

TABLE DES MATIÈRES

I	INTRODUCTION	1
1	MÉCANOBIOLOGIE ET LIAISONS ACCROCHEUSES	3
1.1	La mécanobiologie, un domaine en expansion	4
1.1.1	Aux origines de la mécanobiologie	4
1.1.2	De nouvelles techniques expérimentales	5
1.2	Les liaisons accrocheuses (<i>catch bonds</i>)	8
1.2.1	Définition fonctionnelle	9
1.2.2	Note sur la terminologie	10
1.3	Modélisation des liaisons accrocheuses	11
1.3.1	Modèles conceptuels	11
1.3.2	Modèle historique de DEMBO et al.	13
1.3.3	Modèle contemporain de la liaison glissante	15
1.3.4	Liaison accrocheuse à un état, deux chemins	17
1.3.5	Liaison accrocheuse à deux états	19
1.4	Conclusion	20
	BIBLIOGRAPHIE	22
2	L'ADHÉSINE BACTÉRIENNE FIMH : ÉTAT DE L'ART	27
2.1	Présentation générale	27
2.1.1	Contexte biologique	27
2.1.2	Implication dans les infections urinaires	28
2.1.3	Présentation du pilus de type 1	29
2.2	Vers la découverte d'une protéine grappin	30
2.2.1	Mise en évidence	31
2.2.2	Expériences sur molécule unique	32
2.3	Structure protéique	33
2.3.1	Anatomie générale	33
2.3.2	Variations conformationnelles du domaine lectine	35
2.3.3	Éléments structuraux notables	39
2.4	Modèles proposés	43
2.4.1	Modèle cinétique à deux états	43
2.4.2	Modèle thermodynamique	44
2.5	Conclusion : que reste-t-il à découvrir?	46
	BIBLIOGRAPHIE	49
II	PLASTICITÉ CONFORMATIONNELLE DE FIMH	55
3	EXPLORATION DU PAYSAGE CONFORMATIONNEL	57
3.1	Un outil : la dynamique moléculaire	57
3.1.1	Principe général	57
3.1.2	Paramétrage des champs de force	59
3.1.3	Ambitions et limitations de la dynamique moléculaire	59

3.1.4	Techniques d'échantillonnage renforcé	61
3.2	Préparation du modèle	63
3.2.1	État de protonation	63
3.2.2	Simulations préliminaires et choix du champ de force	65
3.3	Exploration du paysage conformationnel	68
3.3.1	Dynamique moléculaire à échange hamiltonien de répliques	69
3.3.2	Méthodes	71
3.3.3	Résultats	76
3.3.4	Conclusion	84
	BIBLIOGRAPHIE	87
4	THERMODYNAMIQUE D'OUVERTURE DU SITE DE LIAISON	93
4.0.1	Échantillonnage parapluie	93
4.0.2	Méthodes	96
4.0.3	Résultats	99
4.0.4	Conclusion	105
	BIBLIOGRAPHIE	109
	III STABILITÉ DE LA LIAISON ACCROCHEUSE	111
5	MODULATION DE L'AFFINITÉ	113
5.1	Calculs alchimiques d'énergie libre de liaison	113
5.1.1	Enthalpie libre de liaison	113
5.1.2	Approche alchimique par annihilation	114
5.2	Méthodes	117
5.2.1	Protocole de découplage	117
5.2.2	Potentiels de retenue	118
5.2.3	Sélection des configurations initiales	119
5.2.4	Résultats	119
5.3	Conclusion	124
	BIBLIOGRAPHIE	126
6	CINÉTIQUE DE DISSOCIATION	129
6.1	Dynamique moléculaire dirigée	129
6.2	Modèle de dissociation à une étape	130
6.2.1	Méthodes	131
6.2.2	Résultats	136
6.3	Modèle à deux barrières	139
6.3.1	Méthodes	140
6.3.2	Résultats	144
6.4	Conclusion	145
	BIBLIOGRAPHIE	148
	IV CONCLUSION	151
7	CONCLUSION	153
	BIBLIOGRAPHIE	161

V	ANNEXES	165
A	MÉTHODES	167
A.1	Clustering hiérarchique à estimation de densité	167
A.2	Implémentation MV pour le modèle de Markov dépendant de la force	169
B	FIGURES COMPLÉMENTAIRES	171
C	PUBLICATIONS	181
C.1	Dépliement protéique sous force, chaleur et cisaillement fluide	181
C.2	Extension du facteur de von Willebrand	191

TABLE DES FIGURES

FIGURE 1	Dispositifs de spectroscopie de force	7
FIGURE 2	Chambre à écoulement laminaire	8
FIGURE 3	Modèles conceptuels de grappins moléculaires	12
FIGURE 4	Modèle de DEMBO et al.	15
FIGURE 5	Liaison glissante	17
FIGURE 6	Liaison accrocheuse à un état et deux chemins	18
FIGURE 7	Liaison accrocheuse à deux états	20
FIGURE 8	Ligands de FimH	28
FIGURE 9	Structure générale d'un pilus de type 1	30
FIGURE 10	Structure de l'extrémité piliale	34
FIGURE 11	Répertoire conformationnel du domaine lectine	36
FIGURE 12	Structure du site de liaison	40
FIGURE 13	Structure du renflement bêta et du switch alpha	41
FIGURE 14	Structure de la région inter-domaines	42
FIGURE 15	Modèle cinétique à deux états de THOMAS et al.	43
FIGURE 16	Modèle thermodynamique de SAUER et al.	45
FIGURE 17	États de protonation de l'histidine 45	65
FIGURE 18	Flexibilité du domaine lectine avec différents champs de force	68
FIGURE 19	Définition de la DBC	75
FIGURE 20	Variables collectives d'ouverture et d'association	77
FIGURE 21	Distributions de la DBC dans les simulations REST2	78
FIGURE 22	Définition de l'état lié	79
FIGURE 23	Paysage conformationnel exploré au cours de REST2	80
FIGURE 24	Distribution marginale de la variable d'ouverture	82
FIGURE 25	États d'ouverture dans les états A et S	83
FIGURE 26	Évolution de la fraction de configurations liées au cours de REST2	84
FIGURE 27	Environnement de l'histidine 45 et possible ef- fet sur le site de liaison	86
FIGURE 28	Profil d'énergie libre selon la coordonnée d'ou- verture (systèmes diprotonés)	100
FIGURE 29	Profil d'énergie libre selon la coordonnée d'ou- verture (systèmes monoprotés)	101
FIGURE 30	Altération du profil d'énergie libre sous force .	104
FIGURE 31	Prédiction de l'effet de la force à partir des si- mulations non-perturbées	106
FIGURE 32	Cycle thermodynamique pour la méthode al- chimique	115

FIGURE 33	Distribution de la DBC en début et en fin de découplage.	118
FIGURE 34	Enthalpies libres de liaison relatives en fonction des paramètres conformationnels	120
FIGURE 35	Déviations des boucles L2L3 de leur conformation cristallographique	121
FIGURE 36	Cycle thermodynamique reliant état d'ouverture et énergie de liaison	123
FIGURE 37	Trajectoires typiques de simulations à vitesse de charge constante	133
FIGURE 38	Distributions théoriques des forces de rupture pour le modèle de BELL-EVANS	135
FIGURE 39	Force de rupture moyenne en fonction de la vitesse de charge	137
FIGURE 40	Dissociation en deux étapes	140
FIGURE 41	Analyse des trajectoires à force ou vitesse de charge constantes pour le modèle à deux états	143
FIGURE 42	Exposition de la liaison hydrogène N-O2 lors de l'ouverture du site	146
FIGURE B.1	Définition de l'état lié	172
FIGURE B.2	Paysage conformationnel exploré durant REST2 (systèmes diprotonés)	173
FIGURE B.3	Paysage conformationnel exploré durant REST2 (systèmes monoprotonés)	174
FIGURE B.4	Distribution marginale de la variable d'ouverture.	175
FIGURE B.5	Distribution marginale de la variable d'association.	176
FIGURE B.6	Prédiction de l'effet de la force à partir des simulations non-perturbées	177
FIGURE B.7	Exemples de matrices de recouvrement pour le découplage alchimique	178
FIGURE B.8	Écart type de la distribution des forces de rupture en fonction de la vitesse de charge	179

LISTE DES TABLEAUX

TABLE 1	Nomenclature conformationnelle du domaine lectine	36
TABLE 2	Paramètres du modèle cinétique de THOMAS et al.	44
TABLE 3	Comparaison des propriétés de liaison pour plusieurs ligands	45
TABLE 4	Prédiction de l'état de protonation de His45 pour différentes structures	64
TABLE 5	Stabilité du complexe pour différents champs de force	67
TABLE 6	Simulations REST2	72
TABLE 7	Paramètres des fenêtres BEUS	97
TABLE 8	Probabilité des états d'ouverture	102
TABLE 9	Probabilité des états d'ouverture sous force	105
TABLE 10	Paramètres des fenêtres pour le découplage alchimique	117
TABLE 11	Énergies libres de liaison relatives calculées par annihilation du ligand	120
TABLE 12	Énergies de liaison relatives extraites des profils d'ouverture	123
TABLE 13	Estimation des paramètres cinétiques pour le modèle de BELL-EVANS	136
TABLE 14	Estimation des paramètres cinétiques pour le modèle de MARKOV sous force	144
TABLE 15	Estimation des vitesses de dissociation en l'absence de force	145

ACRONYMES

TECHNIQUES EXPÉRIMENTALES

AFM *atomic force microscopy* – microscopie à force atomique

FRET *Förster resonance energy transfer* – transfert d'énergie entre molécules fluorescentes

RMN résonance magnétique nucléaire

SAXS *small angle X-rays scattering* – diffusion des rayons X aux petits angles

LOGICIELS POUR LA SIMULATION

GROMACS *GR*Oningen *MA*chine for *C*hemical *S*imulations

NAMD *NA*noscale *M*olecular *D*ynamics

PLUMED *PLU*gin for *MolE*cular *D*ynamics

MÉTHODES COMPUTATIONNELLES

ABF *adaptive biasing force* – force de biais adaptative

ACP analyse en composantes principales – principal component analysis

AFEP *alchemical free energy perturbation* – perturbation de l'énergie libre par voie alchimique

AUS *adaptive umbrella sampling* – échantillonnage parapluie adaptatif

BEUS *bias exchange umbrella sampling* – échantillonnage parapluie à échange de biais

HREX *hamiltonian replica exchange* – échange de répliques hamiltonien

MBAR *multistate Bennett acceptance ratio*

MC Monte-Carlo

MCNL moindres carrés non-linéaires

MD *molecular dynamics* – dynamique moléculaire

METAD *metadynamics* – métadynamique

MSM *Markov state model* – modèle de MARKOV

MV maximum de vraisemblance

PCA *principal component analysis* – analyse en composantes principales

RMSD *root-mean-square deviation* – racine de l'écart quadratique moyen
 RMSF *root-mean-square fluctuation* – racine des fluctuations quadratiques moyennes
 REMD *replica exchange molecular dynamics* – dynamique moléculaire à échange de répliques
 REST₂ *replica exchange with solute scaling* – échange de répliques avec mise à l'échelle du soluté
 SMD *steered molecular dynamics* – dynamique moléculaire dirigée
 TREM *temperature replica-exchange method* – méthode d'échange de répliques en température
 US *umbrella sampling* – échantillonnage parapluie
 WHAM *weighted histogram analysis method* – analyse par histogramme pondéré

AUTRES

ADN acide désoxyribonucléique
 ARN acide ribonucléique
 BE modèle de BELL-EVANS
 CV *collective variable* – variable collective
 DBC *distance-from-bound-configuration* – distance à la configuration liée
 GPU *graphics processing unit* – processeur graphique
 HM heptyl- α -D-mannose
 IBC *intracellular bacterial community* – communauté bactérienne intracellulaire
 IVU infection des voies urinaires
 MSCRAMM *microbial surface components recognizing adhesive matrix molecules* – composants de la surface microbienne reconnaissant les molécules adhésives de la matrice
 PDB *protein data bank*
 PEL profil d'énergie libre
 PSGL-1 *P-selectin glycoprotein ligand-1* – ligand glycoprotéique-1 de la P-sélectine
 UPEC *uropathogenic Escherichia coli* – *Escherichia coli* uropathogène

Première partie

INTRODUCTION

La première partie de cette thèse est une introduction organisée en deux chapitres. Un premier chapitre remettra en contexte l'émergence du domaine de la mécanobiologie et le développement de nouveaux outils expérimentaux et théoriques ayant conduit à la découverte des liaisons accrocheuses. Le deuxième chapitre s'intéressera plus particulièrement à l'adhésine bactérienne FimH et à l'état de l'art concernant ses propriétés de grappin moléculaire.

Le monde biologique est en perpétuel mouvement. Les organismes vivants sont soumis aux contraintes mécaniques toujours fluctuantes de leur environnement, et sont capables en retour d'exercer des forces afin de se mouvoir. Cette mobilité s'observe à toutes les échelles : les flux organisés de transports vésiculaires au cœur des cellules, orchestrés par de minuscules machines moléculaires, permettent l'acheminement, la maturation et l'excrétion des protéines extracellulaires ; la chorégraphie des chromosomes le long du fuseau mitotique conduit à leur séparation en deux lots identiques, étape indispensable à la division cellulaire ; la contraction concertée de milliers de fibres musculaires génère des forces motrices capables de propulser un être humain à plus de 45 km h^{-1} .

Une propriété fondamentale du vivant est de se maintenir dans un état éloigné de l'équilibre thermodynamique. Le retour spontané vers l'équilibre se décline en plusieurs processus, séparables par la nature des phénomènes observables auxquels ils correspondent et par les échelles de temps qu'ils recouvrent, et dont on peut dégager des équilibres partiels : équilibre chimique, équilibre diffusif, équilibre thermique, équilibre mécanique... À chacun de ces équilibres correspond une ou plusieurs échelles de temps caractéristiques d'évolution du système étudié, qui conditionnent notre capacité expérimentale à isoler et quantifier les processus à l'œuvre. Par exemple, nous pouvons étudier en laboratoire la cinétique d'une réaction enzymatique en faisant l'hypothèse d'un réacteur homogène en termes de composition, température et pression, car les processus de transport chimique, thermique et mécanique sont supposés rapides – ou peu influents – en comparaison à l'avancement de la réaction. Les hypothèses d'équilibres diffusif, mécanique ou thermique ont été centrales dans l'étude de nombreux processus biochimiques en permettant de les modéliser *in vitro* dans les conditions idéalisées de la chimie des solutions. Ces conditions sont pourtant très loin d'être vérifiées dans un grand nombre de contextes physiologiques : les organismes sont des entités anisotropes, contraintes par des forces changeantes et structurées à des échelles qui permettent rarement de faire l'hypothèse d'un mélange homogène et d'un nombre de molécules suffisamment grand pour être traité comme une grandeur continue.

Peut-être en raison de la difficulté à mesurer avec précision les forces s'appliquant aux échelles cellulaires et subcellulaires, l'étude

de l'influence sur les systèmes vivants de leur environnement mécanique est longtemps restée circonscrite au niveau de l'organe, en relative déconnexion avec les apports de la biologie moléculaire et cellulaire en matière de protéomique, génétique et épigénétique. Ces dernières décennies cependant, de nouvelles approches et de nouvelles techniques ont permis de mettre en évidence le rôle crucial des forces mécaniques s'exerçant à l'échelle des macromolécules, des cellules et des tissus, et de comprendre leur intégration en signaux biochimiques interprétés par la machinerie cellulaire, par un processus qualifié de mécanotransduction. L'étude de ces phénomènes s'est progressivement constituée en un nouveau champ, dénommé mécanobiologie.

1.1 LA MÉCANOBIOLOGIE, UN DOMAINE EN EXPANSION

1.1.1 *Aux origines de la mécanobiologie*

L'essor de la mécanobiologie est à relier à la volonté de comprendre l'organisation et l'évolution dynamique de ce vivant fait de compartiments, d'interfaces, de géométries complexes, soumis en permanence à des forces, des frictions, des tensions. Il s'agit d'un champ vaste aux contours mal définis : après tout, d'un point de vue physico-chimique, tout en biologie semble émerger des lois fondamentales de la « mécanique ». Les forces microscopiques impliquées dans les réactions chimiques ne sont pas différentes de celles assurant la cohésion des cellules. Le champ se définit donc davantage par des échelles de temps et d'espace particulières, des techniques expérimentales spécifiques et surtout l'étude de phénomènes biologiques où les concepts macroscopiques de la mécanique classique du solide et des milieux continus (force, viscosité, friction, élasticité) sont pertinents.

Le fondement de la mécanobiologie en tant que champ d'étude prend ses racines dans de nombreuses découvertes au cours de la deuxième moitié du vingtième siècle, dont le mécanisme de contraction musculaire [1, 2], et le domaine s'est épanoui avec la découverte dans les années 1990 des propriétés mécanosensibles des jonctions adhérentes cellule-matrice [3-5]. Ces structures (les contacts focaux), qui assurent l'adhérence des cellules à la matrice extra-cellulaire *via* des protéines filamenteuses nommées intégrines, peuvent croître et se renforcer dans des environnements rigides ou soumis à des forces mécaniques. Les intégrines, en outre, ont la capacité de se lier à des ligands adsorbés sur des surfaces solides, tout en se montrant insensibles aux mêmes ligands en solution. D'un point de vue physico-chimique, de telles propriétés ont de quoi surprendre, les perturbations mécaniques ayant traditionnellement tendance à déstabiliser les assemblages macromoléculaires plutôt qu'à entraîner leur consolidation.

La mécanobiologie se focalise donc historiquement sur l'étude des interactions cellule-cellule et cellule-matrice, avec la perception par les cellules des propriétés mécaniques (rigidité, forme) de leur micro-environnement et ses implications dans l'intégration tissulaire, les migrations cellulaires, le développement et le cancer [5-7]. Un autre aspect concerne l'importance des forces hydrodynamiques dans des phénomènes biologiques, à plusieurs échelles. On peut citer la circulation des érythrocytes et des cellules immunitaires au sein du système vasculaire, la coagulation sanguine, le dépliement ou l'agrégation de protéines (notamment dans les maladies neurodégénératives) et la formation de biofilms bactériens [8-11].

La difficulté de modéliser les processus mécanobiologiques réside dans la multiplicité des échelles de temps et d'espace concernées : comment un stress mécanique à l'échelle de la cellule (μm) voire du tissu (cm) est-il « perçu » par les macromolécules au centre du métabolisme, à savoir les protéines (nm) ? Il en va également du caractère hors-équilibre et non-stationnaire des processus étudiés : comment prédire l'évolution des systèmes étudiés loin des hypothèses simplificatrices d'équilibre mécanique ? Ces questions illustrent le besoin croissant d'approches de modélisation multi-échelle des systèmes biologiques, et notamment à travers les simulations numériques [12].

1.1.2 De nouvelles techniques expérimentales

L'étude mécanique des organismes vivants n'est pas une idée nouvelle en soi. Toutefois, le développement et le perfectionnement de techniques expérimentales adaptées à la mesure et à l'application de forces aux échelles cellulaires, voire moléculaires, a ouvert tout un champ de perspectives sur lesquelles se fonde la mécanobiologie contemporaine.

Ces techniques sont nombreuses et interdisciplinaires : spectroscopie de force, microscopie de force de traction, sondes moléculaires, aspiration par micro-pipette, micropiliers, chambres d'écoulement et autres dispositifs microfluidiques [13-15]. Nous présentons ici rapidement les deux ensembles de techniques qui nous paraissent nécessaires à la compréhension de nos travaux : la spectroscopie de force et les chambres à écoulement laminaire.

Spectroscopie de force

La spectroscopie de force regroupe un ensemble de techniques reposant sur l'utilisation d'une sonde afin d'appliquer une force sur un échantillon et mesurer les déplacements associés. Les trois approches principales sont la microscopie à force atomique, les pinces optiques et les pinces magnétiques [16].

Un actionneur
piézoélectrique
utilise la
déformation induite
par une tension
électrique (effet
piézoélectrique) pour
entraîner un
déplacement.

L'invention de la microscopie à force atomique (*atomic force microscopy*, AFM) en 1985 a démocratisé l'accès expérimental aux échelles sub-nanométriques et à la mesure directe des forces d'interaction qui y règnent. Son principe est le suivant : une sonde en pointe, ou microlevier, dont l'apex est de taille nanométrique, vient balayer un échantillon point par point afin d'en mesurer la topographie. L'échantillon est mû avec une grande précision grâce à un actionneur piézoélectrique. Les mouvements de flexion verticaux dûs aux forces exercées par l'échantillon sur la sonde, dont le module d'élasticité est connu, sont mesurés avec précision par la déflexion d'un laser optique. Cette méthode permet d'imager des surfaces avec une résolution inégalée, proche de l'atome (0,5–1 nm). Par la même occasion, elle autorise la mesure et l'application d'une vaste gamme de forces particulièrement pertinentes aux échelles d'intérêt (10–10⁴ pN).

Cependant, c'est la possibilité de fonctionnaliser l'apex de la sonde et de manipuler des cellules ou molécules uniques qui a ouvert un vaste champ d'application en biologie [16]. L'objectif n'est plus nécessairement de dresser la topographie d'une surface mais d'immobiliser un échantillon biologique entre la sonde et le support (FIGURE 1A). Il est alors possible d'appliquer des forces avec une résolution inférieure au piconewton et de mesurer les déformations associées. Grâce à un système de rétrocontrôle électronique entre le détecteur optique et l'actionneur piézoélectrique plusieurs types de protocoles sont possibles :

- Dans les expériences à force constante (*force clamp*) la distance entre la sonde et le support est constamment ajustée pour maintenir une force constante (c'est-à-dire une déviation du microlevier constante).
- Dans les expériences à rampe de force (*force ramp*), le même principe est utilisé mais la force augmente progressivement au cours du temps.
- Dans les expériences à vitesse constante (*constant speed*) le paramètre de contrôle n'est pas la force mais la distance entre la sonde et le support, qui est accrue linéairement au cours du temps. La force varie librement en réponse aux déformations de l'échantillon. Ce type d'expérience produit typiquement des motifs de variation de la force en « dents de scie », traduisant des changements brusques de la longueur de l'échantillon (par exemple suite au dépliage d'un domaine protéique). Dans la limite où l'échantillon est indéformable ce type d'expérience est équivalent à une augmentation linéaire de la force.

D'autres protocoles plus complexes, par exemple en « zig-zag », ont également été proposés [17].

Reposant sur un principe similaire quoiqu'avec des gammes de paramètres physiques différents, les pinces optiques et pinces magné-

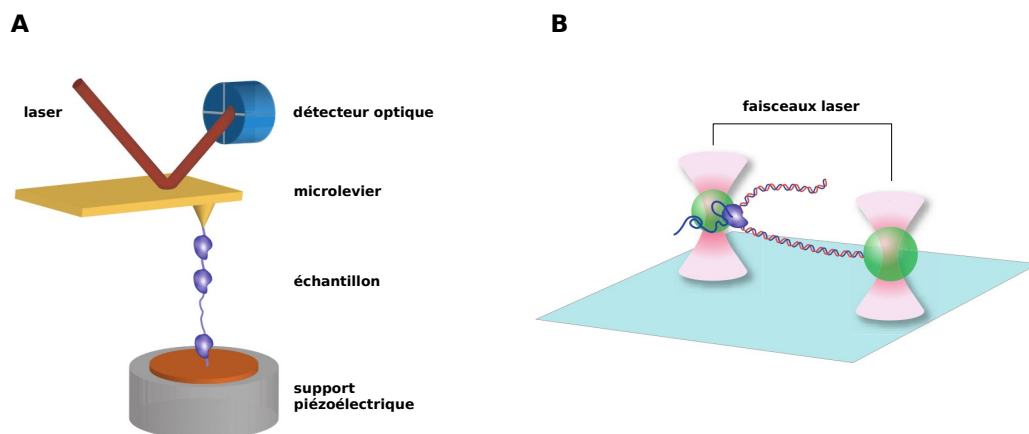


FIGURE 1 – Dispositifs de spectroscopie de force. (A) Schéma d'un dispositif de microscopie à force atomique (AFM) visant à déplier une polyprotéine sous force. L'apex du microlevier est attaché à une extrémité de l'échantillon tandis que l'autre extrémité est fixée à la surface. La hauteur du support piézoélectrique est contrôlée de sorte à ce que le levier exerce une force sur l'échantillon, quantifiée par la déflexion du laser d'un système de détection optique. (B) Schéma d'un dispositif de pince optique. Ici, l'échantillon (un complexe ARN-ARN-polymérase) est maintenu entre deux billes de silice (*en vert*) dont la position est contrôlée par deux faisceaux laser. Repris et adapté de NEUMAN et NAGY, (2008) [16].

tiques (*optical or magnetic tweezers*) permettent également d'avoir accès aux échelles cellulaires et unimoléculaires. Dans ces techniques la sonde n'est pas un microlevier mais une bille micrométrique dont la position est contrôlée à distance, par un faisceau laser formant un piège optique (FIGURE 1B) ou par des champs magnétiques.

De tels dispositifs permettent d'étudier à une échelle très fine les propriétés mécaniques du repliement des biopolymères comme les protéines ou les acides nucléiques [18-22], les changements conformationnels liés à des catalyses enzymatiques [23], ou encore l'adhérence cellule-matrice à l'échelle d'une seule liaison [24].

Chambre à écoulement laminaire

Une autre catégorie de dispositifs expérimentaux utilise les propriétés rhéologiques d'un fluide pour générer des perturbations mécaniques et observer leurs effets sur des cellules. Les chambres à écoulement laminaire (*parallel-plate flow chamber*) permettent de générer un écoulement au voisinage d'une surface et d'étudier les propriétés de certains objets biologiques en présence de contraintes hydrodynamiques.

Le principe est simple : des cellules ou autres particules, généralement porteuses d'un certain récepteur membranaire, sont introduites

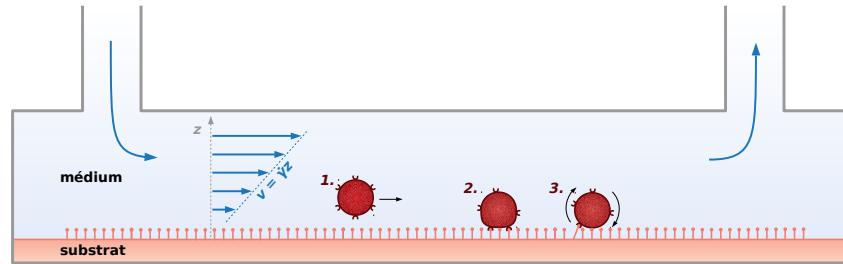


FIGURE 2 – Chambre à écoulement laminaire. L’injection du médium par une pompe génère un profil de vitesse augmentant linéairement avec la distance z au plancher ($v = \gamma z$, avec γ le taux de cisaillement). Les forces de friction résultantes entraînent les cellules en suspension (1) tandis que d’autres adhèrent au substrat et s’immobilisent (2) ou roulent à sa surface (3) en fonction du nombre et de la résistance des interactions liantes récepteur–ligand.

dans une chambre parallélépipédique dont le plancher peut être fonctionnalisé avec des ligands (ou tapissé de cellules exprimant ce ligand). Un flux laminaire est généré par une pompe, résultant en un gradient de vitesse (écoulement laminaire cisailant dit de COUETTE) au niveau de la surface. Les propriétés des objets d’étude en présence des contraintes hydrodynamiques sont observées avec diverses méthodes, généralement par microscopie optique.

Cette technique permet de générer des taux de cisaillement hydrodynamique à la surface de l’ordre de 10^1 – 10^3 s^{-1} , des valeurs comparables à celles trouvées, par exemple, au niveau des endothéliums du système vasculaire [25]. En faisant varier le taux de cisaillement, la viscosité du fluide, la composition du milieu, il est possible d’étudier les propriétés d’adhérence cellule-surface dans diverses conditions. Dans certains cas il est possible de faire l’hypothèse qu’une cellule immobilisée est attachée au substrat par une unique liaison et ainsi avoir accès à des informations cinétiques à l’échelle moléculaire. Ces expériences ont permis d’étudier entre autres des phénomènes tels que le recrutement de leucocytes sanguins par des cellules endothéliales activées dans un contexte inflammatoire [26], l’activation du facteur de von Willebrand, polyprotéine impliquée dans l’hémostase [10], l’adhérence de bactéries pathogènes à divers substrats, et – ce qui va nous intéresser par la suite – les liaisons accrocheuses.

1.2 LES LIAISONS ACCROCHEUSES (*catch bonds*)

Le développement des techniques précédemment mentionnées a notamment permis au cours des dernières décennies de mettre en évidence les propriétés particulières de certains assemblages moléculaires soumis à des forces extérieures. Plutôt que d’être déstabilisés, ces assemblages voient leur durée de vie augmenter en présence d’un

stress mécanique. Un nouveau concept émerge peu à peu au sein des processus biomécaniques : la liaison accrocheuse – ou *catch bond* en anglais.

1.2.1 Définition fonctionnelle

Une définition simple de la liaison accrocheuse est généralement donnée dans la littérature selon ces termes :

Une liaison accrocheuse est une liaison chimique non-covalente dont la durée de vie est accrue lorsqu'elle est soumise à une force de tension mécanique.

Cette définition intuitive présente quelques difficultés qui méritent d'être soulignées.

- La première est inhérente à la notion même de *liaison*. Comme développé par de nombreux et nombreuses auteur·e·s, le concept même de *liaison*, au fondement de la chimie en tant que discipline, n'a pas d'existence propre d'un point de vue physique [27-29]. Il s'agit donc d'une « fiction pratique » que les chimistes mettent à profit pour représenter le caractère plus ou moins durable des structures formées par des ensembles d'atomes, rendues cohésives par un nuage électronique qui ne peut être réduit à un ensemble de traits reliant des points. Ce constat est vrai *a fortiori* pour les liaisons non-covalentes entre macromolécules, qui sont elles aussi soumises à un certain arbitraire du regard de l'expérimentateur et du théoricien. Les multiples interactions interatomiques et la flexibilité interne des partenaires nous contraignent à penser la *liaison* comme un ensemble de configurations (ou un volume de l'espace des phases) respectant un certain critère géométrique. Nous retrouverons cette difficulté lorsqu'il s'agira de définir un *état lié* afin de calculer, par exemple, des énergies libres de *liaison*.
- La seconde est celle de *force de tension mécanique*. La *liaison* ne pouvant pas toujours être représentée comme simple un trait reliant deux points matériels, il faut alors déterminer où la force s'applique, et dans quelle direction, pour pouvoir donner du sens à la *liaison accrocheuse*. En effet, il est possible qu'un même système moléculaire exhibe des propriétés de *liaison* différentes selon qu'on applique une force mécanique dans une direction ou dans une autre, ce qui a donné lieu à la notion de « *liaison accrocheuse directionnelle* » (*directional catch bond*) [33].

La caractéristique non-covalente n'est pas strictement nécessaire pour la notion de liaison accrocheuse. Des travaux récents dans un domaine encore plus jeune, la mécanochimie, mettent en évidence des liaisons accrocheuses moléculaires [30-32] pour lesquelles une réaction de clivage est inhibée par l'application d'une force directionnelle.

1.2.2 Note sur la terminologie

Le terme de "*catch bond*" fut proposé par DEMBO et al. [34] conjointement et par opposition à celui de "*slip bond*". Le caractère (relativement) nouveau de l'étude des *catch bonds* en font une notion encore peu représentée dans la littérature francophone. Notre observation montre que les traductions plus ou moins littérales « liaison accrocheuse » et « liaison glissante » sont généralement employées dans les articles ou travaux de thèse traitant du sujet [35, 36].

Le terme anglais *catch bond* désigne en premier lieu la liaison, relation immatérielle, entre deux entités moléculaires (on dira à titre d'exemple, "*The individual P-selectin-PSGL-1 catch bonds*" ou "*FimH forms catch bonds with mannose*" [37]). Toutefois, il peut également être adjectivé pour qualifier l'un des deux partenaires, généralement celui jugé responsable fonctionnel du caractère particulier de la liaison ("*catch bond-forming protein*" ou plus simplement "*catch-bond protein*" [38]) ou même en tant que substantif, par métonymie (dans "*allosteric site of a catch bond*" [37] ou dans "*mechanical regulation can activate integrins, making them catch bonds*" [39]).

Cette asymétrie, arbitraire d'un point de vue chimique mais compréhensible d'un point de vue biologique, est à retrouver dans la distinction traditionnelle récepteur-ligand.

La langue française ne permet pas autant de flexibilité. Afin d'éviter la lourdeur de « protéine formant des liaisons accrocheuses », nous proposons d'enrichir le vocabulaire existant avec les notions de « grappin moléculaire » ou de « protéine grappin », qui nous semblent constituer une métaphore parlante pour la majorité des systèmes mentionnés.

1.2.2.1 Un phénomène ubiquitaire

Depuis sa proposition théorique par DEMBO et al., (1988) [34], l'existence des liaisons accrocheuses a été progressivement étayée expérimentalement et définitivement confirmée par des expériences AFM sur molécule unique [40]. Depuis, un nombre toujours croissant de couples protéine–ligand impliqués de près ou de loin dans des processus biomécaniques se voient attribuer la capacité à former des liaisons accrocheuses :

- des adhésines bactériennes de la famille MSCRAMM (*microbial surface components recognizing adhesive matrix molecules*), en particulier chez *Staphylococcus*, dont le facteur d'agglutination A (ClfA) ou la protéine de surface SpsD avec le fibrinogène (Fg) humain [24], ou le facteur d'agglutination B (ClfB) avec la loricrine [41],
- la protéine d'adhésion endothéliale sélectine P et son ligand PSGL-1,
- le récepteur des cellules T avec le peptide antigénique du complexe majeur d'histocompatibilité [42],

- des protéines des jonctions d'adhérence avec le cytosquelette (caténine–actine [43], vinculine–actine [33]),
- la glycoprotéine endothéliale ICAM-1 et les érythrocytes infectés par *Plasmodium falciparum* [44],
- des protéines impliquées dans la coagulation sanguine (glycoprotéine plaquettaire GP1BA–facteur de von Willebrand [45]).

Comme nous allons le détailler dans la section suivante, les conditions nécessaires à l'observation de comportements « accrocheurs » sont en réalité peu spécifiques et peu restrictives quand à la nature des entités chimiques impliquées. Il n'est donc pas étonnant que des propriétés de ce type émergent spontanément dans des systèmes macromoléculaires suffisamment complexes, tels que les protéines, et soient sélectionnées et amplifiées par la sélection naturelle.

Dans l'écrasante majorité des cas, les protéines constituent l'un des deux partenaires au moins d'une liaison accrocheuse. Ce constat n'est pas surprenant étant donné la plasticité et la complexité structurelle des polypeptides leur permettant d'acquérir une incroyable diversité fonctionnelle façonnée par l'évolution. Nous mentionnons toutefois la possibilité pour d'autres macromolécules biologique comme les acides nucléiques de constituer des grappins moléculaires : l'ADN, par exemple, dans des systèmes conçus artificiellement [46] ou dans des expériences sous force [47]. Cependant, la pertinence *in vivo* de liaisons accrocheuses sans partenaire protéique n'est, à notre connaissance, pas avérée. Les propriétés mécaniques des acides ribonucléiques (ARN) et leur capacité à former des grappins moléculaires nous semblent largement inexplorées et pourraient avoir des conséquences intéressantes sur l'étude des origines de la vie, notamment dans la fameuse hypothèse du « monde à ARN ».

1.3 MODÉLISATION DES LIAISONS ACCROCHEUSES

Dans cette section nous évoquons les principaux modèles théoriques des liaisons accrocheuses.

1.3.1 Modèles conceptuels

Afin de donner du corps à la notion de liaison accrocheuse et de proposer des structures mécaniques crédibles pouvant produire un tel phénomène, plusieurs modèles conceptuels ont été proposés dans la littérature. Plutôt que de vrais modèles quantitatifs ou mathématisés, ces modèles proposent des analogies structurelles, parfois inspirées d'objets macroscopiques, permettant de se forger une intuition et de former des hypothèses quant à la mécanique microscopique des grappins moléculaires.

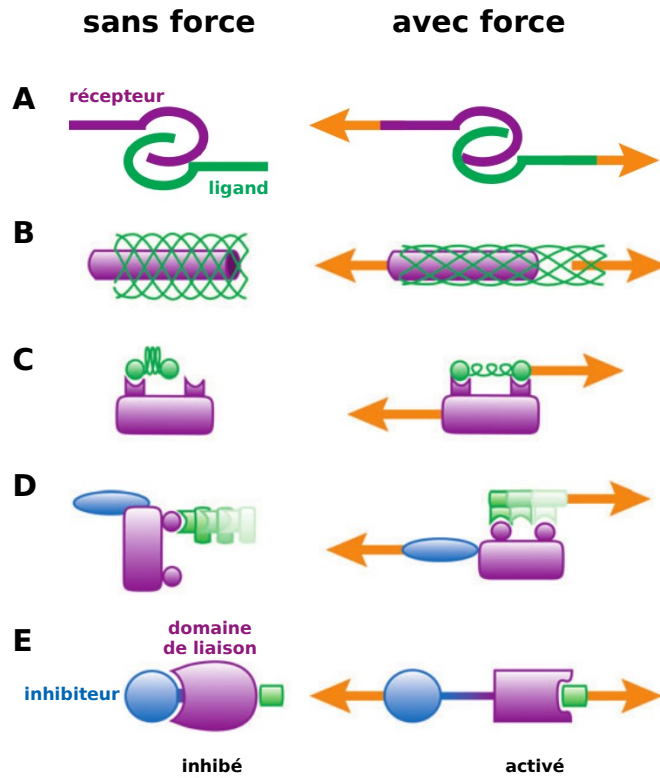


FIGURE 3 – Modèles conceptuels de grappins moléculaires. (A) Harpon, crochet, grappin. (B) Piège à doigts chinois. (C) Modèle de déformation. (D) Modèle de glissement-reliaison. (E) Modèle allostérique. Repris et adapté de THOMAS, VOGEL et SOKURENKO, (2008) [37].

Dans une revue publiée en (2008), THOMAS, VOGEL et SOKURENKO [37] dressent un inventaire de ces modèles conceptuels que nous retranscrivons ci-après. La FIGURE 3 en propose une illustration schématique.

Harpon, crochet ou grappin

Dans ce modèle, comme pour leurs équivalents macroscopiques, les deux partenaires de la liaison peuvent aisément se séparer sous l'effet de forces transversales (générées à l'échelle moléculaire par l'agitation thermique). L'application d'une force provoque l'imbrication des structures, entravant ainsi la dissociation. Si les partenaires sont flexibles la liaison peut également être rompue par déformation sous haute force (FIGURE 3A).

Modèles de déformation

Dans ces modèles, la déformation élastique d'un des partenaires sous l'effet de la force entraîne une augmentation des interactions

liantes et ainsi un renforcement de la liaison. La métaphore du « piège à doigts chinois » en offre une visualisation ludique en faisant référence à ces jouets constitués d'un tube dont l'extension réduit le diamètre, piégeant ainsi le doigt de la victime peu avisée qui tente de s'en extraire en tirant (FIGURE 3B). Dans cet exemple macroscopique, la structure élastique du tube convertit la force d'étirement en forces de frictions tangentielles entre le doigt et le piège. Une structure peut-être plus crédible pour une biomolécule, mais conceptuellement similaire, est représentée en FIGURE 3C, où la déformation du ligand lui confère une meilleure complémentarité de forme avec le récepteur.

Modèle de glissement-reliaison

Ce modèle propose que l'application de la force provoque la réorientation d'une partie d'un partenaire autour d'une charnière. La mise en parallèle de l'interface récepteur-ligand par rapport à la direction de la tension permet alors de générer de nouveaux contacts liants dès que le premier site d'interaction cède, prolongeant ainsi la durée de liaison (FIGURE 3D).

Modèle allostérique

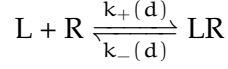
Les modèles précédents sont librement inspirés des intuitions héritées de structures macroscopiques qui nous entourent. Dans l'univers des biologistes et biochimistes, la régulation de l'affinité des protéines pour des ligands solubles s'effectue bien souvent par un mécanisme d'allostérie, dans lequel la fixation d'une molécule régulatrice sur un site secondaire du récepteur entraîne une modification des propriétés du site primaire. Cela conduit naturellement à l'hypothèse que la force, comme un ligand soluble, peut agir sur un site allostérique couplé au site de liaison. Dans un tel modèle, un changement local du site régulateur directement causé par la force – par exemple *via* la séparation d'un domaine inhibiteur (FIGURE 3E) – provoque un changement conformationnel du site de liaison vers un état de plus haute affinité.

Indépendamment de ces modèles conceptuels, des approches plus mathématiques ont été mises en avant, que nous allons maintenant synthétiser.

1.3.2 Modèle historique de DEMBO et al.

Il est intéressant de noter que la notion de liaison accrocheuse proposée par DEMBO et al. [34] est un résultat purement théorique, dont la réalité expérimentale a été confirmée *a posteriori*. Dans ce papier, les auteurs modélisent le processus de liaison-dissociation d'une mo-

lécule adhésive unique entre une membrane portant un récepteur et un support portant un ligand, séparés par une distance d , par un processus chimique stochastique :



où L et R désignent le ligand et le récepteur, et k_{+} et k_{-} les constantes de vitesse, dépendantes de la distance d , des étapes élémentaires de liaison (+) et dissociation (−).

La liaison est alors modélisée comme un simple ressort Hookéen, de constante de raideur κ et de longueur à l'équilibre d_0 . L'énergie libre de liaison G est donc

$$G = G_0 + \frac{1}{2}\kappa(d - d_0)^2 \quad (1)$$

où G_0 est l'énergie libre standard du ressort à vide. En outre la constante d'équilibre en l'absence de stress mécanique est donnée par :

$$K_{eq} = \frac{k_{+}(d_0)}{k_{-}(d_0)}. \quad (2)$$

Afin de respecter la cohérence entre thermodynamique et cinétique on a donc :

$$\frac{k_{+}(d)}{k_{-}(d)} = \exp(-\beta G) = K_{eq} \exp\left[-\beta \frac{1}{2}\kappa(d - d_0)^2\right]. \quad (3)$$

Pour préciser la forme fonctionnelle de $k_{+}(d)$ et $k_{-}(d)$, les auteurs font alors l'hypothèse que l'état de transition a lui aussi une liaison de type harmonique selon d , ne différant de l'état lié que par sa constante de raideur κ^\ddagger .

$$G^\ddagger = G_0^\ddagger + \frac{1}{2}\kappa^\ddagger(d - d_0)^2 \quad (4)$$

On en déduit d'après la théorie de l'état de transition, en supposant le facteur pré-exponentiel constant :

$$k_{-}(d) = k_{-}(d_0) \exp\left[\beta \frac{1}{2}(\kappa - \kappa^\ddagger)(d - d_0)^2\right] \quad (5)$$

On constate alors que la vitesse de détachement k_{-} varie en fonction de l'étirement du ressort. Deux cas se présentent :

- $\kappa > \kappa^\ddagger$: la vitesse de détachement augmente avec la force, la liaison est glissante,
- $\kappa < \kappa^\ddagger$: la vitesse de détachement diminue avec la force, la liaison est accrocheuse.

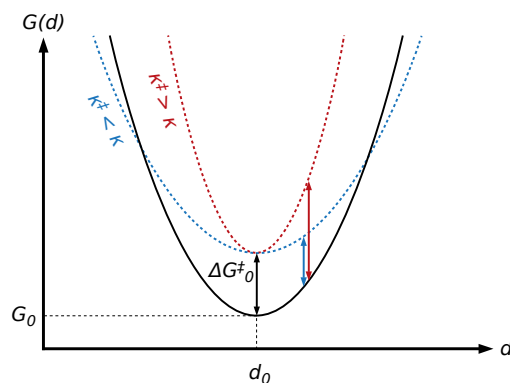


FIGURE 4 – Modèle de DEMBO et al. L'état lié est représenté par un potentiel harmonique (*ligne noire*). La dissociation nécessite une énergie d'activation ΔG^\ddagger qui augmente ou diminue avec l'extension $d - d_0$ selon que la constante de raideur de l'état de transition est supérieure (*en rouge*) ou inférieure (*en bleu*) à celle de l'état lié, conduisant à un comportement accrocheur ou glissant respectivement.

Bien qu'il prédise mathématiquement la possibilité d'existence des liaisons accrocheuses, ce modèle pose plusieurs problèmes :

- La forme harmonique de la surface d'énergie libre des états lié et de transition implique que la vitesse de dissociation tend vers zéro pour une distance de séparation d infinie. Dans ces conditions il est impossible de prédire une réversion vers une liaison glissante au delà d'un certain seuil de force appliquée. Or, une telle propriété est attendue et observée pour des systèmes réels.
- Pour des distances de séparation $d < d_0$ le modèle prédit également une liaison accrocheuse.

En outre les hypothèses sur la forme du paysage énergétique sont contraignantes et pas nécessairement réalistes : nous nous attendons par exemple à ce que l'état de transition se situe plutôt en amont de la coordonnée d ($d_0^\ddagger > d_0$) et non selon une coordonnée parfaitement transversale.

1.3.3 Modèle contemporain de la liaison glissante

Avant de s'intéresser aux modèles plus récents de liaison accrocheuse, commençons par décrire la situation classique d'une liaison glissante. L'un des cadres théoriques les plus rigoureux pour décrire les réactions chimiques au sein des systèmes biomoléculaires est celui hérité des travaux de KRAMERS sur la cinétique chimique suramortie en phase liquide condensée [48]. L'état lié est modélisé par un puits sur une surface d'énergie libre, confiné par une haute barrière énergétique ΔG_0^\ddagger . L'échappement vers l'état non-lié s'effectue selon un chemin préférentiel à travers un col de la barrière. Dans un système

réaliste, il existe potentiellement un grand nombre de tels chemins. Pour simplifier, il est pratique de considérer que l'application d'une force extérieure F sélectionne un de ces chemins, que l'on peut alors projeter selon une unique coordonnée de réaction x . Dans ces conditions, la vitesse de dissociation k s'exprime simplement :

$$k(F) = A \exp[-\Delta G^\ddagger(F)/k_B T] \quad (6)$$

où A est un facteur pré-exponentiel dépendant de la forme et des propriétés diffusives le long du profil énergétique, T est la température et k_B la constante de BOLTZMANN. $\Delta G^\ddagger(F)$ exprime la dépendance de la hauteur de la barrière énergétique vis à vis de la force extérieure. Pour une barrière haute et raide, la forme de la barrière et la position de son sommet peuvent être considérées comme indépendantes de F par hypothèse. En revanche, sa hauteur est affectée proportionnellement à la projection de la barrière sur la direction d'application de la force, c'est-à-dire :

$$\Delta G^\ddagger(F) = \Delta G_0^\ddagger - F\Delta x \quad (7)$$

où $\Delta x = \langle x^\ddagger \cos \theta \rangle$ est la moyenne d'ensemble de la distance x^\ddagger entre l'état lié de départ et l'état de transition le long de la coordonnée x et projetée sur la direction d'application de la force à un angle θ .

La combinaison des équations 6 et 7 permet de réécrire

$$\begin{aligned} k(F) &= A \exp[-(\Delta G_0^\ddagger - F\Delta x)/k_B T] \\ &= A \exp[-\Delta G_0^\ddagger/k_B T] \exp[F\Delta x/k_B T] \\ &= k_0 \exp[F\Delta x/k_B T] \end{aligned} \quad (8)$$

où l'on a noté k_0 la vitesse de réaction en l'absence de force.

L'équation 8 correspond à l'expression phénoménologique de BELL-EVANS [49, 50] et est fréquemment utilisée pour interpréter les résultats de spectroscopie de force sur molécules uniques, par exemple afin d'extrapoler les paramètres k_0 et Δx à partir d'une série d'expériences. Sa validité est empiriquement bien testée pour des gammes de forces faibles à modérées, bien que la signification des paramètres obtenus soit sujette à discussion [51, 52].

L'effet de l'application d'une force extérieure peut être compris comme l'augmentation de la probabilité d'atteindre le haut de la barrière énergétique grâce à l'abaissement de celle-ci d'un facteur $F\Delta x$, correspondant au travail mécanique effectif moyen fourni par cette force entre l'état réactif et l'état de transition. Dans l'hypothèse d'une fréquence pré-exponentielle A indépendante de la force, cela se traduit donc par une augmentation de la vitesse de dissociation qui dépend exponentiellement de l'amplitude de la tension mécanique appliquée.

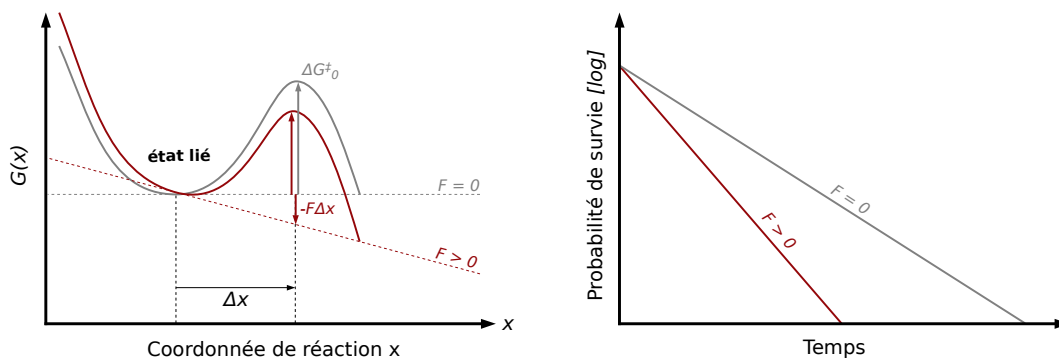


FIGURE 5 – Liaison glissante. (gauche) Projection unidimensionnelle de la surface d'énergie potentielle sur la direction de la force. (droite) Probabilité de survie de la liaison au cours du temps en l'absence et en présence de force.

Notons que l'application d'une force dans la direction opposée provoque au contraire un accroissement de la barrière et donc un allongement de la durée de liaison – ce qui est le propre d'une liaison accrocheuse. Une autre façon équivalente d'envisager cette situation est de considérer un $\Delta x < 0$. Dans une situation concrète toutefois, une force de tension appliquée en deux points situés sur les deux partenaires d'un complexe (ou de manière quasi-équivalente, appliquée en un point sur l'un des partenaires tandis que l'autre est ancré à un point fixe) aura tendance à générer un couple qui aligne de manière unique l'axe de liaison avec la direction d'application. En toute probabilité, le chemin de dissociation d'un tel complexe moléculaire se retrouve plus ou moins colinéaire et de même orientation que la direction de la force. Cela explique l'écrasante majorité des situations dans lesquelles les liaisons se révèlent glissantes quelle que soit la directionnalité de la tension appliquée. Il en résulte que les liaisons accrocheuses correspondent à des exceptions où la directionnalité de la force ne coïncide pas avec celle du chemin réactionnel le plus probable en son absence.

1.3.4 Liaison accrocheuse à un état, deux chemins

La manifestation la plus simple d'une telle exception se produit lorsque le chemin réactionnel sélectionné par la force est différent de celui privilégié sans perturbation mécanique. Dans ce cas on peut modéliser le profil énergétique par deux barrières de part et d'autre de l'état lié (FIGURE 6). On se retrouve dans une situation équivalente à deux liaisons glissantes telles que décrites précédemment, chacune avec ses paramètres de hauteur de barrière $\Delta G^\ddagger_{1,0}$ et $\Delta G^\ddagger_{2,0}$ et de distance effective à l'état de transition Δx_1 et Δx_2 . Le chemin « ac-

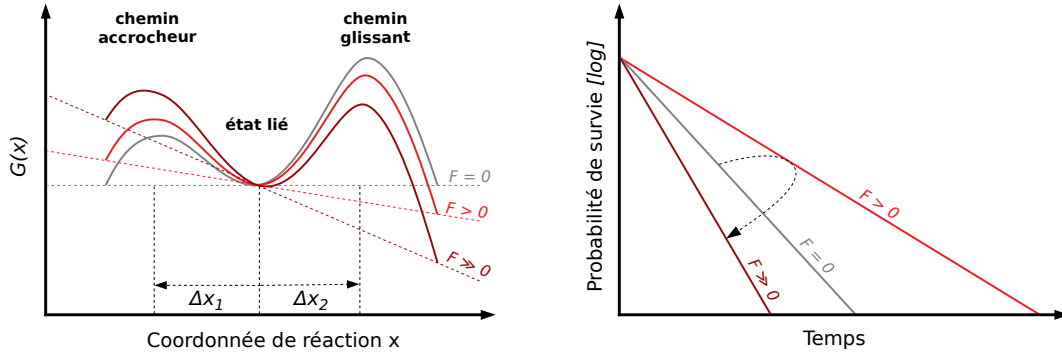


FIGURE 6 – Liaison accrocheuse à un état et deux chemins. (*gauche*) Projection unidimensionnelle de la surface d'énergie potentielle sur la direction de la force. (*droite*) Prédiction de la durée de vie de la liaison pour plusieurs amplitudes de force constante. La vitesse de dissociation augmente avec la force (régime accrocheur) jusqu'à un certain seuil, puis diminue (régime glissant).

crocheur » est alors caractérisé par $\Delta x_1 < 0$ tandis que le chemin « glissant » correspond à $\Delta x_2 > 0$.



Pour obtenir un grappin moléculaire, le chemin « accrocheur » doit dominer à basse force, ce qui suppose $k_{1,0} \gg k_{2,0}$, ce qui se traduit généralement par $\Delta G_{1,0}^\ddagger \gg \Delta G_{2,0}^\ddagger$. Les constantes de vitesse pour chacun des deux chemins de dissociation s'expriment alors

$$\begin{aligned} k_1(F) &= k_{1,0} \exp[F\Delta x_1/k_B T] \\ k_2(F) &= k_{2,0} \exp[F\Delta x_2/k_B T] \end{aligned} \quad (9)$$

En toute rigueur, les facteurs pré-exponentiels A_1 et A_2 n'ont aucune raison d'être égaux.

Il est toutefois raisonnable de les supposer du même ordre de grandeur et de considérer le facteur exponentiel comme dominant dans l'expression des constantes de vitesse.

et la vitesse totale de dissociation est simplement

$$k_{\text{off}}(F) = k_1(F) + k_2(F). \quad (10)$$

Pour des forces petites, le chemin privilégié est le chemin accrocheur car $k_1 \gg k_2$ et $k_{\text{off}} \sim k_1$. Le modèle prédit alors une diminution de la vitesse de dissociation k_{off} avec des forces croissantes. Au delà d'une force critique telle que $k_1 \approx k_2$, le chemin glissant domine ($k_2 \gg k_1$) et la vitesse k_{off} augmente avec la force (FIGURE 6).

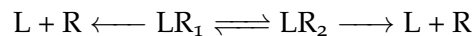
Ce modèle simple est suffisant pour reproduire les données expérimentales de certains grappins moléculaires tels que les L-sélectines [53] ou les myosines [54]. En revanche il ne permet pas de reproduire des systèmes où la distribution de temps de rupture est multimodale.

Plusieurs modèles conceptuels sont compatibles avec le modèle cinétique résultant. Le modèle du crochet ou harpon est parfaitement adapté, avec un chemin accrocheur correspondant justement au « dé-

crochage » et un chemin glissant correspondant à la rupture forcée par déformation des partenaires liés. Une autre possibilité est un modèle allostérique à un état actif et un état inactif incapable de liaison : la dissociation peut s'effectuer *via* l'état actif (chemin accrocheur) ou par transition vers l'état inactif (chemin glissant). Il est donc important de signaler que le modèle cinétique résultant ne permet pas nécessairement de discriminer différents modèles conceptuels ou structuraux.

1.3.5 Liaison accrocheuse à deux états

Si l'on s'autorise un niveau de complexité supplémentaire, on peut construire un modèle de liaison accrocheuse à deux états liés de basse énergie séparés par une barrière. Chacun des deux états se comporte comme une liaison glissante mais de paramètres cinétiques différents. Un tel modèle possède donc quatre chemins possibles : deux pour l'interconversion entre les deux états liés et deux réactions de dissociation.



Ce modèle est compatible avec un modèle conceptuel allostérique à deux états liés. En fonction des différents paramètres du modèle, le comportement de la liaison accrocheuse peut différer. Si la transition entre les états inactif et actif est lente relativement à la dissociation, le système se retrouve rarement dans l'état actif. Lorsque cela se produit, la demi-vie de la liaison augmente significativement. Ce caractère biphasique se traduit par une distribution biexponentielle du temps de dissociation, comme illustré sur la FIGURE 7. Tant que l'augmentation de la force appliquée favorise la transition vers l'état de haute affinité plus qu'elle n'accélère la dissociation de l'état de basse affinité, la distribution du temps de rupture se déplace vers les temps longs. Si la transition entre les deux états est rapide, on se retrouve avec un pré-équilibre et la dissociation suit une seule constante de vitesse apparente. Dans ce cas le modèle peut être indiscernable du modèle à un état vu précédemment.

Comme nous le verrons plus en détail, le modèle à deux états est le plus « simple » pouvant expliquer les propriétés de l'adhésine FimH. Le modèle à un état et deux chemins n'est pas capable d'expliquer la distribution biexponentielle des temps de rupture (ou la distribution bimodale des forces de ruptures pour des expériences sous force à vitesse constante).

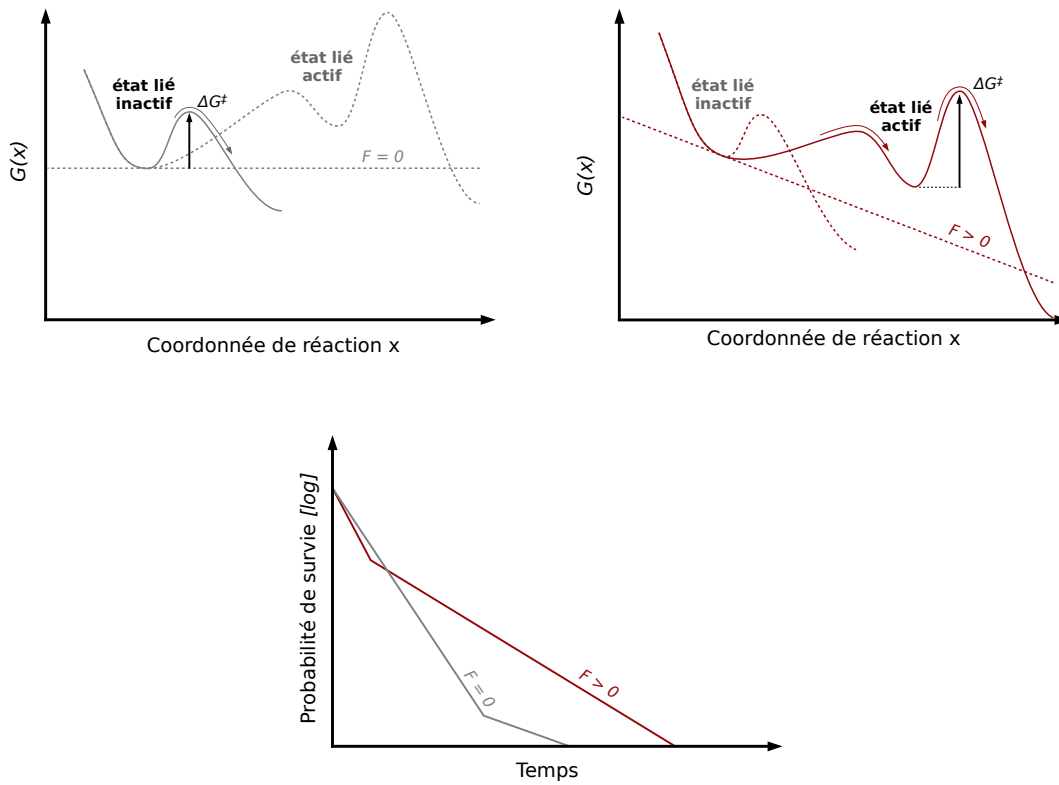


FIGURE 7 – Liaison accrocheuse à deux état. (*haut, gauche*) Profil d'énergie libre en l'absence de force. Le chemin préférentiel est celui ne passant pas par l'état activé. (*haut, droite*) Profil d'énergie libre en présence d'une force. Le chemin préférentiel passe par l'état activé dont la barrière de dissociation est plus élevée. (*bas*) Probabilité de survie de la liaison au cours du temps en l'absence et en présence de force. La coexistence des deux chemins peut se traduire par une distribution biexponentielle des temps de survie.

1.4 CONCLUSION

Cette courte introduction nous a permis d'exposer le concept clé de liaison accrocheuse et comment il semble être au cœur d'un grand nombre de processus étudiés en mécanobiologie.

Nous avons vu que des modèles théoriques relativement simples permettaient de reproduire qualitativement ce phénomène. Il n'est cependant pas toujours évident de discriminer ces modèles entre eux au seul regard des données accessibles expérimentalement. En particulier, il n'existe pas de correspondance directe entre différents paradigmes « conceptuels » ou structuraux et les quelques modèles quantitatifs présentés ici. Des modèles cinétiques en apparence simples peuvent dissimuler des mécanismes conformationnels complexes et l'interprétation de l'origine moléculaire des paramètres (k_0 , Δx , ...) nécessite l'obtention de données structurales en complément des expériences sous force. Ces données peuvent provenir de techniques

expérimentales comme la cristallographie aux rayons X, de la spectroscopie RMN, de la diffusion des rayons X aux petits angles (SAXS), mais aussi – et de plus en plus – de prédictions théoriques, à travers les simulations (dynamique moléculaire et méthodes de Monte-Carlo) ou l'apprentissage machine (nous pensons par exemple aux récents exploits quant à la prédiction *in silico* de la structure native de protéines grâce aux modèles d'apprentissage profond [55, 56]).

Dans le chapitre suivant nous allons nous intéresser de plus près à l'état de l'art d'un exemple particulièrement bien étudié de protéine grappin, l'adhésine FimH, qui fera également l'objet de nos recherches. Nous y verrons comment le modèle cinétique à deux états (1.3.5) parvient, dans l'ensemble, à expliquer son fonctionnement, mais nous verrons aussi que les données structurales suggèrent une réalité microscopique plus complexe.

RÉFÉRENCES

- [1] H. HUXLEY et J. HANSON. Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation. *Nature*, **173**, 973-976 (1954).
DOI : [10.1038/173973a0](https://doi.org/10.1038/173973a0)
- [2] H. E. HUXLEY. The mechanism of muscular contraction. *Science*, **164**, 1356-1365 (1969).
DOI : [10.1126/science.164.3886.1356](https://doi.org/10.1126/science.164.3886.1356)
- [3] U. S. SCHWARZ. Mechanobiology by the numbers : a close relationship between biology and physics. *Nature Reviews Molecular Cell Biology*, **18**, 711-712 (2017).
DOI : [10.1038/nrm.2017.109](https://doi.org/10.1038/nrm.2017.109)
- [4] C. S. CHEN. Forces as regulators of cell adhesions. *Nature Reviews Molecular Cell Biology*, **18**, 715-715 (2017).
DOI : [10.1038/nrm.2017.112](https://doi.org/10.1038/nrm.2017.112)
- [5] T. ISKRATSCH, H. WOLFENSON et M. P. SHEETZ. Appreciating force and shape — the rise of mechanotransduction in cell biology. *Nature Reviews Molecular Cell Biology*, **15**, 825-833 (2014).
DOI : [10.1038/nrm3903](https://doi.org/10.1038/nrm3903)
- [6] J. H. .-C. WANG et B. P. THAMPATTY. An Introductory Review of Cell Mechanobiology. *Biomechanics and Modeling in Mechanobiology*, **5**, 1-16 (2006).
DOI : [10.1007/s10237-005-0012-z](https://doi.org/10.1007/s10237-005-0012-z)
- [7] M.-E. FERNANDEZ-SANCHEZ, T. BRUNET, J.-C. RÖPER et E. FARGE. Mechanotransduction's Impact on Animal Development, Evolution, and Tumorigenesis. *Annual Review of Cell and Developmental Biology*, **31**, 373-397 (2015).
DOI : [10.1146/annurev-cellbio-102314-112441](https://doi.org/10.1146/annurev-cellbio-102314-112441)
- [8] P. SZYMCAK et M. CIEPLAK. Proteins in a shear flow. *The Journal of Chemical Physics*, **127**, 155106 (2007).
DOI : [10.1063/1.2795725](https://doi.org/10.1063/1.2795725)
- [9] I. B. BEKARD, P. ASIMAKIS, J. BERTOLINI et D. E. DUNSTAN. The effects of shear flow on protein structure and function. *Biopolymers*, **95**, 733-745 (2011).
DOI : [10.1002/bip.21646](https://doi.org/10.1002/bip.21646)
- [10] H. FU, Y. JIANG, D. YANG, F. SCHEIFLINGER, W. P. WONG et T. A. SPRINGER. Flow-induced elongation of von Willebrand factor precedes tension-dependent activation. *Nature Communications*, **8**, 324 (2017).
DOI : [10.1038/s41467-017-00230-2](https://doi.org/10.1038/s41467-017-00230-2)
- [11] D. HUBER, A. OSKOEI, X. CASADEVALL I SOLVAS, A. DEMELLO et G. V. KAIGALA. Hydrodynamics in Cell Studies. *Chemical Reviews*, **118**, 2042-2079 (2018).
DOI : [10.1021/acs.chemrev.7b00317](https://doi.org/10.1021/acs.chemrev.7b00317)
- [12] M. CHIRICOTTO, F. STERPONE, P. DERREUMAUX et S. MELCHIONNA. Multiscale simulation of molecular processes in cellular environments. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, **374**, 20160225 (2016).
DOI : [10.1098/rsta.2016.0225](https://doi.org/10.1098/rsta.2016.0225)

- [13] W. J. POLACHEK et C. S. CHEN. Measuring cell-generated forces : a guide to the available tools. *Nature Methods*, **13**, 415-423 (2016).
DOI : [10.1038/nmeth.3834](https://doi.org/10.1038/nmeth.3834)
- [14] P. ROCA-CUSACHS, V. CONTE et X. TREPAT. Quantifying forces in cell biology. *Nature Cell Biology*, **19**, 742-751 (2017).
DOI : [10.1038/ncb3564](https://doi.org/10.1038/ncb3564)
- [15] G. A. SHAMSAN et D. J. ODDE. Emerging technologies in mechano-transduction research. *Current Opinion in Chemical Biology*, **53**, 125-130 (2019).
DOI : [10.1016/j.cbpa.2019.08.002](https://doi.org/10.1016/j.cbpa.2019.08.002)
- [16] K. C. NEUMAN et A. NAGY. Single-molecule force spectroscopy : optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, **5**, 491-505 (2008).
DOI : [10.1038/nmeth.1218](https://doi.org/10.1038/nmeth.1218)
- [17] M. A. NASH. Zig Zag AFM Protocol Reveals New Intermediate Folding States of Bacteriorhodopsin. *Biophysical Journal*, **118**, 538-540 (2020).
DOI : [10.1016/j.bpj.2019.12.004](https://doi.org/10.1016/j.bpj.2019.12.004)
- [18] J. SCHÖNFELDER, R. PEREZ-JIMENEZ et V. MUÑOZ. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nature Communications*, **7**, 11777 (2016).
DOI : [10.1038/ncomms11777](https://doi.org/10.1038/ncomms11777)
- [19] G. STIRNEMANN, D. GIGANTI, J. M. FERNANDEZ et B. J. BERNE. Elasticity, structure, and relaxation of extended proteins under force. *Proceedings of the National Academy of Sciences*, **110**, 3847-3852 (2013).
DOI : [10.1073/pnas.1300596110](https://doi.org/10.1073/pnas.1300596110)
- [20] B. JAGANNATHAN et S. MARQUEE. Protein folding and unfolding under force. *Biopolymers*, **99**, 860-869 (2013).
DOI : [10.1002/bip.22321](https://doi.org/10.1002/bip.22321)
- [21] Y. JAVADI, J. M. FERNANDEZ et R. PEREZ-JIMENEZ. Protein Folding Under Mechanical Forces : A Physiological View. *Physiology*, **28**, 9-17 (2013).
DOI : [10.1152/physiol.00017.2012](https://doi.org/10.1152/physiol.00017.2012)
- [22] G. LEE, L. CHRISSEY et R. COLTON. Direct measurement of the forces between complementary strands of DNA. *Science*, **266**, 771-773 (1994).
DOI : [10.1126/science.7973628](https://doi.org/10.1126/science.7973628)
- [23] A. P. WITTA, R. PEREZ-JIMENEZ, K. A. WALTHER, F. GRÄTER, B. J. BERNE, A. HOLMGREN, J. M. SANCHEZ-RUIZ et J. M. FERNANDEZ. Probing the chemistry of thioredoxin catalysis with force. *Nature*, **450**, 124-127 (2007).
DOI : [10.1038/nature06231](https://doi.org/10.1038/nature06231)
- [24] P. HERMAN-BAUSIER, C. LABATE, A. M. TOWELL, S. DERCLAYE, J. A. GEOGHEGAN et Y. F. DUFRÊNE. *Staphylococcus aureus* clumping factor A is a force-sensitive molecular switch that activates bacterial adhesion. *Proceedings of the National Academy of Sciences*, **115**, 5564-5569 (2018).
DOI : [10.1073/pnas.1718104115](https://doi.org/10.1073/pnas.1718104115)
- [25] A. PIERRES, A.-M. BENOLIEL et P. BONGRAND. Studying Molecular Interactions at the Single Bond Level with a Laminar Flow Chamber. *Cellular and Molecular Bioengineering*, **1**, 247-262 (2008).
DOI : [10.1007/s12195-008-0031-9](https://doi.org/10.1007/s12195-008-0031-9)

- [26] G. KAPLANSKI, C. FARNARIER, O. TISSOT, A. PIERRES, A. M. BENOLIEL, M. C. ALESSI, S. KAPLANSKI et P. BONGRAND. Granulocyte-endothelium initial adhesion. Analysis of transient binding events mediated by E-selectin in a laminar shear flow. *Biophysical Journal*, **64**, 1922-1933 (1993).
DOI : [10.1016/S0006-3495\(93\)81563-7](https://doi.org/10.1016/S0006-3495(93)81563-7)
- [27] P. BALL. Beyond the bond. *Nature*, **469**, 26-28 (2011).
DOI : [10.1038/469026a](https://doi.org/10.1038/469026a)
- [28] R. F. HENDRY. Two Conceptions of the Chemical Bond. *Philosophy of Science*, **75**, 909-920 (2008).
DOI : [10.1086/594534](https://doi.org/10.1086/594534)
- [29] P. HOBZA et K. MÜLLER-DETHLEFS. *Non-covalent interactions*. Avec la coll. de R. S. of CHEMISTRY (GREAT BRITAIN). RSC theoretical and computational chemistry series. OCLC : ocn437083462. Cambridge : Royal Society of Chemistry, 2009. 225 p.
- [30] S. S. M. KONDA, J. N. BRANTLEY, B. T. VARGHESE, K. M. WIGGINS, C. W. BIELAWSKI et D. E. MAKAROV. Molecular Catch Bonds and the Anti-Hammond Effect in Polymer Mechanochemistry. *Journal of the American Chemical Society*, **135**, 12722-12729 (2013).
DOI : [10.1021/ja4051108](https://doi.org/10.1021/ja4051108)
- [31] P. DOPIERALSKI, J. RIBAS-ARINO, P. ANJUKANDI, M. KRUPICKA, J. KISS et D. MARX. The Janus-faced role of external forces in mechanochemical disulfide bond cleavage. *Nature Chemistry*, **5**, 685-691 (2013).
DOI : [10.1038/nchem.1676](https://doi.org/10.1038/nchem.1676)
- [32] R. GROOTE, B. M. SZYJA, F. A. LEIBFARTH, C. J. HAWKER, N. L. DOLTSINIS et R. P. SIJBESMA. Strain-Induced Strengthening of the Weakest Link : The Importance of Intermediate Geometry for the Outcome of Mechanochemical Reactions. *Macromolecules*, **47**, 1187-1192 (2014).
DOI : [10.1021/ma4022339](https://doi.org/10.1021/ma4022339)
- [33] D. L. HUANG, N. A. BAX, C. D. BUCKLEY, W. I. WEIS et A. R. DUNN. Vinculin forms a directionally asymmetric catch bond with F-actin. *Science*, **357**, 703-706 (2017).
DOI : [10.1126/science.aan2556](https://doi.org/10.1126/science.aan2556)
- [34] M. DEMBO, D. C. TORNEY, K. SAXMAN et D. HAMMER. The reaction-limited kinetics of membrane-to-surface adhesion and detachment. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **234**, 55-83 (1988).
DOI : [10.1098/rspb.1988.0038](https://doi.org/10.1098/rspb.1988.0038)
- [35] C. GONZALEZ GUTIERREZ. Forces mécaniques dans la liaison des anticorps à domaine unique développés pour la thérapeutique : réponses moléculaire et cellulaire. Thèse de doctorat. Université d'Aix-Marseille, 2018.
- [36] A. PIERRES, A.-M. BENOLIEL et P. BONGRAND. Liaisons « accrocheuses » et mécanique moléculaire. *Médecine/sciences*, **19**, 1175-1179 (2003).
DOI : [10.1051/medsci/200319121175](https://doi.org/10.1051/medsci/200319121175)
- [37] W. E. THOMAS, V. VOGEL et E. SOKURENKO. Biophysics of Catch Bonds. *Annual Review of Biophysics*, **37**, 399-416 (2008).
DOI : [10.1146/annurev.biophys.37.032807.125804](https://doi.org/10.1146/annurev.biophys.37.032807.125804)

- [38] K. C. DANSUK et S. KETEN. A Simple Mechanical Model for Synthetic Catch Bonds. *Matter*, **1**, 911-925 (2019).
DOI : [10.1016/j.matt.2019.06.005](https://doi.org/10.1016/j.matt.2019.06.005)
- [39] E. V. SOKURENKO, V. VOGEL et W. E. THOMAS. Catch-Bond Mechanism of Force-Enhanced Adhesion : Counterintuitive, Elusive, but ... Widespread? *Cell Host & Microbe*, **4**, 314-323 (2008).
DOI : [10.1016/j.chom.2008.09.005](https://doi.org/10.1016/j.chom.2008.09.005)
- [40] B. T. MARSHALL, M. LONG, J. W. PIPER, T. YAGO, R. P. McEVER et C. ZHU. Direct observation of catch bonds involving cell-adhesion molecules. *Nature*, **423**, 190-193 (2003).
DOI : [10.1038/nature01605](https://doi.org/10.1038/nature01605)
- [41] P. VITRY, C. VALOTTEAU, C. FEUILLIE, S. BERNARD, D. ALSTEENS, J. A. GEOGHEGAN et Y. F. DUFRÊNE. Force-Induced Strengthening of the Interaction between *Staphylococcus aureus* Clumping Factor B and Loricrin. *mBio*, **8**, (2017).
DOI : [10.1128/mBio.01748-17](https://doi.org/10.1128/mBio.01748-17)
- [42] B. LIU, E. M. KOLAWOLE et B. D. EVAVOLD. Mechanobiology of T Cell Activation : To Catch a Bond. *Annual Review of Cell and Developmental Biology*, **37**, 65-87 (2021).
DOI : [10.1146/annurev-cellbio-120219-055100](https://doi.org/10.1146/annurev-cellbio-120219-055100)
- [43] X.-P. XU, S. POKUTTA, M. TORRES, M. F. SWIFT, D. HANEIN, N. VOLKMANN et W. I. WEIS. Structural basis of α E-catenin-F-actin catch bond behavior. *eLife*, **9**, e60878 (2020).
DOI : [10.7554/eLife.60878](https://doi.org/10.7554/eLife.60878)
- [44] Y. B. LIM, J. THINGNA, F. KONG, M. DAO, J. CAO et C. T. LIM. Temperature-Induced Catch-Slip to Slip Bond Transit in *Plasmodium falciparum*-Infected Erythrocytes. *Biophysical Journal*, **118**, 105-116 (2020).
DOI : [10.1016/j.bpj.2019.11.016](https://doi.org/10.1016/j.bpj.2019.11.016)
- [45] T. YAGO et al. Platelet glycoprotein Ib α forms catch bonds with human WT vWF but not with type 2B von Willebrand disease vWF. *The Journal of Clinical Investigation*, **118**, 3195-3207 (2008).
DOI : [10.1172/JCI35754](https://doi.org/10.1172/JCI35754)
- [46] S. van LEUKEN. DNA catch bonds : theory, simulation, and experiments on a simple, humanmade, and tunable catch bond. Student thesis : Master. Eindhoven University of Technology, 2019.
- [47] W. LI, W.-J. WONG, C.-J. LIM, H.-P. JU, M. LI, J. YAN et P.-Y. WANG. Catch-bond behavior of DNA condensate under tension. *Chinese Physics B*, **24**, 128704 (2015).
DOI : [10.1088/1674-1056/24/12/128704](https://doi.org/10.1088/1674-1056/24/12/128704)
- [48] E. EVANS. Probing the Relation Between Force—Lifetime—and Chemistry in Single Molecular Bonds. *Annual Review of Biophysics and Biomolecular Structure*, **30**, 105-128 (2001).
DOI : [10.1146/annurev.biophys.30.1.105](https://doi.org/10.1146/annurev.biophys.30.1.105)
- [49] G. I. BELL. Models for the specific adhesion of cells to cells. *Science*, **200**, 618-627 (1978).
DOI : [10.1126/science.347575](https://doi.org/10.1126/science.347575)

- [50] E. EVANS, D. BERK et A. LEUNG. Detachment of agglutinin-bonded red blood cells. I. Forces to rupture molecular-point attachments. *Biophysical Journal*, **59**, 838-848 (1991).
DOI : [10.1016/S0006-3495\(91\)82296-2](https://doi.org/10.1016/S0006-3495(91)82296-2)
- [51] O. K. DUDKO, J. MATHÉ, A. SZABO, A. MELLER et G. HUMMER. Extracting Kinetics from Single-Molecule Force Spectroscopy : Nanopore Unzipping of DNA Hairpins. *Biophysical Journal*, **92**, 4188-4195 (2007).
DOI : [10.1529/biophysj.106.102855](https://doi.org/10.1529/biophysj.106.102855)
- [52] O. K. DUDKO, G. HUMMER et A. SZABO. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences*, **105**, 15755-15760 (2008).
DOI : [10.1073/pnas.0806085105](https://doi.org/10.1073/pnas.0806085105)
- [53] J. LOU, T. YAGO, A. G. KLOPOCKI, P. MEHTA, W. CHEN, V. I. ZARNITSYNA, N. V. BOVIN, C. ZHU et R. P. McEVER. Flow-enhanced adhesion regulated by a selectin interdomain hinge. *Journal of Cell Biology*, **174**, 1107-1117 (2006).
DOI : [10.1083/jcb.200606056](https://doi.org/10.1083/jcb.200606056)
- [54] B. GUO et W. H. GUILFORD. Mechanics of actomyosin bonds in different nucleotide states are tuned to muscle contraction. *Proceedings of the National Academy of Sciences*, **103**, 9844-9849 (2006).
DOI : [10.1073/pnas.0601255103](https://doi.org/10.1073/pnas.0601255103)
- [55] J. JUMPER et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589 (2021).
DOI : [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2)
- [56] K. TUNYASUVUNAKOOL et al. Highly accurate protein structure prediction for the human proteome. *Nature*, **596**, 590-596 (2021).
DOI : [10.1038/s41586-021-03828-1](https://doi.org/10.1038/s41586-021-03828-1)

L'ADHÉSINE BACTÉRIENNE FIMH : ÉTAT DE L'ART

L'adhésine fimbriale FimH de *Escherichia coli* est l'une des adhésines bactériennes les plus étudiées et les mieux connues, de par la facilité de culture de la bactérie dont elle est issue et l'intérêt qu'elle suscite comme cible thérapeutique pour traiter de nombreuses pathologies. Elle a également un rôle historique dans la démonstration de l'existence des liaisons accrocheuses et la compréhension de leurs mécanismes. Sans prétendre à une revue exhaustive de la littérature, ce chapitre a pour ambition de retracer les grandes lignes des travaux ayant abouti aux modèles actuels du fonctionnement de FimH en tant que protéine grappin.

2.1 PRÉSENTATION GÉNÉRALE

2.1.1 Contexte biologique

La plupart des processus d'adhésion en biologie doivent faire face aux perturbations mécaniques, induites notamment par les forces hydrodynamiques de cisaillement au sein de l'environnement. De nombreuses bactéries à Gram négatif forment des interactions adhésives avec les surfaces qu'elles colonisent par le biais de protéines d'adhésion spécifique, les adhésines, le plus souvent situées à l'extrémité de filaments protéiques multimériques ancrés dans la membrane bactérienne externe, les fimbriae ou pili (du latin poil, cheveu). Ces adhésines se fixent à des récepteurs moléculaires particuliers des surfaces colonisées, comme des glycoprotéines membranaires ou des protéines de la matrice extracellulaire. Impliquées dans la formation de biofilms et dans les mécanismes d'invasion, elles constituent des facteurs de virulence majeurs et font l'objet de nombreuses recherches en tant que cibles thérapeutiques.

La protéine FimH de *E. coli* constitue un modèle bien connu parmi ces adhésines fimbriales. Cette protéine est exprimée à l'extrémité des pili dits de type 1 et leur confère leur capacité distinctive à se lier à des dérivés de mannose (FIGURE 8A–E) ainsi qu'à une grande variété de glycoprotéines fortement mannosylées, dont les uroplakines Ia de la membrane des cellules urothéliales (FIGURE 8F) [2], mais aussi d'autres récepteurs spécifiques des cellules épithéliales

Nous utiliseront pilus et fimbria de façon interchangeable, bien que certain-es auteur-es préfèrent réserver l'usage de « pili » aux longs pili de conjugaison ou pili sexuels, et celui de « fimbriae » aux plus courts pili communs régissant l'attachement aux surfaces [1].

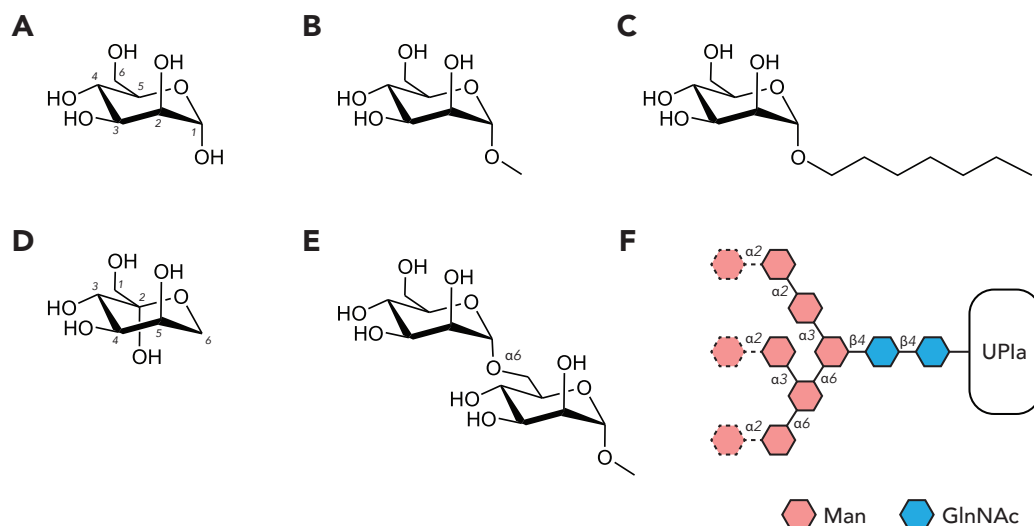


FIGURE 8 – Ligands de FimH. (A) D-mannose. Alkylmannosides : (B) α-méthyl-D-mannose et (C) α-heptyl-D-mannose. (D) fructopyranose représenté de façon à montrer sa ressemblance structurale au mannose. (E) Exemple de dimannoside, α-D-mannose(1 → 6)-α-D-méthylmannose. (F) Structure des N-glycanes hautement mannosylés exposés à la surface de l'uropaline Ia (UPIa), les osides facultatifs sont en pointillés. Man : mannose. GlnNAc : N-acétylglucosamine.

des tubes proximaux rénaux, de la bouche, de l'intestin ou encore des poumons [3].

Les fimbriae de type 1 sont exprimées par 95% des souches connues d'*Escherichia coli*. Leur implication dans diverses pathologies a été abondamment étudiée. Elles jouent notamment un rôle dans les méningites néonatales ou la maladie de Crohn chez l'humain, ainsi que des mastites (inflammation des mamelles) en élevage bovin [5]. Toutefois, leur rôle le mieux élucidé concerne les infections du tractus urinaire.

2.1.2 Implication dans les infections urinaires

Les infections des voies urinaires (IVU) sont des infections bactériennes fréquentes, qui touchent disproportionnellement les femmes. On estime que 40% des femmes en seront victimes au moins une fois au cours de leur vie [6, 7]. Parmi les infections du tractus urinaire inférieur (cystites), entre 20% et 30% des patientes font l'expérience de récurrences dans les 3 à 4 mois après un premier épisode. En l'absence de traitement approprié, les infections peuvent évoluer vers des formes plus graves touchant les voies supérieures (reins), les pyélonéphrites, puis des septicémies.

De 80 à 90% des IVU sont causées par des souches d'*E. coli* spécifiques, qualifiées d'uropathogènes (UPEC). La plupart des UPEC pré-

On retrouve des pili de type 1 homologues chez d'autres entérobactéries telles que *Salmonella*. Bien que l'adhésine ne partage que 15% d'identité de séquence avec son homologue chez *E. coli*, leurs structures et propriétés sont similaires [4].

sentent plusieurs mutations au niveau du locus *fimH* par rapport aux souches commensales de l'intestin, généralement inoffensives. Les fimbriae de type 1 ont un rôle établi dans la survenue d'IVU inférieures. Leur expression permet aux UPEC de s'attacher aux épithéliums de la vessie et de l'urètre et favorise la colonisation bactérienne. L'ancrage à l'urothélium est une étape préliminaire à l'invasion par les bactéries de l'intérieur de la paroi épithéliale, où elles forment des communautés bactériennes intracellulaires (IBC). Ces sortes de biofilms protègent les bactéries du système immunitaire et des antibiotiques et constituent des réservoirs quiescents pouvant conduire à des infections récurrentes.

Les IVU sont généralement bien traitées par l'administration d'antibiotiques de première intention. Cependant, l'émergence de souches bactériennes résistantes ou multirésistantes constitue une préoccupation grandissante. Le rôle clé des pili de type 1 dans le processus d'infection en font une cible thérapeutique très intéressante pour bloquer le processus de colonisation et d'invasion, en complément ou remplacement des antibiothérapies.

Toutefois, la capacité de l'adhésine FimH à former des liaisons accrocheuses pose un sérieux obstacle à la réussite de telles thérapies anti-adhésives. En effet, le mécanisme grappin ne peut être activé qu'avec des ligands glycoprotéiques ancrés à la surface épithéliale, car ceux-ci offrent la résistance nécessaire à l'établissement d'une tension mécanique. Par contraste, des inhibiteurs compétitifs libres en solution sont inaptes à former une liaison accrocheuse. La compréhension fine du mécanisme sous-jacent est donc une étape cruciale au développement de thérapies anti-adhésives efficaces.

2.1.3 Présentation du pilus de type 1

Les fimbriae de type 1 se présentent comme des appendices fibrillaires d'environ 1–2 µm de long et 7 nm de large visibles en microscopie électronique (FIGURE 9A) [5, 8]. Elles sont composées d'une tige (*fimbrial rod*) relativement rigide composée d'un assemblage hélicoïdal de sous-unités FimA (jusqu'à ≈ 3000), ainsi que d'une extrémité distale formée par les pilines adaptatrices FimF-FimG et l'adhésine FimH (FIGURE 9BC). La tige hélicoïdale est extensible et se déroule lorsqu'elle est soumise à une certaine force. Une bactérie piliée exprime de l'ordre de quelques centaines de pili à sa surface.

Les pili de type 1 sont synthétisés par une voie dite « chaperon-placeur » (*chaperone-ushe*r). Les sous-unités piliales sont complexées dans le périplasme par une protéine chaperon (FimC) puis leur excrétion et leur assemblage sont catalysés au niveau de la membrane externe par un canal protéique (FimD) dit « placeur » (*usher*) [1]. Un cluster de neuf gènes *fim* est nécessaire à l'expression des pili de

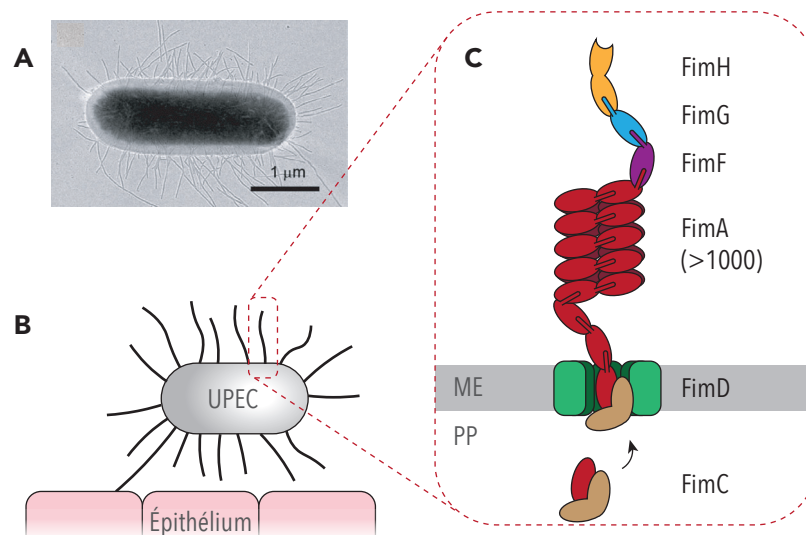


FIGURE 9 – Structure générale d'un pilus de type 1. (A) Micrographe électronique d'une bactérie *Escherichia coli* exprimant des pili de type 1 [10]. (B) et (C) Schéma de l'organisation générale d'un pilus de type 1. ME : membrane externe. PP : périplasma.

type 1. Ceux-ci codent pour l'adhésine spécifique aux mannosides, FimH, ainsi que les pilines FimG, FimF et FimA constituant les autres sous-unités du pilus. FimB, FimE sont des protéines de régulation de l'expression piliale. Enfin, le rôle de FimI demeure incertain mais elle pourrait être un facteur de régulation ou une unité d'ancrage dans la membrane externe [9].

Le contexte général ayant été abordé, nous proposons dans la section suivante une approche historique de l'étude de FimH et son implication dans la découverte et la caractérisation des liaisons accrocheuses.

2.2 VERS LA DÉCOUVERTE D'UNE PROTÉINE GRAPPIN

Dès les années 1970, l'intérêt pour les fimbriae (ou pili) exprimées par certaines souches d'*Escherichia coli* conduit les microbiologistes à étudier leur rôle dans de nombreux processus, dont l'adhérence bactérienne [11]. Les fimbriae dites de type 1 peuvent provoquer une hémagglutination (agrégation des globules rouges) inhibée par des dérivés du mannose, et sont connues comme un facteur de virulence majeur des souches uropathogènes (UPEC). Des recherches approfondies s'intéressent alors à la spécificité des pili de type 1 pour divers mono- et oligosaccharides, notamment dans l'objectif de trouver des inhibiteurs compétitifs à l'adhérence bactérienne aux épithéliums humains.

2.2.1 Mise en évidence

Les premières expériences suggérant la capacité de FimH à former des liaisons accrocheuses se sont intéressées à l'accroissement des propriétés d'adhérence bactérienne en présence d'un stress hydrodynamique. Les forces de cisaillement causées par l'écoulement de fluides au voisinage des surfaces d'ancrage de diverses bactéries, éventuellement pathogènes, notamment dans le système vasculaire, le tube digestif ou les voies urinaires, doivent être supportées par celles-ci grâce à de solides interactions avec leur substrat. Toutefois, un ancrage permanent est un désavantage pour la bactérie, car il limite ses capacités migratoires. Une stratégie évolutive possible est donc de développer des systèmes permettant de moduler la force d'adhérence en fonction du stress mécanique environnant.

Plusieurs études avaient observé l'effet promoteur des forces hydrodynamiques cisailantes sur l'adhérence dans plusieurs contextes biologiques [12]. Par exemple, l'adhérence des plaquettes à des surfaces recouvertes par le facteur de von Willebrand, une protéine plasmatique impliquée dans la coagulation sanguine, nécessite un certain niveau de cisaillement fluide [13]. Les leukocytes ne se lient pas aux surfaces couvertes de ligand glycoprotéique-1 de la P-sélectine (PSGL-1) en conditions statiques, mais s'y attachent et y roulent en présence d'un écoulement [14]. De même, un accroissement du stress hydrodynamique cause une augmentation du nombre de cellules de *Staphylococcus aureus* adsorbées à une surface recouverte de collagène [15]. Quelques mécanismes ont été proposés pour expliquer ces phénomènes dans le paradigme des liaisons glissantes. Par exemple, dans le cas des leukocytes, il a été proposé que le cisaillement accélère la vitesse de formation des liaisons moléculaires plus qu'il n'accélère leur rupture, augmentant ainsi leur durée de vie. L'aplatissement cellulaire engendré par les forces hydrodynamiques pouvait également augmenter la surface de contact et donc la force d'adhérence. Bien que plausibles, ces explications écartent la possibilité pour l'interaction moléculaire elle-même d'être influencée favorablement par une force mécanique.

L'un des premiers travaux à démontrer la capacité propre d'une adhésine, en l'occurrence FimH, à accroître son affinité en présence d'une perturbation mécanique est mené par THOMAS et al. [16]. Les auteur·es y analysent l'effet des cisaillements fluides sur l'agrégation d'érythrocytes de cobaye, qui présentent à leur surface des glycoprotéines mannosylées, en présence de bactéries *E. coli* exprimant FimH. L'agglutination réversible des érythrocytes dans des conditions dynamiques, mais pas en conditions statiques ou en présence d'inhibiteur (α -méthylmannose), constitue un premier indice de l'aptitude de FimH à former des liaisons accrocheuses. Le deuxième type d'ex-

périence mené par THOMAS et al. utilise des chambres à écoulement laminaire : des bactéries exprimant FimH sont immobilisées sur la surface de la chambre et les érythrocytes de cobaye déposés sur le tapis bactérien. Un éluant est alors injecté dans la chambre à plusieurs débits, générant plusieurs contraintes de cisaillement à l'interface. Le mouvement des érythrocytes, suivi par vidéo microscopique, met en évidence plusieurs régimes : aux faibles cisaillements, les érythrocytes attachés à la surface « roulent » sur le tapis bactérien, et leur vitesse augmente avec le débit ; avec des débits croissants en revanche, la vitesse des érythrocytes attachés diminue soudainement, voire s'annule ; pour finir, au delà d'un certain débit, les érythrocytes se décrochent toujours. L'ajout d'inhibiteur dans le milieu diminue la probabilité de nouvelles adhésions mais n'accélère pas le décrochage des érythrocytes déjà immobilisés. Cette observation conduit donc à l'hypothèse que des liaisons uniques, activées sous force, sont à l'origine de l'immobilisation des cellules. Dans une étude consécutive les auteur-es proposent donc un modèle cinétique fondé sur l'hypothèse que FimH forme des liaisons accrocheuses grâce à un mécanisme allostérique à deux états [17], que nous détaillons plus loin (2.4.1).

2.2.2 Expériences sur molécule unique

Bien que les expériences précédentes aient fourni de forts indices en faveur de l'existence des liaisons accrocheuses, une preuve définitive viendrait de l'observation du phénomène à l'échelle de la molécule unique [18]. C'est sur le grappin P-sélectine/PSGL-1 que la première étude exploitant l'AFM pour étirer directement le complexe avec l'application d'une force constante démontre sans ambiguïté l'augmentation de la durée de vie de la liaison avec des forces croissantes [19].

La liaison FimH-mannose isolée est étudiée plus directement par YAKOVENKO et al. [20]. Des expériences AFM à vitesse constante démontrent clairement une distribution bimodale de la force de rupture de la liaison, avec un pic inférieur à 40 pN et un autre avoisinant 140–180 pN, suggérant deux types d'états liés. Les forces de ruptures faibles sont favorisées pour des faibles vitesses de charge (10^2 pN s^{-1}) tandis que des vitesses plus élevées ($10^3\text{--}10^4 \text{ pN s}^{-1}$) augmentent fortement la fréquence des hautes forces de rupture. Ce comportement biphasique n'est pas observé avec le mutant A188D (constitutivement haute affinité), ou en présence d'un anticorps stabilisant FimH dans un état haute affinité, ou encore lorsque la liaison est pré-exposée à une force de tension modérée.

Les travaux sur molécule unique viennent donc confirmer que les propriétés d'adhérence bactérienne en présence d'une contrainte mécanique s'expliquent *a minima* par la capacité de l'adhésine FimH à

former des liaisons accrocheuses. Ils tendent à corroborer un modèle allostérique, dans lequel la protéine présente deux états liés ayant des vitesses de dissociation différentes. Cette allostérie est activée *in vivo* par l'application de la force de tension elle-même, et *in vitro* par des anticorps liant un épitope spécifique à l'une des formes, ou de façon constitutive chez certains mutants.

Les expériences sous force mettent également en lumière le rôle crucial de la tige piliale poly-FimA dans le processus d'adhérence [21-23]. La capacité de celle-ci à se dérouler de manière élastique à des forces modérées (~60 pN), proche de la gamme où la durée de vie de la liaison accrocheuse est la plus longue (60–100 pN), permet d'absorber les chocs mécaniques et de maximiser le temps d'accroche.

2.3 STRUCTURE PROTÉIQUE

Malgré l'impressionnante résolution des dispositifs de spectroscopie de force unimoléculaires, il nous faut détailler la structure de FimH à l'échelle moléculaire afin de comprendre pleinement l'origine de ses propriétés accrocheuses. Nous présentons ici la morphologie globale de la protéine dans sa forme native, avant de détailler la découverte successive de plusieurs de ses états conformationnels.

2.3.1 Anatomie générale

FimH est une protéine de 30 kDa (279 acides aminés) constituée de deux domaines. Le domaine lectine (acides aminés 1 à 159), qui comprend le site de liaison spécifique aux mannosides, est relié au domaine piline (160 à 279) par un segment de liaison (*linker*) flexible (154-160).

Les lectines désignent des protéines se liant de façon spécifique et réversible à des glucides.

À l'instar des autres pilines formant la fimbria, les deux domaines ont une structure d'immunoglobuline formée de 7 brins bêta se repliant en « sandwich » autour d'un cœur hydrophobe. Seul le domaine lectine possède la totalité des 7 brins. Le domaine piline est incomplet et son repliement fonctionnel nécessite la complexation d'un brin exogène apporté par l'unité suivante du pilus (FimG). Ce processus par lequel l'extrémité N-terminale d'une autre piline vient s'insérer de façon non-covalente dans le sandwich- β est nommé « complémentation de brin donneur » (*donor strand complementation*).

Mentionnons ici que le même principe d'assemblage relie FimG à l'unité protéique suivante, FimF, et cette dernière à une unité FimA. Le reste du pilus est constitué par un enchaînement poly-FimA, là-aussi reposant sur la complémentation de brin donneur. Ce type d'architecture assure l'extrême stabilité cinétique du pilus dans le milieu extracellulaire : le complexe FimF-FimG par exemple a une demi-

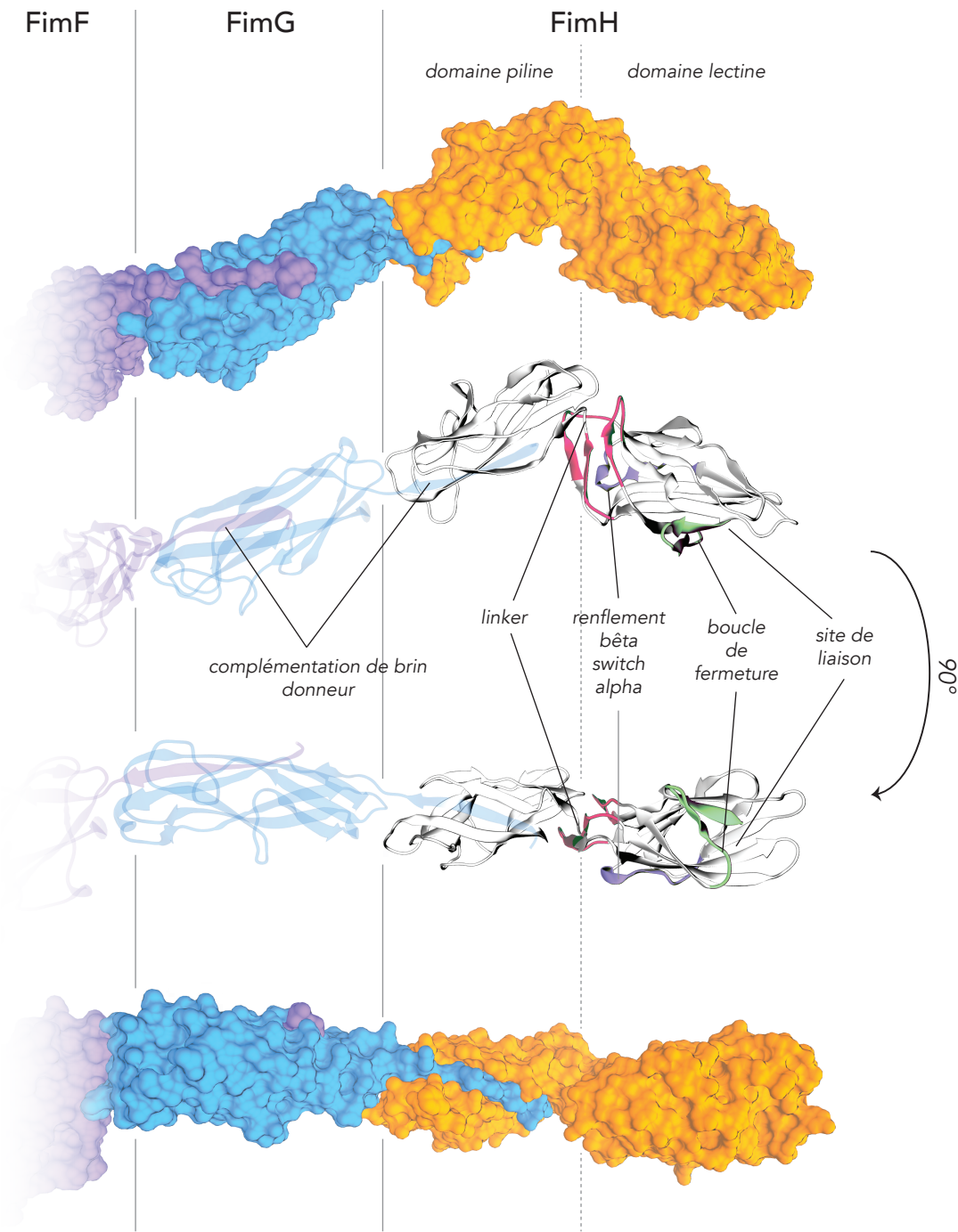


FIGURE 10 – Structure de l'extrémité piliale. La structure cristallographique (PDB: 3JWN) est montrée sous deux angles. En représentation surfacique on peut apprécier l'enchaînement par complémentation de brin donneur des protéines FimF (violet), FimG (bleu) et FimH (orange). La représentation *cartoon* permet de montrer la structure en β -sandwich de chaque domaine. Des régions clés du domaine lectine y ont été colorées : la boucle de fermeture (vert) du site de liaison, le renflement bêta/switch alpha (violet) et la région inter-domaines (rouge) comprenant le *linker*.

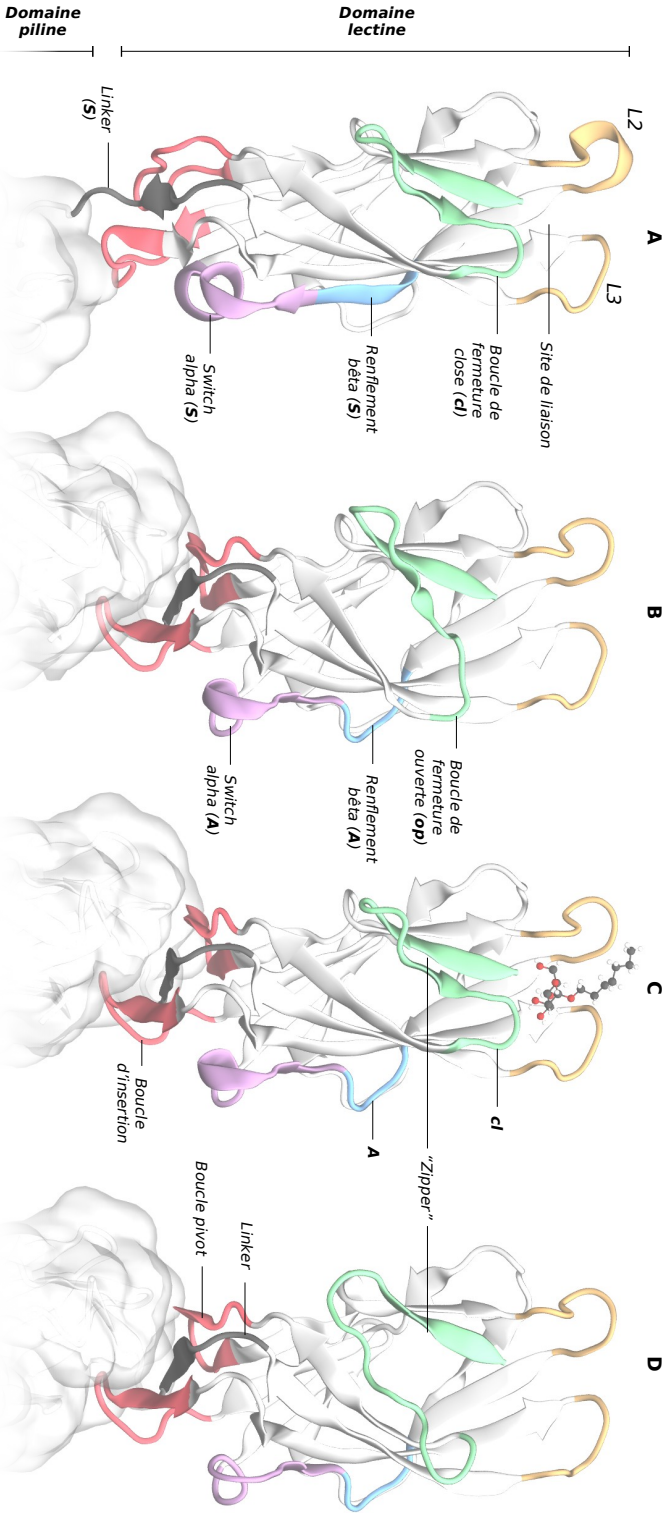
vie de dissociation estimée à plusieurs milliards d'années dans des conditions standard [24], et résiste à des forces de tension très élevées (>200 pN) [22] ce qui en fait l'un des assemblages protéiques non-covalents parmi les plus stables connus à ce jour. La stabilité de FimH est également assurée par deux ponts disulfures, l'un au sein du domaine lectine et voisin du site de liaison, l'autre au sein du domaine piline juste à la base du *linker*. La FIGURE 10 présente la structure de la protéine intégrée à l'extrémité piliale ainsi que certains éléments structuraux notables, qui seront détaillés plus loin.

2.3.2 Variations conformationnelles du domaine lectine

La structure exacte de FimH est demeurée longtemps inaccessible en raison de l'impossibilité de la cristalliser sous sa forme pure [11]. En effet, le mécanisme de complémentation de brin donneur est nécessaire pour assurer le repliement stable du domaine piline. L'idée d'utiliser la protéine chaperonne FimC, qui stabilise FimH *in vivo* avant son assemblage au sein du pilus en jouant un rôle de complémentation temporaire dans le périplasme bactérien, conduit à la première structure cristallographique de FimH par CHOUDHURY et al., (1999) [25]. L'analyse structurale de cette première structure permettra déjà d'entrevoir un bon nombre de caractéristiques fonctionnelles de FimH : c'est notamment sur cette base que THOMAS et al. [16] font l'hypothèse que l'extension du segment de liaison par une force de tension provoque un changement allostérique au niveau du site de liaison, pouvant expliquer le mécanisme de liaison accrocheuse. Pourtant, des preuves cristallographiques d'une telle allostérie n'arriveront que des années plus tard. En effet, la cristallisation du domaine lectine isolé et sa co-cristallisation avec divers mannoses (par exemple dans [26], PDB: 1UWF) ne montrera pas de changements conformationnels majeurs. Il faudra attendre les travaux de TRONG et al. et la cristallisation de l'intégralité de l'apex du pilus (FimH·FimG·FimF, avec la chaperonne FimC complexée à FimF) pour observer FimH avec ses deux domaines en interaction proche [27].

Jusque-là, les structures concernaient soit des domaines lectines isolés et purifiés, soit la protéine entière complexée à FimC, qui s'interpose partiellement entre les deux domaines et empêche leur association (voir structure FIGURE 11A). Dans la nouvelle structure, les deux domaines sont intimement associés et le domaine lectine présente une toute autre conformation. Outre des changements structuraux dans la région inter-domaines, la nouvelle forme présente (entre autres) un site de liaison beaucoup plus ouvert et un renflement bêta (*beta-bulge*) au niveau de la région 59–63 (voir FIGURE 11B). Ces observations tendent à confirmer le scénario d'une régulation allostérique intramoléculaire négative, où l'application d'une force colinéaire à la liaison inter-domaines provoque ou favorise la rupture des interac-

FIGURE 11 – Répertoire conformationnel du domaine lectine.



TRAVAUX PRÉCÉDENTS				
MAGALA et al. [28]	Groupe 1 et 2	Groupe 4	Groupe 3	Groupe 5
TRONG et al. [27]	<i>elongated</i>	<i>compressed</i>		
SAUER et al. [29]	Sbound, Sfree	Afree	Abound	
LIV et al. [30]	<i>active</i>	<i>inactive</i>	<i>inactive</i>	
NOMENCLATURE ADOPTÉE ICI				
État allostérique	S	A	A	A
Ouverture du site	cl	op	cl	op

TABLE 1 – Nomenclature conformationnelle du domaine lectine.

tions lectine–piline, l’extension du *linker* et le changement conformationnel du domaine lectine pour un état de plus haute affinité.

Le tableau sera complété par SAUER et al., qui parviennent à stabiliser un construit de FimH en solution grâce à un peptide jouant le rôle du brin donneur de FimG. Le complexe FimH·DsG (pour *donor strand G*) peut être simultanément cristallisé et ses propriétés biochimiques d’interaction avec des ligands étudiées aisément. Les structures de FimH·DsG complexé à une molécule d’heptylmannose présentent une nouvelle conformation du domaine lectine. Celle-ci est en quelque sorte intermédiaire entre les deux structures précédentes, comparable à la structure « associée » en ce qui concerne la région inter-domaines et le renflement bêta mais possédant un site de liaison fermé (FIGURE 11C). Cela conduira les auteur·es à proposer un mécanisme d’ajustement induit, dans lequel la liaison du ligand s’accompagne par une fermeture de la boucle bordant la cavité du site de liaison (*clamp loop*).

On recense une certaine diversité dans la nomenclature employée dans les publications pour qualifier les différentes conformations de FimH. Certaines caractéristiques émergent cependant, peut-être car elles paraissent les plus intuitives, les plus aisément reconnaissables, et les plus liées aux aspects fonctionnels de l’adhésine. On retrouve ainsi des adjectifs qualifiant des oppositions binaires entre deux sous-groupes conformationnels :

- le site de liaison : ouvert/fermé (*open/closed*) ; relâché/serré (*loose/tight*) ; large/étroit (*wide/narrow*).
- la forme globale du domaine lectine : allongé/compressé (*elongated/compressed*)
- l’interaction entre les deux domaines : associés/séparés (*associated/separated*), abrégé A/S.
- la présence d’un ligand : lié/libre (*bound/free*).
- l’affinité supposée de la conformation pour un ligand : haute/-basse affinité (*high/low affinity*) ; forme active/inactive ; forme relâchée/tendue (*relaxed/tense*), abrégé R/T, en référence à la nomenclature allostérique proposée historiquement par MONOD, WYMAN et CHANGEUX [31].

Il existe aujourd’hui pas moins d’une soixantaine d’entrées dans la PDB concernant l’adhésine FimH. Une étude récente de MAGALA et al. dresse une revue des différentes structures rencontrées et les classe en cinq groupes fondés sur la conformation du domaine lectine, discriminés par RMSD de paires [28] :

1. Le groupe 1 comprend la majorité des modèles (121). Il intègre la plupart des modèles où le domaine lectine est isolé ou distancié du domaine piline (FIGURE 11A).
2. Le groupe 2 est très proche du groupe 1 et correspond à l’ensemble issu d’une étude RMN [32] (FIGURE 11A).

3. Le groupe 3 ne contient que quatre membres (4XOB-A, 4XOB-C, 4XOB-E, 4XOE-A). Il correspond aux domaines lectines complexés avec un mannoside, et au sein d'une protéine FimH complète dans laquelle le domaine piline est stabilisé par un peptide qui joue le rôle de brin donneur (FIGURE 11C).
4. Le groupe 4 contient des modèles ayant la plus large déviation de RMSD avec les groupes 1 et 2. Toutes les structures correspondent à la conformation du domaine lectine dans la protéine entière en l'absence de ligand, où le domaine piline est soit complétement par un peptide brin donneur soit complexé avec FimG, FimF et FimC (FIGURE 11B)
5. Le groupe 5 ne comprend qu'un seul modèle (4J3O-H) est correspond au complexe FimH·FimG·FimF·FimC dans le canal du *usher* FimD. Il diffère du groupe 4 par une conformation ouverte alternative de la boucle de fermeture (FIGURE 11D).

Une bien moindre attention a été apportée aux variations conformationnelles internes du domaine piline, vraisemblablement limitées. Davantage d'intérêt a été porté à l'orientation relative des deux domaines [28, 33]. Dans leur conformation associée, leurs axes principaux forment un angle d'environ $\sim 150^\circ$. Dans les structures où les domaines sont séparés, une grande variété de positions relatives sont observées dans les structures cristallographiques. En raison de la flexibilité intrinsèque du *linker* ces configurations particulières sont à relier au processus d'empilement cristallin et non à des états biologiques remarquables. En effet, des simulations confirment la forte flexibilité dynamique de l'articulation lectine-piline lorsque le *linker* est étendu [33].

Dans nos travaux nous emploierons une nomenclature visant à bien distinguer la composition du système d'une part, et l'état conformationnel du domaine lectine d'autre part :

Nous verrons qu'il s'agit du motif clé limitant dans la transition allostérique.

- L'état allostérique, déterminé par la conformation du renflement bêta, reprendra les termes associé (noté A) et séparé (noté S). Notons que le domaine lectine peut donc être dans une conformation associée A même en l'absence du domaine piline.
- Lorsque nécessaire, nous différencierons les systèmes où les deux domaines sont présents (LP) et les systèmes où le domaine lectine est isolé (L).
- La conformation du site de liaison et sa boucle de fermeture sera qualifiée d'ouverte (op pour *open*) ou de fermée (cl pour *closed*). Nous rencontrerons également un état semi-ouvert (so).
- Au besoin, la présence ou l'absence du ligand sera signalée par un symbole plein (●) ou vide (○) respectivement.

Ces choix prendront tout leur sens quand nous explorerons le paysage conformationnel au chapitre 3. La TABLE 1 résume les dénominations principales utilisées dans la littérature et celles que nous allons

privilégier dans nos travaux pour les 5 groupes de conformations trouvées dans la PDB.

2.3.3 *Éléments structuraux notables*

Un certain nombre d'éléments structuraux de FimH ont retenu l'attention car ils sont soupçonnés d'avoir un rôle particulier dans ses propriétés de liaison ou d'allostérie. Ce sont en particulier les parties du domaine lectine qui interagissent avec ses partenaires (ligand ou domaine piline), celles qui diffèrent fortement d'une conformation à l'autre, ou dont les expériences (de mutagenèse dirigée notamment) ont démontré l'importance *in vitro*. Nous souhaitons attirer l'attention sur ces différents éléments afin de faciliter la discussion ultérieure de nos travaux.

Site de liaison et boucle de fermeture

Le site de liaison est caractéristique des lectines et autres récepteurs aux glucides [34]. La liaison est très spécifique aux α -D-mannopyranosides (et dans une moindre mesure aux fructopyranosides [26]) grâce à un dense réseau de liaisons hydrogène au fond du site de liaison, entre les chaînes latérales de la protéine et les groupements hydroxyles du sucre (FIGURE 12). Le cycle mannose lié interagit par liaison-H directe avec les chaînes latérales de Asp54, Gln133, Asn135, Asp140, avec l'extrémité terminale NH_3^+ (Phe1) et avec la chaîne principale de Asp47. Ce réseau permet de compenser le coût de désolvatation du ligand, très hydrophile, et du site lui-même lors de la formation du complexe. Le site de liaison est très conservé évolutivement et les études de mutagenèse indiquent que les acides aminés interagissant directement avec le mannose sont cruciaux pour assurer spécificité et haute affinité [5, 35].

Le pourtour de la cavité forme au contraire un environnement hydrophobe qui protège les interactions polaires protéine-ligand du solvant et forme une interaction favorable avec le cycle pyranose, en particulier dans les conformations fermées, notamment *via* les résidu Ile13 et Ile52.

Portail à tyrosines

Le portail à tyrosines (*tyrosine gate*) désigne la paire de résidus tyrosines Tyr48 et Tyr137 situés de part et d'autre de la partie supérieure du site de liaison, sur les boucles L2 et L3 respectivement (FIGURE 12).

Cette structure particulière explique l'affinité relativement forte de FimH pour des mannosides substitués avec des groupes aromatiques [26, 36-38]. Les substituants (aryles, biphényles) se positionnent entre

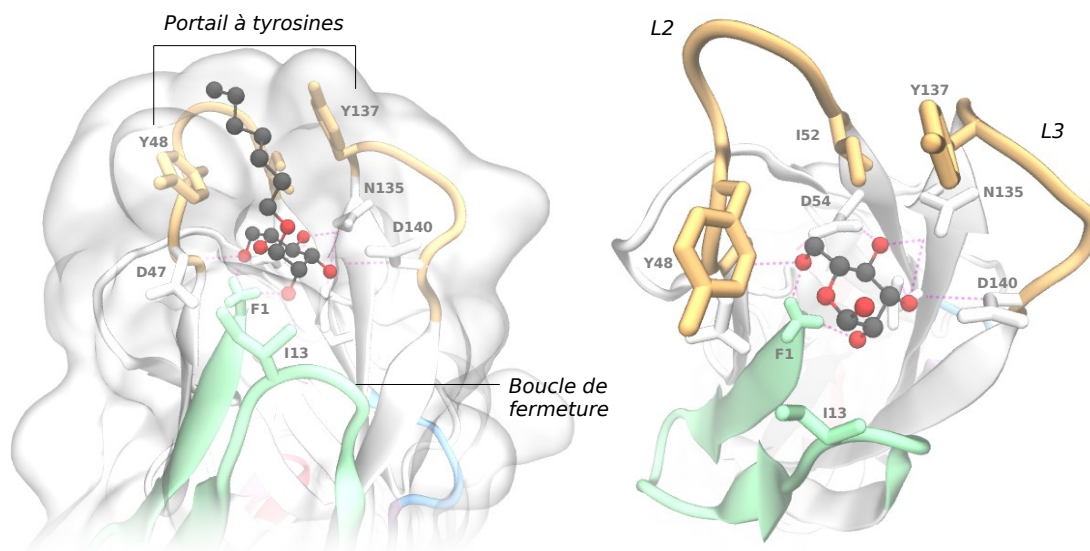


FIGURE 12 – Structure du site de liaison dans le co-cristal avec le ligand heptylmannose. (*gauche*) Vue de face. La surface protéique est représentée en transparence. Les chaînes latérales des acides aminés notables et leurs liaisons hydrogène avec le ligand sont illustrées. (*droite*) Vue du dessus. La chaîne heptyle du ligand a été retirée pour plus de visibilité.

les deux cycles phénoliques, avec lesquels ils forment des interactions hydrophobes et d'empilement π . Elle pourrait également expliquer l'affinité surprenante des alkyl-mannosides dont la queue hydrophobe hautement flexible peut interagir favorablement avec le portail à tyrosines tout en préservant une grande liberté conformationnelle et en limitant ainsi en partie le coût entropique de liaison [5, 26, 39].

Le portail à tyrosines lui-même présente une plus grande diversité conformationnelle dans les structures cristallines que les résidus de la cavité liant le mannose. En comparant les structures de FimH_L lié à plusieurs mannosides, WELLENS et al. identifient au moins trois rotamères pour Tyr48 formant des portails à tyrosine qualifiés d'ouvert, mi-ouvert et fermé (*open*, *half-open*, *closed*) [39]. Cette relative flexibilité permet à l'adhésine de s'accommoder d'un large spectre de glycanes riches en α -D-mannosides terminaux connectés selon plusieurs modalités de liaison glycosidique ($\alpha 1-2$, $\alpha 1-3$, ...) bien qu'avec des affinités variables. Il est aussi suggéré que l'interaction non-optimale avec la partie aglycone de certains mannoside entraîne une perte enthalpique qui pourrait être en partie compensée entropiquement par la dynamique conformationnelle du portail à tyrosines.

Renflement bêta et switch alpha

Dès la publication de la première structure de la forme A [27], TRONG et al. soulignent un changement structural majeur du segment formé par les résidus 59–71 du brin C par rapport à la forme S. Les

résidus 59–62 perdent leur structure de brin β pour former une petite boucle, ou renflement bêta (*beta-bulge*), et rompent la structure du feuillet β formé avec le brin adjacent (F). Les résidus 63–71 qui constituaient une courte hélice alpha deviennent l'enchaînement d'un court brin β (63–64) et d'une hélice 3_{10} (65–71).

Ce changement s'accompagne de l'exposition de chaînes latérales précédemment enfouies au sein du cœur hydrophobe comme celles de Val67, ou au contraire l'enfouissement d'autres telles que Ser62 et surtout Tyr64, toutes deux formant des liaisons hydrogènes avec le squelette protéique (FIGURE 13).

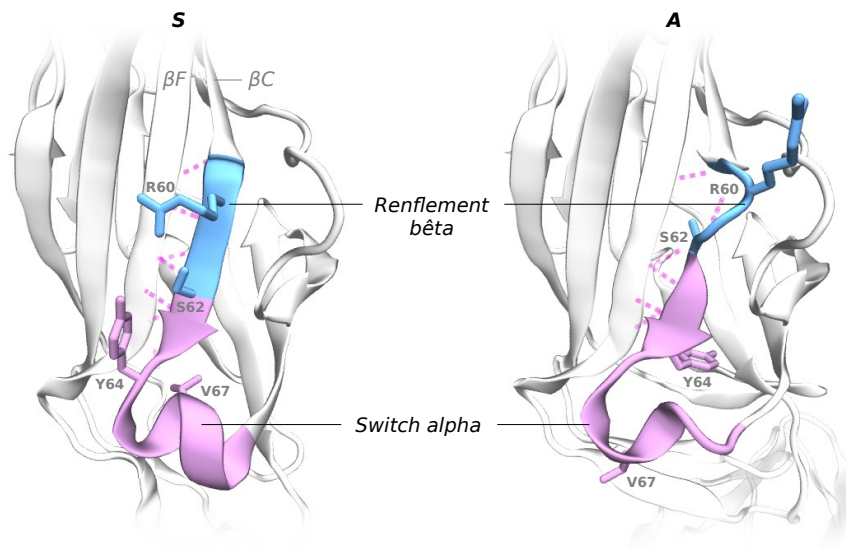


FIGURE 13 – Structure du renflement bêta et du switch alpha. (gauche) Dans l'état allostérique séparé (S). (droite) Dans l'état allostérique associé (A).

Ces modifications structurelles sont supposées centrales dans la propagation de l'information allostérique des boucles inter-domaines au site de liaison. Le « raccourcissement » effectif du brin β C par la formation d'un renflement accommode la torsion du large feuillet β et l'évasement de la poche de liaison, inspirant la métaphore d'un mécanisme en tourne-page (*page-turning*). Plusieurs mutations dans la région affectent les propriétés de liaison en stabilisant l'une ou l'autre des deux conformations majeures [35, 37]. Citons le mutant R60P, pour lequel la proline en position 60 favorise très fortement la formation du renflement bêta, bloquant le domaine lectine dans sa forme compressée (A) même en l'absence du domaine piline, comme indiqué par mesures d'affinité et validé par cristallographie aux rayons X [37].

Région inter-domaines

La région inter-domaines comprend les éléments du domaine lectine en contact direct avec le domaine piline dans la forme associée de la protéine. On distingue la boucle pivot (*swing loop*, résidus 25–34), la boucle d'insertion (*insertion loop*, 111–119) et le segment de liaison (*linker*, 151–158). Ces éléments subissent un réarrangement important lors du passage $S \rightarrow A$, et sont connus pour être particulièrement flexibles en l'absence d'interaction directe avec le domaine piline.

L'association avec le domaine piline est essentiellement médiée par des interactions hydrophobes, notamment des résidus Val30, Ala188, Val28, Val155, Ala115. Il en résulte une certaine fragilité de l'association des deux domaines, probablement nécessaire à leur séparation aisée et rapide sous force.

L'acide aminé 27, par exemple, est un emplacement à fort taux de mutation, sélectionné chez diverses souches d'*E. coli* pour moduler finement les propriétés d'adhérence piliale [35]. À l'instar de R60P, la double mutation V27C/L34C formant un pont disulfure au sein de la boucle pivot parvient à stabiliser le domaine lectine isolé dans l'état A, de basse affinité [37]. La figure FIGURE 14 détaille la structure de l'interface lectine–piline.

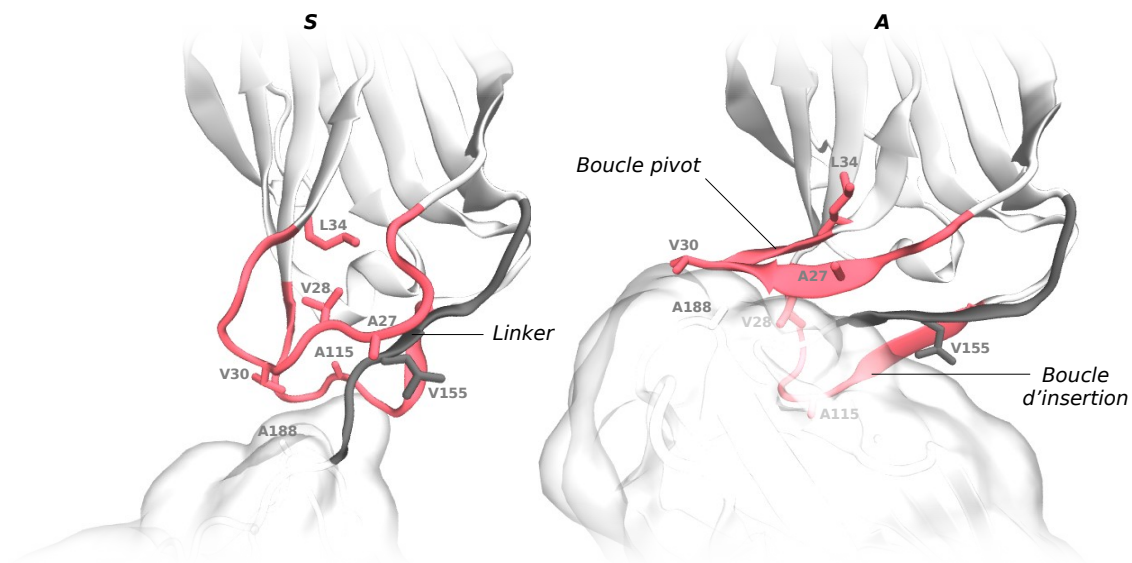


FIGURE 14 – Structure de la région inter-domaines. (gauche) Dans l'état allostérique séparé (S). (droite) Dans l'état allostérique associé (A).

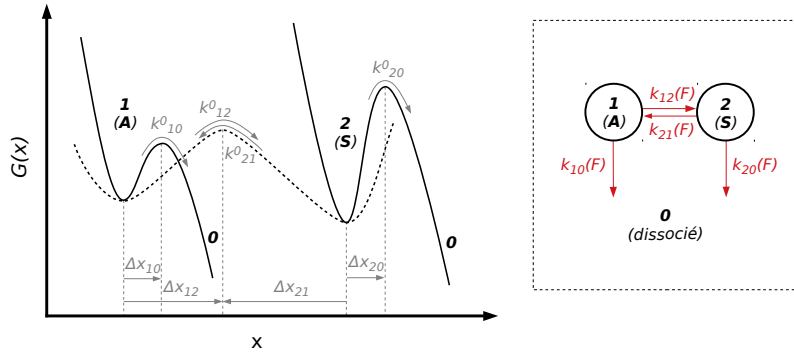


FIGURE 15 – Modèle cinétique à deux états (THOMAS et al.). Inspiré de la publication originale [41].

2.4 MODÈLES PROPOSÉS

Avant de passer à la présentation de nos travaux, nous souhaitons achever cet état de l'art en résumant les deux principaux modèles quantitatifs du mécanisme grappin de FimH publiés dans la littérature à ce jour.

2.4.1 Modèle cinétique à deux états

Dans une étude de (2006), THOMAS et al. analysent le temps de pause de bactéries *E. coli* F18 dans une chambre à écoulement laminaire tapissée d'albumine de sérum bovin mannosylée (*mannose-BSA*). Le déclin biexponentiel de la distribution de temps de pause leur suggère un modèle allostérique à deux états. En faisant l'hypothèse que chaque pause est due à une unique liaison accrocheuse et que celle-ci supporte l'intégralité de la force de traction exercée sur la bactérie, les auteur·es parviennent à paramétrer un modèle cinétique à deux états séparés par une barrière énergétique (voir 1.3.5 et FIGURE 15). Des modèles alternatifs à un état et deux chemins ou même à deux sites distincts ne parviennent pas à expliquer les données.

Les paramètres du modèle ajustés sur les données sont reproduits dans la TABLE 2.

On peut constater la différence de k^0 de trois ordres de grandeur entre l'état de basse affinité (1) et celui de haute affinité (2). La vitesse d'interconversion entre les deux états est intermédiaire. Cela signifie qu'en l'absence de force, le chemin de dissociation favorisé est $1 \rightarrow 0$, bien que l'état lié (2) soit thermodynamiquement favorisé par rapport à l'état lié (1). La valeur de Δx_{12} à 8,58 Å indique que la force accélère considérablement la transition $1 \rightarrow 2$. Il est intéressant de noter que cette valeur est commensurable avec la longueur d'extension du *linker*.

Transition	k^0 [s^{-1}]	Δx [Å]
$1 \rightarrow 0$	6,01	1,37
$2 \rightarrow 0$	0,007	1,76
$1 \rightarrow 2$	0,210	8,58
$2 \rightarrow 1$	0,105	-4,2

TABLE 2 – Paramètres du modèle cinétique de THOMAS et al. [40]. L'état 1 représente l'état lié de basse affinité (état allostérique A) et l'état 2 celui de haute affinité (état S). L'état 0 représente sans distinction l'ensemble des états dissociés.

La valeur négative de Δx_{21} indique au contraire que la réaction retour est défavorisée par la force.

Il apparaît donc qu'un modèle phénoménologique à deux états est suffisant pour expliquer la formation de liaisons accrocheuses par FimH et la diminution de la mobilité bactérienne sous fort cisaillement fluide. Néanmoins, la capacité d'un tel modèle mathématique à expliquer les données et faire des prédictions ne garantit pas sa validité à l'échelle moléculaire. Des modèles structuraux doivent venir étayer le modèle cinétique afin de compléter le mécanisme allostérique.

2.4.2 Modèle thermodynamique

Une modélisation plus poussée du mécanisme d'accroche de l'extrémité piliale est proposée par SAUER et al. [29, 42]. Contrairement au modèle purement cinétique fondé sur les expériences en chambre à écoulement laminaire, ce modèle peut être qualifié de structural car il associe chaque état identifié à une conformation issue des données cristallographiques.

Comme évoqué précédemment, la protéine FimH isolée sans complémentation du domaine piline n'est que marginalement stable et a tendance à s'agréger en conditions physiologiques. SAUER et al. parviennent à stabiliser l'adhésine en la complétant avec le peptide brin donneur de FimG (DsG) *via* une réaction d'échange de brin donneur (*donor strand exchange*, DSE) venant le substituer à la chaperone FimC :



Le complexe FimH·DsG ainsi obtenu est utilisé pour modéliser l'adhésine dans des conditions proches de son positionnement *in vivo* à l'apex du pilus où les deux domaines peuvent librement s'associer. En effet la cristallisation de FimH·DsG donne une structure virtuellement identique à celle observée dans le complexe précédemment observé de l'extrémité piliale FimC·FimF·FimG·FimH [27] (FIGURE 11B).

Ligand	$K_d(\text{H}\cdot\text{DsG})/K_d(\text{H}_L)$	$k_{\text{on}}(\text{H}\cdot\text{DsG})/k_{\text{on}}(\text{H}_L)$	$k_{\text{off}}(\text{H}\cdot\text{DsG})/k_{\text{off}}(\text{H}_L)$	$t_{1/2}(\text{H}\cdot\text{DsG})$ [ms]	$t_{1/2}(\text{H}_L)$ [min]
Man	2100	21	52000	11	10
$\alpha 6\text{Man}_2$	1900	33	76000	11	13
HMan	3300	28	110000	32	58

TABLE 3 – Comparaison des propriétés de liaison de FimH·DsG et FimH_L pour plusieurs ligands. Man : D-mannose. $\alpha 6\text{Man}_2$: $\text{Man}\alpha(1 \rightarrow 6)\text{Man}\alpha\text{Me}$. [42]. HMan : heptyl- α -D-mannose. [29].

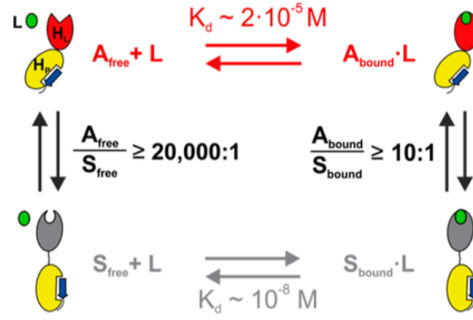


FIGURE 16 – Modèle thermodynamique de SAUER et al. [42]. Les auteurs proposent un cycle thermodynamique à quatre états (deux états liés). Repris de [42].

Les auteurs nomment cet état A_{free} (pour *Associated, ligand-free state*). En revanche la co-cristallisation avec un mannoside (heptyl- α -D-mannose, HM) dévoile une nouvelle conformation avec le site de liaison refermé autour du ligand (FIGURE 11C) baptisée A_{bound} (*Associated, ligand-bound*).

Dans notre nomenclature, $^{\circ}A_{\text{op}}^{\text{LP}}$

$\bullet A_{\text{cl}}^{\text{LP}}$

En utilisant un variant présentant des interactions inter-domaines affaiblies (K12) et en remplaçant le brin DsG par l'équivalent non-apparenté de l'unité FimF, DsF, SAUER et al. parviennent également à isoler une structure cristallographique de la protéine entière ligantée (FimH·DsF·HM) où les deux domaines sont séparés. Le domaine lectine y est alors sous sa forme allongée, identique à celles trouvées précédemment lorsqu'il est exprimé en isolation ou que le domaine piline est stabilisé par FimC (FIGURE 11A). Cet état est baptisé S_{bound} (*Separated, ligand-bound*) et constitue hypothétiquement la forme activée de la liaison accrocheuse lorsque l'adhésine est soumise à une force de tension.

$\bullet S_{\text{cl}}^{\text{LP}}$

Grâce à la caractérisation expérimentale des constantes d'affinité et cinétiques de liaison à HM des construits FimH·DsG d'une part, et du domaine lectine isolé FimH_L, utilisé comme proxy de l'état séparé, d'autre part, un cycle thermodynamique est proposé pour expliquer le mécanisme de grappin moléculaire (FIGURE 16).

La multiplication par $\sim 1900\text{--}3300$ de l'affinité pour divers ligands mannosides dans FimH_L en comparaison à FimH·DsG est principalement liée à une vitesse de dissociation de l'ordre de $10^4\text{--}10^5$ fois plus lente et une vitesse d'association seulement 20 à 50 fois plus lente. Cette différence est remarquable étant donné l'extrême similarité de la conformation du site de liaison dans les structures A_{bound} et S_{bound} . Un élément de réponse est donné par des simulations de dynamique moléculaire montrant une plus grande flexibilité du site actif dans l'état A_{bound} , en particulier au niveau de la boucle de fermeture. Ces fluctuations pourraient en partie expliquer la différence frappante de k_{off} malgré la similarité structurale. Cela conduit les auteurs à employer le concept d'allostérie dynamique pour qualifier l'effet de l'interaction inter-domaines sur les fluctuations au niveau du site de liaison.

La différence d'affinité entre les deux formes s'exprime par une différence d'énergie libre de liaison d'environ 20 kJ mol^{-1} . Les auteurs notent qu'une telle différence est commensurable avec le travail mécanique nécessaire à éloigner les deux domaines d'une distance de 11 \AA (ce qui correspond à l'extension complète du *linker*) sous une force de 40 pN . Dans cette optique, le travail mécanique fourni par l'environnement serait converti en énergie chimique faisant passer l'adhésine de l'état A_{bound} à un état S_{bound} thermodynamiquement moins favorable mais cinétiquement plus stable.

2.5 CONCLUSION : QUE RESTE-T-IL À DÉCOUVRIR ?

Dans ce chapitre nous avons esquissé une brève revue de l'état des connaissances sur l'adhésine FimH et sur le mécanisme à l'œuvre dans sa capacité à former des liaisons accrocheuses. Nous pouvons dire aujourd'hui que FimH est probablement la protéine grappin la plus étudiée et la mieux comprise. Les expériences sur molécule unique combinées à l'élucidation de plusieurs structures cristallographiques ont permis de démontrer sans ambiguïté la pertinence du modèle allostérique à deux états, tant sur le plan cinétique que structural. Toutefois, il serait précipité de conclure que nous en avons terminé avec la compréhension de FimH. Un certain nombre de questions demeurent en suspens.

Le premier point de scepticisme doit être soulevé quant à la correspondance supposée entre les structures cristallographiques observées et les états constitutifs des modèles cinétiques et thermodynamiques. Bien qu'elle nous ouvre une fenêtre inestimable sur la relation structure-fonction, qui est aux fondements de la biologie structurale contemporaine, la cristallographie aux rayons X ne peut et ne doit être considérée comme pleinement représentative des biomolécules *in situ*. Les conditions particulières nécessaires au processus de cristalli-

sation sont par essence différentes de celles rencontrées en solution. Le pH de cristallisation, en particulier, peut influencer les structures obtenues. L'empilement cristallin est susceptible de créer des artefacts en sélectionnant les conformations les plus stables au sein du cristal [43]. En piégeant les molécules dans un puits énergétique particulier et en reposant sur une moyenne, à la fois temporelle et d'ensemble, la cristallographie propose une vision statique des structures observées. Pourtant, à température ambiante, les protéines sont des structures flexibles qui visitent un ensemble de conformations métastables. Cette flexibilité est essentielle à leur fonction biologique. Les facteurs B, ou facteurs de température, sont fréquemment utilisés comme une indication sur la rigidité locale de la structure protéique en mesurant l'atténuation de la diffraction liée aux mouvements atomiques (désordre dynamique) et aux hétérogénéités (désordre statique) au sein du cristal. Toutefois leur interprétation est soumise à caution, et des études comparatives avec des simulations tendent à montrer qu'ils sous-estiment fortement le degré d'hétérogénéité réel [44].

Le problème est illustré dans le cas de FimH par la conformation remarquablement identique du site de liaison dans les structures identifiées aux états $\bullet A_{cl}$ et $\bullet S_{cl}$ [29]. Une telle similarité peut paraître à première vue contradictoire avec plusieurs ordres de grandeurs de différence en termes d'affinité et de k_{off} . Le recours à une explication dynamique s'impose, mais est difficile à valider par cristallographie. Il est donc essentiel de faire appel à d'autres techniques (RMN, MD) pour comprendre cette observation.

Un autre champ d'ignorance concerne l'interprétation des paramètres phénoménologiques des modèles cinétiques tirés des expériences sous force (k^0 et Δx). Ceux-ci ont une signification intuitive lorsqu'on les dérive d'un paysage énergétique simple, mais comment les mettre en relation avec les géométries moléculaires des systèmes réels ? En particulier, les hypothèses d'unicité du chemin réactionnel et de la barrière énergétique peuvent sembler irréalistes au vu de la complexité de la surface d'énergie potentielle des systèmes biologiques. Les protéines sont notoirement connues pour présenter des paysages énergétiques « rugueux », de haute dimensionnalité, constitués d'une multitude d'états métastables, et on peut douter de la validité des modèles théoriques idéalisés pour lesquels des prédictions analytiques ne sont souvent dérivables que dans le cas unidimensionnel. Il est plus crédible de comprendre les paramètres ajustés sur les données expérimentales macroscopiques comme des paramètres phénoménologiques effectifs, qui émergent d'une moyenne d'ensemble d'une multitude de chemins microscopiques possibles [45-47]. Là encore, des techniques permettant d'avoir accès aux échelles fines, tant spatiales que temporelles, sont nécessaires. En la matière, les simulations numériques offrent des voies prometteuses.

Les développements techniques et théoriques récents de la modélisation numérique ont ouvert l'accès à des échelles temporelles de la dynamique macromoléculaire longtemps inaccessibles. Il devient possible de sonder par le calcul des phénomènes moléculaires prenant place sur plusieurs microsecondes voire millisecondes, et ce avec un niveau de détail atomistique. Ce faisant, nous pouvons mettre à l'épreuve, sous un angle nouveau, les résultats théoriques et leur cohérence avec les observations expérimentales.

Pour conclure, résumons simplement les questions que nous tenterons d'aborder par la suite :

- Quelle est la plasticité réelle du site de liaison dans différents états allostériques ? Cette plasticité peut-elle expliquer les différentes stabilités du complexe ?
- À quels ensembles d'états microscopiques correspondent vraiment les états macroscopiques décrits dans les modèles de [40] et [29] ? Les structures cristallographiques sont-elles pleinement représentatives de ces états ? Comment expliquer la différence de k_{off} malgré la similitude des structures $\bullet A_{\text{cl}}$ et $\bullet S_{\text{cl}}$?
- Quel est l'effet de l'application d'une force sur la plasticité conformationnelle de FimH et comment cela impacte-t-il le mécanisme grappin ? À quels aspects structuraux correspondent les paramètres k^0 et Δx dans le modèle cinétique expérimental ?
- Comment s'effectue la transition allostérique $A \leftrightarrow S$ et peut-on l'observer *in silico* ?

RÉFÉRENCES

- [1] G. T. WERNEBURG et D. G. THANASSI. Pili Assembled by the Chaperone/Usher Pathway in *Escherichia coli* and *Salmonella*. *EcoSal Plus*, **8**, (2018).
DOI : [10.1128/ecosalplus.ESP-0007-2017](https://doi.org/10.1128/ecosalplus.ESP-0007-2017)
- [2] G. ZHOU, W.-J. MO, P. SEBBEL, G. MIN, T. A. NEUBERT, R. GLOCKSHUBER, X.-R. WU, T.-T. SUN et X.-P. KONG. Uroplakin Ia is the urothelial receptor for uropathogenic *Escherichia coli* : evidence from in vitro FimH binding. *Journal of Cell Science*, **114**, 4095-4103 (2001).
DOI : [10.1242/jcs.114.22.4095](https://doi.org/10.1242/jcs.114.22.4095)
- [3] F. NEAMATI, R. MONIRI, A. KHORSHIDI et M. SAFFARI. Structural and Functional Characterization of the Fimh Adhesin of Uropathogenic *Escherichia coli* and Its Novel Applications. *Journal of Bacteriology and Mycology*, **7**, 1142 (2020).
- [4] D. I. KISIELA, J. J. KRAMER, V. TCHESNOKOVA, P. APRIKIAN, V. YAROV-YAROVY, S. CLEGG et E. V. SOKURENKO. Allosteric Catch Bond Properties of the FimH Adhesin from *Salmonella enterica* Serovar Typhimurium. *Journal of Biological Chemistry*, **286**, 38136-38147 (2011).
DOI : [10.1074/jbc.M111.237511](https://doi.org/10.1074/jbc.M111.237511)
- [5] S. D. KNIGHT et J. BOUCKAERT. Structure, Function, and Assembly of Type 1 Fimbriae. *Glycoscience and Microbial Adhesion*. Sous la dir. de T. K. LINDHORST et S. OSCARSON. T. 288. Series Title : Topics in Current Chemistry. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009. 67-107. DOI : [10.1007/128_2008_13](https://doi.org/10.1007/128_2008_13)
- [6] B. FOXMAN. Epidemiology of urinary tract infections : Incidence, morbidity, and economic costs. *Disease-a-Month*, **49**, 53-70 (2003).
DOI : [10.1067/mda.2003.7](https://doi.org/10.1067/mda.2003.7)
- [7] M. E. TERLIZZI, G. GRIBAUDO et M. E. MAFFEI. UroPathogenic *Escherichia coli* (UPEC) Infections : Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. *Frontiers in Microbiology*, **8**, 1566 (2017).
DOI : [10.3389/fmicb.2017.01566](https://doi.org/10.3389/fmicb.2017.01566)
- [8] E. HAHN, P. WILD, U. HERMANN, P. SEBBEL, R. GLOCKSHUBER, M. HÄNER, N. TASCHNER, P. BURKHARD, U. AEBI et S. A. MÜLLER. Exploring the 3D Molecular Architecture of *Escherichia coli* Type 1 Pili. *Journal of Molecular Biology*, **323**, 845-857 (2002).
DOI : [10.1016/S0022-2836\(02\)01005-7](https://doi.org/10.1016/S0022-2836(02)01005-7)
- [9] M. L. VALENSKI, S. L. HARRIS, P. A. SPEARS, J. R. HORTON et P. E. ORNDORFF. The Product of the *fimI* Gene Is Necessary for *Escherichia coli* Type 1 Pilus Biosynthesis. *Journal of Bacteriology*, **185**, 5007-5011 (2003).
DOI : [10.1128/JB.185.16.5007-5011.2003](https://doi.org/10.1128/JB.185.16.5007-5011.2003)
- [10] G. CAPITANI, O. EIDAM, R. GLOCKSHUBER et M. G. GRÜTTER. Structural and functional insights into the assembly of type 1 pili from *Escherichia coli*. *Microbes and Infection*, **8**, 2284-2290 (2006).
DOI : [10.1016/j.micinf.2006.03.013](https://doi.org/10.1016/j.micinf.2006.03.013)

- [11] M. HARTMANN et T. K. LINDHORST. The Bacterial Lectin FimH, a Target for Drug Discovery - Carbohydrate Inhibitors of Type 1 Fimbriae-Mediated Bacterial Adhesion. *European Journal of Organic Chemistry*, **2011**, 3583-3609 (2011).
DOI : [10.1002/ejoc.201100407](https://doi.org/10.1002/ejoc.201100407)
- [12] E. V. SOKURENKO, V. VOGEL et W. E. THOMAS. Catch-Bond Mechanism of Force-Enhanced Adhesion : Counterintuitive, Elusive, but ... Widespread ? *Cell Host & Microbe*, **4**, 314-323 (2008).
DOI : [10.1016/j.chom.2008.09.005](https://doi.org/10.1016/j.chom.2008.09.005)
- [13] B. SAVAGE, E. SALDÍVAR et Z. M. RUGGERI. Initiation of Platelet Adhesion by Arrest onto Fibrinogen or Translocation on von Willebrand Factor. *Cell*, **84**, 289-297 (1996).
DOI : [10.1016/S0092-8674\(00\)80983-6](https://doi.org/10.1016/S0092-8674(00)80983-6)
- [14] E. B. FINGER, K. D. PURL, R. ALON, M. B. LAWRENCE, U. H. von ANDRIAN et T. A. SPRINGER. Adhesion through L-selectin requires a threshold hydrodynamic shear. *Nature*, **379**, 266-269 (1996).
DOI : [10.1038/379266a0](https://doi.org/10.1038/379266a0)
- [15] Z. LI, N. MOHAMED et J. ROSS. Shear Stress Affects the Kinetics of *Staphylococcus aureus* Adhesion to Collagen. *Biotechnology Progress*, **16**, 1086-1090 (2000).
DOI : [10.1021/bp000117r](https://doi.org/10.1021/bp000117r)
- [16] W. E. THOMAS, E. TRINTCHINA, M. FORERO, V. VOGEL et E. V. SOKURENKO. Bacterial Adhesion to Target Cells Enhanced by Shear Force. *Cell*, **109**, 913-923 (2002).
DOI : [10.1016/S0092-8674\(02\)00796-1](https://doi.org/10.1016/S0092-8674(02)00796-1)
- [17] W. THOMAS. Catch Bonds in Adhesion. *Annual Review of Biomedical Engineering*, **10**, 39-57 (2008).
DOI : [10.1146/annurev.bioeng.10.061807.160427](https://doi.org/10.1146/annurev.bioeng.10.061807.160427)
- [18] Y. F. DUFRÊNE, A. VILJOEN, J. MIGNOLET et M. MATHELIÉ-GUINLET. AFM in cellular and molecular microbiology. *Cellular Microbiology*, **23**, (2021).
DOI : [10.1111/cmi.13324](https://doi.org/10.1111/cmi.13324)
- [19] B. T. MARSHALL, M. LONG, J. W. PIPER, T. YAGO, R. P. McEVER et C. ZHU. Direct observation of catch bonds involving cell-adhesion molecules. *Nature*, **423**, 190-193 (2003).
DOI : [10.1038/nature01605](https://doi.org/10.1038/nature01605)
- [20] O. YAKOVENKO, S. SHARMA, M. FORERO, V. TCHESNOKOVA, P. APRIKIAN, B. KIDD, A. MACH, V. VOGEL, E. SOKURENKO et W. E. THOMAS. FimH Forms Catch Bonds That Are Enhanced by Mechanical Force Due to Allosteric Regulation. *Journal of Biological Chemistry*, **283**, 11596-11605 (2008).
DOI : [10.1074/jbc.M707815200](https://doi.org/10.1074/jbc.M707815200)
- [21] E. MILLER, T. GARCIA, S. HULTGREN et A. F. OBERHAUSER. The Mechanical Properties of *E. coli* Type 1 Pili Measured by Atomic Force Microscopy Techniques. *Biophysical Journal*, **91**, 3848-3856 (2006).
DOI : [10.1529/biophysj.106.088989](https://doi.org/10.1529/biophysj.106.088989)

- [22] A. ALONSO-CABALLERO, J. SCHÖNFELDER, S. POLY, F. CORSETTI, D. DE SANCHO, E. ARTACHO et R. PEREZ-JIMENEZ. Mechanical architecture and folding of *E. coli* type 1 pilus domains. *Nature Communications*, **9**, (2018).
DOI : [10.1038/s41467-018-05107-6](https://doi.org/10.1038/s41467-018-05107-6)
- [23] A. JACQUOT, C. SAKAMOTO, A. RAZAFITIANAMAHARAVO, C. CAILLET, J. MERLIN, A. FAHS, J. M. GHIGO, C. BELOIN, J. F. L. DUVAL et G. FRANCIUS. Dynamic Modulation of Fimbrial Extension and FimH-Mannose Binding Force on Live Bacteria Under pH Changes : A Molecular Atomic Force Microscopy Analysis. *Journal of Biomedical Nanotechnology*, **10**, 3361-3372 (2014).
DOI : [10.1166/jbn.2014.1905](https://doi.org/10.1166/jbn.2014.1905)
- [24] C. PUORGER, O. EIDAM, G. CAPITANI, D. ERILOV, M. G. GRÜTTER et R. GLOCKSHUBER. Infinite Kinetic Stability against Dissociation of Supramolecular Protein Complexes through Donor Strand Complementation. *Structure*, **16**, 631-642 (2008).
DOI : [10.1016/j.str.2008.01.013](https://doi.org/10.1016/j.str.2008.01.013)
- [25] D. CHOUDHURY, A. THOMPSON, V. STOJANOFF, S. LANGERMANN, J. PINKNER, S. J. HULTGREN et S. D. KNIGHT. X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic *Escherichia coli*. *Science*, **285**, 1061-1066 (1999).
DOI : [10.1126/science.285.5430.1061](https://doi.org/10.1126/science.285.5430.1061)
- [26] J. BOUCKAERT et al. Receptor binding studies disclose a novel class of high-affinity inhibitors of the *Escherichia coli* FimH adhesin : A novel class of FimH high-affinity ligands. *Molecular Microbiology*, **55**, 441-455 (2004).
DOI : [10.1111/j.1365-2958.2004.04415.x](https://doi.org/10.1111/j.1365-2958.2004.04415.x)
- [27] I. L. TRONG et al. Structural Basis for Mechanical Force Regulation of the Adhesin FimH via Finger Trap-like Sheet Twisting. *Cell*, **141**, 645 (2010).
DOI : [10.1016/j.cell.2010.03.038](https://doi.org/10.1016/j.cell.2010.03.038)
- [28] P. MAGALA, R. E. KLEVIT, W. E. THOMAS, E. V. SOKURENKO et R. E. STENKAMP. RMSD analysis of structures of the bacterial protein FimH identifies five conformations of its lectin domain. *Proteins : Structure, Function, and Bioinformatics* (2019).
DOI : [10.1002/prot.25840](https://doi.org/10.1002/prot.25840)
- [29] M. M. SAUER, R. P. JAKOB, J. ERAS, S. BADAY, D. ERIŞ, G. NAVARRA, S. BERNÈCHE, B. ERNST, T. MAIER et R. GLOCKSHUBER. Catch-bond mechanism of the bacterial adhesin FimH. *Nature Communications*, **7**, (2016).
DOI : [10.1038/ncomms10738](https://doi.org/10.1038/ncomms10738)
- [30] J. LIU, L. A. N. AMARAL et S. KETEN. Conformational stability of the bacterial adhesin, FimH, with an inactivating mutation. *Proteins : Structure, Function, and Bioinformatics*, **89**, 276-288 (2021).
DOI : [10.1002/prot.26013](https://doi.org/10.1002/prot.26013)
- [31] J. MONOD, J. WYMAN et J.-P. CHANGEUX. On the nature of allosteric transitions : A plausible model. *Journal of Molecular Biology*, **12**, 88-118 (1965).
DOI : [10.1016/S0022-2836\(65\)80285-6](https://doi.org/10.1016/S0022-2836(65)80285-6)

- [32] S. VANWETSWINKEL, A. N. VOLKOV, Y. G. J. STERCKX, A. GARCIA-PINO, L. BUTS, W. F. VRANKEN, J. BOUCKAERT, R. ROY, L. WYNS et N. A. J. van NULAND. Study of the Structural and Dynamic Effects in the FimH Adhesin upon α -D-Heptyl Mannose Binding. *Journal of Medicinal Chemistry*, **57**, 1416-1427 (2014).
DOI : [10.1021/jm401666c](https://doi.org/10.1021/jm401666c)
- [33] V. KALAS et al. Evolutionary fine-tuning of conformational ensembles in FimH during host-pathogen interactions. *Science Advances*, **3**, e1601944 (2017).
DOI : [10.1126/sciadv.1601944](https://doi.org/10.1126/sciadv.1601944)
- [34] C. P. SAGER, D. ERIŞ, M. SMIEŠKO, R. HEVEY et B. ERNST. What contributes to an effective mannose recognition domain? *Beilstein Journal of Organic Chemistry*, **13**, 2584-2595 (2017).
DOI : [10.3762/bjoc.13.255](https://doi.org/10.3762/bjoc.13.255)
- [35] D. J. SCHWARTZ, V. KALAS, J. S. PINKNER, S. L. CHEN, C. N. SPAULDING, K. W. DODSON et S. J. HULTGREN. Positively selected FimH residues enhance virulence during urinary tract infection by altering FimH conformation. *Proceedings of the National Academy of Sciences*, **110**, 15530-15537 (2013).
DOI : [10.1073/pnas.1315203110](https://doi.org/10.1073/pnas.1315203110)
- [36] B. FIEGE, S. RABBANI, R. C. PRESTON, R. P. JAKOB, P. ZIHLMANN, O. SCHWARDT, X. JIANG, T. MAIER et B. ERNST. The Tyrosine Gate of the Bacterial Lectin FimH : A Conformational Analysis by NMR Spectroscopy and X-ray Crystallography. *ChemBioChem*, **16**, 1235-1246 (2015).
DOI : [10.1002/cbic.201402714](https://doi.org/10.1002/cbic.201402714)
- [37] S. RABBANI et al. Mutation of Tyr₁₃₇ of the universal *Escherichia coli* fimbrial adhesin FimH relaxes the tyrosine gate prior to mannose binding. *IUCr*, **4**, 7-23 (2017).
DOI : [10.1107/S2052252516016675](https://doi.org/10.1107/S2052252516016675)
- [38] M. TOUAIBIA et al. Sites for Dynamic Protein-Carbohydrate Interactions of O- and C-Linked Mannosides on the E. coli FimH Adhesin. *Molecules*, **22**, 1101 (2017).
DOI : [10.3390/molecules22071101](https://doi.org/10.3390/molecules22071101)
- [39] A. WELLENS, M. LAHMANN, M. TOUAIBIA, J. VAUCHER, S. OSCARSON, R. ROY, H. REMAUT et J. BOUCKAERT. The Tyrosine Gate as a Potential Entropic Lever in the Receptor-Binding Site of the Bacterial Adhesin FimH. *Biochemistry*, **51**, 4790-4799 (2012).
DOI : [10.1021/bi300251r](https://doi.org/10.1021/bi300251r)
- [40] W. THOMAS, M. FORERO, O. YAKOVENKO, L. NILSSON, P. VICINI, E. SOKURENKO et V. VOGEL. Catch-Bond Model Derived from Allostery Explains Force-Activated Bacterial Adhesion. *Biophysical Journal*, **90**, 753-764 (2006).
DOI : [10.1529/biophysj.105.066548](https://doi.org/10.1529/biophysj.105.066548)
- [41] W. E. THOMAS, V. VOGEL et E. SOKURENKO. Biophysics of Catch Bonds. *Annual Review of Biophysics*, **37**, 399-416 (2008).
DOI : [10.1146/annurev.biophys.37.032807.125804](https://doi.org/10.1146/annurev.biophys.37.032807.125804)

- [42] M. M. SAUER, R. P. JAKOB, T. LUBER, F. CANONICA, G. NAVARRA, B. ERNST, C. UNVERZAGT, T. MAIER et R. GLOCKSHUBER. Binding of the Bacterial Adhesin FimH to Its Natural, Multivalent High-Mannose Type Glycan Targets. *Journal of the American Chemical Society*, **141**, 936-944 (2019).
DOI : [10.1021/jacs.8b10736](https://doi.org/10.1021/jacs.8b10736)
- [43] A. M. DAVIS, S. J. TEAGUE et G. J. KLEYWEGT. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angewandte Chemie International Edition*, **42**, 2718-2736 (2003).
DOI : [10.1002/anie.200200539](https://doi.org/10.1002/anie.200200539)
- [44] A. KUZMANIC, N. S. PANNU et B. ZAGROVIC. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nature Communications*, **5**, 3220 (2014).
DOI : [10.1038/ncomms4220](https://doi.org/10.1038/ncomms4220)
- [45] O. K. DUDKO, G. HUMMER et A. SZABO. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences*, **105**, 15755-15760 (2008).
DOI : [10.1073/pnas.0806085105](https://doi.org/10.1073/pnas.0806085105)
- [46] R. SATIJA, A. M. BEREZHKOVSII et D. E. MAKAROV. Broad distributions of transition-path times are fingerprints of multidimensionality of the underlying free energy landscapes. *Proceedings of the National Academy of Sciences* 202008307 (2020).
DOI : [10.1073/pnas.2008307117](https://doi.org/10.1073/pnas.2008307117)
- [47] J. SCHÖNFELDER, R. PEREZ-JIMENEZ et V. MUÑOZ. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nature Communications*, **7**, 11777 (2016).
DOI : [10.1038/ncomms11777](https://doi.org/10.1038/ncomms11777)

Deuxième partie

PLASTICITÉ CONFORMATIONNELLE DE FIMH

Dans cette première partie de résultats nous décrivons notre étude de la plasticité conformationnelle de l'adhésine FimH par des méthodes computationnelles fondées sur la dynamique moléculaire. Un premier chapitre abordera l'exploration qualitative du paysage conformationnel du domaine lectine. Le suivant se focalisera plus quantitativement sur l'ouverture et la fermeture du site de liaison.

Nous avons établi que la riche littérature expérimentale sur l'adhésine FimH permettait de construire des modèles cinétiques, thermodynamiques et structuraux relativement détaillés de son mécanisme de grappin moléculaire. C'est notamment grâce à la caractérisation de plusieurs conformations par cristallographie aux rayons X que la mise en relation d'observables macroscopiques (affinité, constantes cinétiques) avec la structure moléculaire est possible. Néanmoins, le paysage conformationnel de la protéine tel que reconstitué par les divers cristallogrammes ne saurait être considéré *a priori* comme complet. En outre, si la cristallographie permet la reconstruction précise d'une conformation proche d'un minimum énergétique, elle donne peu d'information sur la dynamique et les hétérogénéités conformationnelles dans des conditions biologiques.

Ce chapitre a donc pour ambition de présenter les techniques dérivées de la dynamique moléculaire (MD, *molecular dynamics*), et notamment les techniques d'échantillonnage avancé, que nous avons utilisées pour tenter de compléter le tableau du paysage conformationnel de FimH, et en particulier celui du domaine lectine et de son site de liaison. Nous en résumons les grands principes avant d'exposer nos résultats.

3.1 UN OUTIL : LA DYNAMIQUE MOLÉCULAIRE

3.1.1 *Principe général*

L'évolution des systèmes chimiques est décrite – en principe complètement – par les lois de la mécanique quantique à travers la résolution de l'équation de SCHRÖDINGER. L'état du système est alors caractérisé par sa fonction d'onde, à partir de laquelle on peut dériver la probabilité à chaque instant de mesurer une certaine observable, telle que la position ou la quantité de mouvement d'une particule donnée. À l'exception de cas particuliers comprenant généralement une unique particule dans un potentiel simple, l'équation de SCHRÖDINGER est trop complexe pour en trouver une solution analytique et notre seul espoir est de se tourner vers des solutions numériques approchées. Même dans ce cas, décrire l'évolution de chaque particule élémentaire, chaque électron et chaque quark d'un système polyatomique tel qu'une molécule d'ADN, se révélerait d'une complexité

astronomique. Nous devons donc adapter notre description mathématique grâce à plusieurs approximations, pertinentes aux échelles des phénomènes qui nous intéressent.

La dynamique moléculaire (MD) propose de modéliser les systèmes chimiques grâce à deux approximations fondamentales : la première, celle de BORN-OPPENHEIMER, suppose que la dynamique électronique est instantanée par rapport à celle des noyaux, autrement dit que la disposition des électrons dépend de manière unique et indépendante du temps de la donnée des positions nucléaires ; la deuxième suppose que les noyaux peuvent être représentés par des points matériels massifs dans un espace-temps euclidien et obéissant à une mécanique classique newtonienne. Les électrons et autres composants subatomiques ne sont donc pas explicitement représentés mais leurs interactions sont prises en compte de façon émergente au sein d'une surface d'énergie potentielle $U(\mathbf{x})$, correspondant à des forces conservatives, guidant l'évolution des positions \mathbf{x} et quantités de mouvement \mathbf{p} des atomes. Les équations de HAMILTON-JACOBI expriment ces lois du mouvement :

$$\dot{\mathbf{x}} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \quad ; \quad \dot{\mathbf{p}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \quad (11)$$

où on note \mathcal{H} le hamiltonien du système, somme de l'énergie potentielle et de l'énergie cinétique K ,

$$\mathcal{H}(\mathbf{x}, \mathbf{p}) = U(\mathbf{x}) + K(\mathbf{p}). \quad (12)$$

La caractérisation de ce potentiel constitue la partie la plus délicate de cette approche. Les méthodes fondées sur les premiers principes de la mécanique quantique, permettant de recalculer à chaque pas de temps la configuration électronique, et donc les forces s'exerçant sur les noyaux, sont dénommées dynamique moléculaire *ab initio*. Elles offrent l'avantage de pouvoir atteindre de bons degrés de précision et de permettre l'étude de réactions chimiques impliquant rupture et formation de liaisons covalentes, mais sont très gourmandes en ressources de calcul et se limitent actuellement à l'évolution de systèmes de quelques centaines d'atomes sur quelques dizaines de picosecondes. D'autres approches, beaucoup moins coûteuses, cherchent à construire la surface d'énergie potentielle comme une fonction des coordonnées nucléaires, baptisée champ de force (*force-field*), dont la forme et les paramètres sont déterminés de façon « empirique ». Elles constituent ce que l'on appelle la dynamique moléculaire classique et permettent aujourd'hui de modéliser l'évolution sur plusieurs microsecondes de systèmes composés de millions d'atomes, et même dépassant le milliard [1].

3.1.2 Paramétrage des champs de force

Le paramétrage des champs de force classiques fait l'objet d'une recherche intensive [2, 3]. La plupart des champs de force actuels sont une somme additive de termes énergétiques dépendant des positions relatives de groupes d'atomes, distinguant généralement :

- Des interactions liantes, définissant la topologie des molécules, c'est-à-dire des groupes d'atomes connectés par des liaisons covalentes. On y inclut des termes d'élongation (longueur de liaison), de flexion (angle entre deux liaisons), de torsions (angle dièdre propre ou impropre formé par trois liaisons contigües).
- Des interactions non-liantes, comprenant des interactions de paires à courte distance (interactions de VAN DER WAALS) et à longue distance (interactions électrostatiques ou de COULOMB).

Chacun de ces termes suppose une forme fonctionnelle à priori, choisie par les théoriciens dans l'espoir que leur addition reproduise fidèlement l'énergie potentielle totale du système pour toute configuration atomique « réaliste ». L'efficacité computationnelle ou la stabilité numérique sont également des critères faisant privilégier certaines formes mathématiques à d'autres. Les paramètres associés à chacune de ces interactions sont ensuite ajustés par diverses méthodes, incluant :

- des calculs quantiques *ab initio*,
- des distributions empiriques observées, par exemple, dans les structures cristallographiques,
- la comparaison aux mesures expérimentales (calorimétriques, spectroscopiques, ...).

Ce paramétrage est un processus itératif qui nécessite de constants allers-retours entre mesures expérimentales et simulations. La validation et le perfectionnement des champs de force empiriques sont limités tant par la résolution des mesures expérimentales que par la puissance de calcul nécessaire pour générer des prédictions testables à partir du modèle.

3.1.3 Ambitions et limitations de la dynamique moléculaire

La dynamique moléculaire propose de prédire l'évolution déterministe d'un système moléculaire. Une telle ambition est vouée à l'échec de par la nature chaotique des systèmes complexes à N-corps, pour lesquels une incertitude même minime sur la configuration initiale et la forme du potentiel, sans parler de la précision numérique forcément limitée, conduit inévitablement à une divergence rapide des trajectoires obtenues. Fort heureusement les propriétés mesurables à l'échelle macroscopiques dépendent rarement de la trajectoire précise d'une molécule, mais émergent du comportement collectif d'un grand nombre d'entre elles. À condition que la MD produise des tra-

jectoires « typiques » (à défaut d'être « exactes »), il est possible d'en tirer des informations statistiques sur le système à l'étude.

Les outils pour mener à bien et interpréter les simulations de MD sont ceux de la thermodynamique statistique, et permettent de différencier deux types de propriétés des systèmes physico-chimiques :

1. Les propriétés d'équilibre reposent sur le postulat fondamental qu'en l'absence de toute information, l'état microscopique d'un système isolé à N particules, occupant un volume V et possédant une énergie donnée E , a une probabilité égale de se trouver dans n'importe lequel des micro-états possibles de même énergie. L'ensemble de ces micro-états équiprobables est nommé ensemble microcanonique. Si on autorise le système à échanger de l'énergie thermique et du travail mécanique avec son environnement il est ensuite possible de construire les ensembles canonique (N , V , T fixés) ou isobare-isotherme (N , p , T fixés). Ces ensembles thermodynamiques définissent des distributions de probabilité de trouver le système dans un état microscopique d'énergie E , et donc dans une certaine configuration atomique $\{\mathbf{x}, \mathbf{p}\}$. Les propriétés observables (macroscopiques) du système découlent alors de cette distribution de probabilité sur les configurations.

À défaut de pouvoir la calculer exactement (en raison de la très haute dimensionnalité de l'espace des phases), la MD permet d'échantillonner cette distribution de manière efficace. L'hypothèse sous-jacente est que la trajectoire visitera aux temps longs chaque configuration accessible avec une probabilité égale à celle de l'ensemble thermodynamique correspondant. C'est l'hypothèse d'ergodicité, qui stipule que la moyenne temporelle d'une observable $A(t)$ tend aux temps infinis vers la moyenne d'ensemble thermodynamique :

$$\langle A \rangle = \lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t dt' A(t'). \quad (13)$$

Un exemple de propriété d'équilibre est la constante d'affinité d'un complexe chimique : à l'équilibre, quelle est la probabilité de trouver les molécules dans un état lié plutôt qu'en état dissocié ?

2. Le deuxième type de propriété comprend des propriétés cinétiques ou hors d'équilibre, liées à l'évolution temporelle du système à partir d'un état initial déterminé. Elles ne reposent pas seulement sur la probabilité – statique – de trouver le système dans un certain état à l'équilibre, mais sur l'évolution de cette probabilité lorsqu'on part d'une situation hors d'équilibre. Un exemple en est la cinétique de dissociation d'un complexe : à partir d'un état hors d'équilibre (l'état lié), à quelle vitesse et par

quel·s chemin·s le système évolue-t-il vers un autre état (l'état dissocié) ?

Le pari de la MD est donc double, et repose sur l'hypothèse que la moyenne temporelle d'une quantité dépendant de la position ou de la quantité de mouvement des particules convergera aux temps longs vers une valeur d'équilibre. De telles propriétés incluent les énergies libres de liaison, les températures de fusion, entropies et enthalpies associées à diverses transformations. Contrairement aux méthodes de Monte-Carlo (MC), la MD ambitionne en outre de reproduire les propriétés cinétiques et hors d'équilibre thermodynamique des systèmes chimiques. Ces deux aspects, bien que liés, relèvent souvent d'approches distinctes lorsqu'il s'agit de calculer des quantités à partir de simulations moléculaires. Les approches à l'équilibre notamment, comme certains calculs d'énergie libre, exploitent l'évolution temporelle de la MD comme moyen efficace de générer des ensembles configurationnels dans la distribution d'intérêt. L'aspect temporel de leur acquisition n'est, dans ce cas, pas nécessaire à l'analyse. Les approches hors d'équilibre ou cinétiques s'intéressent à l'évolution dynamique et aux propriétés de transport (coefficients de diffusion, vitesses réactionnelles) et reposent intrinsèquement sur la fiabilité de l'évolution temporelle du système.

Cette force de la MD peut également être une limitation. En respectant l'évolution temporelle des systèmes modélisés, nous nous contraignons à limiter les événements dits « rares » et à investir un temps de calcul conséquent à accumuler des données redondantes dans des états métastables. Ce constat a conduit au développement de techniques sacrifiant une part de cohérence temporelle pour accélérer l'échantillonnage des ensembles thermodynamiques d'intérêt.

3.1.4 Techniques d'échantillonnage renforcé

Le principal obstacle à l'exploration de l'ensemble conformationnel des macromolécules par dynamique moléculaire est donc le temps nécessaire au franchissement des barrières énergétiques qui délimitent plusieurs états métastables. De nombreuses techniques ont été proposées afin d'accélérer ce processus au sein des simulations, afin d'obtenir plus rapidement et à moindre coût computationnel les propriétés d'équilibre des systèmes chimiques modélisés.

Un ensemble de ces méthodes repose sur l'idée d'échantillonnage préférentiel (*importance sampling*). L'idée fondamentale est de se départir de la distribution canonique, découlant de l'hamiltonien original, afin d'échantillonner un autre ensemble, qui favorise des régions de l'espace des phase ordinairement peu visitées. L'ensemble modifié peut par exemple augmenter la fréquence des événements rares en abaissant les barrières énergétiques. De telles méthodes nécessitent la

définition d'une ou plusieurs variables collectives (CV, *collective variable*), c'est-à-dire de fonctions agrégeant un ensemble de coordonnées atomiques brutes en une grandeur unidimensionnelle. La complexité de telles CV peut aller d'une simple distance interatomique à une fonction non-linéaire et complexe d'un grand nombre de coordonnées. Idéalement, les CV choisies constituent de bonnes coordonnées de réaction pour le processus d'intérêt et représentent les degrés de liberté lents du système. Une CV peut ainsi être utilisée pour biaiser l'évolution du système en faveur de régions préférentielles. En conservant rigoureusement la trace du biais introduit, il est généralement possible de corriger son influence et se ramener à l'ensemble thermodynamique de départ.

La plus connue de ces méthodes est celle de l'échantillonnage parapluie (US, *umbrella sampling*) [4], dans laquelle un potentiel de biais (le parapluie) est défini pour contenir le système dans une région limitée (« fenêtre ») de la CV. L'accumulation de plusieurs fenêtres suffisamment contigües permet de reconstruire le profil d'énergie libre le long de la coordonnée. Cette méthode a l'inconvénient de nécessiter une certaine connaissance a priori du profil énergétique que l'on souhaite calculer afin de choisir efficacement la position et la forme des potentiels de biais. Des méthodes adaptatives surmontent cette difficulté en construisant un biais de manière incrémentale, en fonction des régions déjà explorées. C'est le cas par exemple de la métadynamique (METAD) [5, 6], où l'on dépose à intervalles réguliers un petit potentiel gaussien à l'endroit où le système se trouve sur la CV. L'accumulation de ces éléments de potentiel dans les régions les plus visitées encourage le système à aller explorer des régions de plus haute énergie. Une technique similaire, la force de biais adaptative (ABF, *adaptive biasing force*) [7, 8], agit plutôt sur le gradient du potentiel, et compense progressivement la force thermodynamique moyenne éprouvée par le système en chaque point de la coordonnée.

Une autre catégorie de méthodes s'économise la définition hautement non-triviale d'une variable collective. Dans ce cas, l'objectif est généralement de modifier l'hamiltonien de façon à échantillonner une gamme d'énergies potentielles plus vaste. Certaines techniques reposent par exemple sur l'augmentation de la température (c'est-à-dire de la partie cinétique de l'hamiltonien) pour accélérer le franchissement des barrières énergétiques. La dynamique moléculaire à échange de répliques (REMD, *replica-exchange molecular dynamics*) [9] utilise la simulation en parallèle d'une échelle de répliques du système thermostatées à différentes températures. Un algorithme Monte-Carlo propose régulièrement des échanges configurationnels entre répliques, et les accepte selon un critère probabiliste (METROPOLIS-HASTINGS) permettant de respecter la distributions canoniques au sein de chaque réplique. Les méthodes à échange hamiltonien de ré-

priques (HREX, *hamiltonian replica exchange*) emploient des fonctions d'énergie potentielle $U(\mathbf{x})$ modifiées afin d'abaisser les barrières et accélérer l'échantillonnage.

Les exemples présentés ci-avant sont très loin d'offrir une vision exhaustive de la myriade de techniques d'échantillonnage renforcé proposées dans la littérature [10, 11]. Il ne semble pas exister de consensus sur la technique ultime à utiliser pour un problème donné, et le choix d'une méthode plutôt qu'une autre dépend d'un grand nombre de facteurs : question scientifique posée, type de système chimique, possibilité ou non de définir une CV pertinente, expérience du chercheur, implémentation dans les moteurs de simulation, performances computationnelles, fréquence d'utilisation dans la communauté, préférences personnelles...

Afin d'éviter une longue description indigeste et déconnectée de nos travaux, nous détaillerons au fil de l'eau les différentes techniques utilisées et la justification de leur emploi.

3.2 PRÉPARATION DU MODÈLE

Après cette présentation générale de la dynamique moléculaire comme outil d'investigation, cette section décrit la construction des modèles moléculaires protéiques qui serviront de base aux simulations suivantes.

3.2.1 État de protonation

La préparation d'une simulation de dynamique moléculaire à partir de structures biomoléculaires issues de la PDB est un procédé semi-automatique, grandement facilité par les outils fournis avec les suites logicielles spécialisées (gmx2pdb pour GROMACS, psfgen pour NAMD, etc.). Un certain nombre de décisions sont prises par défaut par les algorithmes de préparation des topologies, dans l'intention de simplifier la procédure, en se basant sur les cas les plus fréquemment rencontrés comme critère d'arbitrage. Un point crucial est celui du choix de l'état de protonation des sites titrables de la protéine. La cristallographie aux rayons X n'a généralement pas la résolution suffisante pour déterminer la position des atomes d'hydrogène, et ceux-ci doivent être ajoutés à la structure de départ. De plus, certains groupements chimiques peuvent subir des réactions acidobasiques dans les gammes de pH observées *in vivo*, le plus souvent avec l'eau environnante. En fonction du degré d'exposition d'un site titrable au solvant, ces réactions sont généralement rapides et il en résulte une répartition statistique des états de protonation, fonction de la nature chimique des groupements impliqués, du pH, de la force ionique, mais aussi de facteurs plus complexes comme l'environnement chimique local et

les interactions entre sites adjacents. En modifiant l'état acidobasique d'un groupement chimique, sont également modifiés son encombrement stérique, sa charge électrique, sa capacité à donner ou accepter des liaisons hydrogène. Bien que localisées, ces propriétés peuvent avoir un effet dramatique sur l'ensemble conformationnel adopté par la molécule.

En dehors de systèmes spécifiques nécessitant la prise en compte des sites titrables, les théoricien·nes font donc généralement le choix de l'état de protonation le plus représentatif à un pH donné. Par exemple, le programme `gmx2pdb` de la suite Gromacs protone par défaut les résidus Lys, Asp, Glu, ainsi que les extrémités N- et C-terminales en fonction de leur forme majoritaire à pH neutre, et les résidus His *via* la détection automatique d'éventuelles liaisons hydrogène suggérées par la structure cristallographique. Si ces choix sont généralement les plus pertinents pour simuler une protéine à pH neutre, les exceptions ne sont pas rares et nous devons alors recourir à des outils plus avancés de prédiction des états de protonation. Des serveurs en ligne tels que H++ [12-15] ou DelPhiPKa [16-19] permettent d'estimer plus finement les pK_a des résidus titrables grâce à des méthodes basées sur des calculs de potentiel électrostatique.

Ces nuances peuvent échapper au néophyte des simulations biomoléculaires, qui peut alors se fier à l'apparente simplicité des procédures par défaut des programmes de préparation topologique. Nous en avons fait les frais au cours de notre thèse. Notons par exemple que `pdb2gmx` assigne à l'histidine 45 de FimH un état de protonation différent en fonction de la structure PDB d'entrée et ne prédit pas le bon état de protonation tel qu'estimé par le serveur H++ à pH neutre, voir FIGURE 17 et TABLE 4.

Structure	PDB ID	pdb2gmx défaut	H++ (pK_a)
A_{cl}^{LP}	4XOE	HIE	HIP (8,7)
A_{op}^{LP}	4XOD	HIE	HIP (7,9)
S_{cl}^L	4XOC	HID	HIP (8,7)

TABLE 4 – Prédiction de l'état de protonation de His45 pour différentes structures. HID : histidine neutre protonée en δ . HIE : histidine neutre protonée en ϵ . HIP : histidine chargée positivement, diprotonée en δ et ϵ .

Nous déduisons que l'histidine 45 est majoritairement protonée aux pH neutres ou légèrement acides qui correspondent aux conditions physiologiques urinaires (l'urine normale a un pH de référence dans la gamme 5,4 à 7,2 [20]). Le pH intestinal varie de légèrement acide ($\sim 5,5$) à légèrement basique (~ 8). Il est donc possible que la déprotonation aux pH légèrement basiques de His45 ait un rôle biologique dans la modulation de l'adhérence piliale, bien que le pH

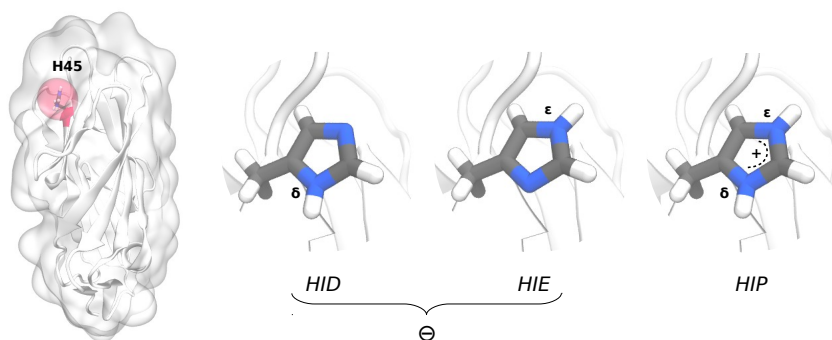


FIGURE 17 – États de protonation de l’histidine 45. (gauche) Emplacement de His45 au sein du domaine lectine. (droite) Représentation des trois états de protonation possibles aux pH physiologiques.

puisse également agir sur les propriétés du reste du pilus et notamment de l’hélice FimA [21]. En outre la différence de pK_a observée entre les états open et closed de la boucle de fermeture du site de liaison est susceptible de causer une modulation pH-dépendante de leurs populations respectives.

Ayant débuté nos travaux en produisant un certain nombre de simulations avec les formes neutres (HID, HIE) nous avons choisi d’en répliquer certaines avec la forme cationique diprotonée (HIP) plus proche de l’état physiologique attendu, faisant de notre « erreur » une source informative de comparaison. Nous prendrons soin à chaque fois de préciser l’état de protonation employé :

- par défaut, nous référerons aux systèmes diprotonés (HIP), dominant aux pH physiologiques,
- les systèmes monoprotés (HID, HIE) seront indiqués par le symbole \ominus , le signe moins faisant référence au proton manquant.

3.2.2 Simulations préliminaires et choix du champ de force

Maintenant que nous disposons de structures, il nous faut sélectionner un champ de force adapté au problème étudié. Le développement et la validation des champs de force décrivant correctement les peptides et protéines est une discipline en soi, et il n’existe à l’heure actuelle pas de consensus et peu de règles générales pour aiguiller notre choix, malgré le développement d’études de comparaison systématique [22-24]. Le problème est d’autant plus marqué lorsqu’on souhaite paramétrer des molécules non peptidiques, par exemple un ligand organique. Chaque molécule doit alors être paramétrée « à la main », à partir de calculs *ab initio*, ou de champs de forces généraux développés en complément des champs de force protéiques.

La solution la plus sûre est bien souvent d’opter pour un champ de force largement utilisé, relativement récent, et si possible d’en comparer plusieurs. C’est la démarche que nous avons adoptée en sélectionnant deux champs de force qui nous paraissaient remplir ces critères au début de notre étude : AMBER ff99SB-ILDN [25] et CHARMM36m [26], en combinaison avec leurs champs de force généralisés pour molécules organiques, GAFF (*general AMBER force field*) et CGenFF (*CHARMM general force field*).

Méthode

Les structures du domaine lectine lié au heptyl- α -D-mannose (HM) dans les conformations $S_{cl}^{L\ominus}$ (PDB: 4xoc) et $A_{cl}^{L\ominus}$ (PDB: 4xoe) ont été préparées avec la suite GROMACS 2018. Pour chaque simulation, le complexe lectine-HM est solvaté dans une boîte de simulation cubique (10 nm) avec conditions périodiques aux limites. Des ions Na^+ et Cl^- sont ajoutés de façon à neutraliser le système et atteindre une concentration d’environ 50 mM. Le système est équilibré à 310 K et 1 bar grâce à une courte simulation (200 ps) où les positions atomiques du soluté sont restreintes. Deux thermostats de Berendsen modifiés [27] sont utilisés pour le soluté (protéine-HM) d’une part et le solvant (eau-ions) d’autre part afin d’éviter les artefacts de « solvant chaud/soluté froid » (“*hot-solvent/cold-solute*”) pouvant advenir dans les simulations [28].

Le champ de force AMBER ff99SB-ILDN fourni nativement avec GROMACS a été utilisé. La topologie du ligand HM pour le champ de force GAFF [29] a été préparée grâce à l’utilitaire ACPYPE [30]. Les fichiers du portage du champ de force CHARMM36m (C36m) pour GROMACS (nov. 2018) ont été téléchargés sur le site du MacKerell Lab [31]. Les fichiers de topologie du ligand HM ont été préparés grâce au serveur en ligne [32] de CGenFF (version 4.1) et convertis avec le script `cgenff_charmm2gmx.py`.

Notre approche initiale a consisté à produire jusqu’à 1 μ s de simulation de dynamique moléculaire pour chaque conformation cristallographique de départ et chaque champ de force, afin de tester la stabilité du modèle.

Résultats

La stabilité conformationnelle du complexe protéine-ligand est évaluée par inspection visuelle des trajectoires. Les deux simulations avec C36m/CGenFF donnent lieu à une dissociation très rapide du complexe (TABLE 5). Par contraste, les deux systèmes utilisant ff99SB-ILDN/GAFF montrent des complexes toujours intacts après 1 μ s de simulation. Cette rapidité de dissociation pour le champ de force

Structure	Champ de force protéine	Champ de force ligand	Durée totale [ns]	Temps de dissociation [ns]	RMSD final [Å]
$S_{cl}^{L\ominus}$	ff99SB-ILDN	GAFF	1000	-	2,1
$A_{cl}^{L\ominus}$	ff99SB-ILDN	GAFF	1000	-	2,1
$S_{cl}^{L\ominus}$	C36m	CGenFF	400	198	2,4
$A_{cl}^{L\ominus}$	C36m	CGenFF	400	298	3,2
$A_{cl}^{L\ominus}$	ff99SB-ILDN	GAFF	400	-	2,1
$A_{cl}^{L\ominus}$	C36m	CGenFF	400	271	2,8

TABLE 5 – Stabilité du complexe pour différents champs de force. Le RMSD final des carbones alpha par rapport à la structure initiales est indiqué.

C36m est très étonnante au regard de la demi-vie du complexe estimée expérimentalement, de l'ordre de la dizaine de millisecondes pour l'état de basse affinité et jusqu'à plusieurs minutes pour l'état de haute affinité (voir TABLE 3, p. 45). Afin de mieux comprendre la provenance de la différence observée, nous avons tenté de marier ff99SB-ILDN au champ de force pour ligand destiné à C36m, et réciproquement. Ces appariements hétérodoxes tendent à montrer que c'est principalement le champ de force protéique qui détermine l'instabilité du complexe (TABLE 5).

L'inspection des trajectoires suggère une flexibilité beaucoup plus élevée de la structure protéique pour C36m que pour ff99SB-ILDN, comme en témoigne le RMSD final plus grand. Nous confirmons cette intuition en calculant le RMSF par résidu des carbones α pour chaque trajectoire (FIGURE 18). La différence est particulièrement marquée dans la zone de la boucle de fermeture (rés. 8–16), ce qui pourrait expliquer l'instabilité plus grande de la liaison au HM.

Cette divergence entre deux champs de force parmi les plus utilisés soulève des questions vis-à-vis de la fiabilité et l'interprétabilité des résultats de simulations moléculaires. Le champ de force CHARMM36m est une évolution de CHARMM36 spécifiquement destinée à améliorer la description des peptides et protéines intrinsèquement désordonnées tout en préservant celle des protéines structurées [26], et il est possible que cet objectif impacte défavorablement la stabilité de complexes protéine-ligand structurellement rigides.

Dans la suite de nos travaux, nous avons choisi de persévérer avec le champ de force a99SB-ILDN, qui montre une stabilité adéquate du complexe FimH-HM. Une investigation plus poussée des différences observées avec la famille des champs de force CHARMM (27, 36 ou 36m), bien qu'instructive à bien des égards, n'est pas l'objectif de nos travaux.

Le RMSD est la distance moyenne des atomes par rapport à une structure de référence, après alignement roto-translationnel optimal. Le RMSF est le RMSD moyen par rapport à une structure moyennée dans le temps.

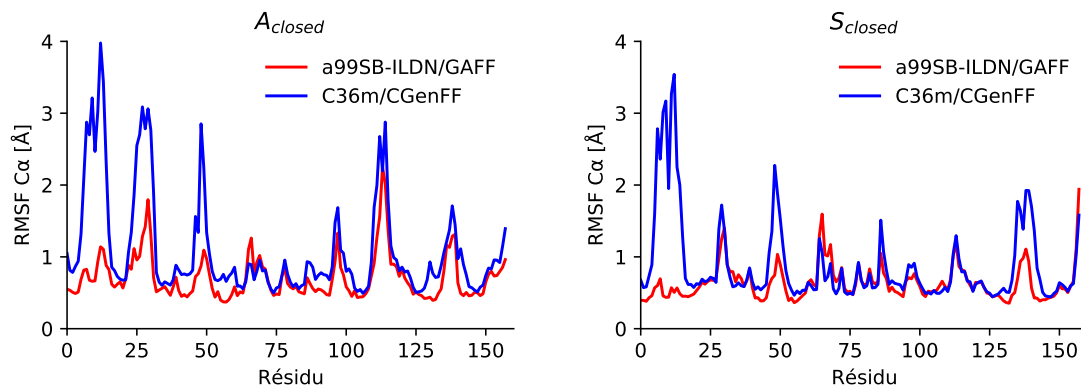


FIGURE 18 – Flexibilité du domaine lectine avec différents champs de force. Le RMSF par résidu est calculé comme la moyenne du RMSD de chaque C α après alignement roto-translationnel de l'ensemble des C α sur la structure moyenne, pour les conformations de départ $A_{cl}^{L\ominus}$ et $S_{cl}^{L\ominus}$ et avec deux champs de forces. Les premières 400 ns de chaque simulation sont utilisées. Notons une flexibilité plus marquée de CHARMM36m, en particulier dans la zone de la boucle de fermeture (rés. 8–16), et pour $A_{cl}^{L\ominus}$ la boucle pivot (rés. 25–34).

3.3 EXPLORATION DU PAYSAGE CONFORMATIONNEL

Guidés par les changements conformationnels repérés durant la simulation μ s, et notamment l'obtention d'un état partiellement ouvert au niveau de la boucle fermant le site de liaison, nous décidons d'approfondir l'exploration du paysage conformationnel du domaine lectine. Notre objectif est multiple :

- évaluer comment l'ouverture du site actif est allostériquement modulée,
- comprendre l'effet de la présence du ligand sur les conformations accessibles à la protéine,
- tenter d'observer une transition entre les états A et S du domaine lectine et ses éventuels intermédiaires.

La dynamique moléculaire traditionnelle ne nous permet d'accéder qu'à des changements conformationnels prenant place sur des échelles de temps relativement courtes. À titre d'exemple, sur un nœud de calcul moderne bénéficiant d'une accélération GPU, il est possible de produire environ 100 ns de trajectoire par jour pour une protéine de la taille de FimH. Cela signifie que seuls les phénomènes prenant place à l'échelle de la microseconde tout au plus peuvent être examinés de manière satisfaisante par MD sans recourir à des techniques avancées.

Comme mentionné en 3.1.4, plusieurs techniques d'échantillonnage renforcé s'offrent à nous pour explorer l'espace conformationnel de FimH. Dans un premier temps, nous nous tournons vers une méthode dépourvue de toute définition de variable collective (CV) afin

d'accélérer sans à priori les degrés de libertés de la protéine : la dynamique moléculaire à échange hamiltonien de répliques, dans sa version REST₂ (*replica exchange with solute scaling*), que nous allons présenter ci-après.

3.3.1 Dynamique moléculaire à échange hamiltonien de répliques

Les techniques d'échantillonnage à échange de répliques sont un ensemble de méthodes de Monte-Carlo reposant sur la simulation en parallèle de plusieurs systèmes, représentant différents ensembles thermodynamiques, et pouvant échanger aléatoirement leurs configurations tout en conservant leurs distributions respectives. L'objectif est de simuler simultanément l'ensemble d'intérêt et d'autres ensembles (qui peuvent être physiquement non-réalistes), dont le rôle est de favoriser et d'accélérer l'exploration de nouvelles configurations.

Échange de répliques en température

La première version historique de cette technique est l'échange de répliques en température (TREM, *temperature replica exchange method*), aussi nommé *parallel tempering*. D'abord développée pour des approches purement MC de physique statistique [33, 34], SUGITA et OKAMOTO en proposent une version adaptée à la MD [9] : l'ensemble d'intérêt est alors l'ensemble canonique d'un système moléculaire à une température trop basse pour permettre des changements conformationnels rapides, et les autres ensembles sont thermostatés à des températures supérieures, dont les distribution diffèrent de celle à l'étude mais facilitent le franchissement des barrières énergétiques. Des échanges configurationnels sont proposés à intervalles réguliers entre deux répliques i et j , et sont acceptés avec une certaine probabilité p_{acc} afin de préserver la convergence vers une distribution canonique dans chacune des répliques,

$$p_{\text{acc}} = \begin{cases} 1 & \text{si } \Delta_{ij} \leq 0 \\ \exp[-\Delta_{ij}] & \text{si } \Delta_{ij} > 0 \end{cases} \quad (14)$$

avec

$$\Delta_{ij} = (\beta_j - \beta_i)[U(\mathbf{x}_j) - U(\mathbf{x}_i)] \quad (15)$$

où $\beta_k = 1/(k_B T_k)$ est la température inverse de la réplique k , U la fonction d'énergie potentielle (commune à toutes les répliques) et \mathbf{x}_k la configuration de la réplique k . On voit alors que si $T_j > T_i$, l'échange est accepté systématiquement si $U(\mathbf{x}_j) \leq U(\mathbf{x}_i)$. Dans le cas contraire, l'échange est possible mais avec une probabilité qui décroît

exponentiellement à mesure que la différence d'énergie potentielle augmente.

L'efficacité de la méthode dépend alors du bon taux d'échange entre les paires de répliques, ce qui nécessite un recouvrement significatif de leurs distributions énergétiques, et donc des températures suffisamment proches. Cela peut nécessiter un grand nombre de répliques intermédiaires pour atteindre une température maximale donnée qui garantit le franchissement aisé des barrières énergétiques limitantes. En effet, la différence d'énergie entre deux répliques croît avec la racine du nombre de particules ($\Delta_{ij} \propto \sqrt{N}$), et pour des systèmes comportant un grand nombre d'atomes – souvent dominé par les molécules de solvant – le coût calculatoire peut rapidement devenir prohibitif.

Échange hamiltonien de répliques REST2

Une approche plus économe repose sur l'utilisation de surfaces d'énergie potentielle modifiées, qui abaissent les barrières énergétiques plutôt que d'augmenter l'énergie cinétique *via* la température. C'est le cas dans la méthode REST2 [35] où les termes d'énergie potentielle de torsion et électrostatiques du champ de force sont mis à l'échelle. L'avantage de cette méthode est qu'elle permet une plus grande souplesse sur les degrés de liberté qui nécessitent une accélération, et en particulier d'exclure le solvant de toute mise à l'échelle. L'hypothèse sous-jacente est alors que les degrés de liberté du solvant sont rapides et non-limitants, contrairement aux degrés de libertés internes du soluté. Ce faisant, le nombre de répliques nécessaire à un bon taux d'échange ne dépend plus que de la taille de la molécule d'intérêt (par exemple une protéine ou un acide nucléique) et peut être considérablement réduit.

L'algorithme peut se résumer comme suit. Soient N répliques (indexées $0 \dots N-1$) du même système (de mêmes topologies et températures T_0) dont l'énergie potentielle peut se décomposer en trois termes

$$U(\mathbf{x}_p, \mathbf{x}_s) = U_{pp}(\mathbf{x}_p) + U_{ps}(\mathbf{x}_p, \mathbf{x}_s) + U_{ss}(\mathbf{x}_s) \quad (16)$$

où \mathbf{x}_p et \mathbf{x}_s désignent les coordonnées de la protéine et du solvant respectivement, U_{pp} l'énergie d'interaction interne à la protéine, U_{ss} l'énergie d'interaction interne au solvant, et U_{ps} l'énergie d'interaction protéine-solvant. Le schéma REST2 propose de mettre à l'échelle l'énergie potentielle de la réplique k

$$U_k = \lambda_k U_{pp} + \sqrt{\lambda_k} U_{ps} + U_{ss} \quad (17)$$

où λ est un facteur de mise à l'échelle pris entre 0 et 1, avec $\lambda_0 = 1$. La probabilité d'acceptation selon le critère de l'équation 14 dépend dorénavant de

$$\begin{aligned}\Delta_{ij} &= \beta_0[U_i(\mathbf{x}_j) + U_j(\mathbf{x}_i) - U_i(\mathbf{x}_i) + U_j(\mathbf{x}_j)] \\ &= \beta_0(\lambda_j - \lambda_i)[U_{pp}(\mathbf{x}_i) - U_{pp}(\mathbf{x}_j)] \\ &\quad + \beta_0(\sqrt{\lambda_j} - \sqrt{\lambda_i})[U_{ps}(\mathbf{x}_i) - U_{ps}(\mathbf{x}_j)].\end{aligned}\tag{18}$$

Nous voyons ainsi que le terme d'interaction solvant-solvant U_{ss} ne contribue pas à la différence d'énergie généralisée Δ_{ij} , ce qui permet de réduire considérablement sa variance et accroître le taux d'échange.

Le facteur de mise à l'échelle λ_k permet de définir une température inverse $\beta_k = \lambda_k \beta_0$, que l'on peut interpréter comme la température inverse effective à laquelle est chauffé le soluté de la réplique k . En effet, pour une distribution canonique il est équivalent de multiplier la température ou de diviser uniformément l'énergie potentielle par un même facteur. Dans le cas présent la mise à l'échelle du potentiel n'est pas uniforme sur tous les degrés de liberté et il n'y a pas d'équivalence stricte, mais la température effective offre une correspondance approximative avec la TREM classique [36].

Le choix des λ_k dépend du nombre de répliques, de la température effective maximale souhaitée T_{\max} , de la température de base T_0 et de la taille du système inclus dans la mise-à-l'échelle. Afin d'obtenir une probabilité d'échange homogène entre répliques adjacentes, l'espace-ment des températures doit suivre une suite géométrique (montrée comme optimale pour un système à capacité calorifique C_v constante [37]) :

$$T_k = T_0 \left(\frac{T_{\max}}{T_0} \right)^{\frac{k}{N-1}} \quad \text{avec } \lambda_k = T_0/T_k,\tag{19}$$

pour $k \in \{0 \dots N-1\}$. Le choix de N et/ou T_{\max} doit s'effectuer de façon à garantir un bon taux d'échange entre deux répliques adjacentes sur l'échelle de températures, idéalement autour de 20 % [38]. Ce choix peut s'effectuer par essai-erreur et en s'inspirant d'autres simulations, le nombre de répliques nécessaires pour assurer un certain taux d'échange croissant avec \sqrt{f} où f est le nombre de degrés de libertés du soluté [35].

3.3.2 Méthodes

Nous appliquons la méthode REST2 à plusieurs systèmes :

- le domaine lectine isolé (L) dans sa forme séparée (S), supposé représentatif de l'état « haute affinité » (PDB : 4xoc [39], voir chapitre 2 FIGURE 11A);

- la protéine FimH entière (LP) stabilisée par le brin donneur DsG, avec le domaine lectine dans sa forme associée (A), supposé représentatif de l'état de « moyenne » ou « basse affinité » (PDB : 4XOE [39], voir FIGURE 11C)
- le domaine lectine isolé (L) tel qu'extraît *in silico* de la structure précédente, sous sa forme A.

Seul le domaine lectine (rés. 1–156) est affecté par la mise-à-l'échelle de REST2. L'eau, les ions, et éventuellement le ligand ou le domaine piline demeurent donc « froids ».

Chaque système est simulé en présence (●) et en l'absence (○) du ligand heptyl- α -D-mannose (HM) complexé dans le site actif. Les systèmes sans ligand sont préparés à partir des structures ligantées par suppression *in silico* du ligand, et leur site de liaison est donc toujours en conformation fermée (à l'exception de la simulation XI où la structure cristallographique apo, PDB : 4XOD [39], voir FIGURE 11B, a été utilisée).

Pour finir, les 3×2 simulations précédentes ont été effectuées avec l'histidine 45 diprotonée (HIP, simulations I–VI) ou monoprotée (HID ou HIE, simulations VII–XII, notées \ominus pour symboliser le proton en défaut). La table 6 récapitule l'ensemble des simulations REST2 effectuées.

N°	PDB	Domaines	Ligand	A/S	Site	His45	Notation	Durée [ns]	N _{rep} *
I	4XOC	L	oui	S	fermé	HIP	●S _{cl} ^L	400	20
II	4XOE	L	oui	A	fermé	HIP	●A _{cl} ^L	400	20
III	4XOE	LP	oui	A	fermé	HIP	●A _{cl} ^{LP}	400	20
IV	4XOC	L	non	S	fermé	HIP	○S _{cl} ^L	400	20
V	4XOE	L	non	A	fermé	HIP	○A _{cl} ^L	400	20
VI	4XOE	LP	non	A	fermé	HIP	○A _{cl} ^{LP}	400	20
VII	4XOC	L	oui	S	fermé	HID	●S _{cl} ^{L\ominus}	500	24
VIII	4XOE	L	oui	A	fermé	HID	●A _{cl} ^{L\ominus}	500	24
IX	4XOE	LP	oui	A	fermé	HIE	●A _{cl} ^{LP\ominus}	500	24
X	4XOC	L	non	S	fermé	HID	○S _{cl} ^{L\ominus}	500	20
XI	4XOD	L	non	A	ouvert	HIE	○A _{op} ^{L\ominus}	500	20
XII	4XOE	LP	non	A	fermé	HIE	○A _{cl} ^{LP\ominus}	500	24

TABLE 6 – Simulations REST2.

*Le nombre de répliques peut varier afin de s'adapter à l'architecture matérielle des supercalculateurs utilisés.

Les systèmes ont été préparés et équilibrés en suivant la même procédure que décrite en 3.2.2. La méthode REST2 a été mise en place grâce au module PLUMED 2.5 [40, 41]. Les fichiers de topologie modifiés pour les répliques mises à l'échelle ont été générés avec le l'utilitaire `partial_tempering` de PLUMED. L'échelle des λ_k est construite

via l'équation 19 afin d'obtenir une gamme de températures effectives entre $T_0 = 300$ K et $T_{\max} = 600$ K. Les tentatives d'échange entre répliques adjacentes ont lieu toutes les 2 ps, en alternant entre couples pairs (0-1, 2-3, ...) et impairs (1-2, 3-4, ...). Cette fréquence est choisie pour favoriser une diffusion rapide dans l'espace des répliques tout en évitant les artefacts possibles pour des intervalles d'échange trop courts [43]. Les configurations de la réplique non-modifiée ($T_0 = 300$ K) sont sauvegardées toutes les 20 ps pour les analyses subséquentes.

Retenue du ligand

En accélérant les degrés de liberté internes de la protéine et en affaiblissant les interactions protéine-ligand, la méthode REST2 risque de précipiter la dissociation du complexe. Si cela se produit, le ligand s'échappe du site de liaison et diffuse librement dans la boîte de simulation, le système échantillonnant alors l'état non-lié. Une telle situation est indésirable si l'on souhaite étudier et comparer les paysages conformationnels dans les états lié et non-lié, car elle rend la convergence des ensembles conformationnels de chaque état tributaire de la dynamique d'association-dissociation du complexe. Il est alors préférable de conduire deux simulations, l'une dans l'état non-lié (sans ligand) et l'autre dans l'état lié, maintenu tel quel par des contraintes, sous la forme d'un potentiel additionnel U_{restr} .

Une approche simple permettant de restreindre les degrés de liberté roto-, trans- et conformationnels d'un même coup a été proposée par SALARI et al. [44]. Elle repose sur une variable collective nommée « distance à la configuration liée » (DBC, *distance from bound configuration*), définie comme le RMSD des atomes du ligand par rapport à une configuration de référence, dans un référentiel attaché au site de liaison. Cette CV représente la déviation de la position du ligand par rapport à sa position de référence après soustraction des mouvements de corps solide du site de liaison. Soient \mathbf{x}_R et \mathbf{x}_L les vecteurs de coordonnées associées aux atomes représentatifs du récepteur et du ligand, et $\mathbf{x}_R^{\text{ref}}$ et $\mathbf{x}_L^{\text{ref}}$ ces vecteurs dans une structure de référence. Notons $\bar{\mathbf{x}}_R$ le centre de masse du jeu de coordonnées \mathbf{x}_R . Après un certain temps, le récepteur aura subi (i) une translation globale $\bar{\mathbf{x}}_R - \bar{\mathbf{x}}_R^{\text{ref}}$ de son centre de masse combinée à (ii) une rotation globale \mathbf{R} autour de son centre de masse et (iii) des changements conformationnels internes. Les coordonnées \mathbf{x}'_i d'un atome i corrigées de la diffusion roto-translationnelle du site de liaison s'écrivent

$$\mathbf{x}'_i = \mathbf{R}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_R) + \bar{\mathbf{x}}_R^{\text{ref}} \quad (20)$$

D'autres schémas d'échange comme l'échantillonnage de GIBBS [42] autorisent simultanément des échanges entre toutes les répliques et pourraient accélérer la convergence mais ne sont pas encore implémentés.

et la DBC est définie comme

$$\delta = \sqrt{\frac{1}{n_L} \sum_{l=1}^{n_L} (\mathbf{x}'_l - \mathbf{x}_l^{\text{ref}})^2} \quad (21)$$

où la somme s'effectue sur les n_L atomes représentatifs du ligand.

La DBC permet de définir un potentiel restreignant le ligand dans une configuration proche de celle liée, de référence. Afin de ne pas perturber la dynamique du ligand au sein-même de l'état lié il est possible d'utiliser un potentiel à fond plat (*flat-well potential*), par exemple semi-harmonique :

$$U_{\text{restr}}(\delta) = \begin{cases} 0 & \text{si } \delta \leq \delta_{\text{max}} \\ \frac{1}{2} \kappa_{\delta} (\delta - \delta_{\text{max}})^2 & \text{si } \delta > \delta_{\text{max}} \end{cases} \quad (22)$$

En choisissant δ_{max} judicieusement on peut laisser suffisamment d'espace au ligand pour fluctuer naturellement dans l'état lié. De plus, pour une constante κ_{δ} suffisamment élevée, le potentiel se comporte comme un « mur » sur lequel le système rebondit lorsque le ligand est sur le point de s'évader.

DÉFINITION DU SITE DE LIAISON : Les carbones α des résidus Phe1, Asn46, Asp54, Tyr95, Asn135 sont choisis pour définir le référentiel du site de liaison. Ces atomes forment un ensemble relativement rigide autour du mannoside lié (la boucle de fermeture, plus mobile, est exclue à dessein).

DÉFINITION DU LIGAND : Pour les atomes représentatifs du ligand HM nous sélectionnons les carbones et oxygènes du cycle manno-pyranose (C1–6, O1–6). Nous excluons la chaîne heptyle en raison de sa flexibilité.

STRUCTURE DE RÉFÉRENCE : La structure cristallographique de FimH·DsG complexé au HM (PDB : 4XOE), soit l'état $\bullet A_{\text{cl}}^{\text{LP}}$, est utilisée.

POTENTIEL DE RETENUE : La largeur δ_{max} nécessaire pour accommoder les fluctuations du ligand dans l'état lié est déterminée grâce aux simulations de MD préliminaires des systèmes $\bullet A_{\text{cl}}^{\text{L}\ominus}$ et $\bullet S_{\text{cl}}^{\text{L}\ominus}$. La distribution de la valeur δ , remarquablement similaire dans les deux états, est négligeable au delà de 2 Å (FIGURE 19). Afin de laisser une confortable marge et autoriser l'exploration de configurations périphériques nous choisissons une valeur $\delta_{\text{max}} = 2,5 \text{ Å}$. La constante harmonique κ_{δ} est fixée à $500 \text{ kJmol}^{-1} \text{ Å}^{-2}$.

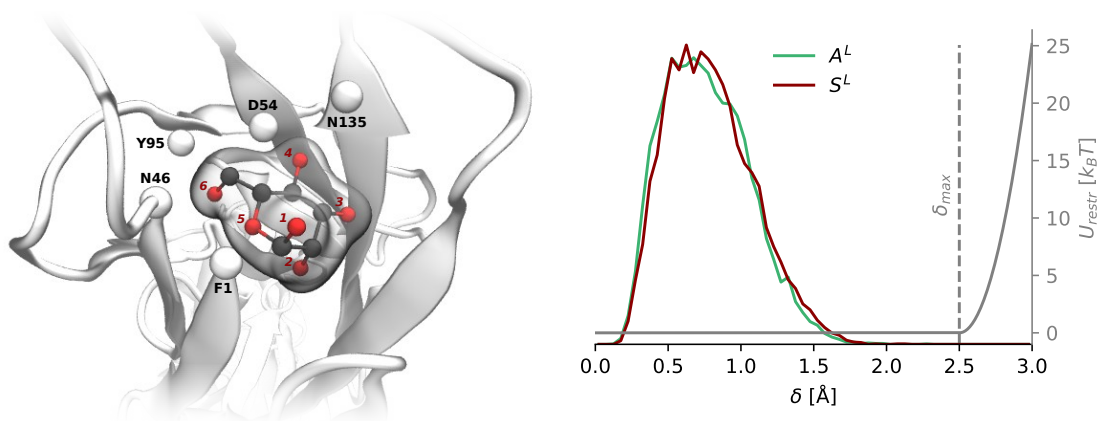


FIGURE 19 – Définition de la DBC. (*gauche*) Illustration de la coordonnée DBC. Des carbones alpha du site de liaison (sphères blanches) définissent le référentiel au sein duquel les atomes du cycle mannopyranoside peuvent fluctuer autour d'une position de référence, dans un hypervolume illustré par la surface translucide. (*droite*) La distribution de la DBC est calculée dans les simulations de MD préliminaires des systèmes $\bullet A_{cl}^{L\ominus}$ et $\bullet S_{cl}^{L\ominus}$. Une valeur δ_{\max} de 2,5 Å est sélectionnée pour la largeur du potentiel à fond plat U_{restr} utilisé dans les simulations suivantes (REST2).

La DBC et le potentiel de retenue sont implémentés grâce au module Colvars [45] récemment porté sur GROMACS. La FIGURE 19 illustre la définition de la DBC dans la configuration de référence.

Restriction de la diffusion rotationnelle

Les systèmes comportant deux domaines (LP) présentent l'inconvénient d'avoir une forme allongée peu adaptée à une boîte de simulation cubique. En effet, afin d'éviter toute interaction de la protéine avec elle-même à travers les conditions périodiques aux limites, il est nécessaire que son grand axe soit plus petit que la largeur de la boîte dans chaque direction, avec une marge d'au moins deux fois la distance de coupure des interactions à courte portée. Cela peut conduire à une boîte cubique comportant un volume de solvant disproportionné.

La solution est d'utiliser une boîte parallélépipédique adaptée à la forme oblongue du soluté tout en restreignant la diffusion rotationnelle de celui-ci [46]. Les carbones α du système sont utilisés pour définir une CV d'orientation. Il s'agit de la composante rotationnelle d'un alignement optimal minimisant le RMSD entre la structure et une structure de référence. Cette rotation exprimée sous la forme d'un quaternion $\mathbf{q} = (q_1, q_2, q_3, q_4)$ peut être contrainte *via* un potentiel de biais harmonique centré sur la rotation identité $\mathbf{q}^I = (1, 0, 0, 0)$:

$$U_{\text{rot}}(\mathbf{q}) = \frac{1}{2} \kappa_{\text{rot}} \sum_i (q_i - q_i^I)^2 \quad (23)$$

où κ_{rot} est une constante harmonique, que nous fixons à 500 kJ mol^{-1} . Cette approche nous permet de limiter la taille de la boîte de simulation à environ $12,5 \times 7,5 \times 7,5 \text{ nm}^3$ et ainsi réduire le volume de solvant simulé.

De manière importante, les degrés de libertés internes à la protéine ne sont pas affectés. Dans nos simulations le potentiel U_{rot} est implémenté grâce à la composante orientation du module Colvars [47].

3.3.3 Résultats

Variables collectives d'analyse

Afin de visualiser et interpréter les ensembles conformationnels générés par l'échantillonnage renforcé, il est utile de définir des variables collectives (CV) sur lesquelles projeter l'espace de très haute dimensionnalité des configurations (3 coordonnées cartésiennes pour chaque atome). La réduction de dimensionnalité et la recherche de CV pertinentes est un problème ouvert et central dans l'analyse des trajectoires de systèmes moléculaires [48]. Nous prenons le parti de définir deux CV à partir de notre connaissance du système et des processus sur lesquels se focalise notre enquête : la dynamique d'ouverture-fermeture du site de liaison et la transition entre les états A et S.

COORDONNÉE D'OUVERTURE : Afin de caractériser une coordonnée d'ouverture robuste aux autres degrés de libertés, nous sélectionnons un ensemble de n_ω carbones alpha appartenant à la boucle de fermeture et au sandwich bêta, en prenant soin d'exclure la région inter-domaines, le renflement bêta, le switch alpha ainsi que les boucles flexibles du portail à tyrosines. La CV d'ouverture baptisée ω (*omega* pour *ouverture*, *opening*) est la projection des positions atomiques sur la composante principale d'ouverture, définie par le vecteur après alignement entre deux conformations extrêmes : S_{cl} et A_{op} . Plus précisément, soient \mathbf{x}_{cl} et \mathbf{x}_{op} les vecteurs des $3 \times n_\omega$ coordonnées dans les deux états de référence. Soit $T_{\text{cl}}(\mathbf{x})$ la superposition rototranslationnelle optimale d'une configuration \mathbf{x} sur \mathbf{x}_{cl} . On note $\Delta\mathbf{x}_{\text{cl} \rightarrow \text{op}} = T_{\text{cl}}(\mathbf{x}_{\text{op}}) - \mathbf{x}_{\text{cl}}$ le vecteur principal d'ouverture. De même on peut écrire le vecteur de déplacement par rapport à \mathbf{x}_{cl} à un instant t : $\Delta\mathbf{x}_{\text{cl} \rightarrow t} = T_{\text{cl}}(\mathbf{x}_t) - \mathbf{x}_{\text{cl}}$. La coordonnée d'ouverture à un instant t est alors la projection de $\Delta\mathbf{x}_{\text{cl} \rightarrow t}$ sur $\Delta\mathbf{x}_{\text{cl} \rightarrow \text{op}}$:

$$\omega_t = \frac{\Delta\mathbf{x}_{\text{cl} \rightarrow t} \cdot \Delta\mathbf{x}_{\text{cl} \rightarrow \text{op}}}{(\Delta\mathbf{x}_{\text{cl} \rightarrow \text{op}})^2} \quad (24)$$

où la normalisation permet d'échelonner les valeurs entre 0 (closed) et 1 (open).

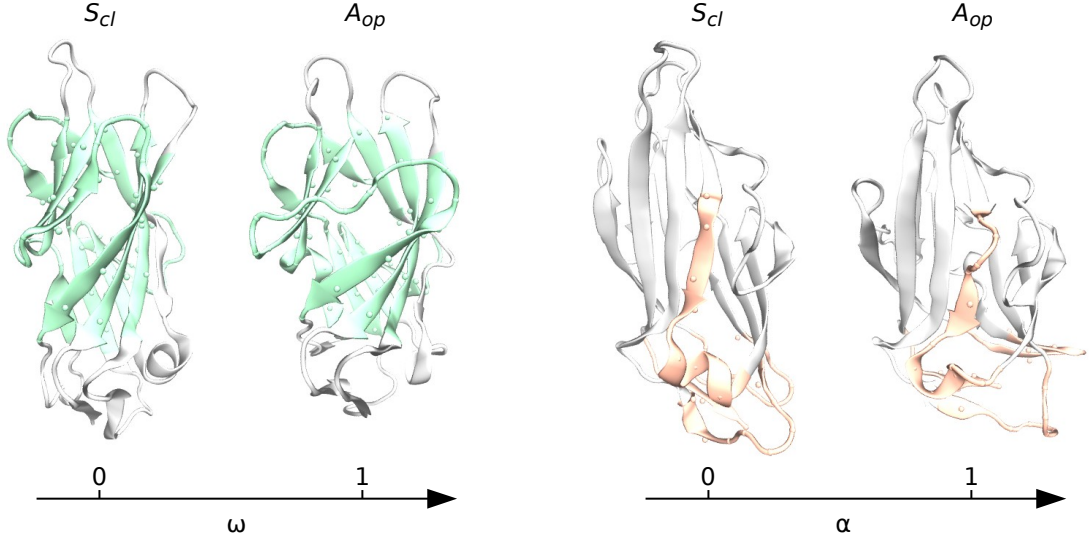


FIGURE 20 – Variables collectives d’ouverture et d’association. Les carbones alpha entrant dans la définition des deux coordonnées sont représentés, et les résidus correspondants sont colorés. (*gauche*) Coordonnée d’ouverture ω . (*droite*) Coordonnée allostérique ou d’association α .

COORDONNÉE ALLOSTÉRIQUE : L’autre phénomène attirant notre attention est la transition éventuelle entre les états A et S. Les carbones alpha des boucles inter-domaines, du renflement bêta et du switch alpha sont utilisés pour définir une CV construite sur le même principe, baptisée α (alpha pour association ou allostérie) :

$$\alpha_t = \frac{\Delta \mathbf{x}_{S \rightarrow t} \cdot \Delta \mathbf{x}_{S \rightarrow A}}{(\Delta \mathbf{x}_{S \rightarrow A})^2}. \quad (25)$$

Les deux coordonnées sont illustrées en FIGURE 20.

Définition de l’état lié

Grâce à l’utilisation de la contrainte DBC nous avons limité l’exploration des systèmes avec ligand aux configurations proches de l’état lié cristallographique. La FIGURE 21 présente les distributions de la DBC pour les trois systèmes ligantés $\bullet A^L$, $\bullet A^{LP}$ et $\bullet S^L$. Le choix d’un mur de potentiel relativement large autorise le ligand à visiter des régions périphériques ($\delta > 2 \text{ \AA}$), non observées dans les simulations préliminaires. Cette tendance est particulièrement marquée pour l’état « basse affinité » $\bullet A^{LP}$, suggérant comme attendu une plus faible stabilité du complexe.

L’observation des trajectoires indique que les poses périphériques sont variables d’une simulation ou d’une réplique à l’autre, et qu’elles forment peu d’interactions spécifiques avec le site de liaison. Bien que

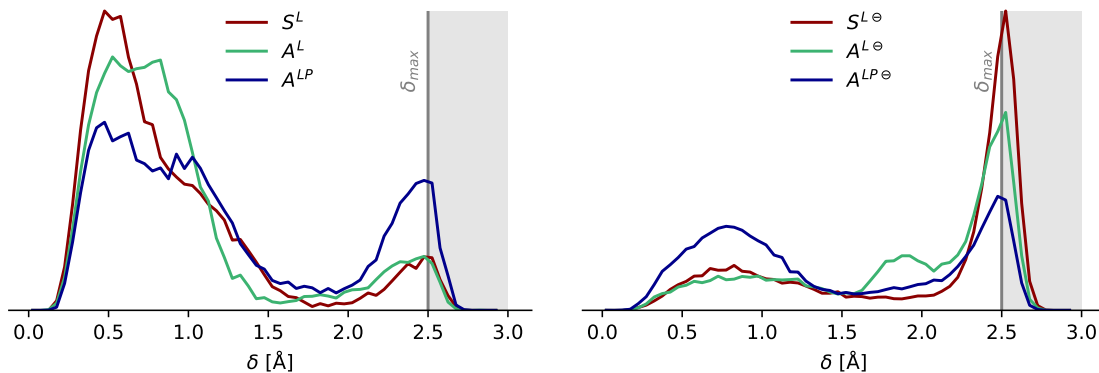


FIGURE 21 – Distributions de la DBC dans les simulations REST2. (*gauche*) Systèmes avec histidine 45 diprotonée. (*droite*) Systèmes avec histidine 45 monoprotonée. On constate le peuplement de configurations s'éloignant de la pose de référence ($\delta = 0 \text{ \AA}$).

les poses périphériques observées puissent représenter des états métastables sur le chemin réactionnel de complexation, il est également possible que les poses se concentrant au niveau de la barrière du potentiel de restrainte soient artéfactuelles et ne représentent pas des états biologiques pertinents. Nous faisons donc pour la suite le choix d'une définition stricte de l'état lié où le ligand maintient l'essentiel des interactions spécifiques (notamment les liaisons hydrogène) visibles dans la pose cristallographique.

Les distributions présentent chacune un minimum entre 1,5 et 2 \AA , mais il est difficile de singulariser une valeur précise permettant de définir la frontière de l'état lié. Dans cet objectif nous définissons une seconde variable collective venant compléter la DBC : la somme des distances entre l'azote N-terminal (résidu Phe1) et les deux oxygènes du mannose avec lesquels il forme des liaisons hydrogène, O2 et O6, notées respectivement $d_{N,O2}$ et $d_{N,O6}$. La projection des trajectoires sur l'espace à deux dimensions formé par δ et $d_{N,O2} + d_{N,O6}$ (FIGURE 22) permet de mieux séparer le bassin énergétique de l'état lié des états périphériques, et ce de manière robuste pour l'ensemble des systèmes étudiés. Le critère final d'appartenance à l'état lié correspond à la fonction indicatrice $\mathbb{1}$:

Les figures pour les autres systèmes sont en annexe B, FIGURE B.1.

$$\mathbb{1}(\delta, d_{N,O2}, d_{N,O6}) = \begin{cases} 1 & \text{si } \delta < 1,5 \text{ \AA} \text{ et } d_{N,O2} + d_{N,O6} < 7 \text{ \AA} \\ 0 & \text{sinon.} \end{cases} \quad (26)$$

Notons que le seuil à 7 \AA correspond à deux fois celui qui est généralement employé pour une liaison hydrogène.

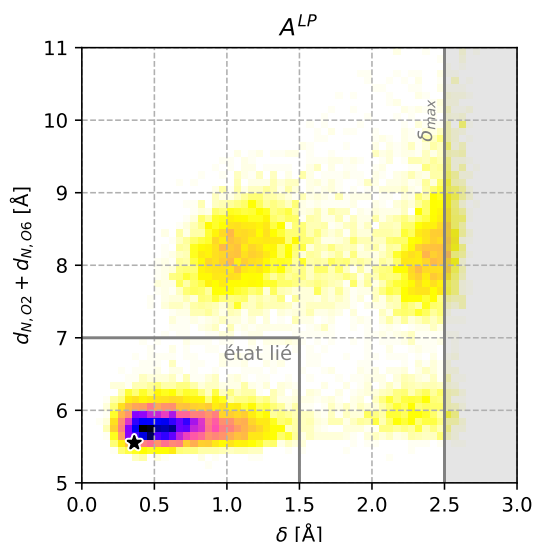


FIGURE 22 – Définition de l'état lié. La distribution jointe formée par δ et $d_{N,O2} + d_{N,O6}$ dans la trajectoire REST₂ (ici $\bullet A^{LP}$) montre une séparation nette du bassin conformationnel correspondant à l'état lié (*en bas à gauche*) des autres configurations explorées. L'histogramme bidimensionnel est coloré par densité de probabilité croissante du blanc au noir. La configuration initiale est dénotée par une étoile (*).

Ensemble conformationnel

Afin de visualiser la diversité conformationnelle explorée dans les simulations REST₂, nous projetons les trajectoires sur l'espace joint des CV d'ouverture (ω) et d'association (α). La FIGURE 23 présente les régions explorées par les trois systèmes (A^L , S^L , A^{LP}) superposées par catégorie (lié ou non, état de protonation). Outre les bassins conformationnels proches des structures cristallographiques de départ, chaque système explore des états comportant divers degrés d'ouverture du site de liaison. Cela est particulièrement surprenant pour les systèmes de configuration initiale S_{cl}^L , qui visitent des conformations ouvertes (S_{op}^L) encore jamais observées par cristallographie.

Les systèmes A^{LP} , où les deux domaines sont associés, ne montrent pas de signe de séparation et restent stables autour d'une valeur $\alpha = 1$, indiquant une certaine rigidité des boucles inter-domaines et du renflement bêta/switch alpha. Cette observation est cohérente avec la rareté d'une séparation spontanée en l'absence de force. De même, pour les domaines isolés (L), malgré une certaine flexibilité le long de la coordonnée d'association (α), la transition complète $S \leftrightarrow A$ ne s'effectue dans aucune des simulations REST₂. Les systèmes A explorent essentiellement des valeurs de α supérieures à 0,6, et inversement pour les systèmes S. Un faible recouvrement est observable autour de valeurs intermédiaires ($\alpha \sim 0,6$). Cependant, ces conforma-

Les trajectoires individuelles projetées sous forme de nuages de points sont visibles en FIGURE B.2 et FIGURE B.3.

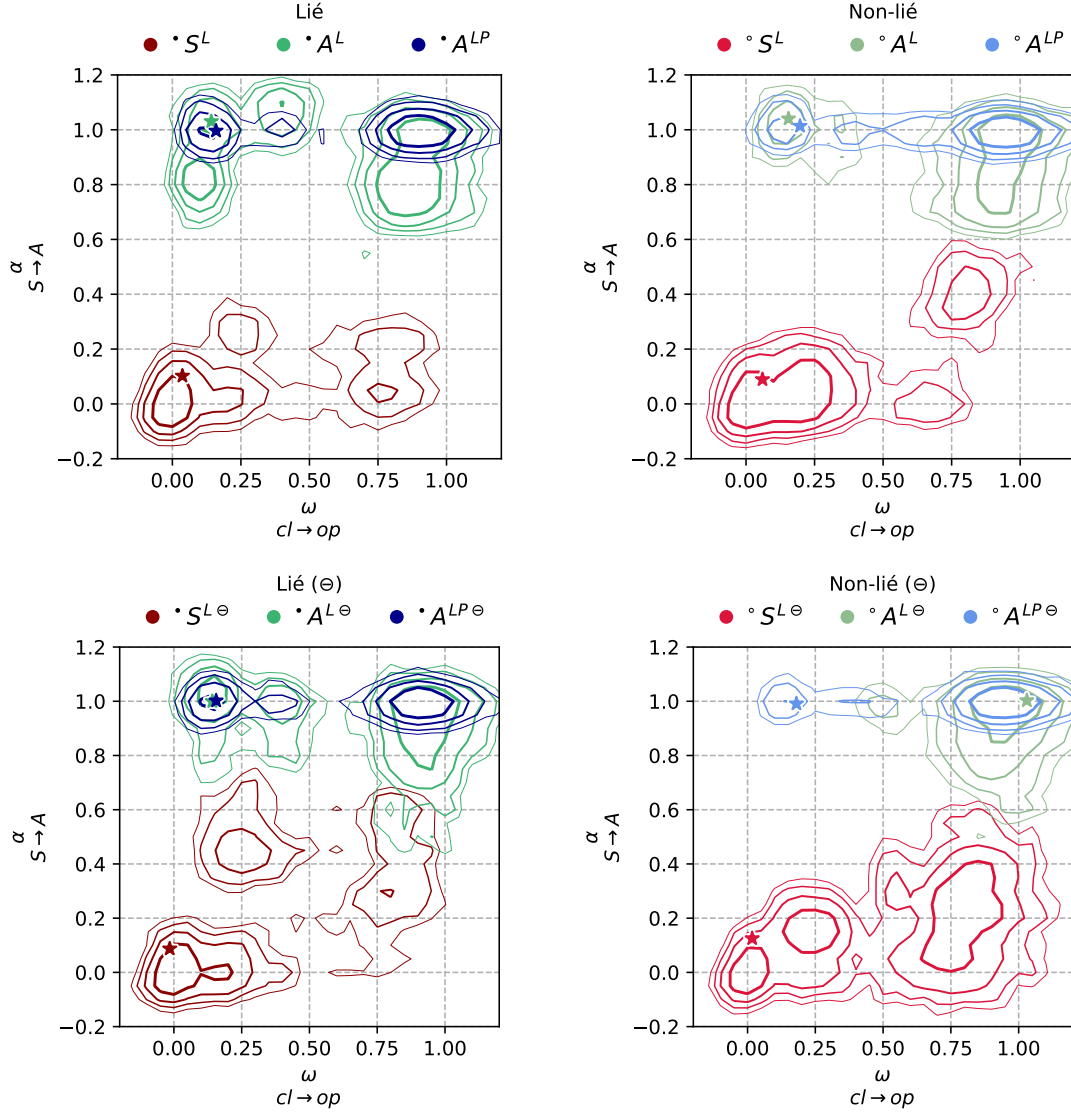


FIGURE 23 – Paysage conformationnel exploré au cours de REST2. Les configurations sont projetées sur l'espace joint des coordonnées d'ouverture (ω) et d'association (α). La densité locale est estimée par noyau gaussien de largeur 0,02 et représentée par un tracé de contour. Les configurations initiales sont dénotées par une étoile (*). (*gauche*) Systèmes ligantés, seules les configurations liées (ÉQUATION 26) sont conservées. (*droite*) Systèmes sans ligand. (*haut*) Systèmes avec histidine 45 diprotonée. (*bas*) Systèmes avec histidine 45 monoprotionée.

tions correspondent à un réarrangement des boucles inter-domaines (boucle pivot, boucle d'insertion, *linker*) mais pas de la région du renflement bêta, qui reste stable dans chaque trajectoire. Ceci est dû au caractère linéaire de la projection de la CV, qui peine à représenter parfaitement la véritable coordonnée de réaction pour la transition $S \leftrightarrow A$, hautement non-linéaire vis-à-vis des coordonnées atomiques.

La trajectoire d'une simulation REST₂ permet difficilement de tirer des informations sur la cinétique des changements conformationnels observés. En effet, les échanges MC produisent des trajectoires discontinues dans l'espace des configuration et cassent l'évolution temporelle du système d'intérêt. Toutefois, il est raisonnable de supposer que les différences d'ordre de grandeur entre les vitesses de plusieurs processus sont approximativement conservées. Dans cette optique, les résultats nous indiquent que l'étape cinétiquement déterminante de la transition $S \leftrightarrow A$ implique la réorganisation du renflement bêta/s-*witch* alpha.

Les degrés de libertés d'ouverture-fermeture du site de liaison et de réorganisation des boucles inter-domaines, eux, sont plus rapides et montrent une relaxation significative durant les 400 ns d'échantillonnage renforcé. Tous les systèmes initialement dans l'état fermé (cl) explorent plusieurs bassins le long de la coordonnée ω . Le système $^{\circ}A^{L\ominus}$, initialisé dans l'état op, ne visite pas l'état fermé. Nous allons rediscuter plus loin de la coordonnée d'ouverture.

Clustering

En complément de la détermination de CV d'analyse fondées sur notre intuition et notre connaissance du système, nous mettons en place un clustering pour identifier les bassins conformationnels de chaque système. Le clustering est une forme de réduction de dimensionnalité où les M points de données (configurations) de la trajectoire sont classifiés dans un ensemble réduit de N_{clust} classes, ou clusters. Nous adoptons une méthode de clustering agglomératif basé sur un estimateur de densité proposée par SITTEL et STOCK [49]. Un jeu de coordonnées internes de la protéine (distances entre paires de carbones alpha) est sélectionné, puis une réduction de dimensionnalité par analyse en composantes principale (PCA) permet de conserver les principaux modes de variance conformationnelle. Une énergie libre locale est calculée pour chaque point de donnée dans cet espace réduit *via* un estimateur de densité à noyau, puis un algorithme construit agglomérativement des clusters d'énergie maximale croissante. Par souci de concision, les détails de la méthode sont exposés en annexe (A.1).

Ce clustering nous permet d'identifier des configurations représentatives des états métastables visités durant REST₂ et de vérifier que la

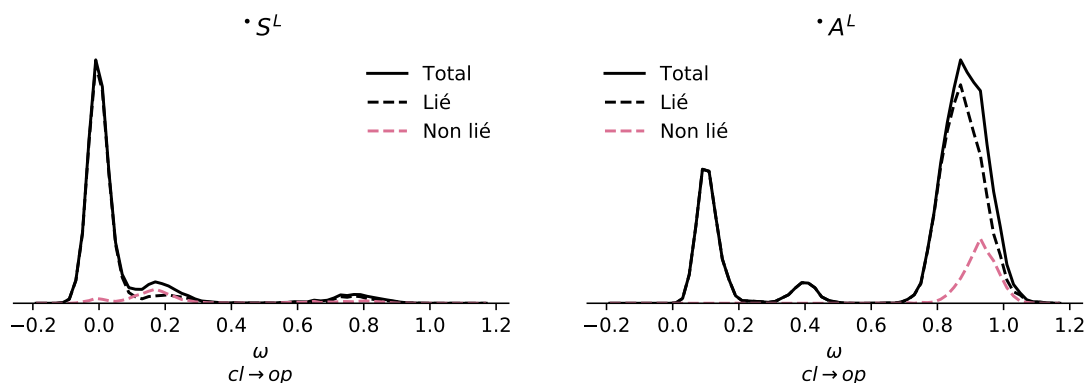


FIGURE 24 – Distribution marginale de la variable d’ouverture dans REST₂. La projection de la trajectoire sur ω permet d’identifier clairement trois pics correspondant à des sites de liaison ouverts (op), semi-ouverts (so) et clos (cl).

projection des trajectoires sur l’espace (ω, α) est relativement robuste pour séparer ces états. La FIGURE 25 illustre le résultat pour les systèmes $\bullet S^L$ et $\bullet A^L$. La visualisation moléculaire des représentants des clusters nous permet d’en comprendre mieux les détails structuraux. En particulier – et c’est ce sur quoi nous allons nous focaliser par la suite – elle permet d’apprécier la plasticité du site de liaison dans la région de la boucle de fermeture.

États d’ouverture

La distribution marginale de la coordonnée d’ouverture ω indique très clairement la présence de trois modes (FIGURE 24 et FIGURE B.4). Les deux pics extrémaux correspondent aux états, ouvert (op) et clos (cl) déjà représentés dans les structures cristallographiques. Il est intéressant de constater la présence d’un troisième pic correspondant à un état d’ouverture métastable intermédiaire, que nous qualifions de semi-ouvert (so). Cet état correspond à une disruption du petit feuillet β formé entre le court brin β de la boucle de fermeture (rés. 10–11) et le brin β N-terminal (rés. 1–5), une structure baptisée *pocket zipper* dans la littérature. La FIGURE 25 représente ces différents états et leur position projetée sur le sous-espace (ω, α) .

Il est intéressant de noter une différence nette de degré d’ouverture à l’issue de REST₂ entre les systèmes diprotonés et monoprotonés (\ominus). La présence de l’histidine 45 sous sa forme cationique stabilise l’état clos comme on peut le constater à la différence frappante entre les distributions des systèmes S^L et $S^{L\ominus}$ (FIGURE B.4), avec ou sans ligand. Cette observation plaide en faveur de l’importance de l’état de protonation vis-à-vis de la complexation entre FimH et les mannosides.

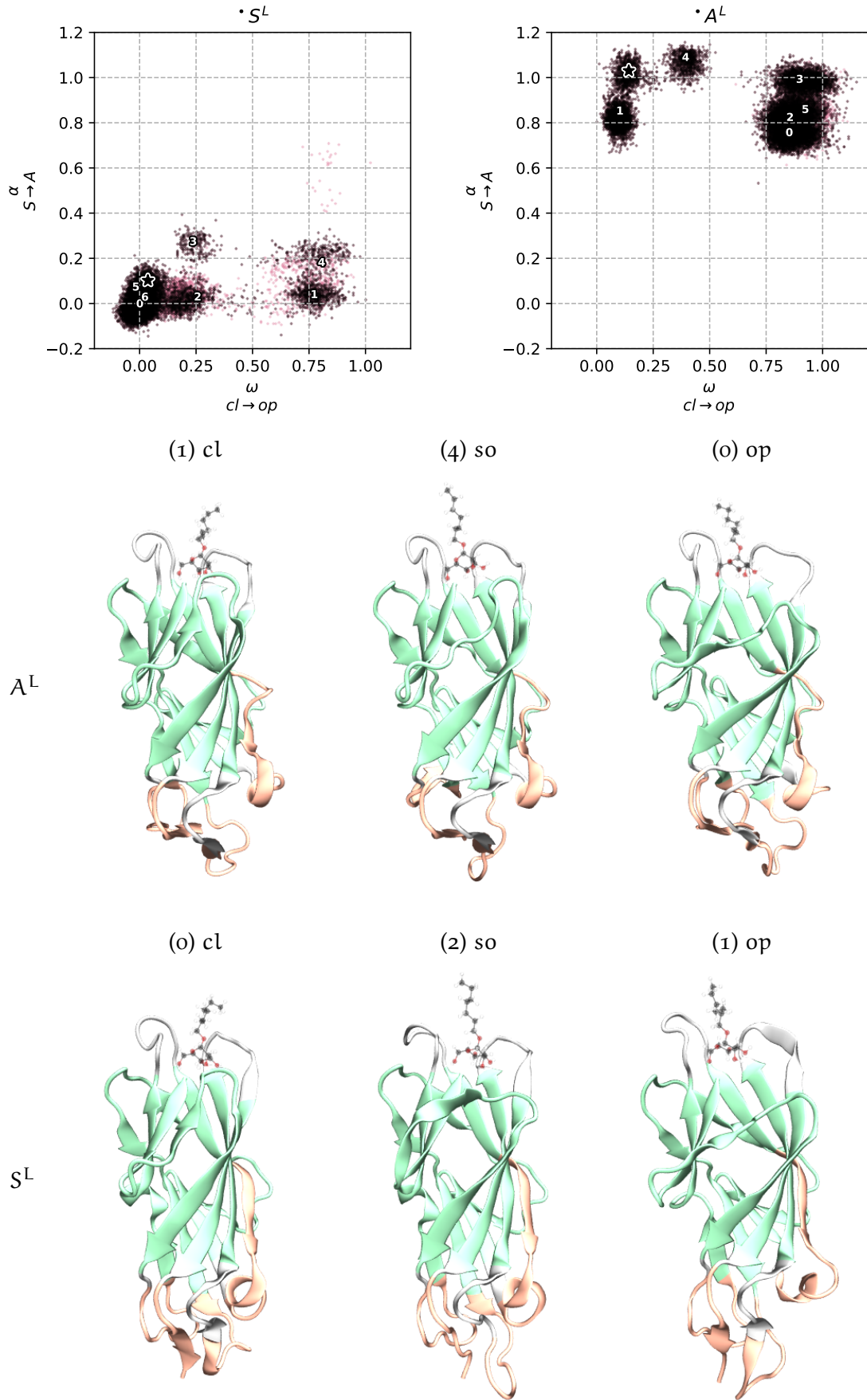


FIGURE 25 – États d'ouverture dans les états A et S. (*haut*) Trajectoires REST₂ de $\bullet S^L$ et $\bullet A^L$ projetées sur l'espace (ω, α) . Les points roses correspondent aux configurations non-liées. Les clusters sont numérotés. Des exemples (*bas*) sont illustrés pour des états fermés (cl), semi-ouverts (so) et ouvert (op).

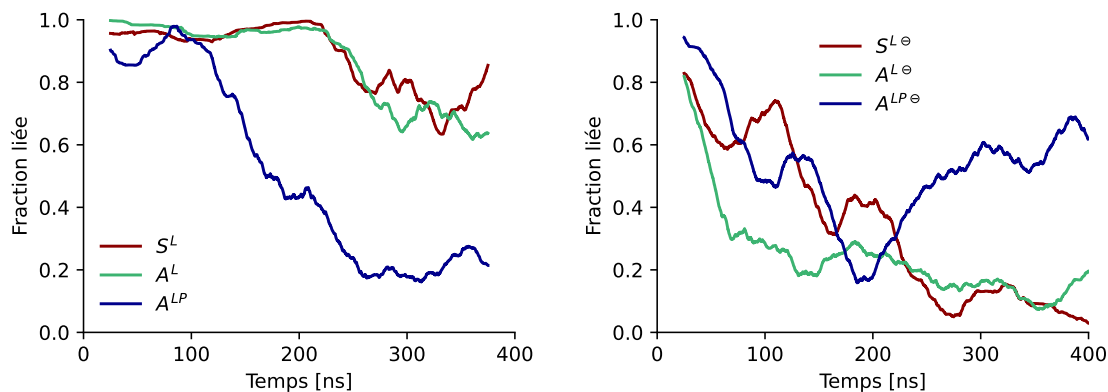


FIGURE 26 – Évolution de la fraction de configurations liées au cours de REST₂. Moyenne sur une fenêtre glissante (50 ns) de la fonction indicatrice $\mathbb{1}$ de l'état lié. (gauche) Systèmes avec histidine 45 diprotonée. (droite) Systèmes avec histidine 45 monoprotée.

Stabilité de l'état lié

De façon remarquable, la coordonnée d'ouverture corrèle avec la fraction de configurations non-liées (selon le critère de l'ÉQUATION 26) bien que le ligand soit retenu proche du site de liaison par la contrainte DBC.

La distribution de la DBC en FIGURE 21 suggère des degrés de stabilité différents du complexe protéine-ligand en fonction de l'état allostérique et de la protonation de His45. En reprenant la définition stricte de l'état lié (ÉQUATION 26), nous pouvons calculer l'évolution de la fraction de configurations liées au cours de la trajectoire sur une fenêtre glissante (FIGURE 26). Nous constatons dans tous les cas une diminution de la fraction liée, avec des disparités selon les systèmes.

Les systèmes avec His45 diprotonée semblent montrer une stabilité plus forte du complexe pour le domaine lectine seul (L) que pour la protéine entière (LP). Ce phénomène peut être relié au degré d'ouverture plus grand du site de liaison dans les systèmes A^{LP} , probable conséquence de la présence du domaine piline. Les systèmes monoprotés (\ominus) paraissent montrer une décroissance globalement plus rapide de la fraction liée. Le caractère bruité des données rend toutefois leur interprétation délicate, et une analyse plus quantitative de la répartition des états d'ouverture et leurs affinités respectives est requise pour confirmer cette tendance.

3.3.4 Conclusion

Dans cette section nous avons appliqué une méthode d'échantillonnage renforcé, REST₂, afin d'accélérer sans à priori les degrés de libertés du domaine lectine dans plusieurs systèmes et conformations

initiales. Nous avons pu échantillonner des régions de l'espace des configuration qui n'étaient pas atteintes dans les simulations préliminaires de dynamique moléculaire traditionnelle à l'échelle de la microseconde. Ainsi, de nouvelles conformations n'ayant pas de précédent dans la littérature ont été mises en évidence. Les plus intéressantes, en ce qui concerne les propriétés d'attachement de FimH aux mannosides, s'organisent autour de la coordonnée d'ouverture du site de liaison : en plus des états ouvert et clos déjà caractérisés par cristallographie du domaine lectine isolé (S_{cl}^L), de la protéine entière FimH·DsG libre (A_{op}^{LP}) et complexée au HM (A_{cl}^{LP}), nous montrons l'existence non-négligeable – avec le champ de force choisi – de l'état S_{op} et d'états intermédiaires « semi-ouverts » A_{so} et S_{so} .

La transition entre les deux états allostériques S et A ne s'effectue pas malgré l'échantillonnage renforcé, ce qui suggère une barrière énergétique importante, et ce même en l'absence du domaine piline pour stabiliser l'état A. Si une certaine flexibilité des boucles inter-domaines (pivot, insertion, *linker*) suggère un début de réorganisation aux échelles de temps accessibles, la région « clé » du renflement bêta et du switch alpha est remarquablement stable et constitue très probablement le motif structural limitant du changement conformationnel. Cela signifie que les ensembles conformationnels obtenus sont davantage représentatifs des bassins conformationnels « locaux », correspondant aux états A et S, que du paysage global, et il est difficile de tirer des conclusions quant à la stabilité relative des deux états allostériques, que ce soit en l'absence ou en présence du domaine piline.

Nos résultats illustrent également l'importance marquée de l'état acidobasique de la protéine vis-à-vis des ensembles obtenus, et incitent à redoubler d'attention dans la sélection d'un état de protonation pertinent pour le contexte physiologique à l'étude. Dans notre cas, la forme de l'histidine 45, prédite acide par des méthodes dédiées mais pas par les outils automatisés de préparation de simulations moléculaires, semble avoir un rôle dans la dynamique d'ouverture du site de liaison et la stabilité du complexe avec le ligand mannoside. Si ses effets quantitatifs restent à déterminer, il est raisonnable d'attendre de ce résidu une influence perceptible, au regard de son positionnement et de son environnement chimique. En effet, His45 appartient à un réseau alterné de paires ioniques à proximité du site et du *pocket zipper*. La FIGURE 27 illustre cet environnement pour des configurations fermées et ouvertes des systèmes S^L et $S^{L\ominus}$. Nous pouvons proposer un début d'interprétation mécanistique en remarquant que la forme acide, chargée positivement, a un effet électrostatique répulsif sur Lys4. Cela pourrait défavoriser le léger mouvement de torsion du brin N-terminal accompagnant l'ouverture de la boucle.

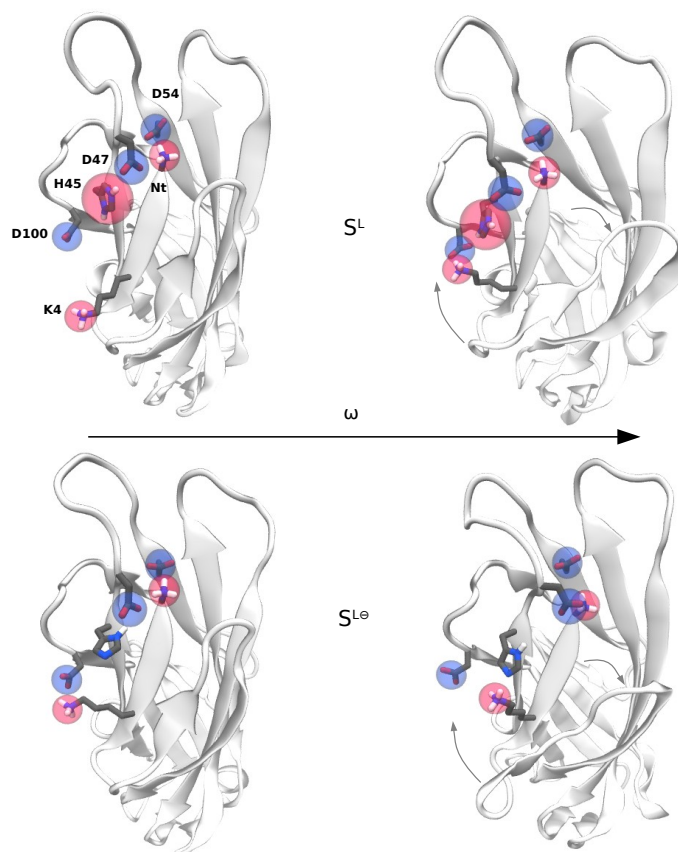


FIGURE 27 – Environnement de l’histidine 45 et possible effet sur le site de liaison. Nous pouvons par exemple supposer que la répulsion électrostatique entre His45 et Lys4, absente dans les formes monoprotonées (\ominus), influence la dynamique du *pocket zipper* et de la boucle de fermeture.

En dépit de l’exploration qualitative du paysage conformationnel, il nous faut insister sur le caractère non quantitatif de notre approche lié au temps limité d’échantillonnage. Nous avons fait le choix d’inclure dans notre analyse l’intégralité des trajectoires REST₂ afin de maximiser la diversité conformationnelle. Cependant, la partie initiale d’une telle trajectoire correspond à une phase transitoire loin de l’équilibre et elle est généralement éliminée pour des analyses quantitatives car elle biaise fortement les populations de l’ensemble conformationnel vers les configurations de départ. Cette phase « d’échauffement » est également connue pour privilégier les états enthalpiquement stables par rapport aux états entropiquement stables [50]. Étant donné l’extension insuffisante de nos simulations pour espérer obtenir un régime de convergence suffisamment long, nous préférons faire appel à des techniques complémentaires pour traiter de façon plus quantitative les populations des états observés. C’est ce que nous allons aborder dans le chapitre suivant.

RÉFÉRENCES

- [1] J. JUNG et al. Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *Journal of Computational Chemistry*, **40**, 1919-1930 (2019).
DOI : [10.1002/jcc.25840](https://doi.org/10.1002/jcc.25840)
- [2] D. van der SPOEL. Systematic design of biomolecular force fields. *Current Opinion in Structural Biology*, **67**, 18-24 (2021).
DOI : [10.1016/j.sbi.2020.08.006](https://doi.org/10.1016/j.sbi.2020.08.006)
- [3] J. MU, H. LIU, J. ZHANG, R. LUO et H.-F. CHEN. Recent Force Field Strategies for Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling*, **61**, 1037-1047 (2021).
DOI : [10.1021/acs.jcim.0c01175](https://doi.org/10.1021/acs.jcim.0c01175)
- [4] G. TORRIE et J. VALLEAU. Nonphysical sampling distributions in Monte Carlo free-energy estimation : Umbrella sampling. *Journal of Computational Physics*, **23**, 187-199 (1977).
DOI : [10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8)
- [5] A. LAIO et M. PARRINELLO. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, **99**, 12562-12566 (2002).
DOI : [10.1073/pnas.202427399](https://doi.org/10.1073/pnas.202427399)
- [6] A. BARDUCCI, G. BUSSI et M. PARRINELLO. Well-Tempered Metadynamics : A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters*, **100**, (2008).
DOI : [10.1103/PhysRevLett.100.020603](https://doi.org/10.1103/PhysRevLett.100.020603)
- [7] E. DARVE, D. RODRÍGUEZ-GÓMEZ et A. POHORILLE. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics*, **128**, 144120 (2008).
DOI : [10.1063/1.2829861](https://doi.org/10.1063/1.2829861)
- [8] J. HÉNIN, G. FIORIN, C. CHIPOT et M. L. KLEIN. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *Journal of Chemical Theory and Computation*, **6**, 35-47 (2010).
DOI : [10.1021/ct9004432](https://doi.org/10.1021/ct9004432)
- [9] Y. SUGITA et Y. OKAMOTO. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, **314**, 141-151 (1999).
DOI : [10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9)
- [10] R. C. BERNARDI, M. C. MELO et K. SCHULTEN. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **1850**, 872-877 (2015).
DOI : [10.1016/j.bbagen.2014.10.019](https://doi.org/10.1016/j.bbagen.2014.10.019)
- [11] Y. I. YANG, Q. SHAO, J. ZHANG, L. YANG et Y. Q. GAO. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics*, **151**, 070902 (2019).
DOI : [10.1063/1.5109531](https://doi.org/10.1063/1.5109531)
- [12] J. C. GORDON, J. B. MYERS, T. FOLTA, V. SHOJA, L. S. HEATH et A. ONUFRIEV. H++ : a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, **33**, W368-W371 (2005).
DOI : [10.1093/nar/gki464](https://doi.org/10.1093/nar/gki464)

- [13] J. MYERS, G. GROTHAUS, S. NARAYANAN et A. ONUFRIEV. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins : Structure, Function, and Bioinformatics*, **63**, 928-938 (2006).
DOI : [10.1002/prot.20922](https://doi.org/10.1002/prot.20922)
- [14] R. ANANDAKRISHNAN, B. AGUILAR et A. V. ONUFRIEV. H++ 3.0 : automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, **40**, W537-W541 (2012).
DOI : [10.1093/nar/gks375](https://doi.org/10.1093/nar/gks375)
- [15] H++ (web-based computational prediction of protonation states and pK of ionizable groups in macromolecules). URL : <http://biophysics.cs.vt.edu/index.php> (visité le 13/09/2021)
- [16] L. WANG, L. LI et E. ALEXOV. pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa : Predicting pKa's of Biomolecules. *Proteins : Structure, Function, and Bioinformatics*, **83**, 2186-2197 (2015).
DOI : [10.1002/prot.24935](https://doi.org/10.1002/prot.24935)
- [17] L. WANG, M. ZHANG et E. ALEXOV. DelPhiPKa web server : predicting pKa of proteins, RNAs and DNAs. *Bioinformatics*, **32**, 614-615 (2016).
DOI : [10.1093/bioinformatics/btv607](https://doi.org/10.1093/bioinformatics/btv607)
- [18] S. PAHARI, L. SUN, S. BASU et E. ALEXOV. DelPhiPKa : Including salt in the calculations and enabling polar residues to titrate. *Proteins : Structure, Function, and Bioinformatics*, **86**, 1277-1283 (2018).
DOI : [10.1002/prot.25608](https://doi.org/10.1002/prot.25608)
- [19] DelPhiPKa Web Server. URL : http://compbio.clemson.edu/pka_webserver/ (visité le 13/09/2021)
- [20] P. DIEUSAERT. *Guide pratique des analyses médicales*. 6e éd. Paris : Maloine, 2015
- [21] A. JACQUOT, C. SAKAMOTO, A. RAZAFITIANAMAHARAVO, C. CAILLET, J. MERLIN, A. FAHS, J. M. GHIGO, C. BELOIN, J. F. L. DUVAL et G. FRANCIUS. Dynamic Modulation of Fimbrial Extension and FimH-Mannose Binding Force on Live Bacteria Under pH Changes : A Molecular Atomic Force Microscopy Analysis. *Journal of Biomedical Nanotechnology*, **10**, 3361-3372 (2014).
DOI : [10.1166/jbn.2014.1905](https://doi.org/10.1166/jbn.2014.1905)
- [22] O. GUVENCH et A. D. MACKERELL. Comparison of Protein Force Fields for Molecular Dynamics Simulations. *Molecular Modeling of Proteins*. Sous la dir. d'A. KUKOL. Réd. par J. M. WALKER. T. 443. Series Title : Methods in Molecular Biology. Totowa, NJ : Humana Press, 2008. 63-88. DOI : [10.1007/978-1-59745-177-2_4](https://doi.org/10.1007/978-1-59745-177-2_4)
- [23] P. ROBUSTELLI, S. PIANA et D. E. SHAW. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, **115**, E4758-E4766 (2018).
DOI : [10.1073/pnas.1800690115](https://doi.org/10.1073/pnas.1800690115)
- [24] J. HUANG et A. D. MACKERELL. Force field development and simulations of intrinsically disordered proteins. *Current Opinion in Structural Biology*, **48**, 40-48 (2018).
DOI : [10.1016/j.sbi.2017.10.008](https://doi.org/10.1016/j.sbi.2017.10.008)

- [25] K. LINDORFF-LARSEN, S. PIANA, K. PALMO, P. MARAGAKIS, J. L. KLEPEIS, R. O. DROR et D. E. SHAW. Improved side-chain torsion potentials for the Amber ff99SB protein force field : Improved Protein Side-Chain Potentials. *Proteins : Structure, Function, and Bioinformatics*, **78**, 1950-1958 (2010).
DOI : [10.1002/prot.22711](https://doi.org/10.1002/prot.22711)
- [26] J. HUANG, S. RAUSCHER, G. NAWROCKI, T. RAN, M. FEIG, B. L. de GROOT, H. GRUBMÜLLER et A. D. MAC KERELL. CHARMM36m : an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, **14**, 71-73 (2017).
DOI : [10.1038/nmeth.4067](https://doi.org/10.1038/nmeth.4067)
- [27] G. BUSSI, D. DONADIO et M. PARRINELLO. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, **126**, 014101 (2007).
DOI : [10.1063/1.2408420](https://doi.org/10.1063/1.2408420)
- [28] M. LINGENHEIL, R. DENSCHLAG, R. REICHOLD et P. TAVAN. The “Hot-Solvent/Cold-Solute” Problem Revisited. *Journal of Chemical Theory and Computation*, **4**, 1293-1306 (2008).
DOI : [10.1021/ct8000365](https://doi.org/10.1021/ct8000365)
- [29] J. WANG, R. M. WOLF, J. W. CALDWELL, P. A. KOLLMAN et D. A. CASE. Development and testing of a general amber force field. *Journal of Computational Chemistry*, **25**, 1157-1174 (2004).
DOI : [10.1002/jcc.20035](https://doi.org/10.1002/jcc.20035)
- [30] A. W. SOUSA DA SILVA et W. F. VRANKEN. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, **5**, 367 (2012).
DOI : [10.1186/1756-0500-5-367](https://doi.org/10.1186/1756-0500-5-367)
- [31] MacKerell Lab. URL : http://mackerell.umaryland.edu/charmm_ff.shtml (visité le 16/09/2021)
- [32] CGenFF Home. URL : <https://cgenff.umaryland.edu/> (visité le 16/09/2021)
- [33] R. H. SWENDSEN et J.-S. WANG. Replica Monte Carlo Simulation of Spin-Glasses. *Physical Review Letters*, **57**, 2607-2609 (1986).
DOI : [10.1103/PhysRevLett.57.2607](https://doi.org/10.1103/PhysRevLett.57.2607)
- [34] D. J. EARL et M. W. DEEM. Parallel tempering : Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, **7**, 3910 (2005).
DOI : [10.1039/b509983h](https://doi.org/10.1039/b509983h)
- [35] L. WANG, R. A. FRIESNER et B. J. BERNE. Replica Exchange with Solute Scaling : A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B*, **115**, 9431-9438 (2011).
DOI : [10.1021/jp204407d](https://doi.org/10.1021/jp204407d)
- [36] G. STIRNEMANN et F. STERPONE. Recovering Protein Thermal Stability Using All-Atom Hamiltonian Replica-Exchange Simulations in Explicit Solvent. *Journal of Chemical Theory and Computation*, **11**, 5573-5577 (2015).
DOI : [10.1021/acs.jctc.5b00954](https://doi.org/10.1021/acs.jctc.5b00954)
- [37] D. A. KOFKE. On the acceptance probability of replica-exchange Monte Carlo trials. *The Journal of Chemical Physics*, **117**, 6911-6914 (2002).
DOI : [10.1063/1.1507776](https://doi.org/10.1063/1.1507776)

- [38] A. KONE et D. A. KOFKE. Selection of temperature intervals for parallel-tempering simulations. *The Journal of Chemical Physics*, **122**, 206101 (2005).
DOI : [10.1063/1.1917749](https://doi.org/10.1063/1.1917749)
- [39] M. M. SAUER, R. P. JAKOB, J. ERAS, S. BADAY, D. ERIŞ, G. NAVARRA, S. BERNÈCHE, B. ERNST, T. MAIER et R. GLOCKSHUBER. Catch-bond mechanism of the bacterial adhesin FimH. *Nature Communications*, **7**, (2016).
DOI : [10.1038/ncomms10738](https://doi.org/10.1038/ncomms10738)
- [40] G. A. TRIBELLO, M. BONOMI, D. BRANDUARDI, C. CAMILLONI et G. BUSSI. PLUMED 2 : New feathers for an old bird. *Computer Physics Communications*, **185**, 604-613 (2014).
DOI : [10.1016/j.cpc.2013.09.018](https://doi.org/10.1016/j.cpc.2013.09.018)
- [41] G. BUSSI. Hamiltonian replica exchange in GROMACS : a flexible implementation. *Molecular Physics*, **112**, 379-384 (2014).
DOI : [10.1080/00268976.2013.824126](https://doi.org/10.1080/00268976.2013.824126)
- [42] J. D. CHODERA et M. R. SHIRTS. Replica exchange and expanded ensemble simulations as Gibbs sampling : Simple improvements for enhanced mixing. *The Journal of Chemical Physics*, **135**, 194110 (2011).
DOI : [10.1063/1.3660669](https://doi.org/10.1063/1.3660669)
- [43] R. IWAI, K. KASAHARA et T. TAKAHASHI. Influence of various parameters in the replica-exchange molecular dynamics method : Number of replicas, replica-exchange frequency, and thermostat coupling time constant. *Biophysics and Physicobiology*, **15**, 165-172 (2018).
DOI : [10.2142/biophysico.15.0_165](https://doi.org/10.2142/biophysico.15.0_165)
- [44] R. SALARI, T. JOSEPH, R. LOHIA, J. HÉNIN et G. BRANNIGAN. A Streamlined, General Approach for Computing Ligand Binding Free Energies and Its Application to GPCR-Bound Cholesterol. *Journal of Chemical Theory and Computation*, **14**, 6560-6573 (2018).
DOI : [10.1021/acs.jctc.8b00447](https://doi.org/10.1021/acs.jctc.8b00447)
- [45] G. FIORIN, M. L. KLEIN et J. HÉNIN. Using collective variables to drive molecular dynamics simulations. *Molecular Physics*, **111**, 3345-3362 (2013).
DOI : [10.1080/00268976.2013.813594](https://doi.org/10.1080/00268976.2013.813594)
- [46] A. AMADEI, G. CHILLEMI, M. A. CERUSO, A. GROTTESI et A. DI NOLA. Molecular dynamics simulations with constrained roto-translational motions : Theoretical basis and statistical mechanical consistency. *The Journal of Chemical Physics*, **112**, 9-23 (2000).
DOI : [10.1063/1.480557](https://doi.org/10.1063/1.480557)
- [47] *Collective Variables Module - Colvars Module - Reference Manual*. URL : <https://colvars.github.io/colvars-refman-gromacs/colvars-refman-gromacs.html> (visité le 22/09/2021)
- [48] G. A. TRIBELLO et P. GASPAROTTO. Using Dimensionality Reduction to Analyze Protein Trajectories. *Frontiers in Molecular Biosciences*, **6**, 46 (2019).
DOI : [10.3389/fmolb.2019.00046](https://doi.org/10.3389/fmolb.2019.00046)

- [49] F. SITTEL et G. STOCK. Robust Density-Based Clustering To Identify Metastable Conformational States of Proteins. *Journal of Chemical Theory and Computation*, **12**, 2426-2435 (2016).
DOI : [10.1021/acs.jctc.5b01233](https://doi.org/10.1021/acs.jctc.5b01233)
- [50] S. WINGBERMÜHLE et L. V. SCHÄFER. *Capturing the Flexibility of a Protein-Ligand Complex with Enhanced Sampling Techniques : Binding Free Energies from Umbrella Sampling, REST2, and Bias Exchange*. preprint. 2019.
DOI : [10.26434/chemrxiv.11106806.v1](https://doi.org/10.26434/chemrxiv.11106806.v1)

THERMODYNAMIQUE D'OUVERTURE DU SITE DE LIAISON

Les résultats de l'exploration par REST₂ nous donnent un aperçu qualitatif de la plasticité conformationnelle du domaine lectine sous diverses conditions (présence ou absence du ligand, état allostérique, présence ou absence du domaine piline). La propension à l'ouverture du site de liaison, y compris dans l'état S laisse penser que les constantes d'affinité et vitesses de réaction expérimentales doivent être interprétées au regard de cette dynamique plutôt qu'à partir des structures fixes héritées de la cristallographie.

Comme nous l'avons souligné, les trajectoires issues de REST₂ ne permettent pas d'obtenir une estimation fiable de la proportion de chaque état d'ouverture observé. Dans cette section nous souhaitons obtenir une vision plus quantitative du profil d'énergie libre le long de la coordonnée d'ouverture. Nous nous tournons donc vers une méthode d'échantillonnage renforcé basée sur la définition d'une variable collective : l'échantillonnage parapluie à échange de biais (BEUS, *bias exchange umbrella sampling*).

4.0.1 Échantillonnage parapluie

L'échantillonnage parapluie ou *umbrella sampling* est l'une des méthodes d'échantillonnage préférentiel les plus anciennes et les plus utilisées [1-4]. Elle consiste à échantillonner les régions difficilement accessibles de la surface d'énergie potentielle le long d'une coordonnée de réaction à l'aide d'un potentiel de biais empêchant le système de rester bloqué dans les régions de plus basse énergie. Il est fréquent d'échantillonner plusieurs ensembles biaisés avec des potentiels se focalisant sur différentes régions de la coordonnée. On parle alors de « fenêtres », dont on peut rassembler les données pour calculer des observables sur la coordonnée entière.

Soit $U_{\text{ref}}(\mathbf{x})$ l'énergie potentielle non-perturbée du système dans une configuration \mathbf{x} , et $\xi(\mathbf{x})$ une variable collective à une dimension. La surface d'énergie potentielle du système biaisé dans une fenêtre i est alors

$$U(\mathbf{x}) = U_{\text{ref}}(\mathbf{x}) + w_i(\xi(\mathbf{x})) \quad (27)$$

où w_i est le potentiel de biais.

La quantité qui nous intéresse est la distribution de probabilité du système non perturbé le long de la coordonnée ξ ,

$$p_{\text{ref}}(\xi) = \frac{\int d^N \mathbf{x} \delta[\xi - \xi(\mathbf{x})] \exp[-\beta U_{\text{ref}}(\mathbf{x})]}{Q_{\text{ref}}} \quad (28)$$

où $Q_{\text{ref}} = \int d^N \mathbf{x} \exp[-\beta U_{\text{ref}}(\mathbf{x})]$ est la fonction de partition et δ la distribution de DIRAC. La grandeur $p_{\text{ref}}(\xi)d\xi$ correspond à la probabilité de trouver le système dans l'intervalle infinitésimal $d\xi$ autour de ξ . Une façon équivalente de décrire cette quantité est de l'exprimer en termes de profil d'énergie libre $F_{\text{ref}}(\xi)$:

$$F_{\text{ref}}(\xi) = -k_B T \ln p_{\text{ref}}(\xi). \quad (29)$$

Supposons maintenant que l'on ne dispose que de $p_i(\xi)$, la distribution du système biaisé avec w_i . On peut montrer que le profil d'énergie libre non-biaisé s'exprime

$$F_{\text{ref}}(\xi) = -k_B T \ln p_i(\xi) - w_i(\xi) + \Delta F_i \quad (30)$$

où $\Delta F_i = -k_B T \ln \langle \exp(-\beta w_i) \rangle_{\text{ref}}$ est la contribution à l'énergie libre liée à l'ajout du biais et ne dépend pas de ξ . Le terme $\ln p_i$ peut être évalué directement par échantillonnage dans la simulation biaisée. À cause du logarithme, l'erreur d'estimation diverge lorsque $p_i(\xi)$ devient petit. Le terme ΔF_i ne peut pas être calculé dans la simulation biaisée. Pour l'utilisation d'une unique fenêtre il peut être ignoré car l'énergie libre n'est définie qu'à une constante arbitraire près. En revanche, si l'on souhaite agréger les données de plusieurs fenêtres, par exemple pour explorer différentes régions de la coordonnée que l'on ne parvient pas à couvrir dans une seule simulation, il nous faut calculer l'ensemble des ΔF_i , ou du moins leurs écarts relatifs. Cela n'est possible que si les distributions $p_i(\xi)$ et $p_j(\xi)$ de deux fenêtres possèdent un recouvrement significatif. Une méthode statistiquement efficace pour estimer les $\{\Delta F_i\}$ d'un ensemble de N répliques est celle de l'analyse par histogramme pondéré ou WHAM (*weighted histogram analysis method*) [5].

Le choix traditionnel en échantillonnage parapluie est de simuler indépendamment un ensemble de N fenêtres avec des biais harmoniques espacés le long de la CV

$$w_i(\xi) = \frac{1}{2} \kappa_i (\xi - \xi_i)^2 \quad (31)$$

où κ_i est la constante harmonique contrôlant la largeur du potentiel et ξ_i la position du minimum. Ces valeurs sont choisies de façon à maintenir le système proche des $\{\xi_i\}$ dans les régions pentues du profil d'énergie libre tout en assurant le recouvrement suffisant des fenêtres adjacentes le long de ξ .

Échantillonnage parapluie adaptatif

Peut on choisir un biais $w_i(\xi)$ plus efficient? Un choix optimal serait précisément de contruire un unique biais permettant d'échantillonner uniformément la section de la coordonnée ξ qui nous intéresse :

$$w_{\text{opt}}(\xi) = k_B T \ln p_{\text{ref}}(\xi) \quad (32)$$

et clairement l'ÉQUATION 30 devient

$$F_{\text{ref}}(\xi) = -w_{\text{opt}}(\xi) + \text{constante} \quad (33)$$

nous enfermant dans une logique quelque peu circulaire. Nous pouvons néanmoins raisonner incrémentalement et proposer un biais $w^{(0)}$ quelconque, puis calculer un profil d'énergie libre approximatif

$$\hat{F}_{\text{ref}}^{(0)} = -k_B T \ln \hat{p}^{(0)} - w^{(0)} + \text{constante} \quad (34)$$

où $\hat{p}^{(0)}$ est une estimation de la distribution dans l'état biaisé, qui n'a pas besoin d'être exacte. Le profil $\hat{F}_{\text{ref}}^{(0)}$ sera probablement une très mauvaise approximation du profil réel dans les régions non échantillonnées. Nous pouvons néanmoins construire un second biais

$$w^{(1)} = -\hat{F}_{\text{ref}}^{(0)} + w^{(0)} \quad (35)$$

et réitérer le processus pour calculer

$$\hat{F}_{\text{ref}}^{(1)} = -k_B T \ln \hat{p}^{(1)} - w^{(1)} + \text{constante} \quad (36)$$

Si tout se passe bien, à chaque itération, le biais $w^{(k)}$ approche w_{opt} et la distribution échantillonnée $\hat{p}^{(k)}$ se rapproche d'une loi uniforme. En pratique, même avec un biais optimal, la qualité de l'échantillonnage le long de ξ dépend d'autres paramètres pouvant devenir limitant tels que le temps de simulation par rapport au temps caractéristique de diffusion selon ξ , le temps de relaxation des degrés de liberté transverses, et conséquemment le choix d'une CV pertinente.

Cette approche, baptisée échantillonnage parapluie adaptatif (AUS, *adaptive umbrella sampling*) [2], peut être utilisée de façon complémentaire à d'autres techniques pour faciliter la convergence d'un calcul. Dans ces travaux nous la combinons à une méthode d'US à échange de biais, décrite ci-après.

Échantillonnage parapluie à échange de biais

L'un des problèmes de l'échantillonnage parapluie traditionnel (US) comportant N biais harmoniques et N simulations indépendantes, est celui de la relaxation des degrés de libertés transverse. En effet, ces degrés de libertés sont en général corrélés avec la variable collective ξ , et on ne peut estimer correctement la distribution locale que si leur échantillonnage est satisfaisant. Ce problème est général aux méthodes d'échantillonnage préférentiel basés sur une CV mais, contrairement à d'autres méthodes telles que la métadynamique, qui permettent la libre diffusion longitudinale selon la coordonnée d'intérêt, les biais harmoniques bloquent le système dans une région restreinte. De ce fait, la diffusion transverse ne peut se faire qu'à valeur de ξ quasi-fixée, et ne peut emprunter d'autres routes, potentiellement plus rapides, aux valeurs voisines. Une telle situation se produit, par exemple lorsque deux chemins réactionnels sont séparés par une haute barrière énergétique : les deux routes affectent la forme du profil d'énergie libre en ξ mais une seule sera échantillonnée durant la simulation. En empêchant le système de retourner vers les produits ou réactifs, le biais harmonique anéantit nos chances d'explorer la seconde route.

Afin de remédier à ce problème, il est possible de faire appel au cadre conceptuel des simulations Monte-Carlo à échange de réplique, présenté dans la section précédente. Les N fenêtres de l'US peuvent être vues comme N répliques dont les hamiltoniens sont modifiés par différents biais. Sur un principe similaire, on les simule simultanément en parallèle, et des échanges configurationnels sont proposés régulièrement et acceptés selon un critère de METROPOLIS-HASTINGS (ÉQUATION 14) avec

$$\Delta_{ij} = \beta_0 [w_i(\xi_j) + w_j(\xi_i) - w_i(\xi_i) - w_j(\xi_j)] \quad (37)$$

Cet US à échange de répliques – ou, selon le point de vue, à échange de biais – permet donc au système de diffuser de fenêtre en fenêtre, ce qui solutionne partiellement le problème énoncé plus haut.

4.0.2 Méthodes

Afin d'élucider le profil d'énergie libre selon la coordonnée d'ouverture (ω) du domaine lectine, nous proposons une combinaison de *umbrella sampling* à échange de biais (BEUS) et adaptatif (AUS). Une première phase de BEUS est réalisée afin d'obtenir un profil d'énergie libre (PEL) approché. En raison du choix non optimal des biais harmoniques, certaines régions du PEL ainsi obtenu souffrent d'un manque de recouvrement entre biais adjacents, visible par un faible taux d'échange. Le PEL obtenu est néanmoins inversé et utilisé comme

biais externe pour une deuxième phase de BEUS. Le recouvrement est alors bien meilleur et nous pouvons calculer le PEL final en retranchant le biais.

Dans le but d'accélérer la phase d'équilibrage de la méthode et la convergence, nous capitalisons sur les simulations précédentes en initialisant les fenêtres de BEUS avec des configurations échantillonnées par REST₂ et proches de la valeur ω_k correspondante. Ce faisant nous espérons partir d'une distribution des coordonnées transverses plus proche de l'équilibre.

Notons que le PEL est calculé dans les états A d'une part et S d'autre part. Nous pouvons le faire sans contrainte supplémentaire, en raison de l'extrême métastabilité de ces états sur les échelles de temps de nos simulations. L'hypothèse implicite est alors que nous atteignons un équilibre local au sein du bassin énergétique de chaque état.

Choix des fenêtres

Le nombre de fenêtres est fixé à 24 pour assurer une fine résolution le long de la CV, et par commodité, en raison de l'architecture matérielle du supercalculateur utilisé pour la simulation en parallèle. Les valeurs extrêmes sont sélectionnées à partir des distributions obtenues dans REST₂ (FIGURE B.4). L'étendue $[\omega_{\min}, \omega_{\max}]$ est légèrement différente dans les états S et A. La TABLE 7 donne les paramètres choisis.

La valeur de la constante harmonique κ_i est fixée égale pour tous les biais à $\kappa = 2000 \text{ kJmol}^{-1}$. Cela permet d'assurer que, hors influence du PEL, un biais harmonique génère une distribution gaussienne d'écart type $\sigma = \sqrt{k_B T / \kappa} = 0.035 \sim \Delta\omega$, et donc un recouvrement suffisant entre les distributions de deux répliques adjacentes à température ambiante.

État	ω_{\min}	ω_{\max}	$\Delta\omega$
S	-0,1	1,05	0,05
A	0,05	1,2	0,05

TABLE 7 – Paramètres des fenêtres BEUS.

Préparation des systèmes

Pour chaque système, soient \mathcal{T} l'ensemble des configurations issues de la trajectoire REST₂ et N fenêtres de BEUS centrées en $\{\omega_1, \dots, \omega_N\}$

et régulièrement espacées de $\Delta\omega$. Soit \mathcal{X}_i le sous ensemble de configurations de T dans un rayon $\Delta\omega$ de ω_i

$$\mathcal{X}_i = \{\mathbf{x} \in \mathcal{T} \mid \omega_i - \Delta\omega < \omega(\mathbf{x}) < \omega_i + \Delta\omega\}. \quad (38)$$

Nous sélectionnons comme configurations initiales $\{\mathbf{x}_1^0, \dots, \mathbf{x}_N^0\}$ au hasard dans l'ensemble \mathcal{X}_i , ou au plus proche de ω_i dans le cas où celui-ci serait vide

$$\begin{cases} \mathbf{x}_i^0 \sim \mathbb{U}(\mathcal{X}_i) & \text{si } \mathcal{X}_i \neq \emptyset \\ \mathbf{x}_i^0 = \arg \min_{\mathbf{x} \in \mathcal{T}} |\omega(\mathbf{x}) - \omega_i| & \text{sinon} \end{cases} \quad (39)$$

où \mathbb{U} désigne la loi uniforme discrète.

Chaque réplique est préparée en suivant le protocole décrit précédemment (voir 3.2.2). Dans les systèmes ligantés, le potentiel de retenue DBC est fixé à $\delta = 1,8 \text{ \AA}$ pour limiter l'échantillonnage des poses périphériques sans perturber l'état lié. Un biais prévenant la diffusion rotationnelle est utilisé pour les systèmes LP (voir 3.3.2).

Simulations

Une première phase de 100 ns de BEUS est réalisée pour l'ensemble des systèmes. Un PEL provisoire est calculé avec l'analyse par histogramme pondérée (WHAM). Un intervalle de 0,0175 est utilisé. Le PEL obtenu est inversé et implémenté comme biais externe grâce à la composante EXTERNAL de PLUMED, avec interpolation polynomiale activée (par défaut). Pour les valeurs de ω à droite et à gauche de la portion où le PEL a été calculé, le biais est continué par des marges constantes. Une seconde phase de 400 ns de BEUS est menée à bien pour le calcul du PEL final. Les trajectoires de chaque réplique dans l'espace de ω sont enregistrées à une fréquence de 100 ps.

Post-traitement

Pour les systèmes liés (●) les configurations ne respectant pas le critère de liaison (ÉQUATION 26) sont éliminées de l'analyse. Le premier quart de la simulation BEUS est également ignorée pour équilibration. L'algorithme d'analyse par histogramme pondérée (WHAM) est utilisé pour reconstituer le profil d'énergie libre à partir des N fenêtres (biais). Nous employons le programme WHAM, une implémentation de l'algorithme éponyme par GROSSFIELD [6].

En raison de l'absence d'échantillonnage de configurations dans la région close ($\omega \approx 0$) à l'issue des simulations REST2 pour $^\circ\text{A}^{\text{L}\ominus}$, ce système est ignoré dans la suite.

Estimation de l'erreur

Il est difficile d'estimer l'erreur commise sur le profil d'énergie libre obtenu. Deux types d'erreurs coexistent :

- L'erreur statistique, due à la taille finie de l'échantillon (les trajectoires de l'ensemble des fenêtres). Cette erreur tend vers zéro pour une taille d'échantillon infinie. Elle est indépendante de la qualité de l'échantillonnage, car elle le présuppose véritablement aléatoire (décorrélé).
- L'erreur liée à la corrélation des sous-échantillons. Cette erreur est à relier à l'hypothèse d'ergodicité. En faisant appel à la MD pour générer des configurations de l'ensemble canonique on obtient une séquence fortement corrélée temporellement. En particulier, les degrés de liberté transverses doivent relaxer rapidement relativement au temps d'acquisition pour une bonne convergence de l'estimation.

Le premier type d'erreur peut être estimé par simple analyse statistique comme un bootstrap : l'ensemble des n échantillons est ré-échantillonné aléatoirement (avec remise) et le PEL est recalculé ; ce processus est réitéré M fois afin de calculer la moyenne et l'écart type du PEL en chaque point. Cette approche est implémentée nativement dans le programme WHAM [6], et nous calculons l'erreur statistique avec $M = 20$.

Le deuxième type d'erreur est plus difficile à analyser. En effet, nous ne disposons d'aucun moyen de garantir la convergence vers la distribution d'équilibre, car des temps caractéristiques de relaxation plus longs que le temps de simulation nous sont par définition invisibles. En réalité, nous exploitons cette propriété pour définir des profils d'énergie libre distincts dans chacun des états métastables S^L , A^L et A^{LP} . Une telle approche a un sens si l'échelle de temps τ_α de la transition $S \leftrightarrow A$ est plus lente et bien séparée des autres degrés de liberté, dont la transition $cl \leftrightarrow op$. Si cela est bien le cas, la convergence du calcul peut être évaluée par l'obtention d'un régime permanent pour des échelles de temps $\ll \tau_\alpha$. Nous évaluons donc cette convergence grâce à une analyse par blocs : la trajectoire d'analyse est divisée en deux et le calcul du PEL est effectué sur chaque bloc temporel. Une faible déviation entre les deux PEL ainsi obtenus indique que l'on a convergé vers un ensemble quasi-stationnaire.

4.0.3 *Résultats*

Les profils d'énergie libre le long de la coordonnée ω sont présentés avec et sans ligand en FIGURE 28 pour les systèmes diprotonés, et FIGURE 29 pour les systèmes monoprotonés. Comme attendu au vu des structures cristallographiques, le domaine lectine seul dans l'état S présente un minimum énergétique dans sa forme close (cl) liée au

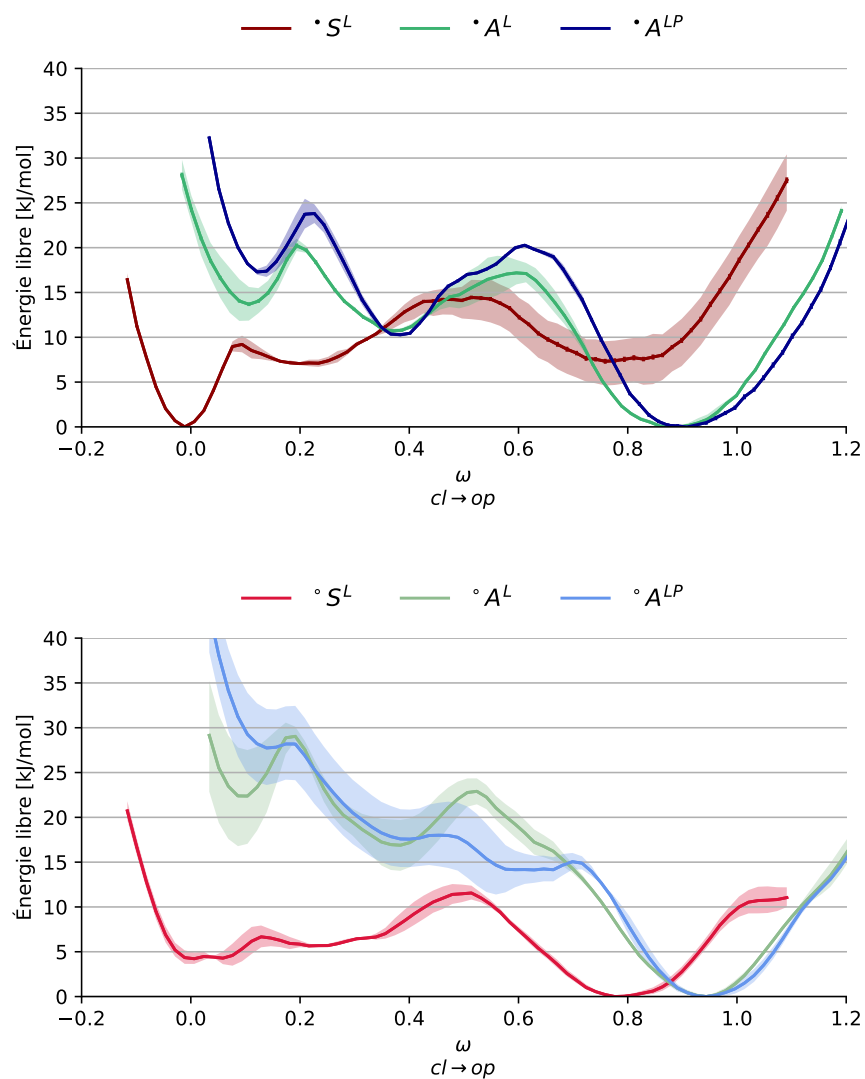


FIGURE 28 – Profil d'énergie libre selon la coordonnée d'ouverture (systèmes diprotonés). (*haut*) Systèmes ligantés. (*bas*) Systèmes non liés. Le minimum est fixé arbitrairement à zéro. Les barres d'erreurs (peu visibles, de l'ordre de l'épaisseur de la courbe) indiquent l'erreur statistique sur l'ensemble de la trajectoire estimée par bootstrap. L'aire entre les PEL estimés par blocs sur deux moitiés de trajectoire est colorée en transparence.

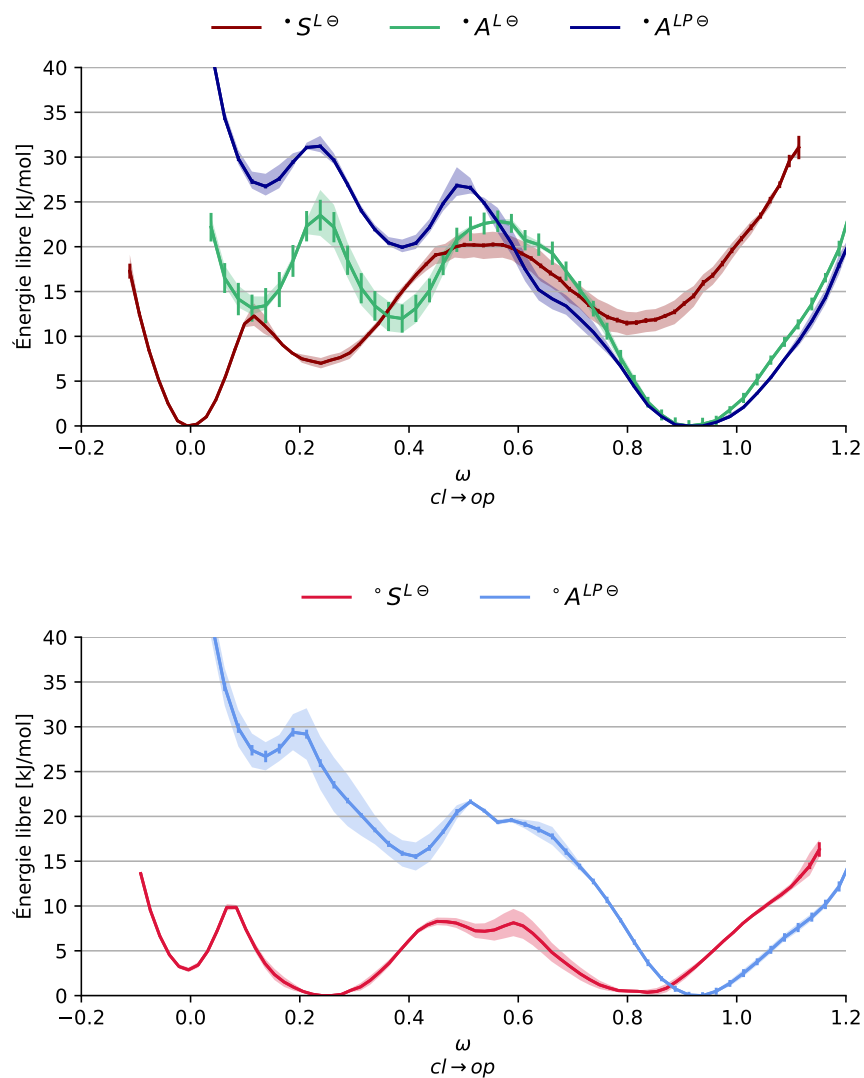


FIGURE 29 – Profil d'énergie libre selon la coordonnée d'ouverture (systèmes monoprotonnés). (*haut*) Systèmes ligantés. (*bas*) Systèmes non liés. Le minimum est fixé arbitrairement à zéro. Les barres d'erreurs (peu visibles, de l'ordre de l'épaisseur de la courbe) indiquent l'erreur statistique sur l'ensemble de la trajectoire estimée par bootstrap. L'aire entre les PEL estimés par blocs sur deux moitiés de trajectoire est colorée en transparence. Le système $^\circ A^{L\ominus}$ n'a pas pu être traité.

mannoside HM. De même, les états A non-liés sont particulièrement stables sous leur forme ouverte (op), également constatée dans les structures.

En revanche, une divergence nette avec les structures natives prédites par cristallographie est observée pour les autres systèmes :

- $^{\circ}\text{S}^{\text{L}}$ présente un minimum énergétique dans l'état ouvert ($\omega \approx 0,8$) alors qu'on l'observe clos dans les structures cristallines
- de même $\bullet\text{A}^{\text{L}}$ et $\bullet\text{A}^{\text{LP}}$ présentent un site de liaison ouvert alors qu'il est fermé dans les co-cristaux avec HM

Afin d'étudier de façon plus quantitative les différences de populations entre chaque état d'ouverture (cl, so et op), nous calculons les probabilités par intégration du PEL obtenu

$$P(\omega_a < \omega < \omega_b) = \frac{\int_{\omega_a}^{\omega_b} d\omega \exp[-\beta F(\omega)]}{\int_{\omega_{\min}}^{\omega_{\max}} d\omega \exp[-\beta F(\omega)]} \quad (40)$$

où ω_a et ω_b sont les bornes inférieure et supérieure de l'état considéré et ω_{\min} , ω_{\max} les bornes où le PEL est connu. Les frontières séparant les états d'ouverture sont définies par les maxima locaux autour des valeurs $\omega = \{0,1;0,5\}$ pour S et $\omega = \{0,2;0,6\}$ pour A. La TABLE 8 présente les populations estimées pour chaque système.

État	P(cl) [%]	P(so) [%]	P(op) [%]
$\bullet\text{S}^{\text{L}}$	75,5	12,0	12,5
$\bullet\text{A}^{\text{L}}$	0,2	1,1	98,7
$\bullet\text{A}^{\text{LP}}$	5×10^{-2}	1,1	98,9
$^{\circ}\text{S}^{\text{L}}$	11,5	9,6	78,9
$^{\circ}\text{A}^{\text{L}}$	7×10^{-3}	0,1	99,9
$^{\circ}\text{A}^{\text{LP}}$	8×10^{-4}	0,5	99,5
$\bullet\text{S}^{\text{L}\ominus}$	88,7	9,3	2,0
$\bullet\text{A}^{\text{L}\ominus}$	0,3	0,6	99,1
$\bullet\text{A}^{\text{LP}\ominus}$	1×10^{-3}	2×10^{-2}	100
$^{\circ}\text{S}^{\text{L}\ominus}$	6,3	45,9	47,8
$^{\circ}\text{A}^{\text{L}\ominus}$	-	-	-
$^{\circ}\text{A}^{\text{LP}\ominus}$	1×10^{-3}	0,1	99,9

TABLE 8 – Probabilité des états d'ouverture.

Ces résultats suggèrent une certaine plasticité conformationnelle à l'équilibre des deux états allostériques du domaine lectine. En particulier, le système $\bullet\text{S}^{\text{L}}$ peuple de manière significative ($\sim 25\%$) des états ouverts (op) et semi-ouverts (so), en complément de l'état clos (cl) majoritaire ($\sim 75\%$) caractérisé dans les structures cristallographiques. De façon plus surprenante, le même système en l'absence de ligand inverse la tendance avec une majorité de configurations op ($\sim 79\%$).

Le même phénomène est observé pour les états A^L et A^{LP} , avec une multiplication par deux ordres de grandeur de la population cl en présence du ligand. Le modèle suggère donc un mécanisme d'ajustement induit avec un déplacement d'équilibre et une stabilisation de l'état clos par complexation aux mannosides. Néanmoins, contrairement à l'image fournie par les structures cristallographiques $\bullet A^{LP}$, la forme cl ajustée au ligand n'est pas prédominante dans nos prédictions.

En outre, il est intéressant de constater une différence systématique de stabilisation de l'état cl entre les systèmes A^L et A^{LP} . La présence physique du domaine piline déstabilise sensiblement l'état cl d'un ordre de grandeur. Cette observation est à relier à la plasticité conformationnelle de la région inter-domaines permise par l'absence de ce dernier et suggère un couplage faible mais significatif entre le site de liaison et la région inter-domaines avant toute réorganisation du renflement bêta/switch alpha. Une interprétation intuitive est que l'immobilisation de la boucle pivot dans une position contrainte par le domaine piline exerce une « tension » défavorisant la fermeture du site de liaison, et que cette « tension » est partiellement relâchée lors de la séparation des deux domaines.

Effet d'une force externe

Le profil d'énergie libre en l'absence de force nous donne une indication sur les propriétés d'équilibre de la protéine libre en solution. Cependant, en conditions physiologiques, la protéine liée à son substrat peut se retrouver soumise à une contrainte mécanique. Nous pouvons modéliser l'action d'une force extérieure F appliquée au grappin moléculaire selon une coordonnée $z(x)$ par un potentiel linéaire venant perturber l'hamiltonien

$$U_F(x) = U_0(x) - Fz(x) \quad (41)$$

où z peut représenter, par exemple, la distance entre les deux points d'application de la force projetée sur la direction de la force.

En supposant une force constante, la surface d'énergie potentielle U_F peut être échantillonnée selon sa distribution de BOLTZMANN grâce aux outils de la MD traditionnelle, en implémentant *in silico* le biais $-Fz(x)$ représentant, par exemple, l'effet de la traction de la bactérie par le fluide environnant et transmise par le pilus jusqu'au complexe entre FimH et un glycan épithélial, ou encore l'action d'un microlevier dans une expérience AFM à force constante.

Afin d'analyser l'effet de l'application d'une force de tension sur le site de liaison nous calculons le profil d'énergie libre selon la coordonnée d'ouverture en présence d'un biais appliqué sur la distance

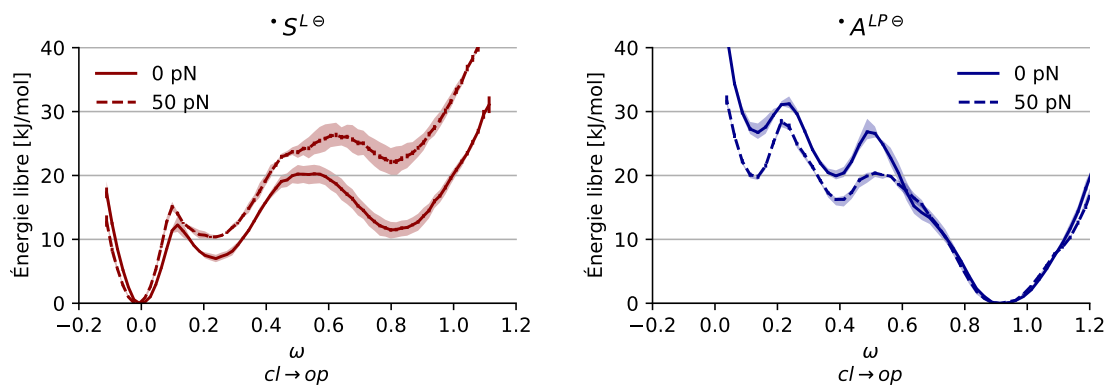
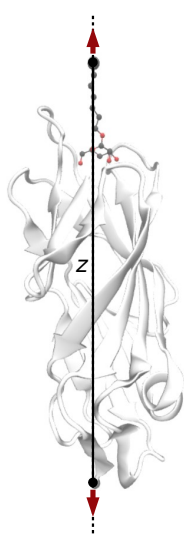


FIGURE 30 – Altération du profil d'énergie libre sous force.



z entre le carbone terminal de la queue heptyle du HM (C13) et le carbone carboxylique de l'extrémité C-terminale du système (Thr158 pour le domaine lectine seul, et Lys14 du peptide DsG pour la protéine entière). Une biais linéaire de pente $F = 30 \text{ kJmol}^{-1}\text{nm}^{-1} \approx 50 \text{ pN}$ est appliqué à z de façon à exercer une tension constante sur le complexe protéine-ligand. Cette valeur correspond au régime intermédiaire où le mécanisme accrocheur de FimH commence à apparaître *in vitro* [7]. La force augmente la propension du complexe à se dissocier, mais la contrainte DBC permet de maintenir le ligand à proximité du site de liaison. Une seule phase de 200 ns de BEUS a été réalisée pour ces calculs et seules les 150 ns terminales sont incluses dans l'analyse. Les PEL sous force pour les systèmes $\bullet S^{L\ominus}$ et $\bullet A^{LP\ominus}$ sont présentés en FIGURE 30.

Pour les deux systèmes étudiés, une stabilisation d'environ 5 à 10 kJmol^{-1} de l'état clos par rapport à l'état ouvert est observée. Cela suggère que dans l'état lié, la présence d'une force de tension stabilise la conformation cl, considérée comme la plus affine pour le ligand. Un tel mécanisme est remarquable car il constitue la base fonctionnelle d'une liaison accrocheuse à l'échelle du site de liaison. Si la vitesse de dissociation (k_{off}) du complexe est faible devant la dynamique d'ouverture, et si la vitesse de dissociation est plus grande dans les formes op que cl, alors les PEL calculés prédisent effectivement un comportement de liaison accrocheuse « locale » au sein même des deux états allostériques A et S, sans réorganisation de la région inter-domaines.

La TABLE 9 compare les populations calculées pour chaque état d'ouverture avec et sans force. On constate qu'une force modérée de 50 pN est suffisante pour faire passer la proportion d'état ouvert $P(\text{op})$ de 2 à 0,03 %

L'influence de la force sur les populations des états d'ouverture est facile à interpréter à partir des simulations non-biaisées. Nous consta-

État	F [pN]	P(cl) [%]	P(so) [%]	P(op) [%]
•S ^L ⊖	0	88,7	9,3	2,0
•S ^L ⊖	50	96,8	3,2	3 × 10 ⁻²
•A ^{LP} ⊖	0	1 × 10 ⁻³	2 × 10 ⁻²	100
•A ^{LP} ⊖	50	1 × 10 ⁻²	0,1	99,9

TABLE 9 – Probabilité des états d'ouverture sous force.

tons en effet que la distance z moyenne est plus grande dans l'état cl que dans l'état op d'environ $\Delta z = 3 \text{ \AA}$. Le travail réversible fourni par une force constante $F = 30 \text{ kJmol}^{-1} \text{ nm}^{-1}$ selon z lors de la transition $\text{op} \rightarrow \text{cl}$ est donc d'environ $W_F = F\Delta z \approx 9 \text{ kJmol}^{-1}$, en accord avec la différence observée sur les PEL FIGURE 30. Plus rigoureusement, la différence d'énergie libre $\Delta G_F(\omega)$ liée à l'introduction du potentiel externe peut s'écrire à une constante près grâce à l'équation de ZWANZIG

$$\begin{aligned} \Delta G_F(\omega) &= -k_B T \ln \langle \exp[-\beta \Delta U_F] \delta(\omega(\mathbf{x}) - \omega) \rangle_0 \\ &= -k_B T \ln \langle \exp[-\beta \Delta U_F] \rangle_{0,\omega} \end{aligned} \quad (42)$$

où $\Delta U_F = -Fz$ et où la moyenne est prise dans l'ensemble non-perturbé. Enfin, le développement en cumulants du logarithme de la moyenne donne

$$\Delta G_F(\omega) = \langle \Delta U_F \rangle_{0,\omega} + \frac{\beta}{2} (\langle \Delta U_F^2 \rangle_{0,\omega} - \langle \Delta U_F \rangle_{0,\omega}^2) + \dots \quad (43)$$

où nous retrouvons à l'ordre (1), $\Delta G_F(\omega) \approx \langle -Fz \rangle_{0,\omega}$. La figure 31 compare l'altération du PEL sous force telle qu'obtenue à partir des simulations BEUS à la prédiction par repondération des trajectoires BEUS en l'absence de force. Malgré une divergence notable pour des valeurs de ω intermédiaires dans l'état •S^L⊖, la tendance générale est reproduite et on obtient un bon accord pour les valeurs extrêmes $\omega \approx 0$ ou $\omega \approx 1$, avec un $\Delta G_F(1) - \Delta G_F(0)$ avoisinant 8–10 kJmol⁻¹.

Notons que ces résultats ont été obtenus sur deux systèmes monoprotonés, vraisemblablement très minoritaires aux pH physiologiques. Néanmoins, il est très probable que des résultats équivalents soient observés avec leurs équivalents diprotonés. En effet les profils $\Delta G_F(\omega)$ prédits pour ces systèmes sont similaires à ceux présentés ici (FIGURE B.6).

4.0.4 Conclusion

Dans cette section nous avons utilisé une méthode d'échantillonnage renforcé, l'échantillonnage parapluie à échange de biais (BEUS), combinée avec une approche adaptative, afin d'obtenir une estima-

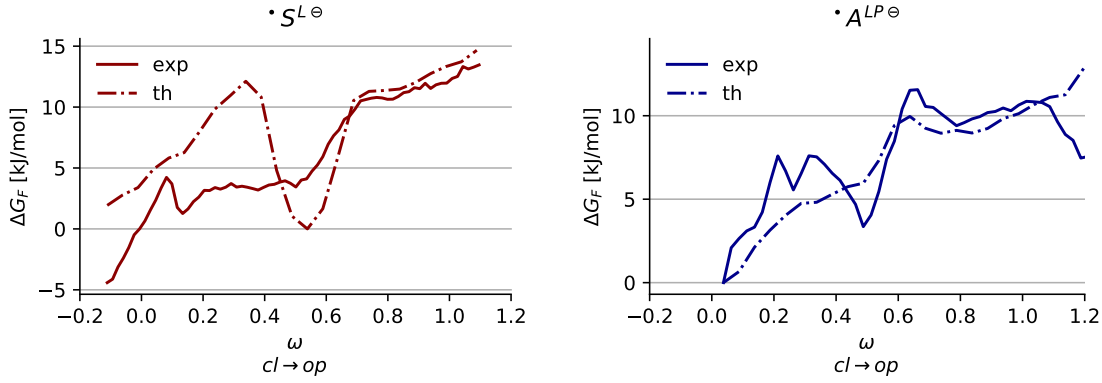


FIGURE 31 – Prédiction de l'effet de la force à partir des simulations non-perturbées. L'altération théorique du PEL sous force ($F = 50$ pN) est calculée (ÉQUATION 42) par repondération des trajectoires en l'absence de force (th) et comparée à la référence obtenue par simulations BEUS biaisées (exp). Le minimum est arbitrairement fixé à $\Delta G_F = 0$.

tion robuste du profil d'énergie libre du domaine lectine (PEL) le long de la coordonnée d'ouverture ω définie précédemment. Cette analyse a été menée sur les deux états allostériques métastables, S et A, et pour ce dernier, en l'absence ou présence du domaine piline associé. En raison des vitesses de séparation inter-domaines ou d'interconversion $S \leftrightarrow A$, très faibles par rapport à la durée de simulation (~ 500 ns), les PEL ainsi calculés correspondent à des pseudo-équilibres locaux pour la dynamique d'ouverture-fermeture du site de liaison.

Les PEL confirment l'existence de trois ensembles d'états d'ouverture du site de liaison séparés par des barrières énergétiques : les états fermés (cl) et ouvert (op), connus des structures publiées dans la PDB, ainsi que des états intermédiaires semi-ouverts (so) non caractérisés jusqu'alors, à notre connaissance. En dehors d'un système ($^\circ S^{L\ominus}$), l'état le plus stable prédit est alternativement cl ou op.

Nos calculs présentent un certain nombre d'accords et de divergences avec les structures natives suggérées par la cristallographie et la RMN :

- le domaine lectine isolé complexé au mannoside ($^\bullet S^L$) et la protéine entière FimH·DsG non-ligantée ($^\bullet A^{LP}$) possèdent des minima énergétiques cohérents avec les structures, respectivement cl et op
- en revanche, le domaine isolé libre ($^\circ S^L$) et la protéine entière liée ($^\bullet A^{LP}$) possèdent des minima énergétiques qui ne correspondent pas aux structures cristallines (op ou so au lieu de cl).

Une telle divergence peut avoir de multiples origines. Il est possible que les artefacts liés aux conditions de cristallisation et à l'empilement cristallin favorisent les états cl au sein du cristal, et que l'état représentatif en solution soit effectivement op. Les petites molécules flexibles sont connues pour cristalliser avec des distorsions conforma-

tionnelles pouvant atteindre 20 kJmol^{-1} [8]. Cependant, les cristaux de protéines présentent des interactions intermoléculaires plus limitées et localisées, ce qui rend de fortes déformations par rapport aux conditions diluées peu probables. La structure du squelette protéique obtenue par cristallographie aux rayons X est généralement perçue comme représentative de la structure native en solution, notamment par l'excellente adéquation avec les structures élucidées par RMN [9]. Outre ce constat général, l'ensemble de structures RMN du domaine lectine isolé dans l'état apo (VANWETSWINKEL et al., (2014), [10]) indique un agrément avec les structures cristallines correspondantes, avec site de liaison fermé ($^{\circ}\text{S}^{\text{L}}$). En ce qui concerne la protéine entière (A^{LP}), bien que nous ne disposions pas d'étude RMN, une différence de plus de 25 kJmol^{-1} en défaveur de la structure cristalline nous paraît pour le moins étonnante.

La deuxième source de désaccord, plus probable, réside dans le paramétrage des champs de force de la protéine et du ligand. Les champs de forces actuels sont connus pour dévier des structures cristallographiques et RMN sur les temps longs et sont plus que suspectés de ne pas toujours reproduire avec fidélité le minimum d'énergie libre du système [11, 12]. L'utilisation de contraintes *ad hoc* pour maintenir la protéine dans le puits énergétique expérimental pourrait même conduire à des ensembles conformationnels et à une dynamique plus réalistes et en meilleure adéquation avec les quantités expérimentales mesurables (facteurs B, déplacements chimiques et temps de relaxation RMN, etc.) [13].

Pour finir, la dernière limitation susceptible de nuancer l'interprétation de nos travaux est d'ordre méthodologique. Le succès de la méthode d'échantillonnage utilisée repose sur un certain nombre d'hypothèses difficiles à vérifier quantitativement. Un choix sous-optimal de coordonnée de réaction limite la diffusion longitudinale des répliques du système entre les divers états métastables et peut conduire à une estimation erronée de leurs populations relatives. Même lorsque la coordonnée est optimale (au sens où elle constitue le mode le plus lent d'évolution du système aux échelles de temps considérées), la présence de coordonnées transverses relativement lentes est également un frein à la convergence du PEL. Nous mitigeons l'impact de ces potentielles lacunes grâce à une approche US permettant un meilleur échange d'information entre les différentes fenêtres le long de la CV (BEUS) et une initialisation au plus proche de l'équilibre capitalisant sur les simulations REST_2 précédentes.

Sans exclure la possibilité d'un biais systématique en faveur de l'état ouvert (op) pour les différents systèmes étudiés, nous pouvons toujours interpréter les résultats dans leurs aspects relatifs et semi-quantitatifs. La comparaison des systèmes liés et non-liés indique une stabilisation d'environ 10 kJmol^{-1} de l'état fermé (cl) par la présence

du ligand HM, compatible avec un modèle d'ajustement induit suggéré par les structures cristallines. Un élément nouveau apporté par notre analyse concerne la stabilisation d'environ 10 kJmol^{-1} supplémentaires de l'état fermé en présence d'une force de tension de 50 pN s'exerçant sur le ligand. À notre connaissance, il n'est pas fait mention d'une telle possibilité dans la littérature. Ces données vont dans le sens d'un mécanisme de liaison accrocheuse à deux échelles, celle de la dynamique d'ouverture du site de liaison d'une part et celle de la dissociation inter-domaines et de la transition allostérique $S \leftrightarrow A$ d'autre part. Il est possible que ce mécanisme de grappin moléculaire « local » permette de prolonger la durée de vie du complexe $\bullet A^{LP}$ le temps que le grappin moléculaire « global », dont la cinétique est plus lente, se mette en place. Une analyse plus détaillée de la dynamique d'interconversion entre les divers états d'ouvertures, allostériques, et leurs vitesses de dissociations associées est requise pour préciser la pertinence biologique de ce phénomène.

RÉFÉRENCES

- [1] G. TORRIE et J. VALLEAU. Nonphysical sampling distributions in Monte Carlo free-energy estimation : Umbrella sampling. *Journal of Computational Physics*, **23**, 187-199 (1977).
DOI : [10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8)
- [2] M. MEZEL. Adaptive umbrella sampling : Self-consistent determination of the non-Boltzmann bias. *Journal of Computational Physics*, **68**, 237-248 (1987).
DOI : [10.1016/0021-9991\(87\)90054-4](https://doi.org/10.1016/0021-9991(87)90054-4)
- [3] J. KÄSTNER. Umbrella sampling : Umbrella sampling. *Wiley Interdisciplinary Reviews : Computational Molecular Science*, **1**, 932-942 (2011).
DOI : [10.1002/wcms.66](https://doi.org/10.1002/wcms.66)
- [4] Y. I. YANG, Q. SHAO, J. ZHANG, L. YANG et Y. Q. GAO. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics*, **151**, 070902 (2019).
DOI : [10.1063/1.5109531](https://doi.org/10.1063/1.5109531)
- [5] S. KUMAR, J. M. ROSENBERG, D. BOUZIDA, R. H. SWENDSEN et P. A. KOLLMAN. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, **13**, 1011-1021 (1992).
DOI : [10.1002/jcc.540130812](https://doi.org/10.1002/jcc.540130812)
- [6] A. GROSSFIELD. WHAM : the weighted histogram analysis method. Version 2.0.9
- [7] O. YAKOVENKO, S. SHARMA, M. FORERO, V. TCHESNOKOVA, P. APRIKIAN, B. KIDD, A. MACH, V. VOGEL, E. SOKURENKO et W. E. THOMAS. FimH Forms Catch Bonds That Are Enhanced by Mechanical Force Due to Allosteric Regulation. *Journal of Biological Chemistry*, **283**, 11596-11605 (2008).
DOI : [10.1074/jbc.M707815200](https://doi.org/10.1074/jbc.M707815200)
- [8] H. P. G. THOMPSON et G. M. DAY. Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chem. Sci.*, **5**, 3173-3182 (2014).
DOI : [10.1039/C4SC01132E](https://doi.org/10.1039/C4SC01132E)
- [9] L. J. BANASZAK. Properties of Biomacromolecules in the Crystalline State. *Foundations of Structural Biology*. Elsevier, 2000. 50-60. DOI : [10.1016/B978-012077700-6/50004-3](https://doi.org/10.1016/B978-012077700-6/50004-3)
- [10] S. VANWETSWINKEL, A. N. VOLKOV, Y. G. J. STERCKX, A. GARCIA-PINO, L. BUTS, W. F. VRANKEN, J. BOUCKAERT, R. ROY, L. WYNS et N. A. J. van NULAND. Study of the Structural and Dynamic Effects in the FimH Adhesin upon α -D-Heptyl Mannose Binding. *Journal of Medicinal Chemistry*, **57**, 1416-1427 (2014).
DOI : [10.1021/jm401666c](https://doi.org/10.1021/jm401666c)
- [11] A. RAVAL, S. PIANA, M. P. EASTWOOD, R. O. DROR et D. E. SHAW. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins*, **80**, 2071-2079 (2012).
DOI : [10.1002/prot.24098](https://doi.org/10.1002/prot.24098)

- [12] O. F. LANGE, D. van der SPOEL et B. L. de GROOT. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophysical Journal*, **99**, 647-655 (2010).
DOI : [10.1016/j.bpj.2010.04.062](https://doi.org/10.1016/j.bpj.2010.04.062)
- [13] Y. XUE et N. R. SKRYNNIKOV. Ensemble MD simulations restrained via crystallographic data : Accurate structure leads to accurate dynamics. *Protein Science : A Publication of the Protein Society*, **23**, 488-507 (2014).
DOI : [10.1002/pro.2433](https://doi.org/10.1002/pro.2433)

Troisième partie

STABILITÉ DE LA LIAISON ACCROCHEUSE

Cette deuxième partie de résultats est consacrée à l'étude de la stabilité du complexe formé entre FimH et les mannosides. Le chapitre 5 adopte les outils de la thermodynamique à l'équilibre pour tenter de quantifier l'affinité de divers états de la protéine pour son ligand. Le chapitre 6 se place sous l'angle hors d'équilibre et s'attache à caractériser la cinétique de dissociation en présence et en absence d'une force.

MODULATION DE L’AFFINITÉ

Les chapitres précédents ont été consacrés à l’exploration qualitative et quantitative des différentes conformations adoptées par FimH. Comment cette plasticité conformationnelle – en particulier selon les axes identifiés d’allostérie et d’ouverture du site de liaison – impacte-t-elle la stabilité du complexe formé avec un ligand mannosidique ?

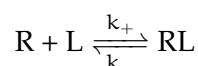
Dans ce chapitre nous explorons la question en essayant de quantifier l’affinité de différents états du domaine lectine pour notre ligand modèle, le heptylmannose (HM). Nous mettons en œuvre une approche fondée sur des calculs d’énergie libre de liaison par méthode alchimique (AFEP, *alchemical free energy perturbation*). Ces calculs sont connus pour leur complexité de mise en place, leur difficile convergence, leur sensibilité au champ de force et aux détails du protocole choisi, ainsi que pour leurs aspects parfois contre-intuitifs qui mènent à des erreurs faciles [1-6]. Ils ont cependant montré de nombreux succès et parviennent dans les meilleurs cas à reproduire les valeurs expérimentales à quelques kJ mol^{-1} près [7].

Nous commençons par décrire brièvement les bases théoriques, puis présentons notre méthodologie et nos résultats préliminaires, obtenus intégralement avec les formes monoprotonnées du système (voir 3.2.1).

5.1 CALCULS ALCHIMIQUES D’ÉNERGIE LIBRE DE LIAISON

5.1.1 Enthalpie libre de liaison

Considérons la réaction de complexation suivante :



Une grandeur physique quantifiant la propension d’un système à former une liaison chimique est l’enthalpie libre standard de liaison $\Delta G_{\text{liais}}^\circ$. Elle est reliée à la constante d’équilibre, grandeur adimensionnée traduisant l’équilibre des concentrations entre réactifs et produits,

$$K_{\text{eq}} = \frac{[RL]_{\text{eq}}/C^\circ}{([R]_{\text{eq}}/C^\circ)([L]_{\text{eq}}/C^\circ)}, \quad (44)$$

Bien que la fonction d’état pertinente dans l’ensemble NpT soit l’enthalpie libre, ou énergie libre de GIBBS, notée G, nous utiliserons par simplicité le terme « énergie libre » de façon interchangeable.

où $C^\circ = 1 \text{ mol L}^{-1}$ désigne la concentration standard, par la relation

$$\Delta G_{\text{liais}}^\circ = -k_B T \ln K_{\text{eq}}. \quad (45)$$

Cette quantité présente deux propriétés remarquable :

- elle ne dépend pas des concentrations initiales d'un système en R, L ou RL,
- elle n'est reliée qu'indirectement aux vitesses de formation et de dissociation du complexe par $K_{\text{eq}} = C^\circ k_+ / k_-$.

Les valeurs de K_{eq} (ou de façon équivalente $\Delta G_{\text{liais}}^\circ$) sont obtenues expérimentalement par titrage. Plus K_{eq} est élevé (resp. $\Delta G_{\text{liais}}^\circ$ est négatif), plus la réaction d'association est favorable.

Les données expérimentales indiquent une différence significative d'énergie libre de liaison avec les mannosides entre le domaine lectine isolé (FimH^L) et la protéine entière complémentée ($\text{FimH}^{LP} \cdot \text{DsG}$) :

$$\begin{aligned} \Delta \Delta G_{\text{liais}}^\circ &= \Delta G_{\text{liais}}^\circ [\text{FimH}^L] - \Delta G_{\text{liais}}^\circ [\text{FimH}^{LP} \cdot \text{DsG}] \\ &\approx -20 \text{ kJ mol}^{-1}. \end{aligned} \quad (46)$$

Une telle différence correspond à une multiplication par 3300 de la constante d'équilibre K_{eq} .

Ces valeurs ont été interprétées à la lumière des structures PDB connues de FimH^L et $\text{FimH}^{LP} \cdot \text{DsG}$ en supposant généralement une unique conformation fermée (cl) du site de liaison en présence du mannoside. Les résultats des chapitres précédents montrent cependant une certaine flexibilité du degré d'ouverture du site, avec un complexe ouvert (op) relativement stable bien que non observé dans les structures cristallines. Si cette flexibilité est confirmée *in vitro* nous devons en conclure que l'affinité expérimentale résulte des contributions conjuguées de ces différents états d'ouverture.

Afin de quantifier la différence d'affinité entre plusieurs états d'ouverture et/ou allostériques du domaine lectine nous mettons en place des calculs d'énergie libre par transformation alchimique.

5.1.2 Approche alchimique par annihilation

Les calculs d'énergie libre par transformation dite « alchimique » reposent sur des transformations non physiques, mais possibles dans le monde virtuel des simulations, permettant de simplifier ou faciliter l'obtention du résultat voulu. Imaginons que nous souhaitions calculer le $\Delta G_{\text{liais}}^\circ$ à partir d'une simulation de dynamique moléculaire du complexe dans une boîte d'eau. Il suffirait en théorie d'une simulation suffisamment longue pour obtenir, après convergence, la distribution à l'équilibre des états lié et dissocié. Le $\Delta G_{\text{liais}} = -k_B T \ln(P_{\text{lié}}/P_{\text{dissocié}})$

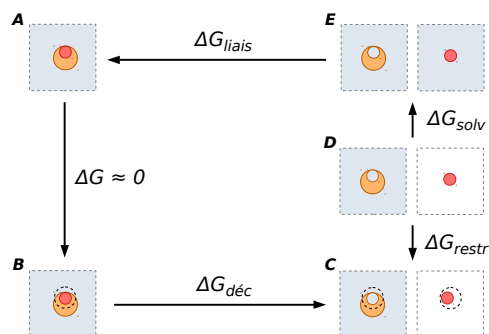


FIGURE 32 – Cycle thermodynamique pour la méthode alchimique.

obtenu permet de se ramener au $\Delta G_{\text{liais}}^{\circ}$ par correction liée aux concentrations standard. Les évènements de complexation et dissociation étant rares, il est très difficile de parvenir à converger le calcul dans un temps raisonnable.

L'énergie libre étant une fonction d'état, sa valeur ne dépend pas du chemin emprunté et nous pouvons décomposer la dissociation en plusieurs états thermodynamiques intermédiaires :

- l'état lié, défini par une fonction indicatrice $\mathbb{I}_{\text{lié}}$ similaire à celle utilisée au chapitre 3, ÉQUATION 26,
- l'état lié avec un potentiel à fond plat U_{restr} restreignant le ligand dans le volume du site de liaison, de façon à ce que le potentiel soit nul partout où $\mathbb{I}_{\text{lié}} = 1$,
- l'état où toutes les interactions entre le ligand et le système (récepteur-solvant) sont supprimées, mais le potentiel U_{restr} confine toujours le ligand dans un volume autour du site de liaison,
- l'état où le ligand et le reste du système n'interagissent pas, et le ligand est alors en phase gaz idéale à la pression standard,
- l'état où le ligand et le récepteurs sont solvatés indépendamment.

Le passage $A \rightarrow B$ correspond à un $\Delta G \approx 0$ puisque le potentiel U_{restr} est choisi suffisamment large pour ne pas travailler dans l'état lié. La transformation $B \rightarrow C$ correspond au découplage des interactions du ligand avec l'environnement du site de liaison, et s'accompagne d'un changement d'enthalpie libre $\Delta G_{\text{déc}}$. Le passage $C \rightarrow D$ correspond à l'enlèvement du potentiel U_{restr} , et s'accompagne d'un changement $-\Delta G_{\text{restr}}$. Pour finir, $D \rightarrow E$ est l'enthalpie libre de solvation du ligand ΔG_{solv} . Le cycle thermodynamique ainsi défini, et illustré en FIGURE 32, permet d'écrire :

$$\Delta G_{\text{liais}} = -\Delta G_{\text{déc}} + \Delta G_{\text{restr}} - \Delta G_{\text{solv}}. \quad (47)$$

Notre objectif n'étant pas de calculer l'enthalpie libre de liaison absolue ΔG_{liais} mais des enthalpies libres de liaison relatives (par exemple entre deux états du récepteur X et Y , $\Delta\Delta G_{\text{liais}}[X \rightarrow Y] = \Delta G_{\text{liais}}[Y] - \Delta G_{\text{liais}}[X]$) remarquons le fait que ΔG_{solv} ne dépend pas de l'état du récepteur et qu'il s'annule exactement dans la différence. De plus, pour un choix de U_{restr} faiblement dépendant des degrés de libertés internes du récepteur nous pouvons également supposer $\Delta\Delta G_{\text{restr}}[X \rightarrow Y] \approx 0$. Sous ces hypothèses nous avons donc :

$$\Delta\Delta G_{\text{liais}} \approx -\Delta\Delta G_{\text{déc}} \quad (48)$$

Les $\Delta G_{\text{déc}}$ peuvent être calculés par perturbation de l'énergie libre. En effet,

$$\begin{aligned} \Delta G_{\text{déc}} &= -k_B T \ln \frac{\int d\mathbf{x} \exp[-\beta(U_B - U_{R,L} - U_{L,s})(\mathbf{x})]}{\int d\mathbf{x} \exp[-\beta U_B(\mathbf{x})]} \\ &= -k_B T \ln \langle \exp[\beta(U_{R,L} + U_{L,s})] \rangle_B, \end{aligned} \quad (49)$$

où $\langle \cdot \rangle_B$ désigne la moyenne d'ensemble dans l'état B , et où nous avons noté $U_{R,L}$ et $U_{L,s}$ les énergies d'interaction récepteur-ligand et ligand-solvant, ainsi que

$$U_B = U_R + U_L + U_s + U_{R,L} + U_{R,s} + U_{L,s} + U_{\text{restr}} \quad (50)$$

l'énergie potentielle totale du système dans l'état B .

Une estimation numérique directe de l'ÉQUATION 49 est impossible, car les états B et C peuplent des espaces de configuration très différents. Les configurations de C ayant une probabilité négligeable dans B , elles ne sont en pratique jamais échantillonnées bien qu'elles contribuent massivement à la moyenne ($U_{R,L} + U_{L,s} \gg 0$). L'approche alchimique consiste à décomposer la transformation $B \rightarrow C$ en une série de petites transformations entre états thermodynamiques suffisamment proches pour avoir un bon recouvrement dans l'espace des phases. Concrètement, le potentiel d'interaction $U_{\text{int}} = U_{R,L} + U_{L,s}$ est dissocié en termes de Coulomb $U_{\text{int}}^{\text{coul}}$ et de van der Waals $U_{\text{int}}^{\text{vdW}}$ paramétrés par des paramètres λ^{coul} et λ^{vdW} entre 1 (couplé) et 0 (découplé).

La transformation passe ainsi par N états $i = 0 \dots N-1$ et la variation d'enthalpie libre pour la transformation totale $0 \rightarrow N-1$ correspond à la somme des variations pour chaque étape,

$$\Delta G_{\text{déc}} = \Delta G[0 \rightarrow 1] + \Delta G[1 \rightarrow 2] + \dots + \Delta G[N-2 \rightarrow N-1], \quad (51)$$

avec

$$\Delta G[i \rightarrow j] = -k_B T \ln \langle \exp[\beta(U_j - U_i)] \rangle_i \quad (52)$$

i	0 (B)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
λ^{coul}	1	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	0.50	0.45	0.40	0.35	0.30	0.25
λ^{vdW}	1							...								1

i	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
λ^{coul}	0.25	0.20	0.15	0.10	0.05	0					...					0
λ^{vdW}	1		...			1	0.975	0.950	0.925	0.900	0.875	0.850	0.825	0.800	0.775	0.750

i	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
λ^{coul}	0							...								0
λ^{vdW}	0.750	0.725	0.700	0.675	0.650	0.625	0.600	0.575	0.550	0.525	0.500	0.475	0.450	0.425	0.400	0.375

i	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60 (c)
λ^{coul}	0							...								0
λ^{vdW}	0.375	0.350	0.325	0.300	0.275	0.250	0.225	0.200	0.175	0.150	0.125	0.100	0.075	0.050	0.025	0

TABLE 10 – Paramètres des fenêtres pour le découplage alchimique.

L'état 0 correspond à l'état B pleinement couplé ($\lambda^{\text{coul}} = \lambda^{\text{vdW}} = 1$)
et l'état N – 1 correspond à l'état c découplé ($\lambda^{\text{coul}} = \lambda^{\text{vdW}} = 0$).

5.2 MÉTHODES

Nous appliquons la méthode du découplage alchimique afin d'estimer les enthalpies libres de liaison relatives entre différents états de la protéine. Précisons à nouveau que ces calculs ont été menés sur les systèmes monoprotônés (\ominus , voir 3.2.1), ce qui rend les résultats peu interprétables par rapport aux valeurs expérimentales, généralement tabulées à pH neutre et où la forme diprotônée est majoritaire. Nous en tirons néanmoins des leçons utiles pour des travaux futurs.

5.2.1 Protocole de découplage

Les simulations sont menées avec la fonctionnalité free-energy de GROMACS. Un total de 61 fenêtres est utilisé, les 21 premières correspondant au découplage des interactions coulombiques et les suivantes aux interactions de van der Waals (TABLE 10).

Chaque fenêtre est simulée séquentiellement, en prenant comme point de départ la configuration finale de la fenêtre précédente, en partant de l'état couplé. Pour une fenêtre i , le système est équilibré pendant 200 ps puis une trajectoire de 2 ns est générée pour l'échantillonnage. Les $\Delta U_{\text{int}} = U_j - U_i$ sont sauvegardés pour l'ensemble des paires $\{U_i, U_j\}$, avec $j \in 0 \dots 60$. Ces valeurs sont traitées avec l'estimateur MBAR (*Multistate Bennett Acceptance Ratio*) [8] pour estimer ΔG_{int} .

Notons que les termes d'énergie d'interaction *intra*-moléculaires du ligand (termes non-liants) sont également découplées durant le processus (option `couple-intramol = yes` de GROMACS). Ceci est fait

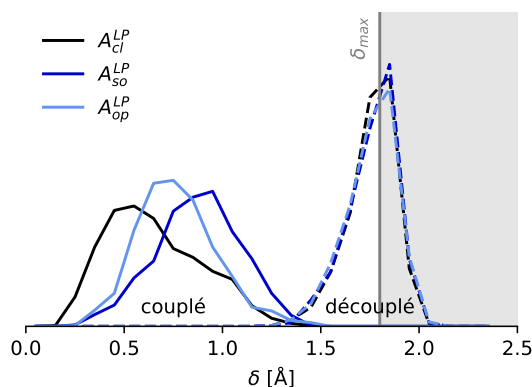


FIGURE 33 – Distribution de la DBC en début et en fin de découplage. Si les distributions diffèrent dans l’état couplé (*lignes continues*), elles sont quasiment confondues dans l’état découplé (*lignes en pointillé*). Trois exemples sont montrés pour les simulations des trois états d’ouverture avec les configurations provenant de BEUS A^{LP} .

dans l’objectif d’éviter que le ligand, une fois découplé de l’environnement protéine-solvant, ne se retrouve cinétiquement piégé dans des états métastables à cause d’interactions intramoléculaires. Des essais préliminaires indiquent en effet qu’une liaison hydrogène métastable peut se former de manière irréversible entre des fonctions alcool du mannoside, induisant une rupture de continuité dans la transformation et un mauvais recouvrement entre les fenêtres antérieures et postérieures à sa formation. L’état découplé n’est donc techniquement pas celui représenté en FIGURE 32C. Cela dit, pour le calcul des $\Delta\Delta G_{\text{liais}}$, la contribution d’enthalpie libre liée au recouplage des interactions intra-ligand en phase gaz s’annule exactement.

5.2.2 Potentiels de retenue

Pendant toute la durée du découplage, un potentiel de restrainte à fond plat basé sur la DBC et utilisé avec un mur semi-harmonique à $\delta_{\text{max}} = 1,8 \text{ \AA}$ et de constante $\kappa_{\delta} = 500 \text{ kJmol}^{-1} \text{ \AA}^{-2}$ (voir 3.3.2, ÉQUATION 22). Le potentiel est suffisamment large pour accommoder les fluctuations du ligand dans l’état couplé. Dans l’état découplé le ligand se déplace librement dans l’hypervolume délimité par le mur et de façon indépendante à la configuration du récepteur (FIGURE 33).

Dans une première série de simulations aucune restriction n’est appliquée aux degrés de liberté internes de la protéine. L’état d’ouverture du site de liaison n’étant pas contraint, nous constatons une transition d’ouverture partielle possible durant le découplage (par exemple $A_{\text{cl}}^{LP\ominus} \rightarrow A_{\text{so}}^{LP\ominus}$). Dans la série suivante, un potentiel à fond plat avec mur semi-harmonique est appliqué pendant le découplage

à la coordonnée ω afin de maintenir les états d'ouverture initiaux tout au long de la transformation alchimique. Les murs sont positionnés autour des barrières énergétiques du PEL obtenu au chapitre 4, $\omega = \{0,1;0,5\}$ pour S et $\omega = \{0,2;0,6\}$ pour A et une constante harmonique $\kappa = 4000 \text{ kJ mol}^{-1}$ est utilisée. De cette façon, le biais perturbe de façon minimale les fluctuations du récepteur dans un état d'ouverture donné, tout en empêchant les interconversions.

5.2.3 Sélection des configurations initiales

Nous avons expérimenté successivement plusieurs approches dans la sélection des points de départ des calculs de découplage. Nous distinguons trois séries :

- (Crist.) Dans ces simulations le découplage est fait directement à partir des structures cristallographiques des complexes. Les états d'ouverture du site sont donc initialement fermés (cl). Toutefois, la coordonnée ω n'est pas restreinte et nous observons une ouverture partielle (cl \rightarrow so) pour les systèmes A.
- (μ sMD) Dans ces simulations le découplage est fait à partir des structures cristallographiques après relaxation par 1 μ s de MD. Au cours de cette relaxation les systèmes A ont transitionné vers l'état semi-ouvert (so). L'ouverture du site n'est pas contrainte mais reste stable pendant le découplage.
- (BEUS) Ces simulations partent de structures issues des trajectoires BEUS pour chaque état d'ouverture. Une configuration est tirée aléatoirement dans la trajectoire de la fenêtre dont le biais est centré sur le minimum énergétique du profil selon ω ($\omega = 0$ pour S_{cl} , $\omega = 0,2$ pour S_{so} , $\omega = 0,8$ pour S_{op} , $\omega = 0,1$ pour A_{cl} , $\omega = 0,4$ pour A_{so}^{LP} , $\omega = 0,9$ pour A_{op}^{LP}). Dans cette série, nous utilisons des murs semi-harmoniques pour maintenir le système dans l'état d'ouverture souhaité. Afin de tester la reproductibilité d'un calcul, nous générons également pour certains systèmes une réplique « bis » (même configuration initiale mais graine aléatoire différente pour le thermostat) ou une autre condition initiale (tirée aléatoirement dans la trajectoire), notée « alt ».

5.2.4 Résultats

La TABLE 11 présente les $\Delta\Delta G_{\text{liais}}$ calculés pour l'ensemble des systèmes étudiés, en prenant pour référence le système le moins affin (μ sMD $A^{L\ominus}$). Ces enthalpies de liaisons relatives peuvent varier considérablement d'un système à l'autre, avec une différence maximale de $53,3 \text{ kJ mol}^{-1}$. Malgré la diversité des protocoles nous pouvons faire plusieurs constats.

Origine	Système	$\langle \omega \rangle_0$	$\langle \omega \rangle_{60}$	RMSD_{L_2, L_3} [Å]	$\Delta \Delta G_{\text{liais}}$ [kJ mol ⁻¹]	
crist.	$S^{LP\ominus*}$	cl	–	0.6	-43.8	◆
crist.	$S^{L\ominus}$	cl	–	0.7	-42.0	◆
crist.	$A^{LP\ominus}$	cl	so	0.6	-26.4	●
crist.	$A^{L\ominus}$	cl	so	0.5	-34.4	●
μsMD	$S^{LP\ominus*}$	cl	–	1.1	-36.8	◆
μsMD	$S^{L\ominus}$	cl	–	1.7	-41.4	◆
μsMD	$A^{LP\ominus}$	so	–	1.6	-16.4	●
μsMD	$A^{L\ominus}$	so	–	1.4	-0.0	○
BEUS	$S^{L\ominus}$	cl	–	2.7	-16.1	◆
BEUS	bis $S^{L\ominus}$	cl	–	2.8	-16.2	◆
BEUS	alt $S^{L\ominus}$	cl	–	1.9	-1.4	◆
BEUS	$A^{L\ominus}$	cl	–	1.3	-11.4	●
BEUS	bis $A^{L\ominus}$	cl	–	1.2	-29.4	●
BEUS	alt $A^{L\ominus}$	cl	–	1.0	-53.8	●
BEUS	$A^{LP\ominus}$	cl	–	1.0	-37.2	●
BEUS	$S^{L\ominus}$	so	–	2.4	-2.0	◆
BEUS	$A^{L\ominus}$	so	–	1.2	-26.0	●
BEUS	$A^{LP\ominus}$	so	–	1.1	-15.7	●
BEUS	$S^{L\ominus}$	op	–	3.1	-23.1	◆
BEUS	$A^{L\ominus}$	op	–	1.9	-8.6	●
BEUS	$A^{LP\ominus}$	op	–	2.4	-20.8	●

TABLE 11 – Énergies libres de liaison relatives calculées par annihilation du ligand. Pour chaque système, l'origine de la configuration initiale est indiquée (voir texte). « bis » : réplique avec configuration identique. « alt » : réplique avec une autre configuration.

*Modèle FimH^{LP}.DsG (F18) avec domaines séparés et lectine dans l'état S, construit par homologie à partir des structures PDB 4XOB, 4XOC, 4XOE.

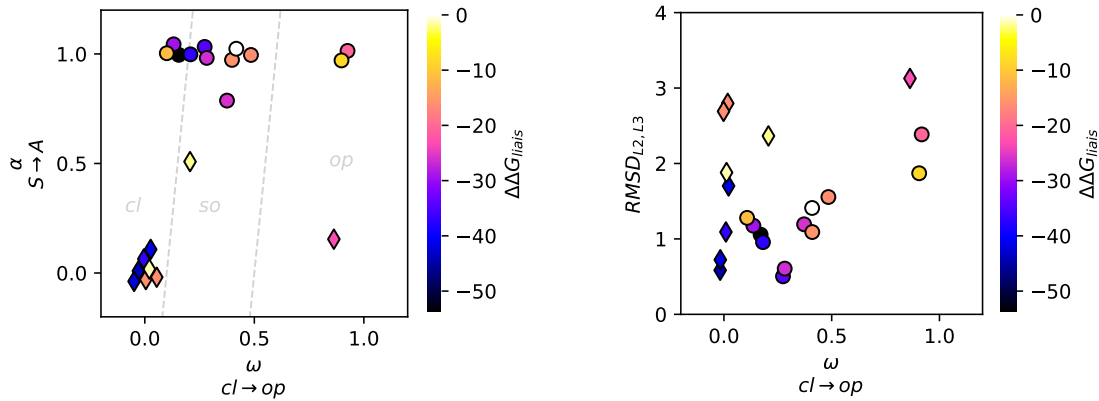


FIGURE 34 – Enthalpies libres de liaison relatives en fonction des paramètres conformationnels. Les symboles sont colorés en fonction du $\Delta \Delta G_{\text{liais}}$. Les ronds représentent l'état A et les losanges l'état S. (droite) Projection sur l'espace d'ouverture/allostérie (ω, α). (gauche) Projection sur l'espace ouverture/RMSD_{L₂,L₃}.

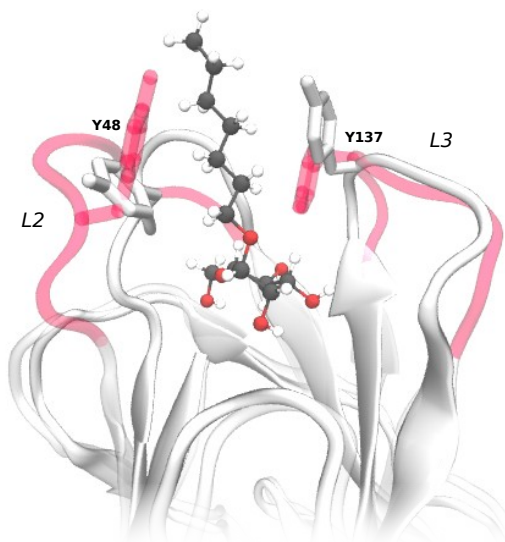


FIGURE 35 – Déviation des boucles L2L3 de leur conformation cristallographique. Nous montrons en exemple la configuration initiale du calcul BEUS S^L_{cl} (en transparence, boucles L2/L3 en rouge) superposée à la configuration cristallographique (cryst. S^L_{cl} , PDB 4XOC). Les chaînes latérales du portail tyrosine sont représentées. La déviation entre les deux conformations permet d’expliquer en partie la variabilité des $\Delta G_{déc}$ calculés.

Nous notons tout d’abord que les complexes les plus stables (avec $\Delta\Delta G_{liais} \approx -50$ à -30 kJ mol $^{-1}$) concernent des conformations fermées (cl). À l’inverse, les conformations ouvertes (op) et semi-ouvertes (so) sont généralement moins affines ($\Delta\Delta G_{liais} > -30$ kJ mol $^{-1}$). Dans la suite logique, les deux systèmes voyant une ouverture partielle du site pendant le découplage (cl \rightarrow so) aboutissent à des valeurs intermédiaires ($\Delta\Delta G_{liais} \approx -30$ kJ mol $^{-1}$). La FIGURE 34 (droite) permet de visualiser cette tendance sur l’espace joint des CV d’ouverture (ω) et d’allostérie (α).

Un ensemble de conformations cl font cependant exception et présentent des $\Delta\Delta G_{liais} > -30$ kJ mol $^{-1}$. Il s’agit de systèmes $S^{L\ominus}$ et $A^{L\ominus}$ issus de configurations BEUS. L’inspection des configurations initiales indique que les boucles bordant le site de liaison (L2-L3) s’y trouvent dans une conformation qui diverge de la structure cristallographique (FIGURE 35). Nous quantifions le RMSD moyen, calculé dans la première fenêtre de découplage, des boucles L2-L3 par rapport à leur conformation cristallographique (voir TABLE 11) et constatons que les ΔG_{liais} les plus bas sont généralement trouvés pour les systèmes où les boucles sont proches de la structure native ($RMSD_{L2L3} \leq 1,2$ Å). Des exceptions existent néanmoins (voir par exemple μ MD $S^{L\ominus}$). La FIGURE 34 (gauche) permet d’apprécier plus visuellement l’effet de l’ouverture du site et de la conformation des boucles sur l’affinité.

Finissons par souligner le manque de reproductibilité d'un calcul à l'autre : ni l'utilisation d'une configuration initiale identique (« bis ») ni d'une alternative obtenue par le même protocole (« alt ») ne garantissent d'obtenir des valeurs proches (voir en particulier les valeurs de BEUS $A^{L\ominus}$ pouvant varier de plus de 40 kJ mol^{-1}). Cette variabilité peut provenir de plusieurs sources, que nous discutons brièvement :

- L'incertitude liée à une taille d'échantillon finie, correspondant à l'incertitude statistique en supposant un échantillonnage parfaitement indépendant de chaque état thermodynamique. L'erreur standard évaluée par MBAR est dans notre cas inférieure à 1 kJ mol^{-1} , ce qui permet d'exclure que la taille d'échantillon soit un facteur majeur de variabilité.
- L'incertitude liée à un échantillonnage corrélé, traduisant la présence de degrés de liberté lents. Ces processus lents par rapport au temps d'acquisition des données dans chaque fenêtre entraînent un échantillonnage insuffisant (non-ergodique) des états successifs le long de la coordonnée alchimique. L'utilisation de MBAR mitige en partie le problème en évaluant la différence d'énergie libre entre deux points de la coordonnée thermodynamique λ de façon distribuée sur l'ensemble des fenêtres. Cela signifie que les configurations échantillonnées aux fenêtres i et $i + 1$ permettent d'évaluer non seulement $\Delta G[i \rightarrow i + 1]$ mais bénéficient également aux fenêtres voisines pourvu que leur recouvrement énergétique soit significatif. Cette condition est en partie vérifiée (voir des exemples de matrices de recouvrement FIGURE B.7) mais ne permet pas de compenser l'effet des degrés de libertés dont les temps caractéristiques de dé-corrélation excèdent la durée de quelques fenêtres d'acquisition ($\sim 10 \text{ ns}$), comme c'est le cas pour la conformation des boucles L2-L3.

Comparaison aux profils d'énergie libre d'ouverture

Nous disposons déjà d'une autre source d'information concernant les énergies de liaison relatives : les profils d'énergie libre selon la coordonnée d'ouverture présentés dans le chapitre 4. En effet, sur le cycle thermodynamique de la FIGURE 36 nous pouvons immédiatement déduire la relation

$$\begin{aligned}
 \Delta\Delta G_{\text{liais}}[X \rightarrow Y] &= \Delta G_{\text{liais}}[Y] - \Delta G_{\text{liais}}[X] \\
 &= G[\bullet Y] - G[^\circ Y] - (G[\bullet X] - G[^\circ X]) \\
 &= G[\bullet Y] - G[\bullet X] - (G[^\circ Y] - G[^\circ X]) \\
 &= \Delta G[\bullet X \rightarrow \bullet Y] - \Delta G[^\circ X \rightarrow ^\circ Y],
 \end{aligned} \tag{53}$$

où X et Y représentent deux états du récepteur (par exemple op et cl).

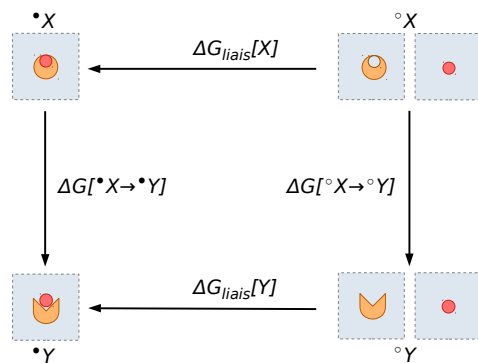


FIGURE 36 – Cycle thermodynamique reliant état d'ouverture et énergie de liaison.

Les valeurs relatives $\Delta G[\bullet X \rightarrow \bullet Y]$ et $\Delta G[^\circ X \rightarrow ^\circ Y]$ sont directement accessibles en utilisant les populations calculées en TABLE 8 par la formule

$$\Delta G[X \rightarrow Y] = -k_B T \ln \frac{P(Y)}{P(X)}. \quad (54)$$

Dans la TABLE 12 nous consignons les $\Delta\Delta G_{\text{liais}}$ en prenant pour référence l'état cl pour chaque PEL.

État	$\Delta\Delta G_{\text{liais}}[\text{cl} \rightarrow X] [\text{kJ mol}^{-1}]$		
	cl	so	op
S^L	0	4.2	9.3
A^L	0	2.7	8.6
A^{LP}	0	8.3	10.3
$S^{L\ominus}$	0	10.6	14.5
$A^{L\ominus}$	–	–	–
$A^{LP\ominus}$	0	4.8	-0.1

TABLE 12 – Énergies de liaison relatives extraites des profils d'ouverture.

À l'exception de $A^{LP\ominus}$, l'ensemble des systèmes s'accordent autour d'une différence d'énergie de liaison avoisinant 10 kJ mol^{-1} entre l'état ouvert et l'état fermé, l'état semi-ouvert obtenant une affinité intermédiaire. Ces résultats confirment qualitativement la tendance obtenue par AFEP, quoiqu'avec des valeurs moins extrêmes. Précisons qu'à la différence des résultats AFEP nous ne pouvons ici pas comparer l'énergie de liaison de deux systèmes distincts (par exemple S_{cl}^L et A_{cl}^L). L'énergie de découplage dans la méthode alchimique offre la référence nécessaire à une telle comparaison.

5.3 CONCLUSION

Les calculs d’énergie libre de liaison par la méthode alchimique d’annihilation du ligand nous ont permis de qualifier, si ce n’est quantifier avec précision, l’impact de la conformation du domaine lectine sur la thermodynamique de sa complexation avec le ligand HM. L’analyse méticuleuse de plusieurs calculs utilisant des protocoles légèrement différents (avec ou sans contrainte sur l’ouverture du site), et surtout des configurations initiales différentes, montre sans surprise que la configuration locale du site de liaison est déterminante pour son affinité avec les mannosides. Ainsi une conformation fermée du site se révèle une condition nécessaire – mais pas suffisante – à l’obtention des énergies de liaison les plus basses.

Nos conclusions mettent cependant en lumière l’existence de degrés de liberté relativement lents que nous avons négligés jusqu’alors dans notre analyse. Il s’agit de la flexibilité des boucles L2 et L3 bordant le site de liaison, incluant le portail à tyrosines (Tyr48-Tyr137). Bien que ces résidus n’interagissent pas directement avec le cycle mannoside, qui assure la spécificité de la liaison, ils forment des contacts hydrophobes stabilisants avec la chaîne aliphatique du HM et d’autres mannosides substitués. L’existence de plusieurs états métastables dont la durée de vie est du même ordre que le temps d’échantillonnage ($\approx 10^2$ ns) conduit à une dépendance de l’énergie libre de découplage avec la géométrie initiale des boucles et leur évolution stochastique le long de la transformation alchimique. Ce problème est bien connu dans la littérature sur les calculs d’énergie libre [5]. Plusieurs solutions peuvent être apportées, incluant des techniques d’échantillonnage avancé comme l’échange de répliques entre états thermodynamiques [9-11], ou l’utilisation de potentiels de retenue. De façon générale, la détermination systématique des coordonnées lentes du système est un problème ouvert, et nous devons bien souvent procéder itérativement par essai-erreur.

Il est intéressant de constater que l’état allostérique (A ou S), essentiellement déterminé par la géométrie du renflement bêta, semble avoir peu d’influence par rapport à l’état d’ouverture et la conformation des boucles L2-L3. Cette observation corrobore le modèle faiblement couplé, dans lequel l’effecteur allostérique (le domaine piline) influence indirectement l’affinité en modifiant les populations des états du site de liaison. Les gammes de $\Delta\Delta G_{\text{liais}}$ obtenues sont tout à fait compatibles avec la différence expérimentale de 20 kJ mol^{-1} observée entre l’affinité de FimH^L et celle de FimH^{LP}.DsG. L’ensemble des données suggère donc que l’équilibre $\text{op} \leftrightarrow \text{so} \leftrightarrow \text{cl}$ est prépondérant dans la détermination du ΔG_{liais} apparent, mais nous ne disposons pour l’instant pas de données suffisamment précises pour reconstituer un modèle complet, et ces travaux ayant été menés sur

les formes monoprotonnées de l'histidine 45 nous ne pouvons pas tirer de conclusion définitive sur la correspondance entre les valeurs calculées et les mesures expérimentales.

En conclusion, malgré des résultats préliminaires encourageants, l'approche présentée pourra être repensée selon plusieurs voies d'amélioration :

- l'identification et la maîtrise des degrés de liberté lents du site de liaison, en particulier la dynamique des boucles L2-L3,
- l'utilisation de la forme diprotinée, à priori majoritaire *in vivo* et expérimentalement,
- le calcul de l'enthalpie libre de solvation du ligand, qui permettrait d'obtenir des valeurs absolues pouvant également être comparée aux données expérimentales.

RÉFÉRENCES

- [1] Y. DENG et B. ROUX. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*, **113**, 2234-2246 (2009).
DOI : [10.1021/jp807701h](https://doi.org/10.1021/jp807701h)
- [2] E. GALLICCHIO et R. M. LEVY. Recent theoretical and computational advances for modeling protein–ligand binding affinities. *Advances in Protein Chemistry and Structural Biology*. T. 85. Elsevier, 2011. 27-80.
DOI : [10.1016/B978-0-12-386485-7.00002-8](https://doi.org/10.1016/B978-0-12-386485-7.00002-8)
- [3] P. V. KLIMOVICH, M. R. SHIRTS et D. L. MOBLEY. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design*, **29**, 397-411 (2015).
DOI : [10.1007/s10822-015-9840-9](https://doi.org/10.1007/s10822-015-9840-9)
- [4] R. SALARI, T. JOSEPH, R. LOHIA, J. HÉNIN et G. BRANNIGAN. A Streamlined, General Approach for Computing Ligand Binding Free Energies and Its Application to GPCR-Bound Cholesterol. *Journal of Chemical Theory and Computation*, **14**, 6560-6573 (2018).
DOI : [10.1021/acs.jctc.8b00447](https://doi.org/10.1021/acs.jctc.8b00447)
- [5] H. M. BAUMANN, V. GAPSYS, B. L. de GROOT et D. L. MOBLEY. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. *The Journal of Physical Chemistry B*, **125**, 4241-4261 (2021).
DOI : [10.1021/acs.jpcb.0c10263](https://doi.org/10.1021/acs.jpcb.0c10263)
- [6] E. DUBOUÉ-DIJON et J. HÉNIN. Building intuition for binding free energy calculations : Bound state definition, restraints, and symmetry. *The Journal of Chemical Physics*, **154**, 204101 (2021).
DOI : [10.1063/5.0046853](https://doi.org/10.1063/5.0046853)
- [7] V. GAPSYS, A. YILDIRIM, M. ALDEGHI, Y. KHALAK, D. van der SPOEL et B. L. de GROOT. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Communications Chemistry*, **4**, 1-13 (2021).
DOI : [10.1038/s42004-021-00498-y](https://doi.org/10.1038/s42004-021-00498-y)
- [8] M. R. SHIRTS et J. D. CHODERA. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, **129**, 124105 (2008).
DOI : [10.1063/1.2978177](https://doi.org/10.1063/1.2978177)
- [9] Y. MENG, D. SABRI DASHTI et A. E. ROITBERG. Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations. *Journal of Chemical Theory and Computation*, **7**, 2721-2727 (2011).
DOI : [10.1021/ct200153u](https://doi.org/10.1021/ct200153u)
- [10] S. WAN, G. TRESADERN, L. PÉREZ-BENITO, H. van VLIJMEN et P. V. COVENEY. Accuracy and Precision of Alchemical Relative Free Energy Predictions with and without Replica-Exchange. *Advanced Theory and Simulations*, **3**, 1900195 (2020).
DOI : [10.1002/adts.201900195](https://doi.org/10.1002/adts.201900195)

- [11] R. EVANS, L. HOVAN, G. A. TRIBELLO, B. P. COSSINS, C. ESTARELLAS et F. L. GERVASIO. Combining Machine Learning and Enhanced Sampling Techniques for Efficient and Accurate Calculation of Absolute Binding Free Energies. *Journal of Chemical Theory and Computation*, **16**, 4641-4654 (2020).
DOI : [10.1021/acs.jctc.0c00075](https://doi.org/10.1021/acs.jctc.0c00075)

CINÉTIQUE DE DISSOCIATION

Dans ce chapitre, nous quittons le monde des calculs à l'équilibre thermodynamique pour nous intéresser à la dynamique de dissociation du complexe entre FimH et un mannoside. Une telle étude requiert un nouveau paradigme d'utilisation des simulations de dynamique moléculaire où l'on s'intéresse aux propriétés temporelles de l'évolution du système, gouvernée par des événements rares.

La première section décrira comment l'occurrence de ces événements rares peut être accélérée grâce à la dynamique moléculaire dirigée (SMD, *steered molecular dynamics*). Nous appliquerons ensuite cette méthode à l'étude de la réaction de dissociation du complexe FimH-HM en reproduisant *in silico* des expériences typiques de microscopie à force atomique sous force constante et à vitesse de charge constante. Nous proposerons alors une approche bayésienne unifiée pour obtenir les paramètres cinétiques à partir de plusieurs types de simulations, à savoir un modèle de MARKOV dépendant de la force.

6.1 DYNAMIQUE MOLÉCULAIRE DIRIGÉE

La dynamique moléculaire dirigée (SMD, *steered molecular dynamics*) propose d'introduire un paramètre dépendant du temps, $\lambda(t)$ et contrôlant l'hamiltonien du système. Par exemple, il peut s'agir d'un potentiel de biais variable au cours du temps :

$$H[\mathbf{x}, \mathbf{p}, \lambda(t)] = H_0(\mathbf{x}, \mathbf{p}) + V[\mathbf{x}, \lambda(t)] \quad (55)$$

La SMD se fonde sur des apports expérimentaux et théoriques [1]. Expérimentaux, tout d'abord, avec le développement des techniques de spectroscopie à force atomique sur molécules uniques offrant un nouvel accès aux échelles microscopiques par la manipulation directe des objets d'étude. Une expérience fondatrice du domaine est notamment la mesure de la force de rupture d'un complexe individuel de biotine-streptavidine (FLORIN, MOY et GAUB, (1994) [2]), complexe d'une extrême stabilité, avec une demie-vie de plusieurs semaines en solution. L'emploi de techniques sous force permet alors d'étudier, aux échelles de temps du laboratoire, des événements moléculaires rares difficilement observables. L'interprétation de ces procédures expérimentales bénéficie alors naturellement de l'implémentation *in silico* de leurs équivalents.

Théoriques, en suite, car l'abandon des conditions de réversibilité pose un problème pour l'interprétation des simulations. L'obtention de propriétés thermodynamiques telles que les énergies libres de liaison, ou la forme d'un potentiel de force moyenne, s'effectue généralement dans le cadre des techniques à l'équilibre (*umbrella sampling*, perturbation de l'énergie libre). L'application d'une perturbation dépendant du temps rompt l'hypothèse d'équilibre quasi-statique et demande un cadre théorique adapté. L'identité dérivée par JARZYNSKI [3] établit un lien rigoureux entre propriétés d'équilibre et transformations hors d'équilibre

$$\exp(-\beta\Delta G) = \langle \exp(-\beta W) \rangle \quad (56)$$

où W est le travail total fourni au système pendant la transformation et la moyenne s'effectue sur l'ensemble des réalisations possibles de la transformation à partir de l'état initial, pris à l'équilibre. Cette approche a été appliquée avec un certain succès mais son utilisation demeure limitée, notamment en raison de sa difficile convergence statistique [4-6].

Une autre possibilité est d'utiliser la SMD non pas pour calculer des propriétés thermodynamiques mais cinétiques. Cette approche est à première vue beaucoup plus simple si l'on souhaite précisément calculer des constantes de vitesse en présence d'une perturbation extérieure comme l'application d'une force. En revanche, la reconstruction de ces propriétés en l'absence de perturbation est plus complexe et ne bénéficie pas de la formulation théorique exacte de la thermodynamique. Certains modèles ont néanmoins été proposés [7-9] et appliqués avec succès à des données simulatoires [10].

Afin de comparer les vitesses de dissociation du complexe formé entre FimH et le ligand HM dans plusieurs états conformationnels de la protéine, nous réalisons une série de simulations SMD. Nous nous focalisons sur les systèmes S^L et A^{LP} dans leurs conformations cl et op.

6.2 MODÈLE DE DISSOCIATION À UNE ÉTAPE

Ce protocole a été conçu dans l'hypothèse d'une réaction de dissociation en une seule étape, c'est-à-dire une seule barrière énergétique. L'objectif est de modéliser la résistance de la liaison au mannose pour divers états d'ouvertures de FimH caractérisés précédemment, et d'en tirer des paramètres permettant d'extrapoler les vitesses de dissociation en l'absence de force afin de les comparer aux données expérimentales.

6.2.1 Méthodes

Dissociation à vitesse de charge constante

Afin de modéliser l'effet d'une force de tension s'appliquant au complexe, nous définissons la variable collective d_z , projection sur l'axe z de la distance entre le carbone terminal de la queue du ligand heptylmannose (C13) et le centre de masse du domaine lectine (C α rés. 1–158). Le centre de masse du domaine lectine est privilégié par rapport à un atome du résidu terminal afin d'éviter que des changements conformationnels de la protéine tels que l'extension du linker ou, pour le système A^{LP} , la séparation des deux domaines, ne viennent interférer avec le processus de dissociation à l'étude. La projection sur z plutôt que la distance elle-même permet de conserver l'alignement du grand axe de la protéine avec la boîte de simulation rectangulaire des systèmes A^{LP} . Un biais harmonique mouvant est appliqué pendant la simulation

$$V(d_z, t) = \frac{1}{2} \kappa [d_z(t) - (d_z^0 + vt)]^2 \quad (57)$$

où κ est la constante harmonique, $d_z^0 \approx \langle d_z \rangle_0$ est le centre du biais harmonique à $t = 0$ et correspond à la distance d_z moyenne dans l'état initial, et v est la vitesse de déplacement du biais.

La force instantanée exercée sur la CV d_z peut s'exprimer

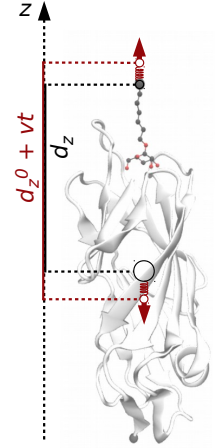
$$F(t) = -\frac{\partial V(d_z, t)}{\partial d_z} = -\kappa [d_z(t) - (d_z^0 + vt)] \quad (58)$$

Pour une structure parfaitement rigide ($d_z = \text{cte}$) ce dispositif revient à accroître linéairement la force à une vitesse de charge κv , exprimée en unité de force par unité de temps (pNs^{-1}). En pratique l'élasticité du complexe et son orientation entraînent l'observation d'une courte période transitoire et l'établissement rapide d'un régime linéaire.

Le biais est implémenté avec la fonction `MOVINGRESTRAINT` de la bibliothèque `PLUMED`. Les valeurs de d_z et F sont sauvegardées toutes les 10 ps pour l'analyse (2 ps pour $\kappa v = 5000 \text{ pNs}^{-1}$).

Configurations initiales

La réalisation d'une trajectoire de dynamique moléculaire est déterministe : une même configuration (\mathbf{x}, \mathbf{p}) de départ suivant la même transformation $\lambda(0) \rightarrow \lambda(t)$ devrait suivre exactement le même chemin dans l'espace des phases. L'observable macroscopique qui nous intéresse, le temps de dissociation moyen, doit donc être compris



comme une moyenne des temps de dissociations sur l'ensemble des conditions initiales, pondérées par leur poids canonique :

$$\bar{t}^* = \langle t^*(\mathbf{x}, \mathbf{p}) \rangle_0. \quad (59)$$

Nous utilisons les trajectoires de l'étude précédente (BEUS) afin d'échantillonner chaque état dans leur distribution d'équilibre local, et de tenir compte des hétérogénéités structurales pouvant impacter la vitesse de dissociation. Concrètement, les trajectoires correspondant au potentiel de biais le plus proche du minimum énergétique local de l'état concerné ($\omega = 0$ pour S_{cl}^L , $\omega = 0,8$ pour S_{op}^L , $\omega = 0,1$ pour A_{cl}^{LP} , $\omega = 0,9$ pour A_{op}^{LP}) sont échantillonnées aléatoirement (avec remise) après filtrage des configurations liées (ÉQUATION 26). Des ensembles de $N = 30$ configurations (positions atomiques du complexe) sont ainsi sélectionnés comme échantillons représentatifs des conditions initiales pour les trajectoires de SMD subséquentes.

Les structures sont orientées de façon à ce que l'axe $C13-C\alpha^{158}$ coïncide avec l'axe z de la boîte de simulation, resolvatées et équilibrées en suivant la procédure habituelle (voir 3.2.2). Dans ces expériences nous n'utilisons pas de biais restreignant la diffusion rotationnelle de la protéine, car celle-ci demeure spontanément alignée avec la direction de la force selon z .

Analyse des trajectoires

Afin d'interpréter les résultats des simulations nous traçons la force exercée par le biais en fonction du temps. Nous observons des trajectoires typiques des expériences de spectroscopie à force atomique sur molécule unique, avec une augmentation linéaire de l'amplitude de la force suivie d'un décrochage brutal lors de la rupture du complexe (FIGURE 37). Afin de réduire le bruit lié aux fluctuations de la force tout en conservant des pics bien définis, les données sont lissées avec un filtre médian de largeur 100 ps.

Nous définissons le temps de premier passage t^* de la dissociation lorsque la distance d_z dépasse le seuil $d_z > 4$ nm. Ce critère coïncide avec la chute brutale de la force suivant la rupture. La force F^* est définie comme le maximum de la force instantanée (lissée) dans les 500 ps précédant t^* .

Nous constatons parfois des irrégularités dans la progression linéaire de la force et l'évolution de d_z , traduisant de possibles événements de dissociation partielle. Dans un premier temps, ces événements seront ignorés et les données analysées avec un modèle de rupture à deux états et une seule barrière énergétique. Nous verrons par la suite que la prise en compte d'un intermédiaire de dissociation est plus à même de rendre compte des données expérimentales.

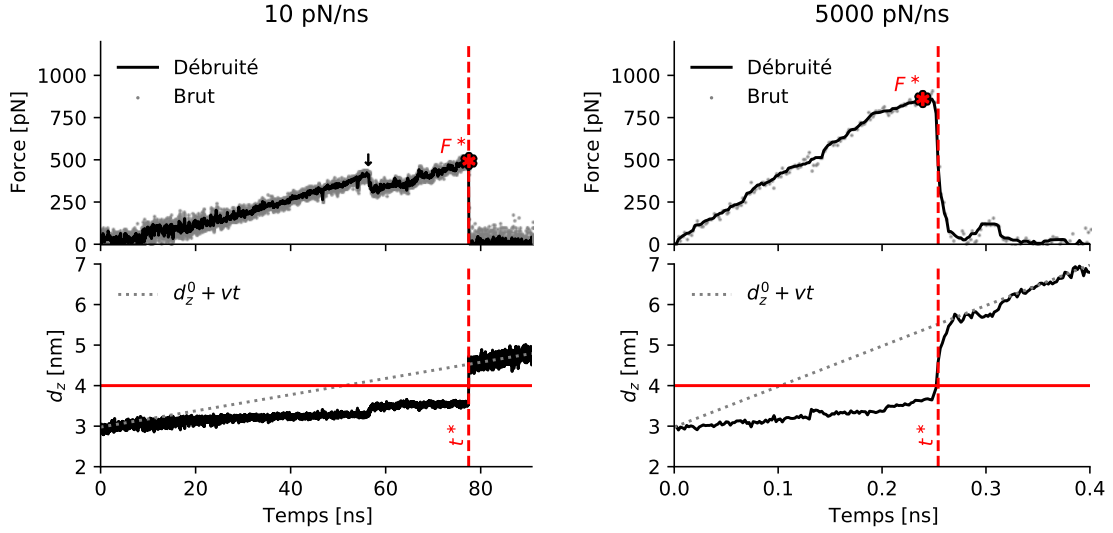


FIGURE 37 – Trajectoires typiques de simulations à vitesse de charge constante. Le déplacement du centre du biais (ligne grise en pointillé) entraîne l’augmentation linéaire de la force selon d_z , présentée pour les données brutes et après débruitage par filtre médian. Un évènement de dissociation au temps t^* est caractérisé par le premier passage $d_z > 4$ nm et la force de rupture F^* par le maximum précédant t^* . Deux exemples sont montrés pour une vitesse de charge de 10 pNns^{-1} (gauche) et 5000 pNns^{-1} (droite). Des évènements de rupture partielle sont souvent observables (\downarrow).

Analyse des paramètres cinétiques

Comme nous l’avons mentionné, l’analyse des expériences de molécule unique sous force dans l’optique d’en récupérer les paramètres cinétiques en l’absence de perturbation est délicate. Des formules analytiques ont été proposées suivant un certain nombre d’approximations [7-9].

Dans un premier temps, nous analyserons les résultats en faisant l’hypothèse d’un processus à une étape :



où la transition s’effectue de manière quasi-instantanée par rapport à la variation de F . On peut, dans ces conditions, exprimer la fonction de survie $S(t)$, probabilité que le complexe est toujours lié au temps t , c’est-à-dire

$$S(t) = \mathbb{P}(t^* > t). \quad (61)$$

La réaction de dissociation s’exprime par la relation

$$p^*(t) = -\frac{dS}{dt}(t) = k_{\text{off}}[F(t)]S(t) \quad (62)$$

où $p^*(t)$ est la densité de probabilité d'obtenir un évènement de rupture au temps t , qui s'intègre dans le cas général

$$S(t) = \exp \left(- \int_0^t k_{\text{off}}[F(t')] dt' \right) \quad (63)$$

et on obtient au final la distribution des temps de rupture

$$p^*(t) = k_{\text{off}}[F(t)] \exp \left(- \int_0^t k_{\text{off}}[F(t')] dt' \right). \quad (64)$$

Pour une force croissant de façon monotone et linéaire, on peut utiliser la relation $p^*(t)dt = p^*(F)dF$ pour exprimer la distribution des forces de rupture

$$p^*(F) = \frac{k_{\text{off}}(F)}{\dot{F}} \exp \left(- \int_0^F \frac{k_{\text{off}}(f)}{\dot{F}} df \right) \quad (65)$$

où on a noté \dot{F} la vitesse de charge.

Afin d'exprimer k_{off} nous faisons appel au modèle de BELL-EVANS (voir chapitre 1, p. 16) :

$$k_{\text{off}}(F) = k_{\text{off}}^0 \exp(\beta F \Delta x) \quad (66)$$

où k_{off}^0 est la vitesse de dissociation en l'absence de force et Δx représente la distance entre l'état lié et l'état de transition. Pour cette dépendance particulière de la vitesse de dissociation vis-à-vis de la force, nous obtenons la distribution de forces de rupture

$$p_{\text{BE}}^*(F) = \frac{k_{\text{off}}^0}{\dot{F}} \exp \left(\beta F \Delta x - \frac{k_{\text{off}}^0}{\beta \dot{F} \Delta x} [\exp(\beta F \Delta x) - 1] \right) \quad (67)$$

La forme de la distribution et sa dépendance vis-à-vis de la vitesse de charge sont illustrées en FIGURE 38. La moyenne et l'écart type de chaque distribution s'expriment en très bonne approximation [7, 11] :

$$\begin{aligned} \langle F \rangle &\approx \frac{1}{\beta \Delta x} \left(\ln \frac{\beta \dot{F} \Delta x}{k_{\text{off}}^0} + \gamma \right) \\ \sigma_{\dot{F}}^2 &\approx \frac{\pi^2}{6 \Delta x^2} \end{aligned} \quad (68)$$

$\gamma = 0,577\dots$

où γ est la constante d'EULER-MASCHERONI.

Il y a plusieurs approches pour estimer les paramètres k_{off}^0 et Δx à partir d'un ensemble de N trajectoires à vitesses de charge $\{\dot{F}_i\}_{i=1}^N$ donnant des forces de rupture $\{F_i^*\}_{i=1}^N$.

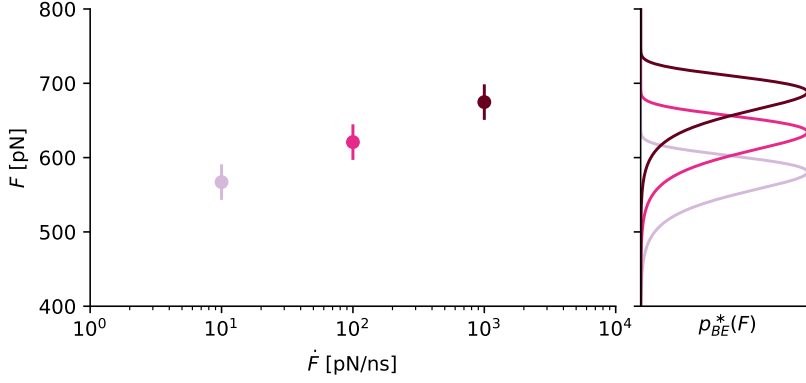


FIGURE 38 – Distributions théoriques des forces de rupture pour le modèle de BELL-EVANS. Sur cet exemple nous avons pris $k_{\text{off}}^0 = 7 \times 10^{-3} \text{ s}^{-1}$ et $\Delta x = 1,76 \text{ \AA}$, valeurs tirées de l'état haute affinité du modèle de THOMAS et al. [12].

Une méthode générale repose sur l'estimateur du maximum de vraisemblance (MV). La fonction de vraisemblance est la probabilité conditionnelle d'obtenir l'ensemble de réalisations (supposées indépendantes) $\{F_i^*\}_{i=1}^N$ sachant les vitesses de charge $\{\dot{F}_i\}_{i=1}^N$ et les paramètres inconnus $\theta \equiv (k_{\text{off}}^0, \Delta x)$:

$$\begin{aligned} \mathcal{L}(\theta) &= P(\{F_i^*\}_{i=1}^N | \{\dot{F}_i\}_{i=1}^N, \theta) \\ &= \prod_{i=1}^N p_{\theta}(F_i^* | \dot{F}_i) \end{aligned} \quad (69)$$

Le meilleur estimateur $\hat{\theta}$ pour le paramètre θ est donné par la maximisation de la fonction de vraisemblance

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta) \quad (70)$$

ou de manière équivalente, pour des raisons de stabilité numérique, son logarithme $\ln \mathcal{L}(\theta)$. En intégrant toute l'information disponible, c'est à dire pas seulement la moyenne des observations mais toute la distribution, la méthode du MV est versatile et permet d'ajuster des modèles complexes comportant plusieurs paramètres et des distributions de formes arbitraires.

Une autre approche consiste à ajuster l'équation de la moyenne $\langle F \rangle(\dot{F})$ (ÉQUATION 68) par la méthode des moindres carrés non-linéaires (MCNL). Dans ce cas, la dispersion des forces de rupture est interprétée comme du bruit gaussien, supposé uniforme pour tous les \dot{F} et n'apportant aucune information pour l'ajustement des paramètres (σ_F est ignoré pour l'estimation de Δx).

Dans la suite les résultats sont analysés selon les deux approches. Nous utilisons la bibliothèque Statsmodels [13] de Python pour l'analyse MV et la fonction `curve_fit` de Scipy [14] pour l'approche MCNL.

6.2.2 Résultats

Les forces de rupture obtenues pour l'ensemble des simulations à vitesse de charge constante sont présentées en FIGURE 39. Il est intéressant de constater que les forces de ruptures moyennes à différentes vitesses \dot{F} sont comparables pour des états d'ouverture similaires : par exemple, la force moyenne avoisine 400 pN à $\dot{F} = 10 \text{ pN s}^{-1}$ pour les états fermés (A_{cl}^{LP} et S_{cl}^L) tandis qu'elle est plus petite d'une centaine de pN pour les états ouverts (A_{op}^{LP} et S_{op}^L). Les valeurs des paramètres estimés sont résumées dans la TABLE 13. Les vitesses de dissociation à force nulle (k_{off}^0) sont significativement plus grandes d'un facteur ~ 4 pour un passage $cl \rightarrow op$. Pour l'estimation MV cette vitesse est influencée principalement par l'état d'ouverture et pas par l'état allostérique (valeurs pour A ou S dans les marges d'erreurs). Ce résultat conforme à l'intuition confirme que la vitesse de dissociation du complexe est influencée localement par le degré d'ouverture du site de liaison plus que par l'état allostérique S ou A (tel que nous l'avons défini par la conformation de l'ensemble inter-domaines/renflement bêta). Nous pouvons donc supposer que la transition allostérique influence indirectement la vitesse de dissociation en modulant la stabilité relative des états d'ouverture.

État	MV		MCNL	
	$k_{off}^0 [10^6 \text{ s}^{-1}]$	$\Delta x [\text{\AA}]$	$k_{off}^0 [10^6 \text{ s}^{-1}]$	$\Delta x [\text{\AA}]$
S_{cl}^L	1,4(5)	0,46(2)	0,5(2)	0,60(5)
A_{cl}^{LP}	0,8(2)	0,48(2)	0,14(6)	0,63(3)
S_{op}^L	3,5(9)	0,57(3)	1,9(7)	0,68(4)
A_{op}^{LP}	4(1)	0,41(2)	0,5(4)	0,74(9)

TABLE 13 – Estimation des paramètres cinétiques pour le modèle de BELL-EVANS. La moyenne et son erreur standard (entre parenthèses, en unités du dernier chiffre significatif) sont calculées par bootstrap (ré-échantillonnage avec remise pour chaque vitesse de charge).

Toutefois, les ordres de grandeur pour les paramètres obtenus ne correspondent pas aux paramètres expérimentaux présentés dans la littérature. Par exemple, dans le modèle cinétique à deux états de THOMAS et al., (2006), les états basse affinité et haute affinité ont des k_{off}^0 respectifs de 6 s^{-1} et $7 \times 10^{-3} \text{ s}^{-1}$ et des Δx de 1,37 et 1,76 Å. Notons que les travaux de THOMAS et al. reposent sur un dispositif très différent (chambre à flux laminaire), mais se fondent également sur la validité du modèle de BELL-EVANS (BE). Les gammes de forces appliquées sur la liaison y sont tout autres (atteignant un maximum estimé

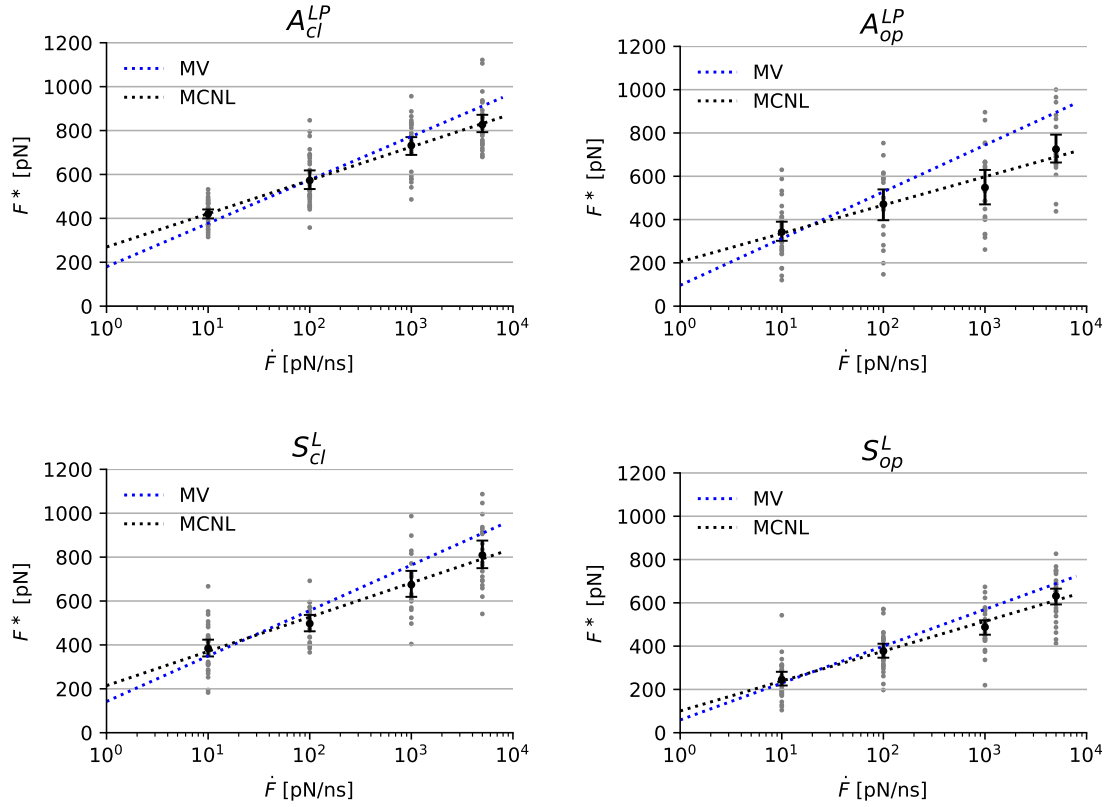


FIGURE 39 – Force de rupture moyenne en fonction de la vitesse de charge. Pour chaque système et vitesse, les données brutes (*points gris*) sont montrées ainsi que leur moyenne empirique (*points noirs*) avec intervalle de confiance à 95% estimé par bootstrap. (*en pointillés*) Moyennes prédites par le modèle de BELL-EVANS après ajustement par maximum de vraisemblance (MV) ou moindres carrés non-linéaires (MCNL).

à ~ 16 pN). Néanmoins, les valeurs de k_{off}^0 mesurées par résonance des plasmons de surface (SAUER et al., (2016) [15]) indépendamment du modèle BE sont cohérentes à un ordre de grandeur près (58 s^{-1} et $3,5 \times 10^{-4} \text{ s}^{-1}$). Ni les valeurs absolues, ni les valeurs relatives ne permettent donc de retrouver la division par 10^3 voire 10^5 de la vitesse de dissociation à force nulle par passage à la forme allostérique activée, telle que suggérée par les mesures expérimentales. Pour comprendre cette incohérence, il nous faut questionner le modèle utilisé.

L'ajustement du modèle de BELL-EVANS par moindres carrés non-linéaires MCNL (ligne pointillée noire FIGURE 39) semble montrer une bonne adéquation avec les moyennes empiriques observées et leur dépendance logarithmique avec la vitesse de charge. Toutefois, cette apparente adéquation avec le modèle est trompeuse car elle laisse de côté l'information pertinente contenue dans la variance de la force de rupture, visible dans la dispersion des points de données. L'ÉQUATION 68 prédit pour le modèle BE une variance indépendante de la vitesse de charge et simplement déterminée par le paramètre Δx . Une variance empirique montrant une dépendance avec la vitesse de charge \dot{F} , ou une valeur incohérente avec le Δx estimé sont des indices d'une mauvaise adéquation du modèle. Nous observons en effet que la valeur théorique déduite de l'approche MCNL conduit à une sous-estimation fréquente de la dispersion des forces de rupture, en particulier pour le système $A_{\text{op}}^{\text{LP}}$ (voir FIGURE B.8).

Au contraire, l'ajustement du modèle de BELL-EVANS par MV prend en compte l'information contenue dans la dispersion des données. Dans ce cas la force moyenne prédite par les paramètres ajustés (ligne pointillée bleue FIGURE 39) ne passe pas systématiquement par les moyennes empiriques. En particulier le modèle ajusté surestime les forces de ruptures aux vitesses de charge élevées. On peut donc comprendre les paramètres estimés dans cette approche comme le fruit d'un compromis entre les deux expressions de l'ÉQUATION 68. La pente $\partial \langle F \rangle / \partial \ln \dot{F}$, mais aussi la variance σ_F^2 sont proportionnelles à Δx^{-1} : la grande variance des observations a donc tendance à accroître la pente aux dépens de l'adéquation avec les moyennes empiriques.

En résumé, quel que soit l'estimateur choisi pour obtenir les paramètres cinétiques, le modèle de BELL-EVANS ne semble pas adapté au mécanisme de dissociation dans la gamme de vitesses de charge étudiée, et conduit à des prédictions en profond décalage avec les grandeurs obtenues dans diverses études expérimentales. Outre la question toujours présente de la validité du champ de force, nous pouvons dégager deux hypothèses pour expliquer cette divergence :

- L'hypothèse d'une réaction à une étape est correcte mais l'expression de BE pour la vitesse de dissociation n'est pas la bonne. DUDKO, HUMMER et SZABO [7] ont démontré une formule plus

générale pour $k_{\text{off}}(F)$ découlant directement de la théorie de KRAMERS, dont le modèle BE n'est qu'un cas particulier :

$$k_{\text{off}}(F) = k_{\text{off}}^0 \left(1 - \frac{\nu F \Delta x}{\Delta G^\ddagger}\right)^{1/\nu-1} \exp\left(\beta \Delta G^\ddagger \left[1 - \left(1 - \frac{\nu F \Delta x}{\Delta G^\ddagger}\right)^{1/\nu}\right]\right) \quad (71)$$

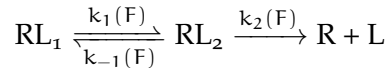
où ν est un paramètre relié à la forme de la surface d'énergie libre et ΔG^\ddagger la hauteur de la barrière énergétique. Contrairement au modèle empirique BE, correspondant à $\nu = 1$, le cas général permet donc d'obtenir des informations sur l'énergie d'activation de la réaction de dissociation. Cette approche a été utilisée avec succès par MILLES et al. [10] pour étudier la cinétique de dissociation d'un complexe impliquant autre adhésine bactérienne (SdrG). Toutefois, la présence de deux nouveaux paramètres libres rend la procédure d'ajustement plus délicate et des essais préliminaires ont montré des problèmes de convergence sur notre jeu de données limité.

- L'hypothèse d'une réaction à une étape est incorrecte et nous ne pouvons pas appliquer la théorie présentée ci-avant, et ce quel que soit le modèle choisi pour exprimer $k_{\text{off}}(F)$. L'observation d'évènements de rupture partielle avec des discontinuités dans l'augmentation linéaire de la force et le plateau de d_z (voir FIGURE 37) corrobore cette hypothèse et nous allons donc l'explorer par la suite.

6.3 MODÈLE À DEUX BARRIÈRES

La visualisation attentive des trajectoires de dissociation nous suggère un intermédiaire réactionnel où le ligand HM présente une pose semi-liée, caractérisée par la disruption de la liaison hydrogène entre l'ammonium N-terminal de la protéine et l'oxygène en position 2 du mannoside (O2) (FIGURE 40). Cet intermédiaire correspond aux discontinuités remarquées précédemment (FIGURE 37) et s'accompagne une différence de $\Delta d_z \approx 0,2 \text{ nm}$. Cette transition entraîne une chute de la force instantanée $\Delta F = \kappa \Delta d_z \approx 100 \text{ pN}$. Nous voyons que l'hypothèse d'une réaction à une seule étape et d'une rampe de force parfaitement linéaire sont de mauvaises approximations.

Nous envisageons donc le modèle cinétique suivant :



Nous faisons là encore l'hypothèse que les constantes de vitesses dépendant de la force $k_r(F)$ suivent un modèle de BELL-EVANS :

$$k_r(F) = k_r^0 \exp(\beta F \Delta x_r) \quad (72)$$

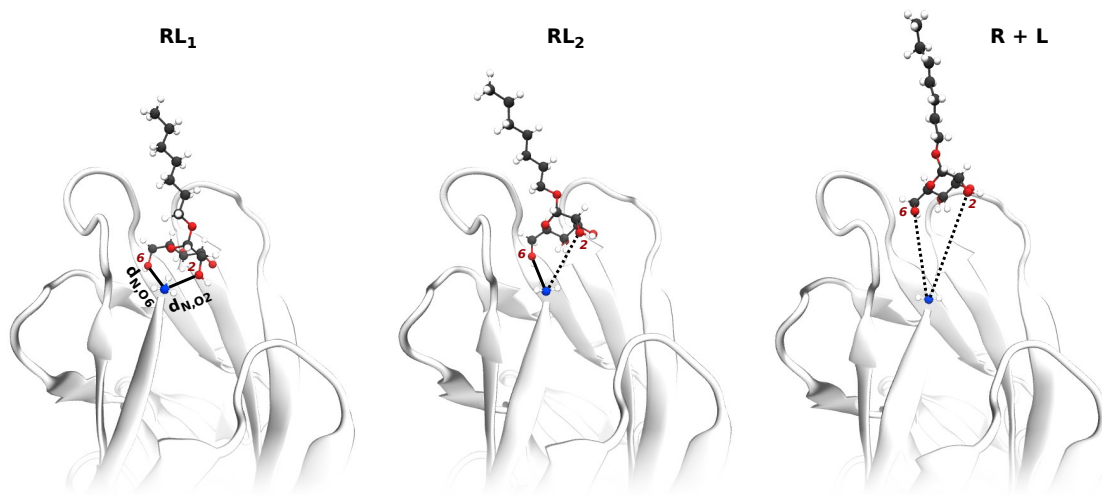


FIGURE 40 – Dissociation en deux étapes.

Le problème est alors d'évaluer les paramètres k_r^0 et Δx_r pour les trois étapes réactionnelles à partir des données collectées. Comme précédemment, nous allons nous appuyer sur l'hypothèse quasi-adiabatique : le temps d'équilibration à l'intérieur d'un des états est rapide par rapport aux transitions entre les états et la vitesse de variation de la force est faible en dehors des transitions elles-mêmes. Cela signifie que le système peut être considéré comme markovien, c'est à dire sans mémoire : une fois dans un état X_i , la dynamique ne dépend pas de l'histoire des autres états visités. Autrement dit, pour un pas de temps δt , la probabilité de trouver le système dans un état X_j à un instant $t + \delta t$ ne dépend que de l'état du système à t :

$$P[X(t + \delta t) | X(t), X(t - \delta t), \dots] = P[X(t + \delta t) | X(t)] \quad (73)$$

6.3.1 Méthodes

Modèle de Markov dépendant de la force

Nous nous appuyons sur la théorie des processus de MARKOV à temps continu pour l'adapter à notre système où la force est variable [16, 17]. Formellement, pour un ensemble fini d'états $\mathcal{S} = \{1, \dots, N\}$, le processus est déterminé par une matrice $N \times N$, matrice de taux de transition, notée \mathbf{K} :

$$\mathbf{K}(F) = \{k_{ij}(F)\} \text{ avec } k_{ii}(F) = - \sum_{j \neq i} k_{ij}(F) \quad (74)$$

où les $\{k_{ij}(F)\}_{i \neq j}$ correspondent aux constantes cinétiques habituelles (« macroscopiques ») des réactions de transition entre états. Dans notre exemple spécifique la matrice de taux de transition prend la forme

$$\mathbf{K}(F) = \begin{bmatrix} -k_1(F) & k_1(F) & 0 \\ k_{-1}(F) & -k_{-1}(F) - k_2(F) & k_2(F) \\ 0 & 0 & 0 \end{bmatrix}. \quad (75)$$

Les probabilités d'observer une transition $i \rightarrow j$ après une durée δt forment une matrice de transition $\mathbf{T}(F)$

$$\begin{aligned} \mathbf{T}(F, \delta t) &= \{P[X(t + \delta t) = j \mid X(t) = i, F(t) = F]\} \\ &= \text{Exp}[\delta t \mathbf{K}(F)] \end{aligned} \quad (76)$$

où Exp désigne l'exponentiation matricielle. Cette relation n'est valable qu'à la condition de considérer que la force est quasi-constante pendant la durée δt .

Dans le modèle BE chaque $k_{ij}(F)$ est paramétré par deux constantes, k_{ij}^0 et Δx_{ij} . Un modèle à N états aura donc $2(N^2 - N)$ paramètres cinétiques (soit deux par transition $i \rightarrow j$ avec $i \neq j$). Notons ces paramètres θ . Ces paramètres peuvent être estimés par maximum de vraisemblance (MV) à partir d'un ensemble de trajectoires. Prenons l'exemple d'une trajectoire particulière suivant le processus, de durée $M\delta t$, discrétisée tous les δt , notée $\mathbf{x} = \{x_0, x_1, \dots, x_M\}$. La probabilité d'observer cette trajectoire sachant l'état initial x_0 et la trajectoire de forces $\mathbf{F} = \{F_0, F_1, \dots, F_M\}$ est simplement le produit des probabilités de transition

$$P_\theta(\mathbf{x} \mid x_0, \delta t, \mathbf{F}) = \prod_{m=0}^{M-1} [T_\theta(F_m, \delta t)]_{x_m, x_{m+1}}. \quad (77)$$

La fonction de vraisemblance pour un ensemble de L trajectoires indépendantes se généralise par le produit

$$\mathcal{L}(\theta) = \prod_{l=1}^L P_\theta(\mathbf{x}^l \mid x_0^l, \delta t^l, \mathbf{F}^l) \quad (78)$$

où l'indice l caractérise chaque trajectoire. On note que l'intervalle de temps δt peut différer entre plusieurs trajectoires pourvu que le critère de lente variation de la force soit respecté.

L'estimateur du MV pour les paramètres est alors comme précédemment,

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta). \quad (79)$$

En pratique les paramètres sont estimés numériquement grâce à un algorithme d'optimisation implémenté en Python. La log-vraisemblance $\ln \mathcal{L}(\theta)$ est maximisée grâce à la fonction `minimize` de la bibliothèque SciPy utilisant l'algorithme L-BFGS-B. Les paramètres k_{ij}^0 sont bornés dans l'intervalle $[0; +\infty[$. La méthode L-BFGS-B est un algorithme itératif de descente de gradient et nécessite l'évaluation répétée du gradient $\nabla_{\theta} \ln \mathcal{L}(\theta)$. Pour des raisons d'efficacité nous implémentons en Cython un calcul approché de la log-vraisemblance et du gradient. Les détails sont explicités en annexe A.2.

Simulations à force constante

Afin d'enrichir le jeu de données et en particulier d'échantillonner mieux la transition aller-retour $LR_1 \rightleftharpoons LR_2$ nous réalisons également des simulations sous force constante. Le protocole est identique (voir 6.2.1) mais un biais linéaire constant est utilisé au lieu d'un biais harmonique mobile. Les forces suivantes ont été utilisées :

- A_{cl}^{LP} : 280, 320, 350, et 380 pN.
- S_{cl}^L : 280, 300, 320, 350, et 380 pN.
- A_{op}^{LP} : 180, 230, 280, et 300 pN.
- S_{op}^L : 180, 230, 280, et 300 pN.

Pour chaque force, 20 trajectoires sont simulées pendant 10–30 ns. Notons que les trajectoires ne s'achèvent pas nécessairement par une dissociation complète. L'analyse par MV bénéficie tout de même de l'information présente dans ces trajectoires.

Analyse des trajectoires

Les distances $d_{N,O2}$ et $d_{N,O6}$ et la force instantanée sont enregistrées avec un intervalle δt . Un filtre de réduction du bruit préservant des événements de rupture nets est appliqué aux données (algorithme de débruitage par réduction de la variation totale [18, 19] implémenté dans la bibliothèque `scikit-image` [20], paramètre de poids = 10). Les états sont discrétisés selon les critères suivants

$$\text{État} \equiv \begin{cases} RL_1 & \text{si } d_{N,O2} < 4 \text{ \AA} \text{ et } d_{N,O6} < 6 \text{ \AA} \\ RL_2 & \text{si } d_{N,O2} > 4 \text{ \AA} \text{ et } d_{N,O6} < 6 \text{ \AA} \\ R + L & \text{sinon.} \end{cases} \quad (80)$$

Une transition n'est prise en compte que si le système demeure au minimum 10 ps dans le nouvel état. La figure 41 présente des trajectoires typiques à vitesse de charge et à force constantes. On peut y apprécier les échanges entre RL_1 et RL_2 avant la rupture irréversible vers $R + L$.

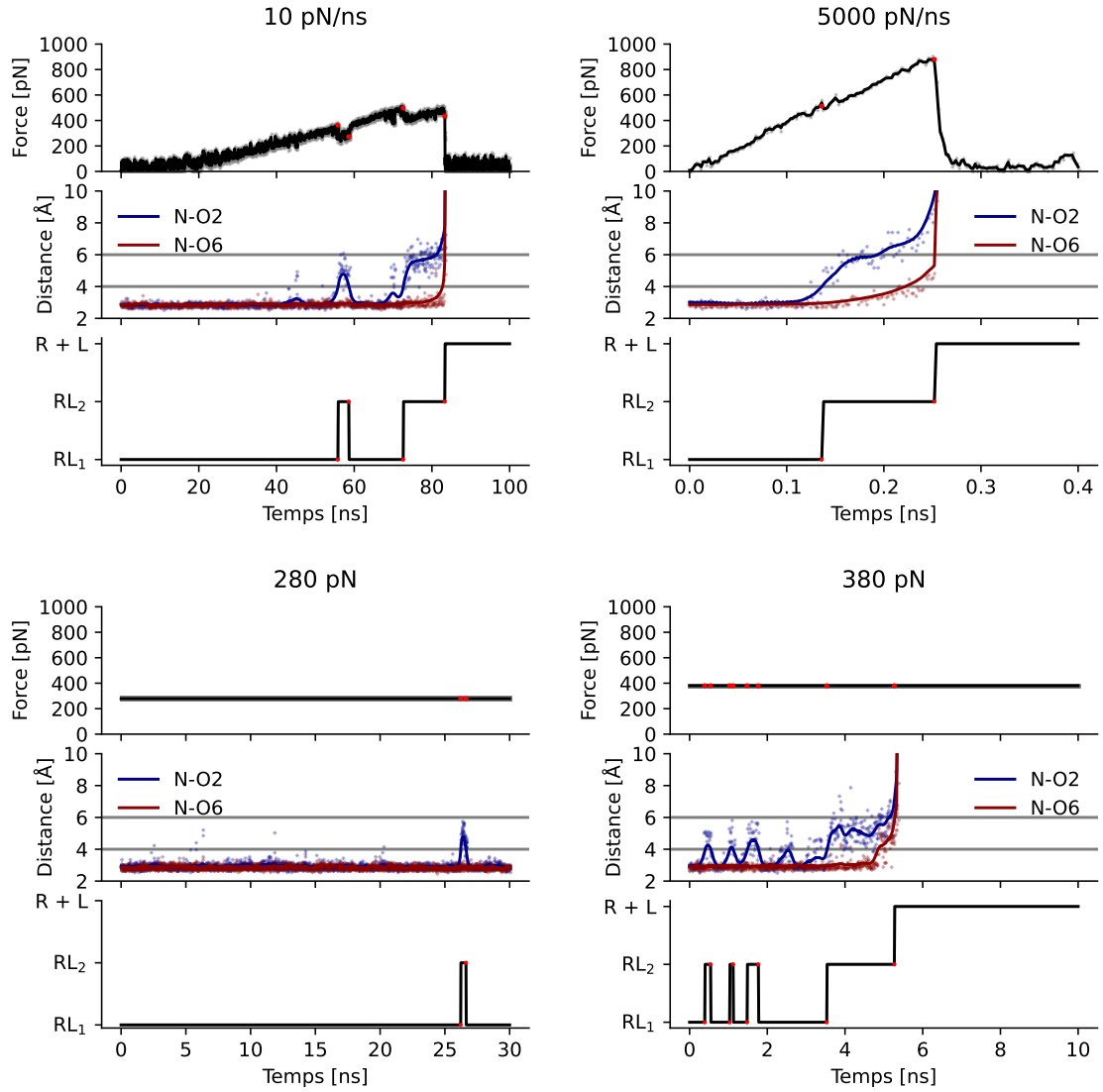


FIGURE 41 – Analyse des trajectoires à force ou vitesse de charge constantes pour le modèle à deux états. Des exemples typiques sont donnés pour le système A_{cl}^{LP} pour des simulations à vitesse de charge constante (*en haut*) et à force constante (*en bas*). Les distances $d_{N,O2}$ et $d_{N,O6}$ sont présentées brutes (*points*) et après lissage (*courbe continue*). L'état RL_1 , RL_2 ou $R + L$ est déduit d'après les seuils (*lignes grises*) de l'ÉQUATION 80. Les forces instantanées précédant chaque transition sont indiquées (*points rouge vif*).

Afin d'évaluer la précision statistique de l'estimation MV des paramètres nous procédons par bootstrap sur les trajectoires indépendantes : pour chaque condition (vitesse de charge ou force) la collection de trajectoires correspondantes est ré-échantillonnée avec remise. La procédure d'ajustement est effectuée sur 10 jeux de données ainsi générés et les paramètres moyens ainsi que leurs écarts types sont calculés.

6.3.2 Résultats

Les constantes cinétiques k_1^0 , k_{-1}^0 et k_2^0 et leurs Δx correspondants sont résumées dans la TABLE 14.

État	k_1^0 [10^6 s^{-1}]	Δx_1 [Å]	k_{-1}^0 [10^6 s^{-1}]	Δx_{-1} [Å]	k_2^0 [10^6 s^{-1}]	Δx_2 [Å]
A_{cl}^{LP}	4(6)	0,50(7)	5387(4032)	-0,32(19)	5(3)	0,38(5)
S_{cl}^L	11(9)	0,42(8)	3411(1952)	-0,21(14)	10(4)	0,34(4)
A_{op}^{LP}	48(8)	0,50(3)	1006(360)	-0,22(4)	12(3)	0,32(3)
S_{op}^L	10(5)	0,60(7)	265(74)	-0,05(8)	47(18)	0,32(7)

TABLE 14 – Estimation des paramètres cinétiques pour le modèle de MARKOV sous force. La moyenne et son erreur standard (entre parenthèses, en unités du dernier chiffre significatif) sont calculées par bootstrap sur les trajectoires (ré-échantillonnage avec remise pour chaque vitesse de charge ou force constante).

Les paramètres Δx sont de 0,3–0,6 Å pour les étapes de rupture de liaison hydrogène (1 et 2). Ces valeurs sont positives, ce qui signifie bien que l'augmentation de la force accélère la rupture. Leur ordre de grandeur est cohérent avec la distance à laquelle on attendrait la barrière énergétique le long de la coordonnée de réaction et qui correspond, à titre d'exemple, à la distance entre le premier pic et le premier minimum de la fonction de distribution radiale de l'eau pure [21]. La précision de l'estimation ne permet pas d'établir de différence significative entre les états.

Comme attendu la réaction retour (-1) possède un Δx généralement négatif, ce qui signifie qu'elle est inhibée par l'augmentation de la force. Leurs valeurs absolues sont légèrement plus faibles que pour les réactions aller mais là encore l'erreur standard nous incite à les comparer avec prudence.

De même, malgré des marges d'incertitude importantes nous voyons que les états fermés et ouverts se distinguent par leur ratio k_1^0/k_{-1}^0 , plus favorable à l'état semi-dissocié lorsque la boucle est ouverte. Il est de $0,7 \times 10^{-3}$ et 3×10^{-3} pour A_{cl}^{LP} et S_{cl}^L respectivement, contre 5×10^{-2} et 4×10^{-2} pour A_{op}^{LP} et S_{op}^L . Nous pouvons, dans le cadre

de l'approximation d'un l'équilibre rapide ($k_1 + k_{-1} \gg k_2$) estimer le k_{off} effectif en l'absence de force

$$k_{\text{off}}^0 \stackrel{\text{AER}}{\approx} \frac{k_1^0}{k_{-1}^0} k_2^0 \quad (81)$$

État	$k_{\text{off}}^0 [10^6 \text{ s}^{-1}]$
$A_{\text{cl}}^{\text{LP}}$	4×10^{-3}
S_{cl}^{L}	3×10^{-2}
$A_{\text{op}}^{\text{LP}}$	6×10^{-1}
S_{op}^{L}	1,8

TABLE 15 – Estimation des vitesses de dissociation en l'absence de force et dans l'approximation d'un pré-équilibre rapide.

Nous constatons que les valeurs de k_{off}^0 pour les formes fermées sont plus faibles de deux ordres de grandeur à celles trouvées par l'approche précédente, en TABLE 13. Il s'agit d'une amélioration en ce qui concerne les valeurs absolues estimées expérimentalement, mais avec une différence toujours impressionnante de 6 à 7 ordres de grandeur pour la forme haute affinité (rappelons que la vitesse de dissociation la plus lente issue des modèles expérimentaux avoisine 10^{-3} – 10^{-4} s^{-1}). La différence entre les formes ouvertes et fermées est également plus marquée avec une multiplication par 150 et 60 de la vitesse pour A^{LP} et S^{L} respectivement.

6.4 CONCLUSION

Dans cette section nous avons établi que le chemin réactionnel de dissociation du ligand comporte un état métastable partiellement dissocié correspondant à la rupture d'une liaison hydrogène entre le site et une fonction hydroxyle du cycle mannosidique. Cette liaison correspond à la partie du réseau d'interactions spécifiques site-ligand juxtaposée à la boucle de fermeture mobile. La boucle en position close (cl) protège la liaison N–O2 du solvant, notamment grâce au résidu hydrophobe Ile13 (FIGURE 42). Au contraire celle-ci est nettement plus accessible dans la forme ouverte (op). Nous pouvons donc supposer que la compétition avec le solvant polaire explique en partie les différences cinétiques observées en TABLE 14.

Grâce à l'étape retour (-1), favorisée pour les états fermés, la durée de vie de la liaison est significativement prolongée par rapport à l'enchaînement des deux seules étapes de dissociation (1) et (2). Du fait de l'action individuelle de la force sur les trois étapes, le modèle prédit un écart à l'augmentation linéaire de la force de rupture moyenne

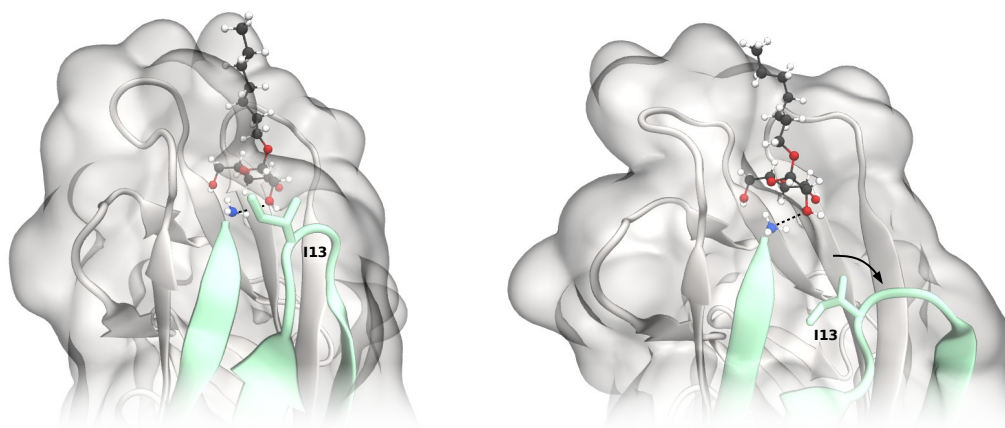


FIGURE 42 – Exposition de la liaison hydrogène N-O₂ lors de l'ouverture du site. Dans l'état clos (*gauche*) la liaison est protégée par un environnement hydrophobe maintenu entre autres par l'isoleucine 13. Le passage à l'état ouvert (*droite*) expose la liaison au solvant.

avec le logarithme de la vitesse de charge. Cette non-linéarité est peu perceptible aux fortes vitesses de charge simulées, où la réaction retour est peu fréquente car rapidement inhibée par l'augmentation de la force.

Ces données suggèrent que la prise en compte d'un mécanisme en deux étapes est pertinente dans les gammes de forces et de vitesses de charge étudiées, mais que des efforts sont encore nécessaires afin de réconcilier les résultats de simulations avec les expériences. L'aspect complexe de cette tâche provient des échelles de temps accessibles aux simulations atomistiques qui contraignent à l'utilisation de forces très grandes et des vitesses de charge de plusieurs ordres de grandeur plus élevées que celles employées *in vitro* dans les expériences de spectroscopie de force. L'extrapolation grâce à un modèle théorique commun des paramètres cinétiques d'un régime à l'autre pose alors plusieurs problèmes :

- le mécanisme (chemin réactionnel) sélectionné par les conditions extérieures peut s'avérer différent et ne pas permettre d'utiliser un modèle commun,
- les hypothèses et approximations (adiabaticité, hauteur et forme de la barrière) justifiant le modèle théorique peuvent ne plus être vérifiées dans certaines conditions extrêmes,
- la dépendance très sensible (souvent exponentielle) des vitesses réactionnelles vis-à-vis de la force entraîne des incertitudes fortes sur les paramètres et rend l'analyse particulièrement sensible à divers biais (fidélité du champ de force, traitement du signal, choix des critères de définition des états...).

Qualitativement, notre étude corrobore l'hypothèse que l'état d'ouverture du site de liaison est un facteur prépondérant dans la ciné-

tique de dissociation du complexe FimH-mannoside. La transition allostérique $A \leftrightarrow S$ module indirectement la vitesse de dissociation macroscopique en changeant les proportions d'états ouverts, semi-ouverts et fermés. Ce modèle rejoint la notion de couplage allostérique faible (*weak coupling*) proposée par certains auteurs [22, 23] suggérant une certaine flexibilité dans la transmission du signal allostérique entre la région inter-domaines et le site de liaison. Il explique pourquoi, en dépit de la forme identiquement close du site de liaison dans les co-cristaux avec HM du domaine lectine seul ($\bullet S_{cl}^L$) et des deux domaines ($\bullet A_{cl}^{LP}$), leurs affinités et k_{off} expérimentales divergent fortement : la dynamique d'ouverture favorisée dans l'état A permet une dissociation beaucoup plus rapide. En ce sens le concept d'allostérie dynamique [15, 23] peut également s'appliquer, bien que la notion historique fasse plutôt référence à des fluctuations thermiques de modes vibrationnels « mous » et « globaux » et moins à la régulation de transitions locales entre états bien distincts [24, 25].

RÉFÉRENCES

- [1] B. ISRALEWITZ, M. GAO et K. SCHULTEN. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, **11**, 224-230 (2001).
DOI : [10.1016/S0959-440X\(00\)00194-9](https://doi.org/10.1016/S0959-440X(00)00194-9)
- [2] E. FLORIN, V. MOY et H. GAUB. Adhesion forces between individual ligand-receptor pairs. *Science*, **264**, 415-417 (1994).
DOI : [10.1126/science.8153628](https://doi.org/10.1126/science.8153628)
- [3] C. JARZYNSKI. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, **78**, 2690-2693 (1997).
DOI : [10.1103/PhysRevLett.78.2690](https://doi.org/10.1103/PhysRevLett.78.2690)
- [4] S. PARK, F. KHALILI-ARAGHI, E. TAJKHORSHID et K. SCHULTEN. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics*, **119**, 3559-3566 (2003).
DOI : [10.1063/1.1590311](https://doi.org/10.1063/1.1590311)
- [5] C. DELLAGO et G. HUMMER. Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy*, **16**, 41-61 (2014).
DOI : [10.3390/e16010041](https://doi.org/10.3390/e16010041)
- [6] H. M. BAUMANN, V. GAPSYS, B. L. de GROOT et D. L. MOBLEY. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. *The Journal of Physical Chemistry B*, **125**, 4241-4261 (2021).
DOI : [10.1021/acs.jpcb.0c10263](https://doi.org/10.1021/acs.jpcb.0c10263)
- [7] O. DUDKO, G. HUMMER et A. SZABO. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Physical Review Letters*, **96**, 108101 (2006).
DOI : [10.1103/PhysRevLett.96.108101](https://doi.org/10.1103/PhysRevLett.96.108101)
- [8] O. K. DUDKO, J. MATHÉ, A. SZABO, A. MELLER et G. HUMMER. Extracting Kinetics from Single-Molecule Force Spectroscopy : Nanopore Unzipping of DNA Hairpins. *Biophysical Journal*, **92**, 4188-4195 (2007).
DOI : [10.1529/biophysj.106.102855](https://doi.org/10.1529/biophysj.106.102855)
- [9] O. K. DUDKO, G. HUMMER et A. SZABO. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences*, **105**, 15755-15760 (2008).
DOI : [10.1073/pnas.0806085105](https://doi.org/10.1073/pnas.0806085105)
- [10] L. F. MILLES, K. SCHULTEN, H. E. GAUB et R. C. BERNARDI. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science*, **359**, 1527-1533 (2018).
DOI : [10.1126/science.aar2094](https://doi.org/10.1126/science.aar2094)
- [11] A. GARG. Escape-field distribution for escape from a metastable potential well subject to a steadily increasing bias field. *Physical Review B*, **51**, 15592-15595 (1995).
DOI : [10.1103/PhysRevB.51.15592](https://doi.org/10.1103/PhysRevB.51.15592)

- [12] W. THOMAS, M. FORERO, O. YAKOVENKO, L. NILSSON, P. VICINI, E. SOKURENKO et V. VOGEL. Catch-Bond Model Derived from Allostery Explains Force-Activated Bacterial Adhesion. *Biophysical Journal*, **90**, 753-764 (2006).
DOI : [10.1529/biophysj.105.066548](https://doi.org/10.1529/biophysj.105.066548)
- [13] S. SEABOLD et J. PERKTOLD. Statsmodels : Econometric and Statistical Modeling with Python. Python in Science Conference. Austin, Texas, 2010. 92-96. DOI : [10.25080/Majora-92bf1922-011](https://doi.org/10.25080/Majora-92bf1922-011)
- [14] P. VIRTANEN et al. SciPy 1.0 : fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**, 261-272 (2020).
DOI : [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- [15] M. M. SAUER, R. P. JAKOB, J. ERAS, S. BADAY, D. ERIŞ, G. NAVARRA, S. BERNÈCHE, B. ERNST, T. MAIER et R. GLOCKSHUBER. Catch-bond mechanism of the bacterial adhesin FimH. *Nature Communications*, **7**, (2016).
DOI : [10.1038/ncomms10738](https://doi.org/10.1038/ncomms10738)
- [16] V. S. PANDE, K. BEAUCHAMP et G. R. BOWMAN. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods*, **52**, 99-105 (2010).
DOI : [10.1016/j.ymeth.2010.06.002](https://doi.org/10.1016/j.ymeth.2010.06.002)
- [17] R. T. MCGIBBON et V. S. PANDE. Efficient maximum likelihood parameterization of continuous-time Markov processes. *The Journal of Chemical Physics*, **143**, 034109 (2015).
DOI : [10.1063/1.4926516](https://doi.org/10.1063/1.4926516)
- [18] L. I. RUDIN, S. OSHER et E. FATEMI. Nonlinear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, **60**, 259-268 (1992).
DOI : [10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F)
- [19] A. CHAMBOLLE. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, **20**, 89-97 (2004).
DOI : [10.1023/B:JMIV.0000011325.36760.1e](https://doi.org/10.1023/B:JMIV.0000011325.36760.1e)
- [20] S. v. d. WALT, J. L. SCHÖNBERGER, J. NUNEZ-IGLESIAS, F. BOULOGNE, J. D. WARNER, N. YAGER, E. GOUILLART et T. YU. scikit-image : image processing in Python. *PeerJ*, **2**, e453 (2014).
DOI : [10.7717/peerj.453](https://doi.org/10.7717/peerj.453)
- [21] A. SOPER. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics*, **258**, 121-137 (2000).
DOI : [10.1016/S0301-0104\(00\)00179-8](https://doi.org/10.1016/S0301-0104(00)00179-8)
- [22] V. B. RODRIGUEZ, B. A. KIDD, G. INTERLANDI, V. TCHESNOKOVA, E. V. SOKURENKO et W. E. THOMAS. Allosteric Coupling in the Bacterial Adhesive Protein FimH. *Journal of Biological Chemistry*, **288**, 24128-24139 (2013).
DOI : [10.1074/jbc.M113.461376](https://doi.org/10.1074/jbc.M113.461376)

- [23] S. RABBANI, B. FIEGE, D. ERIS, M. SILBERMANN, R. P. JAKOB, G. NAVARRA, T. MAIER et B. ERNST. Conformational switch of the bacterial adhesin FimH in the absence of the regulatory domain : Engineering a minimalist allosteric system. *Journal of Biological Chemistry*, **293**, 1835-1849 (2018).
DOI : [10.1074/jbc.M117.802942](https://doi.org/10.1074/jbc.M117.802942)
- [24] A. COOPER et D. T. DRYDEN. Allostery without conformational change. A plausible model. *European biophysics journal : EBJ*, **11**, 103-109 (1984).
DOI : [10.1007/BF00276625](https://doi.org/10.1007/BF00276625)
- [25] S. J. WODAK et al. Allostery in Its Many Disguises : From Theory to Applications. *Structure*, **27**, 566-578 (2019).
DOI : [10.1016/j.str.2019.01.003](https://doi.org/10.1016/j.str.2019.01.003)

Quatrième partie

CONCLUSION

CONCLUSION

RÉCAPITULATIF

Dans cette thèse nous avons tenté de comprendre les détails moléculaires à l'origine du mécanisme grappin de l'adhésine bactérienne FimH par une approche computationnelle fondée sur la dynamique moléculaire. Notre approche peut se résumer en trois temps et deux directions, qui se reflètent dans l'organisation des chapitres. Les trois temps correspondent à une investigation par échelles successives, du général vers le particulier : exploration générale du paysage conformationnel, quantification du profil énergétique selon une coordonnée spécifique régulant l'affinité avec le ligand, évaluation des propriétés de liaison dans chaque état identifié. Les deux axes sont ceux des propriétés d'équilibre thermodynamique et de celles hors équilibre, notamment lors de l'application d'une force extérieure au système.

Dans un premier temps, nous avons exploré l'espace des configurations accessible au domaine effecteur, le domaine lectine, grâce à une méthode d'échantillonnage renforcé, la dynamique moléculaire à échange de répliques hamiltonien. Cette méthode permet d'accélérer l'échantillonnage de l'ensemble conformationnel d'équilibre avec un minimum d'a priori sur les degrés de liberté lents du système. À partir de nos observations et des structures cristallographiques connues, nous caractérisons deux coordonnées nous paraissant pertinentes pour décrire la dynamique du domaine effecteur : une coordonnée « allostérique » et une coordonnée « d'ouverture » du site de liaison. La première décrit le changement conformationnel favorisé par l'association et la dissociation du domaine régulateur, le domaine piline, et se traduit par une réorganisation complexe des boucles à l'interface des deux domaines et d'un motif structural particulier nommé « renflement bêta/switch alpha ». La seconde décrit l'évasement local du site de liaison, l'ouverture d'une boucle bordant son pourtour, et le mouvement plus global de torsion du sandwich bêta constituant le domaine et assurant le couplage allostérique. À partir de deux conformations allostériques extrêmes nommées S (haute affinité) et A (basse affinité), nous constatons que, malgré une flexibilité des boucles inter-domaines, la transition complète est trop lente pour être observée dans nos simulations en dépit de l'échantillonnage accéléré. En revanche, chaque état explore différents sous-états

d'ouverture dont certains n'ont, à notre connaissance, pas encore été identifiés expérimentalement.

Afin de préciser plus quantitativement la corrélation entre état allostérique, présence du ligand et conformation du site, nous avons employé dans un second temps une méthode d'échantillonnage préférentiel le long de la coordonnée d'ouverture précédemment définie, l'échantillonnage parapluie à échange de biais. Contrairement à la phase d'exploration, l'objectif était ici d'accélérer l'échantillonnage dans les régions difficilement accessibles d'une variable collective d'intérêt. Les profils d'énergie libre obtenus permettent de distinguer trois puits correspondant à trois états d'ouverture que nous qualifions d'ouvert, semi-ouvert et fermé, et de calculer leurs populations respectives en fonction de l'état allostérique et de la présence ou non du domaine régulateur. Ces données sont cohérentes avec un modèle par couplage faible entre le site de liaison et l'interface régulatrice, avec une coexistence d'états d'ouverture dont les proportions sont modulées par la transition allostérique. De façon surprenante, nos calculs prédisent que la forme ouverte de la protéine est toujours majoritaire dans l'état A, et majoritaire dans l'état S non-lié, en décalage avec les formes isolées dans les données cristallographiques et suggérées par RMN. Nous notons toutefois que la présence du ligand stabilise significativement les conformations fermées, ce qui est cohérent avec le modèle d'ajustement induit proposé dans la littérature. Dans l'optique d'étudier le fonctionnement de la protéine en conditions de stress mécanique, nous avons également calculé les profils d'énergie libre en présence d'une force constante appliquée au complexe et constatons une stabilisation de l'état fermé. Cette observation est intéressante car elle rend plausible un mécanisme de liaison accrocheuse secondaire, local au site de liaison, et redondant avec celui reposant sur une allostérie inter-domaines. Nous faisons l'hypothèse que la stabilisation sous force de la conformation fermée retarde la dissociation du complexe à moyen terme et donne le temps au système de transitionner vers l'état allostérique activé. Il reste à déterminer en pratique si ce mécanisme a des effets mesurables dans les conditions physiologiques.

Dans un troisième et dernier temps, nous avons cherché à mettre en relation la plasticité du site de liaison avec la stabilité du complexe. Capitalisant sur les états caractérisés précédemment, nous avons réalisé des calculs d'énergie libre par une méthode de découplage alchimique afin d'estimer leurs affinités relatives. Bien que préliminaires, nos résultats confirment le rôle crucial de l'état d'ouverture sur la stabilité du complexe. Ils mettent également en évidence des degrés de liberté lents au niveau des boucles L2 et L3 et du portail à tyrosines entourant le site de liaison, dont la prise en compte pourrait être nécessaire à l'interprétation des valeurs calculées et leur mise en relation

avec les mesure expérimentales. Dans une perspective hors d'équilibre, nous avons également effectué des simulations de dynamique moléculaire dirigée (SMD) pour caractériser la cinétique de dissociation du complexe soumis à une force de traction. Nous analysons les données dans le cadre du modèle phénoménologique de BELL-EVANS afin de retrouver les constantes de vitesse en l'absence de force. Conformément aux attentes, les résultats montrent une cinétique plus lente dans les états fermés par rapport aux états ouverts. Les k_{off} calculés sont cependant surestimés de plusieurs ordres de grandeur par rapport aux données expérimentales, et leur différences relatives sont trop faible. Constatant que le chemin réactionnel est mieux décrit par un modèle cinétique à deux états liés, nous avons conçu un modèle de MARKOV dépendant de la force et ajusté ses paramètres sur nos données. Cette approche nous permet d'améliorer la concordance avec les attentes expérimentales mais conserve un décalage de plusieurs ordres de grandeur, témoignant de la difficulté à concilier les échelles de la dynamique moléculaire atomistique avec celles du laboratoire.

COMPARAISON AUX TRAVAUX THÉORIQUES PRÉCÉDENTS

La dynamique moléculaire a fait les preuves à maintes reprises de sa capacité à rationaliser et même prédire des propriétés et mécanismes structuraux des systèmes moléculaires, en particulier dans le domaine de la mécanique des protéines et l'interprétation des expériences de spectroscopie de force sur molécule unique [1, 2]. La protéine-grappin FimH ne fait pas exception et plusieurs études théoriques préalables à la nôtre ont été menées, en parallèle et en accompagnement des résultats expérimentaux. En raison des limitations inhérentes à la MD et de l'évolution relativement récente des capacités de calcul – notamment avec l'avènement des supercalculateurs pétaflopiques et la démocratisation de l'accélération GPU durant la dernière décennie – ces approches sont longtemps restées qualitatives, avec des simulations dépassant rarement la centaine de nanosecondes, n'utilisant qu'un petit nombre de répliques et ne permettant pas de réelle quantification statistique. Elles ont cependant contribué à étendre la compréhension structurale fondée sur les structures cristallines en y introduisant de la dynamique. Nous pouvons citer des simulations de SMD ayant permis de mettre en évidence l'extension du *linker* et la déformation des boucles inter-domaines sous force [3], d'estimer la force critique à partir de laquelle la séparation des domaines domine par rapport à la dissociation du ligand [4] ou encore de caractériser le mécanisme de rupture de l'assemblage supramoléculaire de l'extrémité piliale [5]. D'autres simulations de MD traditionnelle ont mis en évidence les différences de plasticité du site de liaison et de sa boucle de fermeture, en soulignant l'importance du

pont disulfure 24–65 pour la stabilité sous force [6], ou en constatant la différence d’amplitude des fluctuations entre les états S et A [7–9].

Parmi les études computationnelles plus conséquentes nous pouvons citer les travaux de INTERLANDI et THOMAS, (2016) [10] qui constatent une divergence entre la flexibilité de la boucle de fermeture dans l’état A et celle suggérée par les facteurs B cristallographiques, et l’associent à de probables contacts cristallins. Grâce à des simulations atteignant plusieurs microsecondes exécutées sur le supercalculateur spécialisé Anton [11], les auteur·e·s détaillent les interactions protéine-ligand et attribuent la différence d’affinité des états A_{op} et S_{cl} à une liaison hydrogène supplémentaire dans ce dernier et une surface accessible au solvant moins grande. Des indices de transmission allostérique sont analysés par coefficients de corrélation linéaire entre diverses variables de la région inter-domaines, du feuillet bêta central et du site de liaison. Une approche similaire est reprise par LIU, AMARAL et KETEN, (2021) [12] qui identifient des différences dynamiques entre les deux états allostériques, au niveau des angles dièdres de la chaîne principale localisés dans les régions pré-citées. Ces travaux permettent de mettre en exergue des corrélations dynamiques le long du « chemin allostérique » supposé, sur la base de fluctuations thermiques autour d’un bassin conformationnel métastable. Toutefois, même pour des simulations de plusieurs micro-secondes, ils ne peuvent donner de vision globale et quantitative de l’influence de l’état allostérique sur le paysage énergétique du domaine effecteur.

Notre étude est la première en date à utiliser de véritables techniques d’échantillonnage renforcé nous permettant d’identifier et de quantifier les principaux bassins conformationnels du domaine lectine dans ses deux états allostériques. Nous montrons l’existence d’au moins trois sous-états distincts dont la stabilité est modulée par la transformation du motif clé renflement bêta/switch alpha. Nous pouvons confirmer que ces états de fermeture du site corrélaient avec la stabilité du complexe, que ce soit en termes d’affinité (stabilité thermodynamique) ou de vitesse de dissociation (stabilité cinétique). Enfin, nous appliquons pour la première fois une force *in silico* sur le ligand, ce qui nous permet d’identifier un mécanisme de dissociation en deux étapes, mais également de faire l’hypothèse d’un mécanisme de liaison accrocheuse à deux niveaux.

LIMITATIONS

Modèle

L’une des sources de divergence entre les prédictions de la dynamique moléculaire classique et l’expérience réside dans les approximations et erreurs du champ de force utilisé. On peut distinguer deux

types d'erreur, légèrement différentes du point de vue épistémologique :

- Les erreurs de paramétrage supposent qu'il existe un jeu de paramètres optimal pour une forme fonctionnelle donnée du champ de force. Les erreurs proviennent alors d'un paramétrage sous-optimal, par exemple en raison d'un manque de données d'apprentissage.
- Les erreurs de modélisation, qui caractérisent la difficulté pour le modèle donné de décrire avec exactitude les observations en raison d'hypothèses trop simplificatrices ou d'une forme fonctionnelle inadaptée.

Les efforts en cours dans le domaine de la chimie et de la biochimie théorique visent à améliorer ces deux aspects. Les champs de force traditionnels montrent encore une marge de progression et sont étendus à des territoires encore sous-explorés comme les protéines intrinsèquement désordonnées, les acides aminés non-standard, les modifications post-traductionnelles, les acides nucléiques modifiés, etc. [13, 14]. Malgré le développement de méthodes d'optimisation *ab initio*, l'acquisition et la standardisation de mégadonnées bio-physico-chimiques de référence est nécessaire à leur validation. En outre, la relative simplicité des modèles est un frein à l'optimisation d'un champ de force unique couvrant la variété de situations et d'environnements physico-chimiques en biologie. Le développement de modèles plus complexes, comme les champs de force incorporant une polarisation explicite, pourrait aider à les unifier dans un même cadre. Une autre piste de recherche est l'utilisation de réseaux d'apprentissage profond faisant l'économie d'a priori sur une forme fonctionnelle et pouvant modéliser des champs de force réactifs à des coûts bien moindres que ceux de la dynamique moléculaire *ab initio* [15].

Notre expérience montre que malgré leur vaste utilisation, des différences marquantes persistent d'un champ de force à l'autre. La réplique de nos expériences avec un autre modèle (tel que CHARMM-36m, que nous avons décidé de laisser de côté en début de projet) pourrait être riche d'enseignements mais serait coûteuse en ressources. Les sources d'erreurs méthodologiques et statistiques étant nombreuses, il n'est pas garanti qu'un autre champ de force améliore significativement nos prédictions théoriques.

Définition des variables collectives

L'analyse et l'interprétation des trajectoires hautement dimensionnelles générées par MD est une tâche complexe et sujette aux tâtonnements. L'un des défis est la recherche de descripteurs structuraux pertinents, permettant de décrire de façon simplifiée les processus moléculaires étudiés. Ces variables collectives (CV), par leur distribution statistique et leur dynamique, peuvent être reliées à des mesures

macroscopiques (données spectroscopiques, temps de relaxation RMN, distance entre molécules fluorescentes dans des expériences type FRET, rayon hydrodynamique, température de fusion, ...). Elles sont également indispensables à la mise en place de méthodes d'échantillonnage préférentiel pour faire face aux degrés de liberté lents du système [16].

La recherche de CV est généralement guidée soit par l'intuition et les connaissances des chercheur-e-s, soit par une méthode automatisée de réduction de dimensionnalité. Dans les travaux présentés ici, nous avons majoritairement privilégié la première approche, en raison de l'abondance de données structurales nous permettant de faire des hypothèses raisonnables sur les transformations géométriques importantes du système. Les CV ainsi définies offrent l'avantage d'être intuitives et faciles à interpréter. Nous voyons cependant les limitations de cette démarche lorsqu'elle risque de négliger des degrés de liberté lents venant influencer le processus d'intérêt, comme nous l'avons constaté avec la conformation des boucles L2-L3 et le portail à tyrosines. Notre intuition initiale était que ces régions sans structure secondaire, exposées au solvant et en contact avec la partie aglycone flexible du ligand, relaxeraient rapidement aux échelles de temps étudiées. Il s'avère en réalité qu'elles pourraient être un facteur cinétique important de la stabilité du complexe, au même titre que la coordonnée d'ouverture du site.

Cette expérience nous encourage à compléter notre approche par des méthodes non-supervisées de réduction de dimensionnalité visant à délimiter les principaux états métastables du système. Cette tâche hautement non-triviale est un sujet actif de recherche [17-20] et les développements récents en matière d'apprentissage profond offrent là-aussi des perspectives prometteuses [21, 22].

QUELQUES PERSPECTIVES

Dynamique du site de liaison

Par la nature des méthodes d'échantillonnage renforcé utilisées, notre étude de la plasticité conformationnelle du domaine lectine s'est concentrée sur ses propriétés thermodynamiques. La cinétique d'inter-conversion entre états d'ouverture est cependant une donnée cruciale dans la modélisation détaillée du mécanisme grappin. La théorie cinétique de KRAMERS permet en principe de recouvrer la vitesse de réaction à travers une barrière d'un profil d'énergie libre, à condition (i) que la coordonnée choisie est une coordonnée de réaction raisonnable et (ii) de pouvoir estimer le coefficient de diffusion selon cette coordonnée, le long du chemin réactionnel.

Une autre approche, de plus en plus populaire, est la construction de modèles de MARKOV (MSM, MARKOV *state model*) à partir d'un ensemble de trajectoires non biaisées [23, 24]. Contrairement au modèle simple du chapitre 6, les MSM construits à partir de simulations de macromolécules discrétisent l'espace des configuration en des milliers voire des dizaines de milliers d'états. Sous certaines conditions, le modèle obtenu décrit de façon cohérente les propriétés thermodynamiques et cinétiques du système aux temps longs. L'exploration avancée de l'espace conformationnel dans nos travaux pourrait servir de base à la construction d'un MSM, par exemple en focalisant l'échantillonnage de courtes trajectoires à partir de configurations déjà visitées au voisinage des barrières énergétiques.

Mécanisme de la transition $A \leftrightarrow S$

La description du mécanisme détaillé de la transition allostérique $A \leftrightarrow S$ de FimH demeure encore largement méconnue. À l'instar de bien d'autres protéines, son étude par des méthodes de dynamique moléculaire reste limitée par le caractère rare d'une telle transition [25]. Le modèle cinétique de [26] attribue un temps de vie de plusieurs secondes aux deux états allostériques en l'absence de force. La SMD peut permettre d'accélérer la séparation – relativement aisée – des domaines associés [4], mais la transformation subséquente du renflement bêta/switch alpha s'effectue vraisemblablement selon une coordonnée de réaction hautement non linéaire, difficile à identifier et à biaiser.

Durant notre investigation, nous avons tenté d'appliquer une méthode de recherche de chemin réactionnel d'énergie minimale fondé sur une CV de chemin adaptative combinée à de la métadynamique [27, 28]. Malgré nos efforts, nous ne sommes pas parvenus à produire des résultats stables et concluants et nous avons fait le choix de ne pas inclure cette phase exploratoire dans cette thèse. La recherche de chemin réactionnel pour d'importantes réorganisations macromoléculaires est particulièrement complexe en raison de la difficulté à identifier à priori, à partir des structures statiques, un espace de CV de dimension modeste au sein duquel effectuer une telle recherche. La notion même d'un « chemin » préférentiel, c'est-à-dire une unique « vallée » reliant deux bassins du paysage énergétique, et pouvant être décrit simplement par une unique coordonnée, est sujette à caution. D'autres techniques d'échantillonnage de chemin de transition (*transition path sampling*) reposant sur une vision d'ensemble de chemins et ne nécessitant pas de définir au préalable un espace de CV ont été développées [29, 30] et pourraient être appliquées à FimH.

Modélisation intégrative

Les propriétés physico-chimiques de FimH ne forment qu'un aspect du processus d'adhérence bactérienne en conditions biologiques. La description globale du phénomène nécessite l'intégration d'un grand nombre d'éléments : la forme de la bactérie, le nombre et la répartition des ligands sur la surface d'ancrage, la dynamique d'écoulement du fluide environnant, les propriétés mécaniques de l'ensemble du pilus, la présence éventuelle d'inhibiteurs solubles, etc. Une telle modélisation est intrinsèquement multi-échelle car elle mêle plusieurs niveaux de détail spatio-temporels. L'adhésine et son ligand, par exemple, doivent être traités à l'échelle moléculaire, alors que le fluide et la bactérie peuvent se satisfaire d'une description mésoscopique. Dans cette optique, des méthodes hybrides combinant dynamique des particules et description efficiente de l'hydrodynamique ont été développées, et peuvent être directement appliquées à l'étude multi-échelle des grappins moléculaires, en complément des simulations atomistiques [31-33].

En parallèle de notre investigation sur FimH, nous avons effectué quelques pas dans cette direction et employé ce type d'approche pour étudier la stabilité mécanique des protéines en écoulement cisaillant. Nous présentons en annexe C deux articles issus de ces recherches.

CONCLUSION

L'adhésine bactérienne FimH est une protéine modèle pour l'étude des liaisons accrocheuses, cet intrigant phénomène à la frontière entre la biologie, la chimie et la physique, et qui se révèle central dans l'organisation mécanique du vivant. En dépit de formidables développements expérimentaux permettant de sonder des échelles toujours plus courtes de temps et d'espace, de nombreux processus biochimiques demeurent inaccessibles à l'observation directe. Le « microscope computationnel » [2] que représente la dynamique moléculaire nous en propose un film haute-résolution – pourvu que l'on puisse démêler les faits de la fiction. Grâce à cette fenêtre sur l'infiniment petit, nous avons tenté de décrire avec une précision inégalée la plasticité du site de liaison de FimH, et ses implications sur le mécanisme grappin de l'adhésine. Nous espérons que les connaissances et les méthodes développées dans nos travaux contribueront à mieux appréhender les propriétés d'adhérence des bactéries pathogènes afin d'élaborer de nouvelles thérapies, et plus généralement à affiner la compréhension des processus biomécaniques aux échelles moléculaires.

RÉFÉRENCES

- [1] J. R. PERILLA, B. C. GOH, C. K. CASSIDY, B. LIU, R. C. BERNARDI, T. RUDACK, H. YU, Z. WU et K. SCHULTEN. Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, **31**, 64-74 (2015).
DOI : [10.1016/j.sbi.2015.03.007](https://doi.org/10.1016/j.sbi.2015.03.007)
- [2] E. H. LEE, J. HSIN, M. SOTOMAYOR, G. COMELLAS et K. SCHULTEN. Discovery Through the Computational Microscope. *Structure*, **17**, 1295-1306 (2009).
DOI : [10.1016/j.str.2009.09.001](https://doi.org/10.1016/j.str.2009.09.001)
- [3] W. E. THOMAS, E. TRINTCHINA, M. FORERO, V. VOGEL et E. V. SOKURENKO. Bacterial Adhesion to Target Cells Enhanced by Shear Force. *Cell*, **109**, 913-923 (2002).
DOI : [10.1016/S0092-8674\(02\)00796-1](https://doi.org/10.1016/S0092-8674(02)00796-1)
- [4] P. APRIKIAN et al. The Bacterial Fimbrial Tip Acts as a Mechanical Force Sensor. *PLoS Biology*, **9**, e1000617 (2011).
DOI : [10.1371/journal.pbio.1000617](https://doi.org/10.1371/journal.pbio.1000617)
- [5] A. ALONSO-CABALLERO, J. SCHÖNFELDER, S. POLY, F. CORSETTI, D. DE SANCHO, E. ARTACHO et R. PEREZ-JIMENEZ. Mechanical architecture and folding of E. coli type 1 pilus domains. *Nature Communications*, **9**, (2018).
DOI : [10.1038/s41467-018-05107-6](https://doi.org/10.1038/s41467-018-05107-6)
- [6] L. M. NILSSON, O. YAKOVENKO, V. TCHESNOKOVA, W. E. THOMAS, M. A. SCHEMBRI, V. VOGEL, P. KLEMM et E. V. SOKURENKO. The cysteine bond in the *Escherichia coli* FimH adhesin is critical for adhesion under flow conditions. *Molecular Microbiology*, **65**, 1158-1169 (2007).
DOI : [10.1111/j.1365-2958.2007.05858.x](https://doi.org/10.1111/j.1365-2958.2007.05858.x)
- [7] V. B. RODRIGUEZ, B. A. KIDD, G. INTERLANDI, V. TCHESNOKOVA, E. V. SOKURENKO et W. E. THOMAS. Allosteric Coupling in the Bacterial Adhesive Protein FimH. *Journal of Biological Chemistry*, **288**, 24128-24139 (2013).
DOI : [10.1074/jbc.M113.461376](https://doi.org/10.1074/jbc.M113.461376)
- [8] M. SINGARAVELU, A. SELVAN et S. ANISHETTY. Molecular dynamics simulations of lectin domain of FimH and immunoinformatics for the design of potential vaccine candidates. *Computational Biology and Chemistry*, **52**, 18-24 (2014).
DOI : [10.1016/j.combiolchem.2014.08.002](https://doi.org/10.1016/j.combiolchem.2014.08.002)
- [9] M. M. SAUER, R. P. JAKOB, J. ERAS, S. BADAY, D. ERIŞ, G. NAVARRA, S. BERNÈCHE, B. ERNST, T. MAIER et R. GLOCKSHUBER. Catch-bond mechanism of the bacterial adhesin FimH. *Nature Communications*, **7**, (2016).
DOI : [10.1038/ncomms10738](https://doi.org/10.1038/ncomms10738)
- [10] G. INTERLANDI et W. E. THOMAS. Mechanism of allosteric propagation across a β -sheet structure investigated by molecular dynamics simulations : β -Sheet Allosteric Mechanism. *Proteins : Structure, Function, and Bioinformatics*, **84**, 990-1008 (2016).
DOI : [10.1002/prot.25050](https://doi.org/10.1002/prot.25050)

- [11] D. E. SHAW et al. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, **51**, 91-97 (2008).
DOI : [10.1145/1364782.1364802](https://doi.org/10.1145/1364782.1364802)
- [12] J. LIU, L. A. N. AMARAL et S. KETEN. Conformational stability of the bacterial adhesin, FimH, with an inactivating mutation. *Proteins : Structure, Function, and Bioinformatics*, **89**, 276-288 (2021).
DOI : [10.1002/prot.26013](https://doi.org/10.1002/prot.26013)
- [13] D. van der SPOEL. Systematic design of biomolecular force fields. *Current Opinion in Structural Biology*, **67**, 18-24 (2021).
DOI : [10.1016/j.sbi.2020.08.006](https://doi.org/10.1016/j.sbi.2020.08.006)
- [14] J. MU, H. LIU, J. ZHANG, R. LUO et H.-F. CHEN. Recent Force Field Strategies for Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling*, **61**, 1037-1047 (2021).
DOI : [10.1021/acs.jcim.0c01175](https://doi.org/10.1021/acs.jcim.0c01175)
- [15] S. CHMIELA, H. E. SAUCEDA, K.-R. MÜLLER et A. TKATCHENKO. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, **9**, 3887 (2018).
DOI : [10.1038/s41467-018-06169-2](https://doi.org/10.1038/s41467-018-06169-2)
- [16] F. SITTEL et G. STOCK. Perspective : Identification of collective variables and metastable states of protein dynamics. *The Journal of Chemical Physics*, **149**, 150901 (2018).
DOI : [10.1063/1.5049637](https://doi.org/10.1063/1.5049637)
- [17] D. MENDELS, G. PICCINI et M. PARRINELLO. Collective Variables from Local Fluctuations. *The Journal of Physical Chemistry Letters* (2018).
DOI : [10.1021/acs.jpcllett.8b00733](https://doi.org/10.1021/acs.jpcllett.8b00733)
- [18] H. WU et F. NOÉ. Variational approach for learning Markov processes from time series data. *arXiv :1707.04659 [math, stat]* (2019).
arXiv : [1707.04659](https://arxiv.org/abs/1707.04659)
- [19] G. A. TRIBELLO et P. GASPAROTTO. Using Dimensionality Reduction to Analyze Protein Trajectories. *Frontiers in Molecular Biosciences*, **6**, 46 (2019).
DOI : [10.3389/fmolb.2019.00046](https://doi.org/10.3389/fmolb.2019.00046)
- [20] A. GLIELMO, B. E. HUSIC, A. RODRIGUEZ, C. CLEMENTI, F. NOÉ et A. LAIO. Unsupervised Learning Methods for Molecular Simulation Data. *Chemical Reviews* [acs.chemrev.0c01195](https://doi.org/10.1021/acs.chemrev.0c01195) (2021).
DOI : [10.1021/acs.chemrev.0c01195](https://doi.org/10.1021/acs.chemrev.0c01195)
- [21] A. MARDT, L. PASQUALI, H. WU et F. NOÉ. VAMPnets for deep learning of molecular kinetics. *Nature Communications*, **9**, 5 (2018).
DOI : [10.1038/s41467-017-02388-1](https://doi.org/10.1038/s41467-017-02388-1)
- [22] W. CHEN, H. SIDKY et A. L. FERGUSON. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *The Journal of Chemical Physics*, **150**, 214114 (2019).
DOI : [10.1063/1.5092521](https://doi.org/10.1063/1.5092521)
- [23] B. TRENDELKAMP-SCHROER et F. NOÉ. Efficient Estimation of Rare-Event Kinetics. *Physical Review X*, **6**, 011009 (2016).
DOI : [10.1103/PhysRevX.6.011009](https://doi.org/10.1103/PhysRevX.6.011009)

- [24] W. WANG, S. CAO, L. ZHU et X. HUANG. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *WIREs Computational Molecular Science*, **8**, e1343 (2018). DOI : [10.1002/wcms.1343](https://doi.org/10.1002/wcms.1343)
- [25] S. J. WODAK et al. Allostery in Its Many Disguises : From Theory to Applications. *Structure*, **27**, 566-578 (2019). DOI : [10.1016/j.str.2019.01.003](https://doi.org/10.1016/j.str.2019.01.003)
- [26] W. THOMAS, M. FORERO, O. YAKOVENKO, L. NILSSON, P. VICINI, E. SOKURENKO et V. VOGEL. Catch-Bond Model Derived from Allostery Explains Force-Activated Bacterial Adhesion. *Biophysical Journal*, **90**, 753-764 (2006). DOI : [10.1529/biophysj.105.066548](https://doi.org/10.1529/biophysj.105.066548)
- [27] A. PÉREZ DE ALBA ORTÍZ, A. TIWARI, R. C. PUTHENKALATHIL et B. ENSING. Advances in enhanced sampling along adaptive paths of collective variables. *The Journal of Chemical Physics*, **149**, 072320 (2018). DOI : [10.1063/1.5027392](https://doi.org/10.1063/1.5027392)
- [28] A. PÉREZ DE ALBA ORTÍZ, J. VREEDE et B. ENSING. The Adaptive Path Collective Variable : A Versatile Biasing Approach to Compute the Average Transition Path and Free Energy of Molecular Transitions. *Biomolecular Simulations*. Sous la dir. de M. BONOMI et C. CAMILLONI. T. 2022. New York, NY : Springer New York, 2019. 255-290. DOI : [10.1007/978-1-4939-9608-7_11](https://doi.org/10.1007/978-1-4939-9608-7_11)
- [29] L. T. CHONG, A. S. SAGLAM et D. M. ZUCKERMAN. Path-sampling strategies for simulating rare events in biomolecular systems. *Current Opinion in Structural Biology*, **43**, 88-94 (2017). DOI : [10.1016/j.sbi.2016.11.019](https://doi.org/10.1016/j.sbi.2016.11.019)
- [30] D. W. H. SWENSON, J.-H. PRINZ, F. NOE, J. D. CHODERA et P. G. BOLHUIS. OpenPathSampling : A Python Framework for Path Sampling Simulations. 1. Basics. *Journal of Chemical Theory and Computation*, **15**, 813-836 (2019). DOI : [10.1021/acs.jctc.8b00626](https://doi.org/10.1021/acs.jctc.8b00626)
- [31] F. STERPONE, P. DERREUMAUX et S. MELCHIONNA. Protein Simulations in Fluids : Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. *Journal of Chemical Theory and Computation*, **11**, 1843-1853 (2015). DOI : [10.1021/ct501015h](https://doi.org/10.1021/ct501015h)
- [32] X. ZHAO, Y. LIU, Z. GUO, Y. ZHANG, Y. LI et W. LIU. Mechanical response and deformation mechanics of Type IV pili investigated using steered coarse-grained molecular dynamics simulation. *Journal of Biomechanics*, **56**, 97-101 (2017). DOI : [10.1016/j.jbiomech.2017.03.013](https://doi.org/10.1016/j.jbiomech.2017.03.013)
- [33] F. STERPONE, S. DOUTRELIGNE, T. T. TRAN, S. MELCHIONNA, M. BAADEN, P. H. NGUYEN et P. DERREUMAUX. Multi-scale simulations of biological systems using the OPEP coarse-grained model. *Biochemical and Biophysical Research Communications*, **498**, 296-304 (2018). DOI : [10.1016/j.bbrc.2017.08.165](https://doi.org/10.1016/j.bbrc.2017.08.165)

Cinquième partie

ANNEXES

MÉTHODES

A.1 CLUSTERING HIÉRARCHIQUE À ESTIMATION DE DENSITÉ

Nous décrivons succinctement cette approche, qui se décompose en deux temps :

1. CHOIX D'UN ESPACE DE COORDONNÉES ADAPTÉ : les coordonnées cartésiennes sont peu adaptées à l'analyse conformationnelle des molécules. En effet, elles sont sensibles aux mouvements de rotation et translation globales, qui sont par symétrie indépendants des degrés de liberté internes. Il est possible d'éliminer en partie la composante roto-translationnelle par un alignement optimal (minimisant le RMSD) sur une référence. Toutefois, cette solution est suboptimale car pour un système flexible elle introduit des corrélations non-locales : par exemple, une flexion dans une région charnière changera l'alignement global et donc les coordonnées de l'ensemble des atomes. Pour des systèmes encore plus flexibles (par exemple une protéine intrinsèquement désordonnée) il n'est même pas toujours possible de définir un alignement ayant un sens. Il est donc préférable d'utiliser des coordonnées dites internes, naturellement invariantes vis-à-vis des rotations et translations globales [1]. Nous utilisons les distances entre paires de carbones α du domaine lectine (rés. 1–158) comme base pour le clustering.
2. FILTRAGE DES COORDONNÉES : l'ensemble des distances de paires pour $n_{C\alpha} = 158$ atomes génère pas moins de $(n_{C\alpha})(n_{C\alpha} - 1)/2 = 12403$ coordonnées, ce qui peut rapidement devenir coûteux pour l'analyse subséquente. En réalité, cet ensemble de coordonnées est hautement redondant par rapport aux $3n_{C\alpha} - 6 = 468$ degrés de liberté internes de l'ensemble d'atomes considéré, et beaucoup de distances sont peu informatives si elles ne varient pas entre différents états du système. Afin de réduire la dimension de l'espace considéré et de sélectionner les coordonnées les plus informatives sur les modes lents du système, nous ne considérons que les distances entre atomes formant des contacts transitoires, c'est-à-dire ayant une valeur inférieure à $d_{\text{cont}} = 7 \text{ \AA}$ au moins une fois durant la trajectoire, et au plus 90% du temps. Les paires d'atomes voisins le long de la séquence protéique sont également ignorées jusqu'au troisième voisin inclus.

3. ANALYSE EN COMPOSANTES PRINCIPALE ET RÉDUCTION DIMENSIONNELLE : nous pratiquons une analyse en composantes principales (ACP) de la trajectoire sur l'espace des distances pré-sélectionnées. L'ACP permet de définir une base de nouvelles variables appelées composantes principales, qui sont des combinaisons linéaires des premières et qui maximisent la variance des données. Intuitivement l'ACP identifie un ensemble de directions orthogonales dans l'espace de départ, selon lesquelles l'ensemble des données varie le plus et contient donc le plus d'information. En ne gardant que les premières composantes (classées par ordre décroissant de variance) on obtient un espace de petite dimension qui conserve une part importante de l'information initiale.
4. ESTIMATION DE DENSITÉ : la méthode du clustering repose sur une estimation de la densité par noyau, sous la forme d'une « énergie libre locale », au voisinage de chaque échantillon :

$$F_i = -k_B T \ln \frac{P_i}{P_{\max}} \quad (82)$$

où la population P_i au voisinage du point i est évaluée en comptant les voisins à une distance inférieure à r_{noy} .

5. CONSTRUCTION HIÉRARCHIQUE PAR REMPLISSAGE DE BASSINS: un clustering hiérarchique à K niveaux est construit itérativement en « remplissant les bassins d'énergie libre ». Concrètement, l'algorithme suivant est utilisé :
 - a) au niveau $k = 0$ choisir une énergie seuil $F^0 = \Delta F$
 - b) parmi les points non classés, le point d'énergie minimale est choisi pour former un nouveau cluster
 - c) le cluster est étendu itérativement en agglomérant les points à une distance inférieure à d_{agg} d'au moins un point du cluster, et avec une énergie maximale F^k
 - d) s'il reste encore des points non classés d'énergie $< F^k$, retourner à l'étape (b)
 - e) sinon, si l'ensemble des données ont été classées, arrêter l'algorithme
 - f) sinon, passer au niveau supérieur $k \leftarrow k + 1$ et augmenter le seuil $F^{k+1} \leftarrow F^k + \Delta F$ puis reprendre en (b)

Le processus de clustering aboutit, dans le langage de la théorie des graphes, à une forêt d'arbres enracinés. On peut représenter cet ensemble de graphes en ordonnant les nœuds par niveau k d'énergie seuil F^k . Les feuilles de chaque arbre constituent des bassins locaux qui grandissent avec l'énergie seuil et fusionnent progressivement jusqu'à la racine.

Pour obtenir l'ensemble des clusters finaux :

- on élimine tous les nœuds comprenant moins de 10 membres,
- les clusters sont les derniers ancêtres de chaque feuille avant toute ramification.

Nous choisissons les paramètres $r_{\text{noy}} = 1,5$, $d_{\text{agg}} = 1,5$ et $\Delta F = 0,1k_B T$.

A.2 IMPLÉMENTATION MV POUR LE MODÈLE DE MARKOV DÉPENDANT DE LA FORCE

Soit un jeu de données correspondant à l'ensemble des M transitions observées d'un état i vers un état j , à la force F et avec un intervalle de temps δt (réunissant éventuellement plusieurs trajectoires indépendantes). La fonction de log-vraisemblance s'écrit

$$\begin{aligned} \ln \mathcal{L}(\theta) &= \sum_{m=1}^M \ln \mathbf{T}_{\theta}(F_m, \delta t_m)_{i_m, j_m} \\ &= \sum_{m=1}^M \ln \text{Exp}[\delta t_m \mathbf{K}_{\theta}(F_m)]_{i_m, j_m} \end{aligned} \quad (83)$$

Afin d'éviter toute confusion nous avons noté Exp l'exponentielle matricielle tandis que \exp et \ln dénotent respectivement l'exponentielle et le logarithme naturel élément par élément. La dérivée partielle par rapport à un paramètre θ_u s'exprime

$$\frac{\partial}{\partial \theta_u} \ln \mathcal{L}(\theta) = \sum_{m=1}^M \frac{1}{\mathbf{T}_{\theta}(F_m, \delta t_m)_{i_m, j_m}} \frac{\partial}{\partial \theta_u} \mathbf{T}_{\theta}(F_m, \delta t_m)_{i_m, j_m} \quad (84)$$

Nous exprimons \mathbf{T}_{θ} en faisant l'approximation suivante :

$$\mathbf{T}_{\theta}(F_m, \delta t_m) = \text{Exp}[\delta t_m \mathbf{K}_{\theta}(F_m)] \approx \mathbf{I} + \delta t_m \mathbf{K}_{\theta}(F_m) \quad (85)$$

Cette approximation est raisonnable pour des temps δt_m petits devant les temps caractéristiques $k_{ij}(F_m)$. On en déduit

$$\mathbf{T}_{\theta}(F_m, \delta t_m)_{i_m, j_m} = \begin{cases} 1 - \delta t \sum_{j \neq i_m} k_{i_m, j}^0 \exp(F_m \Delta x_{i_m, j}) & \text{si } i_m = j_m \\ \delta t k_{i_m, j_m}^0 \exp(F_m \Delta x_{i_m, j_m}) & \text{sinon.} \end{cases} \quad (86)$$

Nous pouvons alors calculer les gradients pour divers paramètres :

$$\frac{\partial}{\partial k_{ij}^0} \mathbf{T}_\theta(F_m, \delta t_m)_{i_m, j_m} = \begin{cases} -\delta t \exp(F_m \Delta x_{ij}) & \text{si } i_m = i = j_m = j \\ \delta t \exp(F_m \Delta x_{ij}) & \text{si } i_m = i \neq j_m = j \\ 0 & \text{sinon} \end{cases} \quad (87)$$

$$\frac{\partial}{\partial \Delta x_{ij}} \mathbf{T}_\theta(F_m, \delta t_m)_{i_m, j_m} = \begin{cases} -\delta t k_{ij}^0 F_m \exp(F_m \Delta x_{ij}) & \text{si } i_m = i = j_m = j \\ \delta t k_{ij}^0 F_m \exp(F_m \Delta x_{ij}) & \text{si } i_m = i \neq j_m = j \\ 0 & \text{sinon} \end{cases} \quad (88)$$

Nous implémentons ces gradients dans une fonction compilée en Cython. Pour des raisons numériques les valeurs de \mathbf{T}_θ sont maintenues dans un intervalle $[\epsilon; 1]$ pendant le calcul, où ϵ est un réel très petit ($\epsilon = 10^{-12}$).

FIGURES COMPLÉMENTAIRES

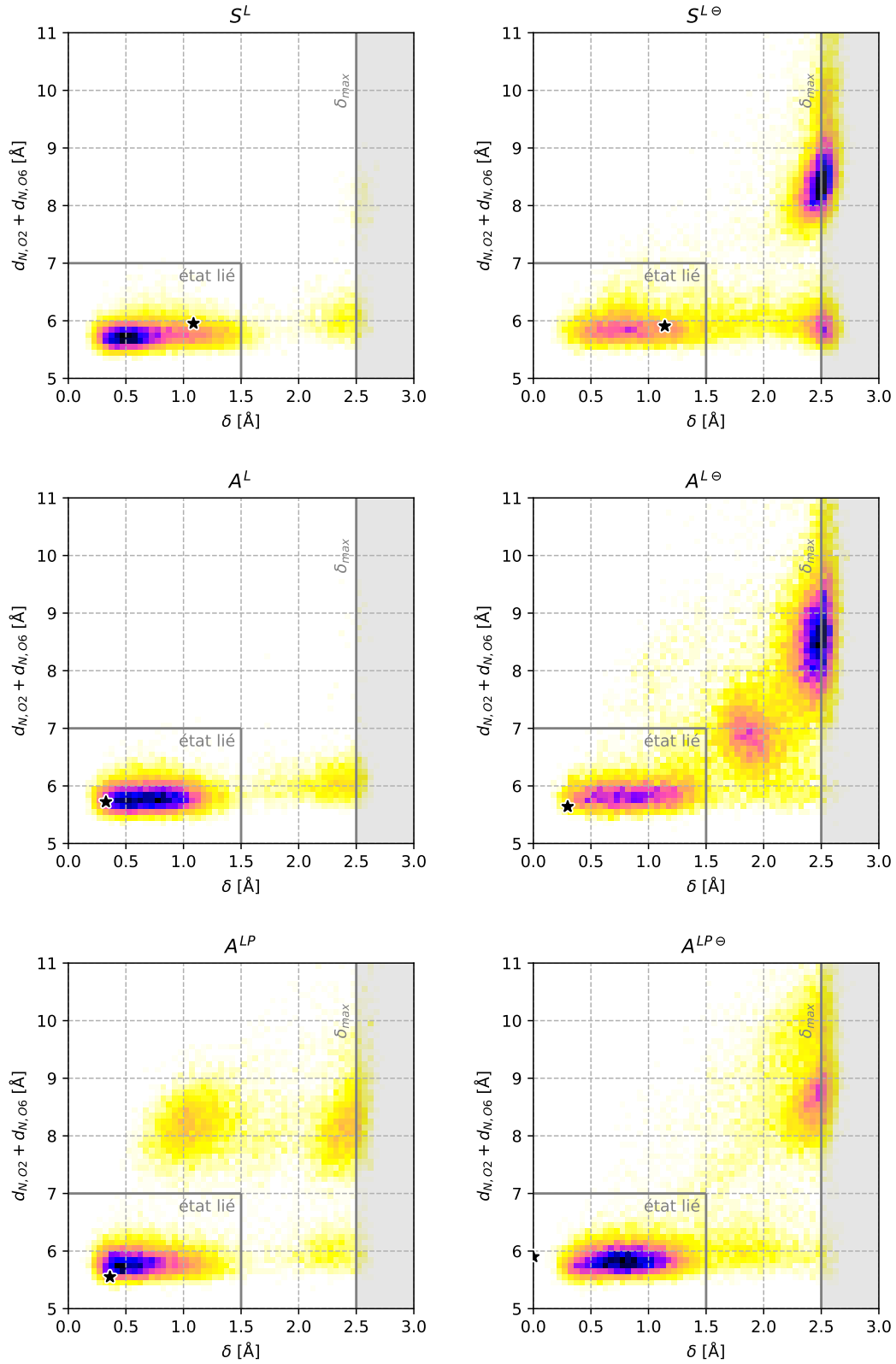


FIGURE B.1 – Définition de l'état lié. La distribution jointe formée par δ et $d_{N,O2} + d_{N,O6}$ dans les trajectoires REST2. La séparation nette du bassin conformationnel correspondant à l'état lié (*en bas à gauche*) des autres configurations explorées est visible dans chaque système. L'histogramme bidimensionnel est coloré par densité de probabilité croissante du blanc au noir. La configuration initiale est dénotée par une étoile (*). Si la position précise dans l'espace $(\delta, d_{N,O2} + d_{N,O6})$ varie d'une structure initiale à l'autre, l'état lié occupe à l'équilibre un bassin très similaire.

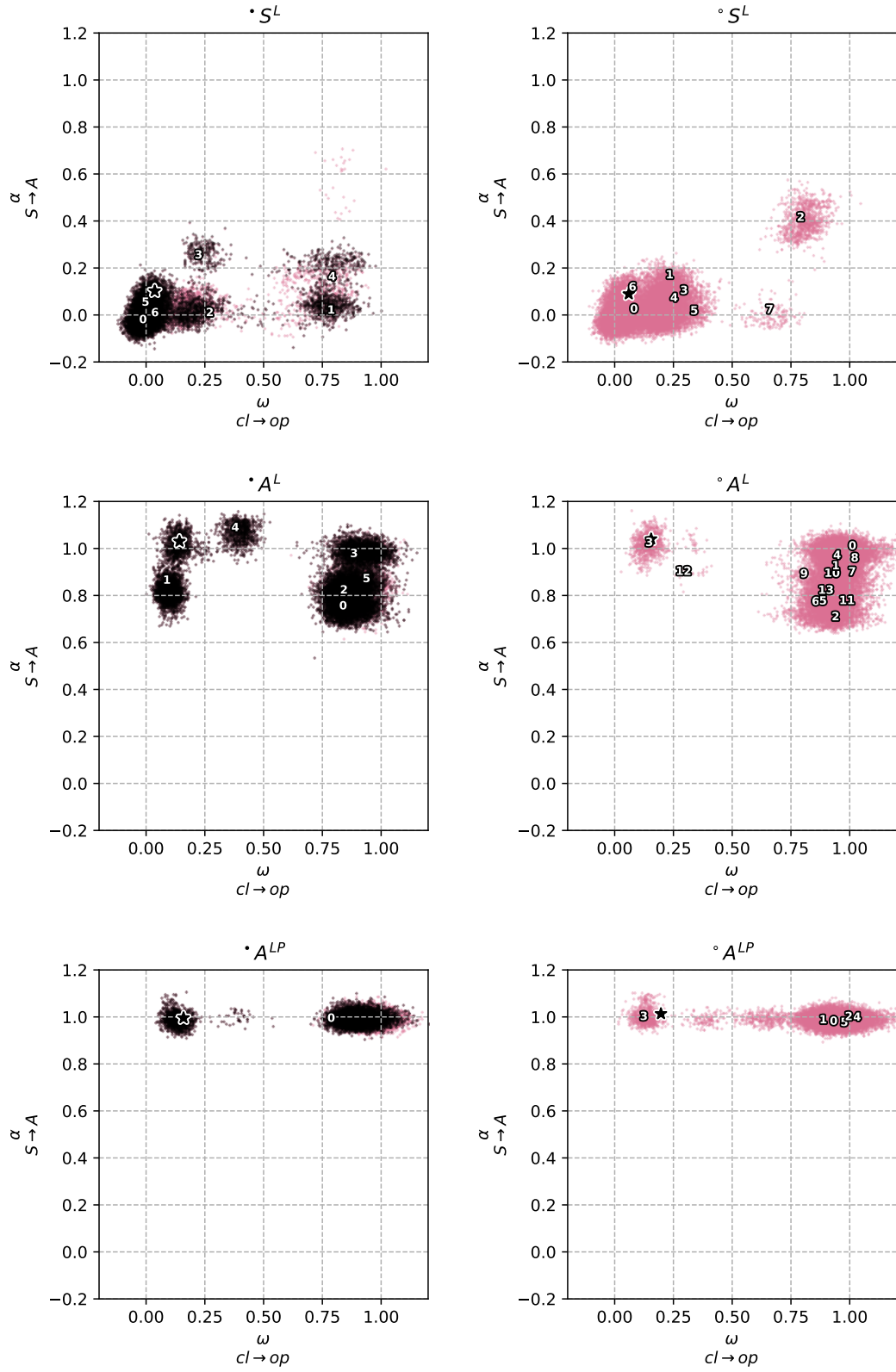


FIGURE B.2 – Paysage conformationnel exploré durant REST₂ (systèmes diprotonés). Trajectoires REST₂ projetées sur l'espace (ω, α) . Les points roses correspondent aux configurations non-liées. Les clusters sont numérotés. La configuration initiale est dénotée par une étoile (★).

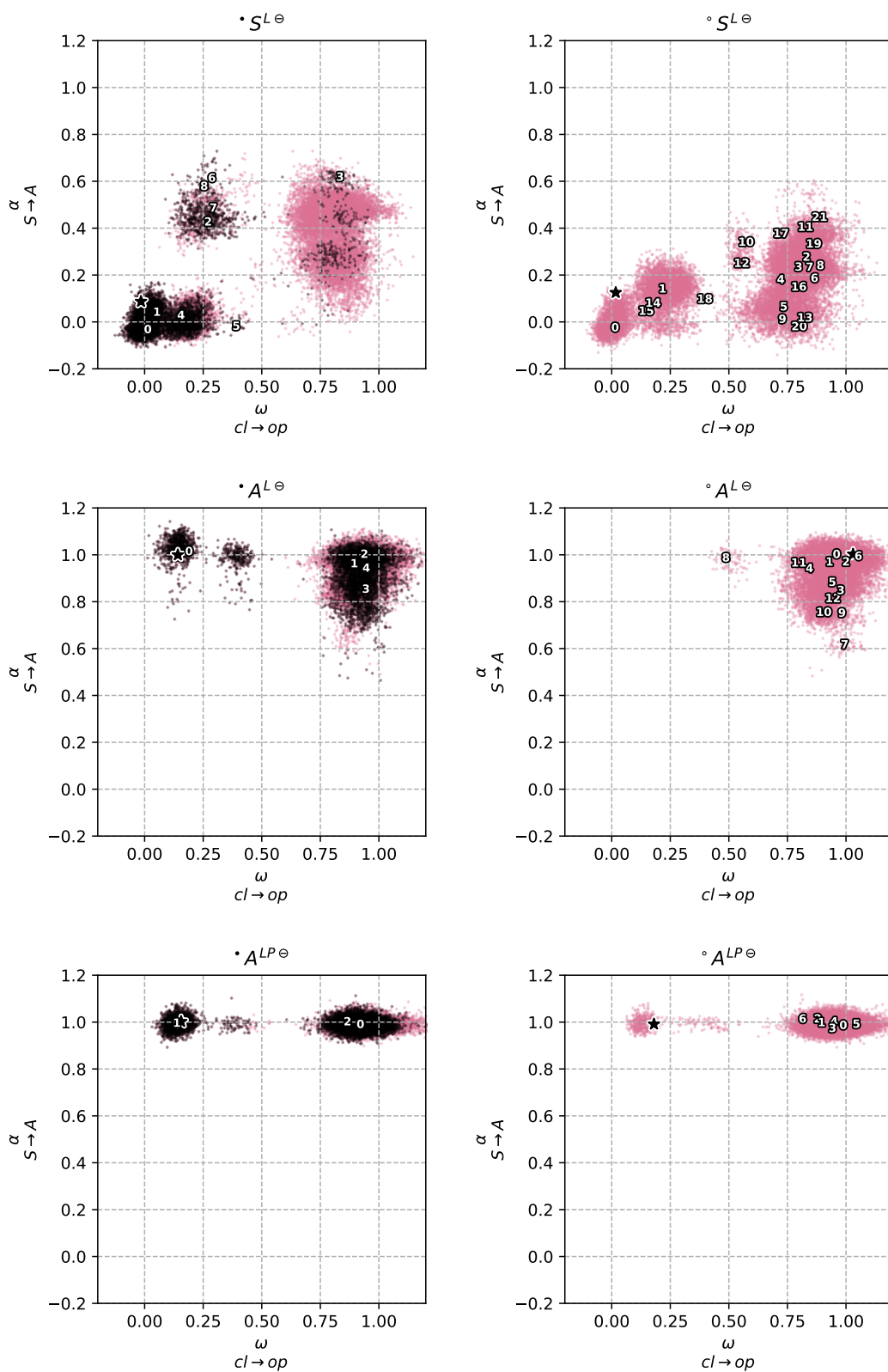


FIGURE B.3 – Paysage conformationnel exploré durant REST2 (systèmes monoprotés). Trajectoires REST2 projetées sur l'espace (ω, α) . Les points roses correspondent aux configurations non-liées. Les clusters sont numérotés. La configuration initiale est dénotée par une étoile (*).

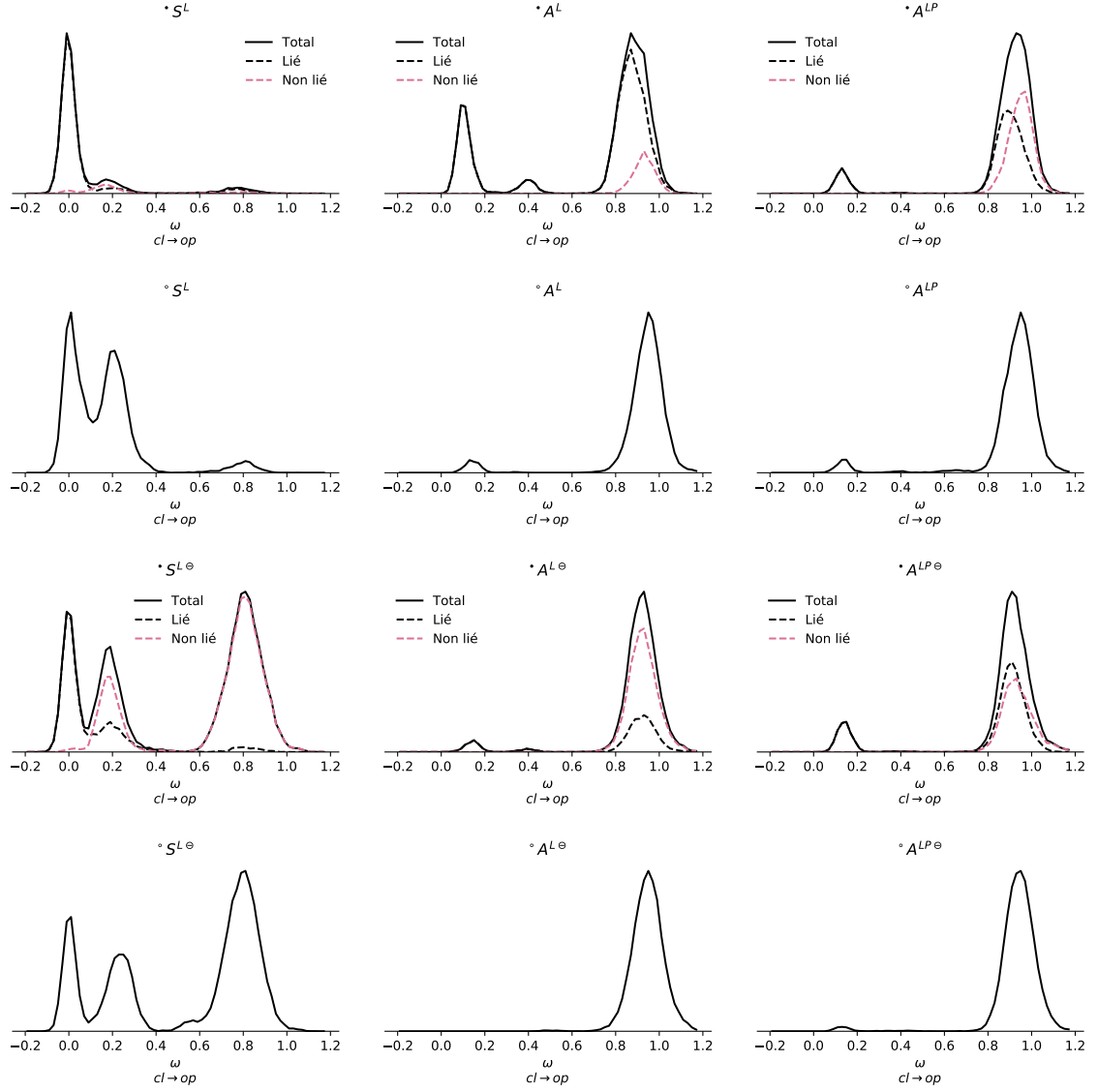


FIGURE B.4 – Distribution marginale de la variable d'ouverture dans REST₂. La projection de la trajectoire sur ω permet d'identifier clairement dans quasiment toutes les trajectoires trois pics correspondant à des sites de liaison ouverts (op), semi-ouverts (so) et clos (cl). Pour les systèmes ligantés, la décomposition des distributions entre conformations liées et non-liées (selon le critère strict de l'ÉQUATION 26) est montrée.

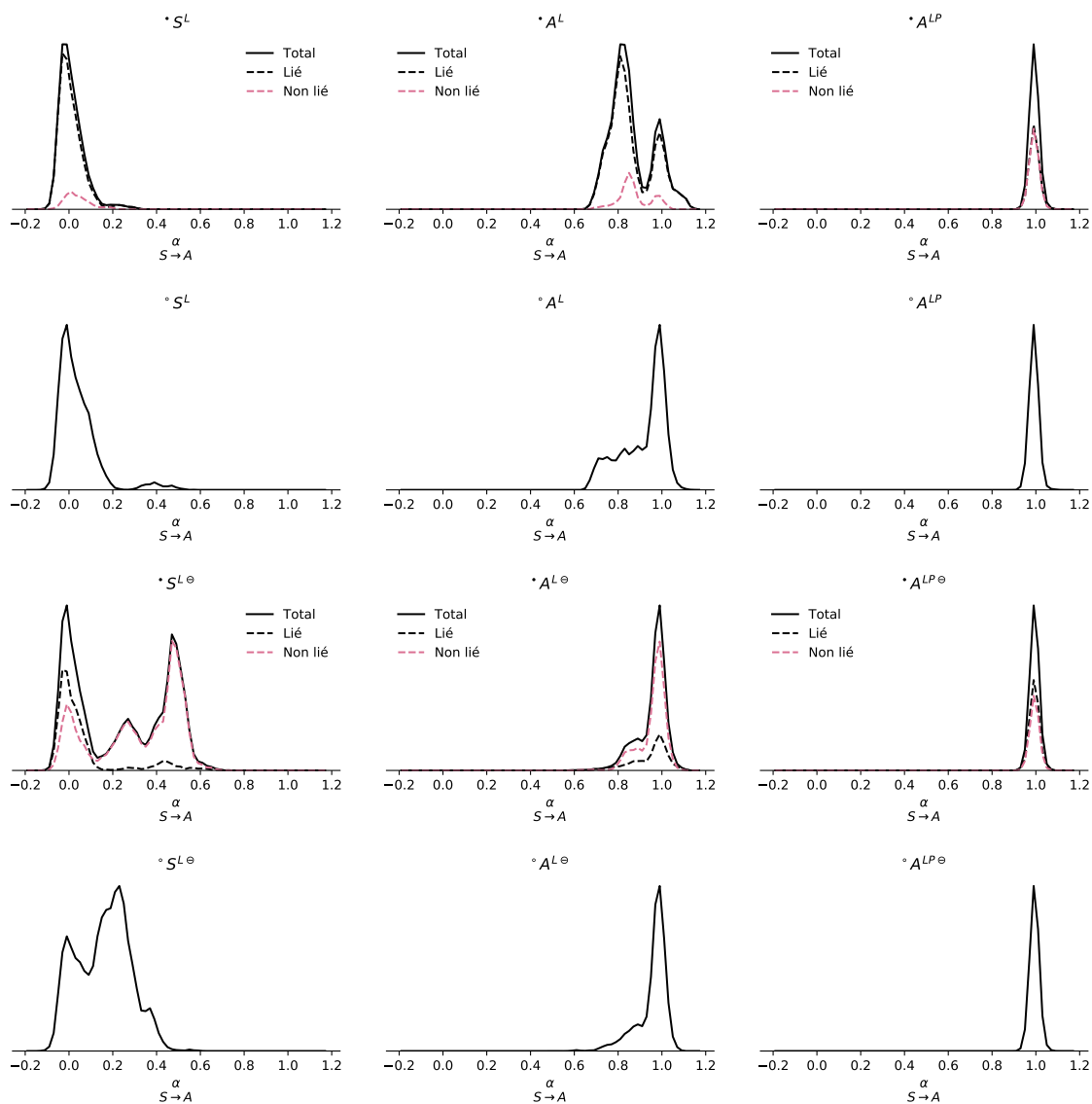


FIGURE B.5 – Distribution marginale de la variable d'association dans REST2. La projection de la trajectoire sur α montre une certaine flexibilité de la région inter-domaines, à l'exception des systèmes A^{LP} pour lesquels la présence du domaine piline stabilise fortement la conformation de référence. Pour les systèmes ligantés, la décomposition des distributions entre conformations liées et non-liées (selon le critère strict de l'ÉQUATION 26) est montrée.

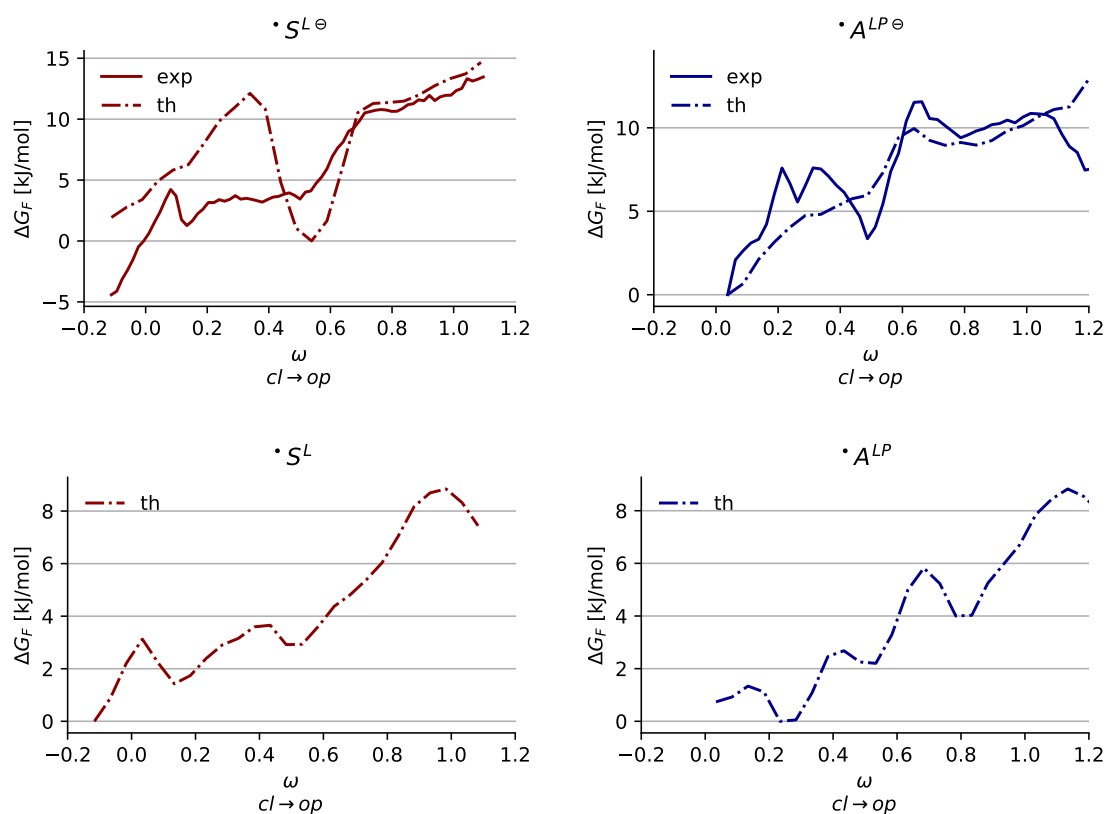


FIGURE B.6 – Prédiction de l'effet de la force à partir des simulations non-perturbées. L'altération théorique du PEL sous force ($F = 50$ pN) est calculée (ÉQUATION 42) par repondération des trajectoires en l'absence de force (th) et comparée à la référence obtenue par simulations BEUS biaisées (exp). Le minimum est arbitrairement fixé à $\Delta G_F = 0$.

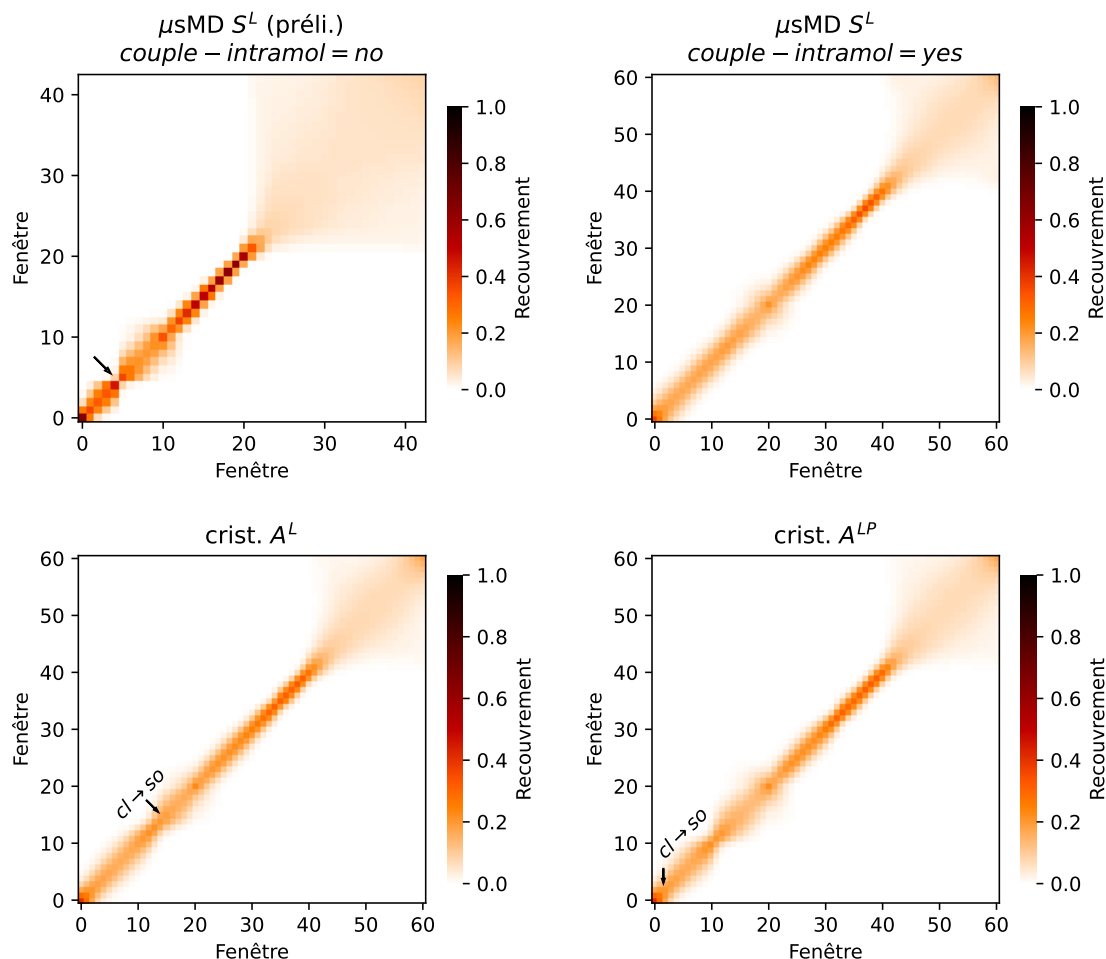


FIGURE B.7 – Exemples de matrices de recouvrement pour le découplage alchimique. Le recouvrement est la probabilité qu’une configuration échantillonnée dans l’état i soit observée dans l’état j . Une valeur diagonale proche de 1 signifie peu de recouvrement avec les états limitrophes. (*haut*) Des essais préliminaires sans découplage des interactions intra-ligand montrent à une discontinuité de recouvrement indiquée par une flèche, ainsi qu’une répartition et un nombre de fenêtres sous-optimaux. Le découplage des interactions intra-ligand (couple-intramol = yes) permet de contourner le problème. (*bas*) L’ouverture partielle du site dans les simulations sans restriction n’entraîne pas de problème de recouvrement notable mais rend difficile l’interprétation du $\Delta G_{\text{déc}}$ calculé.

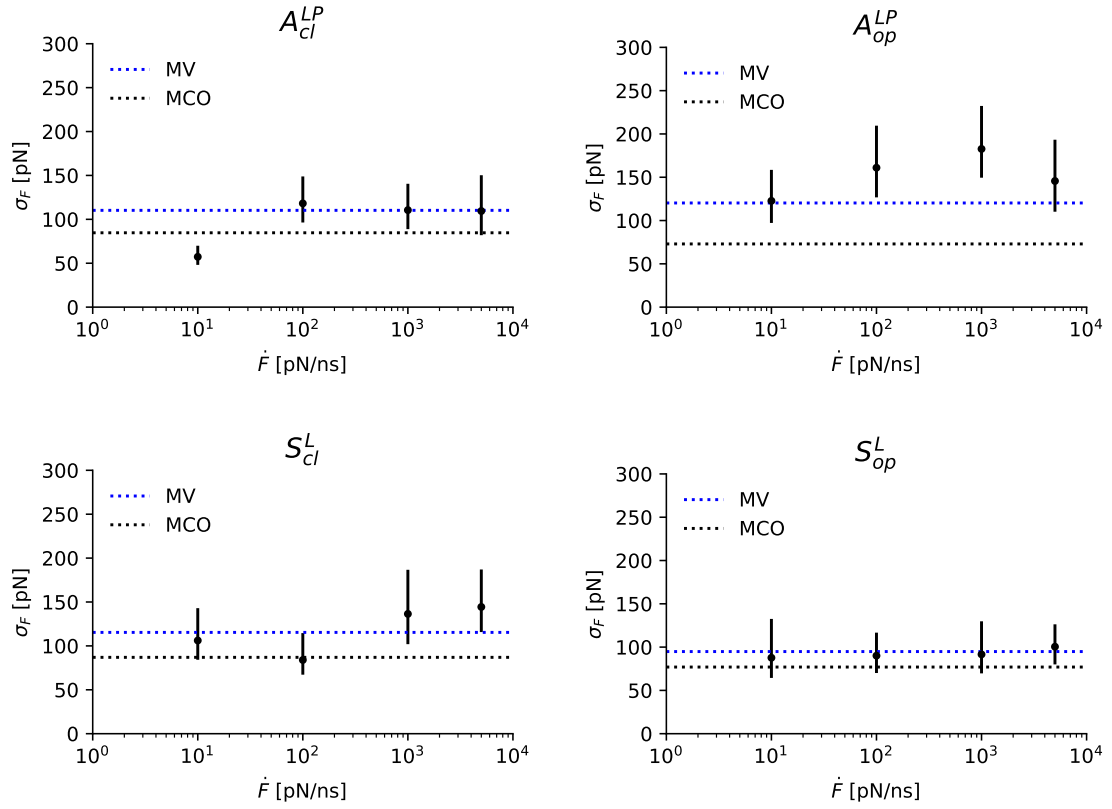


FIGURE B.8 – Écart type de la distribution des forces de rupture en fonction de la vitesse de charge. Les valeurs empiriques (points noirs) sont indiquées avec un intervalle de confiance à 95 % estimé par bootstrap. Les valeurs prédites par le modèle de Bell-Evans après ajustement par maximum de vraisemblance (MV) ou moindres carrés non-linéaires (MCNL) sont indiquées en pointillés.

PUBLICATIONS

Si les travaux de cette thèse ont pour objet principal l'adhésine FimH, nous avons également contribué à la publication de deux articles aux thématiques très voisines que nous souhaitons intégrer dans cette annexe. Ces articles font suite à des recherches sur la stabilité mécaniques des protéines soumises à des écoulements cisailants, initiées en stage de pré-thèse et continuées en parallèle de notre projet principal.

Nous avons mentionné en Introduction l'importance des interactions hydrodynamiques dans de nombreux processus physiologiques. Dans le cas de l'adhérence de bactéries aux surfaces *via* des adhésines situées aux extrémités piliales, nous avons modélisé l'effet de ces écoulements par une force de tension directionnelle appliquée à la liaison. Cependant, certains phénomènes nécessitent une description plus complète des écoulements fluides et de leurs interactions avec les objets biologiques. En particulier, les protéines sont des macromolécules flexibles pouvant réagir à des forces hydrodynamiques.

C.1 DÉPLIEMENT PROTÉIQUE SOUS FORCE, CHALEUR ET CISAILLEMENT FLUIDE

Dans ce premier article nous nous intéressons à la stabilité d'une petite protéine globulaire de réponse au choc froid (*cold-shock protein*) lorsqu'elle est soumise à des perturbations de différentes natures : un fort cisaillement hydrodynamique, une force directionnelle appliquée aux extrémités et une montée en température. À cette fin nous appliquons une approche couplant de la dynamique moléculaire gros-grain à une méthode de BOLTZMANN sur réseau (*lattice Boltzmann*) pour une description mésoscopique du solvant. Nous montrons que chaque perturbation suscite un chemin de dépliement différent, correspondant à un point faible spécifique au sein de la structure native. Nos résultats illustrent l'utilité de développer des approches multi-échelles prenant en compte explicitement les interactions hydrodynamiques pour l'étude de certains phénomènes biologiques, comme la coagulation sanguine ou les liaisons accrocheuses.

Three Weaknesses for Three Perturbations: Comparing Protein Unfolding Under Shear, Force, and Thermal Stresses

Published as part of *The Journal of Physical Chemistry virtual special issue "Deciphering Molecular Complexity in Dynamics and Kinetics from the Single Molecule to the Single Cell Level"*.

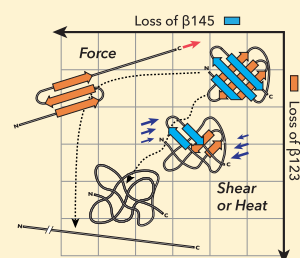
Olivier Languin-Cattoën,[†] Simone Melchionna,[‡] Philippe Derreumaux,[†] Guillaume Stirnemann,^{*,†} and Fabio Sterpone^{*,†}

[†]Laboratoire de Biochimie Théorique, CNRS, Institut de Biologie Physico-Chimique, Sorbonne Paris Cité, PSL University, 13 rue Pierre et Marie Curie, 75005 Paris, France

[‡]ISC–CNR, via dei Taurini 19, 00185 Rome, Italy

Supporting Information

ABSTRACT: The perturbation of a protein conformation by a physiological fluid flow is crucial in various biological processes including blood clotting and bacterial adhesion to human tissues. Investigating such mechanisms by computer simulations is thus of great interest, but it requires development of ad hoc strategies to mimic the complex hydrodynamic interactions acting on the protein from the surrounding flow. In this study, we apply the Lattice Boltzmann Molecular Dynamics (LBMD) technique built on the implicit solvent coarse-grained model for protein Optimized Potential for Efficient peptide structure Prediction (OPEP) and a mesoscopic representation of the fluid solvent, to simulate the unfolding of a small globular cold-shock protein in shear flow and to compare it to the unfolding mechanisms caused either by mechanical or thermal perturbations. We show that each perturbation probes a specific weakness of the protein and causes the disruption of the native fold along different unfolding pathways. Notably, the shear flow and the thermal unfolding exhibit very similar pathways, while because of the directionality of the perturbation, the unfolding under force is quite different. For force and thermal disruption of the native state, the coarse-grained simulations are compared to all-atom simulations in explicit solvent, showing an excellent agreement in the explored unfolding mechanisms. These findings encourage the use of LBMD based on the OPEP model to investigate how a flow can affect the function of larger proteins, for example, in catch-bond systems.



INTRODUCTION

Many biological processes depend on the protein response to the mechanical perturbations arising from a solvent in motion.^{1,2} For example, during blood coagulation, a large concatemeric blood protein, the von Willebrand Factor (vWf), senses the modifications in the bloodstream caused by wounds in the vessels.¹ The local increase of shear flow triggers conformational changes that lead to the exposure of anchor domains to platelets and to the subendothelium, thus allowing the formation of a hemostatic plug.^{1,3–5} Flow-induced conformational changes also activate the so-called catch-bond proteins like FimH that anchors bacterial cells to the epithelium in urinary tract infections.^{1,6,7} In this case, the shear flow acts as an allosteric effector^{8,9} that induces a conformational shift toward protein states of high binding affinity with the membrane-linked substrate.

Beyond these examples of functional activation, it has been debated if and to which extent shear or elongational flows unfold—at least partially—proteins.^{10–14} Several experimental studies and numerical estimates concluded that only very high shear rates, $\dot{\gamma} > 1 \times 10^7 \text{ s}^{-1}$, can unfold midsize globular proteins ($\sim 1 \times 10^2$ amino acids (aa)).^{10,15,16} However, a few studies reported conformational perturbations even at low

shear rate ($\dot{\gamma} \simeq 1 \times 10^2 \text{ s}^{-1}$) for proteins such as lysozyme (~ 130 aa) and bovine serum albumin (~ 580 aa).¹¹ Overall, the alteration of a protein native state by external hydrodynamic forces seems to be highly sensitive to its distinct structural features, global shape, and internal cohesion.^{10,16–18} While physiologically attainable shear rates cannot unfold isolated small globular proteins, for large systems such as the multimeric vWf, the tensile force caused by physiological shear ($1 \times 10^5 \text{ s}^{-1}$)³ acts down at the nanometer length scale on the small functional domains, for example, causing the (partial) unfolding of the A2 domain.¹⁹ Moreover, and notably, the fluid perturbation can trigger protein aggregation, including amyloid- β aggregation in Alzheimer's disease,^{20–22} and stirring is used to accelerate and tune the aggregation path possibly by altering the protein's misfolded states.²³ Studying the detailed mechanisms of such processes is thus of theoretical and practical interest.

From the experimental point of view, the study of protein perturbation under shear has been regularly complemented by

Received: September 6, 2018

Revised: November 16, 2018

Published: November 16, 2018

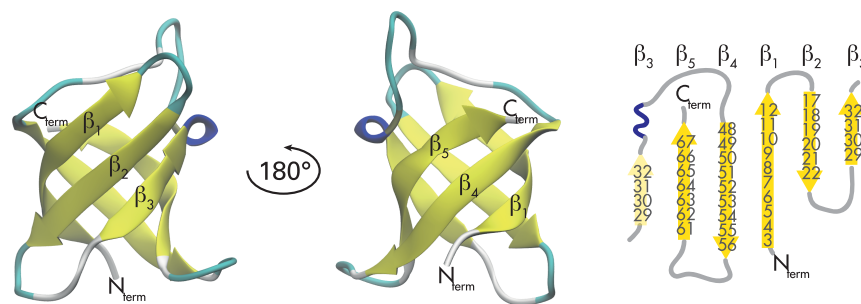


Figure 1. CspA from *E. coli*. (left) The protein is represented by two side views highlighting the secondary structure motifs. (right) The amino acid sequence along the protein structure.

single-molecule experiments, where an external mechanical force is used to perturb the protein fold, with an atomic force microscope (AFM) or optical/magnetic tweezers.^{2,24–26} However, the directional pulling force applied to the protein extremities is very distinct from the delocalized and fluctuating tensile forces generated by a shearing fluid.^{16,27–29} Interestingly, it was shown experimentally that when subjected to pulling forces along different directions, a model protein (SH3) exhibits two alternative unfolding paths.³⁰ It is therefore important to assess how the unfolding mechanism monitored via mechanical pulling compares to that under shear perturbation for small globular proteins as well as for large multimeric protein chains.

Computational approaches are very helpful to investigate the detailed mechanisms of protein unfolding and, therefore, to support the interpretation of single-molecule and ensemble-averaged experiments. For instance, thermal, chemical, and force unfolding processes have been compared using all-atom simulations.^{31–35} In this respect, all-atom simulations are rarely used to study shear effects,^{34,35} mainly because of technical challenges. Therefore, the impact of shear on protein stability has been mostly studied with implicit coarse-grained (CG) models of various complexity coupled to the fluid velocity gradient modeled with several different strategies.^{3,16,28,29,36,37}

In this work, following previous investigations on model peptides¹⁶ and proteins,^{38,39} we deploy the Lattice Boltzmann Molecular Dynamics technique⁴⁰ (LBMD) to investigate the shear unfolding of a cold-shock protein, and we compare it to force and thermal unfolding processes. LBMD allows to include in a natural way hydrodynamic interactions (HI) in implicit-solvent, particle-based molecular simulations, and most importantly, it allows generation of a shear flow instantaneously coupled to the protein conformational dynamics. This study is based on the implicit-solvent, CG model for protein Optimized Potential for Efficient peptide structure Prediction (OPEP).⁴¹ For thermal and force unfolding, the results are also compared to previous all-atom simulations in explicit solvent.³³

Our results show that the unfolding of a model system, the cold-shock protein (CspA), occurs on a time scale of several nanoseconds only at high shear rates, $\dot{\gamma} \approx 1 \times 10^9 \text{ s}^{-1}$, in agreement with our previous estimations for smaller peptides.¹⁶ Within the framework of our description, we find an equivalence between the unfolding time scales observed at these shear rate values and those occurring under a mechanical force in the range of 200–300 pN. However, shear and force unfolding proceed along very different reaction pathways. In

particular, shear unfolding involves protein rotation, so that the delocalized action of the perturbing tensile force is more akin to thermal unfolding, since it challenges different weak spots of the protein scaffold.

METHODS

Lattice Boltzmann Molecular Dynamics. In the last decades, the Lattice Boltzmann method has become very popular to study fluid simulation problems, since it allows to reproduce the Navier–Stokes dynamics with excellent computational efficiency and scalability.^{42,43} The method is based on following a collective group of fluid particles and by characterizing the fluid state by densities of probability f_p to find fluid particles with a given discrete velocity \vec{c}_p at a given place in a three-dimensional lattice. At each step, the probability densities are propagated and updated according to a collision scheme. The Bhatnagar–Gross–Krook (BGK) scheme⁴⁴ is the simplest, yet effective way to recover hydrodynamics behavior and consists of a relaxation to the local equilibrium distribution with a rate τ^{-1} related to the kinematic viscosity of the fluid ν . This scheme can be summarized by the following equation:

$$f_p(\vec{x} + \vec{c}_p \Delta t, t + \Delta t) = f_p(\vec{x}, t) + \tau^{-1}(f_p^{\text{eq}} - f_p) + g_p \quad (1)$$

where the first two terms on the right-hand side represent the streaming and collision contributions, and g_p is a term incorporating external forces acting on the fluid. In LBMD, we couple the molecular motion in continuous space to the fluid evolution in the discrete lattice space.^{45,46} This coupling is caused by a drag force

$$\vec{F}_i^D = -\Gamma(\vec{v}_i - \vec{u}(\vec{x}_i)) \quad (2)$$

where Γ is a friction coefficient, \vec{v}_i is the velocity of the i th particle, and \vec{u} is the fluid velocity extrapolated from the lattice according to the particle position \vec{x}_i . The force acts symmetrically on the fluid, via the term g_p , and on the particles according to the action–reaction principle. For the particle, the drag force adds with the conservative forces derived from the Hamiltonian of the system $F_i^C = -\nabla_i U(\{r\})$, and a random white noise, F_i^R , that represents thermal fluctuations. A full description of LBMD using an implicit solvent CG model for biomolecules is beyond the scope of this article, and the interested reader will find a comprehensive presentation in previous work.⁴⁰

In this study, we used LBMD simulations to study the unfolding of a globular protein under shear flow and to

compare it to force and temperature unfolding. A simulation time step $\tau_{\text{LBMD}} = 1.5$ fs and a friction coefficient $\Gamma = 0.1$ fs were used. Particle and fluid dynamics were evolved synchronously. The lattice space supporting the fluid dynamics was $\Delta x = 2.5$ Å, and the kinematic viscosity of the fluid was set to the value of pure water at ambient conditions (1×10^{-6} m² s⁻¹). A detailed discussion about the size of the grid spacing used to couple LB to a quasi-atomistic force field for biomolecules can be found elsewhere.^{40,47}

The Optimized Potential for Efficient Peptide Structure Prediction Force Field. In our simulations, we used the implicit-solvent force field OPEP to model the protein system. OPEP was developed to model soluble peptides and proteins with a high structural accuracy while remaining free of any biases constraining the secondary structure.^{48–50} It offers the opportunity to efficiently collect large statistics on dynamical and structural features of proteins while keeping a quasi-atomistic resolution for the system of interest. OPEP has been successfully applied to a wide range of problems, from peptide structure prediction to the study of molecular crowding and amyloid aggregation.⁴¹

System. For the purpose of our study, we selected the major cold-shock protein from *Escherichia coli* (CspA, PDB: 1MJC)⁵¹ as a model system. This globular protein consists of 69 residues that fold into a β -barrel made of five β -strands, hereafter denoted from β_1 to β_5 (Figure 1). It belongs to a large family of homologous proteins that exhibit multiple examples of adaptation to high temperatures among thermophilic bacteria. For that reason, mechanical properties of cold-shock proteins have been extensively studied both experimentally and in simulations.^{33,52–55} Moreover, the unfolding of CspA under force and high temperature was previously investigated by our group using all-atom simulations,³³ providing an interesting reference for the present work.

RESULTS AND DISCUSSION

Shear Unfolding. To observe and to quantify the effect of a constant shear flow on CspA, we first generated 20 initial configurations representative of the protein native state by performing a molecular dynamics (MD) simulation at ambient temperature (300 K) in the absence of HI and extracting the protein configuration every 300 ps. For each initial configuration, the protein was then placed in a simulation box of size $L_x = 200$ Å, $L_y = 100$ Å, and $L_z = 150$ Å with a random orientation. Before activating the shear flow, the system was equilibrated for a short time (5 ns) by including HI.

A shear flow was generated in the simulation box by imposing constant forces in opposite directions to the fluid on the top ($z = L_z$) and bottom ($z = 0$) faces. This simulation setup qualitatively reproduces a Couette cell, where the walls enclosing the fluid move with respect to each other so as to generate a laminar flow with a linear transversal velocity gradient.¹⁶ We checked that, in the absence of the protein, transverse velocities vary linearly along the z -axis, with a slope that directly yields the shear rate $\dot{\gamma}$ (in s⁻¹) and which is proportional to the forces set at the boundaries. In the presence of the protein, we monitored the velocity profile in the molecule's vicinity, which underlines the very natural coupling between the particle and fluid dynamics in LBMD simulations; see ref 16. A steady-state profile is observed after a

few picoseconds, which is short compared to the protein time scales.

In our simulations, several shear rate values in the range from 2×10^9 s⁻¹ to 1×10^{10} s⁻¹ were employed. It is important to mention that these values correspond to very high rates that are hardly attainable in physiological or experimental conditions. However, such high rates were necessary to explore the unfolding of CspA on the nanosecond time scale (our simulations were extended to full protein unfolding, which typically occurred in the range of 10–100 ns, depending on the shear rate).

In Figure 2, we show a typical unfolding event of the CspA protein under shear. As already described for soft object and

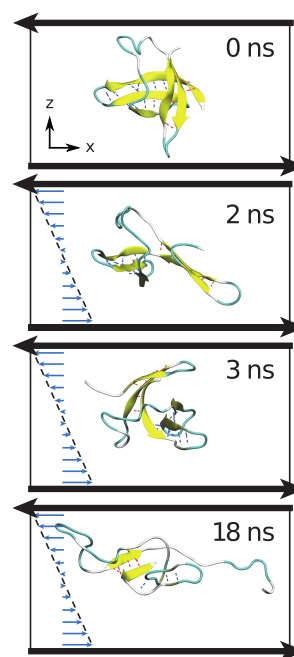


Figure 2. Protein unfolding under shear stress. Snapshots showing the protein characteristic tumbling motion, alternating extension (e.g., at 2 and 18 ns) and compaction (3 ns) phases. The progressive loss of secondary structure can be appreciated.

polymers in shear flow, the protein experiences two characteristic motions as a result of the decomposition of the velocity field.^{16,17,29} The protein is affected by a rotational motion, and the fold is altered by elongational forces acting along the shear eigenvector. These features can be understood by the fact that a simple shear flow velocity field can be expressed in terms of rotational and elongational components, which lead to a periodic tumbling movement,⁵⁶ where the molecule alternates between compact and extended states as the tensile forces align periodically with its axis of weaker mechanical resistance. During the early stage of shear perturbation, the periodic distortion is somehow reversible, and it can be seen as an elastic deformation of the global structure. This characteristic behavior is well-represented by the time evolution of gyration radius R_g that is marked by regular oscillations with periodicity proportional to the inverse of the shear rate, $\tau_R \propto 1/\dot{\gamma}$ (Figure 3).²⁹ The oscillatory behavior can be monitored also in the time evolution of the root-mean-square displacement (RMSD; Figure S1 in the Supporting Information).

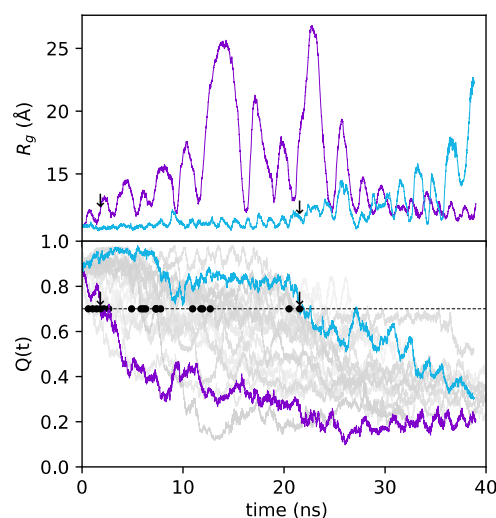


Figure 3. Unfolding under shear. (top) Time evolution of the protein radius of gyration R_g as a function of time for two unfolding trajectories; see also curves highlighted in the bottom panel. (bottom) Decay of the fraction of native contacts under constant shear. Traces of 20 replicas (light gray) from simulations using a shear rate of $4 \times 10^9 \text{ s}^{-1}$. The FPTs at the threshold value $Q = 0.7$ are indicated for each replica by black dots. Two trajectories (purple and blue lines) and their respective FPTs (black arrows) were highlighted for clarity purposes.

The protein softness and ability to recover quickly its original conformation may strongly depend on the nature of its fold. For example, in the case of a simple β -hairpin peptide, the unfolding is triggered by the tensile force acting on the hydrogen-bond (HB) ramp that stabilizes the parallel β -sheet.¹⁶ The intensity of the fluid drag force decreases periodically the barrier for HB separation, eventually leading to the sequential but fast unzipping of all the HBs in the ramp. For the small β -hairpin peptide, no stable intermediate has been observed.¹⁶ On the contrary, in CspA, the β -barrel can be partially opened at the contact between β_3 and β_5 , the rest of the protein being intact, which leads to partially unfolded intermediates.

The sequential separation of secondary structure motifs, and the presence of some intermediate partially unfolded states, has already been observed in the shear unfolding of the A2 domain of the vWf protein.³⁸ For this protein, the partial unfolding may have functional impact by helping the quick exposure of the cleavage site of the protein under blood shear perturbation.² The presence of intermediate states associated with different dissociations of secondary structures is presently a very interesting focus in single-molecule experiments that probe unfolding by applying external directional mechanical forces.⁵⁵

To quantify the progressive unfolding of a protein under shear, a metric that is mostly insensitive to the periodic tumbling is required. Some frequently used collective variables such as the RMSD or the radius of gyration, albeit suitable for thermal or chemical denaturation, are here dominated by the oscillating extension of the global shape (Figure 3 and Figure S1) and do not capture well the irreversible loss of the native structure. Therefore, the unfolding of CspA is better described by considering a native contact analysis based on the average pairwise distances between the backbone C_α extracted from a

short reference simulation with no shear. The fraction of native contacts is then computed for each configuration according to

$$Q(t) = \frac{1}{Q^{\text{ref}}} \sum_i^N \sum_{\substack{j \\ |i-j| \geq 4}}^N Q_{ij}(t) \quad (3)$$

where the matrix element Q_{ij} for residues i and j is equal to 0 if the reference distance between the two C_α atoms is $d_{ij}^{\text{ref}} > r_c$, $e^{-(d_{ij}(t)-d_{ij}^{\text{ref}})^2/2}$ if $d_{ij}^{\text{ref}} < r_c$ and $d_{ij}(t) > d_{ij}^{\text{ref}}$, or 1 if $d_{ij}^{\text{ref}} < r_c$ and $d_{ij}(t) - d_{ij}^{\text{ref}}$, with $r_c = 10 \text{ Å}$. The normalization factor Q^{ref} is the number of contacts between C_α atoms in the reference. The condition $|i-j| \geq 4$ guarantees that neighboring amino acids are not taken into account, and the exponential term allows a smoother evolution of $Q(t)$.⁵⁷

The decay of the native contact fraction for each individual trajectory reveals highly heterogeneous unfolding paths (Figure 3). A mean first passage time (FPT) analysis can be used to characterize the unfolding kinetics and the dependence on the shear perturbation. We selected an arbitrary threshold of $Q = 0.7$, which was found to be a good compromise, since lower values were attained too slowly at weak shear rates, and higher values were too close to the native structure. However, thresholds of $Q = 0.6$ and $Q = 0.8$ yielded qualitatively similar results.

The distributions of inverse FPTs (unfolding rates) and their averages are plotted against the shear rate in Figure 4.

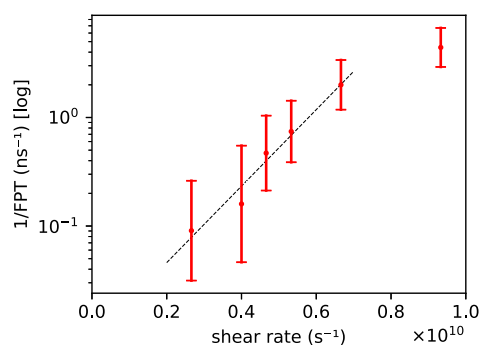


Figure 4. Kinetics of unfolding as a function of shear rate. The inverse first passage times for unfolding are extracted from the fraction of native contacts (criterion $Q = 0.7$; see Figure 3) and plotted against shear rate for sets of replicas. The average measure (red dots; error bars show standard deviation) follows a linear trend at lower rates (dashed line) and seems to reach a plateau above $7 \times 10^9 \text{ s}^{-1}$.

Overall, the average unfolding rate shows a typical logarithmic dependence with the shear rate $\dot{\gamma}$ that is reminiscent of the typical kinetics for protein unfolding under a pulling force.¹⁶ Only the highest value ($9.3 \times 10^9 \text{ s}^{-1}$) departed from the trend. We already observed for other model peptides/proteins¹⁶ that at high shear rates the kinetics reaches a plateau. An explanation can be proposed for the behavior in this high shear rate regime: the unfolding may no longer be determined by the necessity to overcome energetic barriers, but it is limited by the diffusion on the free energy landscape.¹⁶ In fact, a similar behavior is observed when unfolding is monitored by applying an external force, and the force value is so high that the limiting factor is diffusion rather than energetics.^{27,31,58}

The analogy with the force-induced unfolding is strong, but as already noted, the shear drag force acts differently on the protein depending on the tumbling dynamics. As discussed in a previous work,¹⁶ for the simple β -hairpin peptide this rotation exposes the reaction coordinate of unfolding (the axis of the HBs ramp) to the fluid tensile force with a periodicity that is inversely proportional to the shear rate $\dot{\gamma}$. In that case one can describe the mechanical perturbation as a fluctuating external force acting on the reaction coordinate x , for example, $f_x(x, \dot{\gamma}) = f_0(x, \dot{\gamma}) \cdot \cos(\dot{\gamma}t)$. For CspA a clear unfolding reaction coordinate is harder to define, and, as we will show in the following sections, a direct comparison with other unfolding processes, namely, under external force or high temperature, is useful.

Force Unfolding. In this section, we discuss the unfolding of the CspA protein in an *in silico* simulation setup mimicking typical force spectroscopy experiments. Similarly to the shear unfolding, we generated a set of ~ 20 initial configurations from a trajectory at ambient condition in the absence of external force. We then fixed the C_α carbon of the N-terminal residue, and we applied a constant pulling force of 400 pN to the C-terminal C_α . Each simulation was extended to observe partial or total unfolding of the initial configuration. Unfolding was followed as in the experiments by monitoring the protein end-to-end distance; see Figure 5. A molecular view of a typical

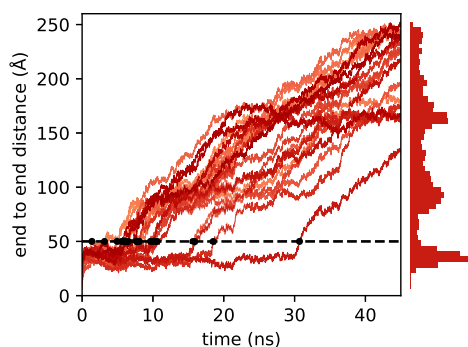


Figure 5. Multiple-step unfolding under force. We report the time evolution of the end-to-end distance using a pulling force of 400 pN, together with the unfolding time for each trajectory (black dots) corresponding to the first crossing of a 50 Å threshold (dashed line). On the lateral side of the graph, it is reported the cumulative histogram (first 45 ns) of the end-to-end distances during the unfolding events. The presence of multiple intermediate states is indicated by the presence of multiple peaks in the histogram.

unfolding trajectory under force is also shown in Figure S2. The unfolding proceeds via multiple extensions because of the presence of long-lived intermediates, as illustrated by the cumulative histogram on the right-hand side of Figure 5. We note that this “staircase” motif, broadly observed among computational studies at constant force,^{27,32,55,59} is hardly observed in experimental measurements, where unfolding is generally well-fitted by a two-state model,⁶⁰ because of the time resolution of the instrumentation (\sim ms). However, long-lasting intermediate states have recently received experimental support from finely controlled force-clamp and force-ramp experiments, suggesting that unfolding generally occurs along highly heterogeneous paths.⁵⁵

In the specific case of CspA, we demonstrated in a previous study³³ based on all-atom steered molecular dynamics using the CHARMM36 force field⁶¹ that the presence of a long-lived

initial intermediate at ~ 3 nm was due to a key structural motif, that is, the loop L_3 connecting the β_3 and β_4 strands. We had shown that, when the protein was folded, this loop hindered the unzipping of the β_4 and β_5 strands. It is only when the L_3 loop moves aside that this obstruction is removed, and that β_4 - β_5 unzipping and thus protein extension can occur. The same mechanism was indeed observed in our very different simulation setup based on the implicit solvent CG model OPEP, as shown in Figure S3, further reinforcing the validity of our modeling.

To examine the force dependence of the unfolding kinetics, we repeated the simulations by varying the force in the range of 300–450 pN (Figure 6). The first passage time was defined

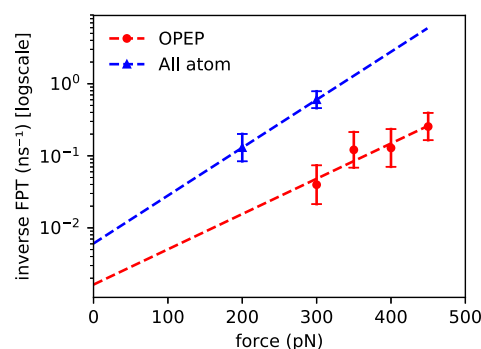


Figure 6. Kinetics of unfolding as a function of the pulling force. The inverse first passage times at an end-to-end distance of 50 Å are plotted against force for CG simulations (red) and atomistic simulation based on CHARMM36 (blue).³³ In both cases the mean FPTs follow a linear trend that matches Bell’s model.

for each trajectory as corresponding to the first crossing of a 50 Å threshold, which is located beyond an initial, long-lived intermediate that corresponds to a slightly distorted but still folded native state. The average unfolding rate shows an exponential dependence upon force, as expected from the Bell

model,⁶² $k = k_0 \exp\left(\frac{-\Delta G^\ddagger + F\Delta x}{k_B T}\right)$ where Δx is a phenomeno-

logical parameter measuring the distance between the reactant and transition states, assumed not to be affected by the force. For the sake of comparison we reported in the same plot the results from atomistic simulations³³ performed at 200 and 300 pN. From the model it is possible to extract the phenomenological parameter $\Delta x \simeq 0.3$ Å, for both CG and all-atom simulations, which is smaller than the typical value estimated for proteins from experiments, that are usually 1 order of magnitude higher.⁵³ The prefactor k_0 can be estimated using standard unidimensional transition-state theory ($\frac{k_B T}{h}$), and in this case the extrapolated rate constant at zero force leads to a reference free-energy barrier $\Delta G^\ddagger = 9$ kcal/mol in the absence of force. For all-atom simulations (blue triangles), the value of the free-energy barrier is close to 6.9 kcal/mol, in good agreement with the present CG simulations (red circles). An alternative approach is to use Kramers’ expression in the high friction limit for the extrapolated rate constant at zero force, in which the reaction rate is expressed after some simplifications (see Supporting Information) as

$k = \frac{D}{2\pi\sigma^2} \exp\left(\frac{-\Delta G^\ddagger}{k_B T}\right)$, where D is the effective diffusion coefficient along the end-to-end coordinate, accounting for

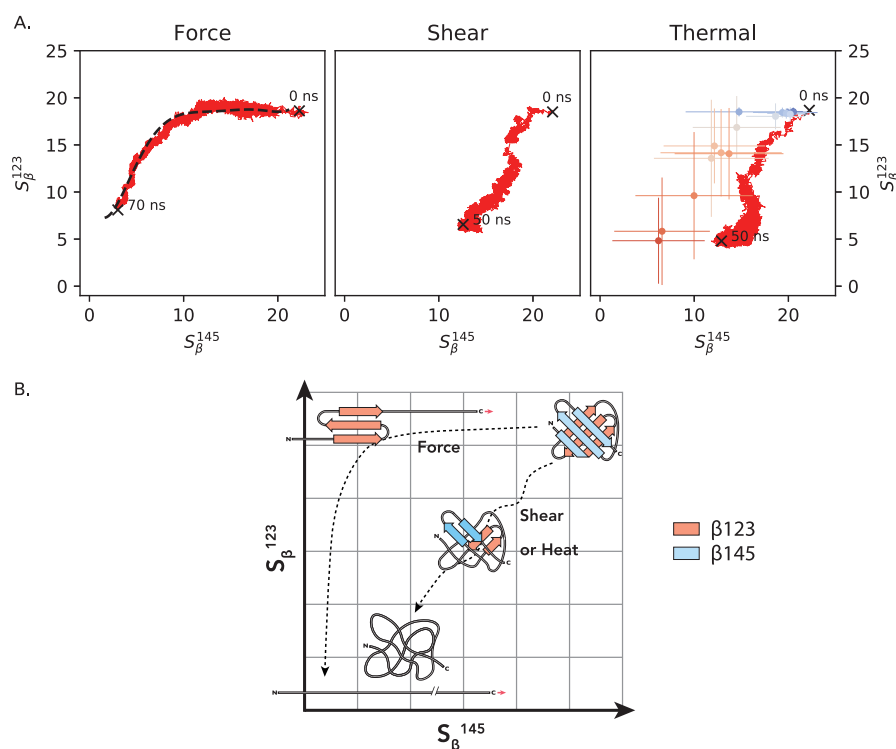


Figure 7. Comparison of the unfolding paths for three different perturbations. (A) The mean unfolding path trajectories are projected on the two-dimensional space defined by the S_{β}^{123} and S_{β}^{145} collective variables. For thermal unfolding, the data extracted from replica-exchange simulations represents the average value at each associated temperature. (B) Schematic representation of the unfolding mechanisms.

both the coupling with the solvent and internal contributions, and which can be numerically estimated using harmonic constraints⁶³ (see [Supporting Information](#)), while σ represents the standard deviation of the end-to-end distance in the reactant state. The resulting free-energy barriers are smaller, $\Delta G^{\ddagger} = 4.4$ and 3.5 kcal/mol, for the CG and all-atom simulations, respectively. It is worth recalling that the energy scale of the CG model OPEP can be modulated so as to tune the propensity to maintain the protein fold stable. The comparison with the all-atom simulations indicates that the actual setup is reasonable for inquiring the mechanical stability of the protein. Experiments on another analogous mesophilic cold-shock protein have reported free-energy barriers of 2.9 kcal/mol for thermal denaturation and 9.1 kcal/mol for the unfolding under force.⁵³ Note however that this last value was obtained assuming a given rate constant prefactor, which is largely setup-dependent,⁶⁴ and that was not specifically determined in these experiments, thereby suggesting that the experimental free-energy barrier might be overestimated.

Comparing Thermal, Force and Shear Unfolding. In this section we compare more extensively the unfolding pathways of CspA under shear, force, and thermal perturbations. To this purpose, a set of simulations were performed at $T = 430$ K to observe thermal unfolding on the nanosecond time scale. CspA consists of a β -barrel structure with two groups of interacting β -sheets that are more easily evidenced in [Figure 1](#), β_1 - β_2 - β_3 and β_4 - β_5 , with some extra HBs between β_1 and β_4 . We therefore decided to project the trajectories of the unfolding events along the two collective coordinates consisting in the number of HBs in the β_1 , β_2 , β_3 and β_1 , β_4 , β_5 sheets, denoted S_{β}^{123} and S_{β}^{145} , respectively:

$$S_{\beta}^X(t) = \sum_{i \in X} \sum_{\substack{j \in X \\ |i-j| \geq 4}} S_{ij}(t) \quad (4)$$

where the sum i and j run over the backbone HB donors (the amide N atoms) and HB acceptors (the carbonyl oxygen atoms) in the β -sheets $X = 123$ or 145. The sums consider only amino acids separated by at least four residues, $|i - j| \geq 4$. The term $S_{ij}(t)$ is a smoothed number of HBs, which only depends on the distance between the HB donor and acceptor. It is equal to 1 if this distance is smaller than the $d_c = 3$ Å, and otherwise it is given by $e^{-d_{ij}(t)^2/2}$.

Using these collective coordinates, we can now compare the three types of perturbation ([Figure 7](#)). In shear and thermal unfolding, the β_1 - β_2 - β_3 interactions break first, and β_1 - β_4 - β_5 is more conserved, at least on the explored time scale (50 ns). Under force, the unfolding proceeds along a different path, with the β_1 - β_4 - β_5 unzipping first and then a progressive unfolding of β_1 - β_2 - β_3 . Interestingly, because of protein tumbling under shear, the protein experiences a tensile perturbation that does not act along a fixed direction, as in pulling experiments, and the mechanical perturbation is much more distributed on the protein fold. The shear perturbation is therefore more equally distributed onto the protein structure, just as thermal excitation. A simplified scheme of the unfolding paths in the reduced free-energy landscape is given in [Figure 7B](#).

For force and thermal unfolding, we also reported the data from all-atom simulations.³³ For force unfolding we extracted the average unfolding paths considering 25 independent trajectories at $F = 200$ pN. For thermal unfolding, we

considered data obtained with replica-exchange simulations performed with the CHARMM36⁶¹ force field, each replica evolving for 500 ns at different effective temperatures in the range of 306–407 K.^{33,65} The agreement between the results from the all-atom model and from the OPEP model is very good, suggesting that the CG OPEP model is able to capture the essential aspects of mechanical and thermal response of proteins.

CONCLUSION

In this work we have compared the unfolding mechanism of a small globular cold-shock protein, CspA, under three external perturbations: shear flow, mechanical pulling, and thermal excitation. While thermal and force unfolding have been extensively investigated along the years and provide a conceptual reference, less is known about the behavior of proteins under shear.^{10–14} However, this aspect is key to understand the functionality of special types of proteins that work under physiological fluid flow.^{1,2}

In this study we have used the LBMD technique, which allows to self-consistently generate the fluid flow perturbing the protein dynamics,^{16,40} combined with the implicit-solvent CG OPEP model for the protein.⁴¹ The validity of our results for the force and thermal unfolding was assessed by comparing the current approach with data obtained from more accurate, but computationally more expensive, all-atom simulations.³³

As already observed for other systems, the unfolding of an isolated small globular protein can occur under Couette flow only at high shear rates, $\dot{\gamma} \approx 1 \times 10^9 \text{ s}^{-1}$, much higher than those in physiological conditions.^{10,16} At these high shear rates, CspA unfolds on the nanosecond time scale, and the unfolding rate is proportional to the shear rate $\dot{\gamma}$. Moreover, we observed that unfolding under a directional mechanical force of $\sim 300 \text{ pN}$ occurred on similar time scales. Extrapolation of the force–extension kinetics at zero force allowed to extract the free-energy barrier for unfolding, which was found to be in very good agreement with that obtained from atomistic simulations.

It is worth noting, however, that for very long multimeric proteins such as vWf, the extension of the chain under shear occurs at physiological shear rates up to $\dot{\gamma} \approx 1 \times 10^5 \text{ s}^{-1}$. The resulting tensile force acts at the nanometer scale on the individual domains and is expected to cause partial unfolding of the weaker ones—in the case of vWf, the functional A2 domain. It was shown¹⁹ that, with a simple mechanical model and assuming an unfolding force of $\sim 10 \text{ pN}$, the shear rate required to cause the chain extension and domain unfolding was estimated to be around $10^5\text{--}10^6 \text{ s}^{-1}$, close to the physiological conditions. Experimentally the force that would cause the partial unfolding of the A2 domain is estimated to be 20 pN . Therefore, the magnitude of the unfolding force for A2 compares to the low force regime in single-molecule experiments, where the characteristic unfolding time is generally within seconds.

We have finally individuated two collective variables that allow to compare the unfolding mechanism under the three perturbations. We have clearly shown that the unfolding pathways under shear have strong similarities with those observed during thermal unfolding. This is to contrast with the mechanical pulling perturbation, where the external force acts anisotropically and oriented along the end-to-end distance of the protein. Vice versa, the tensile force generated by the shear acts isotropically on average because of the protein tumbling.

Therefore, force and shear perturbations probe very different weaknesses of the protein fold, and they lead to distinct unfolding paths. In the future, we will assess how, when moving to a longer multimeric protein as vWf, the shearing acts down at the level of the individual domains, and how the elongation of the protein chain affects the isotropic action of the tensile force.

In conclusion we have demonstrated that the combination of a very effective CG model with a realistic treatment of hydrodynamic interactions via Lattice Boltzmann allows to successfully explore the response of a protein under perturbing fluid conditions and to compare its behavior under thermal and mechanical stresses. This methodology is ready to be deployed to extensively study at the molecular level some peculiar proteins such as catch-bonds or vWf, to characterize how the shear flow imposes local reorganization of the functional domains, and try to extrapolate this behavior at low-shear physiological conditions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b08711.

An additional plot of the time evolution of the RMSD under shear; snapshots of force unfolding trajectories illustrating the unfolding mechanism and the key role of the L₃ loop; a detailed description of the method employed to estimate the free-energy barriers along with related graphs and data (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: stirnemann@ibpc.fr. (G.S.)

*E-mail: fabio.sterpone@ibpc.fr. (F.S.)

ORCID

Philippe Derreumaux: 0000-0001-9110-5585

Guillaume Stirnemann: 0000-0002-5631-5699

Fabio Sterpone: 0000-0003-0894-8069

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the ERC (FP7/2007-2013) Grant No. 258748. Part of this work was performed using HPC resources from GENCI [CINES and TGCC] (Grant No. x20176818). We acknowledge the financial support by the “Initiative d’Excellence” program from the French State (Grant No. “DYNAMO”, ANR-11-LABX-0011-01).

REFERENCES

- (1) Thomas, W. E.; Vogel, V.; Sokurenko, E. Biophysics of Catch Bonds. *Annu. Rev. Biophys.* **2008**, *37*, 399–416.
- (2) Springer, T. A. von Willebrand Factor, Jedi Knight of the Bloodstream. *Blood* **2014**, *124*, 1412–1425.
- (3) Schneider, S. W.; Nuschele, S.; Wixforth, A.; Gorzelanny, C.; Alexander-Katz, A.; Netz, R. R.; Schneider, M. F. Shear-Induced Unfolding Triggers Adhesion of von Willebrand Factor Fibers. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 7899–7903.
- (4) Zhang, Q.; Zhou, Y.-F.; Zhang, C.-Z.; Zhang, X.; Lu, C.; Springer, T. A. Structural Specializations of A2, a Force-Sensing

Domain in the Ultralarge Vascular Protein von Willebrand Factor. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9226–9231.

(5) Fu, H.; Jiang, Y.; Yang, D.; Scheiflinger, F.; Wong, W. P.; Springer, T. A. Flow-Induced Elongation of von Willebrand Factor Precedes Tension-Dependent Activation. *Nat. Commun.* **2017**, *8*, 324.

(6) Cusumano, C. K.; Pinkner, J. S.; Han, Z.; Greene, S. E.; Ford, B. A.; Crowley, J. R.; Henderson, J. P.; Janetka, J. W.; Hultgren, S. J. Treatment and Prevention of Urinary Tract Infection with Orally Active FimH Inhibitors. *Sci. Transl. Med.* **2011**, *3*, 109ra115.

(7) Hartmann, M.; Lindhorst, T. K. The Bacterial Lectin FimH, a Target for Drug Discovery Carbohydrate Inhibitors of Type 1 Fimbriae-Mediated Bacterial Adhesion. *Eur. J. Org. Chem.* **2011**, *2011*, 3583–3609.

(8) Sauer, M. M.; Jakob, R. P.; Eras, J.; Baday, S.; Eris, D.; Navarra, G.; Berneche, S.; Ernst, B.; Maier, T.; Glockshuber, R. Catch-bond Mechanism of the Bacterial Adhesin FimH. *Nat. Commun.* **2016**, *7*, 10738.

(9) Kalas, V.; Pinkner, J. S.; Hannan, T. J.; Hibbing, M. E.; Dodson, K. W.; Holehouse, A. S.; Zhang, H.; Tolia, N. H.; Gross, M. L.; Pappu, R. V.; et al. Evolutionary Fine-Tuning of Conformational Ensembles in FimH During Host-Pathogen Interactions. *Sci. Adv.* **2017**, *3*, e1601944.

(10) Jaspe, J.; Hagen, S. J. Do Protein Molecules Unfold in a Simple Shear Flow? *Biophys. J.* **2006**, *91*, 3415–3424.

(11) Bekard, I. B.; Asimakis, P.; Teoh, C. L.; Ryan, T.; Howlett, G. J.; Bertolini, J.; Dunstan, D. E. Bovine Serum Albumin Unfolds in Couette Flow. *Soft Matter* **2012**, *8*, 385–389.

(12) Bekard, I. B.; Dunstan, D. E. Shear-Induced Deformation of Bovine Insulin in Couette Flow. *J. Phys. Chem. B* **2009**, *113*, 8453–8457.

(13) Bekard, I. B.; Barnham, K. J.; White, L. R.; Dunstan, D. E. α -Helix Unfolding in Simple Shear Flow. *Soft Matter* **2011**, *7*, 203–210.

(14) Ashton, L.; Disting, J.; Imomoh, E.; Balabani, S.; Blanch, E. W. Shear-Induced Unfolding of Lysozyme Monitored In Situ. *Biophys. J.* **2009**, *96*, 4231–4236.

(15) Brückl, L.; Schröder, T.; Scheler, S.; Hahn, R.; Sonderegger, C. The Effect of Shear on the Structural Conformation of rhGH and IgG1 in Free Solution. *J. Pharm. Sci.* **2016**, *105*, 1810–1818.

(16) Sterpone, F.; Derreumaux, P.; Melchionna, S. Molecular Mechanism of Protein Unfolding under Shear: A Lattice Boltzmann Molecular Dynamics Study. *J. Phys. Chem. B* **2018**, *122*, 1573–1579.

(17) Alexander-Katz, A.; Schneider, M. F.; Schneider, S. W.; Wixforth, A.; Netz, R. Shear-Flow-Induced Unfolding of Polymeric Globules. *Phys. Rev. Lett.* **2006**, *97*, 138101.

(18) Alexander-Katz, A.; Netz, R. R. Dynamics and Instabilities of Collapsed Polymers in Shear Flow. *Macromolecules* **2008**, *41*, 3363–3374.

(19) Morabito, M.; Dong, C.; Wei, W.; Cheng, X.; Zhang, X.; Oztekin, A.; W, E. Internal Tensile Force and A2 Domain Unfolding of von Willebrand Factor Multimers in Shear Flow. *Biophys. J.* **2018**, *115* (10), 1860–1871.

(20) Hamilton-Brown, P.; Bekard, I.; Ducker, W. A.; Dunstan, D. E. How Does Shear Affect $\alpha\beta$ Fibrillogenesis? *J. Phys. Chem. B* **2008**, *112*, 16249–16252.

(21) Trumbore, C. N. Shear-Induced Amyloid Formation in the Brain: I. Potential Vascular and Parenchymal Processes. *J. Alzheimer's Dis.* **2016**, *54*, 457–470.

(22) Byington, M. C.; Safari, M. S.; Conrad, J. C.; Vekilov, P. G. Protein Conformational Flexibility Enables the Formation of Dense Liquid Clusters: Tests Using Solution Shear. *J. Phys. Chem. Lett.* **2016**, *7*, 2339–2345.

(23) Kinoshita, M.; Lin, Y.; Dai, I.; Okumura, M.; Markova, N.; Ladbury, J. E.; Sterpone, F.; Lee, Y.-H. Energy Landscape of Polymorphic Amyloid Generation of β 2-Microglobulin Revealed by Calorimetry. *Chem. Commun.* **2018**, *54*, 7995–7998.

(24) Neuman, K. C.; Nagy, A. Single-Molecule Force Spectroscopy: Optical Tweezers, Magnetic Tweezers and Atomic Force Microscopy. *Nat. Methods* **2008**, *5*, 491–505.

(25) Jagannathan, B.; Marqusee, S. Protein Folding and Unfolding under Force. *Biopolymers* **2013**, *99*, 860–869.

(26) Javadi, Y.; Fernandez, J. M.; Perez-Jimenez, R. Protein Folding Under Mechanical Forces: A Physiological View. *Physiology* **2013**, *28*, 9–17.

(27) Szymczak, P.; Cieplak, M. Stretching of Proteins in a Force-Clamp. *J. Phys.: Condens. Matter* **2006**, *18*, L21.

(28) Szymczak, P.; Cieplak, M. Stretching of Proteins in a Uniform Flow. *J. Chem. Phys.* **2006**, *125*, 164903.

(29) Szymczak, P.; Cieplak, M. Protein in Shear Flow. *J. Chem. Phys.* **2007**, *127*, 155106.

(30) Jagannathan, B.; Elms, P.; Bustamante, C.; Marqusee, S. Direct Observation of a Force-Induced Switch in the Anisotropic Mechanical Unfolding Pathway of a Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17820–17825.

(31) Sun, L.; Noel, J. K.; Sulkowska, J. I.; Levine, H.; Onuchic, J. N. Connecting Thermal and Mechanical Protein (Un)folding Landscapes. *Biophys. J.* **2014**, *107*, 2950–2961.

(32) Stirnemann, G.; Kang, S.-g.; Zhou, R.; Berne, B. J. How Force Unfolding Differs from Chemical Denaturation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 3413–3418.

(33) Stirnemann, G.; Sterpone, F. Mechanics of Protein Adaptation to High Temperatures. *J. Phys. Chem. Lett.* **2017**, *8*, 5884–5890.

(34) Chen, Z.; Lou, J.; Zhu, C.; Schulten, K. Flow-Induced Structural Transition in the β -Switch Region of Glycoprotein Ib. *Biophys. J.* **2008**, *95*, 1303–1313.

(35) Lou, J.; Zhu, C. Flow Induces Loop-to-b-Hairpin Transition on the b-Switch of Platelet Glycoprotein Iba. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13847–13852.

(36) Sing, C. E.; Alexander-Katz, A. Elongational Flow Induces the Unfolding of von Willebrand Factor at Physiological Flow Rates. *Biophys. J.* **2010**, *98*, 35–37.

(37) Lemak, A. S.; Chen, J. Z. Y.; et al. Unfolding Proteins in an External Field: Can We Always Observe the Intermediate States? *Phys. Rev. E* **2003**, *67*, 031910.

(38) Chiricotto, M.; Sterpone, F.; Derreumaux, P.; Melchionna, S. Multiscale Simulation of Molecular Processes in Cellular Environments. *Philos. Trans. R. Soc., A* **2016**, *374*, 20160225.

(39) Sterpone, F.; Doutreligne, S.; Tran, T. T.; Melchionna, S.; Baaden, M.; Nguyen, P.; Derreumaux, P. Multi-Scale Simulations of Biological Systems Using the OPEP Coarse-Grained Model. *Biochem. Biophys. Res. Commun.* **2018**, *498*, 296–304.

(40) Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843–1853.

(41) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragolini, T.; Chebaro, Y.; St-Pierre, J.-F.; Kalimeri, M.; Barducci, A.; et al. The OPEP Protein Model: from Single Molecules, Amyloid Formation, Crowding and Hydrodynamics to DNA/RNA Systems. *Chem. Soc. Rev.* **2014**, *43*, 4871–4893.

(42) Succi, S. *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*; Clarendon Press, 2001.

(43) Benzi, R.; Succi, S.; Vergassola, M. The Lattice Boltzmann Equation: Theory and Applications. *Phys. Rep.* **1992**, *222*, 145–197.

(44) Bhatnagar, P.; Gross, E.; Krook, M. A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems. *Phys. Rev.* **1954**, *94*, 511–525.

(45) Ahlrichs, P.; Dünweg, B. Simulation of a Single Polymer Chain in Solution by Combining Lattice Boltzmann and Molecular Dynamics. *J. Chem. Phys.* **1999**, *111*, 8225–8239.

(46) Ahlrichs, P.; Dünweg, B. Lattice-Boltzmann Simulation of Polymer-Solvent Systems. *Int. J. Mod. Phys. C* **1998**, *9*, 1429–1438.

(47) Chiricotto, M.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Hydrodynamic Effects on β -Amyloid (16–22) Peptide Aggregation. *J. Chem. Phys.* **2016**, *145*, 035102.

(48) Derreumaux, P.; Mousseau, N. Coarse-grained Protein Molecular Dynamics Simulations. *J. Chem. Phys.* **2007**, *126*, 025101.


- (49) Chebaro, Y.; Pasquali, S.; Derreumaux, P. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. *J. Phys. Chem. B* **2012**, *116*, 8741–8752.
- (50) Sterpone, F.; Nguyen, P.; Kalimeri, M.; Derreumaux, P. Importance of the Ion-Pair Interactions in the OPEP Coarse-Grained Force Field: Parametrization and Validation. *J. Chem. Theory Comput.* **2013**, *9*, 4574–4584.
- (51) Schindelin, H.; Jiang, W.; Inouye, M.; Heinemann, U. Crystal Structure of CspA, the Major Cold Shock Protein of Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5119–5123.
- (52) Tych, K. M.; Hoffmann, T.; Brockwell, D. J.; Dougan, L. Single Molecule Force Spectroscopy Reveals the Temperature-Dependent Robustness and Malleability of a Hyperthermophilic Protein. *Soft Matter* **2013**, *9*, 9016.
- (53) Tych, K. M.; Batchelor, M.; Hoffmann, T.; Wilson, M. C.; Paci, E.; Brockwell, D. J.; Dougan, L. Tuning Protein Mechanics through an Ionic Cluster Graft from an Extremophilic Protein. *Soft Matter* **2016**, *12*, 2688–2699.
- (54) Tych, K. M.; Batchelor, M.; Hoffmann, T.; Wilson, M. C.; Hughes, M. L.; Paci, E.; Brockwell, D. J.; Dougan, L. Differential Effects of Hydrophobic Core Packing Residues for Thermodynamic and Mechanical Stability of a Hyperthermophilic Protein. *Langmuir* **2016**, *32*, 7392–7402.
- (55) Schönfelder, J.; Perez-Jimenez, R.; Muñoz, V. A Simple Two-State Protein Unfolds Mechanically via Multiple Heterogeneous Pathways at Single-Molecule Resolution. *Nat. Commun.* **2016**, *7*, 11777.
- (56) Guyon, E.; Hulin, J.-P.; Petit, L.; Mitescu, C. D. *Physical Hydrodynamics*; Oxford University Press, 2001.
- (57) Best, R. B.; Hummer, G.; Eaton, W. A. Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874–17879.
- (58) West, D. K.; Brockwell, D. J.; Olmsted, P. D.; Radford, S. E.; Paci, E. Mechanical Resistance of Proteins Explained Using Simple Molecular Models. *Biophys. J.* **2006**, *90*, 287–297.
- (59) Beedle, A. E. M.; Lezamiz, A.; Stirnemann, G.; Garcia-Manyes, S. The Mechanochemistry of Copper Reports on the Directionality of Unfolding in Model Cupredoxin Proteins. *Nat. Commun.* **2015**, *6*, 7894.
- (60) Garcia-Manyes, S.; Badilla, C. L.; Fernández, J. M.; et al. Force-Clamp Spectroscopy of Single-Protein Monomers Reveals the Individual Unfolding and Folding Pathways of I27 and Ubiquitin. *Biophys. J.* **2007**, *93*, 2436–46.
- (61) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Sidechain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (62) Evans, E.; Ritchie, K. Dynamic Strength of Molecular Adhesion Bonds. *Biophys. J.* **1997**, *72*, 1541–1555.
- (63) Stirnemann, G.; Giganti, D.; Fernandez, J. M.; Berne, B. J. Elasticity, Structure, and Relaxation of Extended Proteins under Force. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 3847–3852.
- (64) Berkovich, R.; Hermans, R. I.; Popa, I.; Stirnemann, G.; Garcia-Manyes, S.; Berne, B. J.; Fernandez, J. M. Rate Limit of Protein Elastic Response Is Tether Dependent. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 14416–14421.
- (65) Stirnemann, G.; Sterpone, F. Recovering Protein Thermal Stability Using All-Atom Hamiltonian Replica-Exchange Simulations in Explicit Solvent. *J. Chem. Theory Comput.* **2015**, *11*, 5573–5577.

C.2 EXTENSION DU FACTEUR DE VON WILLEBRAND

Le facteur de von Willebrand (vWf) est une protéine multimérique géante du plasma sanguin impliquée dans l'hémostase, c'est-à-dire la régulation de la coagulation. Dans certaines conditions hydrodynamiques, le domaine A2 du vWf se déplie et expose un site de reconnaissance pour la protéase ADAMTS₁₃. Le clivage du domaine A2 permet la régulation de la longueur des multimères, paramètre crucial au bon fonctionnement hémostatique. Dans ce deuxième article, nous appliquons l'approche LBMD (*lattice Boltzmann molecular dynamics*) au fragment A₁A₂A₃ du vWf afin d'étudier les détails moléculaires de son dépliement en contexte d'écoulement et l'exposition du domaine de reconnaissance. Nous comparons les résultats à l'action d'une force directionnelle typique des expériences de spectroscopie de force et en analysons les divergences. Ces simulations constituent une étape vers la construction de modèles multi-échelles du vWf afin d'étudier son fonctionnement en conditions physiologiques.

Article

Exposure of Von Willebrand Factor Cleavage Site in A1A2A3-Fragment under Extreme Hydrodynamic Shear

Olivier Languin-Cattoën ^{1,†}, Emeline Laborie ^{1,†}, Daria O. Yurkova ², Simone Melchionna ^{3,‡},
Philippe Derreumaux ^{1,†}, Aleksey V. Belyaev ^{2,*}  and Fabio Sterpone ^{1,*,†}

¹ Laboratoire de Biochimie Théorique, CNRS, Université de Paris, UPR 9080, 13 rue Pierre et Marie Curie, F-75005 Paris, France; languin@ibpc.fr (O.L.-C.); laborie@ibpc.fr (E.L.); philippe.derreumaux@ibpc.fr (P.D.)

² Faculty of Physics, Lomonosov Moscow State University, 119991 Moscow, Russia; iurkova.do16@physics.msu.ru

³ Dipartimento di Fisica, Università Sapienza, P.le A. Moro 5, 00185 Rome, Italy; simone.melchionna@gmail.com

* Correspondence: aleksey_belyaev@yahoo.com (A.V.B.); fabio.sterpone@ibpc.fr (F.S.)

† Current address: Institut de Biologie Physico-Chimique-Fondation Edmond de Rothschild, PSL Research University, F-75005 Paris, France.

‡ Current address: Lexma Technology, 1337 Massachusetts Avenue, Arlington, MA 02476, USA.

Abstract: Von Willebrand Factor (vWf) is a giant multimeric extracellular blood plasma involved in hemostasis. In this work we present multi-scale simulations of its three-domains fragment A1A2A3. These three domains are essential for the functional regulation of vWf. Namely the A2 domain hosts the site where the protease ADAMTS13 cleavages the multimeric vWf allowing for its length control that prevents thrombotic conditions. The exposure of the cleavage site follows the elongation/unfolding of the domain that is caused by an increased shear stress in blood. By deploying Lattice Boltzmann molecular dynamics simulations based on the OPEP coarse-grained model for proteins, we investigated at molecular level the unfolding of the A2 domain under the action of a perturbing shear flow. We described the structural steps of this unfolding that mainly concerns the β -strand structures of the domain, and we compared the process occurring under shear with that produced by the action of a directional pulling force, a typical condition of single molecule experiments. We observe, that under the action of shear flow, the competition among the elongational and rotational components of the fluid field leads to a complex behaviour of the domain, where elongated structures can be followed by partially collapsed melted globule structures with a very different degree of exposure of the cleavage site. Our simulations pose the base for the development of a multi-scale in-silico description of vWf dynamics and functionality in physiological conditions, including high resolution details for molecular relevant events, e.g., the binding to platelets and collagen during coagulation or thrombosis.

Keywords: von Willebrand factor; molecular dynamics; coarse-grains; lattice Boltzmann; shear flow; protein unfolding



Citation: Languin-Cattoën, O.; Laborie, E.; Yurkova, D.O.; Melchionna, S.; Derreumaux, P.; Belyaev, A.V.; Sterpone, F. Exposure of Von Willebrand Factor Cleavage Site in A1A2A3-Fragment under Extreme Hydrodynamic Shear. *Polymers* **2021**, *13*, 3912. <https://doi.org/10.3390/polym13223912>

Academic Editor: Brian J. Edwards

Received: 19 October 2021

Accepted: 6 November 2021

Published: 12 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Von Willebrand Factor (vWf) is a giant extracellular blood plasma protein that plays a key role in arterial hemostasis and thrombosis. Normally, in vivo vWf is a linear multimer consisting of 40–200 covalently linked monomers [1]. With a length of 60 to 80 nm, a monomer comprises 2050 amino acids distributed between multiple domains, each a few nanometers in size. One of the essential features of vWf is that these proteins can expand in the bloodstream [2–5] and capture the blood platelets in hydrodynamically adverse conditions, i.e., under extreme forces and mechanical stresses [6]. Vascular trauma, wound, plaque rupture, or inflammation are rapidly followed by the vWf-mediated adhesion and aggregation of blood platelets at the surface of damaged endothelium or exposed collagen [7–10]. After a mechanical stimulus (e.g., elevated shear, elongational stress, attachment

to surface) the vWf multimers change shape from compact to extended, providing adhesive sites for platelets [4,11,12]. The stretching forces can lead to a number of conformational changes within the domains of this protein [13,14]. It is known that the deficiency of vWf concentration in blood, as well as the short length of these macromolecules, causes bleeding disorder [15]. At the same time, ultra-long vWf concatamers may cause thrombotic conditions [16].

The three neighbouring globular domains A1, A2 and A3 attract the most attention as they provide important properties related to platelet adhesion and aggregation. A3 domain is responsible for binding to collagen and thus initiation of thrombosis. A1 domain can establish adhesive bonds with GPIb platelet membrane receptors and is crucial for platelet plug formation under high shear stress [10,17]. Mechanical stability of A2 domain is essential for size control of vWf multimers.

During primary hemostasis in arteries and arterioles, platelet adhesion depends on the length distribution of the vWf multimers [18–20]. The globular A2 domain (unlike A1 and A3) does not contain disulfide bonds between its N- and C-terminal ends [21]. This fact makes vWf-A2 remarkably susceptible to mechanical tension [14]. Control of the multimer sizes in vivo occurs owing to this structural feature: after a forced unfolding of A2 domain and exposure of the scissile bond located between the residues Tyr1605-Met1606, the protein can be cleaved by the metalloprotease ADAMTS13 [22,23]. Such force-induced proteolysis, as observed for the ADAMTS13-vWf system, represents a model for probing the molecular mechanisms underlying the translation of a mechanical stimulus into a chemical response in a biological system [24]. The basic principle of this mechanoenzymatic phenomenon relies on hydrodynamic forces and torques exerted on vWf molecules in a sheared viscous fluid. The evolutionary developed regulation of vWf activity is extremely important for thrombosis and normal hemostasis [16,25]. However, its response depends on protein sequence, globular structure and stability of both the enzyme (ADAMTS13) and the substrate (vWf-A2). The outcomes of an inherited or acquired failure of the vWf activity can lead to severe or even life-threatening consequences [26].

Mechanical stability of A2 domain is thus one of the major points of interest with respect to vWf thrombogenic activity and regulation. Characterization of the mechanical properties of vWf at the molecular scale is important for understanding its hemostatic functions. Prior studies underline the resistive role of the central β -sheet and hydrogen bond networks in case of steered Molecular Dynamics (MD) simulations within the constant loading rate regime [27]. Structural analysis also suggests that calcium ions stabilize the native conformation of vWf-A2 domain [28]. However, experiments with optical tweezers showed that the calcium rather accelerates the refolding, without affecting mechanical unfolding of the globule [29]. Several recent works are devoted to the stability and interactions of C-domains, C-terminal dimeric bouquets, D'D3-, and D4-domains under high and even pathological mechanical load [30–34]. Many theoretical studies (steered MD simulations [27,35,36]) and experimental setups (Atomic Force Microscopy [34,37,38], Optical and Magnetic Tweezers [14,33]) were designed to understand the biomechanical basis for vWf properties by reproducing the unidirectional pulling. In other cases, the effect of shearing was considered by means of extremely coarse-grained simulations [12,39,40], as well as experimentally [3,4,11].

Further insights emphasize the importance of the structure of globular domains and flexible linker segments for understanding biogenesis, shear-induced conformational changes, platelet recruitment abilities, and mechano-chemical regulation of von Willebrand factor [30,41,42]. There is a growing evidence that relative spatial arrangement of neighbouring domains of vWf may lead to substantial functional changes of this protein. Recent studies revealed that interactions between A1 and A2 domains are critical for the attenuation of vWf adhesivity to platelets [35]. On the other hand, in case of adhesion to collagen, the three-domain constructs A1A2A3 may compensate for clinically relevant point mutations in the A3 domains [37]. Several experimental works used the A1A2A3 fragment as an elementary molecular model for vWf functioning, including its proteoly-

sis by ADAMTS13 [38,43]. Therefore, not just the unfolding dynamics of individual A2 domains, but the collective behaviour of A1A2A3 constructs is likely to be involved into the regulation of the hemostatic activity of vWf. The detailed understanding of the exact A2-unfolding pathway and scissile bond exposure under different regimes of mechanical loading (pulling or shearing) is still missing in cases of A1A2A3 constructs and whole vWf multimers.

In the present work we use computer simulations to study molecular details of the A2 unfolding and the exposure of the scissile bond Tyr1605-Met1606 (or, using the indexing that begins from the N-terminus of A2 domain, ¹¹¹Tyr-¹¹²Met). We use the Lattice Boltzmann Molecular Dynamics technique [44] based on the coarse-grained (CG) model OPEP [45]. We first validate the capability of the OPEP force field to reproduce the mechanical unfolding of the A2 domain by comparing CG and all-atom simulations [46]. Then, the shear induced unfolding of the A2 domain in presence of its natural molecular surrounding is studied by modelling of the whole A1A2A3 fragment. We focus on the hydrodynamic consequences of the attachment of A1 and A3 domains to the A2 domain.

The manuscript is organized as follows. We describe the simulation methods and analysis techniques in the Materials and Methods section. The Results section presents and analyses the results concerning the force and shear induced unfolding of the individual A2 domain, and the shear induced unfolding of the A2 domain in the A1A2A3 construct. A final discussion is presented in the Conclusions.

2. Materials and Methods

In this section we describe the model and the methodology used through this work. We first present the flexible coarse-grained model OPEP that we exploited to investigate the unfolding of the A2 domain under tensile force and shear, and the simplified representations introduced to model the elementary three-domains construct A1A2A3 simulated in shear flow. A pictorial representation of the A1A2A3 system is given in Figure 1. We then describe the Lattice Boltzmann Molecular Dynamics (LBMD) technique that we employed to explore the shear-induced unfolding process. Finally, we detail the theoretical models used to extract the unfolding kinetics at different force and shear regimes.

2.1. Protein Structures

In order to construct our systems we used as starting point the crystallographic structures of the three domains A1 (PDB code 1AUQ [47], 207 residues), A2 (PDB code 3GXB [22], 177 residues), and A3 (PDB code 1AO3 [48], 187 residues), see Figure 1a. In order to model the initial configuration of the flexible linkers connecting A1 and A2, and A2 and A3, we generated a linear configuration of the amino-acid chain then relaxed using the OPEP force field as described in the following, see Figure 1b.

2.2. The OPEP Model

The OPEP force field is a multi-resolution coarse-grained model developed to investigate peptide and protein folding without ad-hoc biases, and aggregation [45,49,50]. Namely, small proteins were successfully folded using enhanced sampling simulations based on the OPEP force field [45,50], the experimental different thermal-stability of homologous proteins was successfully reproduced [51], and the mechanical unfolding of small proteins under pulling forces compares very well with the process simulated using an all-atom force field [46]. OPEP is based on the atomistic resolution of the backbone while each amino acid side chain is represented by a single bead (see Figure 1b). Glycine and proline are exceptions and are fully modeled at atomic resolution. The Hamiltonian is composed of two sets of potential energy terms: the bonded terms, that ensure molecular topology as well as preferential dihedral orientation in the backbone, and the non-bonded terms, that implicitly include screened electrostatic interactions, so that all non-bonded interactions are short-range in nature. Specific hydrogen bond cooperative terms are included to favor the formation of secondary structures. A detailed description of the Hamiltonian has

been presented in previous works, see ref. [45]. In the simulations presented hereafter we used the force field version v.4 [45]. The use of a fully flexible force field such as OPEP is mandatory for describing folding/unfolding or large conformational changes but is not necessary when the main conformation of the protein is preserved in time. In this latter case in order to increase the computational efficiency, it is preferable to exploit an elastic network model based on the OPEP force field for the intermolecular non-bonded interactions, and a reduced resolution. The amino acids are then represented using only the alpha carbons (C_α) of the backbone and the side-chain beads. All neighbouring pairs of particles within a given spatial cut-off (in our case 6 Å) are linked by a network of elastic potentials. For each pair of interacting sites, the equilibrium distance of the elastic potential is set to the inter-site distance found in the crystallographic native configuration, and the spring constant is set to 5 kcal/(mol Å²) to ensure sufficient internal rigidity. This elastic representation was implemented to model the A1 and A3 domains since they remain stable under vWf elongation.

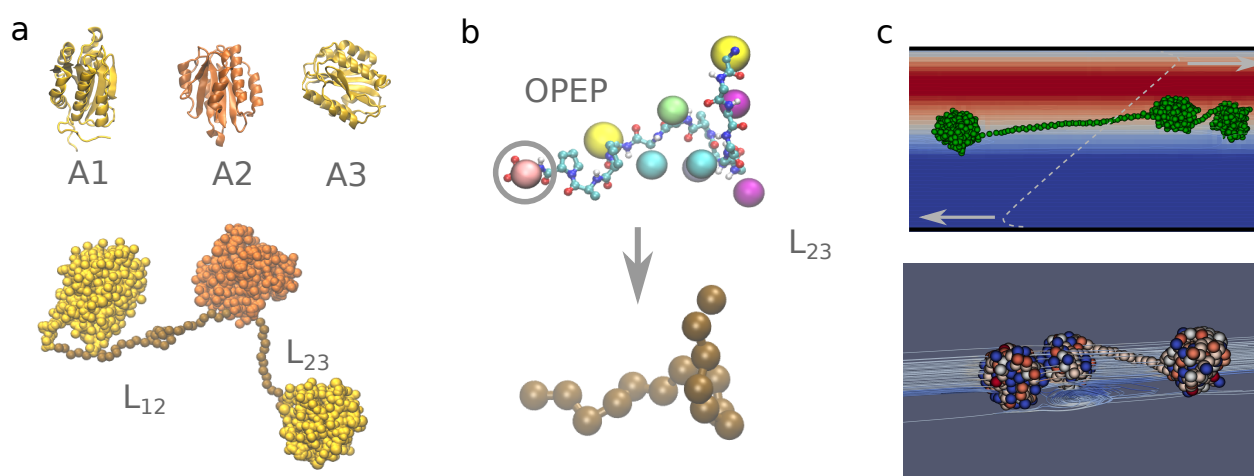


Figure 1. A1A2A3 system. Panel (a). Molecular representation of the A1A2A3 construct modelled in this work. In the top part we show the individual domains (from X-ray PDB structures) represented by their secondary structures. In the bottom part we show the A1A2A3 construct including the flexible amino acid sequences connecting the domains. Panel (b). The linker chain is represented using the OPEP (top) and a single bead (bottom) models for the amino acids. Panel (c). Representation of the Couette flow generated in the simulation box of the A1A2A3 system (top) with the velocity field represented by the colour gradient. In the (bottom) we highlight the local deformation of the velocity streamlines around the protein domains A1A2A3.

Finally, in order to model the linkers between domains A1 and A2 (L_{12}), as well as A2 and A3 (L_{23}), we derived a simplified chain representation based on the OPEP force field. For each linker we first performed a simulation with the full flexible OPEP force field. In this modelling we did not account for the glycosylations of the amino-acids. By focusing on the C_α positions, we extracted the overall bond and angle distributions for three consecutive C_α atoms, and we derived the equilibrium mean values as well as the force constants of the associated harmonic potentials via Boltzmann inversion, see SI Figure S1. The obtained values are, $r_0 = 3.8$ Å and $k_b = 149$ kcal/(mol Å²) for the bond potential, and $\theta_0 = 110^\circ$ and $k_\theta = 1.5 \times 10^{-3}$ kcal/(mol) for the angular potential. A torsional potential based on typical OPEP backbone parameters ($\phi_0 = 180^\circ$, multiplicity $n = 2$ and force constant $k_\phi = 1$ kcal/mol) was also introduced to ensure a correct extension of the chain. Finally, in order to provide a chemical flavor to the simplified representation of the linkers, we added a non-bonded potential between the beads. First, we classified the amino acids in two main groups, polar (P) and hydrophobic (H) ones, then we employed the OPEP potential interactions parameters to obtain an averaged interaction potential for the interactions of P-P, H-H, and P-H beads.

2.3. The Lattice Boltzmann Molecular Dynamics Framework

The LBMD, or in a more general term Lattice Boltzmann particle dynamics [44], has been previously introduced in the context of polymer physics to incorporate hydrodynamic effects in simulations based on implicit-solvent molecular representations [52–54]. The approach has also recently been extended to the investigation of biological systems at quasi-atomistic resolution [55,56]. In particular, the coupling between particle and fluid dynamics was effective in modeling amyloid aggregation [57,58], crowded protein solutions [45,59], nanoscale vesicles [60], and protein unfolding under shear flow [46,61].

The coupling between particles and solvent arises from a Stokes-like drag force acting on each particle:

$$\vec{F}_i^D = -\gamma(\vec{v}_i - \vec{u}_i) \quad (1)$$

where \vec{v}_i is the i -th particle's velocity, \vec{u}_i is the fluid velocity \vec{u} smeared over a finite extension of the i -th particle, and γ is the frictional coupling, an adjustable parameter in the methodology. The drag force adds up to the usual conservative forces derived from the Hamiltonian of the system, $\vec{F}_i^C = -\vec{\nabla}_i U(\{r\})$ and to a random white noise, \vec{F}_i^R , that represents thermal fluctuations.

In our simulations, the LB implementation uses the BGK (Bhatnagar-Gross-Krook) collisional operator [62] with a lattice spacing of 5 Å, a resolution needed to sufficiently resolve local hydrodynamic interactions for macromolecular systems. The solvent kinematic viscosity was set to the value for bulk water at ambient conditions.

For the dynamics of the A1A2A3 system we used a multiple time stepping to separate the integration of the bonded interactions involved in the flexible A2 domain, the A1/A3 elastic network domains and L₁₂/L₂₃ linkers (timestep of 3 fs), and the non-bonded interactions (timestep of 9 fs). The molecular and fluid dynamics were evolved synchronously at the largest molecular timestep. The LBMD simulations were carried out using the code MUPHY [63]. Technical aspects concerning the method and its numerical setup as well as shear generation have been detailed in previous works [55,58,61]. The solvent undergoes an external perturbation and produces a Couette flow. This, in turn, is perturbed by the presence of the vWF protein and its dynamics (two-way exchange). This modifies locally the linear velocity gradient of the Couette flow (see Figure 1c and SI Figure S2). In order to avoid finite size effects along the elongation direction we set up a simulation box of size 80 × 20 × 20 nm.

2.4. All-Atom Molecular Dynamics

All-atom simulations of clamp-force unfolding of the A2 domain were carried out in Gromacs 2018.4 patched with PLUMED 2.5, using the Amber a99SB-disp force-field for disordered and ordered proteins together with its TIP4P-disp water model [64]. The structure of the A2 domain (PDB:3gxb) was processed using the Gromacs package tools and solvated in a $\sim 7 \times 7 \times 50$ nm simulation box with 10 Na⁺ ions to ensure charge neutrality. The system was minimized then equilibrated in the NPT ensemble (300 K, 1 bar) using the canonical velocity-rescaling thermostat and the Parrinello-Rahman barostat. In order to obtain uncorrelated unfolding trajectories, five 10 ns equilibration trajectories were produced using different random seeds and snapshots at 6, 7, 8, 9 and 10 ns were extracted from each one in order to generate 25 starting configurations for the following SMD experiments. The end-to-end distance projection on the z-axis was biased with a constant force using PLUMED. In all cases the protein very fastly aligned along the pulling direction. An upper wall was set at 46 nm to prevent extension passed the box size. Simulations were run from 20 ns up to 180 ns so that a majority achieved unfolding of the cleavage site. The z-projected end-to-end distance was recorded every 10 ps and unfolding kinetics were analyzed using the Maximum Likelihood (ML) framework described bellow.

2.5. Kinetic Models for Unfolding

The unfolding kinetics of a protein subjected to a pulling tensile force has been often described using a two-state irreversible Markovian model for the underlying evolution.

In this model, the first passage time τ to move from the folded configuration toward the unfolded one follows an exponential probability distribution that depends on the forward reaction rate k :

$$f(\tau; k) = k \exp(-k\tau) \quad (2)$$

According to Bell's model [65], when an external force is applied to one terminus of the protein, the barrier-limited unfolding transition is accelerated because of the added mechanical work along the unfolding path, $k = k_0 e^{\beta F \delta x}$, where k_0 is the rate at null force, $\beta = 1/k_B T$, and δx is the distance between the reactant state and the transition state along the pulling direction. This expression is valid for a two-state, thermally activated reaction happening on a one-dimensional energy landscape, conditions that can be challenged in realistic systems [66]. Moreover, in MD simulations that generally allow to access molecular processes at the $ns - \mu s$ timescale, computational limitations impose the use of pulling forces higher than those usually employed in experiments. In the high-force, diffusion-dominated regime, one can instead model the unfolding process as a Brownian motion with drift along the unfolding coordinate [67]. In that case, the time τ for going from the reactant state (at $x = 0$) to some unfolded state (at $x = \alpha$) follows an inverse Gaussian distribution that reads as follows:

$$f(\tau; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi\tau^3}} \exp\left(-\frac{\lambda(\tau - \mu)^2}{2\mu^2\tau}\right) \quad (3)$$

with the parameters of the distribution defined as $\mu = \alpha/\nu$ and $\lambda = \alpha^2/\sigma^2$. They physically relate to the displacement amplitude of the underlying Brownian motion σ , and to the drift ν that incorporates the combined effect of the external force and of diffusivity, i.e., the action of friction along the path.

2.6. Maximum Likelihood Analysis of Unfolding Kinetics

In order to characterize the unfolding dynamics of the A2 domain under the action of an external pulling force, or fluid shear, we followed the time evolution of several independent runs and introduced a suitable order parameter to describe the time-dependent unfolding. At a given threshold of the order parameter, an unfolding event is recorded, yielding for that simulation a first unfolding time τ . In cases where the unfolding criterion was not reached before the end of a trajectory, the final time τ^c was registered instead. Finally, using the extracted distributions of unfolding times or simulation times we estimate the parameters of the two kinetic models described above using a maximum likelihood (ML) framework for survival analysis. Let $f(\tau; \theta)$ be the probability density function for first unfolding time τ in a given model (exponential model or inverse gaussian model), with parameters θ . Let $\{\tau_i\}_{i=1\dots N}$ be the sample of observed unfolding times and $\{\tau_j^c\}_{j=1\dots M}$ the ending times of trajectories that did not unfold yet (that is right censored data). The likelihood function is defined as:

$$\mathcal{L}(\theta) = \prod_{i=1}^N f(\tau_i; \theta) \prod_{j=1}^M \int_{\tau_j^c}^{+\infty} d\tau f(\tau; \theta) \quad (4)$$

The best estimate $\hat{\theta}$ for the model parameters is then given by:

$$\hat{\theta} = \operatorname{argmax}_{\theta} \mathcal{L}(\theta) = \operatorname{argmax}_{\theta} \ln \mathcal{L}(\theta) \quad (5)$$

where the logarithm of the likelihood function is used for computational convenience.

For the exponential model (2), each force is analyzed independently using the corresponding dataset ($\theta \equiv k_F$). For the inverse Gaussian model (3), the scale parameter λ is estimated jointly for all forces while the location parameter μ is estimated for each of the n forces ($\theta \equiv \{\mu_{F_1}; \mu_{F_2}; \dots; \mu_{F_n}; \lambda\}$).

The analysis of the simulation data was performed using the Python scripting language. The parameters were estimated using the *minimize* function of the SciPy package.

3. Results

3.1. Validating the Unfolding of the A2 Domain

In this section we discuss the unfolding of the A2 domain under the action of an external mechanical force. We compare the results from CG and all-atom simulations. The comparison of the unfolding kinetics and structural unfolding paths between the two sets of simulations is presented as validation of the CG model for the study of the extended A1A2A3 fragment under shear flow.

3.1.1. Single A2 Domain: Pulling with a Constant Force

We carried out steered Molecular Dynamics simulations at 300 K on the A2 domain of vWf modelled using the OPEP force-field. Coupling with the fluid was not considered during these simulations. The domain was simulated in a box of dimension $14 \times 7 \times 7$ nm. In the simulation a constant force is applied to one terminus of the domain while the other is kept fixed in space during the dynamics. We explored a range of forces between 300 pN and 750 pN. Generally, the simulations were stopped once unfolding was achieved. Otherwise, we extended the trajectories up to 100–500 ns, depending on the force value. For each value of the pulling force, up to 20 independent simulations were performed. We also performed a set of atomistic simulations used as a basis of comparison for the unfolding mechanism. These MD simulations were based on the Amber99SB-disp force field [64], and using pulling forces of 400 pN and 600 pN.

In order to quantify the accessibility of the cleavage site for ADAMTS13, we define unfolding as the full exposure of the core amino acid ^{111}Tyr . We further define a collective variable counting the number of backbone atoms within 10 \AA of any of ^{111}Tyr atoms (excluded), this variable measures how the residue is buried in the protein matrix before getting exposed to the solvent, and referred as coordination number. This definition allows to compare the results from coarse-grained and all atom simulations. When plotted against time (see left panels in Figure 2), the coordination number of ^{111}Tyr decreases as parts of A2 unfold and the tyrosine gets exposed to the solvent. Along a typical unfolding trajectory, in agreement with past studies [36], the domain visits several metastable states, whose lifetimes increase when the pulling force is lowered. It should be noted that for the lowest tested force, 300 pN, none of the trajectories displayed significant exposure of ^{111}Tyr on the simulation timescale.

When the number of neighbouring atoms is under 50, we consider ^{111}Tyr fully exposed and the protein unfolded. Since unfolding is a stochastic process, some trajectories, especially at low force values, did not reach this criterion before the end of the simulation. While in these cases we cannot devise a first unfolding time, they are nonetheless informative since they yield a lower bound for it. Hence, we use a maximum-likelihood approach to analyze together both the “unfolded” trajectories and the “not unfolded” ones (see Methods). For each force value F the best estimate for the unfolding rate $\hat{k}(F)$ is expressed as the inverse of the mean first unfolding time, and is computed for the two kinetic models described in the Method section. For the exponential model, each set of simulations for a given force value F is analyzed independently to evaluate $\hat{k}(F)$. For the inverse Gaussian model we could not proceed with an independent evaluation of both $\hat{\mu}(F)$ and $\hat{\lambda}(F)$ at each force given our sample sizes. Instead, we further constrained the optimization problem by assuming a force-independent $\hat{\lambda}$. Such an assumption is consistent with the interpretation that λ relates to the elementary displacement of the underlying Brownian motion (σ) and the escaping distance (α), none of which should depend on the applied force. On the other hand, independent parameters $\mu(F)$ are provided for each force value F . The resulting set of parameters was optimized globally using a single ML procedure. The obtained results are shown in Figure 2. The central chart shows the unfolding kinetics obtained using the two models. The panels in the right part of the figure report the unfolding time

distributions associated with the two models. The results from atomistic simulations are also shown for comparison and were fitted with the exponential model only, since the inverse Gaussian model could not be reliably fitted on this limited dataset.

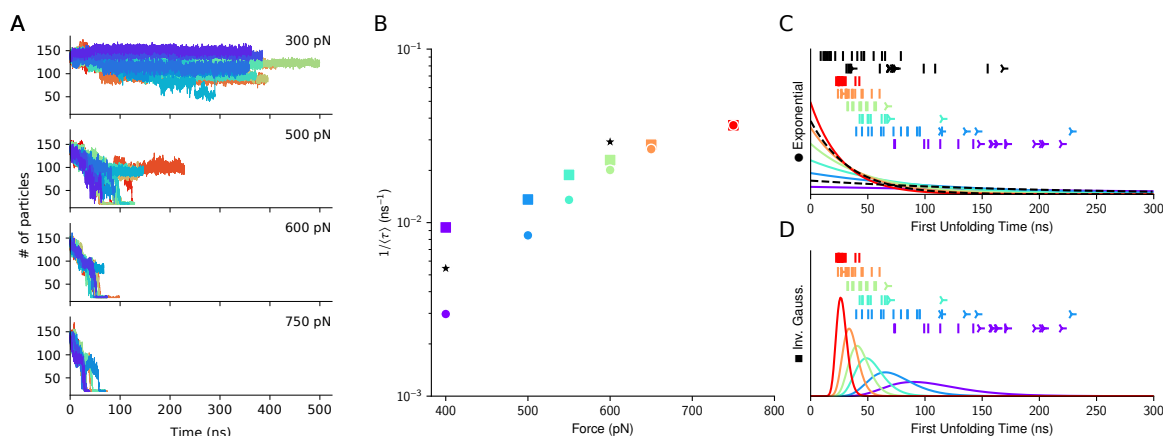


Figure 2. Exposure of the cleavage site upon force-clamp unfolding. Panel (A). Time evolution of the coordination number of ¹¹¹Tyr under the action of mechanical forces of different magnitudes in coarse-grained simulations. At the beginning the residue is buried in the protein matrix, and the number of protein backbone atoms in its proximity is the highest. During unfolding this number decreases as a sign of progressive exposure out of the protein matrix. For each force value, up to 20 independent simulations were run. Panel (B). Kinetics for the cleavage site exposure. For the CG model the inverse mean first unfolding times is estimated for the exponential (circles) and inverse Gaussian (squares) distributions. For all-atom simulations only the exponential model is used (black stars). Panels (C,D). The estimated distributions are shown (solid lines) together with the simulated unfolding times (vertical bars) or end of not-yet-unfolded trajectories (three-pointed stars).

The comparison between the distribution shapes and the empirical data strongly suggests that exponential decay does not apply well to the high force regime investigated, 400–750 pN. A lower force of 300 pN was also simulated but hardly reached the unfolding criterion on the simulation time frame. In this regard, the inverse Gaussian model offers much more realistic predictions. Moreover, in the exponential fit, estimated rates strongly deviate from the exponential dependence on applied force predicted by Bell's model. Taken together, the results suggest that the dynamics of the system is highly diffusive in this high-force regime and that the observed kinetics cannot be simply related to a phenomenological energy barrier extracted from lower-force experiments. Hence, the observed dependence on force does not arise from easier crossing of a single well-defined free-energy barrier, but can be understood as increased drift velocity in the driven Brownian model. This agrees to what already observed in other molecular simulations, where the A2 mechanical unfolding is a multi-step process proceeding via several metastable states [36].

3.1.2. Single A2 Domain: Structural Analysis of Force Unfolding

In order to characterize the progress of the unfolding that leads to the exposure of ¹¹¹Tyr, we visually inspected the trajectories and singled out three sequential structural events: (i) the unzipping of β strands 5 (res. 129–136) and 6 (res. 155–157); (ii) the unzipping of β strands 4 (res. 108–114) and 5; and (iii) the unzipping of β strands 1 (res. 3–10) and 4. We note that the unfolding of the A2 domain in the force-clamp setup begins from the C-terminus and demonstrates transient intermediate states in accordance with prior experimental and structural studies [14,22]. A pictorial representation of this sequence is given in Figure 3, left panels.

To support that intuition we analyzed the number of native hydrogen bonds formed between each β strands pair and involving the backbone atoms. To this purpose, native

contacts between the backbone O and H_N atoms are defined using a 3 Å cutoff in the reference crystal structure. Then, the number of contacts is defined as:

$$C(t) = \sum_{\text{contacts}} \frac{1}{1 + \exp\{a(d(t) - d_{\text{ref}})\}} \quad (6)$$

where a is a smoothing factor (set at 0.5 \AA^{-1}), $d(t)$ is the O-H_N distance at time t and d_{ref} is the distance in the reference structure.

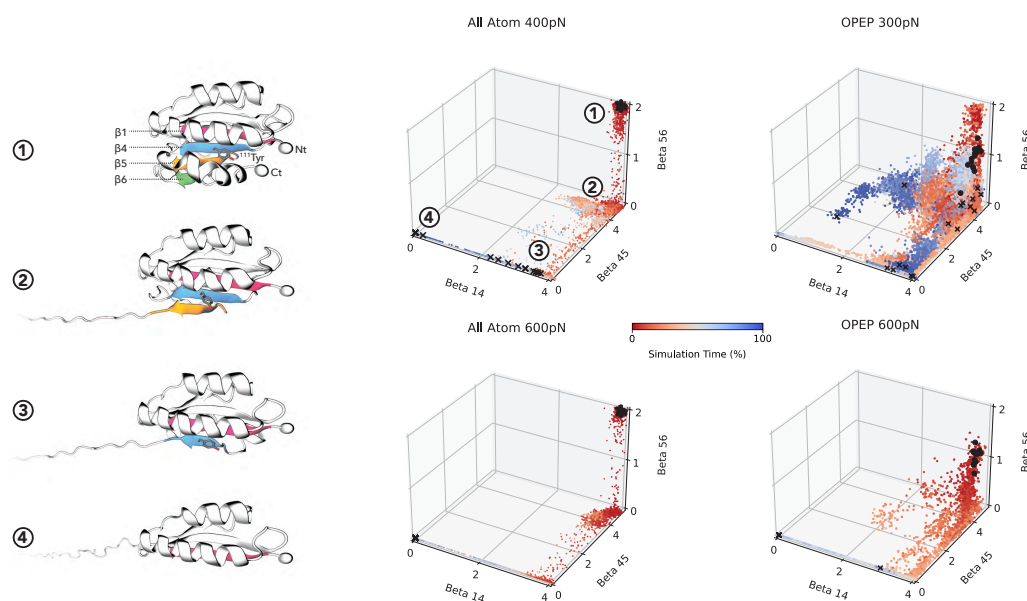


Figure 3. A2 unfolding path. **(Left).** From top to bottom, typical unfolding pathway of A2 in force-clamp simulation. As beta strands 6 (green), 5 (orange), 4 (blue) and 1 (red) unfold, residue ¹¹¹Tyr is more and more exposed. Snapshots were rendered in VMD and numbered 1–4. **(Center, Right).** The unfolding trajectories are projected on the 3-dimensional CV space of the native H-bonds between beta strands 1 and 4, 4 and 5, 5 and 6. Black dots and crosses denote the beginning and ending of the trajectories, respectively. Trajectories are colored from red (beginning) to blue (ending). In the first chart, the number 1 to 4 indicate the typical protein structure along the unfolding path shown in the left of the figure.

The number of contacts for the three pairs of β strands are accumulated for all trajectories at a given force and shown on Figure 3 as 3D scatter plots. For the sake of comparison we have reported the results from OPEP CG simulations and all atoms simulations. The sequential breaking of each part of the β sheet is very clear for all-atom simulations. While OPEP exhibits a more flexible behaviour, the overall sequence of event is respected on average. Trajectories that do not reach full unfolding stay in metastable states either at the second (β 4–5) or the third (β 1–4) steps, hinting for the key role of these structural motifs in the mechanical resistance of A2.

This sequence of unfolding detected in our simulations is consistent with that described in previous work using SMD simulations [27,36].

3.1.3. Exposure of the Cleavage Site

It is worth noticing that the cleavage of vWf-A2 is a multistage process that obeys the known “key-lock” principle. The active domain of metalloproteinase ADAMTS13 has relatively low proteolytic activity against human von Willebrand factor, and they are sometimes considered as a non-optimal enzyme-substrate pair [68]. Before the metal ion cuts the bond, the enzyme must anchor to the unfolded vWf, recognize correctly the location of the cleavage site and place the active site properly [43]. The exposure of the cleavage site is a necessary, but not a sufficient condition.

In the disintegrin-like domain of ADAMTS13 a disordered region called V-loop is present. Although being variable in length and sequence, such region is also found in different ADAM-family proteins [69]. In ADAMTS13 this region contains certain charged residues located near Leu350 and Arg349, which should interact with a few specific residues, most probably Ala1612 and Asp1614 (^{118}Ala and ^{120}Asp in the local notation), of the unfolded A2 domain [24,70,71]. Recently, it was suggested that these charged residues may collaboratively create a vWf-binding exosite on the surface of the metalloprotease [24]. This hypothetically allows to position the scissile bond Tyr1605-Met1606 for cleavage, noticeably affecting the rate constant and catalytic efficacy of proteolysis [43]. Given the distance between the aforementioned residues in ADAMTS13, the uncoiled part of A2 should not only contain the scissile bond itself, but also a segment of approximately 25 Å in the direction to the C-terminal end, which we here refer to as the *recognition site*. In case of force-pulling since the force is applied at the C-terminus the recognition site is exposed before the cleavage site, see Figure 4. As mentioned above, under real conditions, the unfolding of the A2 domain is however triggered by the shear forces. Hence the unfolding directionality and the associated sequence of events might differ. For a set of unfolding simulations of A2 under shear flow we have monitored the exposure of the recognition and cleavage sites, see SI Figure S3. It appears that the correlation among the two sites exposures is lost, and less evident than in the force unfolding. This is somehow not surprising since the isolated domain rotates in the shear field and different structural elements are periodically exposed to the elongation component of the flow that causes the unfolding to proceed along a variety of pathways. This is further confirmed by looking at the unfolding of the elementary secondary structures as reported in SI Figure S4. The variability of unfolding pathways under shear flow when compared to the unfolding caused by a directional pulling force was already observed for other proteins, and relates to the cyclic application of the elongation force to different portions of the unfolding protein structure [46,61].

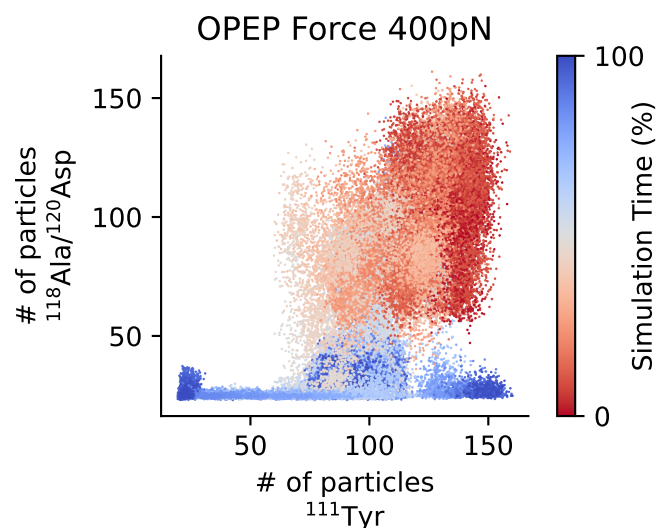


Figure 4. Exposure of the cleavage and the recognition sites of A2 domain under force. The CG trajectories of force unfolding (400 pN) are projected as a scatter plot on the 2D space of the coordination number of ^{111}Tyr (cleavage site) and that of $^{118}\text{Ala}/^{120}\text{Asp}$ (recognition site). The sequential exposure of the recognition site before the cleavage site can be appreciated. Trajectories are colored according to progress from red (beginning) to blue (ending).

3.2. The Three Domains A1A2A3: The Unfolding of A2 in Shear Flow

In the previous section we have shown that the OPEP model describes the force induced unfolding paths and kinetics of the isolated A2 domain very similarly to what obtained by an all-atom model in explicit solvent. Here, we investigate a more realistic system, with the A2 domain linked to the neighbouring domains, A1 and A3, as found in the vWf. The complete system is described in the Methods section. In order to study the

effect of shear flow we first generated a set of independent initial states by equilibrating the system in the absence of shear perturbation and by extracting a configuration every 1 ns. In these initial configurations the A1 and A3 domains are not in contact with A2, but the linkers are not fully extended. Each configuration was used to start a simulation where the fluid shear flow was activated. First we considered the high shear regime that induces unfolding and exposure of the cleavage site on a timescale of tenths of nanoseconds.

The exposure dynamics of ^{111}Tyr are monitored as described in the previous section by considering the decrease of the number of A2 domain backbone atoms contained in a sphere of radius $r_c = 10 \text{ \AA}$ centered at the cleavage site. The evolution of the coordination number of ^{111}Tyr is reported in Figure 5 for several trajectories and for different shear rates. While at the highest shear rate almost all trajectories lead to the complete exposure of ^{111}Tyr , weaker perturbations lead in many cases to incomplete unfolding and the cleavage site remains buried.

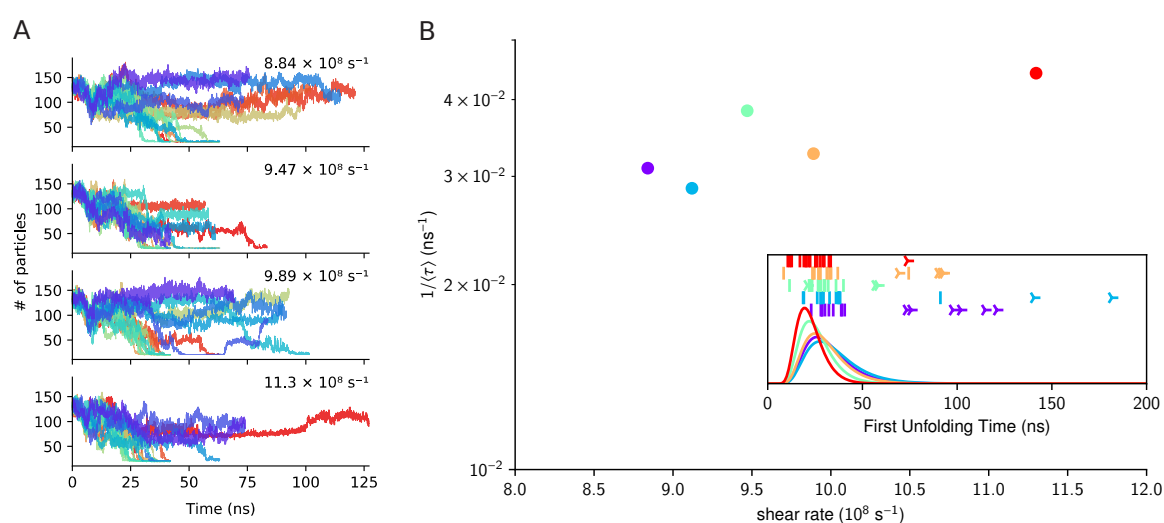


Figure 5. Exposure of cleavage site in A1A2A3 under shear flow. Panel (A). Time evolution of the coordination number of the cleavage site in A2 expressed as the number of backbone atoms inside a sphere of radius $r_c = 10 \text{ \AA}$ centered on ^{111}Tyr . In the four panels we report data for several shear rates. Panel (B). Shear-dependent mean exposure kinetics derived from the estimate of the first unfolding time obtained from maximum likelihood approach (ML). In the inset chart the ML derived distributions for the exposure time are reported for the different shear rates (solid lines) together with the simulated unfolding times (vertical bars) or the end of not-yet-unfolded trajectories (three-pointed stars).

The elongation dynamics in shear flow can also be followed by monitoring the chain extension of the A2 domain expressed by its end-to-end distance, see Figure S5. We observe two typical scenarios: the first one, when a continuous diffusive-like evolution of the extension leads to unfolding and, therefore, to full exposure of the cleavage site (the contour length of A2 is $\approx 65 \text{ nm}$), and the second one, when the extension is limited by a metastable local configurational lock, and eventually reverses back. This latter type of event is caused by the rotational dynamics of the whole complex induced by the Couette velocity field and that leads to partial collapse of the three domains A1, A2 and A3. This compaction of the three domains reduces the tensile force acting on the A2 domain, which acquires an unfolded molten globule conformation instead of an extended linear one. In these cases the cleavage site ^{111}Tyr is not accessible. A molecular view of the two scenarios is given in Figure 6 for complete exposure, and in SI Figure S6 for partial exposure.

By using the same strategy adopted for the force-induced unfolding, we quantified the unfolding kinetics by applying the diffusive model (the most adequate to fit the data), see Figure 5. By comparing the characteristic unfolding time with those obtained for the simulation of A2 under force, we deduce a semi-quantitative match among the shear values $9\text{--}10 \times 10^9 \text{ s}^{-1}$ and a directional mechanical force of 500–700 pN.

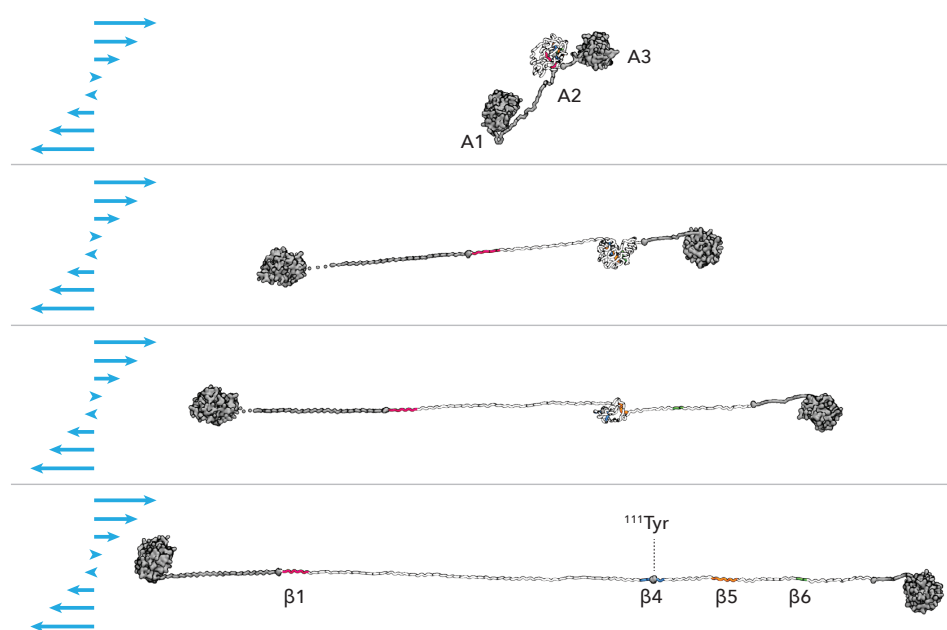


Figure 6. Molecular view of the A1A2A3 construct under shear. Sequence of simulation snapshots that represents the molecular steps of A2 extension under shear. The sets of amino-acids forming in the native state the β strands β_1 , β_4 , β_5 , β_6 are represented with different colors. The cleavage site ^{111}Tyr is explicitly represented.

We have then verified the simultaneous exposure of the cleavage and recognition sites, see Figure 7. For the three domains system A1A2A3 the exposure of the recognition site anticipates slightly that of the cleavage site. One may extrapolate this observation on the vWf multimers: for longer chains the mechanical load on A2 domain is more resemblant to the directional pulling. This is especially pronounced for the grafted multimers, which usually experience greater local tension as compared to the free-flowing ones [12,72]. We also see that under a higher shear stress the extension of the A2 domain has a tendency to follow the complete unfolding pathway, rather than the reassembly pathway, so that the probability for the correct ADAMTS13-vWf complex formation is higher.

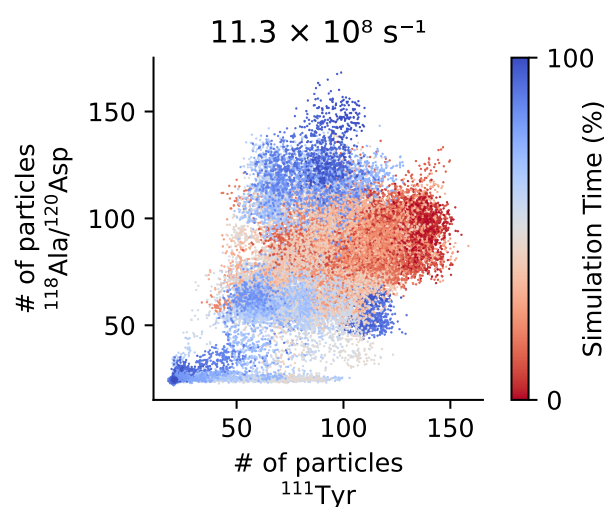


Figure 7. Exposure of the cleavage and the recognition sites of A2 domain under shear. The trajectories of shear unfolding ($\dot{\gamma} = 11.3 \times 10^8 \text{ s}^{-1}$) are projected as a scatter plot on the 2D space of the coordination number of ^{111}Tyr (cleavage site) and that of $^{118}\text{Ala}/^{120}\text{Asp}$ (recognition site). The sequential exposure of the recognition site before the cleavage site can be appreciated. Trajectories are colored according to progress from red (beginning) to blue (ending).

3.3. The Three Domains A1A2A3: The Shear-Induced Unfolding and Mechanical Tension

In this section we relate more explicitly the fluid shear flow to the mechanical tensile force that triggers the unfolding of the A2 domain. We recall that, at longer length scales, and considering specifically the multimeric vWf, the analysis of the tension profile recorded along the vWf chain modelled as a polymer moving under the action of shear flow and hydrodynamics interactions, suggested a general mechanism (the creation of a local protrusion) that leads to the transition from the collapsed to the elongated shape [40]. By combining scaling analysis with a two-state model that describes the A2 domain as open or close, it was proposed that the rate for ADAMTS13 cleavage activity is limited by the fraction of cleavage sites that get exposed because of the tensile action of the shear flow on the protrusions [39]. When focusing at the protein length scale, by using a simple mechanical model, and energetic considerations, Jaspe and Hagen [73] set the critical shear rate to unfold a mid-size globular protein to $\dot{\gamma} > 10^7 \text{ s}^{-1}$. Numerical simulations based on the OPEP CG model confirmed this, and showed that in order to observe the unfolding of small/mid size proteins in the nanosecond timescale shear rates should be indeed as high as $\sim 10^9 \text{ s}^{-1}$ [46,61]. At this shear rate value, the fluid acts on the rotating proteins applying an elongation mechanical drag along the unfolding reaction coordinate (e.g., the HB ramp in β -hairpin) that can spike at values in the order of several 10^2 pN . This is coherent with the typical values for the forces applied in pulling MD simulations. Moreover, in the previous sections by comparing the kinetics of unfolding, we found an empirical correlation among the force and the shear rate required to unfold a protein in the nanosecond timescale [46].

We now quantify at the level of the A2 terminus, where the unfolding starts, how the action of the shear translates in a mechanical pulling force. In our model the terminal residues of the A2 domain are attached to the initial beads of the linkers L_{12} and L_{23} via a harmonic bond, $\frac{1}{2}k(d - d_0)^2$. Under the perturbation of the shear flow this bond is stretched, and it is possible to record the instantaneous stressing force $k(d(t) - d_0)$. This stress can be averaged over a time interval τ_w in order to estimate the evolution of the local mechanical force acting on the A2 terminal sites along the unfolding process, $\bar{f} = \langle k(d(t) - d_0) \rangle_{\tau_w}$.

In Figure 8 we show an exemplar plot, where for the shear rate $\dot{\gamma} = 1.13 \times 10^9 \text{ s}^{-1}$ we report the force acting on the A2 terminus via the L_{12} linker during the unfolding event. The force evolution is contrasted with the evolution of the coordination number of the ^{111}Tyr . During a first initial transient phase of about 5 ns the force starts raising as effect of the fully elongation of the linkers that transfer the force to A2. In this phase the system is in tension and the first extension of the A2 domain occurs. After 10 ns the force peaks at a value of about 500 pN inducing a further extension of the domain. Such values agree with the estimations for local tension in multimeric vWf under shear or elongational flow [14]. The action of the force extends for the following 30 ns, with the tension progressively decreasing till a final extension jump that leads to the full exposure of ^{111}Tyr .

In SI Figures S7 and S8 we report similar plots for both the L_{12} and the L_{23} connections to the A2 domain obtained from several independent runs. We generally observe that the resulting pulling force is more intense at one edge than the other. This asymmetric response is due to the different lengths of the linkers, 37 and 15 respectively, and the different velocity field experienced by the attached domains. The longer linker L_{12} is more easily subject to tension and generates a loading force at the edge of the A2 domain higher than that generated at the other terminus by the shorter L_{23} linker. However, it must be noted, as also shown in Figure 6, that the unfolding of A2 proceeds at both C- and N-ter, and the protein is exposed to a total load that combines the pulling forces at the extremities.

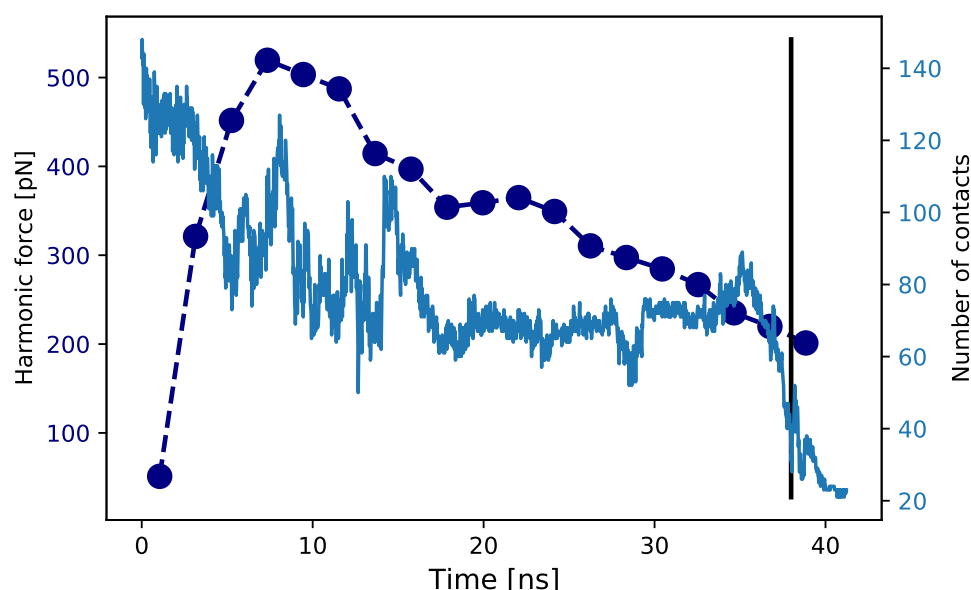


Figure 8. Tensile force. Time evolution of the mechanical force acting at the A2 terminus connected to the L_{12} linker (dark blue). The force is estimated as a moving average over a time window of 2.1 ns. We also report the time evolution of the coordination number of ^{111}Tyr (light blue) until the full exposure criterion is reached, as shown by the vertical line at 38 ns.

3.4. The Three Domains A1A2A3: The Rotational Dynamics

In order to better characterize the competitive effects on the three domains of the extensional and rotational components of the shear flow, we built a second model of the A1A2A3 system. In this model, all three domains are described by an elastic network so that also the A2 domain maintains its folded structure. The model allows us to span longer timescales in the simulation since we can increase the integration time-step and reduce the intramolecular degrees of freedom for A2. Therefore we can investigate the overall rotational dynamics of the molecular construct, and the inter-domains interactions, even at small shear rate values (10^5 – 10^6 s^{-1}).

First, we quantify the rotational dynamics of the two peripheral domains, A1 and A3, with respect to the central A2 domain. This is done by computing for each simulated system i the first order orientational time correlation function (TCF) $c(\tau) = \langle \cos(\theta(\tau)) \rangle_i$, where $\cos(\theta(\tau))$ is defined as the scalar product between two instances, separated by a lag time τ , of the vector connecting the centers of mass of A1(A3) and A2 domains, $\cos(\theta(\tau)) = \mathbf{R}_{12(32)}(t + \tau) \cdot \mathbf{R}_{12(32)}(t)$.

The obtained TCFs decay in time but their trend do not fit to standard decay models (e.g., exponential or stretched exponential). Therefore, from each TCF, we have individuated the value when it reaches zero the first time. These characteristic times averaged over independent trajectories and as a function of the shear rates are plotted in panel A and B of Figure 9 for the rotation of the two domains A1 and A3 with respect to the central one A2. The obtained results clearly show that below a critical shear rate of $\sim 10^8 \text{ s}^{-1}$, the rotational dynamics of the peripheral domains is significantly slowed down (a factor 3–4 compared to the high shear values). In order to resolve such dynamics much longer simulations should be performed, but it seems that given the characteristic size of the A1A2A3 construct a shear threshold for fluid activated rotation exists. One would expect that for a longer multimer this threshold should downshift to a smaller shear rate value. Moreover, it is very intriguing that below the threshold the evolution of the relative rotation of this realistic molecular system deviates from the theoretical relationship between the field rotational period and shear rate value, $T_r = 4\pi/\dot{\gamma}$, see SI Figure S9. The decoupling between external field rotational flow and molecular rotation is due to the different contribution of the thermal motion and fluid induced rotation, with the former dominating at low shear rates, $< 10^6 \text{ s}^{-1}$.

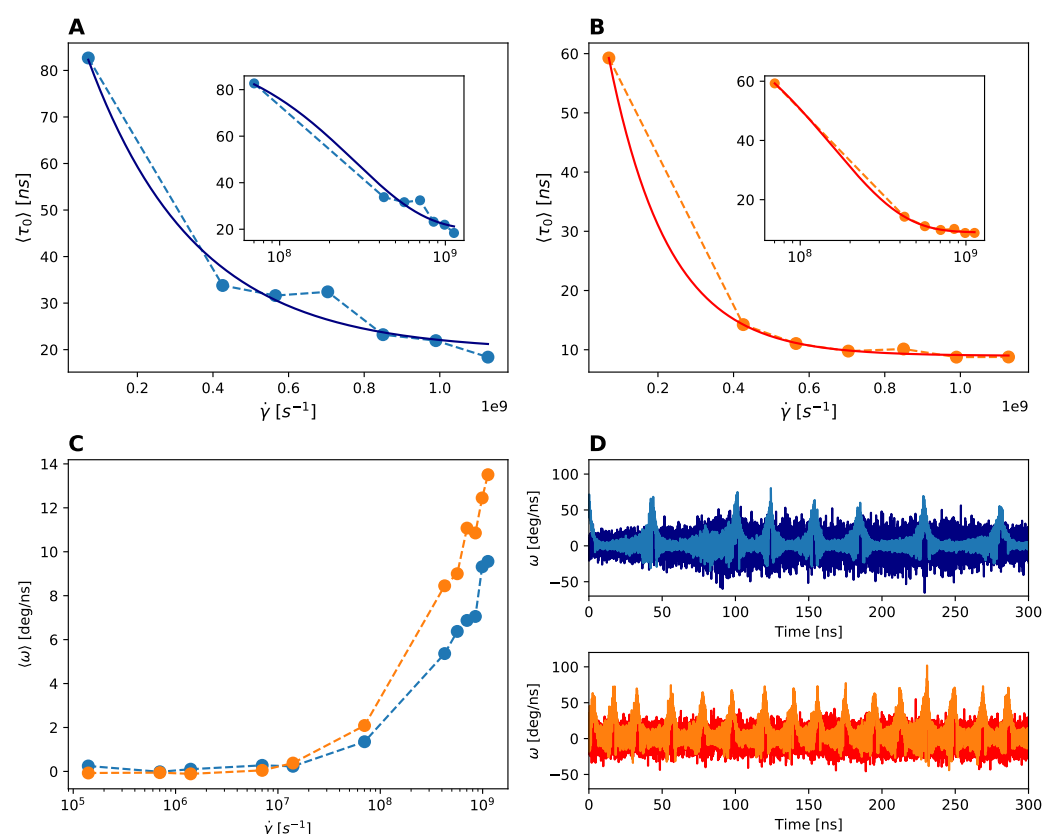


Figure 9. Rotational dynamics. Panels (A,B). Characteristic rotational time of the A1 (left panel) and A3 (right panel) domains with respect to the central A2 domain as a function of the simulated shear rates $\dot{\gamma}$. In the inset the data are represented in log-x scale for shear rates values $\dot{\gamma} > 10^7 s^{-1}$. Panel (C). Angular velocity computed for the A1 (blue) and A3 (orange) domain rotation relative to A2 as a function of the shear rate. Panel (D). Time evolution of the angular velocity for the A1 (top) and A3 (bottom) domains, for two shear rates $\dot{\gamma} = 8.5 \times 10^8 s^{-1}$ (light blue and orange) and $\dot{\gamma} = 1.4 \times 10^5 s^{-1}$ (dark blue and red).

The rotational dynamics implies that the molecular three-domain system is exposed cyclically, at least at the highest shear rates, to the tensile force and to angular acceleration. The angular velocity of the peripheral domains A1 and A3 with respect to the central A2 is computed by considering the variation of the vector connecting the centers of mass in the unit of time Δt , $\Delta \mathbf{R} = \mathbf{R}_{12(32)}(t + \Delta t) - \mathbf{R}_{12(32)}(t)$, and having for small displacements the angular variation $\omega = \Delta \mathbf{R}_{12(32)} / \mathbf{R}_{12(32)}$. The angular velocity is then computed on the plane of the rotation (XZ), and reported as the average over time and replicas versus the shear rate in Figure 9, panel C. Again, we notice the activation of the domain relative rotation above a critical shear rate, $\dot{\gamma}_c \sim 10^7$ – $10^8 s^{-1}$. On panel D of the same figure, we report for A1 (top panel) and A3 (bottom panel) rotations the time evolution of the angular velocity for an individual trajectory and at two representative shear values, $1.4 \times 10^5 s^{-1}$ and $8.5 \times 10^8 s^{-1}$. When the rotation is activated by the shear flow at $\dot{\gamma} = 8.5 \times 10^8 s^{-1}$, the angular velocity shows a typical cyclic pattern, with spikes periodically occurring every 30–40 ns. On the contrary for the lowest shear rate, the angular velocity fluctuates uniformly in time as result of the molecular thermal motion only.

In the same spirit of previous work [40], but at a very different scale, we inquire how the threshold individuated for the rotational dynamics, transposes in a neat transition of the A1A2A3 fragment conformational state. Namely, we quantified the propensity of the three domains to stay detached and elongated, or collapsed. The results are summarised in Figure 10 where we show the time-averaged value of the inter-domain distances as a function of the shear rate. It is clear that at low shear rate the three domains stay in a collapsed configuration and move as a compact unique entity. On the contrary, at high

shear rate, the three domains extend cyclically (see also Figure S10). The individuated threshold coincides with the that of the rotational dynamics, $\dot{\gamma}_c \sim 10^7\text{--}10^8\text{ s}^{-1}$. For a longer chain constituted by a larger number of domains, this threshold should decrease (although different models predict different scaling [40,73]). However, the structure of a vWf monomer [41], and the sequential linking of its domains, is expected to generate a hierarchical scale of detachment, elongation, and unfolding events. The collapsed structure observed at low shear rate, and also the cyclic collapse/extension dynamics typical of the high shear rate regime, point on the critical role of the relative domain interaction for self-inhibition toward the cleavage.

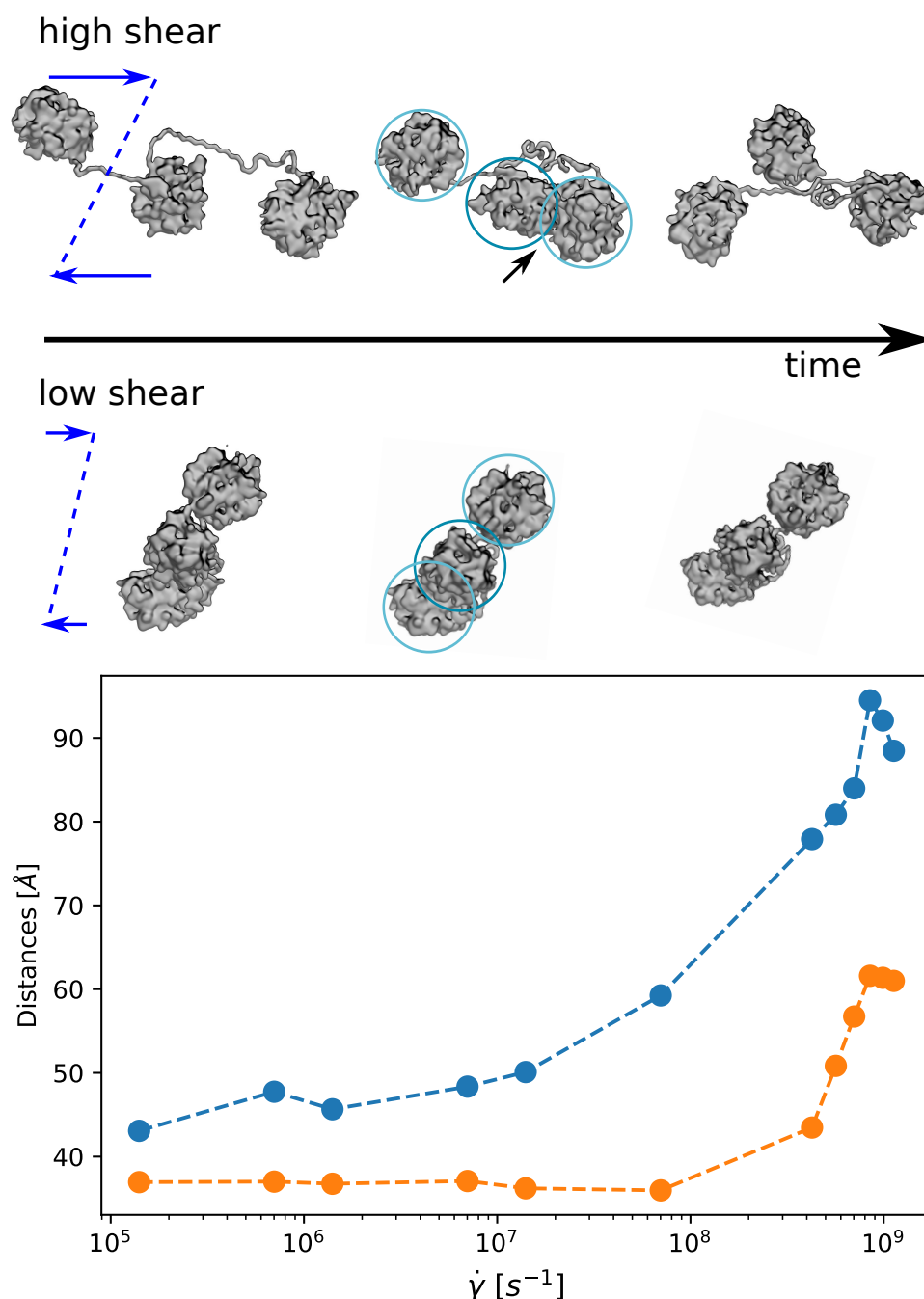


Figure 10. Elongation/Collapse dynamics. Top. Molecular representation of the A1A2A3 system moving in the shear flow at high and low shear rates. At low shear rate the three domains stay collapsed, while at high shear rate they cyclically extend. Bottom. Average inter-domain distances for A1A2 (blue) and A1A3 (orange) as a function of the shear rate.

4. Discussion

In the present paper we used both full-atomistic and coarse-grained computer simulations to explore the unfolding of the A2 domain of vWf in two perturbative regimes typical of experimental setups: the AFM-like unidirectional pulling with a constant force, and the Couette flow typical for near-wall regimes of micro-fluidic flow chambers and rotational viscosimeters. We considered the isolated A2 domain and the A1A2A3 tri-domain construct of the vWf monomer. Our results show that the unfolding pathway of the A2 domain is controlled by the sequential disruption of the β -strands in the core of the globule, while the α -helices give almost no resistance to the deformation.

The unfolding kinetics under extremely large forces, according to the presented analysis, deviate from Bell's law and are better approximated by the inverse Gaussian model. The exposure of the Tyr1605-Met1606 cleavage site may correspond to partial unfolding of the A2 domain. However, under stronger forcing, complete unfolding events leading to a fully linear unfolded protein structure have been observed in simulations over tens of nanoseconds timescale.

The unfolding pathway of the A2 domain depends on the deformation mode, applied forcing, and, importantly, on whether the individual A2 domain or the tri-domain A1A2A3 structure is loaded. For the A1A2A3 assembly in Couette flow the deformation pathway consists of the stretching of the A2 domain punctuated by relatively long periods of rotation-induced refolding, at least under high shear stress. Due to the tumbling motion, A2 can not only adopt a linear, extended conformation, but also a melted globule one that hinders ADAMTS13 binding and subsequent proteolysis. This contrasts with the behaviour of the individual A2 domain under hydrodynamic shear, in which we mostly detected transient periods of partial unfolding. Moreover, the results of CG simulations suggest that, if the shear stress is strong enough, the rotational component of the flow dominates over the protein's Brownian motion above a critical threshold. In the shear flow we observed two different scenarios of A2 extension, revealing a complex interplay between fluid dynamics and molecular structure.

Several intriguing questions emerge, including the refolding pathway of the A2 domain after deformation, the reversibility of the unfolding, the structure of the compacted globule and influence of the protein's molecular surroundings on this process. According to [14], refolding to the correct native state of the A2 domain when the tension is released is important for vWf functions in vivo, as an incorrectly refolded protein could become less resilient and eventually permit cleavage by ADAMTS13. We have observed that under a high shear rate in fluid, the A1A2A3 construct does not exhibit the native structure of A2 domain after the rotation-induced collapse. To understand what happens in a steady fluid or under a mild shear rate is a prominent goal for forthcoming research. Moreover, in our analysis we have focused on the complete exposure of the cleavage site as a prerequisite for ADAMTS13 intervention. Most studies generally assume complete unraveling of the A2 domain [74]. However, some uncertainty remains regarding the relevant A2 conformational state for cleavage. For example, Baldauf et al. [36] propose that the partially unfolded intermediate following β 5 unzipping might suffice for recognition and proteolysis.

Finally, by focusing on the relative inter-domain dynamics, we have individuated a shear-rate controlled dynamical configurational transition among a stable collapsed structure of the three domains (low shear) and a cyclically alternated collapsed/elongated structure (high shear). The preferential inter-domain interactions visible in the collapsed structures highlight the key role of these interactions in self-inhibition to cleavage.

The present study is a step toward a better understanding of the biophysical mechanisms that govern mechanoreception and mechanotransduction in proteins at the molecular level.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/polym13223912/s1>, Figure S1: Coarse-grain potential for inter-domains linker. Figure S2: A1A2A3 under shear. Figure S3: Exposure of the cleavage and the recognition sites of A2 domain. Figure S4: Unfolding path of the A2 domain under shear. Figure S5: Extension of the A2 domain within the A1A2A3 construct under shear. Figure S6: A1A2A3 under shear. Figure S7: Effective load force on A2 terminals. Figure S8: Effective load force on A2 terminals. Figure S9: A1A2A3 rotation under shear. Figure S10: A1A2A3 collapse/extension dynamics.

Author Contributions: Conceptualization, A.V.B., S.M. and F.S.; methodology, F.S., S.M. and P.D.; software, S.M. and F.S.; formal analysis, O.L.-C., E.L., A.V.B. and F.S.; investigation, O.L.-C., E.L., D.O.Y., A.V.B. and F.S.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: A.V.B. and D.O.Y. acknowledge the support from Russian Foundation for Basic Research (grant 19-01-00480-a), the Interdisciplinary Scientific and Educational School of Lomonosov Moscow State University “Photonic and Quantum Technologies. Digital Medicine”. F.S. acknowledge supports from “Initiative d’Excellence” program from the French State (Grant “DYNAMO”, ANR-11-LABX-0011-01). S.M. acknowledge support from the Gruppi di Ricerca—LazioInnova project “EMOTAC” (B83117000080002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: A.V.B. acknowledges the computing resources of the shared research facilities of HPC at Lomonosov Moscow State University [75]. F.S. acknowledges the support from GENCI (french national HPC resources) and from local resources at LBT.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Springer, T.A. von Willebrand Factor, Jedi Knight of the Bloodstream. *Blood* **2014**, *124*, 1412–1425. [\[CrossRef\]](#)
2. Slayter, H.; Loscalzo, J.; Bockenstedt, P.; Handin, R.I. Native Conformation of Human von Willebrand Protein. Analysis by Electron Microscopy and Quasi-Elastic Light Scattering. *J. Biol. Chem.* **1985**, *260*, 8559–8563. [\[CrossRef\]](#)
3. Novák, L.; Deckmyn, H.; Damjanovich, S.; Hársfalvi, J. Shear-Dependent Morphology of von Willebrand Factor Bound to Immobilized Collagen. *Blood* **2002**, *99*, 2070–2076. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Schneider, S.W.; Nuschele, S.; Wixforth, A.; Gorzelanny, C.; Alexander-Katz, A.; Netz, R.R.; Schneider, M.F. Shear-Induced Unfolding Triggers Adhesion of von Willebrand Factor Fibers. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7899–7903. [\[CrossRef\]](#)
5. Rack, K.; Huck, V.; Hoore, M.; Fedosov, D.; Schneider, S.; Gommer, G. Margination and Stretching of von Willebrand Factor in the Blood Stream enable adhesion. *Sci. Rep.* **2017**, *7*, 14278. [\[CrossRef\]](#)
6. Nesbitt, W.S.; Westein, E.; Tovar-Lopez, F.J.; Tolouei, E.; Mitchell, A.; Fu, J.; Carberry, J.; Fouras, A.; Jackson, S.P. A Shear Gradient-Dependent Platelet Aggregation Mechanism Drives Thrombus Formation. *Nat. Med.* **2009**, *15*, 665–673. [\[CrossRef\]](#)
7. Stalker, T.J.; Traxler, E.A.; Wu, J.; Wannemacher, K.M.; Cermignano, S.L.; Voronov, R.; Diamond, S.L.; Brass, L.F. Hierarchical Organization in the Hemostatic Response and its Relationship to the Platelet-Signaling Network. *Blood* **2013**, *121*, 1875–1885. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Jackson, S.P.; Nesbitt, W.S.; Westein, E. Dynamics of Platelet Thrombus Formation. *J. Thromb. Haemost.* **2009**, *7*, 17–20. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Coburn, L.; Damaraju, V.; Dozic, S.; Eskin, S.; Cruz, M.; McIntire, L. Rolling under Shear Stress Shows Differences between Type 2B and 2M von Willebrand Disease. *Biophys. J.* **2011**, *100*, 304–312. [\[CrossRef\]](#)
10. Yago, T.; Lou, J.; Wu, T.; Yang, J.; Miner, J.J.; Coburn, L.; Lopez, J.A.; Cruz, M.A.; Dong, J.F.; McIntire, L.V.; et al. Platelet Glycoprotein Ibα Forms Catch Bonds with Human WT vWF but not with Type 2B von Willebrand Disease vWF. *J. Clin. Investig.* **2008**, *118*, 3195–3207. [\[CrossRef\]](#)
11. Fu, H.; Jiang, Y.; Yang, D.; Scheiflinger, F.; Wong, W.P.; Springer, T.A. Flow-Induced Elongation of von Willebrand Factor Precedes Tension-Dependent Activation. *Nat. Commun.* **2017**, *8*, 324. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Kushchenko, Y.K.; Belyaev, A.V. Effects of Hydrophobicity, Tethering and Size on Flow-Induced Activation of von Willebrand Factor Multimers. *J. Theor. Biol.* **2020**, *485*, 110050. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Lancellotti, S.; Sacco, M.; Basso, M.; Cristofaro, R.D. Mechanochemistry of von Willebrand Factor. *Biomol. Concepts* **2019**, *10*, 194–208. [\[CrossRef\]](#)

14. Zhang, X.; Halvorsen, K.; Zhang, C.Z.; Wong, W.P.; Springer, T.A. Mechanoenzymatic Cleavage of the Ultralarge Vascular Protein von Willebrand Factor. *Science* **2009**, *324*, 1330–1334. [[CrossRef](#)] [[PubMed](#)]
15. Denis, C.; Methia, N.; Frenette, P.S.; Rayburn, H.; Ullman-Culleré, M.; Hynes, R.O.; Wagner, D.D. A Mouse Model of Severe von Willebrand Disease: Defects in Hemostasis and Thrombosis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9524–9529. [[CrossRef](#)]
16. Dong, J.F.; Moake, J.L.; Nolasco, L.; Bernardo, A.; Arceneaux, W.; Shrimpton, C.N.; Schade, A.J.; McIntire, L.V.; Fujikawa, K.; López, J.A. ADAMTS-13 Rapidly Cleaves Newly Secreted Ultralarge von Willebrand Factor Multimers on the Endothelial Surface Under Flowing Conditions. *Blood* **2002**, *100*, 4033–4039. [[CrossRef](#)]
17. Kim, J.; Zhang, C.Z.; Zhang, X.; Springer, T.A. A Mechanically Stabilized Receptor-Ligand Flex-Bond Important in the Vasculature. *Nature* **2010**, *466*, 992–995. [[CrossRef](#)]
18. Lippok, S.; Obser, T.; Müller, J.P.; Stierle, V.K.; Benoit, M.; Budde, U.; Schneppenheim, R.; Rädler, J.O. Exponential Size Distribution of von Willebrand Factor. *Biophys. J.* **2013**, *105*, 1208–1216. [[CrossRef](#)]
19. Lippok, S.; Radtke, M.; Obser, T.; Kleemeier, L.; Schneppenheim, R.; Budde, U.; Netz, R.R.; Rädler, J.O. Shear-Induced Unfolding and Enzymatic Cleavage of Full-Length VWF Multimers. *Biophys. J.* **2016**, *110*, 545–554. [[CrossRef](#)]
20. Belyaev, A.V. Long Ligands Reinforce Biological Adhesion Under Shear Flow. *Phys. Rev. E* **2018**, *97*, 042407. [[CrossRef](#)]
21. Butera, D.; Passam, F.; Ju, L.; Cook, K.M.; Woon, H.; Aponte-Santamaria, C.; Gardiner, E.; Davis, A.K.; Murphy, D.A.; Bronowska, A.; et al. Autoregulation of von Willebrand Factor Function by a Disulfide Bond Switch. *Sci. Adv.* **2018**, *4*, eaq1477. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, Q.; Zhou, Y.F.; Zhang, C.Z.; Zhang, X.; Lu, C.; Springer, T.A. Structural Specializations of A2, a Force-Sensing Domain in the Ultralarge Vascular Protein von Willebrand Factor. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9226–9231. [[CrossRef](#)]
23. Sadler, J.E. Biochemistry and Genetics of von Willebrand Factor. *Ann. Rev. Biochem.* **1998**, *67*, 395–424. [[CrossRef](#)] [[PubMed](#)]
24. Akiyama, M.; Takeda, S.; Kokame, K.; Takagi, J.; Miyata, T. Crystal Structures of the Noncatalytic Domains of ADAMTS13 Reveal Multiple Discontinuous Exosites for von Willebrand Factor. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19274–19279. [[CrossRef](#)]
25. Li, Y.; Choi, H.; Zhou, Z.; Nolasco, L.; Pownall, H.J.; Voorberg, J.; Moake, J.L.; Dong, J.F. Covalent Regulation of ULVWF String Formation and Elongation on Endothelial Cells Under Flow Conditions. *J. Thromb. Haemost.* **2008**, *6*, 1135–1143. [[CrossRef](#)]
26. Hovinga, J.A.K.; Lämmle, B. Role of ADAMTS13 in the Pathogenesis, Diagnosis, and Treatment of Thrombotic Thrombocytopenic Purpura. *Hematology* **2012**, *2012*, 610–616. [[CrossRef](#)]
27. Chen, W.; Lou, J.; Zhu, C. Molecular Dynamics Simulated Unfolding of von Willebrand Factor A Domains by Force. *Cell. Mol. Bioeng.* **2009**, *2*, 75–86. [[CrossRef](#)]
28. Zhou, M.; Dong, X.; Baldauf, C.; Chen, H.; Zhou, Y.; Springer, T.A.; Luo, X.; Zhong, C.; Gräter, F.; Ding, J. A Novel Calcium-Binding Site of von Willebrand Factor A2 Domain Regulates its Cleavage by ADAMTS13. *Blood* **2011**, *117*, 4623–4631. [[CrossRef](#)] [[PubMed](#)]
29. Xu, A.J.; Springer, T.A. Calcium stabilizes the von Willebrand factor A2 domain by promoting refolding. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3742–3747. [[CrossRef](#)] [[PubMed](#)]
30. Zhou, Y.F.; Eng, E.T.; Nishida, N.; Lu, C.; Walz, T.; Springer, T.A. A pH-Regulated Dimeric Bouquet in the Structure of von Willebrand Factor. *EMBO J.* **2011**, *30*, 4098–4111. [[CrossRef](#)]
31. O'Brien, H.E.R.; Zhang, X.F.; Sanz-Hernandez, M.; Chion, A.; Shapiro, S.; Mobayen, G.; Xu, Y.; Simone, A.D.; Laffan, M.A.; McKinnon, T.A.J. Blocking von Willebrand Factor Free Thiols Inhibits Binding to Collagen Under High and Pathological Shear Stress. *J. Thromb. Haemost.* **2020**, *19*, 358–369. [[CrossRef](#)]
32. Sacco, M.; Lancellotti, S.; Ferrarese, M.; Bernardi, F.; Pinotti, M.; Tardugno, M.; Candia, E.D.; Gennaro, L.D.; Basso, M.; Giusti, B.; et al. Noncanonical Type 2B von Willebrand Disease Associated with Mutations in the VWF D/D3 and D4 Domains. *Blood Adv.* **2020**, *4*, 3405–3415. [[CrossRef](#)] [[PubMed](#)]
33. Löf, A.; Müller, J.P.; Benoit, M.; Brehm, M.A. Biophysical Approaches Promote Advances in the Understanding of von Willebrand Factor Processing and Function. *Adv. Biol. Regul.* **2017**, *63*, 81–91. [[CrossRef](#)] [[PubMed](#)]
34. Müller, J.P.; Mielke, S.; Löf, A.; Obser, T.; Beer, C.; Bruetzel, L.K.; Pippig, D.A.; Vanderlinden, W.; Lipfert, J.; Schneppenheim, R.; et al. Force Sensing by the Vascular Protein von Willebrand Factor is Tuned by a Strong Intermonomer Interaction. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1208–1213. [[CrossRef](#)] [[PubMed](#)]
35. Aponte-Santamaria, C.; Huck, V.; Posch, S.; Bronowska, A.K.; Grässle, S.; Brehm, M.A.; Obser, T.; Schneppenheim, R.; Hinterdorfer, P.; Schneider, S.W.; et al. Force-Sensitive Autoinhibition of the von Willebrand Factor Is Mediated by Interdomain Interactions. *Biophys. J.* **2015**, *108*, 2312–2321. [[CrossRef](#)]
36. Baldauf, C.; Schneppenheim, R.; Stacklies, W.; Obser, T.; Pieconka, A.; Schneppenheim, S.; Budde, U.; Zhou, J.; Gräter, F. Shear-induced unfolding activates von Willebrand factor A2 domain for proteolysis. *J. Thromb. Haemost.* **2009**, *7*, 2096–2105. [[CrossRef](#)]
37. Posch, S.; Obser, T.; König, G.; Schneppenheim, R.; Tampé, R.; Hinterdorfer, P. Interaction of von Willebrand Factor Domains with Collagen Investigated by Single Molecule Force Spectroscopy. *J. Chem. Phys.* **2018**, *148*, 123310. [[CrossRef](#)]
38. Wu, T.; Lin, J.; Cruz, M.A.; Dong, J.F.; Zhu, C. Force-Induced Cleavage of Single VWF A1A2A3 tridomains by ADAMTS-13. *Blood* **2010**, *115*, 370–378. [[CrossRef](#)] [[PubMed](#)]
39. Radtke, M.; Lippok, S.; Rädler, J.O.; Netz, R.R. Internal Tension in a Collapsed Polymer under Shear Flow and the Connection to Enzymatic Cleavage of von Willebrand factor. *Eur. Phys. J. E* **2016**, *39*, 32. [[CrossRef](#)]

40. Alexander-Katz, A.; Schneider, M.F.; Schneider, S.W.; Wixforth, A.; Netz, R.R. Shear-Flow-Induced Unfolding of Polymeric Globules. *Phys. Rev. Lett.* **2006**, *97*, 138101. [[CrossRef](#)]
41. Belyaev, A.V. Intradimer Forces and their Implication for Conformations of von Willebrand Factor Multimers. *Biophys. J.* **2021**, *120*, 899–911. [[CrossRef](#)] [[PubMed](#)]
42. Okhota, S.; Melnikov, I.; Avtaeva, Y.; Kozlov, S.; Gabbasov, Z. Shear Stress-Induced Activation of von Willebrand Factor and Cardiovascular Pathology. *Int. J. Mol. Sci.* **2020**, *21*, 7804. [[CrossRef](#)]
43. Li, Z.; Lin, J.; Sulchek, T.; Cruz, M.A.; Wu, J.; Dong, J.F.; Zhu, C. Domain-Specific Mechanical Modulation of VWF–ADAMTS13 Interaction. *Mol. Biol. Cell.* **2019**, *30*, 1920–1929. [[CrossRef](#)] [[PubMed](#)]
44. Bernaschi, M.; Melchionna, S.; Succi, S. Mesoscopic simulations at the physics-chemistry-biology interface. *Rev. Mod. Phys.* **2019**, *91*, 025004. [[CrossRef](#)]
45. Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragolini, T.; Chebaro, Y.; St-Pierre, J.F.; Kalimeri, M.; Barducci, A.; et al. The OPEP Protein Model: From Single Molecules, Amyloid Formation, Crowding and Hydrodynamics to DNA/RNA Systems. *Chem. Soc. Rev.* **2014**, *43*, 4871–4893. [[CrossRef](#)]
46. Languin-Cattoen, O.; Melchionna, S.; Derreumaux, P.; Stirnemann, G.; Sterpone, F. Three Weaknesses for Three Perturbations: Comparing Protein Unfolding under Shear, Force, and Thermal Stresses. *J. Phys. Chem. B* **2018**, *122*, 11922–11930. [[CrossRef](#)]
47. Emsley, J.; Cruz, M.; Handin, R.; Liddington, R. Crystal Structure of the von Willebrand Factor A1 Domain and Implications for the Binding of Platelet Glycoprotein IB. *J. Biol. Chem.* **1998**, *273*, 10396. [[CrossRef](#)]
48. Bienkowska, J.; Cruz, M.; Atiemo, A.; Handin, R.; Liddington, R. The von Willebrand Factor A3 Domain Does not Contain a Metal Ion-dependent Adhesion Site Motif. *J. Biol. Chem.* **1997**, *272*, 25162–25167. [[CrossRef](#)]
49. Chebaro, Y.; Pasquali, S.; Derreumaux, P. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. *J. Phys. Chem. B* **2012**, *116*, 8741–8752. [[CrossRef](#)] [[PubMed](#)]
50. Sterpone, F.; Nguyen, P.; Kalimeri, M.; Derreumaux, P. Importance of the Ion-Pair Interactions in the OPEP Coarse-Grained Force Field: Parametrization and Validation. *J. Chem. Theory Comput.* **2013**, *9*, 4574–4584. [[CrossRef](#)]
51. Kalimeri, M.; Derreumaux, P.; Sterpone, F. Are coarse-grained models apt to detect protein thermal stability? The case of OPEP force field. *J. Non-Cryst. Solids* **2015**, *407*, 494–501. [[CrossRef](#)]
52. Ahlrichs, P.; Dünweg, B. Simulation of a Single Polymer Chain in Solution by Combining Lattice Boltzmann and Molecular Dynamics. *J. Chem. Phys.* **1999**, *111*, 8225–8239. [[CrossRef](#)]
53. Ahlrichs, P.; Dünweg, B. Lattice-Boltzmann Simulation of Polymer-Solvent Systems. *Int. J. Mod. Phys. C* **1998**, *9*, 1429–1438. [[CrossRef](#)]
54. Dünweg, B.; Ladd, A.J.C. Lattice Boltzmann Simulations of Soft Matter Systems. In *Advances in Polymer Science*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–78.
55. Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843–1853. [[CrossRef](#)] [[PubMed](#)]
56. Chiricotto, M.; Sterpone, F.; Derreumaux, P.; Melchionna, S. Multiscale Simulation of Molecular Processes in Cellular Environments. *Philos. Trans. A* **2016**, *374*, 20160225. [[CrossRef](#)]
57. Chiricotto, M.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Multiscale Aggregation of the Amyloid AB16–22 Peptide: From Disordered Coagulation and Lateral Branching to Amorphous Prefibrils. *J. Phys. Chem. Lett.* **2019**, *10*, 1594–1599. [[CrossRef](#)]
58. Chiricotto, M.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Hydrodynamic Effects on Beta-Amyloid (16–22) Peptide Aggregation. *J. Chem. Phys.* **2016**, *145*, 035102. [[CrossRef](#)]
59. Timr, S.; Gnut, D.; Ebbinghaus, S.; Sterpone, F. The Unfolding Journey of Superoxide Dismutase 1 Barrels under Crowding: Atomistic Simulations Shed Light on Intermediate States and Their Interactions with Crowders. *J. Phys. Chem. Lett.* **2020**, *11*, 4206–4212. [[CrossRef](#)]
60. Brander, A.; Timr, S.; Melchionna, S.; Derreumaux, P.; Baaden, M.; Sterpone, F. Modelling lipid systems in fluid with Lattice Boltzmann Molecular Dynamics simulations and hydrodynamics. *Sci. Rep.* **2019**, *9*, 16450. [[CrossRef](#)]
61. Sterpone, F.; Melchionna, S.; Derreumaux, P. Molecular Mechanism of Protein Unfolding Under Shear: A Lattice Boltzmann Molecular Dynamics Study. *J. Phys. Chem. B* **2018**, *122*, 1573–1579. [[CrossRef](#)]
62. Benzi, R.; Succi, S.; Vergassola, M. The lattice Boltzmann equation: Theory and applications. *Phys. Rep.* **1992**, *222*, 145–197. [[CrossRef](#)]
63. Bernaschi, M.; Melchionna, S.; Succi, S.; Fyta, M.; Kaxiras, E.; Sircar, J. MUPHY: A parallel Multi PHYsics/scale code for high performance bio-fluidic simulations. *Comput. Phys. Comm.* **2009**, *180*, 1495–1502. [[CrossRef](#)]
64. Robustellia, P.; Piana, S.; Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4758–E4766. [[CrossRef](#)]
65. Bell, G.I. Models for the Specific Adhesion of Cells to Cells. *Science* **1978**, *200*, 618–627. [[CrossRef](#)]
66. Satija, R.; Berezhkovskii, A.M.; Makarov, D.E. Broad Distributions of Transition-Path Times are Fingerprints of Multidimensionality of the Underlying Free Energy Landscapes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 27116–27123. [[CrossRef](#)]
67. Luccioli, S.; Imparato, A.; Mitternacht, S.; Irback, A.; Torcini, A. Unfolding Times for Proteins in a Force Clamp. *Phys. Rev. E* **2010**, *81*, 010902. [[CrossRef](#)]
68. Muia, J.; Zhu, J.; Greco, S.C.; Vanhoorelbeke, K.; Gupta, G.; Westfield, L.A.; Sadler, J.E. Phylogenetic and Functional Analysis of ADAMTS13 Identifies Highly Conserved Domains Essential for Allosteric Regulation. *Blood* **2019**, *133*, 1899–1908. [[CrossRef](#)]

-
69. Takeda, S. ADAM and ADAMTS Family Proteins and Snake Venom Metalloproteinases: A Structural Overview. *Toxins* **2016**, *8*, 155. [[CrossRef](#)]
 70. de Groot, R.; Bardhan, A.; Ramroop, N.; Lane, D.A.; Crawley, J.T.B. Essential role of the disintegrin-like domain in ADAMTS13 function. *Blood* **2009**, *113*, 5609–5616. [[CrossRef](#)]
 71. Petri, A.; Kim, H.J.; Xu, Y.; de Groot, R.; Li, C.; Vandenbulcke, A.; Vanhoorelbeke, K.; Emsley, J.; Crawley, J.T.B. Crystal Structure and Substrate-Induced Activation of ADAMTS13. *Nat. Commun.* **2019**, *10*, 3781. [[CrossRef](#)]
 72. Belyaev, A.V. Catching Platelets from the Bloodflow: The Role of the Conformation of von Willebrand Factor. *Mat. Mod. Nat. Phenom.* **2018**, *13*, 44. [[CrossRef](#)]
 73. Jaspe, J.; Hagen, S.J. Do Protein Molecules Unfold in a Simple Shear Flow? *Biophys. J.* **2006**, *91*, 3415–3424. [[CrossRef](#)]
 74. Xiang, Y.; de Groot, R.; Crawley, J.T.B.; Lane, D.A. Mechanism of von Willebrand factor scissile bond cleavage by a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 (ADAMTS13). *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11602–11607. [[CrossRef](#)]
 75. Voevodin, V.; Antonov, A.; Nikitenko, D.; Shvets, P.; Sobolev, S.; Sidorov, I.; Stefanov, K.; Voevodin, V.; Zhumatiy, S. Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community. *Supercomput. Front. Innov.* **2019**, *6*, 4–11. [[CrossRef](#)]

Ce document a été mis en page en L^AT_EX grâce au modèle libre de droits classicthesis développé par André Miede. Le style en est inspiré par l'ouvrage de Robert Bringhurst sur la typographie, *"The Elements of Typographic Style"*.

<https://bitbucket.org/amiede/classicthesis/>

Version du 18 mars 2022



OLIVIER LANGUIN-CATTOËN
PLASTICITÉ DU SITE DE LIAISON
ET MÉCANISME GRAPPIN
DE L'ADHÉSINE FIMH

RÉSUMÉ

Les liaisons « accrocheuses » sont des liaisons non-covalentes qui voient leur durée de vie s'allonger lorsqu'elles sont soumises à des forces de tension. L'une des liaisons accrocheuses les mieux documentées est celle formée par l'adhésine bactérienne FimH de *E. coli*. FimH se lie spécifiquement aux dérivés de mannose trouvés en abondance sur certaines glycoprotéines des épithéliums humains et se révèle être un facteur de virulence pour certaines souches impliquées dans des infections urinaires.

Le mécanisme de liaison accrocheuse de FimH est celui d'une allostérie intramoléculaire négative, où l'application d'une force entraîne la séparation des deux domaines constituant la protéine, suivie d'une transition allostérique au sein du domaine comportant le site de liaison aux mannosides. Toutefois, de nombreux points d'ombre demeurent sur les mécanismes précis de cette régulation interdomaine et sur l'origine structurelle de la différence frappante d'affinité entre les deux états allostériques.

En nous appuyant sur une abondante littérature expérimentales, nous menons une étude fondée sur des simulations de dynamique moléculaire et des méthodes d'échantillonnage renforcé. Ces techniques nous permettent d'explorer le paysage conformationnel du domaine effecteur en présence et en l'absence de domaine régulateur, d'identifier de nouveaux états d'ouvertures du site de liaison et d'estimer leurs populations relatives. Nous mettons en évidence l'importance de la dynamique d'ouverture du site dans les affinités apparentes mesurées expérimentalement, l'influence locale de l'application d'une force sur la dynamique d'ouverture, ainsi qu'un mécanisme de dissociation en deux étapes nécessaire pour rendre compte des constantes de vitesses expérimentales. Dans l'ensemble, nos résultats dressent un portrait complexe et multi-échelle du grappin moléculaire formé par FimH et soulignent le rôle crucial des études simulatoires pour compléter l'interprétation des données expérimentales, tout en questionnant leurs limitations.

MOTS CLÉS

liaison accrocheuse, adhésine, UPEC, dynamique moléculaire, échantillonnage renforcé, dynamique moléculaire dirigée

ABSTRACT

Catch bonds are non-covalent bonds that increase their lifetime when subjected to tensile forces. One of the most documented examples of a catch bond protein is the bacterial adhesin FimH from *E. coli*. FimH specifically binds to mannose derivatives found in abundance on glycoproteins of human epithelia, and is a virulence factor for some *E. coli* strains involved in urinary tract infections.

The catch-bond mechanism of FimH relies upon negative intramolecular allostery, where the application of a force results in the separation of the two domains constituting the protein, followed by an allosteric transition within the binding site domain. However, many questions remain about the precise mechanism of the interdomain regulation and the structural origin of the allosteric states' striking difference in affinity.

Drawing on the extensive experimental literature, we conduct a study based on molecular dynamics simulations and enhanced sampling methods. These techniques allow us to explore the conformational landscape of the effector domain in the presence and absence of the regulatory domain. We identify new opening states of the binding site and estimate their relative abundance. We shed light on the importance of site opening dynamics in interpreting the experimental affinities. We show a local influence of force on the dynamics, and propose a two-step mechanism that may be necessary to account for the experimental rate constants. Overall, our results display a complex and multiscale picture of FimH catch-bond mechanism and highlight the role of computational studies to interpret experimental data, while questioning their limitations.

KEYWORDS

catch bond, adhesin, UPEC, molecular dynamics, enhanced sampling, steered molecular dynamics