 acidic amino acids Glu10, Glu22 and Asp152 (GIII.2 numbering). These structures were then neutralized with the appropriate number of Na⁺ ions (from 14 for the triply-protonated arm to 17 for the non-protonated arm) and solvated with TIP3P water [24] under periodic condition with a non-cubic box with a distance between the protein and the edge of the box of at least 1.3 nm. The structures obtained after energy minimization were used to initiate molecular dynamics simulations. An integration time step of 2 fs was used during the calculation and the neighbour list was updated every 25th step. Short range non bonded Van der Waals (Lennard-Jones) and Coulomb interactions were calculated within a cut-off radius of 1 nm. The long range electrostatic interactions were calculated with the particle mesh Ewald (PME) method [25] with a grid spacing of 0.16 nm. The long range Lennard-Jones interactions were analytically corrected for in the calculation of the pressure and the energy.

Water molecules were constrained using the SETTLE algorithm [26] and the covalent bonds in the proteins were constrained using the P-LINCS algorithm [27].

Two NVT equilibrations of 500 ps were performed, one at 150°K followed by another one at 300°K. The solvent and the protein were coupled separately to an external heat bath at 300 K with the velocity-rescaling thermostat [28] using a time constant of 0.1 ps.

Then we performed a NPT equilibration of 500ps. The pressure of the simulation box was kept at an average of 1 bar using the isotropic Parrinello-Rahman barostat [29] with a time constant of 2 ps and a compressibility of $4.5e^{-5}$ bar⁻¹.

After the equilibration, the position restraints were removed. All other simulation parameters were the same as during the equilibration.

VMD 1.9.2 was used for RMSD calculations and Gromacs tool “rmsf” for the RMSF.

Atomic model derived from the GII.10 cryo-EM map

An extended model was generated for the GI.1 crystallographic dimer (pdb ID: 1IHM, chains A and B). The cryo-EM map used is the available one for the GII.10 norovirus VLP to about 10 Å resolution (EMDB ID: 5374)[8]. The interstitial regions (residues 220 to 230) of the crystallographic dimer were removed and the P dimer and S dimer were fitted separately with Chimera program version 1.10 [30]. The S dimer was first placed manually so that chain A was at its approximate position as one of the five A molecules around a fivefold axis. The S dimer was then adjusted by the auto-fit function of Chimera with a simulated map of 10 Å to finally reach a correlation coefficient of 0.89. The same protocol was used to fit the P dimer above the positioned S dimer, with a final correlation coefficient for the P dimer of 0.93. The S and P dimers were then connected by their two interstitial regions (residues 220 to 230) with Modeller.

Simulations for the P/S domain interaction study

For this study, we used the same simulation protocol as for N-terminal arms study (see above) for replicas 1 and 2. The distance between the protein and the edge of the box was increased from 1.3 nm to 2 nm for replicas 3–5. We initially set a pre-production time of 5 ns before a production period of 100 ns but finally included the whole 105 ns as production.

The principal axis of the S domain was calculated with the Single Value Decomposition module of numpy [23]. This method is adapted from the UCSF Chimera 1.10 [30]. Only residues in beta sheets of the S domain (67–70, 79–84, 107–114, 122–128, 147–151, 158–162, 181–186, 202–210) were considered to minimize the noise due to the thermic fluctuation of the

inter-strand loops. The height of domain S was calculated by aligning the P dimer axis on the Z axis. The center of mass of the jelly roll of the simulated protein was compared with the corresponding center of mass of the crystallographic GI.1 structure that was set to $z = 0$. Changing interactions between S and P domains were monitored with VMD [31] and plotted with matplotlib [32].

Sequence conservation

29 norovirus capsid protein full sequences were retrieved, each one the representative sequence of a single genotype [6]. Genbank accession codes for GI members are: M87661 (GI.1, crystal structure), L07418, U04469, AB042808, AJ277614, AF093797, AJ277609, AF538679; GII members are U07611, X81879, U02030, X76716, AJ277607, AJ277620, AJ277608, AF195848 (GII.10, cryo-EM map), AY038599, AF427118, AB074893, AJ277618, AY113106, AY130761, AY130762, AY502010, AY502009; GIII members are AJ011099, AF320625 (GIII.2, TR-SAXS); GIV member is AF195847 and GV member is AY228235. An alignment of these 29 sequences was generated within Jalview software v2.7 [33] with the MUSCLE [34] webserver and default parameters. From this alignment, a sequence conservation score was computed at each position in Jalview. The score is based on the one used in AMAS [35] and in case of variation takes into account the number of conserved physico-chemical properties at a position. The conservation is scored from 0 (not conserved) to 11 (completely conserved).

A dendrogram for these 29 sequences was calculated on the phylogeny.fr webserver [36] with PhyML 3.1/3.0 [37] with the aLRT statistical test [38] and WAG substitution model [39]. The tree was then rendered with drawtree 3.66 [36].

SAXS

VLP production, electron microscopy and dissociation. The capsid gene of norovirus GI.1 Norwalk virus (AY502016) was cloned into a baculovirus expression system as previously described [40]. The VLPs were harvested from Hi5 cells at five days post infection. The clarified supernatant was pelleted and applied to a 15–45% sucrose-PBS ultracentrifugation gradient (Beckmann SW40-Ti rotor) for 2 h at 4°C. VLP fractions were confirmed using EM and homogenous particles were pooled and concentrated to 2 mg/ml in PBS (pH 7.4). The VLP samples were applied to EM grids, washed once in distilled water, and then stained with 4% uranyl acetate. The samples were examined using EM (Zeiss EM 910). Prior to SAXS VLPs were dissociated into dimers through overnight dialysis at 4°C in CHES 0.01 M, pH 9. Protein concentration after dialysis was 1 mg/ml.

Data acquisition and analysis. Data were collected at the SWING beamline of Synchrotron Soleil (Saint-Aubin, France). Serial dilutions of the protein were measured (initial concentration, dilution 2 and dilution 4). The dialysis buffer was measured before and after each dilution. All buffer measurements were found to be superimposable and were averaged and subtracted from the protein solution measurements. Subsequent analyses were performed using various tools from the ATSAS suite, version 2.8.3 [41], as indicated. The three dilutions were analyzed in PRIMUS and found to be superimposable after scaling. We further analyzed the data from the undiluted sample. Experimental radius of gyration and Dmax were computed from the pair distribution function $p(r)$ with GNOM from the AutoRg routine in PRIMUS, to be consistent with our previous analysis of the GIII.2 dimer data [9]. We also generated a real space envelope using an automated Python script that streamlines and standardizes the classical procedure of multiple envelope averaging. 30 envelopes were generated with

DAMMIN and aligned with SUPCOMB. We finally took the average envelope and the lowest envelope corresponding to the envelope with all common points between all envelopes.

Rigid body modeling. We used EOM [42] for modeling with several conformations and DADIMODO [43] for modeling with a single conformation. To allow for sufficient flexibility we chose as flexible slightly enlarged versions of the termini and connecting segments. Thus in both cases the three rigid bodies were defined as residues 232 to 520 of both chains A and B (dimer of P), residues 29 to 217 of chain A and residues 29 to 217 of chain B. For DADIMODO that uses complete all-atom models, the N-terminus and C-terminus regions were added to the GI.1 crystallographic A-B dimer with Modeller 9.14. The DADIMODO genetic algorithm was used with the following parameters: At each step 20 offspring are generated by changing a random main chain dihedral angle in the flexible regions by a maximum of 30° and minimizing the resulting conformation. A population of 10 conformations is then selected by their fit to the SAXS data as computed with CRY SOL, with increasing selection pressure over a maximum of 300 generations.

Structure images

Images and movies were rendered with Pymol 1.8.2.0 [14].

Results

Simulations on the GIII.2 VP1 show an asymmetrical behavior of the dimer and protonation-dependent conformations in the N-terminal arm

Homology modeling from the GI.1 VP1 dimer and simulations. In the absence of a crystal structure of the GIII.2 VP1 of which we experimentally characterized the self-assembly behavior, both statically [44] and dynamically by TR-SAXS [9], we generated a homology model of the GIII.2 VP1 dimer from the GI.1 VLP crystal structure. The amino acid sequence of chain B is more complete in the crystal structure, with residues 10–520 (out of 530) present in the model, corresponding to residues 10–511 (out of 522) of the GIII.2 VP1. We therefore used chain B as a template for both molecules A and B of the GIII.2 model, adding the missing 9 N-terminal and 11 C-terminal residues in the modeling. We performed a 100 ns MD simulation to assess stability of this model. S and P domains maintained both their internal fold and their initial interactions P to P and S to S across the two-fold axis. RMSD of non-hydrogen atoms for this VP1 dimer core, excluding N and C-termini, plateaued to less than 4 Å (Fig 2A, black curve). The C-termini were dynamically flexible (not shown), as expected from parts not seen in any of the three crystallographic molecules. The N-terminal arms strongly diverged from the starting structure (Fig 2A, green curve), including residues 10–30, the counterparts of which are fully ordered in crystallographic molecule B. A simulation of the GI.1 crystal structure A-B dimer with termini added *ab initio* that we performed in parallel behaved very similarly, plateauing slightly below 3 Å RMSD for core P and S but with N-termini diverging (not shown).

We previously noted that in the GI.1 crystal structure ordering of the N-terminal arm from residue 10 in chain B (while it is disordered up to residue 29 in chains A and C [3]) involves clustering of three acidic amino-acids that seem conserved in GIII.2 VP1 and whose charges do not seem compensated by nearby basic residues [44]. Indeed, the initial disordering of the N-terminal arms in the simulations seems linked to initial electrostatic repulsion between negatively charged amino acids among this cluster (D13-E24-D156 for GI.1 and D10-E22-E152 for GIII.2). Furthermore, the N-terminal arm of chain B displayed higher mobility than that of chain A (Fig 2B).

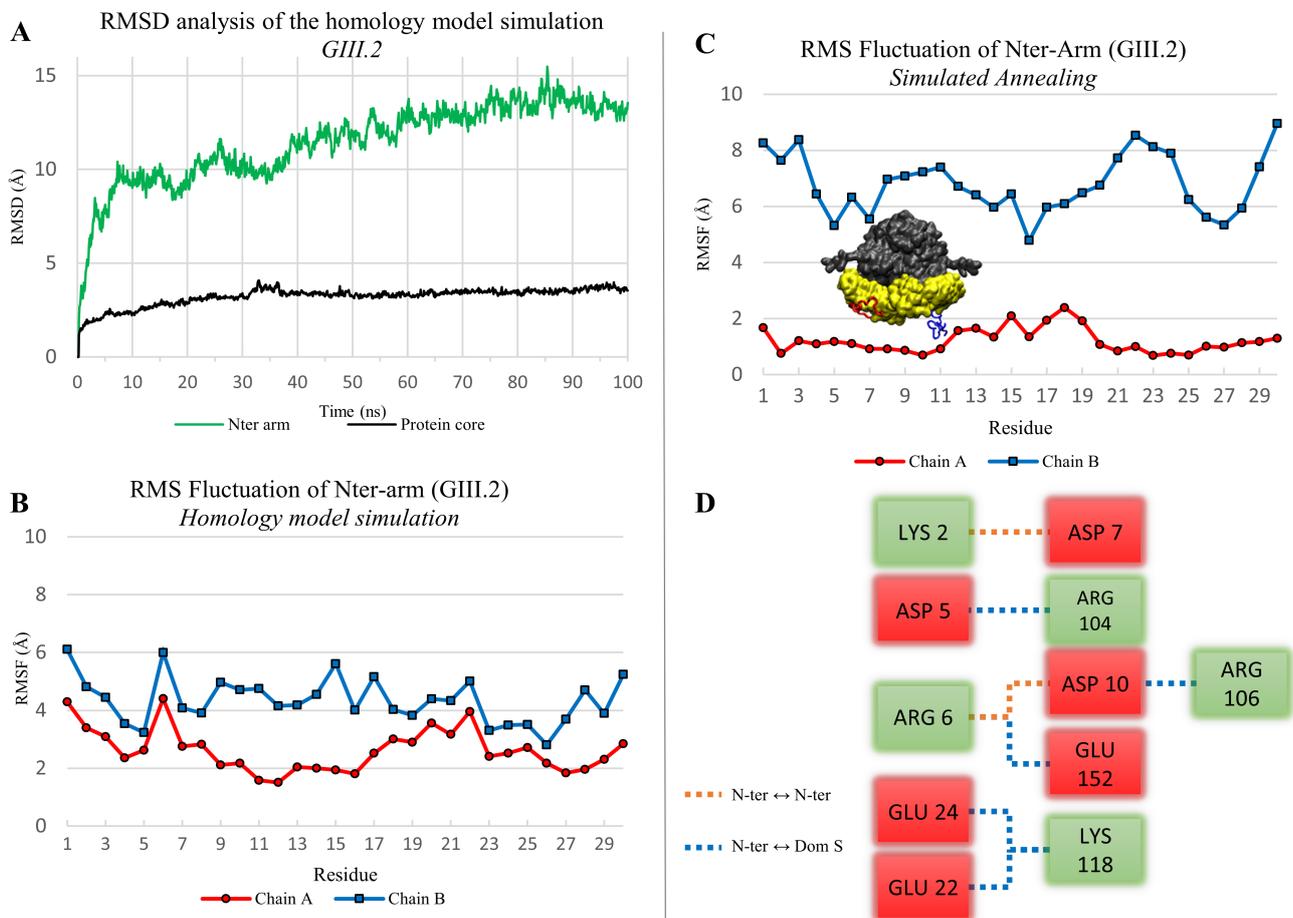


Fig 2. Dynamic properties and conformations of the N-terminal arm in a homology model of the GIII.2 dimer. (A) RMSD analysis of the homology model simulation (100 ns). The protein core (black curve) is defined by residues 31 to 512 (without N-terminal and C-terminal arms). The N-terminal arm is analysed separately (green curve). (B) RMS fluctuation of the N-terminal arms (residues 1 to 30) in the homology model simulation. Heavy atoms only were considered for the RMSD and RMSF calculation. (C) RMS fluctuation of N-terminal arms in the 199 endpoints of simulated annealing. A representative position of each arm is also displayed on the protein picture. (D) Salt bridge network found for the arm of chain A in 84% of the simulated annealing endpoints. Interactions within the arm are represented in orange and interactions with S domain in blue. The color of the box is defined by the charge of the amino acid (negative in red, positive in green).

<https://doi.org/10.1371/journal.pone.0182056.g002>

Simulated annealing finds an alternate and stable N-terminal arm conformation. In view of this, we sought to generate in a GIII.2 VP1 dimer alternate conformations of the N-terminal arms remote from their initial, unstable conformation. Our goal was to obtain starting points from which to simulate various alternate charge states of the arms. To this end we performed simulated annealing on our minimized homology model (initially perfectly symmetrical), keeping residues 31–522 fixed (S1 Movie).

We used our homemade trajectory clustering tool [21] to analyze the conformations of the N-terminal arms during the 199 simulated annealing iterations. Using the same cut-off on hierarchical clustering on both chains we find a larger number of distinct clusters on chain B than on chain A (S1 Fig). For the B chain's N-terminus, a cut-off of 8 arbitrary units yields 7 clusters (4 mains) with an average RMSD between the members of the clusters (spread) of 8.2 Å (Table 1). In contrast, for chain A only three clusters are present at this hierarchical distance

Table 1. Simulated annealing clustering results. For each cluster the size is the number of endpoints and the spread the average RMSD between all endpoints composing this cluster. The spread can also be represented by the compactness of nodes in a dendrogram (see S1 Fig). a, the three members of this cluster are the first three endpoints.

CLUSTER	SIZE	SPREAD (Å)
Chain A		
1	3 ^a	5.3
2	29	1
3	167	0.5
Chain B		
1	3	6.8
2	14	10.1
3	12	8.2
4	52	8.5
5	35	8.2
6	43	8.1
7	40	7.7

<https://doi.org/10.1371/journal.pone.0182056.t001>

and a preferred conformation emerges. There are two clusters with a spread of 0.5 Å and 1.0 Å constituting 83.9% and 14.5% respectively of all the chain A N-terminal conformations. The single sharp difference between the two clusters is a loop between residues 11 to 19 that occupies alternate positions in either cluster. Accordingly, RMSF along the N-terminal arm is low for chain A in the simulated annealing, with a single bump for this loop, while it is high for chain B (Fig 2C).

On examination of the endpoints, the N-terminal arm makes few interactions with its S domain for chain B. In contrast, in chain A there are numerous salt bridges between the N-terminal arm and the S domain (Fig 2D). Particularly, D10, E22 and E152 pair with R106 (corresponding to GI.1 R110), K118 (GI.1 K122) and R6 (GI.1 K6), respectively.

In summary, assuming basal protonation states and thus all negative charges of the D10-E22-E152 triad, not only is their clustering initially unstable, but the N-terminal arm can find a stable conformation (possibly a local minimum) where the three residues contribute three anchor points through salt bridges with the S domain.

Simulations reveal a protonation-dependent propensity to come back to the GI.1 arm conformation. Since we found a favored conformation of the fully deprotonated arm by simulated annealing (the one accounting for 84% of endpoints in chain A, see above), we next assessed the protonation dependence of the N-terminal arm conformation and mobility by starting from this conformation and varying charge states of the arm. The arm makes only intramolecular contacts in both the crystallographic and simulated annealing conformations. Therefore for this experiment only we worked with a monomer. We generated the eight possible protonation states of the triad, from D10 / E22 / E152 (all three residues deprotonated, as in the above analyses) to D10-H / E22-H / E152-H (all three residues protonated). We ran a 50-ns simulation for each of these eight combinations. Particularly we monitored the three difference distances between each pair of side chain carboxyls compared to the corresponding pair in the GI.1 crystal structure, for instance for D10 and E22 of GIII.2 (corresponding to D13 and E24 of GI.1) we considered

$$d(D13, E24) - d(D10, E22)$$

where $d(x, y)$ is the distance between the carbon of the side chain carboxyl of residue x and that of residue y (Fig 3A).

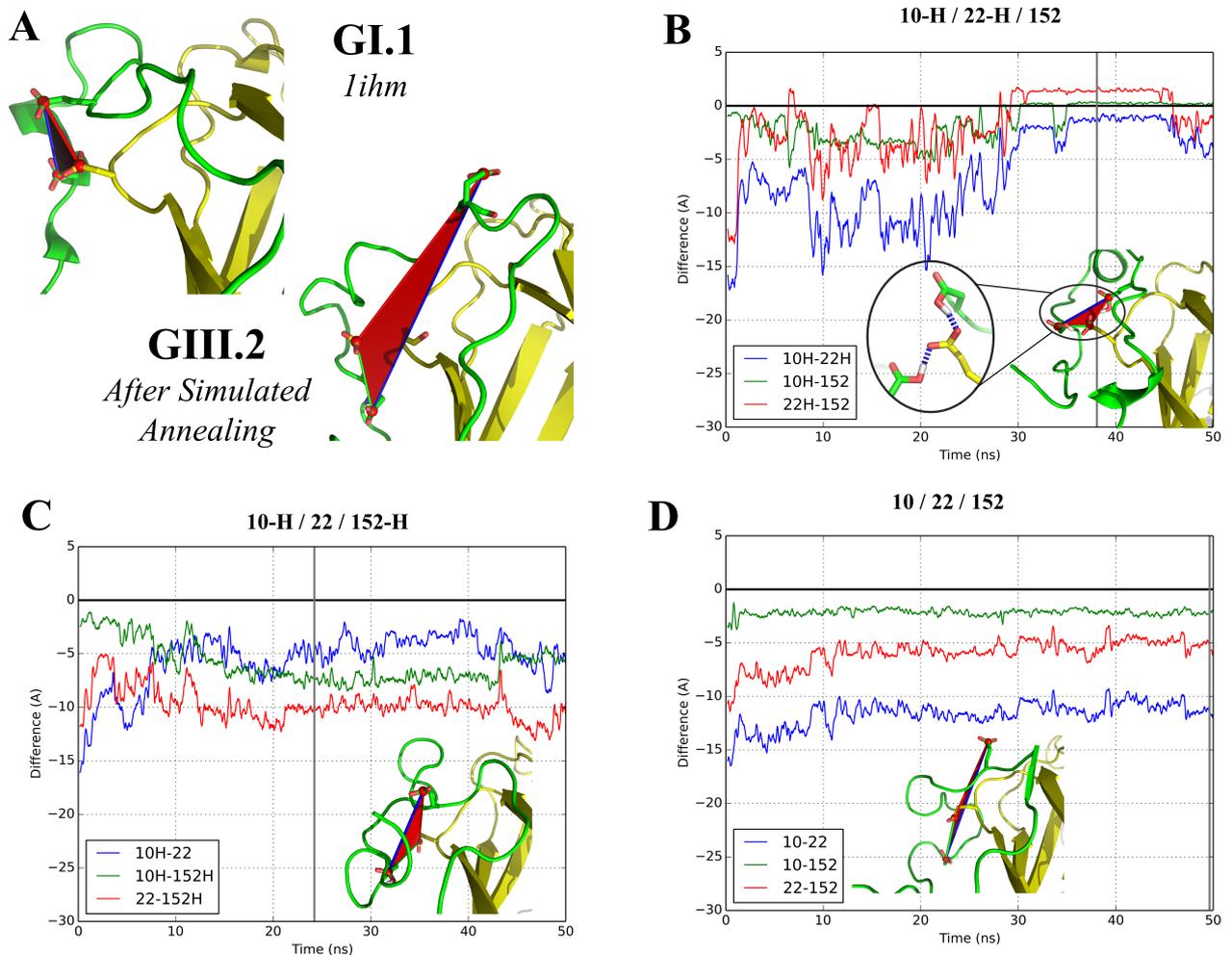


Fig 3. Molecular dynamics simulations in different sets of protonation for the three acidic residues clustered in the Gl.1 VP1. (A) Comparison of N-terminal arms in the Gl.1 crystal structure (chain B) and after simulated annealing of our GIII.2 model. Color code as in Fig 1A. The side chains of the three acidic residues considered are displayed. (B-D) Evolution of the difference distances between D10, E22 and E152. A value of 0 corresponds to restoration of the initial distance in the Gl.1 crystal structure while a negative value indicates a larger distance (see text for details). The illustrations represent the geometry of a representative frame of the largest cluster of each trajectory (S2 Fig). The gray vertical line indicates the frame from which the picture originates.

<https://doi.org/10.1371/journal.pone.0182056.g003>

In the case where all three residues are deprotonated (Fig 3D, S2 Movie), we find that the arm conformation relaxes into a position that remains close to the simulated annealing position. This is evidenced by the representative frame that is found at the end of the simulation and by the two difference distances that remain stable at high negative values, as E22 remains far from the other two residues of the triad. Overall the pattern of salt bridges depicted in Fig 2D remains in place. Indeed D10 and E152 draw slightly closer because they simultaneously interact with R6. The same behavior is found when D10 and only D10 is protonated (S3B Fig). In contrast, large changes occur as soon as either E22 or E152 are protonated. Two cases are then found. When either D10 or E22 remains deprotonated (Fig 3C, S2 Movie and S3B, S3C, S3D and S3E Fig), the N-terminal arm becomes dynamically disordered. As interactions between protonated residues and basic residues (formerly salt bridges) break, the corresponding

acidic residues drift away. The difference distances fluctuate along the simulations and generally remain below -4 \AA , indicating a lack of interaction between the acidic residues. However, when both D10 and E22 are protonated (Fig 3B, S2 Movie and S3F Fig), the initial fluctuation leads to grouping of the D10-E22-E152 triad into a stable cluster as in the GI.1 crystal structure.

These results suggest this clustering is likely a general property of the norovirus VP1 but also indicate that two protonation events are necessary in order for the N-terminal arm to fold into its final state in chain B. At the neutral pH at which the simulations were performed, this double event is unlikely without modifications of the chemical environment. In the discussion section below we discuss possible chemical clues that could be involved during capsid assembly.

Simulations on an extended form of the GI.1 dimer reveal a propensity for alternate, asymmetrical S-P interactions

Simulations starting from an extended dimer model derived from cryo-EM data of the GII.10 VLP. The compact form of VP1 with P domain interacting with S domain (Fig 1A) has been found so far only with the GI.1 VLP crystal structure. Our simulations show that this form is stable in the isolated dimer at the 100 ns timescale for both GI.1 VP1 and a homology model of GIII.2 VP1 (Fig 2A). However, we previously found by SAXS that the GIII.2 dimer has a larger radius of gyration than the GI.1 crystallographic dimer and that the GIII.2 VLP has a larger diameter than the GI.1 VLP [44]. Both GV virion and GII VLP have been imaged by EM as a double shell, with a cage of P dimers raised off the S shell and no contact between S and P [7][8]. We thus wanted to assess whether the GI.1 dissociated dimer could assume a similarly extended conformation. We took advantage of the availability of the EM map for GII.10 VLP to generate a starting extended GI.1 dimer model by successively fitting a GI.1 S dimer then a P dimer into this map (Fig 4A). As in previous reports we find the P dimer and S dimer rotated and separated compared to the crystal structure [7][8]. If we take the P dimer as a reference, the axes of the S domains are at an angle of some 53° (Fig 4B) and their center of masses are some 8 \AA below a plane orthogonal to the dimer axis (XY plane in Fig 4C).

We performed five 105 ns distinct simulations starting from this extended model (this time without the termini modeled in) to investigate the stability of the dimer in this conformation (S3 Movie). We monitored the angles and heights of S domains with reference to the crystal structure A-B dimer and examined the endpoints of the simulations (Fig 5). All five simulations quickly diverge from the initial extended model and its initial values of 53° and -8 \AA (see for instance Fig 5A and the angles in Fig 5C). There is a common behavior in all five simulations in that in all of them an S domain quickly comes into contact with a P domain and stays attached to it (Fig 5B), as evidenced by stabilisation of its relative height (Fig 5C).

Furthermore, the contact between S domains, that is found in both crystal dimers, is also maintained throughout the five simulations. However, no two simulations are the same and none comes back to a conformation similar to the crystallographic A-B dimer that we keep as a reference. The first simulation comes closest and indeed it is the only one in which both domains S come into contact with their P domains. In the other four simulations one S domain (either of chain A or of chain B) comes into contact with a P domain at an early stage and subsequently remains in interaction with it. Simulation 4 is an outlier in that the interaction is across molecules, between domain S of chain A and domain P of chain B. In all other simulations the interaction is within molecules as in the crystal structure, either within chain A (simulations 3 and 5), within chain B (simulation 2), or within both (simulation 1).

S-P interaction sites and their sequence conservation. We next looked at the residues involved in the interactions between the S and P domains along all five simulations (S4 Fig)

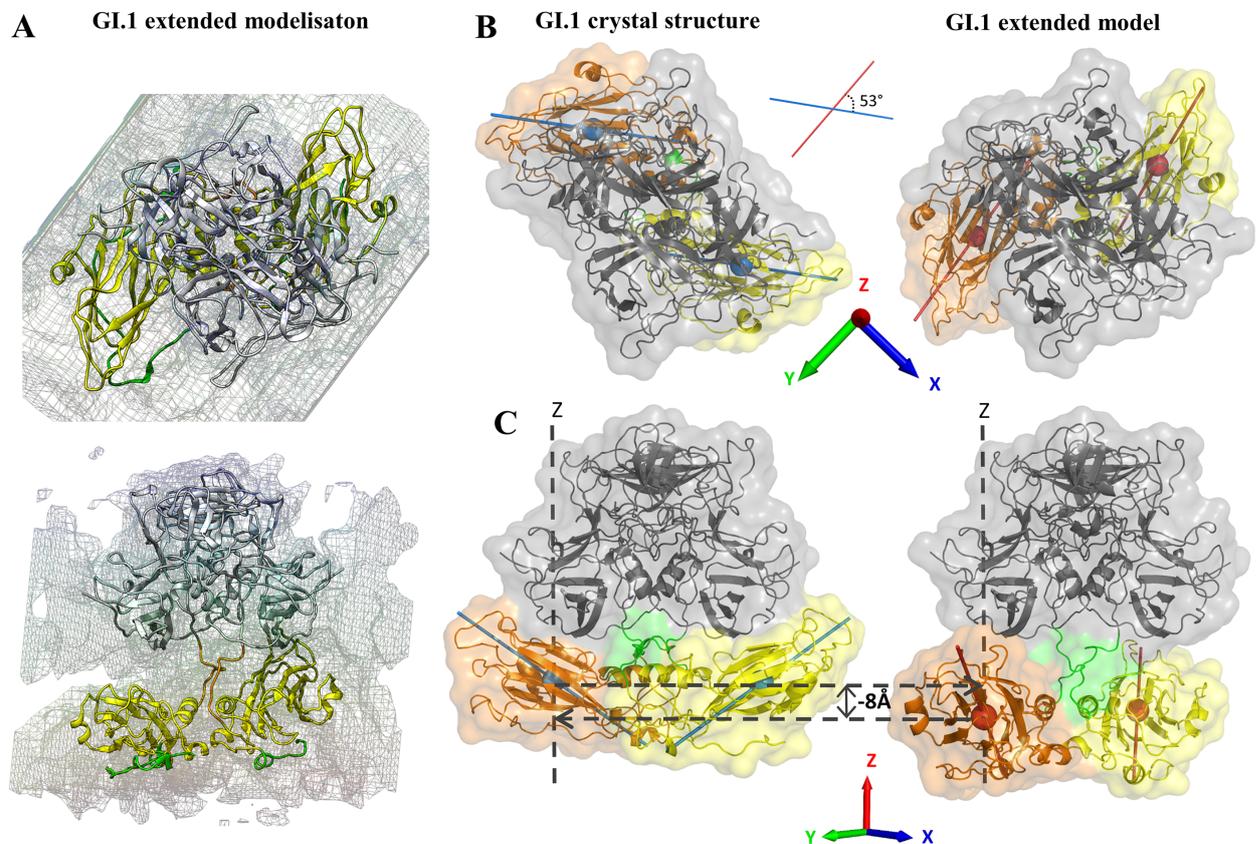


Fig 4. Construction of an extended form of GI.1 VP1 dimer based on the available cryo-EM data. (A) Fitting of GI.1 A-B dimer in the GII.10 cryo-EM Map (EMDB: 5374 [8]). The connecting segments were generated with modeller 9.14. The termini were not modeled. Color code as in Fig 1A. (B) Definition of angle and height of S domains relative to the crystal structure's. The reference is the dimer of P, the height is of the center of mass of S along the P dimer axis (OZ) and the angle is between the S axes as defined in the methods section. In this panel, S of chain A is in orange and of chain B in yellow.

<https://doi.org/10.1371/journal.pone.0182056.g004>

and compared them to the GI.1 VLP A, B and C molecules. In the simulations, after stabilisation of the S-P contacts, we find overall five segments in S interacting with four segments in P (Fig 6A). If we set aside the outlying simulation 4, in which the interaction is intermolecular and involves a contact segment (around residue 404) not found in other simulations, a consistent pattern emerges (Fig 6B and 6C). There are two interacting regions at the top of domain S made of segments 174–178 and 131–136 on the one hand and of a continuous patch made of neighboring segments 64–70, 71–82 and 198–205 on the other hand. These two regions are contacted by three loops at the bottom of domain P, a long central segment 504–518 and its flanking loops 449–458 and 465–469. Depending on the relative orientation of S and P, 504–518 may contact either or both S regions.

Despite different orientations of the domains, a similar pattern is found also in the crystallographic molecules A, B and C (Fig 6D). The segments involved are the same, and alternate interactions of loop 504–518 reflect the small conformational differences in the three quasi-equivalent molecules. The only exception is the short loop 465–469, belonging to P domain, that is involved in S contacts in all simulations but does not contact the S domain in any of the three crystallographic molecules. Instead in the VLP it caps segments 71–82 and 131–136 of

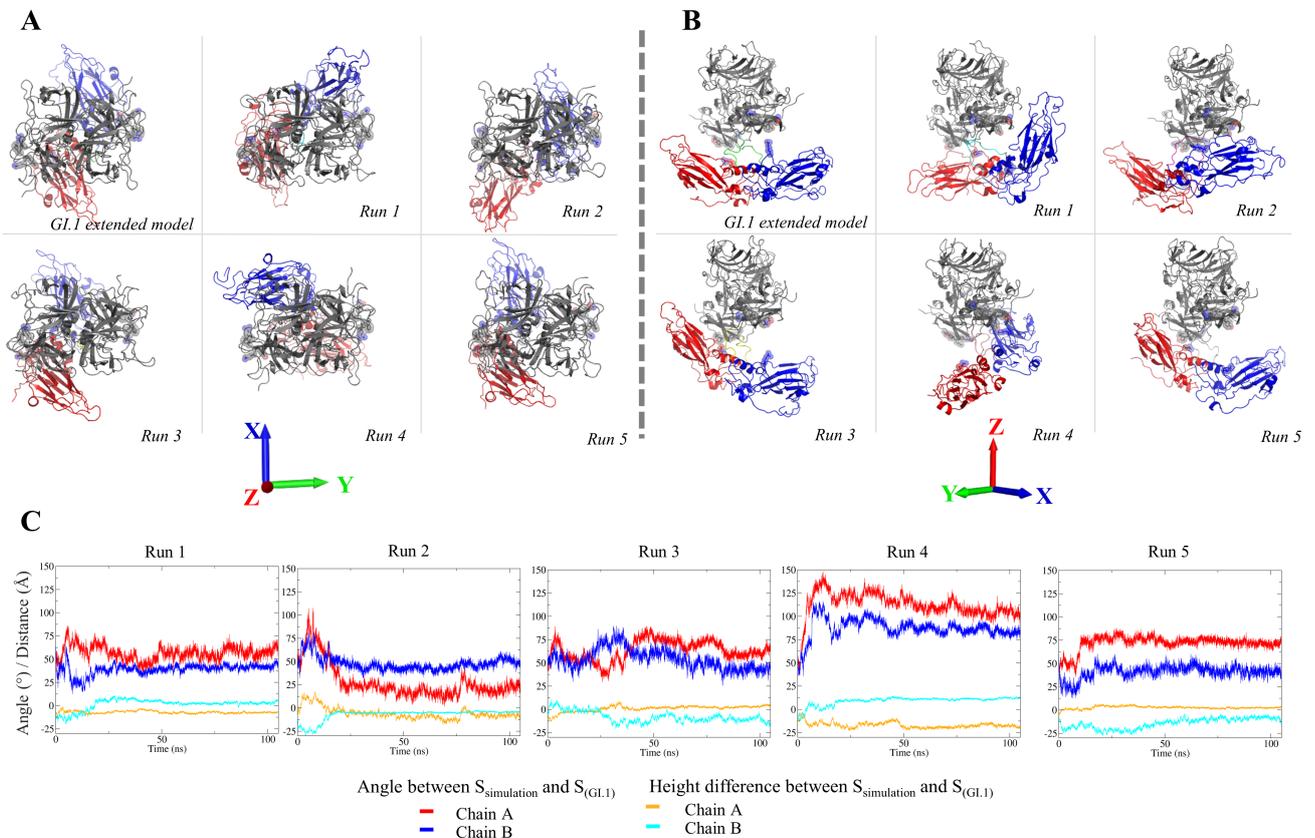


Fig 5. Simulations of the extended GI.1 dimer. Top (A) and front (B) view for the extended model of Fig 4 and the endpoints of 5 simulations. S domains are red for chain A and blue for chain B. (C) Evolution of the angles and heights of S domains as defined in Fig 4. Values of 0 correspond to the S conformation of the GI.1 crystal structure.

<https://doi.org/10.1371/journal.pone.0182056.g005>

one molecule in a neighboring dimer. Segment 401–410 is also involved in interactions (between P domains) in the VLP. Thus, all the P-S interacting segments that are found in our GI.1 dimer simulations participate in key interactions in the GI.1 VLP crystal structure, and all but one are the elements that position domain P relative to domain S. We examined their sequence conservation in the five major genogroups of noroviruses (Fig 7).

Most contain several very conserved residues, above 8 on the discrete scale (where 11 is strictly conserved and 10 very conserved with physico-chemical properties strictly conserved), while the average score for VP1's entire sequence is 5.84. This higher conservation is particularly noteworthy for the loops in domain P since conservation is lower in this domain (average score of 4.8). However, the first domain S interacting region (segments 131–136 and 174–178) is surprisingly nonconserved. In particular the polar segment 174–178, involved in the interaction with P and/or the interstitial region, displays very low conservation. For instance, R175-N176 is an amino acid pair that is most often prominently involved in P/S interactions in our simulations (S3 Fig) and in two of the three crystallographic molecules (Fig 6D). Yet in our sample of 29 norovirus genotypes it is present only in GI.1 and GII.10.

New SAXS data reveal a very extended and asymmetric GI.1 VP1 dimer. To further probe the conservation and differences between dimers of different genogroups, we acquired SAXS data on the GI.1 dimer dissociated from purified VLP by high pH and low ionic

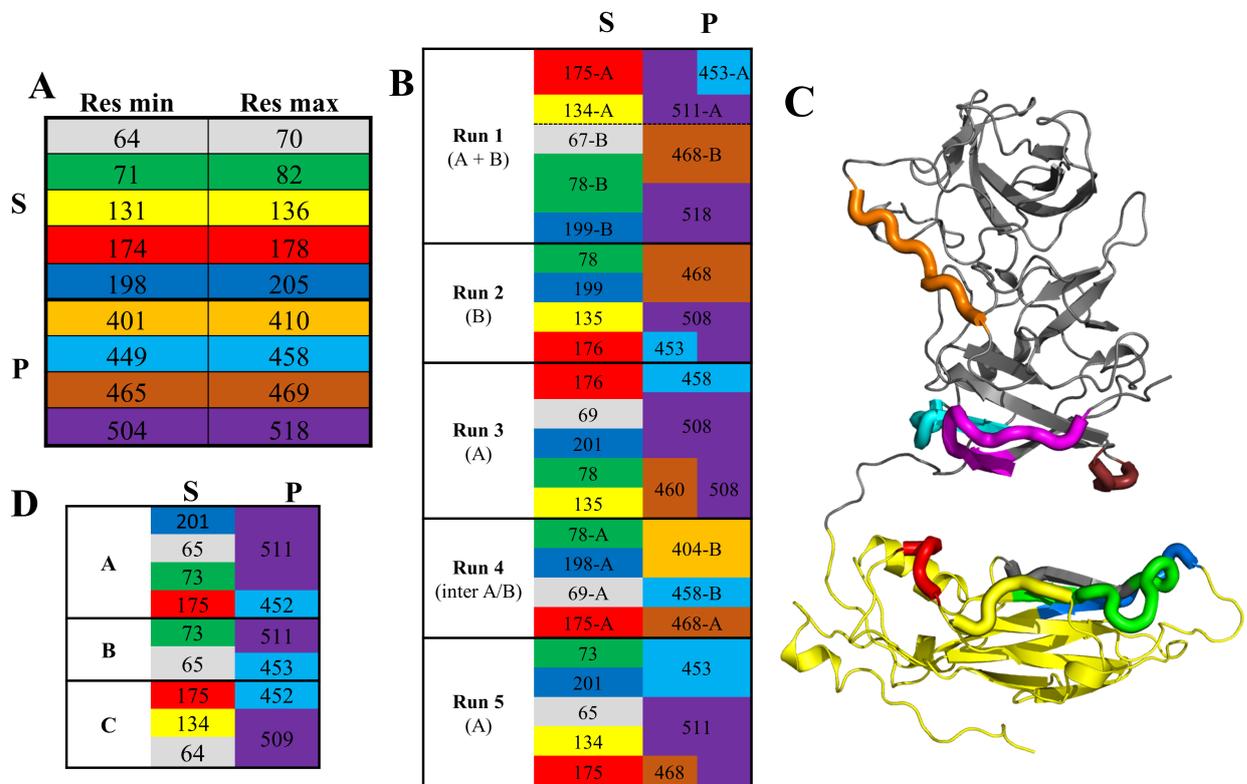


Fig 6. Segments of S and P domains interacting with one another. (A) Limits of amino acid segments found in S (top) interacting with P or in P (bottom) interacting with S in any of the 5 simulations of Fig 5. Each segment is associated with a color kept in the rest of the figure. (B) Interactions between segments in all simulations. Interacting segments between S (left) and P (right) are presented on the same line. The segment's residue indicated is the one with the strongest and/or most frequent interaction (S4 Fig). Chains are indicated as needed (run 1, where both chains have S-P interactions; run 4, where cross-chain interactions are observed). (C) Location of the segments on the (extended) GI.1 VP1 structure. (D) Same table as (B), but for the three crystallographic molecules.

<https://doi.org/10.1371/journal.pone.0182056.g006>

strength, as we did previously for the GIII.2 dimer [9]. We could use a sample setup more suited to a 115-kDa protein than in our previous GIII.2 work, yielding more accurate data on a more appropriate range of wavenumbers ($0.01 \text{ \AA}^{-1} < q < 0.42 \text{ \AA}^{-1}$). In their common range the dimer data are comparable between GI.1 and GIII.2 (S5 Fig) and their analysis in terms of distance distributions yield similar radii of gyration of 42 Å and 43 Å, respectively, for a maximum dimension of 150 Å. This shows that the two dimers are comparable in their extension. The derived SAXS molecular envelopes are also comparable for GI.1 and GIII.2 as cylinders with a central bulge (not shown). However, the radii of gyration and maximum dimensions are much higher than computed from the X-ray crystallography or cryo-EM derived experimental model (Table 2). In all our simulations from the cryo-EM model (described above), radii of gyration also keep well below 42 Å, settling at or near the cryo-EM value (S6 Fig).

The limitations of the GIII.2 data did not allow a more precise definition of the solution dimer form, as they were consistent with several qualitatively different associations of domains (not shown). The higher quality of the present GI.1 data (S5 Fig) allows more discriminating reconstruction of the solution structure by rigid body modeling. We set the N- and C-termini and the two connecting segments as flexible while keeping two S domains and one P dimer as rigid (see material and methods). We either used the strategy of generating 10000 random

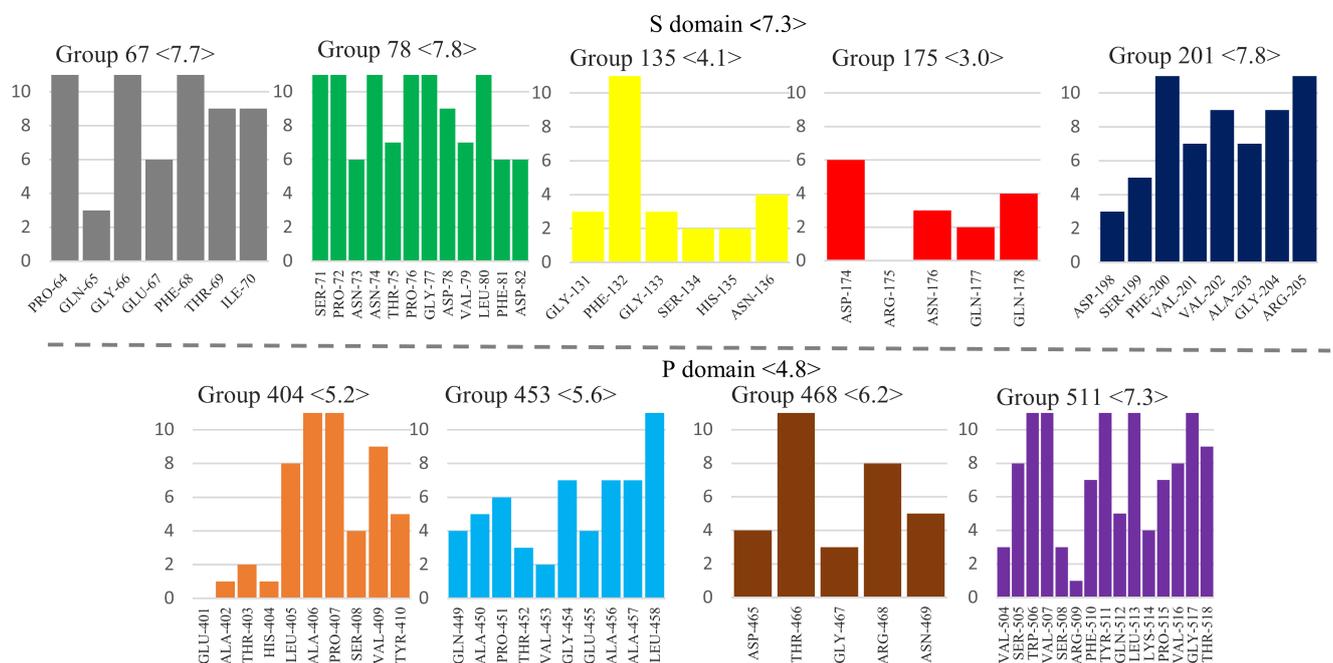


Fig 7. Sequence conservation among noroviruses of the S-P interaction segments. The color code is the same as Fig 6. The conservation score was calculated with Jalview [33] on a scale from 0 to 11 based on one representative VP1 sequence for each of 29 genotypes [6]. Averages for segments and domains are indicated in brackets.

<https://doi.org/10.1371/journal.pone.0182056.g007>

conformers and selecting a subset best matching the data [42], or the strategy of single-molecule modeling by perturbation from a starting conformation then selection based on fit to the data of intermediate models [45][43]. The former strategy was less successful than the latter in getting a good fit to the data, either when constraining models to two-fold symmetry (Fig 8A) or when generating non-symmetrical models (Fig 8B).

These searches result in a selection of 3 symmetric models whose combination yields a χ^2 goodness of fit of 4 or of 2 asymmetric ones with a χ^2 near 5, respectively (Table 3). In contrast, we obtain single models that fit our data better (χ^2 in the range 2–2.5) by using the genetic algorithm implemented in the DADIMODO software, where generation of populations of random conformational variants is cycled with selection on their fit to the data [43]. The model with the best fit is shown on (Fig 8C). In all cases, we obtain models with the S domains separated and coming on either side of the P dimer, either as major components of multi-model ensembles (Fig 8A and 8B, Table 3) or as the single-model solution (Fig 8C). When not constrained to

Table 2. Radii of gyration and maximum dimensions of crystallographic dimer, cryo-EM based model, SAXS data and Dadimodo model. Radii of gyration and Dmax were calculated with Crysol for structural models, for which termini were included. P(R) analysis was used for experimental data.

	Radii of gyration (Å)	Dmax (Å)
SAXS GI.1	43.3	150
SAXS GIII.2	42.0	150
GI.I Xray	32.7	99.6
Cryo-EM Model	35.1	110.9
Dadimodo	43.7	202.1

<https://doi.org/10.1371/journal.pone.0182056.t002>

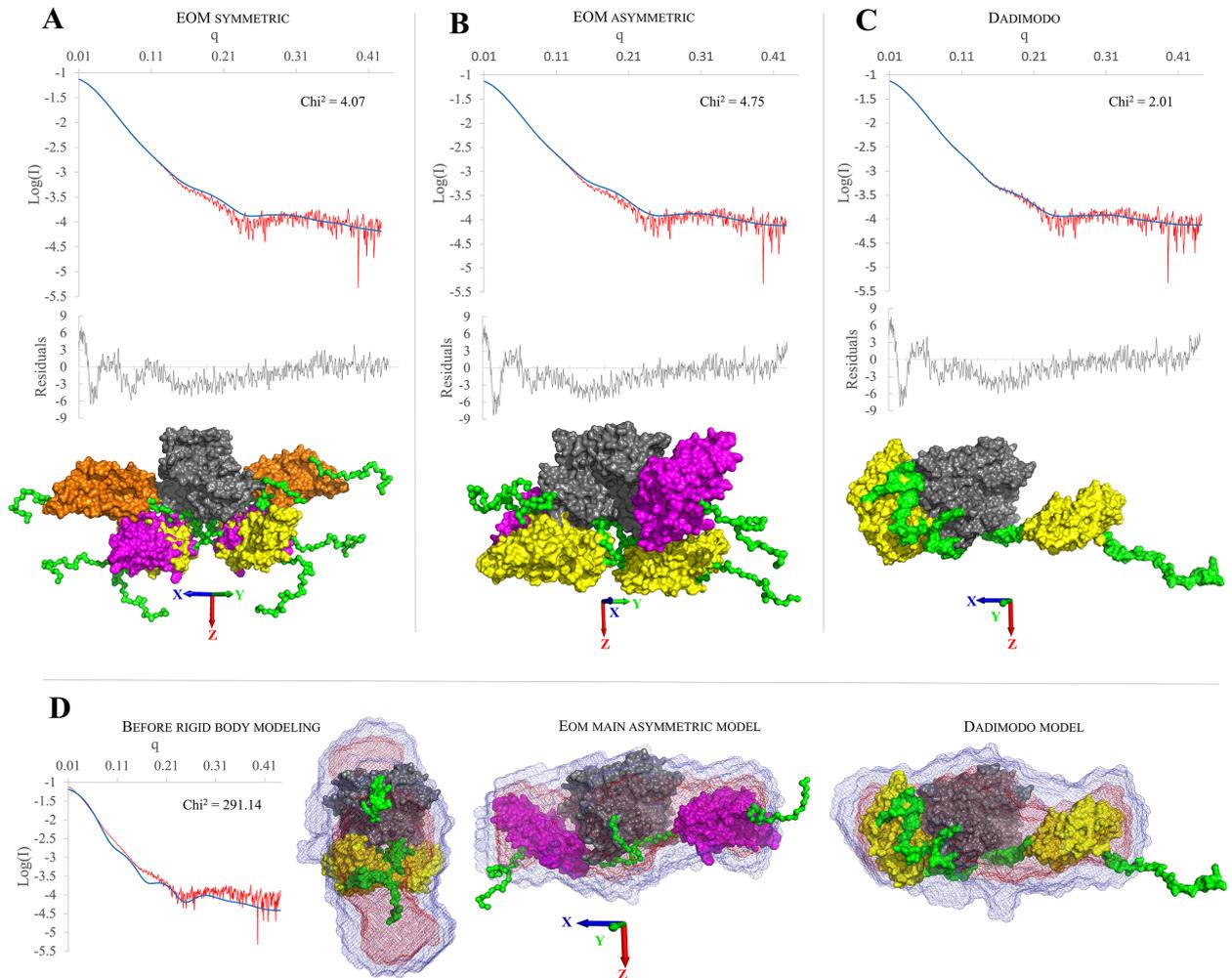


Fig 8. Solution structures of dissociated GI.1 dimer obtained by rigid body modeling of SAXS data. The fit of the curve computed from the model (in blue) to the experimental data (in red) and their associated reduced residuals (in gray) are displayed atop the models. (A) EOM results for 10000 symmetric models. The population of 3 selected models is displayed superposed on their P domains (in gray) with the 3 pairs of S domains in yellow, orange and magenta. (B) EOM results for 10000 asymmetric models. The population of 2 selected models is displayed superposed on their P domains (in gray) with the 2 pairs of S domains in yellow and magenta. (C) Fit of the single Dadimodo model. (D) Comparison of the GI.1 dimer before rigid body modeling (1IHM with arms modeled in, the fit to the experimental data is also shown) with the two main models from EOM (asymmetric) and Dadimodo. The best superposition of the *ab initio* envelope computed from the SAXS data is displayed for each model.

<https://doi.org/10.1371/journal.pone.0182056.g008>

two-fold symmetry solution (Fig 8B and 8C) the solutions also display a pronounced asymmetry with one S close to the P dimer while the other extends away from it, a shape that matches well the molecular envelope while the crystallographic dimer does not (Fig 8D).

Discussion

The norovirus protein VP1 forms stable dimers of which 90 self-assemble to form the major protein part of the capsid. VP1 is a two-domain protein harboring two points of flexibility, as seen in the crystal structure of the GI.1 VLP (Fig 1A). The first point of flexibility is a *ca* 10-residue segment between domains S and P that allows S dimers to assume either a "bent" conformation (dimer A-B) or a "flat" conformation (dimer C-C) [3]. The classical quasi-equivalence

Table 3. Individual χ^2 for EOM models, GI.1 crystallographic dimer and Dadimodo model. EOM models are presented with their fraction used to fit the experimental data. χ^2 are calculated with Crystol [46]. a, for this model only the program FOXS [47] is used for computing χ^2 . The RMSD was calculated with the VMD visualizer tools using the Dadimodo model as reference. All proteins were aligned on the P domain and the RMSD was calculated based on the S/P domains position.

Symmetric models			
Model	Used for (%)	χ^2	RMSD (Å)
1191	29	30.455	45.2
3878	57	13.51	60.7
8091	14	35.53	33.4
combination		4.07	
Asymmetric models			
1590	27	24.4	38.2
3350	73	7.1	12.2
combination		4.75	
Other models			
GI.1 xray		326 ^a	31.6
Dadimodo		2.01	0

<https://doi.org/10.1371/journal.pone.0182056.t003>

in the norovirus capsid allows environments of all molecules to be almost identical and thus A, B and C molecules display very limited adjustments in GI.1 VLP. From this observation Prasad et al [3] drew the most simple model of norovirus capsid self-assembly: A-B and C-C dimers would interconvert in solution, the proper conformation being incorporated at each point into a growing capsid. Indeed, our 100-ns simulations of a GI.1 crystallographic dimer (not shown) or of the GIII.2 model based on it (Fig 2A) would be consistent with this view of limited conformational range. However, we showed by SAXS that for the GIII VP1, compared to the GI.1 crystal structure, a more expanded VLP assembles from more extended dimers [44][9]. Here we show by SAXS that similarly dissociated GI.1 VP1 dimers assume as extended a conformation (S5 Fig). This shows that for GI.1 too the available range of conformations of VP1 goes well beyond the slight differences between A-B and C-C dimers in the GI.1 VLP. Indeed, rigid body modeling shows that the solution structure of the dimer is not simply a separation of P dimer and S dimer as for instance in the GII VLP [8], combined to extension of the mobile termini (Table 2). The solution structure is actually far out of the range of conformations assumed by VP1 in assembled capsids. A major difference is a prominent break of two-fold symmetry with one S domain in contact with the P dimer while the other projects away from it (Fig 8). Simulations starting from an extended GI.1 dimer based on the available GII.10 cryo-EM map (Fig 4) also feature asymmetric conformations, most of the time with only one S interacting with P (Fig 5, S3 Movie). The interactions involve the same segments of S and P as in the GI.1 crystal structure, but some of the contacts are not found in the final capsids (Fig 6). This further establishes the capability of the GI.1 VP1 dimer to assume conformations not seen in the crystal structure.

These results raise the question of the possible status of strongly asymmetric dimers in norovirus capsid assembly. From the large contact area between the five S domains around the five-fold axis, Prasad et al surmised that the initial assembly nucleus would be a pentamer of dimers and naturally considered that this nucleus would grow symmetrically [3]. Thus the next step would be a decamer of dimers with five-fold symmetry preserved, continuing the growth of a spherical piece of capsid. With pronouncedly asymmetrical dimers, the first step is not compromised (indeed the extension of one S domain away from the rest of the dimer may facilitate formation of an initial S-based pentamer of dimers), but symmetrical capture of five

further dimers becomes problematic. Indeed, we found that the major intermediate in GIII.2 capsid self-assembly, although a 10- or 11-mer of dimers, is not a spherical piece of capsid. Instead it is very elongated and consistent with two connected pentamers of dimers [9]. The tendency for symmetry break of VP1 dimers we find here gives a possible structural basis for that unexpected observation.

Subsequent formation of the quasi-sixfold axes that feature in the architecture of the T = 3 norovirus capsid relies on the ability of the S domain to form also hexamers with almost the same pairwise contacts as in S pentamers. Interestingly, differences with the pentamers involve the second point of flexibility in VP1: The N-terminal arm that is disordered up to residue 29 in molecules A and C in the GI.1 VLP. Its ordering from residue 10 in molecule B actually cements the S hexamer by setting up on the one hand interactions between arms around the quasi-sixfold axis and on the other hand interactions with the nearby domains S of C molecules. However we show here that this arm conformation is unstable as is (Fig 2) due to a cluster of acidic amino-acids, two in the switching part of the arm and one on the surface of the S domain (Fig 3). For the GIII.2 dimer, we even find an alternate arm position that is highly stable (S1 Movie) due to salt bridges with domain S. Interestingly, we come back to clustering of the three acidic residues if and only if we protonate the two arm residues (Fig 3, S2 Movie, S3 Fig). Our findings thus highlight a switch in the VP1 N-terminal arm that requires neutralization of two of its negative charges to assume the conformation seen around quasi-sixfold axes. Such a mechanism requires time and/or the right triggers for formation of this part of the capsid. This would explain our biophysical observation that there is a slow step in the self-assembly of VP1 VLP going from the aforementioned (and presumably pentamer-based) intermediate to the full capsid [9]. In that previous work we triggered self-assembly by a salinity and pH jump from 9 to 6, so protonation of the two acidic residues still requires a shift in their pKa brought by a changing environment during assembly. In the GI.1 VLP crystal structure, protonation and stabilization of the B molecule N-terminal arm may have been favored by the crystallization pH of 4.8 [3]. But during assembly, that presumably occurs in a non-acidic cellular compartment, two alternate possibilities may be more relevant. First, ions may intervene to neutralize the charges of the two arm residues. Second, these charges may also be neutralized by the small basic norovirus protein VP2. VP2 is present in infectious norovirus particles in a small number of copies and is thought to mediate genome packaging during assembly [48]. Its involvement also in neutralization of the N-terminal arm is a particularly attractive speculation as it would provide a simple way to link capsid completion to genome encapsidation.

In conclusion, we find that the dissociated norovirus VP1 dimer tends to assume asymmetric conformations in solution through the two flexible points in its two-domain structure. We propose that this feature is functionally important in the assembly process.

Supporting information

S1 Fig. Clustering of N-terminal arm simulated annealing endpoints (A) for chain A and (B) for chain B. Top, clustering dendrograms. The cutoff used to generate clusters from hierarchical classification is represented by a gray bar (8 for both chains). Bottom, each successive frame is assigned its cluster color. Graphics were generated with the TrajectoryClustering program (available at <https://github.com/tubiana/TrajectoryClustering>). (TIF)

S2 Fig. Clustering of N-terminal arm conformations along simulations with different protonation states. For every cluster, the size (number of frames in the cluster) and the spread (average rmsd between all cluster's frames) are given. The number of the cluster depends on

its order of appearance in the trajectory. A color code was attributed to each cluster.
(TIF)

S3 Fig. Molecular dynamics simulations in different sets of protonation for the three acidic residues clustered in the GI.1 VP1. As for Fig 3, (A-F) Evolution of the difference distances between D10, E22 and E152. A value of 0 corresponds to restoration of the initial distance in the GI.1 structure while a negative value indicates a larger distance (see text for details). The illustrations represent the geometry of a representative frame according to the clusterisation of each trajectory (S2 Fig). The gray vertical line indicates the frame from which the picture originates.
(TIF)

S4 Fig. Timelines of S domain interactions with P domain for simulations from an extended form of the GI.1 dimer. Residues inside a rectangle with the same colour belong to a single segment as defined in Fig 6. Interactions were sampled every 100 ps with VMD [31] and graphics made with matplotlib [32].
(TIF)

S5 Fig. Comparison of SAXS results of the GI.1 (blue) and GIII.2 (orange) VP1 dimer.
(TIF)

S6 Fig. Timelines of radii of gyration during simulations from an extended form of the GI.1 dimer. Values for the crystallographic model and our cryo-EM derived model (without termini) are indicated.
(TIF)

S1 Movie. Simulated annealing trajectory. Representation as in Fig 2C, with the N-terminal arms of chains A and B in green and emerald, respectively. Rendered with VMD.
(MP4)

S2 Movie. 50 ns molecular dynamics simulations in three different sets of protonation for the three acidic residues clustered in the GI.1 VP1. The protonation states and representation are the same as for Fig 3.
(MP4)

S3 Movie. 105 ns simulations of the extended GI.1 dimer. The 5 replicas are displayed as in Fig 4.
(MP4)

Acknowledgments

This work was granted access to the high performance computing resources of TGCC under the allocation t2016077583 made by GENCI. We thank Anna Koromyslova and Grant Hansman for their generous gift of the GI.1 VLP sample. We acknowledge synchrotron SOLEIL for allocation of beamtime, Aurélien Thureau for invaluable assistance in SAXS data collection and the automated DAMMIN script, Olga Roudenko for access to SOLEIL computing infrastructure and DADIMODO computations, and Marc Baaden for invaluable advice.

Author Contributions

Conceptualization: Yves Boulard, Stéphane Bressanelli.

Data curation: Thibault Tubiana.

Formal analysis: Thibault Tubiana.

Funding acquisition: Stéphane Bressanelli.

Investigation: Thibault Tubiana.

Methodology: Thibault Tubiana.

Project administration: Yves Boulard, Stéphane Bressanelli.

Software: Thibault Tubiana.

Supervision: Yves Boulard, Stéphane Bressanelli.

Visualization: Thibault Tubiana.

Writing – original draft: Stéphane Bressanelli.

Writing – review & editing: Thibault Tubiana, Yves Boulard, Stéphane Bressanelli.

References

- Green KY. Caliciviridae: The Noroviruses. In: Krieger DM, Howley PM, editors. *Fields Virology*. Philadelphia: Lippincott, Williams, and Wilkins; 2013. pp. 582–608.
- Jiang X, Wang M, Graham DY, Estes MK. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J Virol*. 1992; 66: 6527–6532. PMID: [1328679](#)
- Prasad BVV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. X-ray Crystallographic Structure of the Norwalk Virus Capsid. *Science*. 1999; 286: 287–290. <https://doi.org/10.1126/science.286.5438.287> PMID: [10514371](#)
- Caspar DL, Klug A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*. 1962; 27: 1–24. PMID: [14019094](#)
- Kroneman A, Vega E, Vennema H, Vinjé J, White PA, Hansman G, et al. Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol*. 2013; 158: 2059–2068. <https://doi.org/10.1007/s00705-013-1708-5> PMID: [23615870](#)
- Zheng D-P, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS. Norovirus classification and proposed strain nomenclature. *Virology*. 2006; 346: 312–323. <https://doi.org/10.1016/j.virol.2005.11.015> PMID: [16343580](#)
- Katpally U, Wobus CE, Dryden K, Virgin HW, Smith TJ. Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles. *J Virol*. 2008; 82: 2079–2088. <https://doi.org/10.1128/JVI.02200-07> PMID: [18094184](#)
- Hansman GS, Taylor DW, McLellan JS, Smith TJ, Georgiev I, Tame JRH, et al. Structural Basis for Broad Detection of Genogroup II Noroviruses by a Monoclonal Antibody That Binds to a Site Occluded in the Viral Particle. *J Virol*. 2012; 86: 3635–3646. <https://doi.org/10.1128/JVI.06868-11> PMID: [22278249](#)
- Tresselt G, Le Coeur C, Bryche J-F, Tatou M, Zeghal M, Charpilienne A, et al. Norovirus Capsid Proteins Self-Assemble through Biphasic Kinetics via Long-Lived State-like Intermediates. *J Am Chem Soc*. 2013; 135: 15373–15381. <https://doi.org/10.1021/ja403550f> PMID: [23822934](#)
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan M s, Eramian D, Shen M, et al. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Protein Science*. John Wiley & Sons, Inc.; 2001. Available: <http://onlinelibrary.wiley.com/doi/10.1002/0471140864.ps0209s50/abstract>
- Sali A. Comparative protein modeling by satisfaction of spatial restraints. *Mol Med Today*. 1995; 1: 270–277. [https://doi.org/10.1016/S1357-4310\(95\)91170-7](https://doi.org/10.1016/S1357-4310(95)91170-7) PMID: [9415161](#)
- Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci Publ Protein Soc*. 2006; 15: 2507–2524. <https://doi.org/10.1110/ps.062416606> PMID: [17075131](#)
- Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins Struct Funct Bioinforma*. 2003; 50: 437–450. <https://doi.org/10.1002/prot.10286> PMID: [12557186](#)
- Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
- Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1–2: 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>

16. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinforma*. 2010; 78: 1950–1958. <https://doi.org/10.1002/prot.22711> PMID: 20408171
17. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. *J Comput Chem*. 2005; 26: 1668–1688. <https://doi.org/10.1002/jcc.20290> PMID: 16200636
18. Case DA, Darden TA, Cheatham TE, Lii CLS, Wang J, Duke RE, et al. AMBER 9. San Francisco: University of California; 2006.
19. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun*. 1995; 91: 1–41. [https://doi.org/10.1016/0010-4655\(95\)00041-D](https://doi.org/10.1016/0010-4655(95)00041-D)
20. Case D.A., Darden T.A., Cheatham T.E. III, Simmerling C.L., Wang J, Duke R.E., Luo R., Merz K.M., Pearlman D.A., Crowley M., Walker R.C., Zhang W., Wang B., Hayik S., Roitberg A., Seabra G., Wong K.F., Paesani F., Wu X., Brozell S., Tsui V., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Beroza P., Mathews D.H., Schafmeister C., Ross W.S., and Kollman P.A. (2006), AMBER 9, University of California, San Francisco.
21. Tubiana T, Boulard Y, Bressanelli S. TTClust: A versatile graphical molecular simulation trajectory clustering program. *Bioinforma Oxf Engl. soumis*;
22. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*. 2015; 109: 1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015> PMID: 26488642
23. Walt S van der, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng*. 2011; 13: 22–30. <https://doi.org/10.1109/MCSE.2011.37>
24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79: 926–935. <https://doi.org/10.1063/1.445869>
25. Darden T, York D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993; 98: 10089–10092. <https://doi.org/10.1063/1.464397>
26. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem*. 1992; 13: 952–962. <https://doi.org/10.1002/jcc.540130805>
27. Hess* B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput*. 2008; 4: 116–122. <https://doi.org/10.1021/ct700200b> PMID: 26619985
28. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007; 126: 14101. <https://doi.org/10.1063/1.2408420> PMID: 17212484
29. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*. 1981; 52: 7182–7190. <https://doi.org/10.1063/1.328693>
30. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25: 1605–1612. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
31. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996; 14: 33–38, 27–28. PMID: 8744570
32. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng*. 2007; 9: 90–95. <https://doi.org/10.1109/MCSE.2007.55>
33. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinforma Oxf Engl*. 2009; 25: 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033> PMID: 19151095
34. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
35. Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Bioinformatics*. 1993; 9: 745–756. <https://doi.org/10.1093/bioinformatics/9.6.745>
36. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. 2008; 36: W465–W469. <https://doi.org/10.1093/nar/gkn180> PMID: 18424797
37. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol*. 2010; 59: 307–321. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638

38. Anisimova M, Gascuel O. Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Syst Biol*. 2006; 55: 539–552. <https://doi.org/10.1080/10635150600755453> PMID: [16785212](https://pubmed.ncbi.nlm.nih.gov/16785212/)
39. Whelan S, Goldman N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol*. 2001; 18: 691–699. <https://doi.org/10.1093/oxfordjournals.molbev.a003851> PMID: [11319253](https://pubmed.ncbi.nlm.nih.gov/11319253/)
40. Koromyslova AD, Hansman GS. Nanobody binding to a conserved epitope promotes norovirus particle disassembly. *J Virol*. 2015; 89: 2718–2730. <https://doi.org/10.1128/JVI.03176-14> PMID: [25520510](https://pubmed.ncbi.nlm.nih.gov/25520510/)
41. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr*. 2012; 45: 342–350. <https://doi.org/10.1107/S0021889812007662> PMID: [25484842](https://pubmed.ncbi.nlm.nih.gov/25484842/)
42. Tria G, Mertens HDT, Kachala M, Svergun DI. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*. 2015; 2: 207–217. <https://doi.org/10.1107/S205225251500202X> PMID: [25866658](https://pubmed.ncbi.nlm.nih.gov/25866658/)
43. Evrard G, Mareuil F, Bontems F, Sizun C, Perez J. DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *J Appl Crystallogr*. 2011; 44: 1264–1271. <https://doi.org/10.1107/S0021889811035758>
44. Tresset G, Decouche V, Bryche J-F, Charpilienne A, Le Cœur C, Barbier C, et al. Unusual self-assembly properties of Norovirus Newbury2 virus-like particles. *Arch Biochem Biophys*. 2013; 537: 144–152. <https://doi.org/10.1016/j.abb.2013.07.003> PMID: [23871846](https://pubmed.ncbi.nlm.nih.gov/23871846/)
45. Petoukhov MV, Svergun DI. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys J*. 2005; 89: 1237–1250. <https://doi.org/10.1529/biophysj.105.064154> PMID: [15923225](https://pubmed.ncbi.nlm.nih.gov/15923225/)
46. Svergun D, Barberato C, Koch MHJ. CRY SOL—a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr*. 1995; 28: 768–773. <https://doi.org/10.1107/S0021889895007047>
47. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys J*. 2013; 105: 962–974. <https://doi.org/10.1016/j.bpj.2013.07.020> PMID: [23972848](https://pubmed.ncbi.nlm.nih.gov/23972848/)
48. Glass PJ, White LJ, Ball JM, Leparç-Goffart I, Hardy ME, Estes MK. Norwalk Virus Open Reading Frame 3 Encodes a Minor Structural Protein. *J Virol*. 2000; 74: 6581–6591. PMID: [10864672](https://pubmed.ncbi.nlm.nih.gov/10864672/)

2.2 Figures supplémentaires

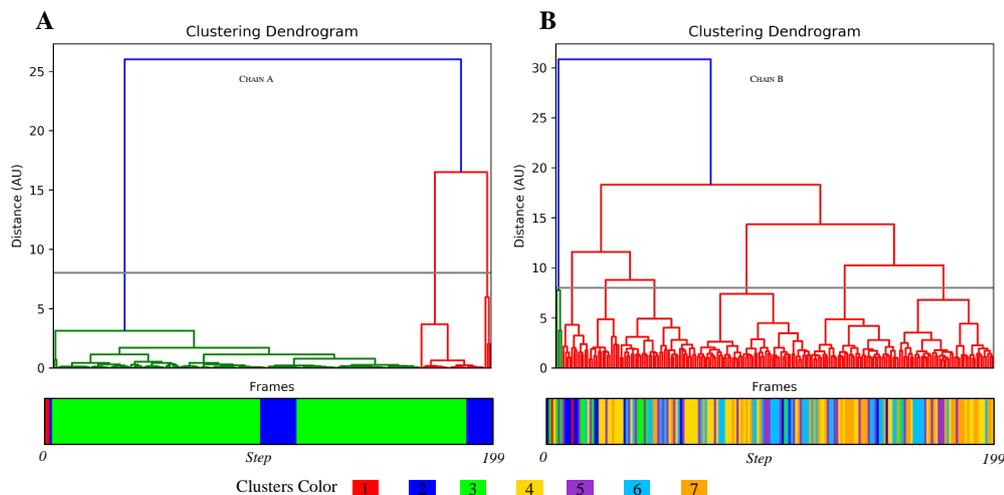


Fig. S. 1: Clustering of N-terminal arm simulated annealing endpoints (A) for chain A and (B) for chain B. Top, clustering dendrograms. The cutoff used to generate clusters from hierarchical classification is represented by a gray bar (8 for both chains). Bottom, each successive frame is assigned its cluster color. Graphics were generated with the TrajectoryClustering program (available at <https://github.com/tubiana/TTclust>).

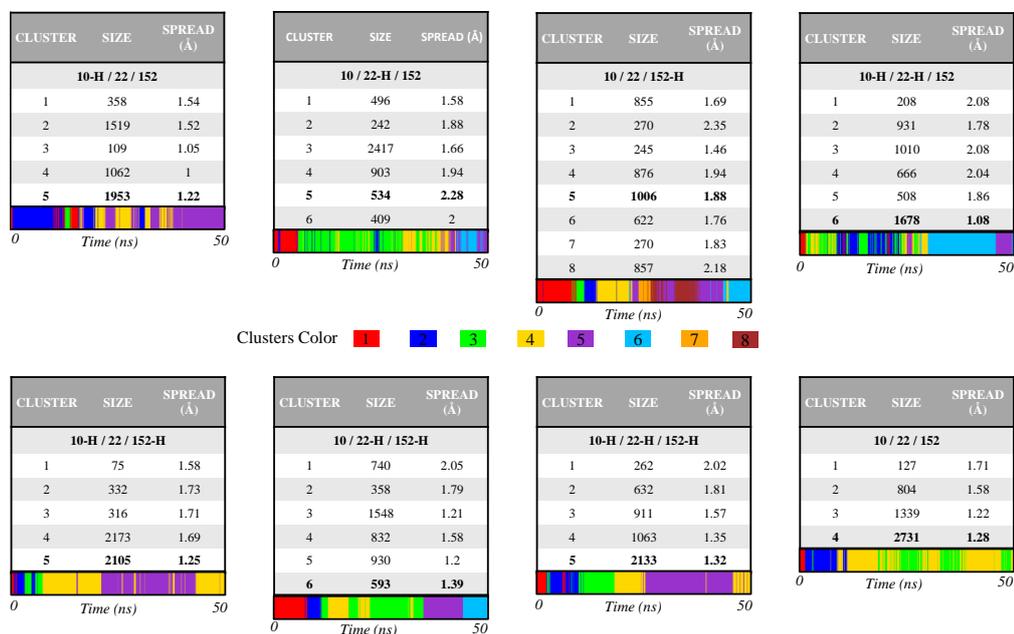


Fig. S. 2: Clustering of N-terminal arm conformations along simulations with different protonation states. For every cluster, the size (number of frames in the cluster) and the spread (average rmsd between all cluster's frames) are given. The number of the cluster depends on its order of appearance in the trajectory. A color code was attributed to each cluster.

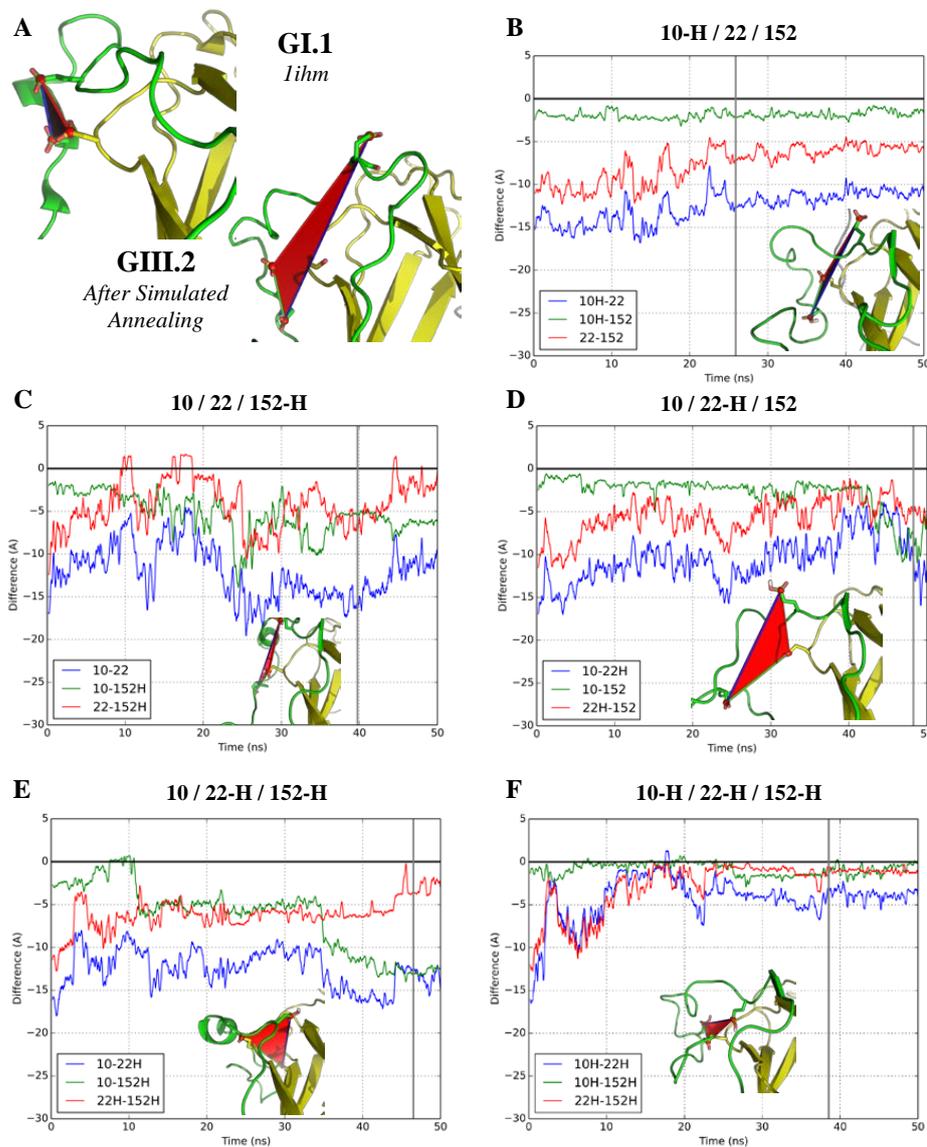


Fig. S. 3: Molecular dynamics simulations in different sets of protonation for the three acidic residues clustered in the GI.1 VP1. As for Fig. 3, (A-F) Evolution of the difference distances between D10, E22 and E152. A value of 0 corresponds to restoration of the initial distance in the GI.1 structure while a negative value indicates a larger distance (see text for details). The illustrations represent the geometry of a representative frame according to the clusterisation of each trajectory (S2 Fig). The gray vertical line indicates the frame from which the picture originates.

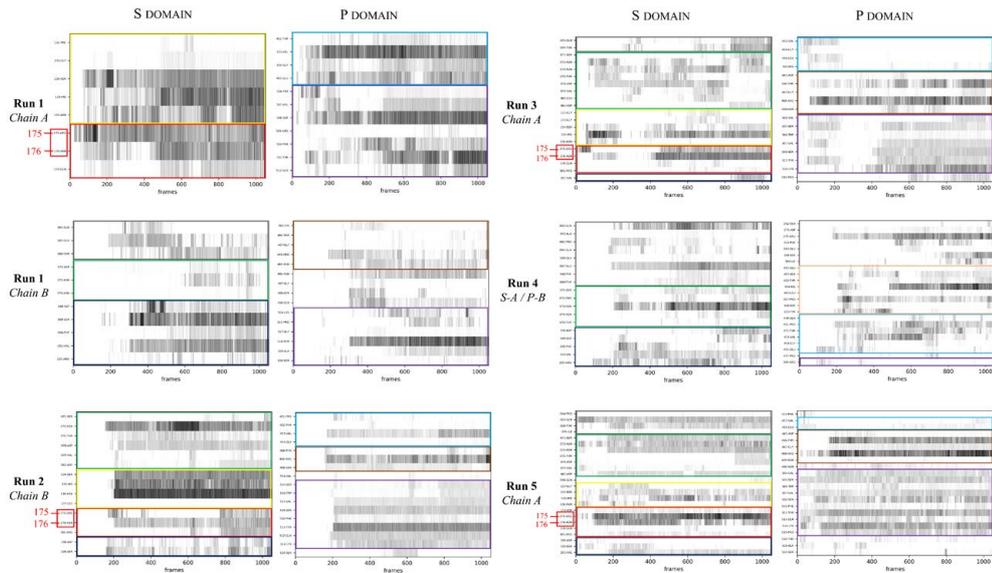


Fig. S. 4: Timelines of S domain interactions with P domain for simulations from an extended form of the GI.1 dimer. Residues inside a rectangle with the same colour belong to a single segment as defined in Fig. 6. Interactions were sampled every 100 ps with VMD [31] and graphics made with matplotlib [32].

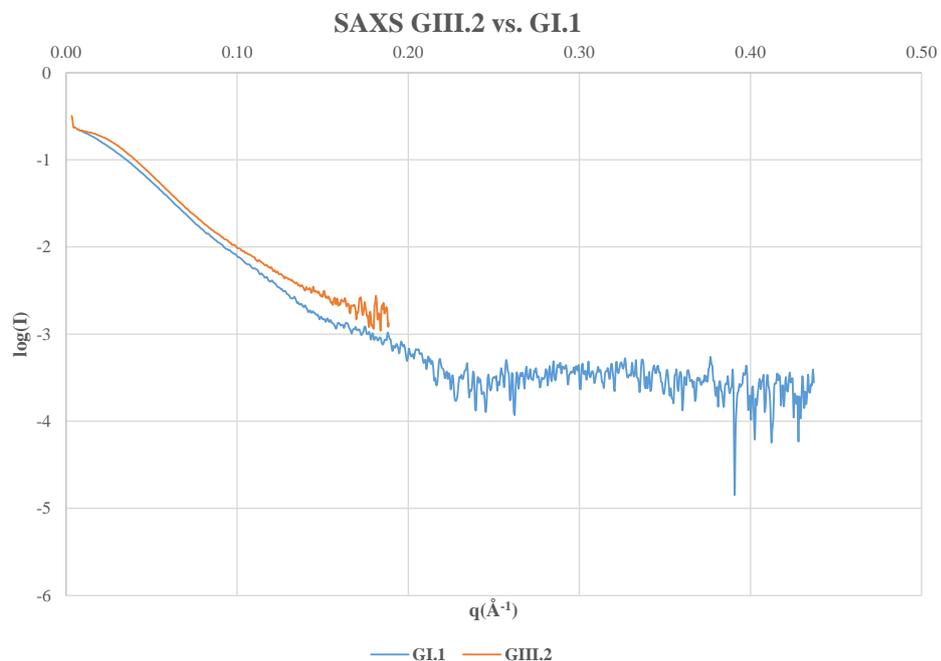


Fig. S. 5: Comparison of SAXS results of the GI.1 (blue) and GIII.2 (orange) VP1 dimer.

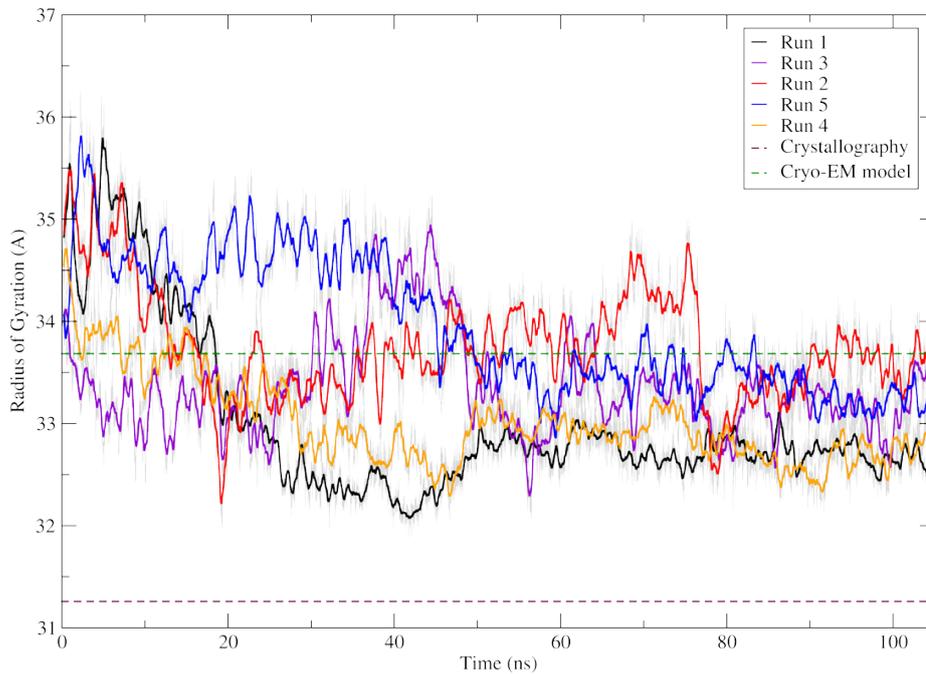


Fig. S. 6: Timelines of radii of gyration during simulations from an extended form of the GI.1 dimer. Values for the crystallographic model and our cryo-EM derived model (without termini) are indicated.

<https://doi.org/10.1371/journal.pone.0182056.s007>

VID. S. 1: Simulated annealing trajectory. Representation as in Fig 2C, with the N-terminal arms of chains A and B in green and emerald, respectively. Rendered with VMD.

<https://doi.org/10.1371/journal.pone.0182056.s008>

VID. S. 2: 50 ns molecular dynamics simulations in three different sets of protonation for the three acidic residues clustered in the GI.1 VP1. The protonation states and representation are the same as for Fig 3.

<https://doi.org/10.1371/journal.pone.0182056.s009>

VID. S. 3: 105 ns simulations of the extended GI.1 dimer. The 5 replicas are displayed as in Fig 4.

Complément à l'article

3.1 Simulation d'un monomère

3.1.1 Contexte

Nos simulations ont quasiment toutes été effectuées sur le dimère de VP1 (conformation A/B). Seules les simulations portant sur l'étude de la protonation de trois résidus acides sur le bras N-terminal ont été réalisées sur une forme monomérique. Nous devons donc ajouter à cette étude une analyse du monomère de VP1 afin de déterminer si celui-ci reste stable durant la dynamique ou non.

Nous avons décidé d'initier les calculs de DM avec la conformation C du monomère car c'est pour celui-ci que l'angle formé entre le domaine S et le domaine P est le plus important (voir [VP1](#), Fig. 1.13). Par conséquent, c'est cette structure qui devait présenter la plus grande amplitude de mouvement. Éventuellement, nous voulions également voir si la conformation C peut se stabiliser en une conformation B ou A au cours du temps.

3.1.2 Préparation

Les régions N-ter (résidus 1→28) et C-ter (521→530) ont été rajoutées avec Modeller 9.16 [112] et une prédiction de protonation à pH9 a été réalisée afin d'être en accord avec les expériences de SAXS (à pH9, VP1 est sous forme dimérique). Le protocole de simulation est le même utilisé que pour les dimères (cf. part. 1.4.1).

3.1.3 Étude de la stabilité d'un monomère C

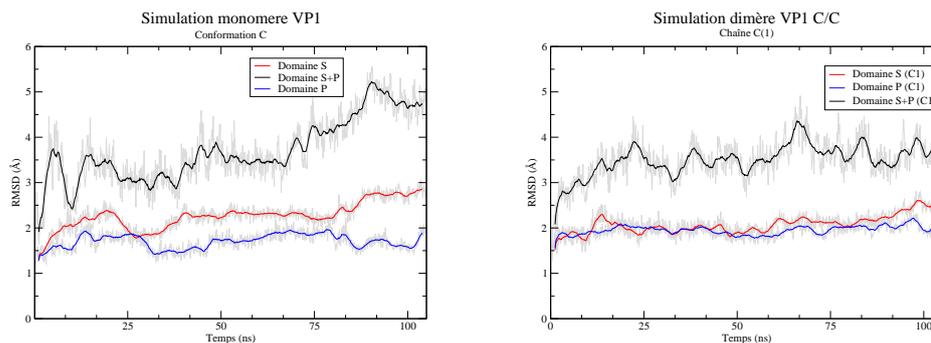
3.1.3.1 Analyse du RMSD

Après un période de production de 100 ns, nous avons calculé les RMSD des domaines S, P et de la protéine sans les régions N-ter et C-ter (Fig. 3.2a). À la fin de la période de production, nous remarquons que la structure du monomère n'est pas encore totalement stable. Au cours de la simulation, chaque domaine apparaît cependant stable indépendamment. On observe donc que les deux domaines structuraux sont stables mais qu'ils peuvent bouger, l'un par rapport à l'autre, grâce au segment interstitiel qui est flexible. Dans le cas du dimère CC, là encore, chacun des domaines S et P des deux monomères reste stable au cours de la simulation mais on remarque également que chacun des monomères est stabilisé sous la forme dimérique. Ceci reflète les interactions inter-domaines P-P et S-S de la forme dimérique qui limitent les mouvements observés sous la forme monomérique.

3.1.3.2 Analyse de l'angle formé entre le domaine S et P

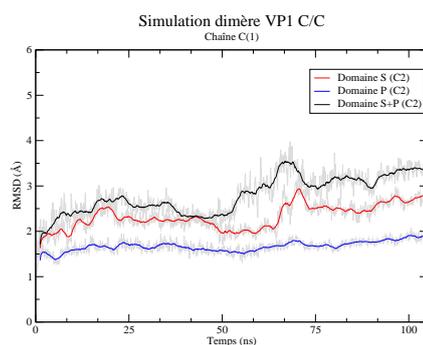
Nous avons également regardé l'angle formé entre le domaine S et l'axe Z. Pour cela, nous avons aligné l'axe principal de la conformation A (cristallographique, sans les régions N et C-terminales) sur l'axe Z (cf. Fig. 3.3a) grâce à l'outil Orient de VMD (http://www.ks.uiuc.edu/Research/vmd/script_library/scripts/orient/). Afin de garder la même référence, les conformations B et C de la structure cristallographique ainsi que les trajectoires du monomère C et de chaque monomère du dimère C/C (C1 et C2) ont été alignées sur le domaine P de la conformation A préalablement alignée sur l'axe Z.

L'angle des monomères C durant la MD est calculé par rapport à l'axe Z grâce à un outil décrit dans la partie 3.3 (page 83) et comparé avec les valeurs des conformations cristallographiques A, B et C (87.4°, 88.1° et 94.8° respectivement).



(a) RMSD - monomère VP1 (conf. C)

(b) RMSD - Dimère C/C VP1 (C1)



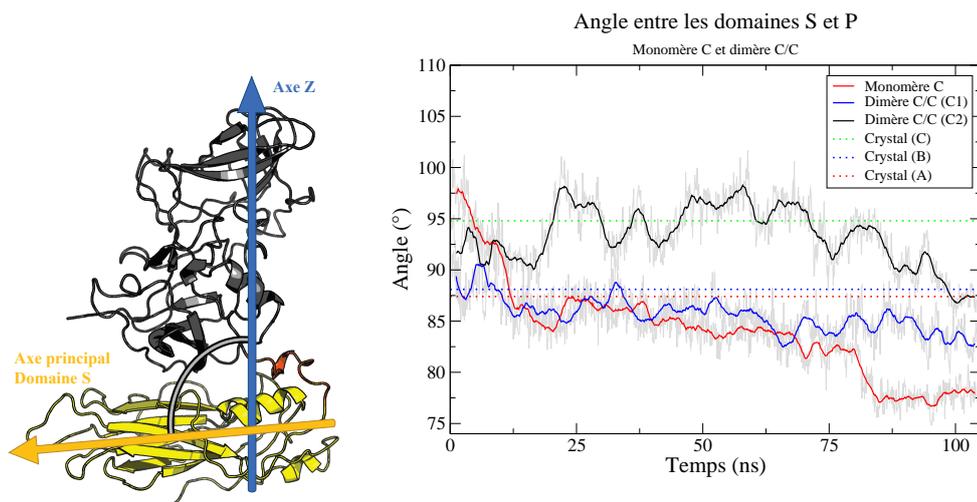
(c) RMSD - Dimère C/C VP1 (C2)

Figure 3.2: RMSD durant la dynamique du monomère C (3.2a) ainsi que d'un dimère C/C. Le RMSD a été calculé sur le domaine S (rouge), domaine P (bleu) et les résidus 30 à 520 (noir). Les domaines ont été alignés sur eux-mêmes avant le calcul des RMSD.

L'angle formé entre le domaine S et le domaine P diminue tout au long de la dynamique (Fig. 3.3b). Il évolue de 95° à 77° en 100 ns. Il est à noter que l'angle final obtenu en fin de simulation est largement inférieur à celui mesuré pour les conformations A, B et C du monomère.

3.1.3.3 Analyse en composante principale

Une [Analyse en Composante Principale*](#) a été réalisée sur les résidus 30 à 520 (carbones alpha uniquement) tout au long de la dynamique (Fig. 3.4). Cette méthode est utilisée pour extraire des informations sur la collectivité des mouvements des atomes. 10 modes ont été calculés avec ProDy [169] et visualisés avec VMD [122]. Bien que la plupart de ces modes représente des



(a) Représentation de l'angle calculé dans la figure 3.3b (b) Angle formé entre l'axe Z et le domaine S durant la dynamique

Figure 3.3: Angle formé entre l'axe Z du monomère et l'axe principal du domaine S. La protéine a été auparavant alignée sur l'axe Z. Les valeurs d'angle mesurées sur les conformations A, B et C au sein du cristal sont également indiquées.

vibrations minimales au sein de la protéine, 3 ont retenus notre intérêt. Nous pouvons tout d'abord observer sur le mode 1 (Fig. 3.4A) des mouvements en accord avec les angles calculés au long de la dynamique (Fig. 3.3b) : il y a une rotation du domaine S qui va venir se rabattre sous la spicule. Les modes 7 et 8 sont très semblables : ils montrent une flexibilité des acides aminés au niveau de l'interface dimérique du domaine S. Nous n'avons pas observé de flexibilité important au niveau de l'interface dimérique du domaine P. Ceci semble indiquer que le domaine S est plus stable sous forme dimérique.

3.1.4 Conclusion de la simulation du monomère C

Nous avons pu voir avec l'analyse par RMSD que le monomère C n'est pas totalement stable après 100 ns de production. Ceci est provoqué par le segment interstitiel qui permet à chaque domaine S et P d'évoluer l'un par rapport à l'autre (Fig. 3.2a). Les résidus à l'interface dimérique du domaine S semblent bien plus flexibles. Cette flexibilité n'est plus observée dans la

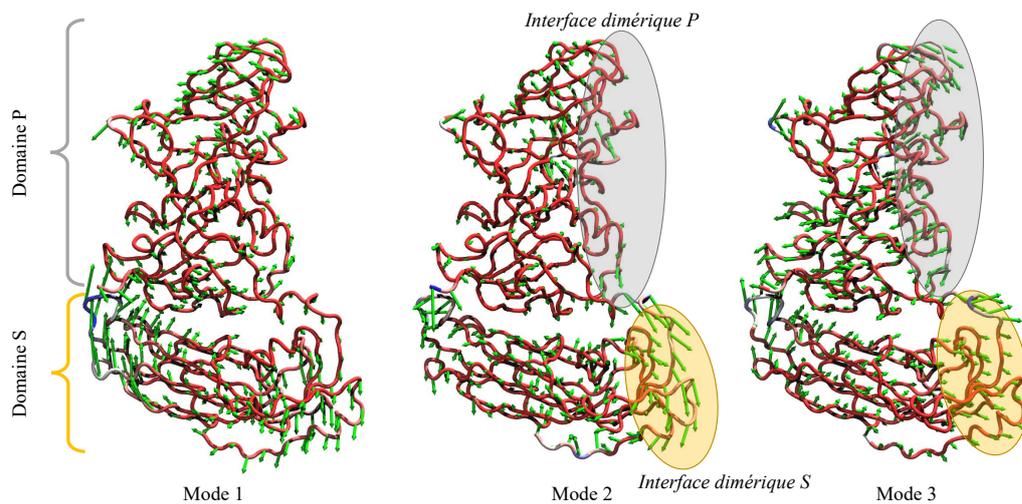


Figure 3.4: Résultats d'une ACP sur la dynamique du monomère C. 10 modes ont été calculés, uniquement les modes 1,7 et 8 sont représentés. La direction des atomes dans les modes est représentée par des vecteurs (flèches vertes) à double sens (Bien que les flèches montrent qu'une direction). La flexibilité des atomes est indiquée par un dégradé de rouge (peu flexible) à bleu (très flexible).

simulation du dimère C/C. De plus, les domaines S ne vont pas se rabattre contre la spicule. Ils sont donc plus stables en forme dimérique.

Sous forme monomérique, le domaine S est donc beaucoup plus flexible mais en restant structuré : c'est tout le domaine qui va bouger d'un bloc. Malgré la faible surface d'interaction entre deux domaines S d'un dimère, cette interaction semble suffisamment forte pour assurer une certaine stabilité des domaines S. Nous savons que la dimérisation de VP1 s'effectue par les domaines P [56, 65]. Nous pouvons donc émettre l'hypothèse à la vue de ces résultats qu'une fois que les domaines S interagissent ensemble, l'angle et la distance au domaine P va s'ajuster pendant la formation de la capsid afin de correspondre aux critères géométriques d'une forme icosaédrique.

3.2 Modélisation de la capsid de norovirus de GIII.2

Depuis l'écriture de l'article, de nouvelles cartes de cryo-EM ont été obtenues par l'équipe de microscopie de Jean Lepault en collaboration avec notre équipe. Julien Maufront lors de son Master 2 a ainsi obtenu trois nouvelles cartes de la capsid de norovirus de GIII.2, une à 20 Å, une à environ 9.5 Å et une à 8.9 Å :

- La carte avec une résolution de 20 Å a été obtenue avec 10 mM d'EDTA, un chélateur d'ions divalents. Cette basse résolution semble être révélatrice d'une déstructuration de la capsid. En effet, l'enveloppe reste peu contrastée et les spicules très élevées par rapport au domaine S.
- La carte avec une résolution de 9.5 Å a été obtenue avec 0.5 mM de Ca^{2+} . Cette résolution permet d'obtenir certains détails : l'orientation des domaines P est différente de la structure cristallographique (PDB 1IHM [56]), et similaire à la structure cryo-EM disponible (EMDB 5374 [170]) : une rotation de 40 à 45° dans le sens horaire ainsi qu'une surélévation.
- La carte avec une résolution de 8.9 Å a été obtenue avec 10 mM de Ca^{2+} . La densité obtenue montre un accord avec la structure cristallographique quant à la hauteur et l'orientation des domaines P. De plus, le domaine S semble avoir une meilleure résolution, de l'ordre de 6 Å.

Ainsi, Julien Maufront et Jean Lepault ont montré que la capsid de norovirus de GIII.2 pouvait adopter deux conformations, une conformation étendue et une conformation compacte.

En janvier 2016, la structure du domaine P pour le GIII.2 a été co-cristallisée avec des Zn^{2+} a été résolue, 5E6T [171]. À partir de cette structure et d'une approche de modélisation par homologie, j'ai réalisé une reconstruction atomique de la capsid de Norovirus de GIII.2 basée sur la nouvelle carte de densité obtenue avec une résolution de 8.9 Å.

3.2.1 Modélisation des domaines

Bien que le domaine P soit résolu par cristallographie (PDB 5E6T[171]), le domaine S seul ne l'est pas. J'ai donc tout d'abord réalisé un modèle du dimère de GIII.2 avec pour références le domaine S du GI.1 (PDBid 1IHM[56]) et le domaine P du GIII.2 (PDBid 5E6T [171]) avec le serveur SWISS-MODEL[105]. La modélisation a été réalisée avec les références décrites dans la table 3.1. Pour réaliser la structure référence, hybride des deux structures cristallographiques, les domaines S d'1IHM et domaines P de 5E6T ont été alignés sur l'unité asymétrique d'1IHM (pour les chaînes A, B et C).

Table 3.1: Références cristallographiques pour chaque région modélisée de VP1 GIII.2. La numérotation du modèle et de 5E6T est celle du GIII.2, la numérotation d'1ihm est celle du GI.1.

RÉGION MODÉLISÉE	RÉFÉRENCE
Domaine S (A) 25 → 212	1IHM (A) 29 → 219
Domaine S (B) 10 → 212	1IHM (B) 12 → 219
Domaine S (C) 25 → 212	1IHM (A) 29 → 219
Domaine P (AC) 220 → 504	5E6T (A) 220 → 504
Domaine P (B) 220 → 506	5E6T (B) 220 → 506
Région interstitielle (ABC) 213 → 219	Denovo

SWISS-MODEL produit ensuite trois modèles distincts basés respectivement sur les chaînes A,B et C. Le modèle de la chaîne A a produit une hélice dans la région 206 → 213. Par souci de symétrie avec la chaîne B et de flexibilité dans la région interstitielle, cette hélice a été déstructurée avec la fonction Sculpting de PYMOL.

L'unité asymétrique du GIII.2 de VP1 a été ensuite reconstruite en regroupant ces trois modèles.

3.2.2 Alignements dans la carte de densité

VEDA[<https://goo.gl/jnFdv8>] (*Visual Environment for Docking Algorithms*) est un environnement visuel développé pour adapter les modèles atomiques interactivement aux reconstructions 3D. Les calculs sont effectués dans l'espace réciproque et la symétrie de la reconstruction est prise en compte. VEDA est donc très adapté à la reconstruction de capsides virales (cf. Fig. 3.5).

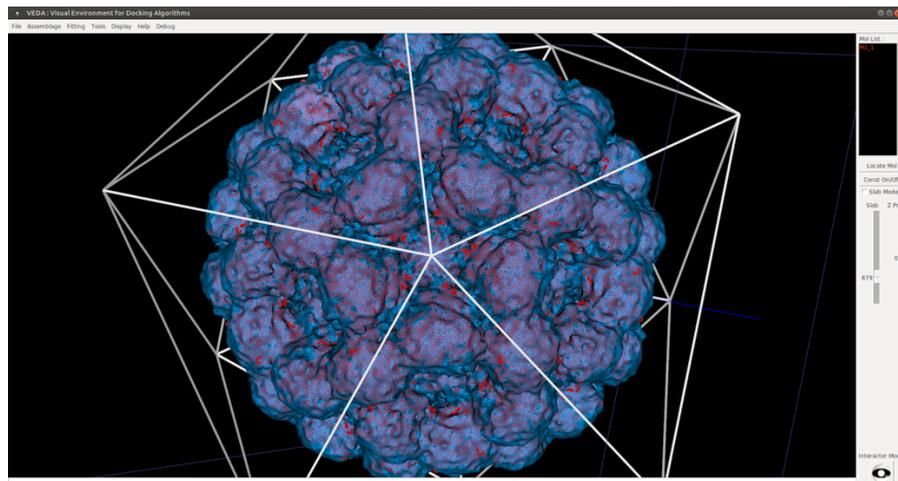


Figure 3.5: Interface utilisateur de VEDA (Visual Environment for Docking Algorithms). Le maillage bleu correspond à la densité électronique de la carte de cryo-microscopie et en rouge à l'intérieur se trouve un modèle atomique de la capside de GIII.2 de norovirus.

Il a d'abord fallu trouver dans laquelle des 26 orientations de l'icosaèdre était la carte de microscopie électronique. C'est l'orientation 2Z2Y.2 qui a été appliquée au vu de la correspondance des éléments de symétrie, notamment les axes 5 (Fig. 3.5) L'unité asymétrique a ensuite été ajustée à l'intérieur de la densité. Un affinement a été réalisé par la suite avec une résolution limite de 6 Å à 8.9 Å afin de retirer un maximum de clashes stériques formés. Cependant, comme le domaine S a été modélisé depuis la structure cristallographique du GI.1 (1IHM), certains clashes persistaient entre les domaines S au niveau des axes 6. Nous les avons donc supprimés avec l'outil de minimisation géométrique de PHENIX[172].

La capside a ensuite été reconstruite en appliquant les 60 matrices de la symétrie icosaédrique sur l'unité asymétrique. Cette étape consiste à rajouter des REMARK 350 dans l'entête du fichier PDB. Ces REMARK contiennent pour chaque chaîne des vecteurs de transformation qui permettront au logiciel de

traitement de copier la chaîne dans une nouvelle position. L'ensemble des chaînes copiées représente la capsidie complète.

3.2.3 Ajustement avec MD-FLEXIBLE FITTING

Nous avons voulu par la suite ajuster le modèle de la capsidie obtenue dans la carte de cryo-microscopie électronique. Nous avons utilisé l'approche MDFF (voir [Ajustement par dynamique moléculaire flexible](#) page 57) dans le vide afin de limiter les fluctuations thermiques et augmenter la flexibilité du système.

Nous avons tenté d'améliorer la structure obtenue par VEDA en testant plusieurs conditions de flexibilité. Nous avons utilisé des restrictions de symétries basées sur la capsidie complète (voir [Ajustement par dynamique moléculaire flexible](#) page 57)

3.2.3.1 Toute l'unité asymétrique flexible

Dans un premier temps nous avons testé l'approche MDFF sur l'unité asymétrique sans contrainte, tout le système était flexible. Nous avons constaté une déstructuration rapide de la capsidie (cf. Fig. 3.6). En effet, la résolution de la carte n'est pas suffisante pour permettre à MDFF un bon ajustement.

Cependant, les structures secondaires restent assez conservées entre la structure initiale et après 50000 étapes de simulation par MDFF (cf. Fig. 3.7). Ceci montre que ce n'est pas un "débobinement" des protéines, mais les boucles ont plutôt tendance à être étirées, entraînant les blocs protéiques structurés avec eux. Le même comportement a été observé en fixant le domaine S et en laissant le domaine P flexible.

Comme décrit dans la partie [VP1](#) (page 17 (Fig. 1.11)), le domaine P est composé en deux sous domaines : le domaine P1 et le domaine P2 qui est moins ordonné que le reste de la protéine. C'est en effet ce domaine qui va avoir tendance à s'étirer, entraînant une déformation de la protéine. Il est donc plus difficile d'utiliser des méthodes d'ajustement flexibles par

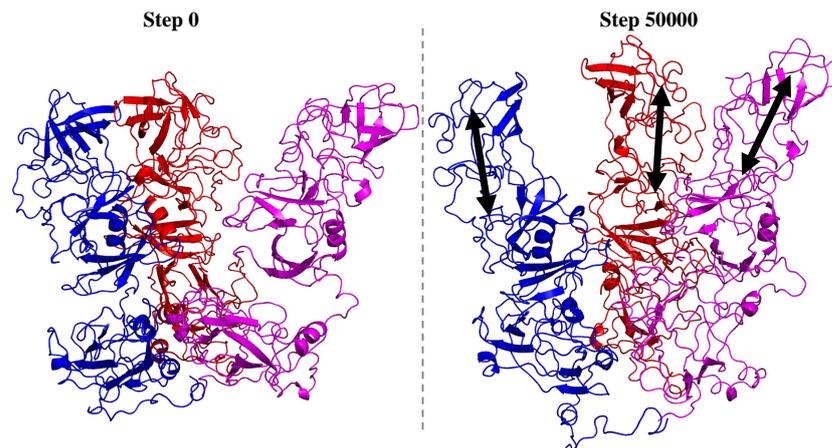


Figure 3.6: Comparaison entre l'unité asymétrique du GIII.2 de la capside de Norovirus avant MDFF et l'unité asymétrique après 50000 pas de simulation par MDFF. Toute l'unité asymétrique est flexible. La chaîne A est représenté en rouge, la B en bleu et la C en magenta.

simulation de dynamique moléculaire sur le domaine P dans une carte de densité de faible résolution sans entraîner une déformation importante.

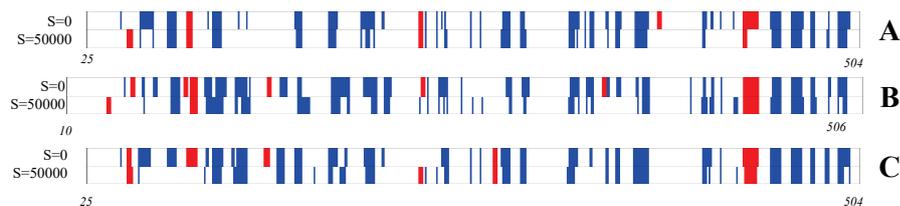


Figure 3.7: Conservation des structures secondaire à $S = 0$ et $S = 50000$ pendant la simulation par MDFF pour l'unité asymétrique du GIII.2 complètement flexible. Les hélices sont en rouge, les brins en bleu et le coil en blanc.

3.2.3.2 Domaine S flexible

Étant donné la plus faible résolution au niveau des spicules, nous avons décidé de laisser le domaine P rigide et le domaine S flexible dont la densité était de meilleure résolution (6.0 Å). Bien que les structures secondaires soient majoritairement conservées (Fig. 3.8B), nous observons une fois de plus des changements majeurs suite aux 50000 étapes de simulation par MDFF : la face supérieure et inférieure du domaine S vont s'éloigner formant ainsi un "vide" au coeur du domaine S (écartement des deux feuillets constitutifs du tonneau bêta, Fig. 3.8A).

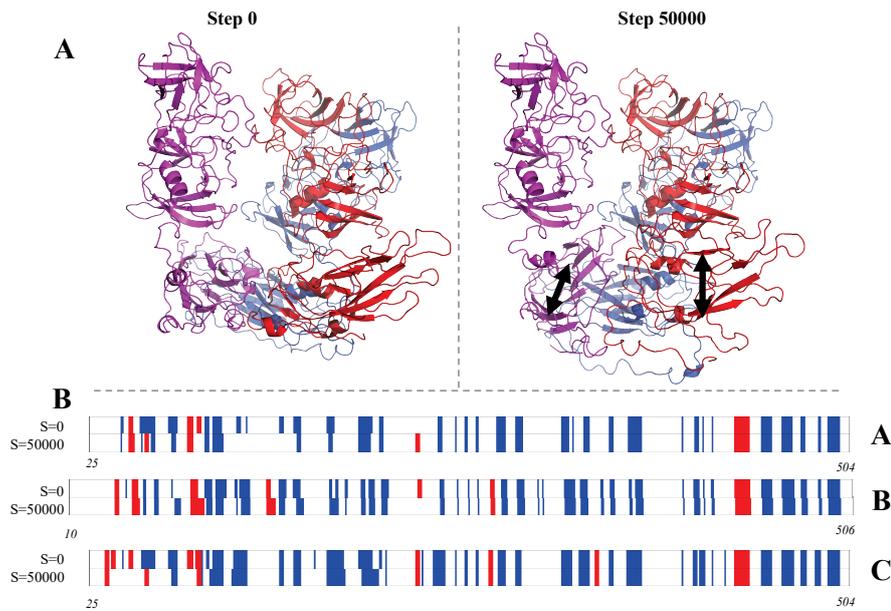


Figure 3.8: Idem que pour les figures 3.6 et 3.7 mais avec uniquement le domaine S flexible.

3.2.3.3 Coil du domaine S flexible

Finalement nous avons décidé de garder uniquement les acides aminés du domaine S, qui n'étaient pas impliqués dans des hélices ou des brins, flexibles (cf. Fig. 3.9). Ceci nous permet de réaliser un ajustement uniquement sur les boucles flexibles. Le nombre d'étapes a été augmenté à 500000 afin de laisser plus de temps aux boucles de s'ajuster dans la densité électronique.

Les résultats de cet affinement (cf. Fig. 3.10) montrent une faible modification de la position des boucles indiquant ainsi que seul un affinement local a été réalisé. Les structures secondaires n'ont pas été altérées pendant ces étapes de simulation. Un calcul du coefficient de corrélation dans CHIMERA entre l'unité asymétrique à $s = 0$ et $s = 500000$ avec une densité simulée de 6 Å montre une légère amélioration de 0.8456 à 0.8531. Finalement, cette amélioration est très minimale. Avec une carte de cette résolution et une symétrie complexe présente dans la capsid, MDFFF éprouve des difficultés à ajuster le modèle dans la carte.

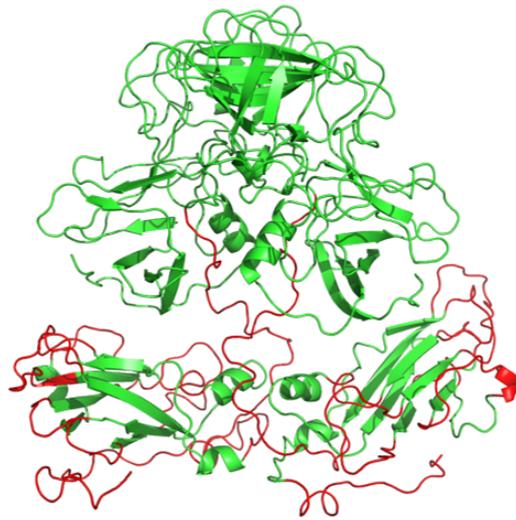
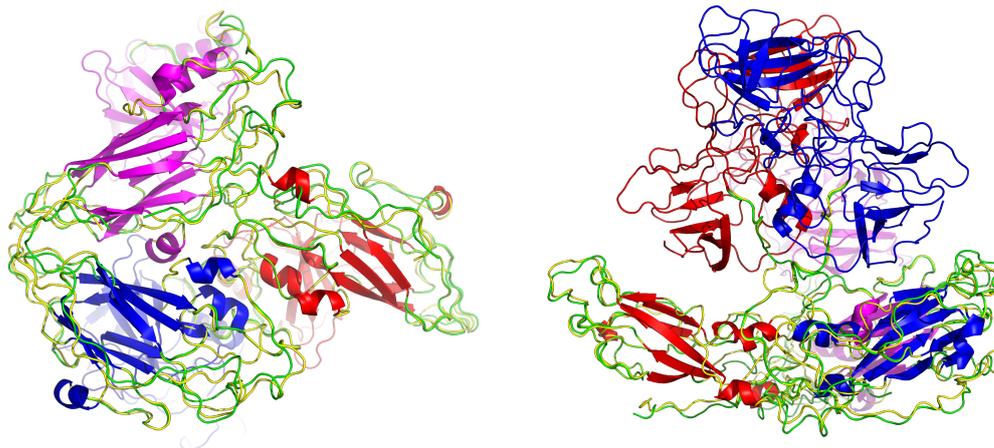


Figure 3.9: Parties flexibles (rouge) et rigides (vert) utilisées dans le protocole de MDFF.

3.2.4 Reconstruction atomique et améliorations possibles

Nous avons pu voir que la résolution de la carte de densité électronique (6 Å à 8.9 Å) n'est pas suffisante pour permettre à MDFF d'effectuer un bon ajustement du modèle dans la densité électronique. Nous pourrions penser que cela est dû spécifiquement au domaine P qui est moins bien défini que S, mais nous observons le même résultat pour le domaine S bien que la résolution de la carte dans cette région soit meilleure. Nous avons néanmoins été en mesure d'affiner la position des boucles du domaine S en limitant les régions flexibles à celles-ci et produire un modèle de la capsid du norovirus de GIII.2 (Fig. 3.11). Nous pouvons observer autour des axes 5 et quasi-6 (Fig. 3.11AB) que les boucles se sont bien ajustées à l'intérieur de la densité électronique, malgré une perte de symétrie exacte pour les chaînes latérales de certains acides aminés.

En comparant les capsides de GI.1 et de GIII.2 (Fig. 3.11D) nous pouvons voir que la capsid de GIII.2 est légèrement plus grande que la structure cristallographique (d'environ 1.2 nm). Cependant, contrairement à la capsid de norovirus de GII.10 (plus grande d'environ 2.5 nm par rapport au GI.1), les spicules sont dans la même orientation.



(a) Vue du dessous

(b) Vue de côté

Figure 3.10: Comparaison entre la structure de l'unité asymétrique avant (vert) et après (jaune) 500000 pas de simulation de MDFF. Les structures secondaires et domaine P étant fixées, elle ont été colorées par chaîne : A-rouge, B-bleue, C-magenta.

Ces résultats peuvent encore être améliorés, du fait que la résolution des images utilisées pour la reconstruction de la carte de densité électronique ait été divisée par deux pour des raisons calculatoires, ainsi que par l'utilisation d'autres méthodes. AMBER intègre depuis la version 14 un mode capable d'ajuster une structure dans une carte de densité. Nous pouvons également penser à une méthode comme BIOSPRING utilisant une approche linéaire combinée à des réseaux élastiques. Cette dernière méthode, développée initialement pour le SAXS pourrait permettre d'ajuster les domaines P à l'intérieur d'une densité électronique de faible résolution.

La plus faible résolution au niveau du domaine P dans les cartes de densité électronique de cryo-microscopie indique une forte flexibilité de ce domaine. Cette flexibilité pourrait nuire à la cristallisation de la forme complète de VP1, sous forme dimérique ou associée en capsid.

3.3 Simulation de Dynamique moléculaire dirigée - Gl.1

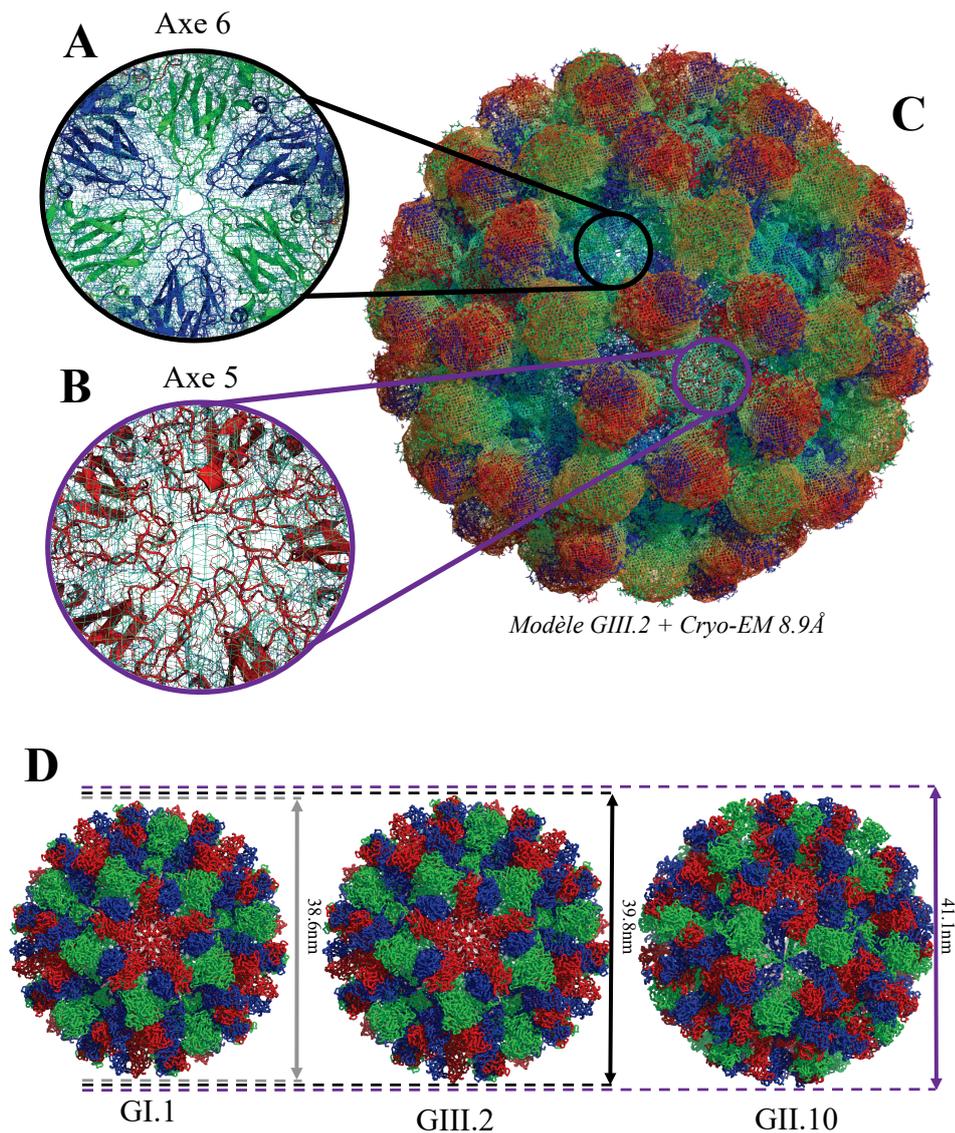


Figure 3.11: Résultat de la modélisation de la capside de GIII.2 à partir de l'échantillon le plus ordonné et la meilleure carte de cryo-EM obtenue par Julien Maufront et Jean Lepault. (A-B) Résultat de l'affinement du domaine S dans les axes 5 et 6. (C) Représentation atomique de la capside de GIII.2 dans sa densité électronique. (D) Comparaison entre les capsides du GI.1, GIII.2 et du GII.10. Les chaînes A sont en rouge, les chaînes B en bleu et C en vert. La densité est colorée avec une coloration radiale de bleu (densité plus proche du centre de la capside) à rouge (densité la plus éloignée du centre de la capside)

3.3.1 Contexte

La capside du norovirus est donc capable de passer d'une conformation compacte à une conformation étendue. Naturellement nous voulions savoir

comment, et surtout pourquoi la forme compacte du génogroupe I.1 était si stable. Dans nos travaux sur le dimère [168], nous avons mis en évidence certains acides aminés du GI.1 peu conservés, mais participant au maintien du domaine P au domaine S. Nous avons ainsi pu définir une zone de contact entre les deux domaines avec des groupes d'acides aminés (cf. [Article Fig. 6 et 7C](#)). Nous avons voulu, par la suite, affiner ce résultat et tenter de déterminer si seulement quelques acides aminés pouvaient être responsables du maintien de la conformation compacte du génogroupe I.1.

La simulation de dynamique moléculaire dirigée (TMD) était donc un outil de choix pour tenter de quantifier l'impact que pouvait avoir certaines mutations dans la transition de la conformation compacte à la conformation étendue.

3.3.2 Réalisation des deux modèles

Ce travail a été réalisé avec le programme AMBER 14[173] sur le supercalculateur OCCIGEN (CINES).

Nous avons pris la structure cristallographique 1IHM pour la forme compacte du dimère de VP1. La forme étendue est le modèle de GI.1 basé sur la carte de Cryo-EM du GII.10 (EMDB 5374) tel que décrit dans l'article [168] (voir [Dynamics and asymmetry in the dimer of the norovirus major capsid protein](#) page 89). Deux systèmes sont préparés pour chaque conformation : un système sauvage et un système mutant. Les résidus mutés sont : S134V; H135E; R175Q; Q226S et S508N. Ils ont été choisis en fonction de leur forte implication dans l'interaction entre les domaines S et P (cf. Fig. S4 article), leur faible conservation et plus particulièrement le fait qu'ils diffèrent entre le génogroupe I et les autres (Cf. Fig. 7 Article) et le changement physico-chimique induit par la mutation. La mutation correspond au résidus consensus identifié suite à l'alignement de séquences représentatives de chaque génogroupe/génotype (voir [Dynamics and asymmetry in the dimer of the norovirus major capsid protein](#), partie "Sequence conservation")

Chaque système a ensuite été neutralisé avec des ions Na⁺ (28 pour le système sauvage et 31 pour le système avec mutation) et hydraté avec de l'eau TIP3P [136] dans tleap. Afin de réaliser une TMD dans AMBER, il faut

que le système de départ et le système cible soient décrits par le même fichier de topologie (l'enchaînement des atomes doit être identique avec le même nombre d'atomes dans la protéine, d'ions, de molécules d'eau). Il faut donc ajuster la taille des boîtes des systèmes, en jouant sur la distance minimum entre le bord de la boîte et la protéine) pour les conformations compactes et étendues de façon à ce que le nombre de molécules d'eau soit proche. Il faut ensuite supprimer les molécules d'eau en surplus dans un des systèmes. La taille des boîtes de simulation ainsi que le nombre de molécules d'eau supprimées sont décrits dans la table 3.2.

Table 3.2: Description des boîtes de simulation pour les systèmes sauvage et mutant : distance minimum entre la protéine et les bords de la boîte et le nombre de molécules d'eau supprimées dans la conformation compacte.

Système	conformation	distance minimum entre protéine et bord de la boîte (Å)	Nombre de molécules d'eau supprimées
Sauvage	Compacte	12.1	13
	Étendue	10	0
Mutant	Compacte	12.3	19
	Étendue	10	0

Les systèmes sont ensuite minimisés en 6 étapes successives de steepest descent de 20000 pas chacune où les contraintes sur les molécules d'eau sont relâchées progressivement.

Pour chaque conformation de chaque système, six simulations de TMD ont été réalisées avec des forces croissantes : 0.005, 0.0075, 0.01, 0.02, 0.04 kcal.mol⁻¹.Å²) avec un RMSD objectif de 0.0 Å calculé sur toute la protéine. L'objectif est de définir la force minimale nécessaire pour passer de la conformation compacte à la conformation étendue et inversement.

3.3.3 Évolution du RMSD sur 1000 ps

Nous pouvons déjà observer l'évolution du RMSD pour le système sauvage. Sur les simulations compactes→étendues (Fig. 3.12A), les conformations compactes arrivent à passer à des conformations étendues avec une force relativement faible, Il semble donc que la forme étendue soit accessible à GI.1.

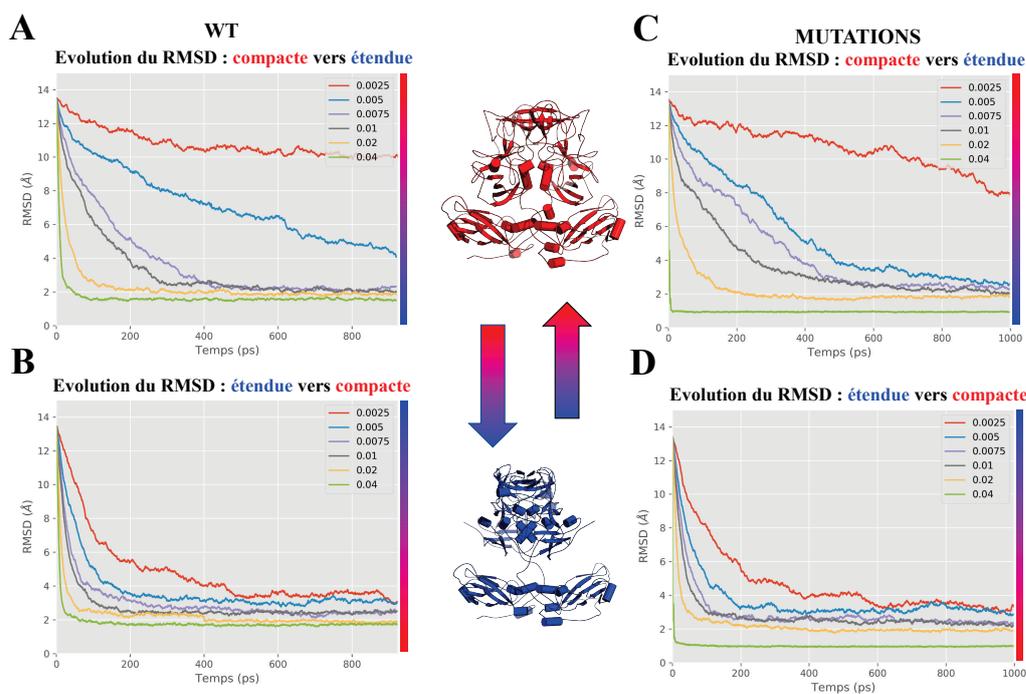


Figure 3.12: Évolution du RMSD entre les simulations compactes→étendues (rouge) et étendues→compactes (bleu) pour la forme sauvage et mutante du dimère de VP1.

Cependant, nous pouvons voir qu'une force de $0.005 \text{ kcal.mol}^{-1}.\text{Å}^2$ (courbe bleue) n'est pas suffisante pour atteindre la conformation compacte en moins de 1000 ps. Bien que la transition soit toujours en train de se produire, il n'y a pas encore de convergence (pas de plateau). Par contre, la force supérieure ($0.0075 \text{ kcal.mol}^{-1}.\text{Å}^2$) est suffisante pour atteindre la forme étendue du dimère de VP1, en 400 ps environ (courbe violette). La force minimale afin de passer de la forme compacte→étendue en 1000 ps est donc comprise entre 0.005 et $0.0075 \text{ kcal.mol}^{-1}.\text{Å}^2$.

Les simulations étendues→compactes (Fig. 3.12B) avec les mêmes forces appliquées montrent une convergence vers la conformation compacte depuis la conformation étendue bien plus rapide que dans le sens compacte → étendue. La forme compacte est donc favorisée puisque la transition étendue → compacte est plus facile que compacte → étendue. S'il y a un équilibre entre les deux conformations du dimère, celui-ci est donc déplacé vers la forme compacte. Si nous appliquons la même observation pour le système mutant (Fig. 3.12C), nous pouvons voir que la transition étendue → compacte (Fig. 3.12D) semble peu modifiée. Cependant, la transition compacte → étendue

(Fig. 3.12C) semble être plus facile : la force minimale se trouve entre 0.0025 et 0.005 kcal.mol⁻¹.Å². De plus, avec une force de 0.04 kcal.mol⁻¹.Å², nous pouvons observer pour la forme mutée une convergence vers la conformation étendue plus rapide (figure 3.12C,D) et plus proche (0.92 Å contre 1.55 Å pour la forme sauvage)

Ces résultats indiquent que la transition de la forme compacte vers la forme étendue est plus exigeante énergétiquement pour la forme sauvage que pour la forme mutée. Les résidus S134, H135, R175, Q226 et S508 ont donc bien un rôle important dans le maintien du domaine S au domaine P.

Ceci permet d'enrichir l'étude réalisée sur le dimère de VP1 et de mieux comprendre sa dynamique. Cela permet également de mieux comprendre pourquoi la capsid du norovirus de GI.1 est la seule structure cristallographique à ce jour.

3.3.4 Perspectives

Les travaux en TMD ont été réalisés très peu de temps avant l'écriture du manuscrit. Aussi, nous n'avons pas eu le temps de faire une étude aussi approfondie que nous l'aurions souhaitée. En effet, Julien Maufront et Jean Lepault ont mis en évidence que la concentration en calcium a un impact fort sur la conformation de la capsid ainsi que sa flexibilité. La structure du domaine P de GIII.2 obtenue récemment montre des ions Zn²⁺ fixés dans des sites de liaisons. Diverses méthodes existent également [174–176] afin de prédire des sites d'ions divalents. Nous pourrions affiner nos simulations de TMD en incluant des ions dans les sites de liaisons possibles afin d'étudier leur influence sur les changements de conformation de la capsid.

Il aurait été également intéressant de réaliser diverses combinaisons de mutations de ces 5 acides aminés afin de vérifier si certaines mutations n'ont pas une incidence négligeable sur la force minimum nécessaire à la transition entre la forme compacte et étendue.

Nous pourrions enfin tester les mutations identifiées sur l'assemblage du GI.1 *in vitro* en comparant, par exemple, des cartes de cryo-EM obtenues pour une forme sauvage et une forme mutante.

Notre travail *in silico* et expérimental publié en 2017 [168] a permis de mettre en évidence que le dimère avait tendance à adopter une forme asymétrique en solution grâce à une région flexible entre les deux domaines S et P.

Afin de compléter cette étude, nous avons réalisé des simulations de MD supplémentaires sur un monomère et le dimère de conformation C (voir [Simulation d'un monomère](#) page 116). Les résultats indiquent que les acides aminés du domaine S, présents à l'interface dimérique sont plus mobiles dans le monomère que le dimère (Fig. 3.4), montrant que cette région est stabilisée quand la protéine est sous sa forme dimérique. Le domaine S interagit avec le domaine P et a tendance à se rabattre contre celui-ci (Fig. 3.3b). Ce mouvement est davantage observé pour la forme monomérique. Cependant, dans le cas du dimère, le rabattement du domaine S reste important car il est similaire, voir supérieur, à celui observé pour la conformation A (la plus aigüe) de la structure cristallographique. Les domaines S ont donc toujours tendance à se replier sur les domaines P une fois qu'il sont sous forme de dimère.

La TMD a permis de montrer qu'il était plus facile de passer de la conformation étendue du dimère à sa conformation compacte, que l'inverse. La mutation de certains résidus peu conservés, identifiés comme favorisant l'interaction entre les domaines S et P durant les MD de la conformation étendue, a permis de faciliter la transition de la conformation compacte → étendu. Certains de ces acides aminés sont uniquement présents dans le GI.1. Ils facilitent donc le maintien du domaine S au domaine P permettant ainsi des interactions plus fortes entre les deux domaines.

Les cartes de densités de cryo-microscopie électronique obtenue par Julien Maufront et Jean Lepault (voir [Modélisation de la capsid de norovirus de GIII.2](#) page 121) montrent une plus faible résolution au niveau des spicules de la VLP de GIII.2. Une concentration plus élevée en calcium permet de rabattre

les spicules et d'obtenir une capsidie plus stable (et ainsi, une meilleure résolution). Un modèle atomique a été réalisé sur la carte de meilleure résolution (8.9 Å voir Fig. 3.11), mais nécessite cependant d'être encore amélioré par d'autres méthodes *in silico* (par exemple la fonction d'ajustement dans des cartes de densité intégrée à partir d' AMBER 14), mais également en utilisant des images non compressées pour le calcul de la densité.

Néanmoins, durant l'acquisition des données de cryo-EM, il a été signalé que la forte flexibilité des spicules était une gêne pour obtenir des données de haute résolution. Le fait que le dimère de P isolé cristallise, mais pas le dimère complet tend à montrer que cette flexibilité est un frein à la cristallisation de la capsidie du norovirus.

Nous émettons donc l'hypothèse que les quelques résidus très peu conservés et impliqués dans les interactions du domaine S au domaine P pourraient favoriser la cristallisation de la capsidie du norovirus et l'obtention de données de meilleure résolution en cryo-EM. Il serait donc très intéressant de réaliser des mutations *in vitro* sur ces résidus sur le GIII.2, en les remplaçant par ceux du GI.1 afin de vérifier expérimentalement si nous arrivons à obtenir des données de meilleure résolution avec les mêmes conditions expérimentales.

En intégrant la totalité des résultats, nous pouvons imaginer un processus dans lequel les spicules au sein de la capsidie se rabattent d'abord sur le domaine S en fonction des conditions physico-chimiques. De fortes interactions se forment alors entre les domaines S et les domaines P. Ces interactions étant les plus fortes pour le GI.1, elles stabiliseraient la capsidie qui serait alors plus stable pour permettre sa cristallisation.

Les résultats de SAXS publiés montrent que les domaines S du dimère en solution sont séparés (Article Fig. 8), et nos résultats sur le monomère C montrent que le domaine S (Fig. 3.4) est stabilisé sous la forme dimérique. Prasad *et al.* ont proposé en 1999 [56] que l'assemblage de la capsidie se ferait d'abord par la formation d'un POD via les domaines S. Les domaines S "flottant", ils pourraient alors plus facilement sonder l'environnement afin de trouver d'autres domaines S avec lesquels interagir.

IV

Étude d'intermédiaire d'assemblage de la capsid du Norovirus.

” *"Mundi placet et spiritus minima", ça n'a aucun sens, mais on pourrait très bien imaginer une traduction du type : "Le roseau plie, mais ne cède... qu'en cas de pépin". Ce qui ne veut rien dire non plus.*

— **Le roi Loth**

Kaamelott, Livre VI, Lacrimosa

Contexte

A la suite du travail sur la brique d'assemblage de la capsid, à savoir le dimère de VP1, nous avons voulu aborder la problématique de la formation de l'intermédiaire identifié par TR-SAXS dans le travail de Tresset *et al* [75]. Cet intermédiaire, de forme allongée, est une structure moléculaire formée de 10 ou 11 dimères. Il apparaît très rapidement (plus de 4 ms mais moins de 200 ms après le déclenchement de l'assemblage) et est suivi par la formation lente de la capsid (à partir de quelques minutes et sur plusieurs heures), sans qu'aucun autre intermédiaire ne soit détectable pendant toute la cinétique d'assemblage. Les auteurs ont proposé pour ce 10-11mer de dimères un modèle composé de deux POD pontés par un dimère interstitiel (??). Cette proposition a l'avantage d'être en accord avec la précédente étude de Prasad *et al.*, dans laquelle les auteurs proposaient à partir de l'analyse de la structure cristallographique de la capsid que le POD serait la base d'assemblage du fait de sa plus grande stabilité comparée au HOD à cause d'interactions plus fortes entre les domaines S (cf. [Modèles d'assemblage du norovirus](#) page 24). Cependant pour Prasad *et al.*, la suite de la formation de la capsid impliquait une croissance isotropique due à la présence de 5 sites de fixation équivalents pour un dimère interstitiel sur le POD symétrique (cf. Fig. 1.19). Cette hypothèse naturelle apparaît à la lecture des résultats de Tresset *et al* incompatible avec l'identification de l'intermédiaire de forme allongée qui ne s'explique qu'avec une croissance dans une direction privilégiée. Au vu de ces résultats, il apparaît que si le POD est bien le premier capsomère à se former en solution, ses propriétés structurales dirigent l'assemblage vers la capsid de façon anisotrope. Un hexamère de dimères (HOD) ne peut pas être totalement exclu à ce stade car l'intermédiaire pourrait correspondre à 10 dimères (20mer), dans ce cas, il n'y a plus de raison de privilégier le POD. J'ai donc cherché à construire un 10mer de dimère allongé basé sur un HOD mais sans succès. En se basant sur l'observation de Tresset *et al.* d'un 10-11mer de dimères allongé comme seule espèce intermédiaire de durée de vie importante, il se fixerait sur le POD ou le HOD un dimère interstitiel qui pourrait rapidement soit amarrer un second POD, soit capturer

successivement 4 ou 5 autres dimères. C'est donc cette hypothèse de travail que nous avons voulu aborder par MD. La taille des complexes devenant importante, nous avons choisi une stratégie gros grains pour construire les structures à simuler.

Nous nous sommes d'abord posé la question de savoir si le POD montrait une plus grande stabilité que le HOD puis si à partir du POD il était possible d'expliquer cette croissance dans une direction privilégiée. Pour la fixation du dimère supplémentaire nous avons utilisé le docking moléculaire.

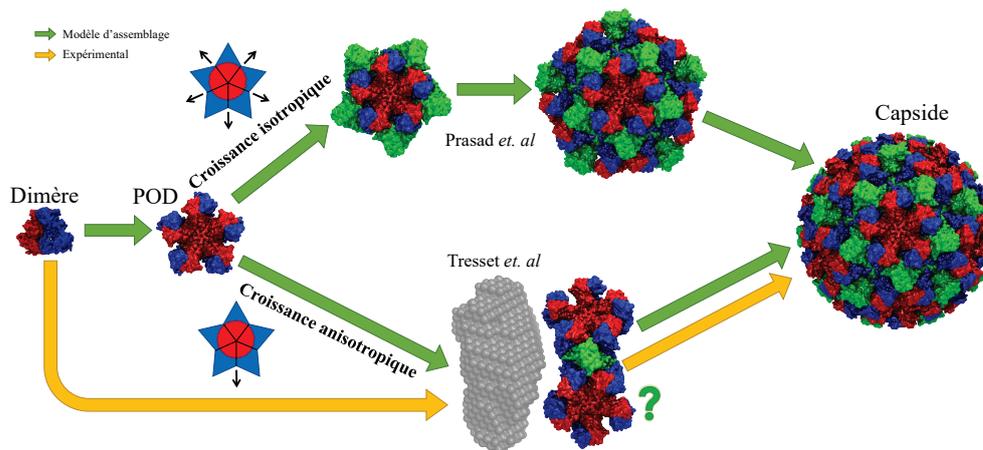


Figure 4.1: Schéma récapitulatif des modèles d'assemblage de Prasad *et al.* [56] et de Tresset *et al.* [75]. Les chemins verts sont des modèles et les chemins jaunes, des transitions observées expérimentalement.

L'étude qui suit a été réalisée par Jean-Charles Carvaillo, stagiaire du Master 2 de Bioinformatique de Paris 7, sous ma supervision. Les résultats (simulation et graphiques) ont été produits et réalisés par ses soins.

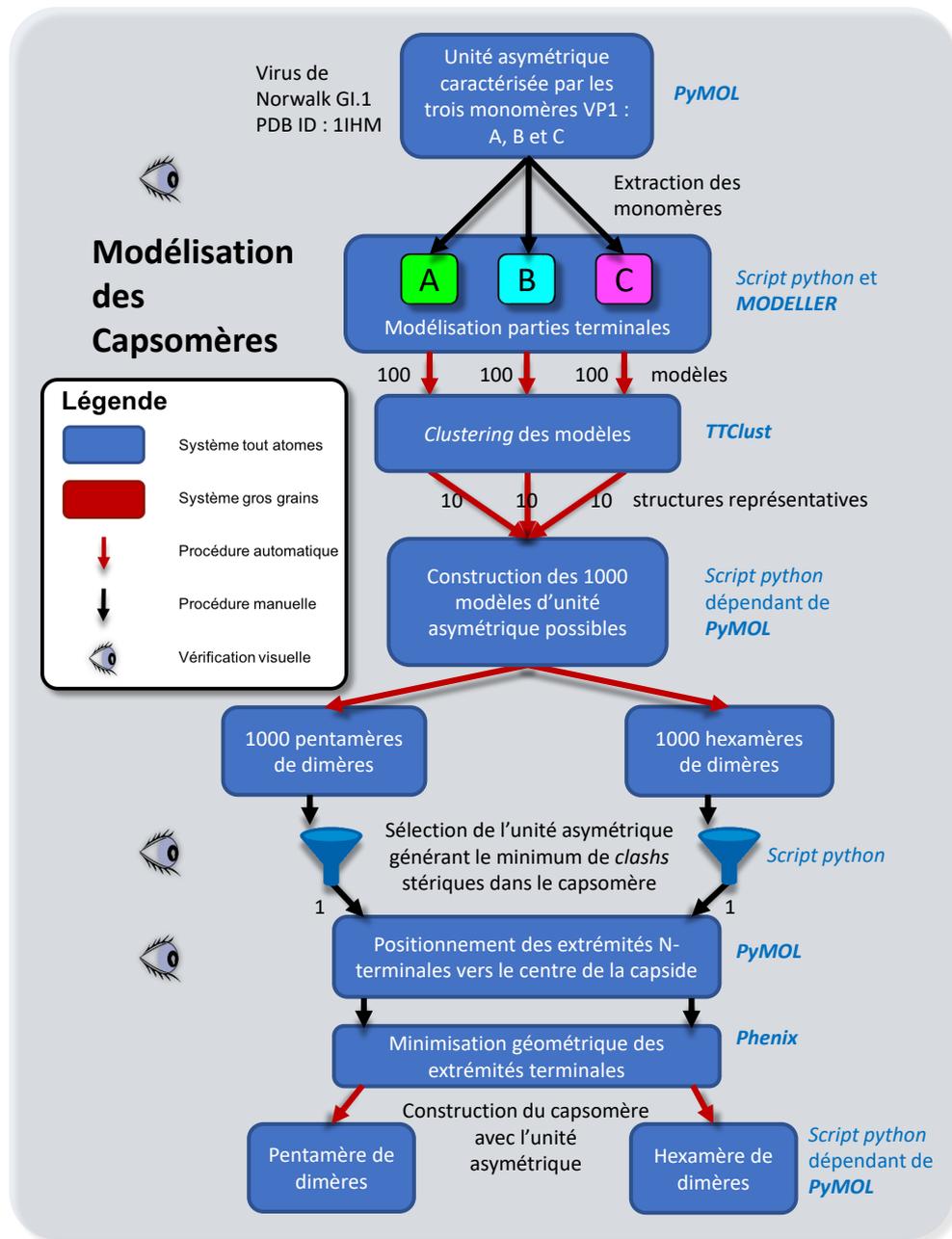


Figure 4.2: Schéma de la préparation des capsomères (POD et HOD). Réalisé par Jean-Charles Carvaillo.

L'étude réalisée en TR-SAXS en 2013 par Tresset *et al.* [167] pose l'hypothèse qu'une relocalisation des bras N-terminaux des dimères peut influencer sur l'assemblage de la capsid. L'étude sur le dimère de VP1 [168] montre que le

cluster acide pointé par Tresset *et al.* constitue certainement un interrupteur conformationnel du bras. Au regard de ces résultats, nous avons voulu utiliser un modèle complet du POD et du HOD, incluant les régions C-terminales et N terminales afin de ne pas biaiser nos simulations par rapport à ces études. Nous ne pouvions donc pas prendre directement les capsomères cristallographiques et nous avons dû alors passer par une étape de modélisation.

2.1 Modélisation de l'unité asymétrique

Nous avons choisi comme stratégie de modéliser les résidus manquants sur les chaînes A,B et C avec Modeller 9.16 [112] en produisant 100 modèles pour chaque monomère afin de générer une diversité de conformation. Ne pouvant pas tester toutes les combinaisons de conformation, et tenant compte du fait qu'il y avait des bras dans des positions très similaires, nous avons clusterisé avec TTclust (voir [TTclust](#) page 75) les structures générées afin d'obtenir 10 structures par chaîne. 1000 unités asymétriques ont ainsi été produites. Nous avons alors reconstruit les POD et HOD en nous basant sur la structure cristallographique, afin d'évaluer le nombre de clashes stériques formés par la modélisation des bras N et C-terminaux. Nous avons ensuite sélectionné l'unité asymétrique avec la plus petite quantité de clashes stériques après les reconstructions. La position des bras N-terminaux au sein de l'unité asymétrique a enfin été optimisée manuellement avec la fonction `sculpting` de Pymol 1.8.2 [165] puis minimisée avec l'outil de minimisation géométrique de PHENIX[172].

2.2 Modélisation des capsomères

Les capsomères ont été obtenus en dupliquant les chaînes A,B et C de l'unité asymétrique et en les alignant sur les chaînes correspondantes du POD et HOD de la structure cristallographique.

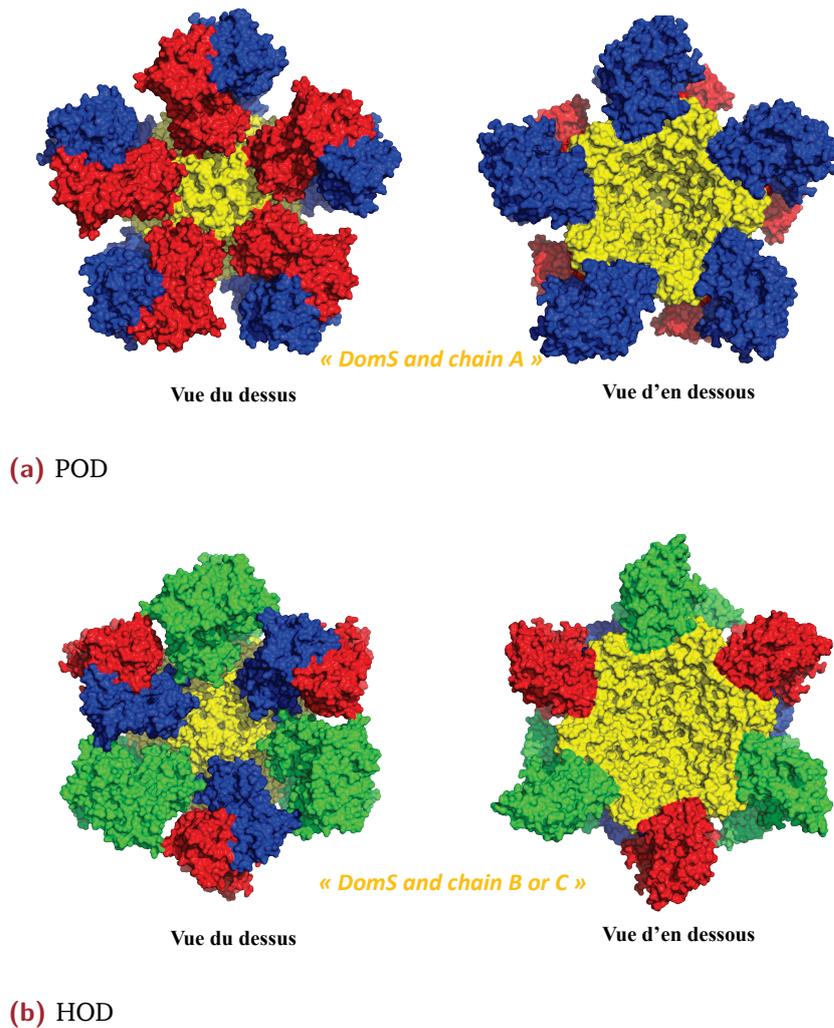


Figure 4.3: Précision de terminologie sur le POD et le HOD. La partie en jaune représente les domaines S centraux.

2.3 Précision de terminologie

Afin de faciliter la compréhension de certaines figures durant cette étude, nous rappelons que le POD est formé de dimères A/B avec les conformations A au centre. Le HOD est lui formé de 3 dimères A/B (conformations B au centre) et 3 dimères C/C. Afin de différencier les chaînes lors des calculs, la chaîne C périphérique est rebaptisée "D".

"DomS and chain A" représente donc les domaines S centraux du POD (Fig. 4.3a) et "DomS and chain B or C" désigne de la même façon les domaines S centraux du HOD (Fig. 4.3b).

Dynamiques du pentamère et de l'hexamère de dimère VP1

3.1 Simulations gros grains

Deux simulations de 10 μs ont été réalisées sur chaque capsomère. Le protocole de simulation est décrit dans la partie [Simulations gros grains](#) (page 60). Une seconde simulation (réplique) a été effectuée afin de confirmer ou d'infirmer les observations faites sur la première trajectoire.

3.1.1 Évolution du RMSD

3.1.1.1 Évolution globale

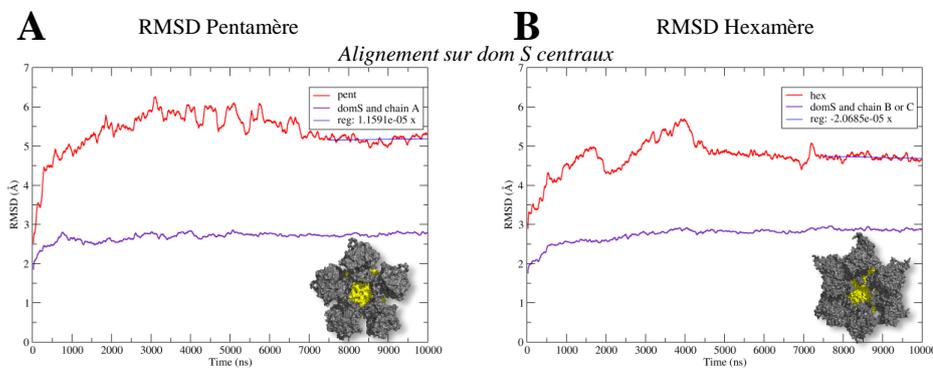


Figure 4.4: Flexibilité du POD (A) et du HOD (B) pour 10 μs de simulation. Les structures ont été alignées sur les domaines S centraux et les RMSD sont calculés sur le *backbone*. La courbe rouge correspond au RMSD sur l'ensemble de la structure (sans les régions N et C-terminales), la violette uniquement sur les domaines S centraux.

Nous avons tout d'abord étudié la flexibilité du POD et du HOD afin de vérifier que les systèmes étaient stables durant les simulations. Pour cela, nous avons aligné les trajectoires des POD et du HOD sur les domaines S centraux, car le domaine S est le plus conservé et les interactions entre dimères se font

essentiellement par la formation de contacts hydrophobes et électrostatiques entre domaines S [56].

Les résultats, présentés dans la figure 4.4 montrent le RMSD sur l'ensemble du POD (Fig. 4.4A) et HOD (Fig. 4.4B) en rouge et sur les domaines S centraux uniquement en violet. Les régions N et C-terminales ont été exclues des calculs et de l'alignement. Les 4 courbes convergent toutes vers une valeur pour atteindre un plateau, montrant que les systèmes se stabilisent avant la fin du temps simulé (7 μs pour le POD et 4 μs pour le HOD) (cf.4.4A,B courbe rouge). Les domaines S centraux convergent beaucoup plus vite, ils se stabilisent dès la première microseconde pour un RMSD d'environ 2.5 Å par rapport à la structure à $t=0 \mu s$.

3.1.1.2 Évolution des dimères

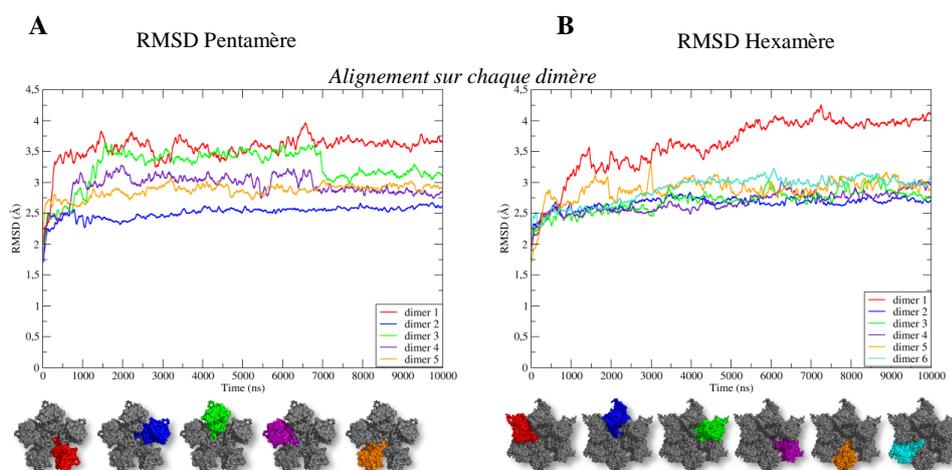


Figure 4.5: Flexibilité pour chaque dimère du POD (A) et du HOD (B) sur 10 μs de simulation. Chaque dimère a été aligné sur lui-même et les RMSD sont calculés sur les *backbones*.

Nous avons ensuite suivi le comportement de chaque dimère au sein des capsomères (4.5). Pour chaque calcul de RMSD, un alignement sur le dimère concerné a été effectué afin d'observer la flexibilité intrasèque au dimère.

Les dimères du POD (Fig. 4.5A) ont un RMSD compris entre 2 et 3.5 Å environ, après 2.5 μs . Quant aux dimères du HOD (Fig. 4.5B), le RMSD de 5

dimères sur 6 se stabilise autour de 2.5 Å, excepté pour un dimère qui voit son RMSD augmenter tout au long de la dynamique (courbe rouge).

Prasad *et al.* ont émis l'hypothèse que le POD se forme d'abord parce que les contacts entre dimères sont les plus importants autour des axes 5 [56]. Pour les auteurs, la formation ou la stabilité du POD serait donc favorisée par rapport au HOD. Cependant, au vu de nos analyses, qui sont confirmées par les répliques, nous ne pouvons pas conclure que le HOD une fois formé est moins stable que le POD.

3.1.2 Distance entre dimères

3.1.2.1 Domaines S et P

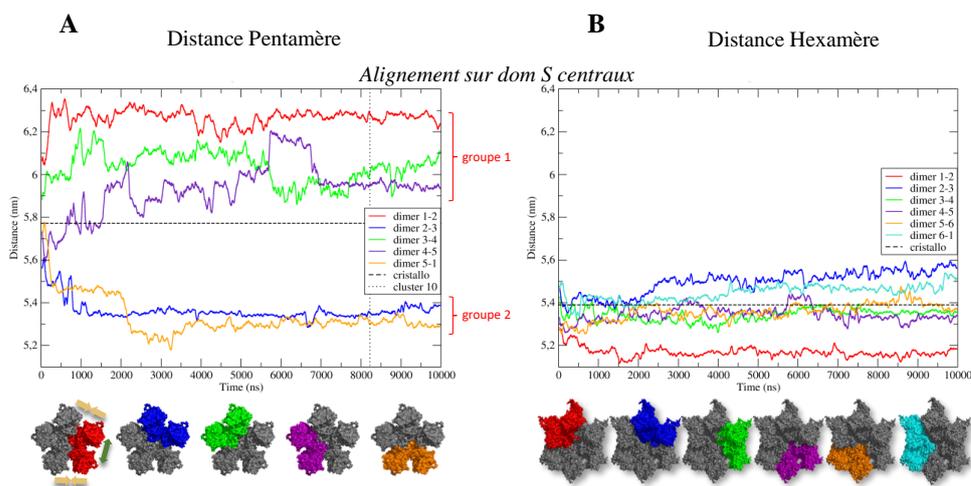


Figure 4.6: Distance entre chaque centre de masse des dimères du POD (A) et du HOD (B) sur 10 μ s de simulation. Les structures ont été alignées sur les domaines S centraux.

Lors de la visualisation de la trajectoire de MD du POD, il était clair que certains dimères pouvaient se rapprocher et d'autres s'éloigner les uns des autres. Nous avons donc voulu quantifier cette observation en mesurant la distance entre les centres de masses de chaque dimère voisin (cf. 4.6A). Pour le PDO, nous pouvons voir deux groupes se former très rapidement :

1. un groupe où les distances interdimériques augmentent par rapport à la structure cristallographique (courbe pointillée)

- un groupe où les distances interdimériques diminuent par rapport à la structure cristallographique.

Nous avons également vérifié si nous pouvions observer la même chose pour le HOD (4.6B), mais clairement il apparaît pas de groupe distinct, les distances entre les dimères évoluent autour de la distance cristallographique.

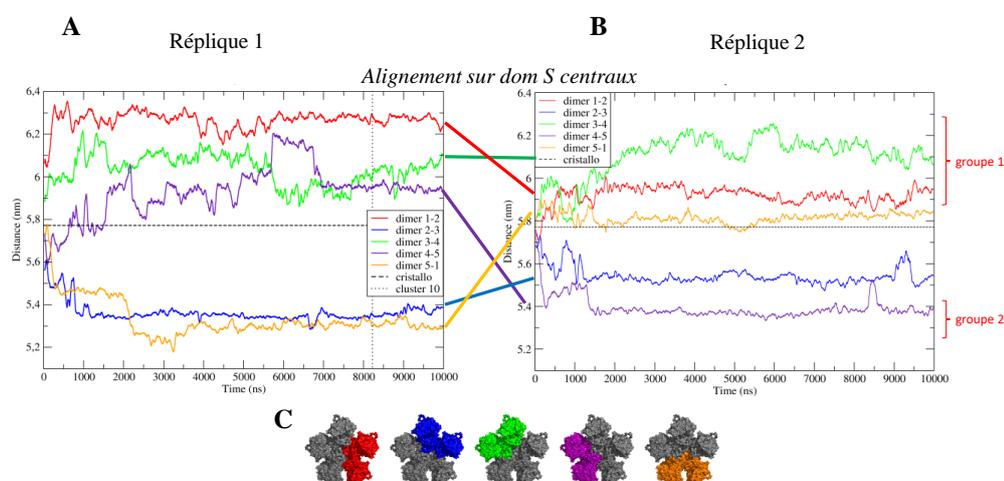
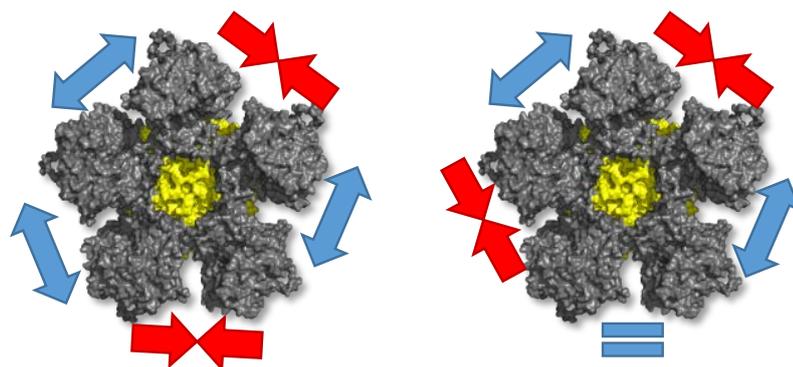


Figure 4.7: Distance entre chaque dimère du POD pour la réplique 1 (A) et 2 (B) sur $10 \mu\text{s}$ de simulation. (C) correspondance courbes/dimères. Les structures ont été alignées sur les domaines S centraux.

Si nous reproduisons la même analyse pour la deuxième réplique (Fig. 4.7), nous pouvons observer le même comportement : trois dimères vont s'éloigner ou rester à une distance similaire à celle observée dans la structure cristallographique (groupe 1) et deux dimères vont se rapprocher (groupe 2) (Fig. 4.7B). Cependant, il y a un changement de position dans les dimères se rapprochant et s'éloignant : la courbe violette montre que les dimères violets se rapprochent dans la réplique 1 alors qu'ils s'éloignaient dans la réplique 2. A l'inverse, les dimères jaunes, qui se rapprochaient dans la réplique 1 semblent garder leur distance initiale dans la seconde. Si nous regardons l'ordre et la position des dimères s'éloignant/se rapprochant, nous pouvons voir qu'elle est identique, mais avec une rotation (Fig. 4.8a et 4.8b).



(a) Simulation

(b) Réplique

Figure 4.8: Alternance des positions des éloignements et rapprochements entre dimères lors de la simulation du POD et de sa réplique. Les flèches bleues indiquent un éloignement, les rouges un rapprochement. Le égal bleu montre une distance inter dimère similaire à la structure cristallographique. Dans les deux cas, nous avons une alternance de $2 \leftrightarrow$, $1 \rightarrow \leftarrow$, $1 \leftrightarrow$ et $1 \rightarrow \leftarrow$.

3.1.2.2 Domaines S périphériques

Le domaine S est clairement le domaine d'assemblage de VP1. Cette conclusion basée sur la structure de Prasad *et al.* a été largement confirmée depuis. De plus, les simulations réalisées précédemment sur le dimère (voir [Dynamics and asymmetry in the dimer of the norovirus major capsid protein](#) page 89) ont permis de mettre en évidence une forte flexibilité entre les domaines S et les domaines P. Ces observations sont confirmées par les conformations alternatives, compacte et étendue, des spicules dans la capsid. Il peut donc être plus approprié de considérer uniquement les domaines S dans l'assemblage, comme l'avaient fait Prasad *et al.* Nous avons donc reproduit les analyses de distances interdimériques en considérant uniquement les domaines S périphériques (chaîne B) et ce sur la première dynamique.

Nous pouvons voir sur la figure 4.9 que deux groupes se forment où 3 domaines S semblent s'éloigner et 2 autres semblent se rapprocher ou rester à une distance similaire à celle observée dans la structure cristallographique (67.6 Å). Il est à noter que ces simulations sont similaires à celles faites en considérant tout le dimère (Fig. 4.7A) mais un espace interdimérique semble tout de même différent : le 2-3 (Fig. 4.9, courbe bleue). Les deux domaines

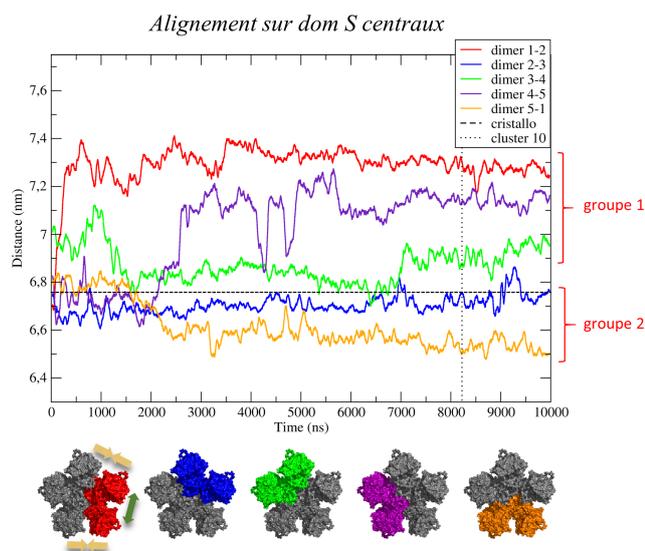


Figure 4.9: Distance entre les centres de masse des domaines S périphériques (chaîne B) durant la première simulation sur le POD.

vont légèrement se rapprocher, mais ensuite retrouver, au bout de $10 \mu\text{s}$, un écartement similaire à celui de la conformation du POD cristallographique. En basant le calcul de distance sur les centres de masses du dimère entier (Fig. 4.7A), les dimères vont se rapprocher de 2 \AA sans jamais retourner à une distance similaire à celle de la conformation cristallographique.

Ces analyses nous révèlent une rupture de symétrie dans le POD avec la même alternance d'ouverture/fermeture. La rotation de la séquence d'ouverture/fermeture dans la réplique indique qu'il ne semble pas avoir de position privilégiée quant à la localisation de la rupture de symétrie.

Nous pouvons également rajouter que sous forme pentamérique, les dimères de S et P semblent rester assez flexibles l'un par rapport à l'autre, ce qui est cohérent avec les simulations réalisées sur le dimère isolé.

3.2 Analyse des modes normaux sur le pentamère et l'hexamère de dimères

Afin d'en savoir plus quant à la flexibilité liée aux mouvements de grandes amplitudes, nous avons réalisé une analyse de modes normaux en utilisant la

méthode NOLB [177]. Cette méthode est non linéaire et permet de garder la protéine en bloc rigide afin de ne pas avoir de phénomène d'étirement.

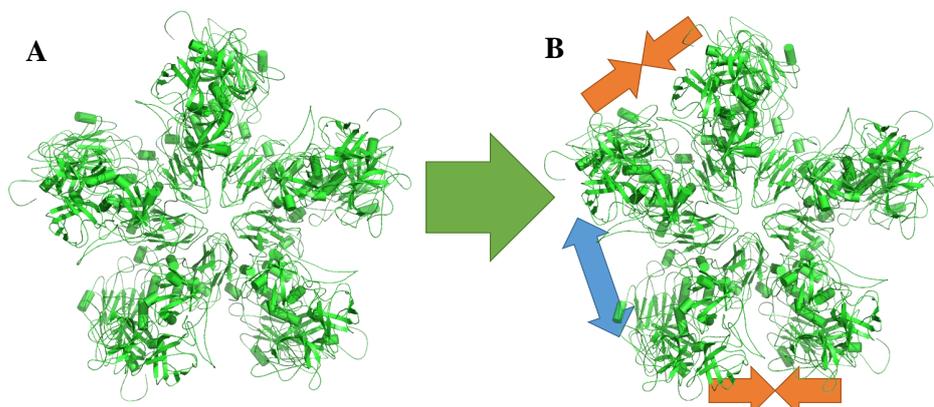


Figure 4.10: Mode 8 de l'analyse par modes normaux sur le POD complet. (A) structure initiale (B) image 3/10.

Nous avons calculé 10 modes sur le POD utilisé dans les simulations (bras N et C terminaux inclus). Celui-ci est parfaitement symétrique initialement. Parmi ces 10 modes, la moitié concernent des mouvements des régions terminales et l'autre moitié les mouvements de grandes amplitudes du POD. Nous pouvons voir sur le mode 8 un rapprochement de deux paires de dimères 4.10, exactement comme la première dynamique sur le POD. En comparant ce mouvement avec les autres modes, nous en avons isolé un deuxième : le 9^{ème}. Là encore, nous observons le même enchaînement rapprochement/éloignement de dimères à une rotation près (Fig. 4.11b)

Les modes normaux nous ont donc permis de retrouver des mouvements similaires à la dynamique sur le POD. Les mouvements observés ne sont donc pas des artefacts, mais bien des mouvements intrinsèques au POD qui peuvent jouer un rôle important dans l'assemblage de la capsid du norovirus.

3.3 Conclusion de l'étude de la dynamique des capsomères.

Les simulations nous ont permis de mettre en évidence une rupture de la symétrie du POD, que nous ne retrouvons pas dans le HOD. Nous observons deux dimères se rapprocher, laissant ainsi une plus grande distance inter

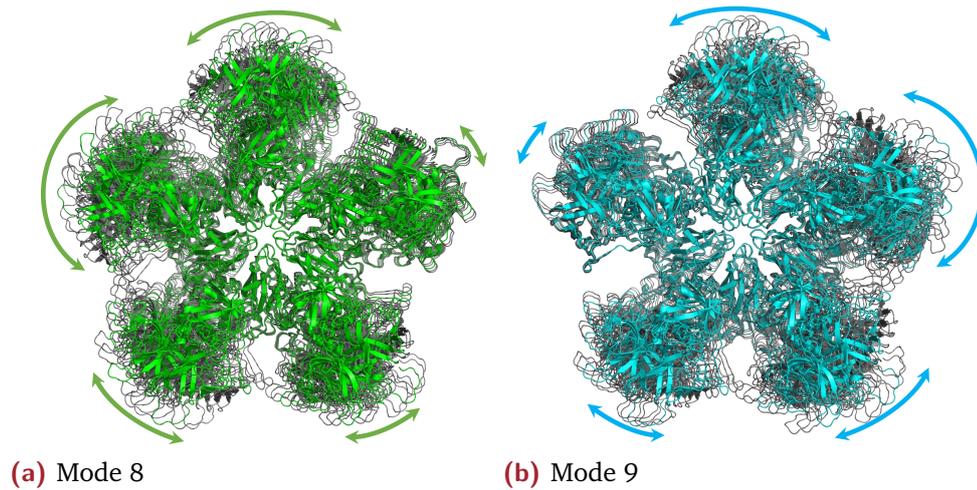


Figure 4.11: Comparaison des modes 8 (4.11a) et 9 (4.11b) de l'analyse des modes normaux sur le POD.

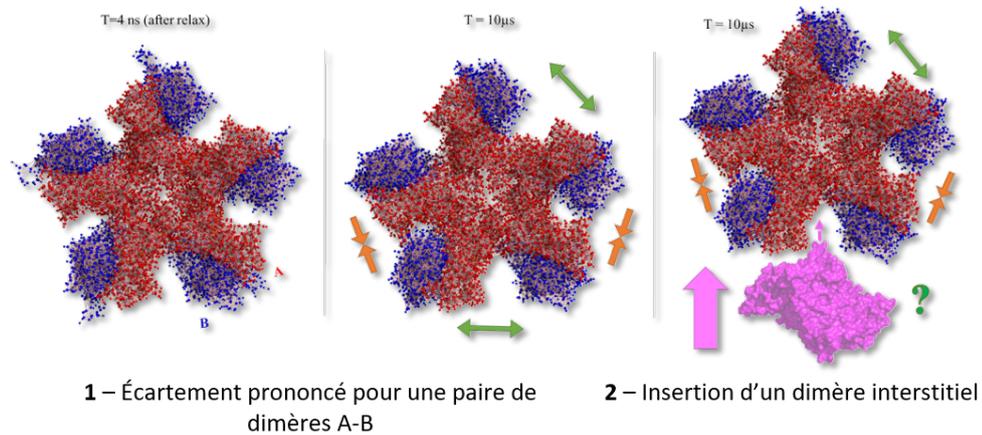


Figure 4.12: Hypothèse d'extension à partir d'un POD.

dimère entre 3 paires de dimères. Nous pouvons émettre l'hypothèse que l'espace ainsi formé pourrait favoriser l'insertion d'un dimère interstitiel (cf. Fig. 4.12) et ainsi commencer à étendre le POD en un oligomère plus grand, comme l'intermédiaire d'assemblage (22mer) [75].

Docking d'un dimère interstitiel sur un pentamère de dimères

Dans un POD extrait de la structure cristallographique, il y a cinq sites de fixations identiques pour un dimère interstitiel. L'idée est de tester si une conformation de POD présentant des distances interdimériques plus grandes, permettrait de favoriser une direction lors de l'extension du POD. Ceci nous permettrait de comprendre la formation de l'intermédiaire observé en TR-SAXS par Tresset *et al.* [75] et de faire le lien avec le modèle d'assemblage émis par Prasad *et al.* en 1999 [56] (cf. partie [Modèles d'assemblage du norovirus](#) page 24).

4.1 Clustering de la trajectoire

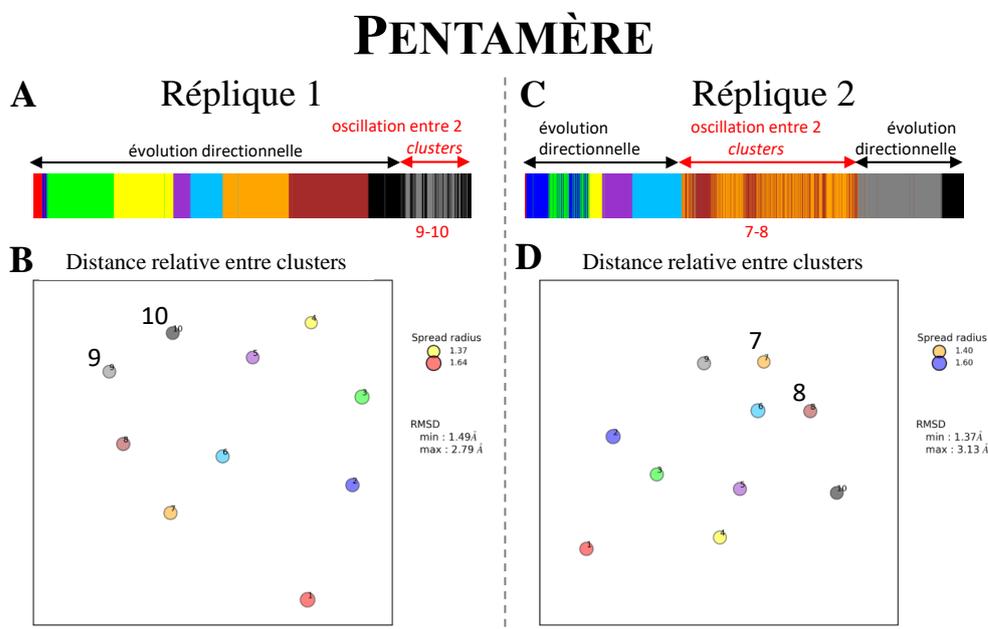


Figure 4.13: Résultats du clustering des trajectoires de MD du POD. (A-C) Graphique linéaire de l'appartenance aux clusters durant la trajectoire pour la réplique 1 (A) et 2 (D). (B-D) Distance relative entre chaque cluster pour les deux répliques. Réalisé avec TTCLUST (cf. [TTclust](#) page 75).

Afin de réaliser le docking, nous devons d'abord choisir une structure représentative dans la trajectoire des simulations sur le POD. Pour cela, nous avons utilisé une approche de clustering par TTKLUST sur la trajectoire de MD du POD et de sa réplique.

Nous pouvons voir pour la première dynamique du POD (4.13A,B) que le système passe d'une conformation à une autre sans retour en arrière pendant les 8 premières microsecondes. Au cours des dernières 2 μs , la structure du POD oscille entre les clusters 9 (gris) et 10 (noir). Pour la réplique 2, l'évolution du système est directionnelle en début (0 μs \rightarrow 3.5 μs) et à la fin (7.5 μs \rightarrow 10 μs) de la dynamique. Le POD oscille entre deux conformations proches entre 3.5 μs et 7.5 μs .

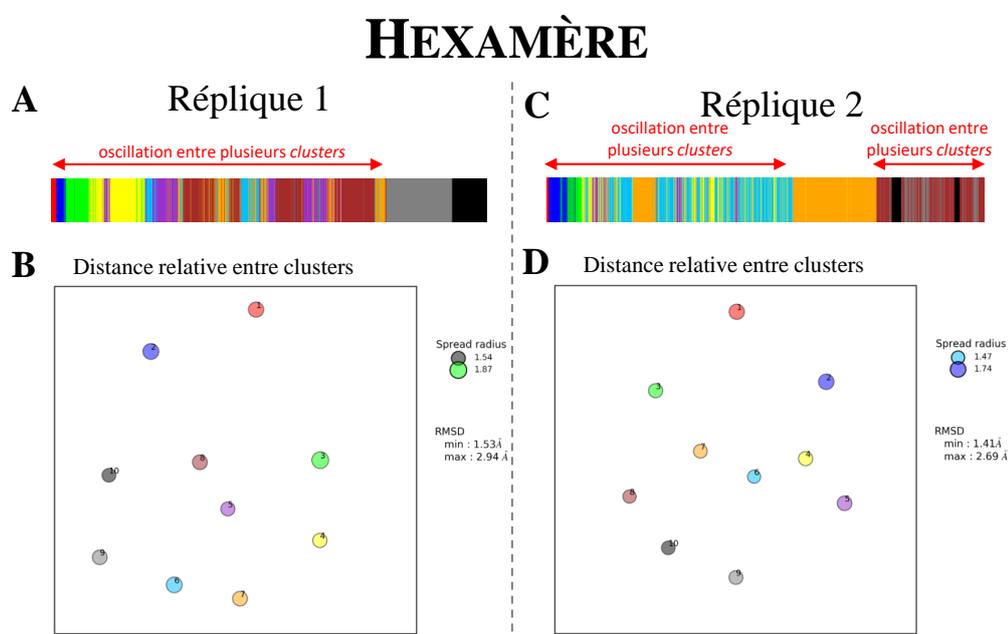


Figure 4.14: Résultats du clustering des trajectoires de MD du HOD. (A-C) Graphique linéaire de l'appartenance aux clusters durant la trajectoire pour la réplique 1 (A) et 2 (C). (B-D) Distance relative entre chaque cluster pour les deux répliques.

Concernant la première dynamique et la réplique du HOD, nous pouvons observer des comportements différents que pour le POD (cf. Fig. 4.14). Des oscillations entre conformations se produisent sur la grande majorité des dynamiques et certaines conformations sont présentes par intermittence durant la dynamique. De plus, certaines conformations présentes en début de dynamique sont retrouvées plus tard dans la simulation de MD (cluster 8

pour la réplique 1, cluster 7 pour la seconde réplique). Enfin, il semble qu'il n'y ait pas de chemin privilégié quant aux changements de conformation.

Nous avons donc décidé de prendre pour le *docking* un POD, et pour cela nous avons choisi la structure représentative du cluster 10 pour la réplique 1 et du cluster 8 pour la réplique 2.

4.2 Docking avec ClusPro

Nous avons utilisé le webserveur CLUSPRO 2.0 [178–181] afin de réaliser nos dockings. Ce serveur réalise une transformation de Fourier rapide prenant en compte une fonction d'énergie spécifique. Un clustering puis une étape de minimisation est ensuite réalisée sur les structures dockées.

4.2.1 Préparation

Ce serveur nécessite une structure tout atomes afin de fonctionner. Étant donné que la méthode backward (reconstitution de structure tout atomes à partir d'une structure gros grains) de MARTINI ne fonctionne pas sur un système utilisant un réseau élastique ELNEDYN, nous avons fait le choix d'aligner 10 monomères S (30-220) et 5 dimères de P (230-520) d'un POD cristallographique sur les structures sélectionnées (voir [Clustering de la trajectoire](#) page 150). Afin de faciliter la reconstruction, nous avons omis les régions terminales et les segments interstitiels. Les structures ont été ensuite minimisées par PHENIX 1.10.1 [172] afin de retirer les clashes stériques.

Les paramètres standards du serveur CLUSPRO 2.0 ont été utilisés. La structure réceptrice est la structure reconstruite tout atomes du POD et la structure dockée est un dimère CC cristallographique.

4.2.2 Résultats

Nous avons classé les résultats de *docking* en deux groupes : "good" représentant une structure avec une position après *docking* similaire à celle trouvée dans la structure cristallographique, et à l'inverse, "bad" où la position des structures est trop différente. Ainsi, nous pouvons observer pour le résultat du

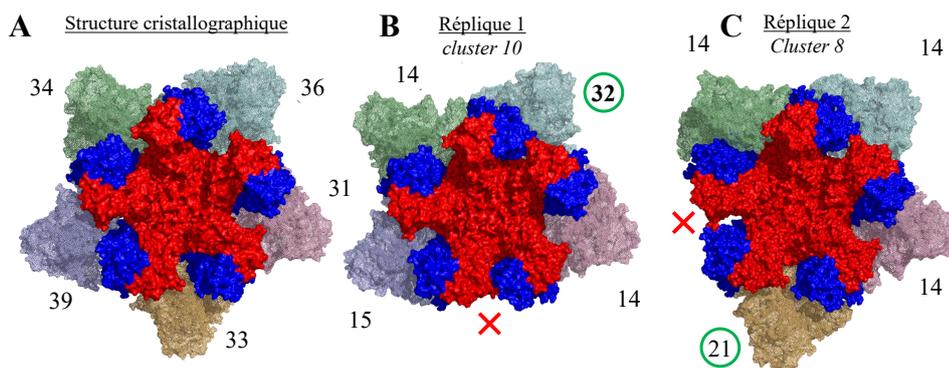


Figure 4.15: Résultats du *docking* sur le POD d'un dimère interstitiel de type C/C, nombre de résultats sur 1000 tirages. Seules les positions dockées favorablement sont présentées.

docking sur la structure cristallographique (4.15A) qu'il n'y a pas de position privilégiée. Il y a approximativement le même nombre de clusters pour les 5 positions possibles. Quand on regarde les positions dockées pour la structure représentative du cluster 10 de la réplique 1 (4.15B) et du cluster 8 de la réplique 2 (4.15C) nous pouvons voir cette fois-ci qu'une position semble être favorisée par rapport aux autres (cercle vert) et une position est totalement ignorée.

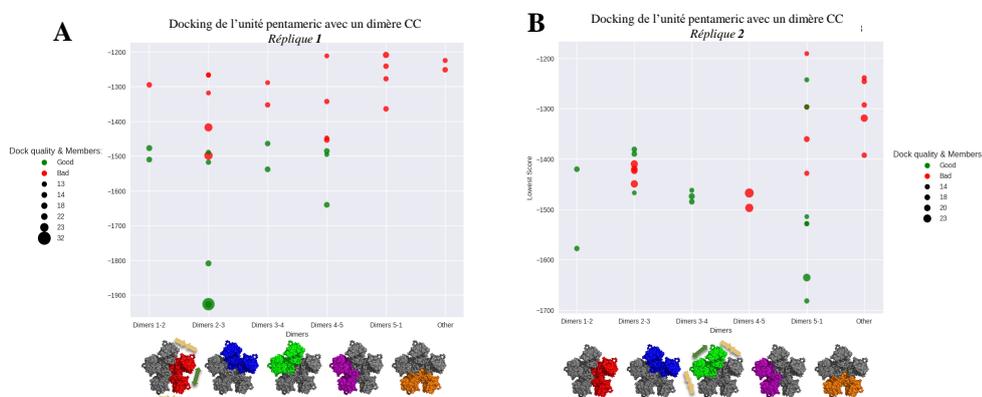


Figure 4.16: Résultats du *docking* entre un dimère CC et POD avec la conformation représentative du cluster 10 pour la réplique 1 et du cluster 8 pour la réplique 2. Les points rouges correspondent aux structures qui sont "mal" dockées et les points verts aux structures dont la position après *docking* est similaire à celle trouvée dans la structure cristallographique.

En se basant sur les scores d'énergies propres à ClusPro 2.0 (Fig. 4.16), une position plus favorable de *docking* ressort. Il s'agit, pour la réplique 1 (Fig. 4.16A), du site entre les dimères 2 et 3 dont le score est inférieur à

-1900 et regroupant 32 résultats sur 1000 tirages (Fig. 4.15B). Ce site de *docking* correspond à un espace interdimérique réduit (Fig. 4.7A), mais les domaines S périphériques sont à une distance similaire de celles trouvées dans la structure cristallographique (Fig. 4.9).

La position 5-1, quant à elle, ne constitue pas une position de *docking* favorable. Cette position correspond à un espace interdimérique (Fig. 4.7A) et inter S périphérique réduit (Fig. 4.9).

La catégorie "Other" regroupe les résultats de *docking* qui n'ont pas une position similaire au dimère CC de la structure cristallographique (position sous le POD, sur des domaines P...) Pour la réplique 2 (Fig. 4.16B), la favorisation d'une position (position 5-1) et l'exclusion d'une autre sont reproduites, ainsi que la cohérence de ces résultats avec l'étude sur l'espace inter dimérique (Fig. 4.7B)

4.2.3 Conclusion

Nous pouvons tirer de ces simulations une conclusion importante : les cinq espaces inter dimères du POD, initialement identiques, évoluent spontanément pour devenir non équivalents. Dans la dynamique et sa réplique, il apparaît un site nettement plus favorable pour le *docking*. En comparant ces résultats avec la valeur d'écartement des dimères (4.7) nous constatons que l'écartement le plus grand ne correspond pas à la meilleure position de *docking*, comme nous l'avions envisagé. Cependant, le rapprochement des domaines S est lui toujours très défavorable au *docking* d'un dimère interstitiel de type C/C. La distance minimale requise entre deux domaines S pour permettre le *docking* d'un dimère supplémentaire semble être la distance trouvée dans le POD cristallographique (environ 68 Å)

Conclusions sur les intermédiaires d'assemblage de la capsidie du norovirus.

Pour chaque capsomère du norovirus (POD ou HOD), deux MD gros grains ont été réalisées avec un champ de force MARTINI couplé au réseau élastique ELNEDYN corrigé par l'algorithme domELNEDIN pour garder toute la flexibilité entre domaines S et P. L'étude de l'évolution des capsomères et des dimères les formant ne nous a pas permis de mettre en évidence une réelle différence de stabilité une fois formés (Fig. 4.4 et 4.5).

Cependant, l'étape de clustering nous indique que le HOD et le POD ont des évolutions qualitativement différentes : à l'inverse du HOD, le POD est caractérisé par une évolution irréversible sur la durée de la dynamique se finissant par une alternance entre deux conformations proches (Fig. 4.13 et 4.14). Les analyses de distance entre dimères au sein des capsomères ainsi que l'analyse des modes normaux révèlent une rupture de symétrie du POD avec des mouvements importants des 5 domaines S périphériques. Ces mouvements induisent des rétrécissements de certains espaces inter dimères et des élargissements d'autres. L'étude par *docking* protéine/protéine a montré qu'une conséquence potentiellement importante est qu'une position devient favorisée pour l'insertion d'un dimère, ce qui suggère l'apparition d'une direction privilégiée de croissance dès le stade POD. Cependant cette position favorisée n'est pas aisément détectable : ce n'est pas celle qui a la plus grande distance inter dimère, ni forcément celle qui a la distance entre domaines S périphériques la plus proche de la valeur de la structure cristallographique. La gamme d'ouverture permettant le bon *docking* d'un dimère interstitiel n'est donc pas directement corrélée à l'écartement de deux dimères. Cependant, une ouverture minimum des domaines S périphériques, correspondant à l'ouverture entre deux dimères retrouvée dans un POD cristallographique, est nécessaire à l'insertion d'un dimère interstitiel.

D'après nos résultats, une croissance isotrope à partir d'un POD (proposé par Prasad *et al.*) semble peu favorable, mais confirme indirectement la première partie du modèle de Prasad *et al.* (formation du POD). Nos résultats sont donc cohérents avec l'observation d'un intermédiaire d'assemblage de forme allongée tel qu'observé par Tresset *et al.*. Il reste à déterminer si et pourquoi cet intermédiaire est bien formé de deux PODs reliés par un dimère interstitiel. D'ores et déjà on peut proposer que la formation d'un POD est suivie par l'amarrage d'un dimère dans une position favorisée, puis par interaction avec d'autres dimères, par une extension dans cette direction favorisée de 4-5 dimères supplémentaires.



Discussion

” *It's more like a big ball of wibbly wobbly...
time-y wimey... stuff.*

— **Doctor Who**
Doctor Who, Saison 3, *Blink*

Discussion

Le but de cette thèse est de mieux comprendre les mécanismes moléculaires impliqués dans la formation de la capsidie des novovirus. La partie protéique du virion est formée pour l'essentiel de 180 copies d'une protéine structurale VP1. VP1 est composée de deux domaines (S, composant la coque de la capsidie et P formant des spicules au-dessus des domaines S). Contrairement à beaucoup de petits virus à ARN simple brin de polarité positive, VP1 n'interagirait pas directement avec l'ARN viral par son domaine interne. C'est pourquoi elle forme spontanément des capsides vides identiques aux virions, qui constituent donc par leur simplicité un excellent modèle d'assemblage pour des études biophysiques et *in silico*. Pour mener à bien ce travail, nous avons utilisé diverses méthodes de MD (classiques, CG, TMD mais également du *docking*) et nous nous sommes appuyés sur des données provenant de différentes techniques biostructurales (Cristallographie, SAXS, Cryo-EM) afin de construire des modèles structuraux pour différents génogroupes avec différentes conformations.

En 1999, Prasad *et al.* proposent un modèle en se basant sur les observations de la structure nouvellement résolue de la capsidie vide du génotype I.1 : la brique d'assemblage est un dimère de VP1. Ces dimères de VP1 vont s'auto-assembler d'abord en POD (formant un axe 5). A partir de ce capsomère, une croissance isotropique formerait la capsidie. En 2013, Tresset *et al.* ont désassemblé et réassemblé *in vitro* la VP1 du génotype III.2. Le désassemblage en dimères a été effectué en conditions non physiologiques à pH 9 et très faible force ionique pour forcer la dissociation par répulsion électrostatique. Le retour par mélange rapide dans des conditions plus physiologiques provoque le réassemblage en capsidie. Tresset *et al.* ont mis en évidence un seul

intermédiaire d'assemblage, de stoechiométrie 10-11 dimères et de forme très allongée. Cet intermédiaire serait interprétable comme deux POD pontés par un dimère interstitiel.

Nous nous sommes donc intéressés tout d'abord à la brique élémentaire de la capsid : le dimère de VP1. En particulier nous avons focalisé notre attention sur une partie du bras N-terminal contenant, dans la molécule B uniquement, une grappe de trois résidus acides non neutralisés. Nos résultats ont tout d'abord montré que le bras N-terminal pouvait se relocaliser suivant l'état de protonation de ces résidus. Une protonation des résidus D10 et E22 est suffisante pour que le bras N-terminal retrouve une conformation similaire à celle observée dans la structure cristallographique. Ce bras N-terminal, dans la chaîne B de la structure cristallographique de la capsid, établit des interactions autour de l'axe quasi-6 qui ne sont pas présentes à l'axe 5. Il se pourrait que pendant l'assemblage, un changement d'environnement élève le pKa de ces acides aminés, provoquant leur protonation à pH neutre et induisant une relocalisation du bras N-terminal permettant ainsi à un dimère de "verrouiller" son interaction avec ses dimères voisins autour de l'axe quasi-6. L'assemblage des VLP pouvant être médié par un saut de pH (pH9→pH neutre ou légèrement acide), la protonation de ces résidus peut donc permettre une relocalisation des bras, mais serait alors probablement un événement tardif dans l'assemblage de la capsid, comme proposé dans Tresset *et al.* [167].

Par la suite, Jean Lepault et Julien Maufront ont travaillé à l'obtention de cartes de cryo-EM de VLP de GIII.2 en fonction de plusieurs conditions expérimentales. En faisant varier la concentration en calcium ou en ajoutant un chélateur d'ions divalents, ils ont été en mesure de jouer sur la compacité de la capsid ainsi que sur l'orientation des spicules. Ces résultats ont montré une indépendance relative entre les dimères de P et les domaines S avec un équilibre entre les conformations compacte-étendue suivant les conditions. Le passage d'une conformation compacte à une conformation étendue de la capsid est donc possible.

Nos modélisations et simulations sur une conformation étendue de VP1 du GI.1 montrent que seul, le dimère peut adopter une grande gamme de

conformations. L'équilibre à deux états observé en cryo-microscopie n'est donc pas retrouvé ici, indiquant qu'il est peu probable que le dimère se trouve dans une conformation étendue en solution. Cependant, ces simulations nous ont également permis d'identifier certaines zones de contact entre les domaines S et P qui comprennent des acides aminés peu conservés et présents majoritairement dans le GI (voire GI.1). Ces simulations ont également révélé la présence d'une asymétrie dans les conformations que pouvait adopter le dimère.

Par la suite, nos résultats de TMD ont montré que la transition d'une conformation étendue à une conformation compacte est plus facile que l'inverse. De plus, la mutation de certains résidus peu conservés aux interfaces S/P (S134, H135, R175, Q226 et S508) ont un impact dans les interactions du domaine P au domaine S. Lorsqu'ils sont mutés par les résidus consensus parmi tous les génogroupes de VP1, le passage d'une conformation compacte à une conformation étendue requiert moins de force, indiquant que ces résidus ont un impact dans le maintien du domaine P au domaine S ainsi qu'un équilibre déplacé vers la forme compacte dans le GI.1. Il est également à noter que les résidus mutés en TMD, bien que peu conservés, se trouvent majoritairement dans le domaine S qui lui est plus conservé que le domaine P. Ils peuvent donc avoir un rôle bien plus important que de lier les domaines S et P entre eux.

Les résultats de nos expériences SAXS réalisées sur le dimère de VP1 du GI.1 en conditions non physiologiques de dissociation montrent une conformation différente des trois conformations cristallographiques (A,B et C). Nos modèles basés sur les données expérimentales montrent un dimère où les domaines S sont séparés, indépendants et positionnés de part et d'autre du dimère de P de manière asymétrique. Cette asymétrie récurrente dans les résultats sur le dimère libre laisse suggérer que cela peut jouer un rôle structural notamment dans des interactions entre dimères. Les résultats de Tresset *et al.* (JACS, 2013) sur l'assemblage de la capsid relèvent qu'à 4 ms dans des conditions proches de physiologiques, le dimère garde un rayon de giration beaucoup plus élevé que le dimère cristallographique [75], suggérant l'existence d'un dimère de GIII.2 en solution similaire à nos observations. Il est également

possible que VP1 soit également en équilibre entre des formes étendues et des formes compactes (similaires à la structure cristallographique).

Cette observation change la vision des premières étapes de l'auto-assemblage de la capsid, car jusqu'à présent, nous réfléchissions avec la structure disponible : le dimère dans ses conformations cristallographiques. Plus généralement, cela montre que dans certains cas, il ne faut pas toujours réfléchir uniquement avec les structures cristallographiques telles quelles. La dynamique moléculaire montre souvent que ces structures sont très flexibles et peuvent adopter plusieurs conformations différentes.

Afin de mieux distinguer et expliquer l'assemblage de l'intermédiaire de Tresset *et al.* (JACS, 2013) [75] et le chemin d'assemblage proposé par Prasad *et al.* (Science, 1999) [56] à partir d'un POD, nous avons réalisé des simulations gros grains sur le POD. Nous avons observé une rupture de symétrie très nette avec rétrécissement de certains espaces interdimériques et élargissement d'autres (que nous retrouvons également dans les analyses de modes normaux). Le *docking* d'un dimère supplémentaire sur la structure du POD issu de la MD montre une position de *docking* privilégiée indiquant également une direction privilégiée pour une croissance anisotrope à partir d'un POD, comme proposé par Tresset *et al.*. La croissance isotrope proposée par Prasad *et al.* ne semble donc pas possible à partir d'un POD. Ceci ouvre la question cruciale de la suite de la croissance à partir d'un POD : Quelle population sera privilégiée ? D'après Tresset *et al.* un intermédiaire formé de 10 dimères est aussi possible qu'un 11^{er} mer de dimères.

En intégrant l'ensemble de ces résultats, nous sommes en mesure de proposer un chemin d'assemblage de la capsid plus complet, tenant compte de la première partie du modèle d'assemblage de Prasad *et al.* en 1999 et des observations faites par Tresset *et al.* en 2013.

Nous proposons ainsi le modèle d'assemblage de la capsid du norovirus suivant 5.1 :

1. Des dimères en solution vont s'assembler, probablement via l'interaction de domaines S libres, pour former un POD comme le pensait Prasad *et al.* [56].

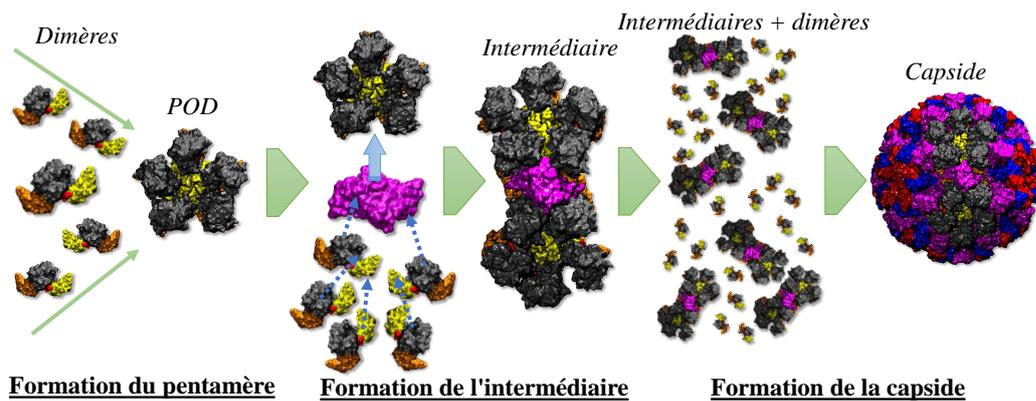


Figure 5.1: Schéma d'assemblage proposé du dimère à la capsid.

2. Rapidement, un dimère interstitiel va venir s'insérer entre deux dimères du POD.
3. Toujours via des interactions entre domaines S, le nouveau capsomère formé va s'étendre pour aboutir à un intermédiaire d'assemblage allongé composé de 10-11 dimères [75].
4. La formation de la capsid se fait ensuite par assemblage des intermédiaires. Si l'intermédiaire est bien composé de 2 POD et un dimère interstitiel, cela implique que 6 intermédiaires et 24 dimères interstitiels se combinent.

Cependant, le passage de cet intermédiaire putatif (composé de 2 POD + 1 dimère) à la capsid n'a pas été abordé durant cette thèse. Une rapide analyse en modes normaux sur l'intermédiaire d'assemblage (Fig. 5.2) montre des mouvements, entre les dimères des POD de l'intermédiaire, similaires à ceux observés sur le POD seul mais à ce stade, on ne peut pas exclure un intermédiaire de 10 dimères (qui est compatible avec le TR-SAXS de Tresset *et al.*).

Si à pH physiologique les domaines S de certains dimères sont toujours séparés (comme observé à pH 9 durant nos expériences en SAXS), il se pourrait que chaque domaine S d'un dimère libre interagisse avec un domaine S d'un POD d'un intermédiaire, faisant comme un lien entre les deux (cf. Fig. 5.3).

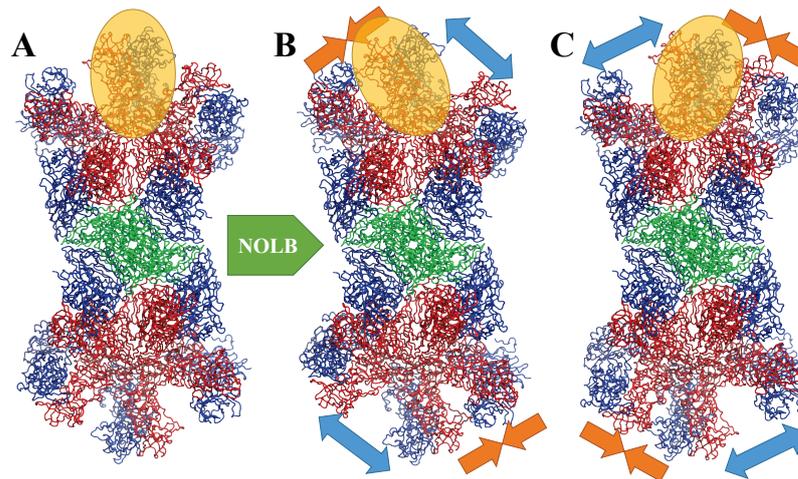


Figure 5.2: Analyse de modes normaux sur l'intermédiaire d'assemblage formé de 11 dimères (A - mode N°7). (B-C) Le dimère entouré d'une ellipsoïde semi-transparente jaune va se rapprocher d'un de ses dimères voisin tout en s'éloignant d'un second dimère voisin. Le programme utilisé est NOLB [177]. Les chaînes A sont représentées en rouge, B en bleue et C en vert.

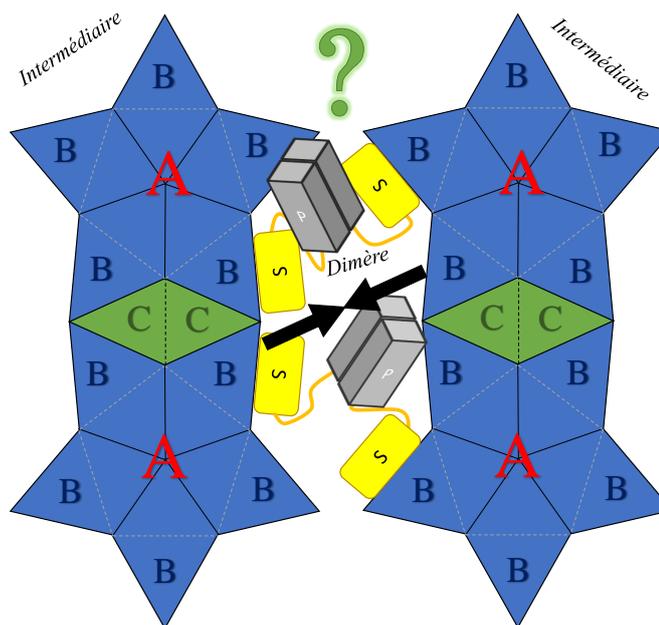


Figure 5.3: Hypothèse d'interaction entre intermédiaires d'assemblage médiée par des dimères étendus.

Pour que cette hypothèse soit compatible avec le fait que pendant l'étude de Tresset *et al.* sur la cinétique d'assemblage de la VLP des norovirus, aucune autre espèce n'est visible que dimères, 10-11 mers et capsides, il faut que la capside soit ensuite formée rapidement par l'effondrement des complexes

d'intermédiaires et de dimères. Ceci pourrait être induit par un mouvement de rotation des domaines P sur eux-mêmes (comme un tour de vis) entraînant un enroulement des régions interstitielles et rapprochant les deux domaines S jusqu'à former un dimère de S (cf. Fig. 5.4).

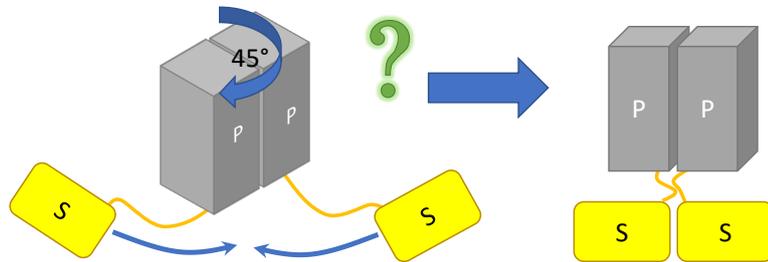


Figure 5.4: Hypothèse de dimérisation du domaine S.

Enfin, pour remettre ces résultats dans le contexte de la formation du virion infectieux *in vivo*, en 2013, Vongpunsawad *et al.* ont montré que la protéine VP2 s'associait à une région du domaine S de VP1 (voir [VP2](#) page 21) qui se trouve très proche à la fois de l'interface entre deux domaines S d'un dimère et du bras N-terminal. Nous pouvons donc poser la question de comment VP1, VP2 (protéine structurale mineure, voir page 21) et VPg (coiffe de l'ARNm, voir page 13) interagissent ensemble et à quel moment de l'assemblage ? Est-ce que le phénomène de relocalisation du bras N-terminal énoncé dans Tresset *et al.* (ABB 2013) [167] et analysé pendant cette thèse (Tubiana *et al.*, 2017 [168]) peut également être induit par un changement de l'environnement du domaine S par la protéine VP2 qui est très basique[51] ? VP2 peut-elle promouvoir le rapprochement de deux domaines S dans un dimère, déclenchant ou accélérant ainsi certaines étapes de l'assemblage ? VP2 est présente en quelques copies seulement dans la capsid. Est-ce que cela est suffisant pour induire la relocalisation des bras ou le rapprochement des domaines S dans toute la capsid ou bien est-ce que ces phénomènes sont uniquement locaux ? Bien qu'on ne sache toujours pas comment VP2 influe sur l'encapsidation du génome, la piste la plus probable reste une interaction avec la protéine VPg qui se trouve en 5' de l'ARN génomique. Si tel est le cas, comment VP2 interagit-elle avec VPg ? A pH physiologique, des calculs d'électrostatique nous ont montré que l'intérieur de la capsid est majoritairement négatif, comme l'ARN. Est-ce que les ions calciums peuvent intervenir ici aussi pour écranter les charges négatives de la surface interne

de la capside ? Dans ce cas nous pourrions penser qu'en présence d'ARN, une diminution du calcium peut avoir un fort effet déstabilisant pour le virus et être le déclencheur de la dissociation du norovirus dans la cellule. Des analyses *in silico*, aussi bien génomiques que structurales, pourraient nous permettre de mieux comprendre le lien entre ces trois protéines.

VI

Perspectives

” *"Ça prouve que j'ai de l'ubiquité..."*

— **Perceval**

Kaamelott, Livre III, L'étudiant

Perspectives

Les travaux réalisés pendant cette thèse ont permis de mieux comprendre les mécanismes moléculaires impliqués dans l'assemblage de la capsid des norovirus. Cependant, certaines questions restent en suspens, et de nouvelles ont émergé.

Il faut d'abord comprendre comment se forment les intermédiaires. Pourquoi la croissance s'arrête-t-elle à 10 ou 11 dimères ? Comment s'assemblent-ils en capsid ensuite ? Nous avons mis en place une méthodologie combinant des simulations gros grains et du *docking*. Nous pourrions l'adapter et l'utiliser pour étudier la formation d'un premier noyau (POD ou autre) puis sa croissance. Nous proposons donc le protocole suivant (cf. également Fig. 6.1) :

1. Simuler la formation du POD. Pour cela nous pourrions simuler plusieurs dimères de VP1 ensemble, en contact, à proche ou moyenne distance. L'utilisation d'un système CG pourrait nous permettre de simuler sur un temps plus long ou bien une méthode de MD biaisée afin de former un POD à partir de cinq dimères.
2. Docking d'un dimère supplémentaire sur un POD. Cette étape a déjà été réalisée dans le cadre de cette thèse (voir [Docking d'un dimère interstitiel sur un pentamère de dimères](#) page 150)
3. Réaliser une MD afin d'étudier l'évolution du complexe obtenu par *docking*. Cette étape est en cours de production.
4. Réaliser plusieurs types de docking :
 - a) *Docking* itératif isotropique autour d'un POD + 1 dimère. Nous pourrions alors démontrer *in silico* qu'une croissance isotropique

est bien impossible si nous n'arrivons pas à docker de dimères supplémentaires.

- b) *docking* itératif anisotropique en prolongation du POD + 1 dimère. Nous pourrions éventuellement être en mesure de démontrer l'existence d'un intermédiaire formé de 10 dimères si le docking d'un 11e n'est pas réalisable.
 - c) *docking* d'un POD supplémentaire. À ce stade, Le docking de deux POD autour d'un dimère interstitiel ne peut pas être exclu. Nous pourrions donc tester cette éventualité.
5. Les étapes de *docking* seront là encore suivies d'une étape de simulation afin de tester le comportement de chaque nouveau modèle. Dans le cas d'un *docking* itératif (4a, 4b) une nouvelle étape de docking sera réalisée après la simulation en fonction de comment a évolué le système.

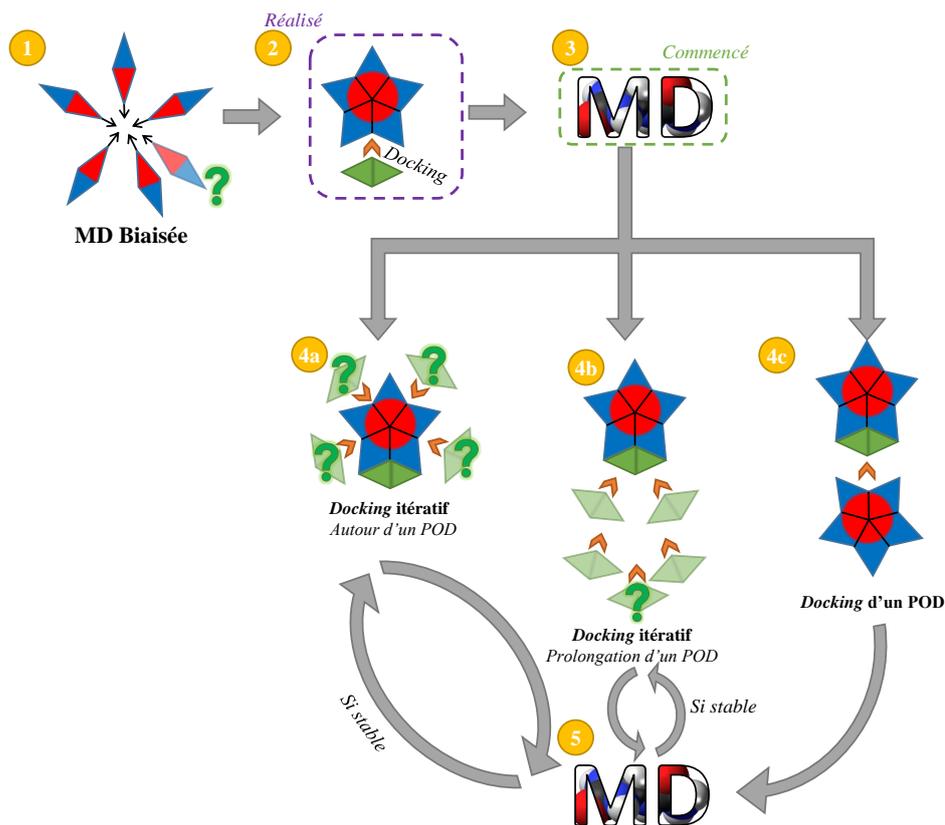


Figure 6.1: Protocole proposé pour l'étude de la formation d'un POD et d'un intermédiaire d'assemblage. Les étoiles à cinq branches représentent des POD, les losanges verts des dimères.

Nous pourrions ainsi continuer à simuler artificiellement la croissance du POD vers l'intermédiaire allongé, et en fonction des résultats de simulation et de docking, comprendre pourquoi la croissance s'arrête à 10 ou 11 dimères. Un intermédiaire à 10 dimères est une hypothèse séduisante car la stoechiométrie de l'assemblage devient simple : pour aboutir à la capsid il n'y a plus besoin de dimères supplémentaires (9 intermédiaires de 10 dimères sont suffisants pour former la capsid). Nous n'avons pas réussi à définir de décimère de dimères allongé dont la combinaison forme une capsid. Nous espérons que le protocole décrit ci-dessus pourrait nous guider vers une solution.

Nous pourrions également développer une méthode similaire à celles décrites en introduction (voir [Quelques simulations de capsides virales](#) page 27) en rajoutant un terme au champ de force afin de favoriser les interactions entre des modèles simplifiés de dimère. En biaisant le champ de force afin de favoriser les interactions entre dimères, nous pourrions alors être en mesure d'observer des assemblages de dimères en POD, et peut-être voir une direction privilégiée pour une croissance vers la capsid. Cette approche a déjà été utilisée dans l'étude de l'assemblage de la capsid du VIH par Grime *et al.* (2016)[95].

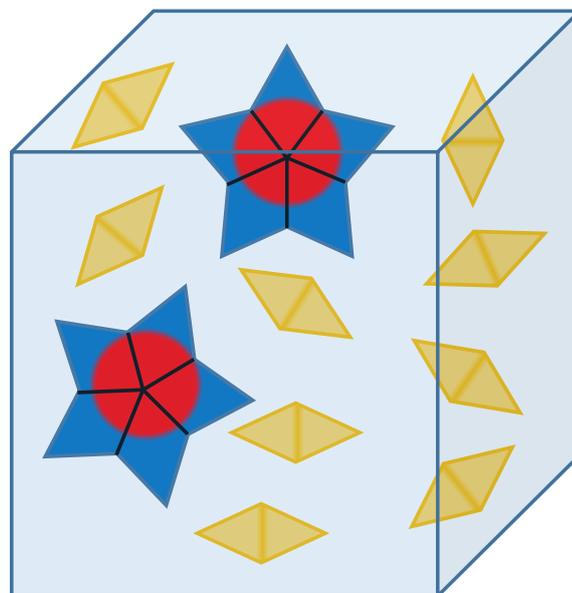


Figure 6.2: Simulation proposée pour étudier les interactions entre POD et dimères. Les losanges jaunes sont des dimères en conformation SAXS.

De plus, nos résultats de SAXS montrent que les domaines S du dimère en solution dans des conditions non physiologiques sont séparés. Nous pourrions alors tester des simulations avec un POD et des domaines S indépendants, ou bien avec la conformation SAXS du dimère à la place d'une conformation cristallographique (cf. Fig. 6.2). Une fois la nature de l'intermédiaire établie, nous pourrions également simuler l'interaction de plusieurs intermédiaires, éventuellement liés par des dimères interstitiels, également en simulation gros grains (tel que présenté dans la figure 5.3.) afin d'étudier si de telles interactions permettent un assemblage rapide des intermédiaires.

Il serait intéressant de vérifier expérimentalement les observations faites en TMD (page 56) : les résidus testés n'étant présents que dans le Génogroupe I. Nous pourrions muter les résidus correspondant dans le GIII.2 par ceux du GI.1 et vérifier par cryo-EM si les capsides sont plus stables avec ces résidus. Ceci confirmerait l'hypothèse disant que ces résidus participent au maintien des domaines P aux domaines S. Si cette hypothèse est vérifiée, nous pourrions même obtenir des cartes de densité avec une meilleure résolution et pourquoi pas, des cristaux.

Julien Maufront et Jean Lepault ont observé qu'en ajoutant des ions calcium avec des VLP, les domaines P se rapprochaient des domaines S permettant ainsi d'avoir des capsides plus stables. Il serait intéressant d'étudier *in silico* l'effet d'ions divalents sur le dimère dans le contexte de la capside. La structure d'un domaine P de GIII.2 a été co-cristallisée avec des ions Zn^{2+} (Singh *et al.*, 2016, PDBID : 5E6T [171]), mais nous pourrions trouver d'autres sites d'ions divalents avec des méthodes prédictives [174–176]. Des simulations sur un dimère de P avec des sucres ont déjà été réalisées [98, 99] afin d'étudier le comportement du site de liaison aux HBGA se trouvant sur le domaine P2. Il serait intéressant d'étendre ces études avec des dimères complets afin de vérifier quelle incidence la présence d'un sucre au site de liaison de la protéine a sur la conformation du dimère de VP1.

Actuellement Volker Lohmann travaille sur la mise au point de systèmes de culture cellulaire pour des norovirus humains, et Grant Hansman sur les conformations de la capside des norovirus humains ainsi que leur neutralisation. Il serait alors intéressant d'étudier les mutations D10N et E22Q du bras

N-terminal en culture cellulaire ou sur protéine recombinante afin d'étudier l'importance de ces résidus et d'une relocalisation du bras N-terminal dans l'assemblage de la capsidie dans le cadre d'une collaboration commune. Si nous arrivions à comprendre l'assemblage de la capsidie, ne serait-ce que les premières étapes (formation de l'intermédiaire), nous pourrions alors tenter de concevoir un inhibiteur d'assemblage pour se protéger du norovirus. Par exemple, si l'assemblage débute bien par interactions entre domaines S pour former un POD, comme suggéré par Prasad *et al.*, nous pourrions concevoir une molécule empêchant ces interactions.

Enfin, nous avons beaucoup travaillé sur l'assemblage de la capsidie du norovirus mais son désassemblage est tout aussi important dans le processus d'infection de l'hôte. En 2012, Larsson *et al.* ont réalisé des simulations d'1 μ s sur la capsidie du virus de la mosaïque du tabac (T=1) avec et sans les ions calciums structuraux se trouvant notamment aux axes 3. Ils ont alors remarqué un processus d'ouverture de la capsidie aux axes 3 en l'absence de ces ions [182]. En 2011, Bhella *et al.* ont montré par microscopie électronique pour un calicivirus félin que la liaison de la capsidie par un récepteur entraînait une perte de symétrie icosaédrique pouvant être le début d'une étape de dissociation [183]. Nous pourrions imaginer un processus similaire pour le norovirus où la liaison avec un sucre et une diminution du calcium signaleraient l'internalisation du virus dans sa cellule cible et entraîneraient la dissociation de la capsidie. La résolution de la structure cristallographique de la capsidie du norovirus n'est pas suffisante pour observer la présence d'ions divalents, mais nous pourrions cependant simuler la capsidie complète avec les domaines P liés à des sucres et avec différentes concentrations en Ca^{2+} afin d'étudier le phénomène de dissociation de la capsidie du norovirus et mieux comprendre le processus d'infection de l'hôte.

Bibliographie

- [1]Gaelle Debret. „Etude par modélisation moléculaire des propriétés mécaniques d’un système membranaire : le canal mécanosensible MscL au sein de bicouches lipidiques modèles“. fr. Thèse de doct. Université Paris-Diderot - Paris VII, oct. 2007 (cf. p. [xxii](#)).
- [2]Benjamin A. Lopman, Duncan Steele, Carl D. Kirkwood et Umesh D. Parashar. „The Vast and Varied Global Burden of Norovirus : Prospects for Prevention and Control“. In : *PLOS Medicine* 13.4 (2016), e1001999. DOI : [10.1371/journal.pmed.1001999](#) (cf. p. [4](#)).
- [3]Ming Tan et Xi Jiang. „Nanoparticles of Norovirus“. In : *Viral Nanotechnology*. CRC Press, juin 2015, p. 363–370 (cf. p. [4](#), [12](#), [13](#), [21](#)).
- [4]J. Zahorsky. „Hyperemesis hemis or the winter vomiting disease“. In : *Arch Pediatr* (1929) (cf. p. [5](#)).
- [5]Joseph I. Waring. „THE VOMITING DISEASE“. In : *American Journal of Diseases of Children* 64.3 (sept. 1942), p. 482–484. DOI : [10.1001/archpedi.1942.02010090084009](#) (cf. p. [5](#)).
- [6]A. Z. Kapikian, R. G. Wyatt, R. Dolin et al. „Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis“. eng. In : *Journal of Virology* 10.5 (nov. 1972), p. 1075–1081 (cf. p. [5](#), [6](#)).
- [7]J. L. Adler et R. Zickl. „Winter vomiting disease“. eng. In : *The Journal of Infectious Diseases* 119.6 (juin 1969), p. 668–673 (cf. p. [5](#)).
- [8]R. Dolin, N. R. Blacklow, H. DuPont et al. „Transmission of acute infectious nonbacterial gastroenteritis to volunteers by oral administration of stool filtrates“. eng. In : *The Journal of Infectious Diseases* 123.3 (mar. 1971), p. 307–312 (cf. p. [5](#)).
- [9]Elizabeth Robilotti, Stan Deresinski et Benjamin A. Pinsky. „Norovirus“. en. In : *Clinical Microbiology Reviews* 28.1 (jan. 2015), p. 134–164. DOI : [10.1128/CMR.00075-14](#) (cf. p. [6](#)).

- [10] Du-Ping Zheng, Tamie Ando, Rebecca L. Fankhauser et al. „Norovirus classification and proposed strain nomenclature“. In : *Virology* 346.2 (mar. 2006), p. 312–323. DOI : [10.1016/j.virol.2005.11.015](https://doi.org/10.1016/j.virol.2005.11.015) (cf. p. 6, 7).
- [11] X. Jiang, M. Wang, K. Wang et M. K. Estes. „Sequence and genomic organization of Norwalk virus“. eng. In : *Virology* 195.1 (juil. 1993), p. 51–61. DOI : [10.1006/viro.1993.1345](https://doi.org/10.1006/viro.1993.1345) (cf. p. 6, 13).
- [12] P. R. Lambden, E. O. Caul, C. R. Ashley et I. N. Clarke. „Sequence and genome organization of a human small round-structured (Norwalk-like) virus“. en. In : *Science* 259.5094 (jan. 1993), p. 516–519. DOI : [10.1126/science.8380940](https://doi.org/10.1126/science.8380940) (cf. p. 6).
- [13] *Family - Caliciviridae*. San Diego : Elsevier, 2012 (cf. p. 6).
- [14] Andrew MQ King, Elliot Lefkowitz, Michael J. Adams et Eric B. Carstens. *Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses*. en. Elsevier, nov. 2011 (cf. p. 6).
- [15] G. S. Hansman, L. T. P. Doan, T. A. K Nguyen et al. „Detection of norovirus and sapovirus infection among children with gastroenteritis in Ho Chi Minh City, Vietnam“. eng. In : *Archives of Virology* 149.9 (sept. 2004), p. 1673–1688 (cf. p. 6).
- [16] Vito Martella, Eleonora Lorusso, Niccola Decaro et al. „Detection and molecular characterization of a canine norovirus“. eng. In : *Emerging Infectious Diseases* 14.8 (août 2008), p. 1306–1308. DOI : [10.3201/eid1408.080062](https://doi.org/10.3201/eid1408.080062) (cf. p. 7).
- [17] Tung Gia Phan, Kunio Kaneshi, Yuichi Ueda et al. „Genetic heterogeneity, evolution, and recombination in noroviruses“. en. In : *Journal of Medical Virology* 79.9 (sept. 2007), p. 1388–1400. DOI : [10.1002/jmv.20924](https://doi.org/10.1002/jmv.20924) (cf. p. 7).
- [18] Ming Tan et Xi Jiang. „Norovirus and its histo-blood group antigen receptors : an answer to a historical puzzle“. English. In : *Trends in Microbiology* 13.6 (juin 2005), p. 285–293. DOI : [10.1016/j.tim.2005.04.004](https://doi.org/10.1016/j.tim.2005.04.004) (cf. p. 7, 19).
- [19] Peter Graber. *Norovirus*. 2010 (cf. p. 7, 19).
- [20] Grant S. Hansman, Syed Shahzad-ul Hussan, Jason S. McLellan et al. „Structural Basis for Norovirus Inhibition and Fucose Mimicry by Citrate“. en. In : *Journal of Virology* 86.1 (jan. 2012), p. 284–292. DOI : [10.1128/JVI.05909-11](https://doi.org/10.1128/JVI.05909-11) (cf. p. 7, 8, 13, 19, 21).
- [21] Stefan Weichert, Anna Koromyslova, Bishal K. Singh et al. „Structural Basis for Norovirus Inhibition by Human Milk Oligosaccharides“. en. In : *Journal of Virology* 90.9 (mai 2016), p. 4843–4848. DOI : [10.1128/JVI.03223-15](https://doi.org/10.1128/JVI.03223-15) (cf. p. 7, 19, 21).

- [22] Salvador F. Ausar, Thomas R. Foubert, Mary H. Hudson, Thomas S. Vedvick et C. Russell Middaugh. „Conformational Stability and Disassembly of Norwalk Virus-like Particles : EFFECT OF pH AND TEMPERATURE“. en. In : *Journal of Biological Chemistry* 281.28 (juil. 2006), p. 19478–19488. DOI : [10.1074/jbc.M603313200](https://doi.org/10.1074/jbc.M603313200) (cf. p. 7, 87).
- [23] Rainer Fretz-Männer, Andreas Baumgartner, Thomas M. Lüthi et al. *Norovirus : caractéristiques biologiques, épidémiologie, tableau clinique, prévention et recommandations pour la gestion des épidémies*. Rapp. tech. Suisse : Office fédéral de la santé publique suisse, 2005 (cf. p. 7, 9).
- [24] M. R. Evans, R. Meldrum, W. Lane et al. „An outbreak of viral gastroenteritis following environmental contamination at a concert hall.“ In : *Epidemiology and Infection* 129.2 (oct. 2002), p. 355–360 (cf. p. 8, 9).
- [25] J. S. Cheesbrough, L. Barkess-Jones et D. W. Brown. „Possible prolonged environmental survival of small round structured viruses“. English. In : *Journal of Hospital Infection* 35.4 (avr. 1997), p. 325–326. DOI : [10.1016/S0195-6701\(97\)90230-9](https://doi.org/10.1016/S0195-6701(97)90230-9) (cf. p. 8).
- [26] Anna D. Koromyslova, Peter A. White et Grant S. Hansman. „Treatment of norovirus particles with citrate“. In : *Virology* 485 (nov. 2015), p. 199–204. DOI : [10.1016/j.virol.2015.07.009](https://doi.org/10.1016/j.virol.2015.07.009) (cf. p. 8, 13).
- [27] Rebecca L. Fankhauser, Stephan S. Monroe, Jacqueline S. Noel et al. „Epidemiologic and molecular trends of “Norwalk-like viruses” associated with outbreaks of gastroenteritis in the United States“. In : *The Journal of infectious diseases* 186.1 (2002), p. 1–7 (cf. p. 8).
- [28] Sarbelio Moreno-Espinosa, Tibor Farkas et Xi Jiang. „Human caliciviruses and pediatric gastroenteritis“. en. In : *Seminars in Pediatric Infectious Diseases* 15.4 (oct. 2004), p. 237–245. DOI : [10.1053/j.spid.2004.07.004](https://doi.org/10.1053/j.spid.2004.07.004) (cf. p. 9).
- [29] J. S. Cheesbrough, J. Green, C. I. Gallimore, P. A. Wright et D. W. Brown. „Widespread environmental contamination with Norwalk-like viruses (NLV) detected in a prolonged hotel outbreak of gastroenteritis“. eng. In : *Epidemiology and Infection* 125.1 (août 2000), p. 93–98 (cf. p. 8).
- [30] Catherine Makison Booth. „Vomiting Larry : a simulated vomiting system for assessing environmental contamination from projectile vomiting related to norovirus infection“. In : *Journal of Infection Prevention* 15.5 (sept. 2014), p. 176–180. DOI : [10.1177/1757177414545390](https://doi.org/10.1177/1757177414545390) (cf. p. 8).
- [31] Peter F.M. Teunis, Christine L. Moe, Pengbo Liu et al. „Norwalk virus : How infectious is it?“ en. In : *Journal of Medical Virology* 80.8 (août 2008), p. 1468–1476. DOI : [10.1002/jmv.21237](https://doi.org/10.1002/jmv.21237) (cf. p. 8).

- [32]Kim Y. Green, Gaël Belliot, Jean Lin Taylor et al. „A Predominant Role for Norwalk-like Viruses as Agents of Epidemic Gastroenteritis in Maryland Nursing Homes for the Elderly“. In : *The Journal of infectious diseases* 185.2 (jan. 2002), p. 133–146. DOI : [10.1086/338365](https://doi.org/10.1086/338365) (cf. p. 9).
- [33]Elaine H Cramer, David X Gu et Randy E Durbin. „Diarrheal disease on cruise ships, 1990–2000“. en. In : *American Journal of Preventive Medicine* 24.3 (avr. 2003), p. 227–233. DOI : [10.1016/S0749-3797\(02\)00644-X](https://doi.org/10.1016/S0749-3797(02)00644-X) (cf. p. 9).
- [34]Elmira T. Isakbaeva, Marc-Alain Widdowson, R. Suzanne Beard et al. „Norovirus Transmission on Cruise Ship“. In : *Emerging Infectious Diseases* 11.1 (jan. 2005), p. 154–157. DOI : [10.3201/eid1101.040434](https://doi.org/10.3201/eid1101.040434) (cf. p. 9).
- [35]T. G. Metcalf, J. L. Melnick et M. K. Estes. „Environmental virology : from detection of virus in sewage and water by isolation to identification by molecular biology-a trip of over 50 years“. In : *Annual Reviews in Microbiology* 49.1 (1995), p. 461–487 (cf. p. 10).
- [36]Cécile Le Mennec, Sylvain Parnaudeau, Myriam Rumebe et al. „Follow-Up of Norovirus Contamination in an Oyster Production Area Linked to Repeated Outbreaks“. en. In : *Food and Environmental Virology* 9.1 (mar. 2017), p. 54–61. DOI : [10.1007/s12560-016-9260-6](https://doi.org/10.1007/s12560-016-9260-6) (cf. p. 10).
- [37]B. A. Lopman, D. W. Brown et M. Koopmans. „Human caliciviruses in Europe“. eng. In : *Journal of Clinical Virology : The Official Publication of the Pan American Society for Clinical Virology* 24.3 (avr. 2002), p. 137–160 (cf. p. 10).
- [38]Christian Beuret, Andreas Baumgartner et Jakob Schlupe. „Virus-Contaminated Oysters : a Three-Month Monitoring of Oysters Imported to Switzerland“. In : *Applied and Environmental Microbiology* 69.4 (avr. 2003), p. 2292–2297. DOI : [10.1128/AEM.69.4.2292-2297.2003](https://doi.org/10.1128/AEM.69.4.2292-2297.2003) (cf. p. 10).
- [39]Sarah M. Bartsch, Benjamin A. Lopman, Sachiko Ozawa, Aron J. Hall et Bruce Y. Lee. „Global Economic Burden of Norovirus Gastroenteritis“. In : *PLOS ONE* 11.4 (2016), e0151219. DOI : [10.1371/journal.pone.0151219](https://doi.org/10.1371/journal.pone.0151219) (cf. p. 10–12).
- [40]Ming Tan et Xi Jiang. „Vaccine against norovirus“. en. In : *Human Vaccines & Immunotherapeutics* 10.6 (juin 2014), p. 1449–1456. DOI : [10.4161/hv.28626](https://doi.org/10.4161/hv.28626) (cf. p. 12).
- [41]X Jiang, M Wang, D Y Graham et M K Estes. „Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein.“ In : *Journal of Virology* 66.11 (nov. 1992), p. 6527–6532 (cf. p. 12).
- [42]Emily M. Plummer et Marianne Manchester. „Viral nanoparticles and virus-like particles : platforms for contemporary vaccine design“. en. In : *Wiley Interdisciplinary Reviews : Nanomedicine and Nanobiotechnology* 3.2 (mar. 2011), p. 174–196. DOI : [10.1002/wnan.119](https://doi.org/10.1002/wnan.119) (cf. p. 12).

- [43] Judith M. Ball, Michele E. Hardy, Robert L. Atmar, Margaret E. Conner et Mary K. Estes. „Oral Immunization with Recombinant Norwalk Virus-Like Particles Induces a Systemic and Mucosal Immune Response in Mice“. In : *Journal of Virology* 72.2 (fév. 1998), p. 1345–1353 (cf. p. 12).
- [44] Roberto A. Guerrero, Judith M. Ball, Sharon S. Krater et al. „Recombinant Norwalk Virus-Like Particles Administered Intranasally to Mice Induce Systemic and Mucosal (Fecal and Vaginal) Immune Responses“. In : *Journal of Virology* 75.20 (oct. 2001), p. 9713–9722. DOI : [10.1128/JVI.75.20.9713-9722.2001](https://doi.org/10.1128/JVI.75.20.9713-9722.2001) (cf. p. 13).
- [45] Yuanmei Ma et Jianrong Li. „Vesicular Stomatitis Virus as a Vector To Deliver Virus-Like Particles of Human Norovirus : a New Vaccine Candidate against an Important Noncultivable Virus“. In : *Journal of Virology* 85.6 (mar. 2011), p. 2942–2952. DOI : [10.1128/JVI.02332-10](https://doi.org/10.1128/JVI.02332-10) (cf. p. 13).
- [46] Charles Richardson, Robert F. Bargatze, Robert Goodwin et Paul M. Mendelman. „Norovirus virus-like particle vaccines for the prevention of acute gastroenteritis“. In : *Expert Review of Vaccines* 12.2 (fév. 2013), p. 155–167. DOI : [10.1586/erv.12.145](https://doi.org/10.1586/erv.12.145) (cf. p. 13).
- [47] Lucy G. Thorne et Ian G. Goodfellow. „Norovirus gene expression and replication“. eng. In : *The Journal of General Virology* 95.Pt 2 (fév. 2014), p. 278–291. DOI : [10.1099/vir.0.059634-0](https://doi.org/10.1099/vir.0.059634-0) (cf. p. 14).
- [48] Gaël Belliot, Stanislav V. Sosnovtsev, Kyeong-Ok Chang et al. „Norovirus proteinase-polymerase and polymerase are both active forms of RNA-dependent RNA polymerase“. eng. In : *Journal of Virology* 79.4 (fév. 2005), p. 2393–2403. DOI : [10.1128/JVI.79.4.2393-2403.2005](https://doi.org/10.1128/JVI.79.4.2393-2403.2005) (cf. p. 14).
- [49] Gaël Belliot, Stanislav V. Sosnovtsev, Tanaji Mitra et al. „In Vitro Proteolytic Processing of the MD145 Norovirus ORF1 Nonstructural Polyprotein Yields Stable Precursors and Products Similar to Those Detected in Calicivirus-Infected Cells“. In : *Journal of Virology* 77.20 (oct. 2003), p. 10957–10974. DOI : [10.1128/JVI.77.20.10957-10974.2003](https://doi.org/10.1128/JVI.77.20.10957-10974.2003) (cf. p. 14).
- [50] Gaël Belliot, Stanislav V. Sosnovtsev, Kyeong-Ok Chang, Peter McPhie et Kim Y. Green. „Nucleotidylylation of the VPg Protein of a Human Norovirus by its Proteinase-Polymerase Precursor Protein“. In : *Virology* 374.1 (avr. 2008), p. 33–49. DOI : [10.1016/j.virol.2007.12.028](https://doi.org/10.1016/j.virol.2007.12.028) (cf. p. 14).
- [51] Michele E. Hardy. „Norovirus protein structure and function“. In : *FEMS Microbiology Letters* 253.1 (déc. 2005), p. 1–8. DOI : [10.1016/j.femsle.2005.08.031](https://doi.org/10.1016/j.femsle.2005.08.031) (cf. p. 14, 21, 164).
- [52] Gregor Meyers. „Translation of the Minor Capsid Protein of a Calicivirus Is Initiated by a Novel Termination-dependent Reinitiation Mechanism“. en. In : *Journal of Biological Chemistry* 278.36 (sept. 2003), p. 34051–34060. DOI : [10.1074/jbc.M304874200](https://doi.org/10.1074/jbc.M304874200) (cf. p. 14).

- [53]Frederick A. Murphy, E. Paul J. Gibbs, Marian C. Horzinek et Michael J. Studdert. *Veterinary Virology*. en. Academic Press, oct. 1999 (cf. p. 15).
- [54]D. L. Caspar et A. Klug. „Physical principles in the construction of regular viruses“. eng. In : *Cold Spring Harbor Symposia on Quantitative Biology* 27 (1962), p. 1–24 (cf. p. 16, 20).
- [55]B. V. Venkataram Prasad et Michael F. Schmid. „Principles of virus structural organization“. eng. In : *Advances in Experimental Medicine and Biology* 726 (2012), p. 17–47. DOI : [10.1007/978-1-4614-0980-9_3](https://doi.org/10.1007/978-1-4614-0980-9_3) (cf. p. 17, 24).
- [56]B. V. V. Prasad, Michele E. Hardy, Terje Dokland et al. „X-ray Crystallographic Structure of the Norwalk Virus Capsid“. en. In : *Science* 286.5438 (oct. 1999), p. 287–290. DOI : [10.1126/science.286.5438.287](https://doi.org/10.1126/science.286.5438.287) (cf. p. 18, 20, 24, 25, 39, 87, 120–122, 135, 138, 143, 144, 150, 161).
- [57]Sheng Cao, Zhiyong Lou, Ming Tan et al. „Structural Basis for the Recognition of Blood Group Trisaccharides by Norovirus“. en. In : *Journal of Virology* 81.11 (juin 2007), p. 5949–5957. DOI : [10.1128/JVI.00219-07](https://doi.org/10.1128/JVI.00219-07) (cf. p. 18, 19, 21, 34).
- [58]Mikael Nilsson, Kjell-Olof Hedlund, Margareta Thorhagen et al. „Evolution of Human Calicivirus RNA In Vivo : Accumulation of Mutations in the Protruding P2 Domain of the Capsid Leads to Structural Changes and Possibly a New Phenotype“. en. In : *Journal of Virology* 77.24 (déc. 2003), p. 13117–13124. DOI : [10.1128/JVI.77.24.13117-13124.2003](https://doi.org/10.1128/JVI.77.24.13117-13124.2003) (cf. p. 19).
- [59]Jae-Mun Choi, Anne M. Hutson, Mary K. Estes et B. V. Venkataram Prasad. „Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus“. eng. In : *Proceedings of the National Academy of Sciences of the United States of America* 105.27 (juil. 2008), p. 9175–9180. DOI : [10.1073/pnas.0803275105](https://doi.org/10.1073/pnas.0803275105) (cf. p. 19, 21, 35).
- [60]Serge Pérez, Thibault Tubiana, Anne Imberty et Marc Baaden. „Three-dimensional representations of complex carbohydrates and polysaccharides—SweetUnityMol : A video game-based computer graphic software“. In : *Glycobiology* 25.5 (mai 2015), p. 483–491. DOI : [10.1093/glycob/cwu133](https://doi.org/10.1093/glycob/cwu133) (cf. p. 19).
- [61]Zhihan Lv, Alex Tek, Franck Da Silva et al. „Game On, Science - How Video Game Technology May Help Biologists Tackle Visualization Challenges“. In : *PLOS ONE* 8.3 (mar. 2013), e57990. DOI : [10.1371/journal.pone.0057990](https://doi.org/10.1371/journal.pone.0057990) (cf. p. 19).
- [62]B. V. Prasad, M. E. Hardy, X. Jiang et M. K. Estes. „Structure of Norwalk virus“. eng. In : *Archives of Virology. Supplementum* 12 (1996), p. 237–242 (cf. p. 20).
- [63]Yury Khudyakov et Paul Pumpens. *Viral Nanotechnology*. CRC Press, 2015 (cf. p. 21).

- [64]A. Bertolotti-Ciarlet, L. J. White, R. Chen, B. V. V. Prasad et M. K. Estes. „Structural Requirements for the Assembly of Norwalk Virus-Like Particles“. en. In : *Journal of Virology* 76.8 (avr. 2002), p. 4044–4055. DOI : [10.1128/JVI.76.8.4044-4055.2002](https://doi.org/10.1128/JVI.76.8.4044-4055.2002) (cf. p. 21).
- [65]Ming Tan, Rashmi S. Hegde et Xi Jiang. „The P Domain of Norovirus Capsid Protein Forms Dimer and Binds to Histo-Blood Group Antigen Receptors“. In : *Journal of Virology* 78.12 (juin 2004), p. 6233–6242. DOI : [10.1128/JVI.78.12.6233-6242.2004](https://doi.org/10.1128/JVI.78.12.6233-6242.2004) (cf. p. 21, 120).
- [66]E. L. Seah, I. C. Gunesekere, J. A. Marshall et P. J. Wright. „Variation in ORF3 of genogroup 2 Norwalk-like viruses“. en. In : *Archives of Virology* 144.5 (mai 1999), p. 1007–1014. DOI : [10.1007/s007050050563](https://doi.org/10.1007/s007050050563) (cf. p. 21).
- [67]C. Wirblich, H. J. Thiel et G. Meyers. „Genetic map of the calicivirus rabbit hemorrhagic disease virus as deduced from in vitro translation studies.“ en. In : *Journal of Virology* 70.11 (nov. 1996), p. 7974–7983 (cf. p. 21).
- [68]Stanislav V. Sosnovtsev, Gaël Belliot, Kyeong-Ok Chang, Oge Onwudiwe et Kim Y. Green. „Feline Calicivirus VP2 Is Essential for the Production of Infectious Virions“. en. In : *Journal of Virology* 79.7 (avr. 2005), p. 4012–4024. DOI : [10.1128/JVI.79.7.4012-4024.2005](https://doi.org/10.1128/JVI.79.7.4012-4024.2005) (cf. p. 21).
- [69]Roger I. Glass, Jacqueline Noel, Tamie Ando et al. „The Epidemiology of Enteric Caliciviruses from Humans : A Reassessment Using New Diagnostics“. In : *The Journal of Infectious Diseases* 181.Supplement_2 (mai 2000), S254–S261. DOI : [10.1086/315588](https://doi.org/10.1086/315588) (cf. p. 21).
- [70]S. Vongpunsawad, B. V. Venkataram Prasad et M. K. Estes. „Norwalk Virus Minor Capsid Protein VP2 Associates within the VP1 Shell Domain“. en. In : *Journal of Virology* 87.9 (mai 2013), p. 4818–4825. DOI : [10.1128/JVI.03508-12](https://doi.org/10.1128/JVI.03508-12) (cf. p. 21, 22).
- [71]Pamela J. Glass, Laura J. White, Judith M. Ball et al. „Norwalk Virus Open Reading Frame 3 Encodes a Minor Structural Protein“. en. In : *Journal of Virology* 74.14 (juil. 2000), p. 6581–6591. DOI : [10.1128/JVI.74.14.6581-6591.2000](https://doi.org/10.1128/JVI.74.14.6581-6591.2000) (cf. p. 21).
- [72]Stanislav V. Sosnovtsev et Kim Y. Green. „Identification and Genomic Mapping of the ORF3 and VPg Proteins in Feline Calicivirus Virions“. In : *Virology* 277.1 (nov. 2000), p. 193–203. DOI : [10.1006/viro.2000.0579](https://doi.org/10.1006/viro.2000.0579) (cf. p. 21).
- [73]Batsal Devkota, Anton S. Petrov, Sébastien Lemieux et al. „Structural and electrostatic characterization of Pariacoto virus : Implications for viral assembly“. en. In : *Biopolymers* 91.7 (juil. 2009), p. 530–538. DOI : [10.1002/bip.21168](https://doi.org/10.1002/bip.21168) (cf. p. 23).
- [74]Vladimir A. Belyi et M. Muthukumar. „Electrostatic origin of the genome packing in viruses“. In : *Proceedings of the National Academy of Sciences* 103.46 (2006), p. 17174–17178 (cf. p. 23, 24).

- [75]Guillaume Tresset, Clémence Le Coeur, Jean-François Bryche et al. „Norovirus Capsid Proteins Self-Assemble through Biphasic Kinetics via Long-Lived State-like Intermediates“. In : *Journal of the American Chemical Society* 135.41 (oct. 2013), p. 15373–15381. DOI : [10.1021/ja403550f](https://doi.org/10.1021/ja403550f) (cf. p. [25](#), [87](#), [137](#), [138](#), [149](#), [150](#), [160–162](#)).
- [76]Stewart A. Adcock et J. Andrew McCammon. „Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins“. In : *Chemical Reviews* 106.5 (mai 2006), p. 1589–1615. DOI : [10.1021/cr040426m](https://doi.org/10.1021/cr040426m) (cf. p. [26](#)).
- [77]Sebastian Kmiecik, Dominik Gront, Michal Kolinski et al. „Coarse-Grained Protein Models and Their Applications“. In : *Chemical Reviews* 116.14 (juil. 2016), p. 7898–7936. DOI : [10.1021/acs.chemrev.6b00163](https://doi.org/10.1021/acs.chemrev.6b00163) (cf. p. [26](#), [81](#)).
- [78]Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller et Edward Teller. „Equation of State Calculations by Fast Computing Machines“. en. In : *The Journal of Chemical Physics* 21.6 (juin 1953), p. 1087–1092. DOI : [10.1063/1.1699114](https://doi.org/10.1063/1.1699114) (cf. p. [26](#)).
- [79]B. J. Alder et T. E. Wainwright. „Phase Transition for a Hard Sphere System“. In : *The Journal of Chemical Physics* 27.5 (nov. 1957), p. 1208–1209. DOI : [10.1063/1.1743957](https://doi.org/10.1063/1.1743957) (cf. p. [27](#)).
- [80]B. J. Alder et T. E. Wainwright. „Studies in Molecular Dynamics. I. General Method“. In : *The Journal of Chemical Physics* 31.2 (août 1959), p. 459–466. DOI : [10.1063/1.1730376](https://doi.org/10.1063/1.1730376) (cf. p. [27](#)).
- [81]Esra Karakas. *Simulations de dynamique moléculaires du complexe collecteur de lumière de type 2 d'une bactérie pourpre dans différents environnements micellaires et membranaire*. Paris Saclay, mai 2016 (cf. p. [27](#)).
- [82]J. A. McCammon, B. R. Gelin et M. Karplus. „Dynamics of folded proteins“. eng. In : *Nature* 267.5612 (juin 1977), p. 585–590 (cf. p. [27](#)).
- [83]A. Warshel et M. Levitt. „Theoretical studies of enzymic reactions : dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme“. eng. In : *Journal of Molecular Biology* 103.2 (mai 1976), p. 227–249 (cf. p. [27](#)).
- [84]Gongpu Zhao, Juan R. Perilla, Ernest L. Yufenyuy et al. „Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics“. In : *Nature* 497.7451 (mai 2013), p. 643–646. DOI : [10.1038/nature12162](https://doi.org/10.1038/nature12162) (cf. p. [27](#), [28](#)).
- [85]Tyler Reddy, David Shorthouse, Daniel L. Parton et al. „Nothing to Sneeze At : A Dynamic and Integrative Computational Model of an Influenza A Virion“. In : *Structure(London, England :1993)* 23.3 (mar. 2015), p. 584–597. DOI : [10.1016/j.str.2014.12.019](https://doi.org/10.1016/j.str.2014.12.019) (cf. p. [27](#), [28](#)).

- [86]Isseki Yu, Takaharu Mori, Tadashi Ando et al. „Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm“. en. In : *eLife* 5 (nov. 2016), e19274. DOI : [10.7554/eLife.19274](https://doi.org/10.7554/eLife.19274) (cf. p. 27).
- [87]Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson et Klaus Schulten. „Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus“. In : *Structure* 14.3 (mar. 2006), p. 437–449. DOI : [10.1016/j.str.2005.11.014](https://doi.org/10.1016/j.str.2005.11.014) (cf. p. 28).
- [88]Xue Wang, Fengting Xu, Jiasen Liu et al. „Atomic Model of Rabbit Hemorrhagic Disease Virus by Cryo-Electron Microscopy and Crystallography“. In : *PLOS Pathogens* 9.1 (jan. 2013), e1003132. DOI : [10.1371/journal.ppat.1003132](https://doi.org/10.1371/journal.ppat.1003132) (cf. p. 28).
- [89]Boon Chong Goh, Juan R. Perilla, Matthew R. England et al. „Atomic Modeling of an Immature Retroviral Lattice Using Molecular Dynamics and Mutagenesis“. In : *Structure* 23.8 (août 2015), p. 1414–1425. DOI : [10.1016/j.str.2015.05.017](https://doi.org/10.1016/j.str.2015.05.017) (cf. p. 28).
- [90]Y. Andoh, N. Yoshii, A. Yamada et al. „All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution“. In : *The Journal of Chemical Physics* 141.16 (oct. 2014), p. 165101. DOI : [10.1063/1.4897557](https://doi.org/10.1063/1.4897557) (cf. p. 28).
- [91]Anton Arkhipov, Peter L. Freddolino et Klaus Schulten. „Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling“. In : *Structure* 14.12 (déc. 2006), p. 1767–1777. DOI : [10.1016/j.str.2006.10.003](https://doi.org/10.1016/j.str.2006.10.003) (cf. p. 28, 29).
- [92]Bo Chen et Robert Tycko. „Simulated Self-Assembly of the HIV-1 Capsid : Protein Shape and Native Contacts Are Sufficient for Two-Dimensional Lattice Formation“. In : *Biophysical Journal* 100.12 (juin 2011), p. 3035–3044. DOI : [10.1016/j.bpj.2011.05.025](https://doi.org/10.1016/j.bpj.2011.05.025) (cf. p. 29, 30).
- [93]Barbie K. Ganser, Su Li, Victor Y. Klishko, John T. Finch et Wesley I. Sundquist. „Assembly and Analysis of Conical Models for the HIV-1 Core“. en. In : *Science* 283.5398 (jan. 1999), p. 80–83. DOI : [10.1126/science.283.5398.80](https://doi.org/10.1126/science.283.5398.80) (cf. p. 30).
- [94]Jiangguo Liu, Farrah Sadre-Marandi, Simon Tavener et Chaoping Chen. „Curvature Concentrations on the HIV-1 Capsid“. In : *Molecular Based Mathematical Biology* 3.1 (jan. 2015). DOI : [10.1515/mlbmb-2015-0003](https://doi.org/10.1515/mlbmb-2015-0003) (cf. p. 30).
- [95]John M. A. Grime, James F. Dama, Barbie K. Ganser-Pornillos et al. „Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly“. en. In : *Nature Communications* 7 (mai 2016), ncomms11568. DOI : [10.1038/ncomms11568](https://doi.org/10.1038/ncomms11568) (cf. p. 31, 169).

- [96]Iain G. Johnston, Ard A. Louis et Jonathan P. K. Doye. „Modelling the self-assembly of virus capsids“. en. In : *Journal of Physics : Condensed Matter* 22.10 (2010), p. 104101. DOI : [10.1088/0953-8984/22/10/104101](https://doi.org/10.1088/0953-8984/22/10/104101) (cf. p. 31, 32).
- [97]Kevin J. Boyd, Prakhar Bansal, Jun Feng et Eric R. May. „Stability of Norwalk Virus Capsid Protein Interfaces Evaluated by in Silico Nanoindentation“. English. In : *Frontiers in Bioengineering and Biotechnology* 3 (2015). DOI : [10.3389/fbioe.2015.00103](https://doi.org/10.3389/fbioe.2015.00103) (cf. p. 32, 33).
- [98]Chaitanya A. K. Koppisetty, Waqas Nasir, Francesco Strino et al. „Computational studies on the interaction of ABO-active saccharides with the norovirus VA387 capsid protein can explain experimental binding data“. en. In : *Journal of Computer-Aided Molecular Design* 24.5 (mai 2010), p. 423–431. DOI : [10.1007/s10822-010-9353-5](https://doi.org/10.1007/s10822-010-9353-5) (cf. p. 32, 34, 170).
- [99]Magnus Lundborg, Eunus Ali et Göran Widmalm. „An in silico virtual screening study for the design of norovirus inhibitors : fragment-based molecular docking and binding free energy calculations“. In : *Carbohydrate Research. Polysaccharide Structure, a Special Issue in Memory of Lennart Kenne and Malcolm Perry* 378 (août 2013), p. 133–138. DOI : [10.1016/j.carres.2013.03.012](https://doi.org/10.1016/j.carres.2013.03.012) (cf. p. 32, 34, 170).
- [100]Eunüs S. Ali, Harinda Rajapaksha, Jillian M. Carr et Nikolai Petrovsky. „Norovirus drug candidates that inhibit viral capsid attachment to human histo-blood group antigens“. en. In : *Antiviral Research* 133 (sept. 2016), p. 14–22. DOI : [10.1016/j.antiviral.2016.07.006](https://doi.org/10.1016/j.antiviral.2016.07.006) (cf. p. 34, 35).
- [101]Weiming Bu, Aygun Mamedova, Ming Tan et al. „Structural basis for the receptor binding specificity of Norwalk virus“. eng. In : *Journal of Virology* 82.11 (juin 2008), p. 5340–5347. DOI : [10.1128/JVI.00135-08](https://doi.org/10.1128/JVI.00135-08) (cf. p. 35).
- [102]F. C. Bernstein, T. F. Koetzle, G. J. Williams et al. „The Protein Data Bank : a computer-based archival file for macromolecular structures“. eng. In : *Journal of Molecular Biology* 112.3 (mai 1977), p. 535–542 (cf. p. 37).
- [103]Nuala A. O’Leary, Mathew W. Wright, J. Rodney Brister et al. „Reference sequence (RefSeq) database at NCBI : current status, taxonomic expansion, and functional annotation“. In : *Nucleic Acids Research* 44.D1 (jan. 2016), p. D733–D745. DOI : [10.1093/nar/gkv1189](https://doi.org/10.1093/nar/gkv1189) (cf. p. 37).
- [104]Daisuke Kihara, éd. *Protein Structure Prediction*. T. 1137. Methods in Molecular Biology. New York, NY : Springer New York, 2014 (cf. p. 38–41).
- [105]Marco Biasini, Stefan Bienert, Andrew Waterhouse et al. „SWISS-MODEL : modelling protein tertiary and quaternary structure using evolutionary information“. In : *Nucleic Acids Research* 42.W1 (juil. 2014), W252–W258. DOI : [10.1093/nar/gku340](https://doi.org/10.1093/nar/gku340) (cf. p. 38, 122).

- [106]Jianyong Yang, Renxiang Yan, Ambrish Roy et al. „The I-TASSER Suite : protein structure and function prediction“. In : *Nature methods* 12.1 (2015), p. 7–8 (cf. p. 38).
- [107]S F Altschul, T L Madden, A A Schäffer et al. „Gapped BLAST and PSI-BLAST : a new generation of protein database search programs.“ In : *Nucleic Acids Research* 25.17 (sept. 1997), p. 3389–3402 (cf. p. 38).
- [108]Robert D. Finn, Jody Clements et Sean R. Eddy. „HMMER web server : interactive sequence similarity searching“. In : *Nucleic Acids Research* 39.suppl_2 (juil. 2011), W29–W37. DOI : [10.1093/nar/gkr367](https://doi.org/10.1093/nar/gkr367) (cf. p. 39).
- [109]Johannes Söding, Andreas Biegert et Andrei N. Lupas. „The HHpred interactive server for protein homology detection and structure prediction“. In : *Nucleic Acids Research* 33.Web Server issue (juil. 2005), W244–W248. DOI : [10.1093/nar/gki408](https://doi.org/10.1093/nar/gki408) (cf. p. 39).
- [110]Marco Wiltgen et Gernot P. Tilz. „Homology modelling : a review about the method on hand of the diabetic antigen GAD 65 structure prediction“. en. In : *Wiener Medizinische Wochenschrift* 159.5-6 (mar. 2009), p. 112–125. DOI : [10.1007/s10354-009-0662-z](https://doi.org/10.1007/s10354-009-0662-z) (cf. p. 39, 40).
- [111]Burkhard Rost. „Twilight zone of protein sequence alignments“. In : *Protein Engineering, Design and Selection* 12.2 (fév. 1999), p. 85–94. DOI : [10.1093/protein/12.2.85](https://doi.org/10.1093/protein/12.2.85) (cf. p. 39).
- [112]Narayanan Eswar, Ben Webb, Marc A. Marti-Renom et al. „Comparative Protein Structure Modeling Using MODELLER“. en. In : *Current Protocols in Protein Science*. John Wiley & Sons, Inc., 2001 (cf. p. 40, 41, 116, 140).
- [113]A. D. MacKerell, D. Bashford, M. Bellott et al. „All-atom empirical potential for molecular modeling and dynamics studies of proteins“. eng. In : *The Journal of Physical Chemistry. B* 102.18 (avr. 1998), p. 3586–3616. DOI : [10.1021/jp973084f](https://doi.org/10.1021/jp973084f) (cf. p. 40).
- [114]G. Marius Clore, Axel T. Brünger, Martin Karplus et Angela M. Gronenborn. „Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination : a model study of crambin“. In : *Journal of molecular biology* 191.3 (1986), p. 523–551 (cf. p. 40).
- [115]Min-yi Shen et Andrej Sali. „Statistical potential for assessment and prediction of protein structures“. In : *Protein Science : A Publication of the Protein Society* 15.11 (nov. 2006), p. 2507–2524. DOI : [10.1110/ps.062416606](https://doi.org/10.1110/ps.062416606) (cf. p. 41).
- [116]G. N. Ramachandran, C. Ramakrishnan et V. Sasisekharan. „Stereochemistry of polypeptide chain configurations“. In : *Journal of Molecular Biology* 7.1 (juil. 1963), p. 95–99. DOI : [10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6) (cf. p. 41).

- [117]Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo et al. „Improved side-chain torsion potentials for the Amber ff99SB protein force field“. en. In : *Proteins : Structure, Function, and Bioinformatics* 78.8 (juin 2010), p. 1950–1958. DOI : [10.1002/prot.22711](https://doi.org/10.1002/prot.22711) (cf. p. 45).
- [118]P. P. Ewald. „Die Berechnung optischer und elektrostatischer Gitterpotentiale“. en. In : *Annalen der Physik* 369.3 (jan. 1921), p. 253–287. DOI : [10.1002/andp.19213690304](https://doi.org/10.1002/andp.19213690304) (cf. p. 45).
- [119]Christophe Chipot. *Les méthodes numériques de la dynamique moléculaire*. Rapp. tech. Nancy : UMR 7565, 2003 (cf. p. 46).
- [120]W. Pauli. „Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren“. de. In : *Zeitschrift für Physik* 31.1 (fév. 1925), p. 765–783. DOI : [10.1007/BF02980631](https://doi.org/10.1007/BF02980631) (cf. p. 47).
- [121]F. London. „Zur Theorie und Systematik der Molekularkräfte“. de. In : *Zeitschrift für Physik* 63.3-4 (mar. 1930), p. 245–279. DOI : [10.1007/BF01421741](https://doi.org/10.1007/BF01421741) (cf. p. 47).
- [122]W. Humphrey, A. Dalke et K. Schulten. „VMD : visual molecular dynamics“. eng. In : *Journal of Molecular Graphics* 14.1 (fév. 1996), p. 33–38, 27–28 (cf. p. 48, 118).
- [123]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola et J. R. Haak. „Molecular dynamics with coupling to an external bath“. In : *The Journal of Chemical Physics* 81.8 (oct. 1984), p. 3684–3690. DOI : [10.1063/1.448118](https://doi.org/10.1063/1.448118) (cf. p. 50).
- [124]Mark Abraham, Berk Hess, David van der Spoel et Erik Lindahl. *GROMACS User Manual, Version 5.0.7*. 2015 (cf. p. 50–52).
- [125]Giovanni Bussi, Davide Donadio et Michele Parrinello. „Canonical sampling through velocity rescaling“. In : *The Journal of Chemical Physics* 126.1 (jan. 2007), p. 014101. DOI : [10.1063/1.2408420](https://doi.org/10.1063/1.2408420) (cf. p. 50, 60).
- [126]Hans C. Andersen. „Molecular dynamics simulations at constant pressure and/or temperature“. In : *The Journal of Chemical Physics* 72.4 (fév. 1980), p. 2384–2393. DOI : [10.1063/1.439486](https://doi.org/10.1063/1.439486) (cf. p. 51).
- [127]Shūichi Nosé. „A molecular dynamics method for simulations in the canonical ensemble“. In : *Molecular Physics* 52.2 (juin 1984), p. 255–268. DOI : [10.1080/00268978400101201](https://doi.org/10.1080/00268978400101201) (cf. p. 51).
- [128]William G. Hoover. „Canonical dynamics : Equilibrium phase-space distributions“. In : *Physical Review A* 31.3 (mar. 1985), p. 1695–1697. DOI : [10.1103/PhysRevA.31.1695](https://doi.org/10.1103/PhysRevA.31.1695) (cf. p. 51).

- [129]Mark James Abraham, Teemu Murtola, Roland Schulz et al. „GROMACS : High performance molecular simulations through multi-level parallelism from laptops to supercomputers“. In : *SoftwareX* 1–2 (sept. 2015), p. 19–25. DOI : [10.1016/j.softx.2015.06.001](https://doi.org/10.1016/j.softx.2015.06.001) (cf. p. [52](#), [59](#), [60](#)).
- [130]Ryan Bradley et Ravi Radhakrishnan. „Coarse-Grained Models for Protein-Cell Membrane Interactions“. en. In : *Polymers* 5.3 (juil. 2013), p. 890–936. DOI : [10.3390/polym5030890](https://doi.org/10.3390/polym5030890) (cf. p. [53](#)).
- [131]Luca Monticelli, Senthil K. Kandasamy, Xavier Periole et al. „The MARTINI Coarse-Grained Force Field : Extension to Proteins“. In : *Journal of Chemical Theory and Computation* 4.5 (2008), p. 819–834 (cf. p. [53](#), [62](#)).
- [132]Wolfgang Kabsch et Christian Sander. „Dictionary of protein secondary structure : Pattern recognition of hydrogen-bonded and geometrical features“. en. In : *Biopolymers* 22.12 (1983), p. 2577–2637 (cf. p. [53](#)).
- [133]Leonardo G. Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank et Klaus Schulten. „Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics“. In : *Structure (London, England : 1993)* 16.5 (mai 2008), p. 673–683. DOI : [10.1016/j.str.2008.03.005](https://doi.org/10.1016/j.str.2008.03.005) (cf. p. [57](#), [58](#)).
- [134]Ryan McGreevy, Abhishek Singharoy, Qufei Li et al. „xMDFF : molecular dynamics flexible fitting of low-resolution X-ray structures“. In : *Acta Crystallographica Section D : Biological Crystallography* 70.Pt 9 (août 2014), p. 2344–2355. DOI : [10.1107/S1399004714013856](https://doi.org/10.1107/S1399004714013856) (cf. p. [57](#)).
- [135]Claudine Darnault, Anne Volbeda, Eun Jin Kim et al. „Ni-Zn-[Fe4-S4] and Ni-Ni-[Fe4-S4] clusters in closed and open subunits of acetyl-CoA synthase/carbon monoxide dehydrogenase“. eng. In : *Nature Structural Biology* 10.4 (avr. 2003), p. 271–279. DOI : [10.1038/nsb912](https://doi.org/10.1038/nsb912) (cf. p. [58](#)).
- [136]William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey et Michael L. Klein. „Comparison of simple potential functions for simulating liquid water“. In : *The Journal of Chemical Physics* 79.2 (juil. 1983), p. 926–935. DOI : [10.1063/1.445869](https://doi.org/10.1063/1.445869) (cf. p. [59](#), [63](#), [130](#)).
- [137]Tom Darden, Darrin York et Lee Pedersen. „Particle mesh Ewald : An $N \cdot \log(N)$ method for Ewald sums in large systems“. In : *The Journal of Chemical Physics* 98.12 (juin 1993), p. 10089–10092. DOI : [10.1063/1.464397](https://doi.org/10.1063/1.464397) (cf. p. [59](#)).
- [138]Berk Hess*. „P-LINCS : A Parallel Linear Constraint Solver for Molecular Simulation“. In : *Journal of Chemical Theory and Computation* 4.1 (2008), p. 116–122. DOI : [10.1021/ct700200b](https://doi.org/10.1021/ct700200b) (cf. p. [59](#), [62](#)).
- [139]Shuichi Miyamoto et Peter A. Kollman. „Settle : An analytical version of the SHAKE and RATTLE algorithm for rigid water models“. en. In : *Journal of Computational Chemistry* 13.8 (oct. 1992), p. 952–962. DOI : [10.1002/jcc.540130805](https://doi.org/10.1002/jcc.540130805) (cf. p. [60](#)).

- [140]M. Parrinello et A. Rahman. „Polymorphic transitions in single crystals : A new molecular dynamics method“. In : *Journal of Applied Physics* 52.12 (déc. 1981), p. 7182–7190. DOI : [10.1063/1.328693](https://doi.org/10.1063/1.328693) (cf. p. 60).
- [141]Iwona Siuda et Lea Thøgersen. „Conformational flexibility of the leucine binding protein examined by protein domain coarse-grained molecular dynamics“. eng. In : *Journal of Molecular Modeling* 19.11 (nov. 2013), p. 4931–4945. DOI : [10.1007/s00894-013-1991-9](https://doi.org/10.1007/s00894-013-1991-9) (cf. p. 60, 62).
- [142]Djurre H. de Jong, Gurpreet Singh, W. F. Drew Bennett et al. „Improved Parameters for the Martini Coarse-Grained Protein Force Field“. In : *Journal of Chemical Theory and Computation* 9.1 (2013), p. 687–697 (cf. p. 62).
- [143]D. A. Case, T. A. Darden, T. E. Cheatham et al. *AMBER 9*. San Francisco : University of California, 2006 (cf. p. 63).
- [144]David A. Case, Thomas E. Cheatham, Tom Darden et al. „The Amber biomolecular simulation programs“. en. In : *Journal of Computational Chemistry* 26.16 (déc. 2005), p. 1668–1688. DOI : [10.1002/jcc.20290](https://doi.org/10.1002/jcc.20290) (cf. p. 63).
- [145]David A. Pearlman, David A. Case, James W. Caldwell et al. „AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules“. In : *Computer Physics Communications* 91.1 (sept. 1995), p. 1–41. DOI : [10.1016/0010-4655\(95\)00041-D](https://doi.org/10.1016/0010-4655(95)00041-D) (cf. p. 63).
- [146]Dmitri I. Svergun et Michel H. J. Koch. „Small-angle scattering studies of biological macromolecules in solution“. en. In : *Reports on Progress in Physics* 66.10 (2003), p. 1735. DOI : [10.1088/0034-4885/66/10/R05](https://doi.org/10.1088/0034-4885/66/10/R05) (cf. p. 68).
- [147]D I Svergun. „Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.“ In : *Biophysical Journal* 76.6 (juin 1999), p. 2879–2886 (cf. p. 68, 74).
- [148]D. I. Svergun. „Determination of the regularization parameter in indirect-transform methods using perceptual criteria“. en. In : *Journal of Applied Crystallography* 25.4 (août 1992), p. 495–503. DOI : [10.1107/S0021889892001663](https://doi.org/10.1107/S0021889892001663) (cf. p. 69, 74).
- [149]D. Svergun, C. Barberato et M. H. J. Koch. „CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates“. In : *Journal of Applied Crystallography* 28.6 (déc. 1995), p. 768–773. DOI : [10.1107/S0021889895007047](https://doi.org/10.1107/S0021889895007047) (cf. p. 70, 74).
- [150]Guillaume Evrard, Fabien Mareuil, Francois Bontems, Christina Sizun et Javier Perez. „DADIMODO : a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints“. In : *Journal of Applied Crystallography* 44.6 (déc. 2011), p. 1264–1271. DOI : [10.1107/S0021889811035758](https://doi.org/10.1107/S0021889811035758) (cf. p. 71, 74).

- [151] Maxim V. Petoukhov et Dmitri I. Svergun. „Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data“. In : *Biophysical Journal* 89.2 (août 2005), p. 1237–1250. DOI : [10.1529/biophysj.105.064154](https://doi.org/10.1529/biophysj.105.064154) (cf. p. 72).
- [152] Maxim V. Petoukhov, Daniel Franke, Alexander V. Shkumatov et al. „New developments in the ATSAS program package for small-angle scattering data analysis“. In : *Journal of Applied Crystallography* 45.2 (avr. 2012), p. 342–350. DOI : [10.1107/S0021889812007662](https://doi.org/10.1107/S0021889812007662) (cf. p. 72, 74).
- [153] G. Tria, H. D. T. Mertens, M. Kachala et D. I. Svergun. „Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering“. en. In : *IUCrJ* 2.2 (mar. 2015), p. 207–217. DOI : [10.1107/S205225251500202X](https://doi.org/10.1107/S205225251500202X) (cf. p. 73, 74).
- [154] Anna D. Koromyslova et Grant S. Hansman. „Nanobody Binding to a Conserved Epitope Promotes Norovirus Particle Disassembly“. In : *Journal of Virology* 89.5 (déc. 2014), p. 2718–2730. DOI : [10.1128/JVI.03176-14](https://doi.org/10.1128/JVI.03176-14) (cf. p. 73).
- [155] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch et D. I. Svergun. „PRIMUM : a Windows PC-based system for small-angle scattering data analysis“. en. In : *Journal of Applied Crystallography* 36.5 (oct. 2003), p. 1277–1282. DOI : [10.1107/S0021889803012779](https://doi.org/10.1107/S0021889803012779) (cf. p. 74).
- [156] M. B. Kozin et D. I. Svergun. „Automated matching of high- and low-resolution structural models“. en. In : *Journal of Applied Crystallography* 34.1 (fév. 2001), p. 33–41. DOI : [10.1107/S0021889800014126](https://doi.org/10.1107/S0021889800014126) (cf. p. 74).
- [157] Daniel R. Roe et Thomas E. Cheatham. „PTRAJ and CPPTRAJ : Software for Processing and Analysis of Molecular Dynamics Trajectory Data“. In : *Journal of Chemical Theory and Computation* 9.7 (juil. 2013), p. 3084–3095. DOI : [10.1021/ct400341p](https://doi.org/10.1021/ct400341p) (cf. p. 76).
- [158] David A Case Ross C Walker UCSD et GSK E Thomas Cheatham III Carlos Simmerling Adrian Roitberg Kenneth M Merz Ray Luo Tom Darden Junmei Wang Robert E Duke. „Amber 2017 Reference Manual“. In : (2017) (cf. p. 76).
- [159] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang et al. „UCSF Chimera—A visualization system for exploratory research and analysis“. en. In : *Journal of Computational Chemistry* 25.13 (oct. 2004), p. 1605–1612. DOI : [10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084) (cf. p. 76).
- [160] Lawrence A. Kelley, Stephen P. Gardner et Michael J. Sutcliffe. „An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies“. In : *Protein engineering* 9.11 (1996), p. 1063 (cf. p. 76).

- [161] Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan et al. „MDTraj : A Modern Open Library for the Analysis of Molecular Dynamics Trajectories“. In : *Biophysical Journal* 109.8 (oct. 2015), p. 1528–1532. DOI : [10.1016/j.bpj.2015.08.015](https://doi.org/10.1016/j.bpj.2015.08.015) (cf. p. 76).
- [162] Eric Jones, Travis Oliphant, Pearu Peterson et others. *SciPy : Open source scientific tools for Python*. 2001 (cf. p. 76).
- [163] Nathalie Scrima, Célia Caillet-Saguy, Michel Ventura et al. „Two Crucial Early Steps in RNA Synthesis by the Hepatitis C Virus Polymerase Involve a Dual Role of Residue 405“. en. In : *Journal of Virology* 86.13 (juil. 2012), p. 7107–7117. DOI : [10.1128/JVI.00459-12](https://doi.org/10.1128/JVI.00459-12) (cf. p. 80).
- [164] Dina Schneidman-Duhovny, Michal Hammel, John A. Tainer et Andrej Sali. „Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments“. In : *Biophysical Journal* 105.4 (2013), p. 962–974. DOI : [10.1016/j.bpj.2013.07.020](https://doi.org/10.1016/j.bpj.2013.07.020) (cf. p. 82).
- [165] Schrödinger, LLC. „The PyMOL Molecular Graphics System, Version 1.8“. Nov. 2015 (cf. p. 84, 140).
- [166] S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler et G. Bricogne. „Tomato bushy stunt virus at 2.9 | [angst] | resolution“. en. In : *Nature* 276.5686 (nov. 1978), p. 368–373. DOI : [10.1038/276368a0](https://doi.org/10.1038/276368a0) (cf. p. 87).
- [167] Guillaume Tresset, Vincent Decouche, Jean-François Bryche et al. „Unusual self-assembly properties of Norovirus Newbury2 virus-like particles“. In : *Archives of Biochemistry and Biophysics* 537.1 (sept. 2013), p. 144–152. DOI : [10.1016/j.abb.2013.07.003](https://doi.org/10.1016/j.abb.2013.07.003) (cf. p. 87, 88, 139, 159, 164).
- [168] Thibault Tubiana, Yves Boulard et Stéphane Bressanelli. „Dynamics and asymmetry in the dimer of the norovirus major capsid protein“. eng. In : *PloS One* 12.7 (2017), e0182056. DOI : [10.1371/journal.pone.0182056](https://doi.org/10.1371/journal.pone.0182056) (cf. p. 88, 130, 134, 139, 164).
- [169] Ahmet Bakan, Lidio M. Meireles et Ivet Bahar. „ProDy : Protein Dynamics Inferred from Theory and Experiments“. In : *Bioinformatics* 27.11 (juin 2011), p. 1575–1577. DOI : [10.1093/bioinformatics/btr168](https://doi.org/10.1093/bioinformatics/btr168) (cf. p. 118).
- [170] G. S. Hansman, D. W. Taylor, J. S. McLellan et al. „Structural Basis for Broad Detection of Genogroup II Noroviruses by a Monoclonal Antibody That Binds to a Site Occluded in the Viral Particle“. en. In : *Journal of Virology* 86.7 (avr. 2012), p. 3635–3646. DOI : [10.1128/JVI.06868-11](https://doi.org/10.1128/JVI.06868-11) (cf. p. 121).
- [171] Bishal Kumar Singh, Anna Koromyslova et Grant S. Hansman. „Structural analysis of bovine norovirus protruding domain“. In : *Virology* 487 (jan. 2016), p. 296–301. DOI : [10.1016/j.virol.2015.10.022](https://doi.org/10.1016/j.virol.2015.10.022) (cf. p. 121, 122, 170).

- [172] Paul D. Adams, Pavel V. Afonine, Gábor Bunkóczi et al. „PHENIX : a comprehensive Python-based system for macromolecular structure solution“. In : *Acta Crystallographica Section D : Biological Crystallography* 66.Pt 2 (fév. 2010), p. 213–221. DOI : [10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925) (cf. p. [123](#), [140](#), [152](#)).
- [173] DA Case, V Babin, JT Berryman et al. *Amber* 14. 2014 (cf. p. [130](#)).
- [174] M Nayal et E Di Cera. „Predicting Ca(2+)-binding sites in proteins.“ In : *Proceedings of the National Academy of Sciences of the United States of America* 91.2 (jan. 1994), p. 817–821 (cf. p. [133](#), [170](#)).
- [175] Joost W. H. Schymkowitz, Frederic Rousseau, Ivo C. Martins et al. „Prediction of water and metal binding sites and their affinities by using the Fold-X force field“. en. In : *Proceedings of the National Academy of Sciences of the United States of America* 102.29 (juil. 2005), p. 10147–10152. DOI : [10.1073/pnas.0501980102](https://doi.org/10.1073/pnas.0501980102) (cf. p. [133](#), [170](#)).
- [176] Weizhuang Zhou, Grace W. Tang et Russ B. Altman. „High Resolution Prediction of Calcium-Binding Sites in 3D Protein Structures Using FEATURE“. In : *Journal of Chemical Information and Modeling* 55.8 (août 2015), p. 1663–1672. DOI : [10.1021/acs.jcim.5b00367](https://doi.org/10.1021/acs.jcim.5b00367) (cf. p. [133](#), [170](#)).
- [177] Alexandre Hoffmann et Sergei Grudinin. „NOLB : Nonlinear Rigid Block Normal-Mode Analysis Method“. In : *Journal of Chemical Theory and Computation* 13.5 (mai 2017), p. 2123–2134. DOI : [10.1021/acs.jctc.7b00197](https://doi.org/10.1021/acs.jctc.7b00197) (cf. p. [148](#), [163](#)).
- [178] Dima Kozakov, David R. Hall, Bing Xia et al. „The ClusPro web server for protein-protein docking“. en. In : *Nature Protocols* 12.2 (fév. 2017), p. 255–278. DOI : [10.1038/nprot.2016.169](https://doi.org/10.1038/nprot.2016.169) (cf. p. [152](#)).
- [179] Dima Kozakov, Dmitri Beglov, Tanggis Bohnuud et al. „How good is automated protein docking?“ en. In : *Proteins : Structure, Function, and Bioinformatics* 81.12 (déc. 2013), p. 2159–2166. DOI : [10.1002/prot.24403](https://doi.org/10.1002/prot.24403) (cf. p. [152](#)).
- [180] S. R. Comeau, D. W. Gatchell, S. Vajda et C. J. Camacho. „ClusPro : an automated docking and discrimination method for the prediction of protein complexes“. en. In : *Bioinformatics* 20.1 (jan. 2004), p. 45–50. DOI : [10.1093/bioinformatics/btg371](https://doi.org/10.1093/bioinformatics/btg371) (cf. p. [152](#)).
- [181] Dima Kozakov, Ryan Brenke, Stephen R. Comeau et Sandor Vajda. „PIPER : An FFT-based protein docking program with pairwise potentials“. en. In : *Proteins : Structure, Function, and Bioinformatics* 65.2 (août 2006), p. 392–406. DOI : [10.1002/prot.21117](https://doi.org/10.1002/prot.21117) (cf. p. [152](#)).
- [182] Daniel S. D. Larsson, Lars Liljas et David van der Spoel. „Virus Capsid Dissolution Studied by Microsecond Molecular Dynamics Simulations“. In : *PLOS Computational Biology* 8.5 (mai 2012), e1002502. DOI : [10.1371/journal.pcbi.1002502](https://doi.org/10.1371/journal.pcbi.1002502) (cf. p. [171](#)).

- [183]D. Bhella et I. G. Goodfellow. „The Cryo-Electron Microscopy Structure of Feline Calicivirus Bound to Junctional Adhesion Molecule A at 9-Angstrom Resolution Reveals Receptor-Induced Flexibility and Two Distinct Conformational Changes in the Capsid Protein VP1“. en. In : *Journal of Virology* 85.21 (nov. 2011), p. 11381–11390. DOI : [10.1128/JVI.05621-11](https://doi.org/10.1128/JVI.05621-11) (cf. p. [171](#)).

Titre : Dynamique d'assemblage de la capsid des norovirus

Mots clés : norovirus, capsid, simulation de dynamique moléculaire, assemblage, in silico, SAXS

Résumé : Le norovirus est le virus de la gastroentérite virale aiguë chez les humains et les animaux. Chaque année, il est responsable de la mort de près de 220 000 personnes et un coût de près de 65 milliards d'euros.

C'est un petit virus à ARN simple brin de polarité positive. Sa capsid icosaédrique est composée de 90 dimères d'une seule protéine structurale (VP1) à deux domaines : le domaine de spicule (P) exposé à l'environnement biologique, et le domaine *shell* (S) qui est le module d'assemblage de la capsid. Un intermédiaire d'assemblage serait le pentamère de dimères, mais expérimentalement c'est un intermédiaire de 10-11 dimères qui a été rapporté.

En utilisant des approches computationnelles (modélisation par homologie, recuit simulé, simulation de dynamique moléculaire avec des

systèmes tout atomes ou en gros grains) nous cherchons à déterminer les bases moléculaires de l'assemblage de la capsid des norovirus.

Nos résultats sur la brique d'assemblage de la capsid (dimère de VP1) indiquent la présence d'une asymétrie pouvant avoir un rôle dans les premières étapes d'autoassemblage, que nous confirmons expérimentalement par SAXS.

Nos résultats sur de plus gros assemblages (pentamère et hexamère de dimère) révèlent également une rupture de symétrie dans le pentamère de dimère, laissant une position privilégiée pour l'insertion d'un dimère supplémentaire et favorisant une croissance anisotropique.

Nous complétons et précisons ainsi le modèle d'assemblage initialement publié par notre équipe en 2013.

Title : Norovirus capsid assembly

Keywords : norovirus, capsid, molecular dynamic simulation, assembly, in silico, SAXS

Abstract : Noroviruses are the leading cause of acute viral gastroenteritis in humans and animals. Each year, they are responsible for 220 000 deaths and cost close to 65 billion euros.

It's a small single-stranded positive sense RNA virus. The capsid is composed of 90 dimers of a single structural protein (VP1) which consists of two main domains: the protruding domain (P) exposed to the biological environment, and the shell domain (S) which is the assembly module of the capsid.

Using computational approaches such as homology modelling, simulated annealing, molecular dynamic simulations in all atoms and coarse grains systems, we are looking

to determine the molecular basis of norovirus capsid assembly.

Our results on the capsid assembly *brick* (dimer of VP1) indicate the presence of an asymmetry that can have a role in the first stages of self-assembly, which we confirm experimentally by SAXS.

Our results on larger VP1 assemblages (pentamer and hexamer of dimers) also reveal a disruption of symmetry in the pentamer of dimer, leaving a preferred position for the insertion of an additional dimer and promoting anisotropic growth.

We complete and specify the assembly model originally published by our team in 2013.

