

Imloul Nesrine

BIF-4000/7000 Modélisation Biomoléculaire

Devoir #2

Rapport présenté à

Patrick Lague

Dans le cadre du cours

BIF-4000/7000 Modélisation Biomoléculaire

BIF-4000

13-03-2025



1. Construction du système

A) Construction du Système

La première étape de cette étude consiste à préparer un système de simulation de dynamique moléculaire pour la protéine assignée. Comme indiqué dans les instructions, nous avons utilisé CHARMM-GUI pour générer les fichiers nécessaires à la simulation.

La construction du système repose sur plusieurs critères essentiels :

- **Absence de ligand** : Conformément aux consignes, le ligand présent dans la structure PDB originale a été retiré pour cette simulation.
- **Immersion dans une boîte d'eau** : La protéine a été placée dans une boîte d'eau dont les dimensions dépassent de **10 Å** celles de la protéine afin d'assurer une solvatation adéquate.
- **Ajout d'ions NaCl** : Une concentration de **0.150 M de NaCl** a été ajoutée pour reproduire un environnement physiologique.

L'ensemble des fichiers de préparation a été généré et exporté depuis **CHARMM-GUI**, et les fichiers essentiels, tels que step3_input.psf et step3_input.pdb, ont été obtenus pour la suite des simulations.

B) Paramètres du Système

Le système de simulation a été construit en respectant les exigences du devoir et en suivant les étapes décrites dans CHARMM-GUI. L'objectif était d'obtenir un environnement réaliste pour la simulation de dynamique moléculaire de la protéine, incluant une solvatation correcte et une concentration physiologique en ions.

Les paramètres du système sont les suivants :

- **Nombre total d'atomes** : **61752**
- **Nombre de molécules d'eau** : **19440**
- **Nombre d'ions** : **202**
- **Type de boîte** : **Cubique**
- **Dimensions de la boîte** :
 - **a=87 Å**
 - **b=87 Å**

- $c=87 \text{ \AA}$
- **Angles :**
 - $\alpha=90.0^\circ$
 - $\beta=90.0^\circ$
 - $\gamma=90.0^\circ$

Ces valeurs ont été extraites directement des fichiers **step3_input.psf** et **step3_input.str**, générés lors de la préparation du système dans **CHARMM-GUI**.

L'ajout de **19440 molécules d'eau** assure une solvatation adéquate de la protéine dans l'environnement de simulation, tandis que la présence de **202 ions** permet d'atteindre une concentration en **NaCl de 0.150 M**, ce qui correspond à des conditions physiologiques.

Enfin, la boîte cubique de **87 Å de côté** garantit un espace suffisant pour permettre les mouvements de la protéine tout en minimisant les effets de bord lors de la simulation de dynamique moléculaire.

C) Informations sur les Histidines et Résidus Chargés

L'analyse de l'état de protonation des résidus histidine (His), glutamate (Glu) et aspartate (Asp) a été réalisée pour la protéine 3QPS à l'aide du serveur PDB2PQR. Ce processus est essentiel en bio-informatique et en modélisation moléculaire, car il permet de prédire les charges des acides aminés en fonction du pH du milieu, ce qui influence la structure tridimensionnelle et la dynamique de la protéine.

Pour cette analyse, le fichier 3QPS.pqr généré par PDB2PQR a été utilisé, car il inclut les informations de protonation des acides aminés, bien que la position des chaînes latérales ne soit pas optimisée. Les valeurs de pKa de chaque résidu ont été calculées par le serveur, ce qui permet d'attribuer un état de protonation basé sur les conditions physiologiques de la simulation.

L'état de protonation des résidus His, Glu et Asp peut être vérifié visuellement à l'aide du logiciel PyMOL, en affichant les atomes d'hydrogène sur les acides aminés titrables. En observant les structures des résidus His, on remarque que la plupart adoptent la forme HSD, c'est-à-dire que l'atome d'hydrogène est porté par l'azote delta (δ) du cycle imidazole. Quelques histidines sont sous la forme HSE, où l'hydrogène est porté par l'azote epsilon (ϵ), mais aucun résidu n'est sous forme HSP (forme complètement protonée).

Pour les acides aminés Glu et Asp, un seul état de protonation est possible à un pH neutre. Contrairement à l'histidine, qui peut exister sous trois formes différentes (HSD,

HSE et HSP), les résidus Glu et Asp ont des pKa inférieurs à 4.5, ce qui signifie qu'ils sont majoritairement déprotonés à pH 7. Cette observation est confirmée par l'analyse de leur structure dans PyMOL, qui montre l'absence d'atomes d'hydrogène sur leurs chaînes latérales, confirmant leur état chargé négativement.

Le tableau ci-dessous résume l'état de protonation des résidus His, Glu et Asp dans la protéine 3QPS, en fonction des valeurs de pKa prédictes :

Tableau 1 : l'état de protonation des résidus de la protéine CmeR.

Résidu	Chaîne	pKa	Model-pKa	État attribué
HIS 72	A	6.80	6.50	HSE
HIS 74	A	5.10	6.50	HSD
HIS 125	A	5.40	6.50	HSD
HIS 174	A	7.10	6.50	HSE
HIS 175	A	4.90	6.50	HSD
HIS 193	A	6.85	6.50	HSE
ASP 38	A	3.20	4.00	Déprotoné (-)
ASP 52	A	3.50	4.00	Déprotoné (-)
ASP 66	A	5.20	4.00	Protoné (0)
ASP 67	A	3.10	4.00	Déprotoné (-)
ASP 121	A	4.80	4.00	Protoné (0)
ASP 123	A	3.60	4.00	Déprotoné (-)
GLU 16	A	4.10	4.40	Déprotoné (-)
GLU 24	A	4.30	4.40	Déprotoné (-)
GLU 33	A	4.20	4.40	Déprotoné (-)
GLU 58	A	4.00	4.40	Déprotoné (-)
GLU 63	A	5.50	4.40	Protoné (0)

GLU 82	A	3.80	4.40	Déprotoné (-)
GLU 90	A	3.90	4.40	Déprotoné (-)
GLU 101	A	4.05	4.40	Déprotoné (-)
GLU 107	A	4.15	4.40	Déprotoné (-)
GLU 131	A	4.25	4.40	Déprotoné (-)
GLU 159	A	4.50	4.40	Déprotoné (-)
GLU 171	A	4.60	4.40	Déprotoné (-)
GLU 189	A	3.85	4.40	Déprotoné (-)
GLU 192	A	5.00	4.40	Protoné (0)
GLU 195	A	3.75	4.40	Déprotoné (-)

E) Visualisation du système

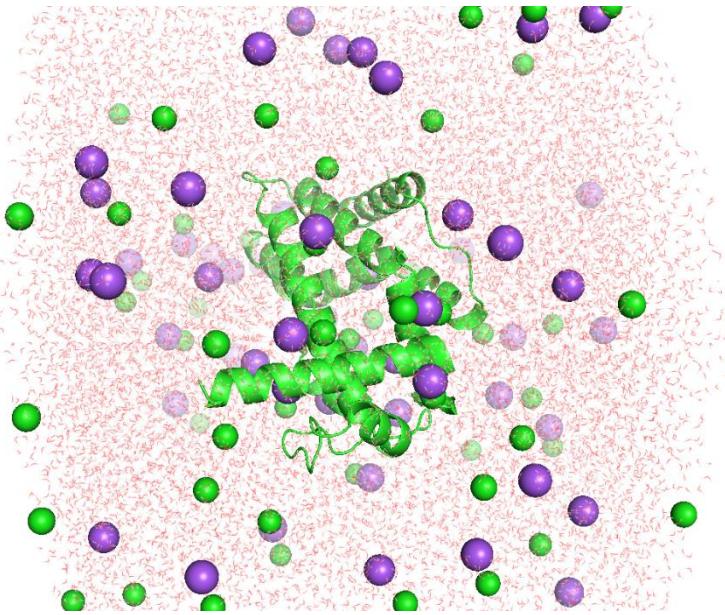


Figure1 : vue globale du système construit, avec les molécules d'eau visibles.

2. Minimisation, équilibration et production de la trajectoire

Les fichiers générés par le site www.charmm-gui.org ont été copiés sur le serveur de calcul Québec. Le paramètre de la fréquence de sauvegarde de coordonnées de la trajectoire (dcdfreq) a été modifié, en modifiant le paramètre de valeur par défaut de 50000 par la valeur 5000, dans le fichier step5_production.inp. L'ordonnancement des tâches sur les grappes de Calcul Canada se fait à l'aide du logiciel Slurm Workload Manager. Un fichier de commandes bash nommé « namd.sh » a été utilisé pour la réalisation de la simulation. Ce script orchestre le processus de simulation, accomplissant plusieurs tâches critiques: il vérifie l'avancement actuel de la simulation, renomme les fichiers de manière séquentielle pour assurer un suivi ordonné, évalue si la simulation se déroule comme prévu et, en cas de problème, l'interrompt. De plus, il ajuste les paramètres de simulation dans le fichier modèle step5_production.inp en fonction des besoins.

3. Analyse: Extraction et analyse de quantités physiques.

Les graphiques montrent l'évolution de différentes propriétés physiques d'un système au cours des étapes d'intégration d'une simulation réalisée avec CHARMM (Chemistry at HARvard Macromolecular Mechanics), un programme de simulation biomoléculaire.

A. Volume

L'évolution du volume montre une stabilisation après une forte fluctuation initiale. Au début de la simulation, une chute brutale du volume est observée, ce qui correspond probablement à l'ajustement initial du système à son environnement simulé. Une fois ce pic passé, le volume oscille autour d'une valeur stable, indiquant que le système atteint une phase d'équilibre.

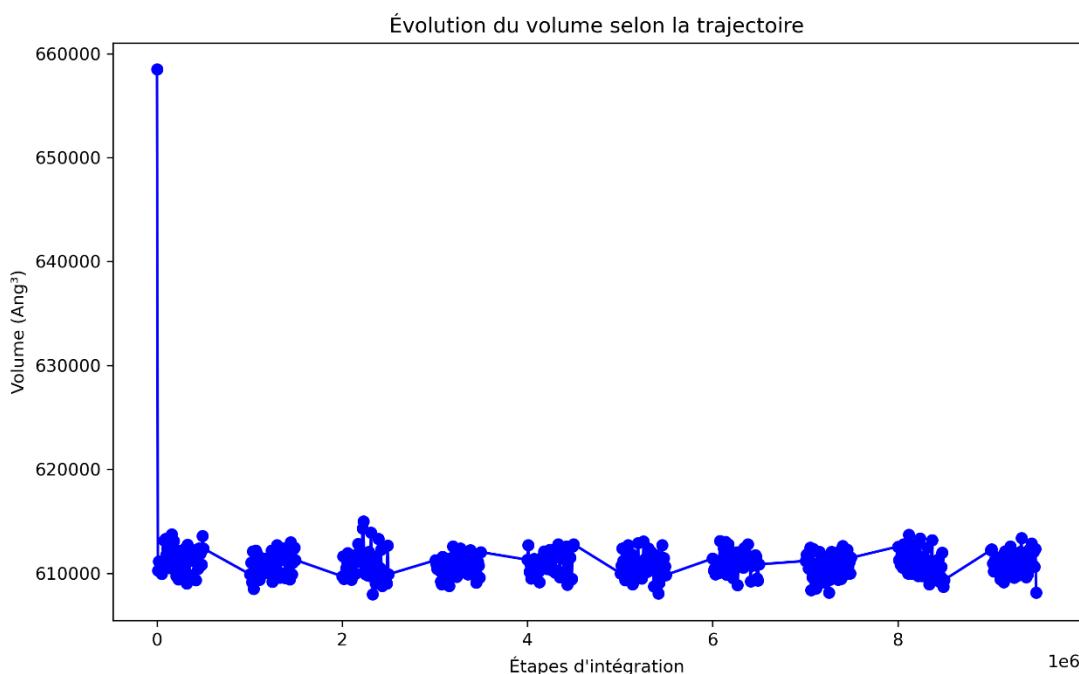


Figure 2 : Graphique de l'évolution du volume du système en fonction de la trajectoire.
Le volume se stabilise après une courte durée.

B. Température :

La température oscille entre **299 K et 305 K**, ce qui est cohérent avec une simulation en conditions physiologiques (~300 K). Les variations sont normales et témoignent de la dynamique thermique du système. Aucune tendance significative ne montre une dérive, ce qui suggère une bonne régulation thermique du système.

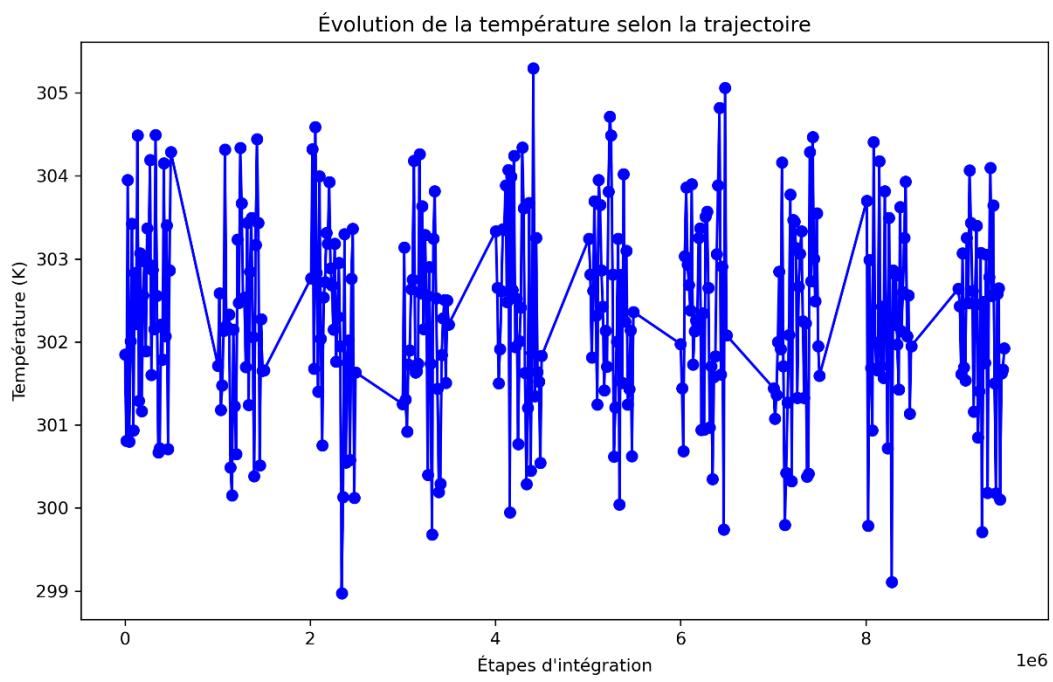


Figure 3 : Graphique de l'évolution et de la variation de la température en fonction de la trajectoire.

C. Energie totale

L'énergie totale suit une tendance similaire au volume, avec une baisse initiale rapide suivie d'une stabilisation. Cette évolution est typique d'une phase d'équilibration, où le système se réorganise avant d'atteindre un état stable. Une fois cette phase passée, l'énergie fluctue autour d'une valeur moyenne, ce qui confirme que le système ne présente pas de variations anormales.

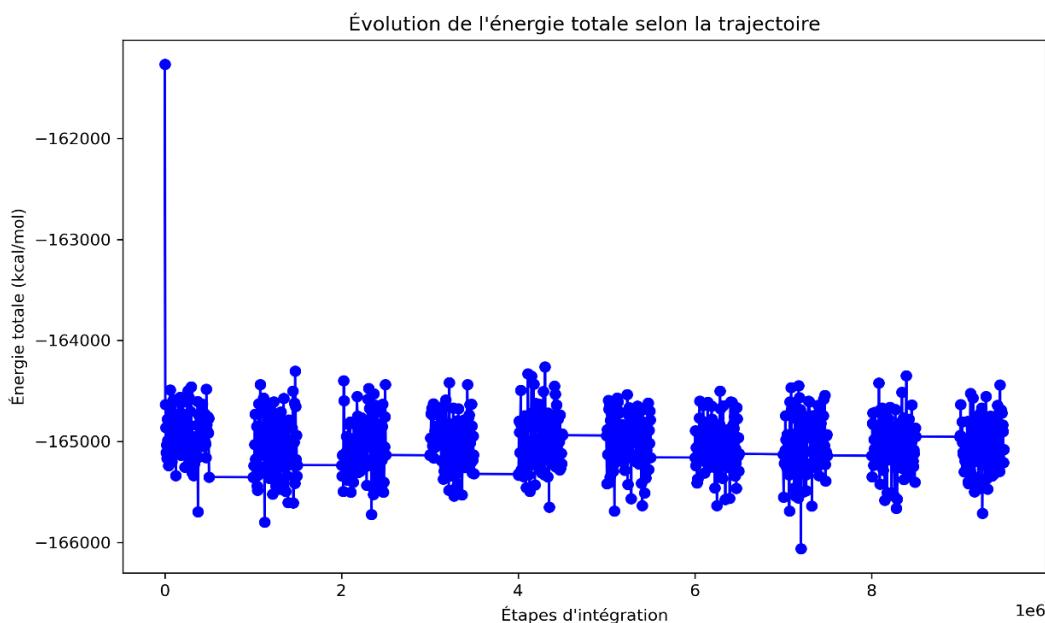


Figure 4: Graphique de l'évolution et de la variation de l'énergie totale en fonction de la trajectoire.

D. RMSD total

L'analyse du **Root Mean Square Deviation (RMSD)** montre une augmentation progressive, indiquant que la structure de la protéine évolue au cours du temps. Après environ **60 unités de temps**, la courbe atteint un plateau, suggérant que la conformation globale de la protéine se stabilise. Une telle évolution est attendue dans une simulation de dynamique moléculaire, car les protéines explorent leur espace conformationnel avant d'atteindre une configuration stable.

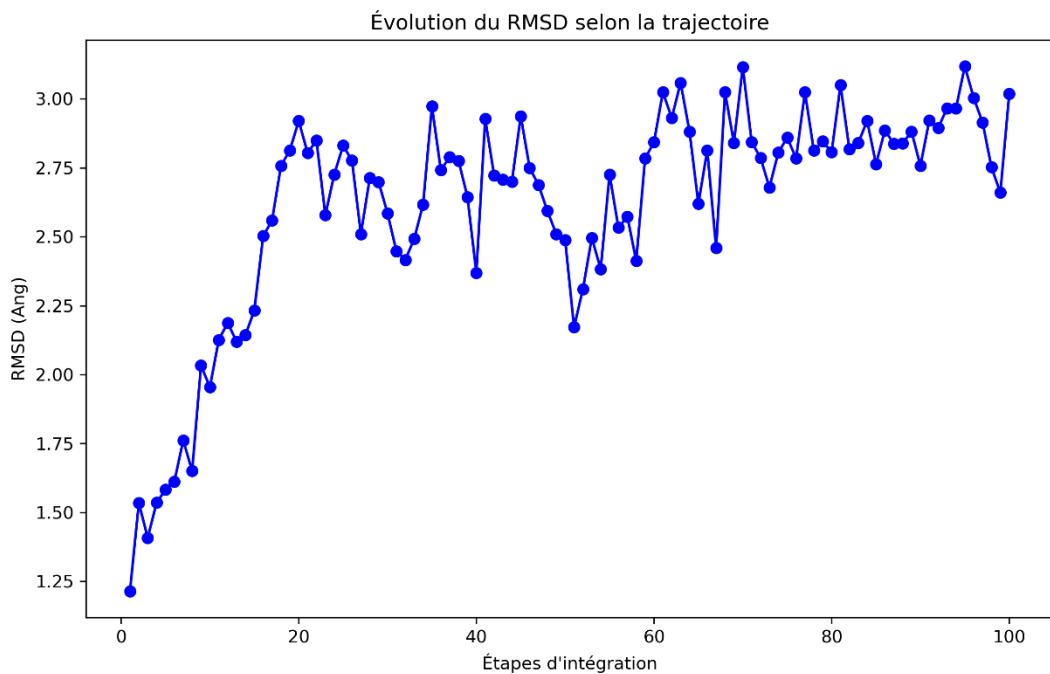


Figure 5: Graphique de l'évolution de RMSD en fonction de la trajectoire.

Le système atteint l'équilibre après environ **60 unités de temps**, comme le montrent l'évolution de la courbe RMSD et celle de l'énergie totale. Après cette phase d'équilibration, les fluctuations observées restent dans des limites normales, indiquant ainsi une simulation bien contrôlée. De manière générale, les résultats montrent une phase initiale de réajustement, suivie d'une stabilisation progressive. Cela confirme que le système a atteint un état d'équilibre thermique et structurel.

4. Analyse: calcul de certains angles dièdres en fonction du temps

Dans cette section, nous analysons l'évolution de l'angle dièdre **x1** (défini par les atomes **N**, **CA**, **CB** et **CG**) pour trois acides aminés sélectionnés en fonction de leur environnement dans la protéine.

A) Acide aminé situé au site de liaison du ligand : Leu 65

L'angle dièdre de **Leu 65**, qui interagit avec le ligand via un **lien hydrogène**, présente deux conformations distinctes : une majorité du temps à **environ 160°**, avec une transition brusque vers **-150°** entre 50 et 80 étapes d'intégration. Cette variabilité suggère une flexibilité importante dans le site actif, probablement liée aux interactions avec le ligand.

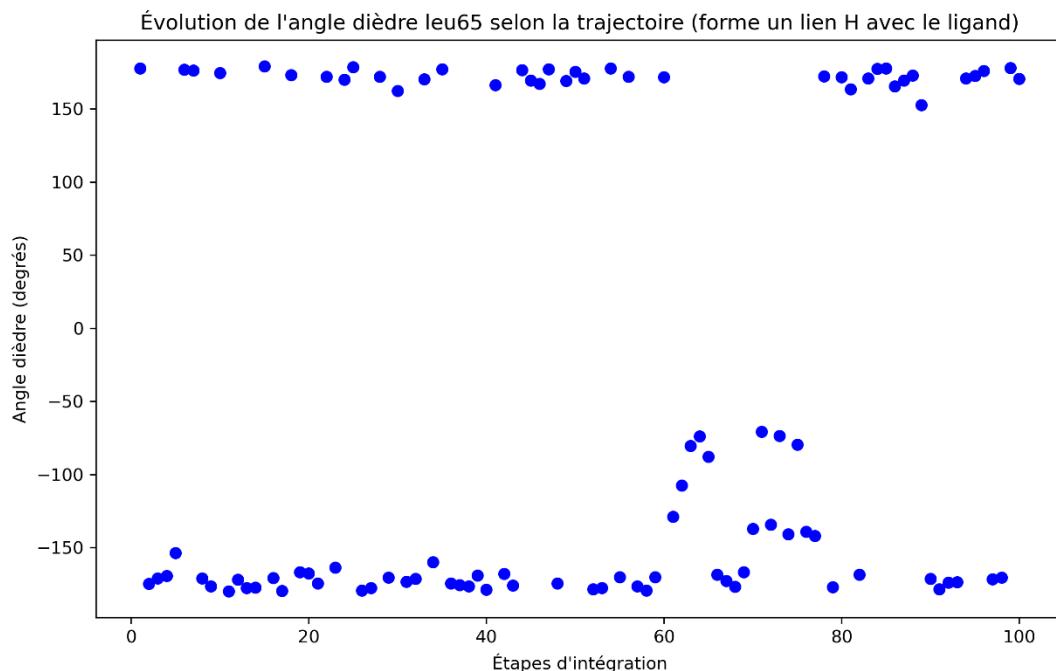


Figure 6: Évolution de l'angle dièdre de la Leu65 en fonction de la trajectoire. La leucine est en contact avec le ligand.

B) Acide aminé enfoui dans la protéine, moins solvaté : Ile 68

L'angle dièdre de **Ile 68**, enfoui dans le cœur hydrophobe de la protéine, oscille autour de **-70° à -90°** avec peu de variations significatives. Cette stabilité reflète l'environnement restreint de l'isoleucine, dont la chaîne latérale est contrainte par des interactions hydrophobes avec les autres résidus du noyau protéique

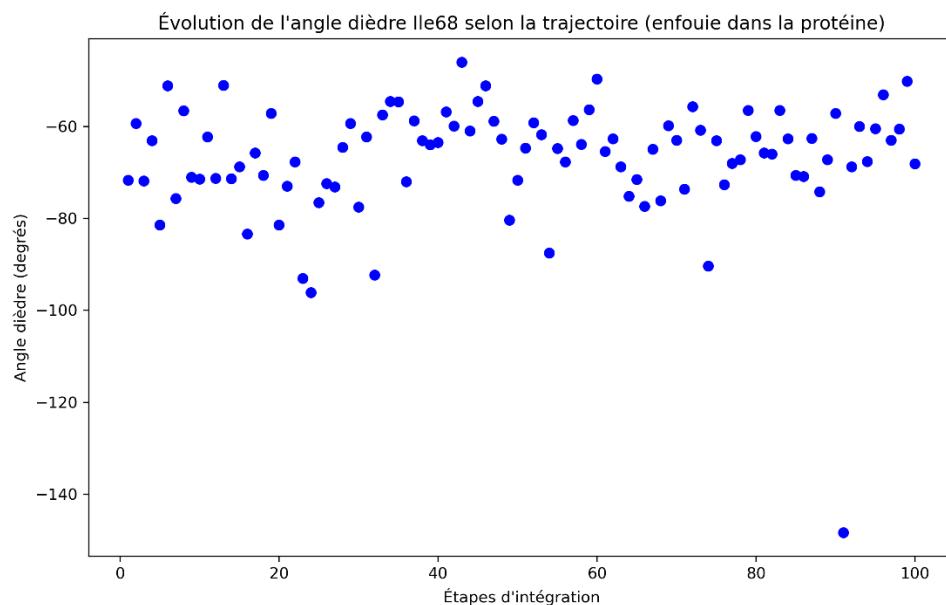


Figure 9 : Évolution de l'angle dièdre de la Ile68 en fonction de la trajectoire. La Ile est enfouie à l'intérieur de la protéine.

C) Acide aminé en surface, bien solvaté : Ser 207

L'angle dièdre de **Ser 207**, exposé au solvant, varie de **40° à 80°** sans adopter de conformation préférentielle. Cette dispersion indique que la chaîne latérale de la sérine est en mouvement constant, ce qui est attendu pour un résidu exposé au solvant et libre de réorienter sa chaîne latérale.

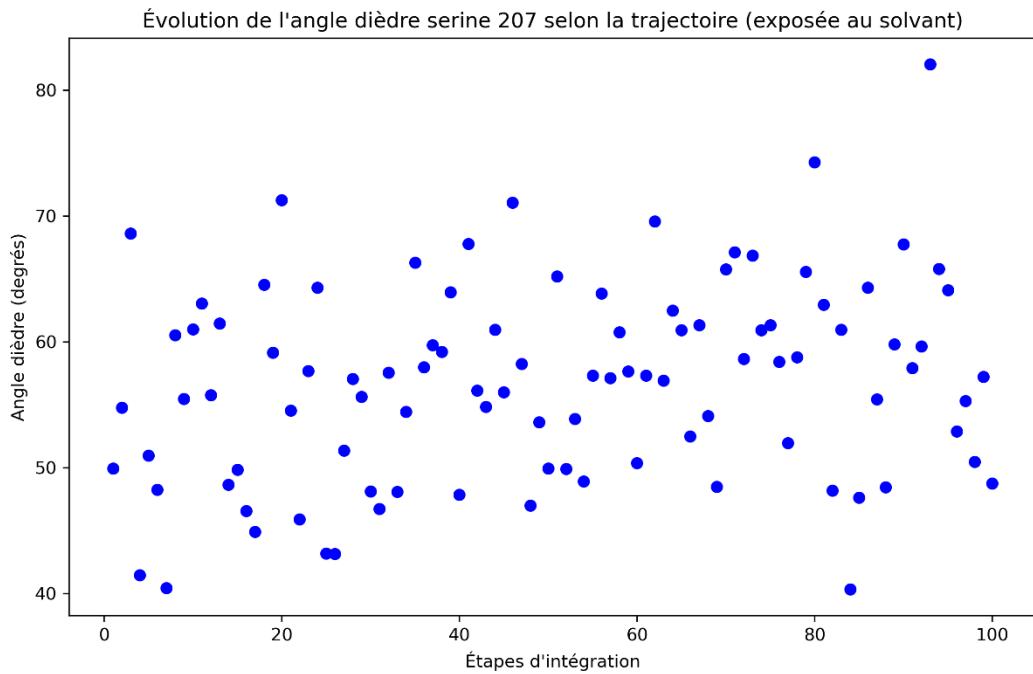


Figure 8: Évolution de l'angle dièdre de Ser 207 en fonction de la trajectoire. La serine est exposée en surface en contact avec le solvant.

Leu 65 situé dans le site actif, subit des changements conformationnels marqués, ce qui reflète son adaptation pour interagir avec le ligand. Ser 207, exposé au solvant en surface, présente une grande variabilité de son angle dièdre, témoignant de sa flexibilité. En revanche, Ile 68, enfoui dans la structure de la protéine, reste stable tout au long de la simulation, ce qui est typique des résidus hydrophobes profondément intégrés. Ces observations illustrent l'influence de l'environnement local sur la dynamique des résidus : ceux exposés au solvant sont plus mobiles, tandis que ceux enfouis présentent des mouvements plus restreints.

5. Analyse de la fluctuation des résidus

L'analyse du **Root Mean Square Fluctuation (RMSF)** permet d'évaluer la mobilité de chaque résidu tout au long de la simulation de dynamique moléculaire. Le graphique obtenu met en évidence plusieurs observations importantes concernant la stabilité et la flexibilité des différentes régions de la protéine.

Tout d'abord, les **extrémités de la protéine** présentent les **fluctuations les plus élevées**. Ce comportement est attendu, car les régions terminales d'une protéine sont souvent moins stabilisées par des interactions intra-moléculaires et ont une plus grande liberté de mouvement. Ces fluctuations importantes peuvent être liées à l'absence d'interactions fortes avec le cœur de la protéine ou à un enroulement partiel qui se défait et se reforme au cours de la simulation.

Ensuite, plusieurs pics de fluctuations internes apparaissent autour des **résidus 10, 50, 120 et 200**, indiquant des segments de la protéine ayant une dynamique plus marquée. Une analyse plus approfondie, basée sur le fichier **comparaison_structures_secondaires.txt**, permet d'identifier la structure secondaire associée à ces résidus :

- **Le résidu 10** est situé dans une **hélice α (H)** et présente une fluctuation notable, bien que généralement, les hélices α soient relativement rigides. Cette variation peut être attribuée à une réorganisation locale ou à une interaction avec un autre élément de la structure protéique.
- **Le résidu 50** se trouve dans une **boucle (L)**, ce qui explique sa forte fluctuation. Les boucles sont généralement **plus mobiles** que les autres structures secondaires, car elles ne sont pas stabilisées par des liaisons hydrogène régulières comme les hélices α ou les feuillets β.
- **Le résidu 120**, également situé dans une **hélice α**, montre des variations modérées. Cela pourrait être dû à une flexibilité locale ou à une interaction transitoire avec un ligand ou une autre partie de la protéine.
- **Le résidu 200**, situé dans une **boucle**, présente une très forte fluctuation, ce qui est cohérent avec la nature flexible de ces régions.

Ces résultats illustrent bien l'influence de la structure secondaire sur la dynamique des résidus. En général, les **hélices α** sont plus rigides et présentent moins de fluctuations, car elles sont stabilisées par des liaisons hydrogène internes et des interactions hydrophobes. En revanche, les **boucles** sont plus souples et peuvent adopter plusieurs conformations au cours de la simulation.

Enfin, la forte fluctuation des extrémités de la protéine confirme leur **caractère flexible** en l'absence de contraintes stabilisantes. Cette flexibilité peut jouer un rôle fonctionnel important, par exemple dans l'adaptation conformationnelle de la protéine ou dans des interactions avec d'autres molécules.

L'analyse du RMSF met en évidence des zones de haute mobilité dans la protéine, principalement situées aux extrémités et dans les boucles. Ces résultats montrent que la dynamique moléculaire est fortement influencée par l'environnement structural des

résidus : les boucles et les extrémités sont plus mobiles, tandis que les hélices α conservent une certaine rigidité. Cette information est essentielle pour comprendre les mouvements de la protéine et son comportement fonctionnel dans un environnement dynamique.

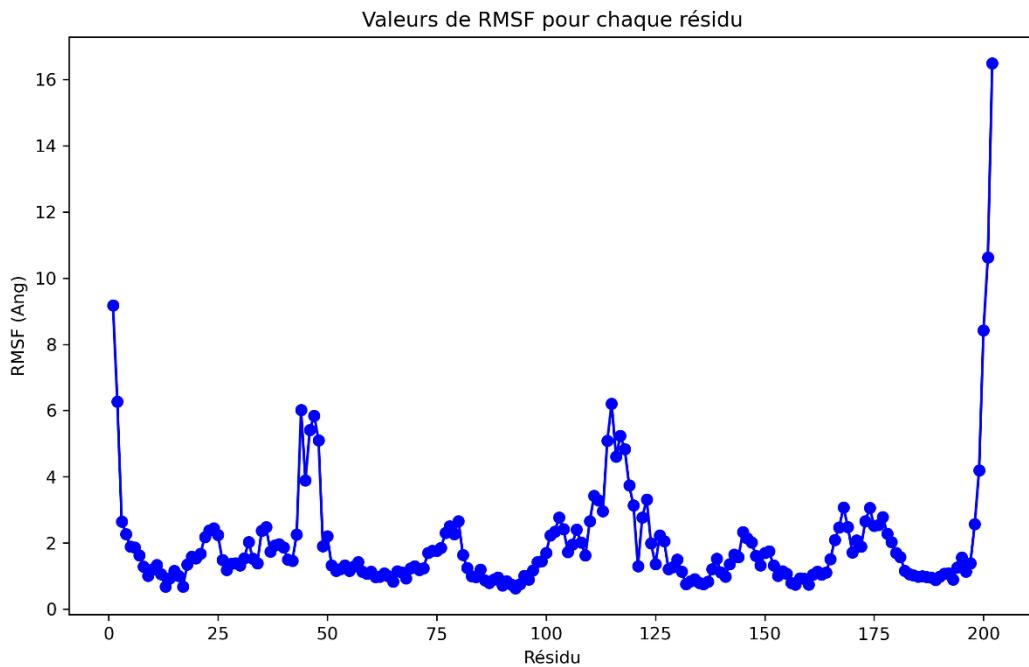


Figure 9: Fluctuation des résidus de la protéine au cours de la simulation (RMSF)

6. Analyse: comparaison de la structure secondaire finale avec celle initiale

L'analyse de la structure secondaire a pour but d'évaluer les changements entre l'état initial et l'état final de la protéine après la simulation de dynamique moléculaire. Cette comparaison permet de déterminer si certaines régions de la protéine ont conservé leur structure, si elles se sont stabilisées ou si elles ont subi des modifications significatives. Pour ce faire, nous avons extrait la structure secondaire de chaque résidu avant et après la simulation à l'aide des fichiers finale.pdb et comparison_structures_secondaires.txt.

L'analyse a révélé plusieurs changements notables dans la structure de la protéine. Certains résidus qui formaient une hélice dans la structure initiale ont perdu cette organisation et sont devenus des boucles. À l'inverse, d'autres résidus, qui faisaient partie de régions flexibles en boucle, ont adopté une conformation en hélice après la

simulation. Ces modifications peuvent être interprétées comme des réarrangements structuraux liés aux interactions intra-moléculaires et aux interactions avec le solvant.

Plusieurs résidus ont subi une transition d'une hélice α vers une boucle. C'est notamment le cas des résidus 7, 8, 31, 32, 33 et 81. Cette perte de structure hélicoïdale peut s'expliquer par une diminution des interactions hydrogène qui stabilisaient initialement ces hélices. Lorsque certaines parties de la protéine deviennent plus exposées au solvant, elles peuvent perdre leur organisation structurale rigide et adopter une configuration plus flexible, ce qui est souvent observé dans les simulations de dynamique moléculaire. Cela signifie que ces régions de la protéine sont plus susceptibles de subir des mouvements conformationnels au cours du temps.

À l'inverse, d'autres résidus ont subi une transition inverse, passant d'une boucle à une hélice. C'est le cas des résidus 45, 46, 47 et 48. Ce phénomène peut être attribué à la formation de nouvelles interactions stabilisatrices, comme des liaisons hydrogène ou des interactions hydrophobes, qui favorisent l'adoption d'une structure plus organisée. Une telle transition peut indiquer que la protéine a trouvé une conformation plus stable dans l'environnement simulé. Ce type de modification peut être particulièrement important si ces résidus sont situés dans des régions fonctionnelles de la protéine, influençant ainsi ses propriétés structurales et dynamiques.

L'emplacement de ces changements est également un élément clé de l'analyse. Les résidus qui ont perdu leur structure hélicoïdale semblent se trouver principalement dans des régions exposées au solvant, ce qui explique leur plus grande flexibilité. À l'inverse, les résidus qui ont acquis une structure en hélice pourraient être situés dans des zones plus internes de la protéine, où la formation de nouvelles interactions stabilisatrices a pu favoriser leur rigidification. Cette réorganisation structurale pourrait influencer les interactions de la protéine avec d'autres molécules, en modifiant la conformation des sites de liaison ou en affectant sa stabilité globale.

D'un point de vue fonctionnel, ces changements peuvent avoir des implications importantes. Une augmentation de la flexibilité dans certaines régions peut être associée à une adaptation aux conditions simulées, permettant à la protéine de s'ajuster à son environnement. À l'inverse, l'apparition de nouvelles hélices pourrait indiquer une tendance à stabiliser certaines parties de la structure, possiblement en réponse à des contraintes thermodynamiques. Ces observations sont cohérentes avec ce que l'on attend des simulations de dynamique moléculaire, où la protéine explore différentes conformations jusqu'à atteindre un équilibre structurel.

Pour visualiser ces changements, une représentation en cartoon de la protéine finale a été réalisée à l'aide du logiciel PyMOL (Figure 8). La Figure 7 illustre la superposition des structures initiale et finale, mettant en évidence les régions où des modifications de la structure secondaire ont eu lieu. Cette visualisation permet d'identifier clairement les

résidus affectés et d'analyser les variations structurales résultant de la dynamique moléculaire.

En conclusion, l'analyse de la structure secondaire a révélé des réarrangements significatifs dans certaines régions de la protéine. La perte de certaines hélices et la formation de nouvelles structures hélicoïdales montrent que la protéine a subi des ajustements structurels durant la simulation. Ces modifications sont cohérentes avec les processus de stabilisation et de flexibilité qui se produisent lors des simulations de dynamique moléculaire, soulignant ainsi l'importance de ces approches pour comprendre l'adaptabilité et la stabilité des protéines.

Tableau 2 : différence entre la structure secondaire finale et initiale de la protéine 3QPS

Résidu	Structure initiale	Structure finale
7	H	L
8	H	L
31	H	L
32	H	L
33	H	L
45	L	H
46	L	H
47	L	H
48	L	H
81	H	L

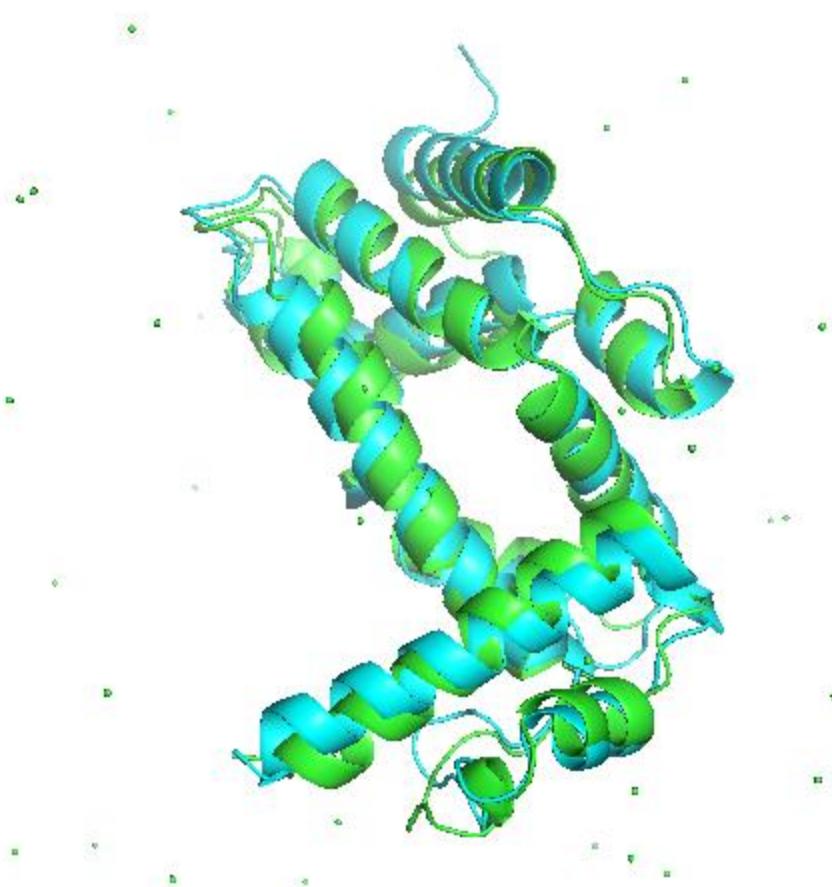


Figure 10 : Superposition des structures initiale (cyan) et finale (vert) de la protéine après la dynamique moléculaire.



Figure 11 : Représentation en cartoon de la structure finale de la protéine après la dynamique moléculaire.