

IMLOUL Nesrine

BIF-4000/7000 Modélisation Biomoléculaire

Devoir #3

Rapport présenté à

Patrick Lague

Dans le cadre du cours

BIF-4000/7000 Modélisation Biomoléculaire

BIF-4000

01-04-2025

©IMLOUL Nesrine, 2025-04-01

1. Introduction

Dans le cadre de ce devoir, deux complexes protéine-ligand ont été étudiés à l'aide de simulations de dynamique moléculaire. Les ligands utilisés sont ceux du devoir #1, portant les codes PDB **3QPS** (ligand 1) et **3QQA** (ligand 2), se liant à la protéine **CmeR**. L'objectif principal est d'analyser la stabilité et les interactions de chacun de ces ligands avec la protéine cible, à travers des simulations en **solvant explicite** et **solvant implicite (GBIS)**. Pour cela, les systèmes ont été construits avec CHARMM-GUI, les ligands paramétrisés avec CGenFF, puis les trajectoires ont été analysées selon plusieurs critères : RMSD, angles dièdres, interactions hydrogène et température d'éjection. Ce rapport présente la méthodologie utilisée et les résultats obtenus pour chacun des deux complexes.

2. Construction des systèmes

Dans le cadre de cette étude, nous avons construit deux systèmes protéine-ligand en solvant explicite, en utilisant le portail [CHARMM-GUI](#). Les complexes étudiés impliquent la protéine **CmeR** liée aux ligands **3QPS** et **3QQA**.

La protéine CmeR est une enzyme monocaténaire composée de 166 résidus d'acides aminés. Lors de la construction des systèmes, la chaîne complète de la protéine a été utilisée. Des mutations ont été introduites pour refléter l'état de protonation approprié des résidus, conformément aux conditions physiologiques. Les ligands ont été intégrés dans le site actif de CmeR en fonction des structures cristallographiques disponibles.

Les systèmes ont été solvés dans une boîte d'eau cubique avec une marge de 10 Å autour du complexe, et des ions sodium (Na^+) et chlorure (Cl^-) ont été ajoutés pour neutraliser la charge globale du système et reproduire les conditions ioniques physiologiques. Les paramètres détaillés de chaque système sont présentés dans le tableau 1.

Tableau 1 : Paramètres des systèmes pour la simulation des complexes CmeR-3QPS et CmeR-3QQA

Paramètre	3QPS	3QQA
Nombre total d'atomes	61535	61535
Nombre d'ions	Na :55; Cl : 85	Na :55; Cl : 85
Dimensions de la boîte	10 Å autour du complexe	10 Å autour du complexe

3. Minimisation, équilibration et production de la trajectoire pour la simulation en solvant explicite

Les fichiers générés par le site www.charmm-gui.org pour les complexes protéine-ligand avec le ligand CHD et la protéine CmeR ont été transférés sur le serveur de calcul Québec afin de réaliser les simulations en solvant explicite. Le paramètre relatif à la fréquence de sauvegarde des coordonnées de la trajectoire (dcdfreq) a été modifié dans le fichier step5_production.inp, passant de la valeur par défaut de 50000 à 5000, afin d'obtenir une meilleure résolution temporelle des données issues de la simulation.

L'ordonnancement des tâches sur les grappes de Calcul Canada a été effectué avec le logiciel Slurm Workload Manager. Un fichier de commandes bash nommé « namd.sh » a été utilisé pour orchestrer le processus de simulation. Ce script effectue plusieurs tâches essentielles : il vérifie l'avancement actuel de la simulation, renomme les fichiers de manière séquentielle pour un suivi clair, évalue si la simulation se déroule correctement et, en cas de problème, l'interrompt automatiquement. De plus, il ajuste les paramètres spécifiques du fichier modèle step5_production.inp selon les besoins précis de chaque simulation.

4. Paramétrisation des ligands

4.1 Description des ligands CHD et TCH

Lors du premier devoir, la protéine CmeR portant le code PDB 3QPS a été assignée pour ce projet. Cette protéine forme un complexe avec le ligand CHD. Après une recherche approfondie dans les bases de données, un second ligand, TCH, formant un complexe similaire avec la même protéine CmeR (code PDB 3QQA), a été identifié. Le choix de ce deuxième ligand, TCH, est justifié par sa meilleure affinité ($K_d = 1,5 \mu\text{M}$) pour la protéine CmeR, comparativement au ligand CHD qui présente une affinité légèrement inférieure ($K_d = 2,5 \mu\text{M}$).

Le ligand CHD est un composé naturel, chimiquement caractérisé par un noyau stéroïde cholane, comportant trois groupements hydroxyles en positions 3α , 7α et 12α et un groupe carboxyle terminal. Ce ligand interagit principalement avec des zones hydrophobes dans le site actif de la protéine.

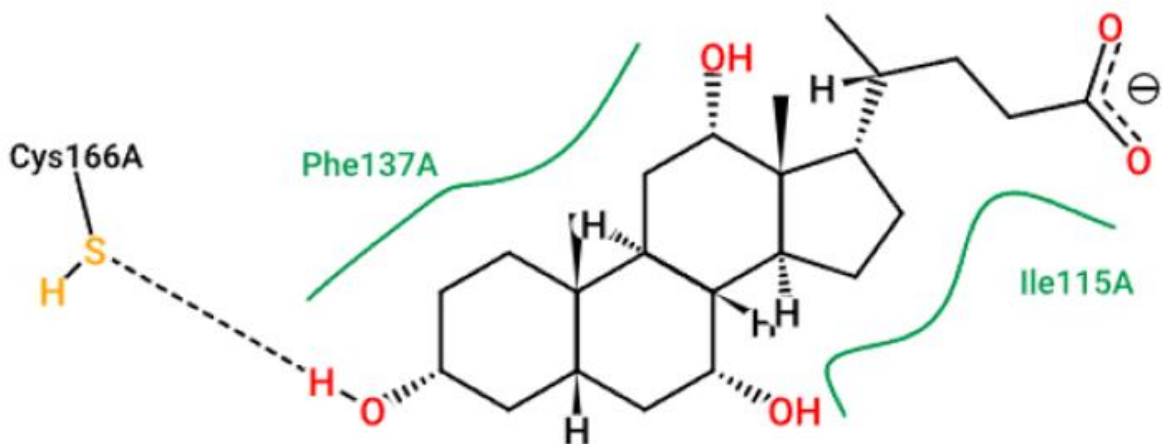


Figure 1: site de liaison du ligand CHD sur la protéine CmeR

Le ligand TCH est un dérivé conjugué du ligand CHD, obtenu par la liaison de l'acide cholique avec la taurine via une liaison amide. Cette liaison introduit un groupe sulfonique augmentant la polarité du composé. Cette différence chimique confère au ligand TCH des propriétés distinctes d'interaction avec la protéine CmeR, lui permettant d'établir des interactions supplémentaires, telles que des liaisons hydrogène et des interactions électrostatiques.

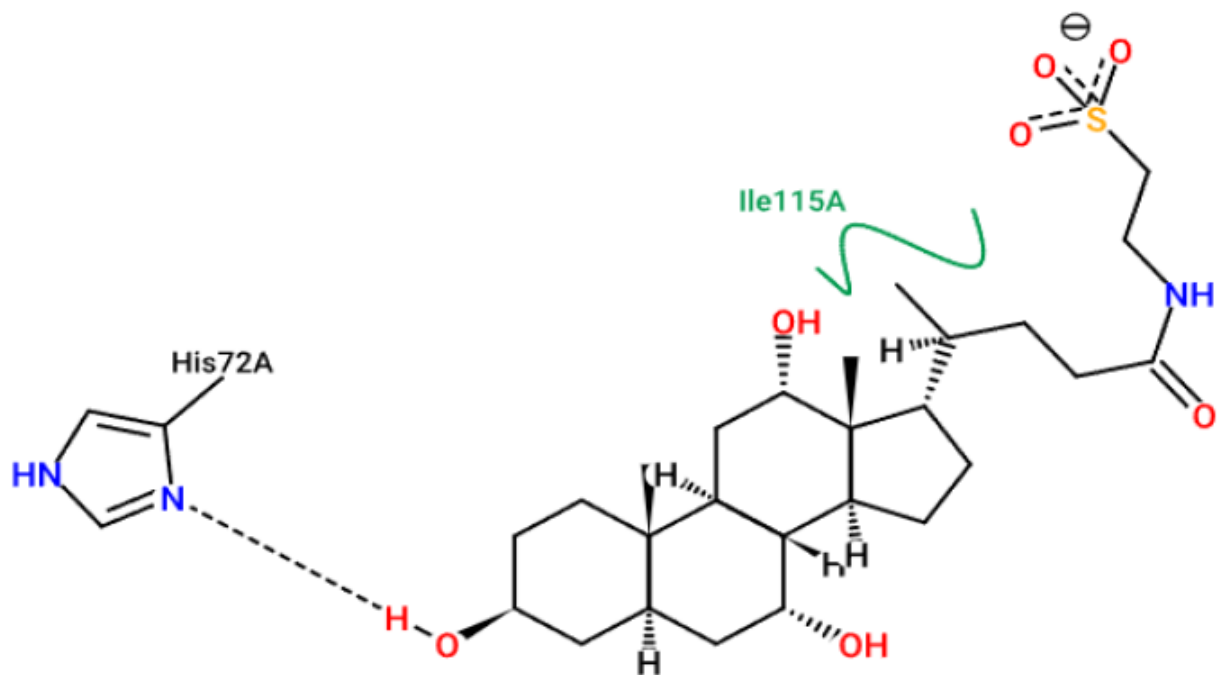


Figure 2 : site de liaison du ligand TCH sur la protéine CmeR.

4.2 Discussion du niveau global de confiance dans les paramètres obtenus à l'aide de CGenFF pour les deux ligands

Afin d'évaluer le niveau global de confiance dans les paramètres des ligands CHD et TCH obtenus à l'aide de CHARMM General Force Field (CGenFF), les fichiers portant les extensions « .rtf » et « .prm » des deux ligands ont été analysés. Une pénalité inférieure à 10 indique une très bonne fiabilité, nécessitant peu ou pas d'optimisation. Une pénalité entre 10 et 50 indique qu'une légère optimisation pourrait être requise, tandis qu'une pénalité supérieure à 50 indique qu'une optimisation intensive est nécessaire.

Pour le ligand CHD, tous les paramètres analysés dans les fichiers « .rtf » et « .prm » présentent une pénalité égale à zéro, indiquant que ces paramètres sont très fiables et qu'aucune optimisation supplémentaire n'est requise.

Pour le ligand TCH, l'analyse montre également que l'ensemble des paramètres issus des fichiers « .rtf » et « .prm » possèdent des pénalités nulles, confirmant ainsi un excellent niveau de fiabilité.

En conclusion, les paramètres générés par CGenFF pour les deux ligands CHD et TCH présentent un niveau global de confiance très élevé, ne nécessitant aucune optimisation supplémentaire avant la réalisation des simulations moléculaires.

5. Simulations avec solvant explicite

5.1. RMSD

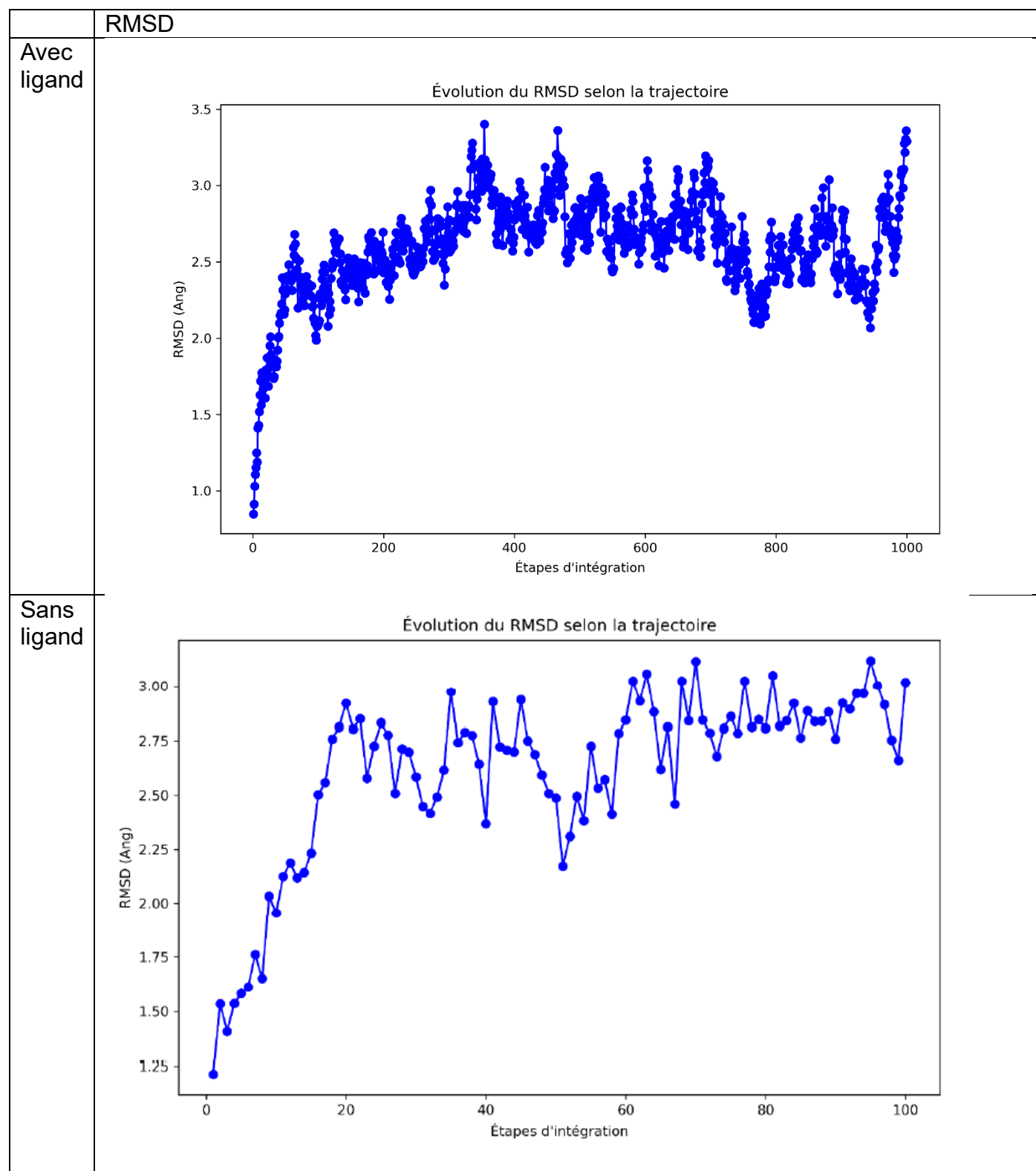


Figure 3 : Analyse de l'évolution de la RMSD pour le ligand CHD en solvant explicite, avec et sans ligand.

L'évolution du Root-Mean-Square Deviation (RMSD) au cours des simulations de dynamique moléculaire est représentée sur les graphiques ci-dessus. Il est important de noter que les deux graphiques ne sont pas directement comparables en raison des différences d'échelle et du nombre d'étapes d'intégration utilisées.

Pour la simulation en présence du ligand, le RMSD fluctue entre environ 1,0 et 3,5 Å, présentant une stabilisation progressive après environ 300 étapes d'intégration. Cette fluctuation initiale plus marquée suggère une réorganisation structurale significative au début de la simulation, possiblement induite par la présence du ligand dans le site actif.

Dans le cas de la simulation sans ligand, le RMSD oscille entre environ 1,2 et 3,0 Å et atteint une relative stabilité après environ 40 étapes d'intégration. Bien que les fluctuations soient moins importantes en amplitude comparativement à la simulation avec ligand, la protéine semble atteindre une conformation stable plus rapidement.

Ces observations indiquent que la présence du ligand pourrait influencer significativement la dynamique structurale de la protéine CmeR, induisant potentiellement des changements conformationnels plus importants et une période d'équilibre plus longue. Ces effets pourraient être liés à la modulation des interactions spécifiques dans le site actif en présence du ligand.

5.2. Angles dièdres

5.2 Angles dièdres

Les graphiques des angles dièdres des résidus Leu65, Ile68 et Ser207 montrent des différences marquées entre les simulations avec et sans ligand. Il est toutefois essentiel

de considérer que le nombre d'étapes d'intégration (frames) est différent pour les graphiques avec et sans ligand, rendant ainsi la comparaison directe limitée.

Pour Leu65, en présence du ligand (nombre d'étapes plus élevé), l'angle dièdre présente une fluctuation très étendue, oscillant entre environ -180 et +180 degrés, indiquant une grande flexibilité ou une rotation fréquente autour de cette liaison. En contraste, sans ligand (moins d'étapes), l'angle dièdre de Leu65 affiche une variation plus restreinte, principalement autour de deux régions distinctes (~ -150 et ~ 150 degrés), suggérant des états conformationnels préférentiels plus définis.

Concernant Ile68, malgré le nombre différent d'étapes, la présence du ligand stabilise visiblement l'angle dièdre, avec une majorité de points concentrés entre -80 et -40 degrés. À l'inverse, sans ligand, les fluctuations observées couvrent une plage légèrement plus étendue, révélant une flexibilité conformationnelle accrue.

Pour Ser207, l'effet stabilisateur du ligand est encore plus marqué. En présence du ligand, l'angle dièdre se stabilise nettement autour de deux états principaux (environ 60 degrés au début, puis stabilisé autour de -50 degrés pour la majorité de la trajectoire), tandis que sans ligand, l'angle dièdre fluctue librement entre 40 et 80 degrés, témoignant d'une plus grande liberté conformationnelle.

Ces différences observées dans les angles dièdres, malgré les limites imposées par le nombre d'étapes différent, suggèrent que le ligand influence fortement la dynamique interne et la flexibilité des résidus clés au sein de la protéine CmeR. En se liant au site actif, le ligand CHD pourrait induire ou stabiliser des conformations spécifiques, réduisant ainsi la variabilité des angles dièdres et contribuant potentiellement à la fonction biologique de la protéine en stabilisant son état actif ou en modulant son interaction avec d'autres molécules.

	Leu65
--	-------

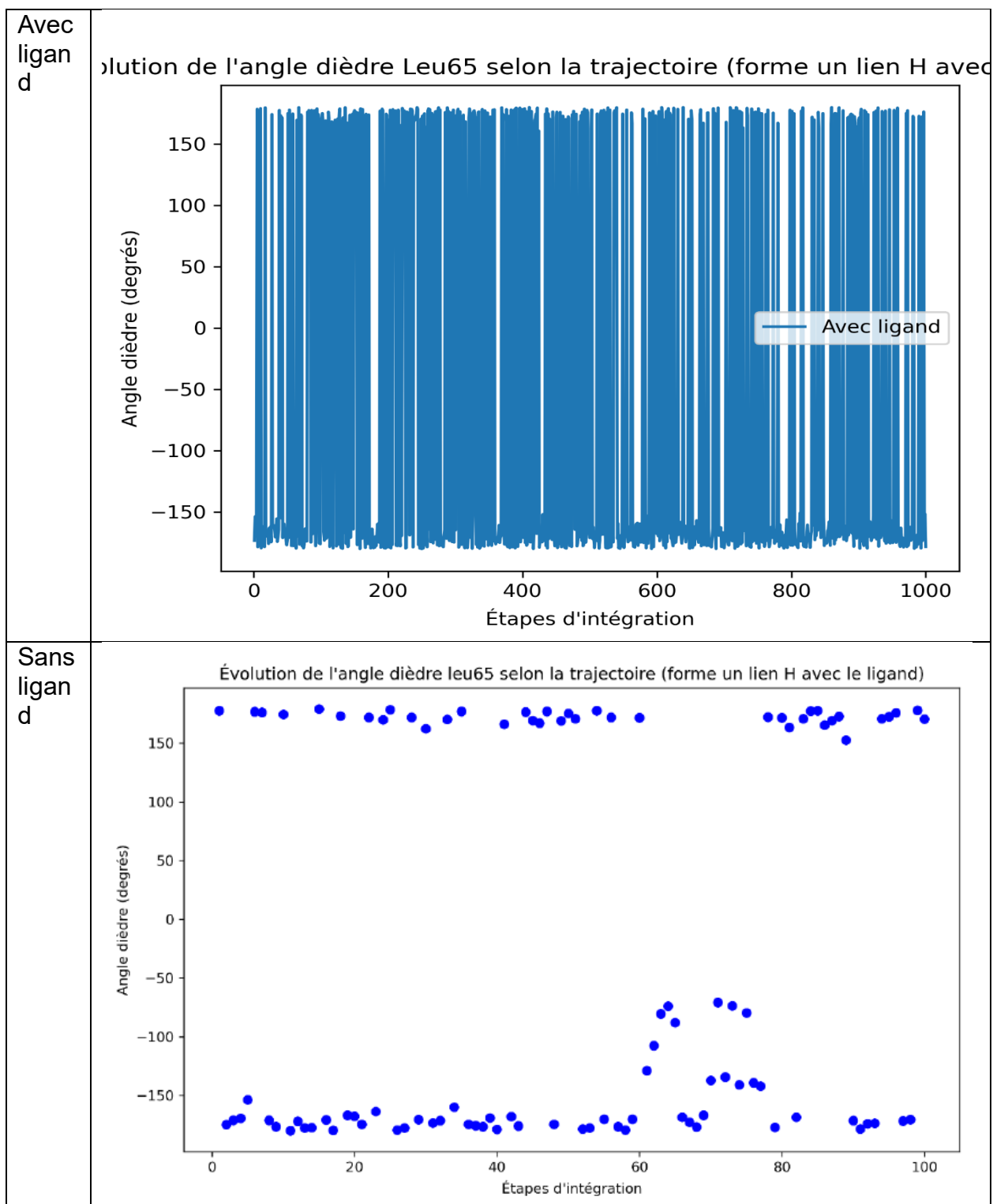


Figure 4 : Analyse de l'évolution d'évolution de l'angle dièdre Leu65 en contact avec le ligand. Selon la trajectoire pour le ligand CHD en solvant explicite, avec et sans ligand.

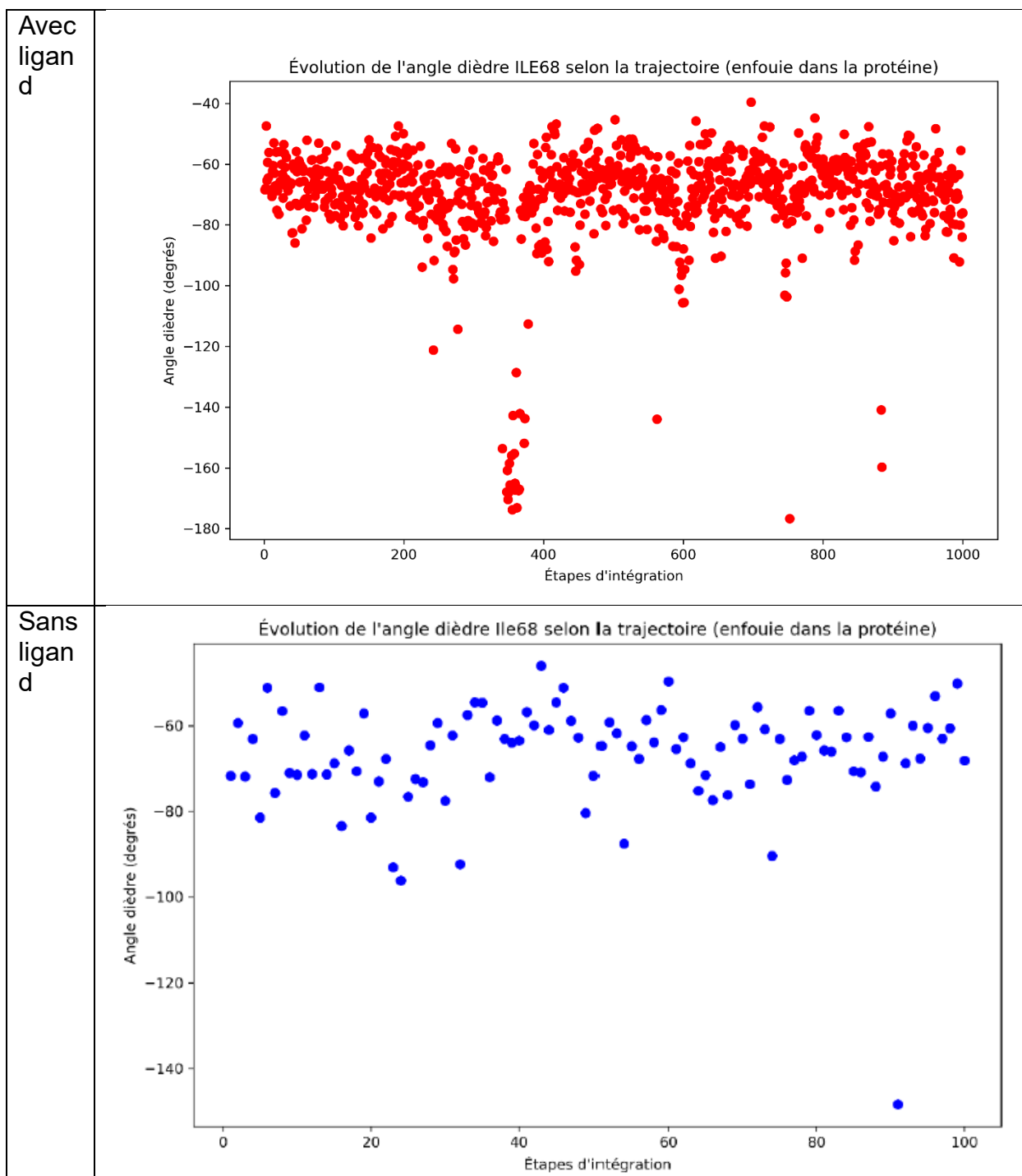


Figure 5 : Analyse de l'évolution d'évolution de l'angle dièdre Ile68 enfouie à l'intérieur de la protéine. Selon la trajectoire pour le ligand CHD en solvant explicite, avec et sans ligand.

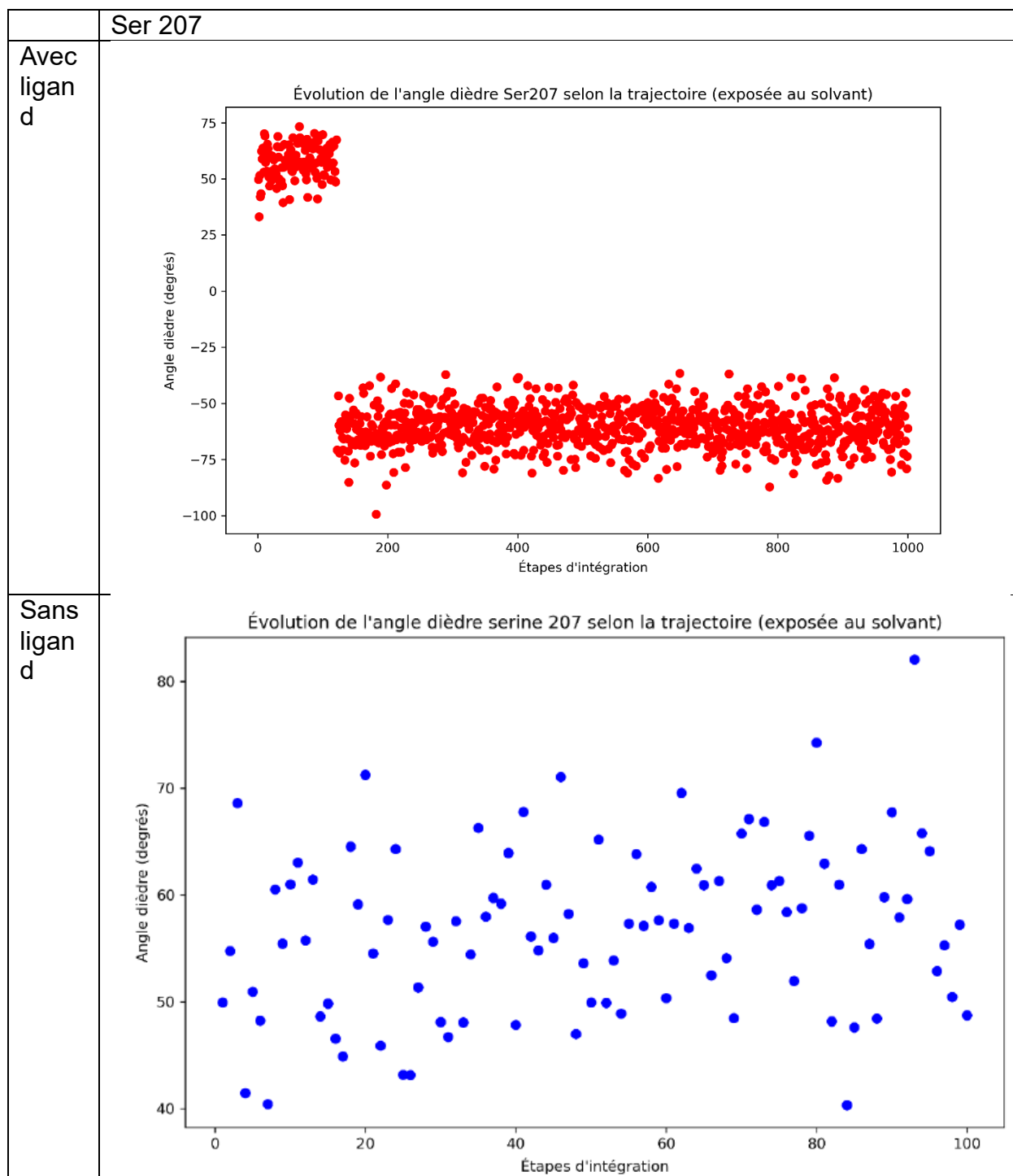


Figure 6 : Analyse de l'évolution d'évolution de l'angle dièdre Ser207 exposé au solvant. Selon la trajectoire pour le ligand CHD en solvant explicite, avec et sans ligand.

En conclusion, les différences observées dans la dynamique des angles dièdres et du RMSD en présence et en absence du ligand pourraient s'expliquer par l'interaction directe du ligand CHD avec des résidus clés du site actif de la protéine CmeR. Le ligand semble imposer des contraintes structurelles spécifiques, réduisant ainsi la flexibilité de certains résidus. À l'inverse, en absence de ligand, la protéine adopte une plus grande variété de conformations, reflétant une flexibilité accrue et une absence de contraintes externes. Ces résultats suggèrent que le ligand CHD joue un rôle stabilisateur déterminant sur la structure et potentiellement sur la fonction biologique de la protéine CmeR.

5.3. Analyse des interactions

La structure finale du complexe protéine-ligand, visualisée à l'aide du logiciel PyMOL, est présentée à la Figure 7. Cette représentation illustre le site de liaison du ligand CHD sur la protéine CmeR après 10 ns de dynamique moléculaire en solvant explicite. Pour plus de clarté, les molécules de solvant et les ions ont été retirés, laissant uniquement la protéine, le ligand, et les résidus importants. Les acides aminés Cys166A, Phe137A et Ile115A, prédits comme étant en interaction avec le ligand dans le devoir #1 via *PoseView*, ont été colorés respectivement en jaune, cyan, et magenta, et affichés en sticks. Des distances ont été mesurées entre ces résidus et le ligand afin d'évaluer la présence potentielle de liaisons hydrogène à la fin de la trajectoire.

Les résultats montrent qu'aucune liaison hydrogène claire n'est observée entre les résidus Cys166A, Phe137A ou Ile115A et le ligand CHD à la fin de la simulation. Les distances mesurées dépassent généralement la limite conventionnelle pour une interaction H (souvent $< 3.5 \text{ \AA}$). Ceci indique une perte des interactions hydrogène prédite initialement par *PoseView*. De plus, l'orientation du ligand semble légèrement décalée par rapport à la structure de départ, suggérant une certaine mobilité dans le site de liaison.

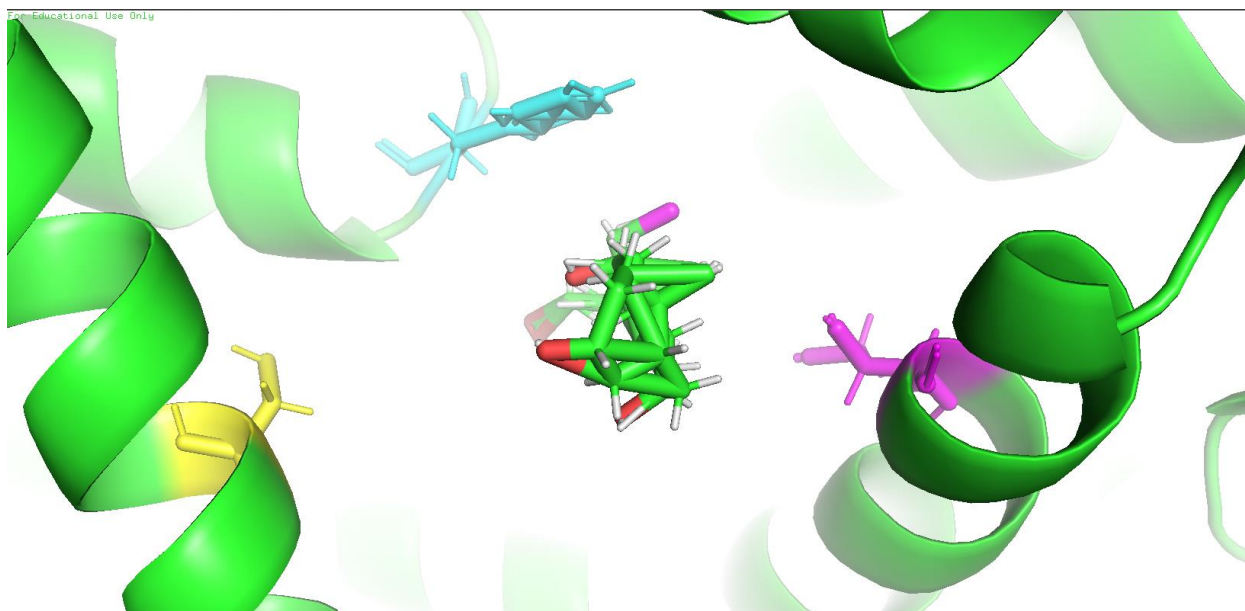


Figure 7: Capture de la structure finale, générée à l'aide du logiciel PyMol, du site de liaison du ligand CHD sur la protéine CmeR et les interactions avec les résidus Cys166, Phe137 et Ile115. Cys166 est présenté en jaune et Phe137 est présenté en cyan et Ile est présenté en magenta. La structure de la protéine et présentés en cartoon en verte et le ligand en bâtonnet.

Tableau 2 : Comparaison des interactions ligand-protéine prévues par PoseView et observées après la simulation

Acide aminé	Interaction prédite avec PoseView	Interaction observée à la fin de la trajectoire
Cys166	Liaison hydrogène avec un OH du ligand	Aucune interaction observée
Phe137	Interaction hydrophobe	Aucune interaction observée
Ile115	Interaction hydrophobe	Aucune interaction observée

6. Simulations GBIS avec réchauffement

RMSD

L'analyse de la déviation quadratique moyenne (RMSD) permet de suivre l'évolution conformationnelle des complexes CmeR-ligand au cours des simulations en solvant implicite avec réchauffement (méthode GBIS). Ces graphes nous renseignent sur la stabilité structurale des complexes et peuvent indiquer des événements critiques tels que la dissociation du ligand du site actif.

Pour les deux ligands étudiés (CHD, associé à 3QPS, et TCH, associé à 3QQA), trois trajectoires ont été effectuées afin de mieux cerner la variabilité due aux vitesses initiales aléatoires.

Complexe 3QPS (ligand CHD)

Les courbes RMSD obtenues pour les trois trajectoires du complexe 3QPS montrent une tendance globale à la hausse. Au début de la simulation, les valeurs RMSD restent relativement stables (environ 1.0-2.0 Å), suggérant que le complexe conserve sa conformation initiale. Toutefois, à mesure que la température augmente, la RMSD augmente aussi de manière progressive, atteignant parfois 3.5 Å voire plus dans certaines trajectoires.

Cette élévation soutenue de la RMSD peut refléter des changements conformationnels dans la protéine, potentiellement liés à un affaiblissement des interactions ligand-protéine ou à une dissociation partielle du ligand.

Complexe 3QQA (ligand TCH)

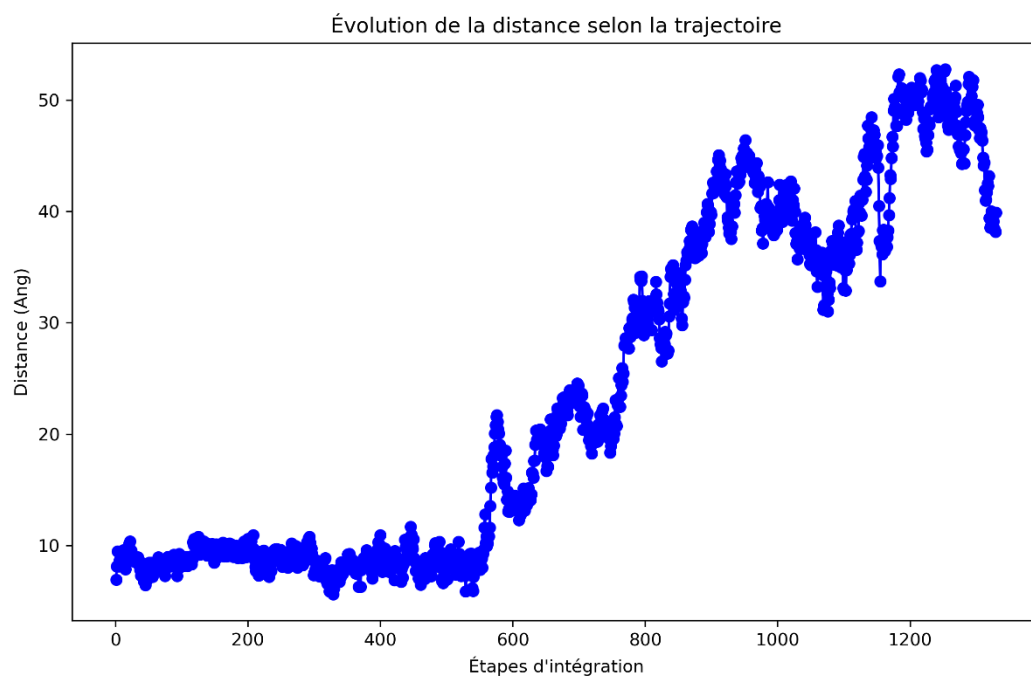
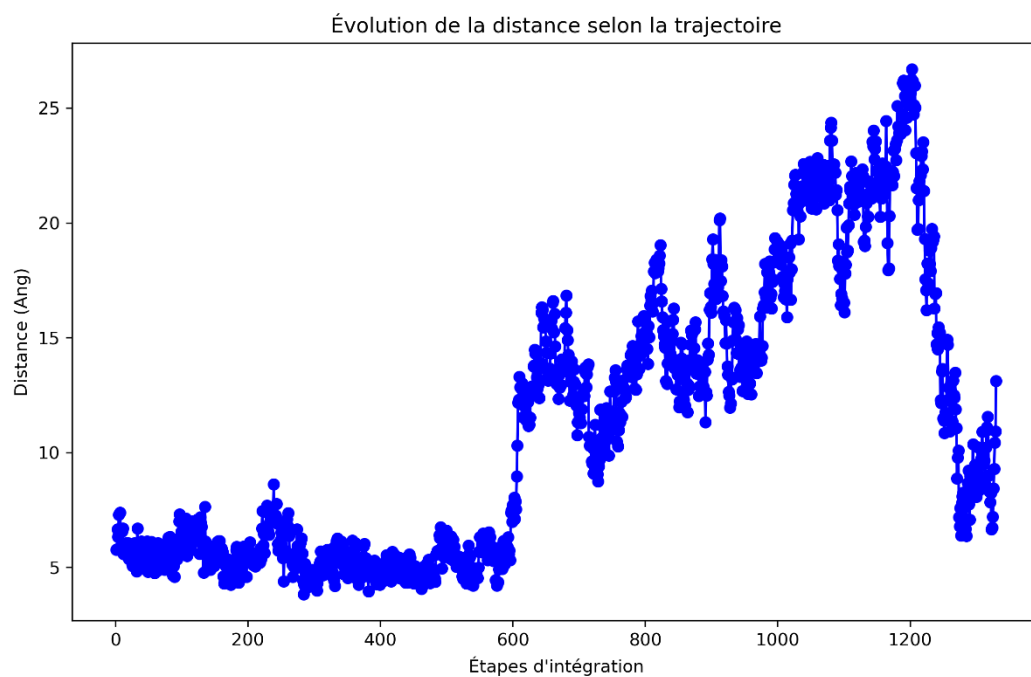
Les trajectoires RMSD du complexe 3QQA révèlent un comportement légèrement différent. Dans la majorité des cas, les RMSD sont plus faibles et plus stables par rapport au complexe 3QPS, oscillant souvent entre 1.0 et 2.5 Å. Cela suggère une plus grande stabilité du complexe au fil de la simulation.

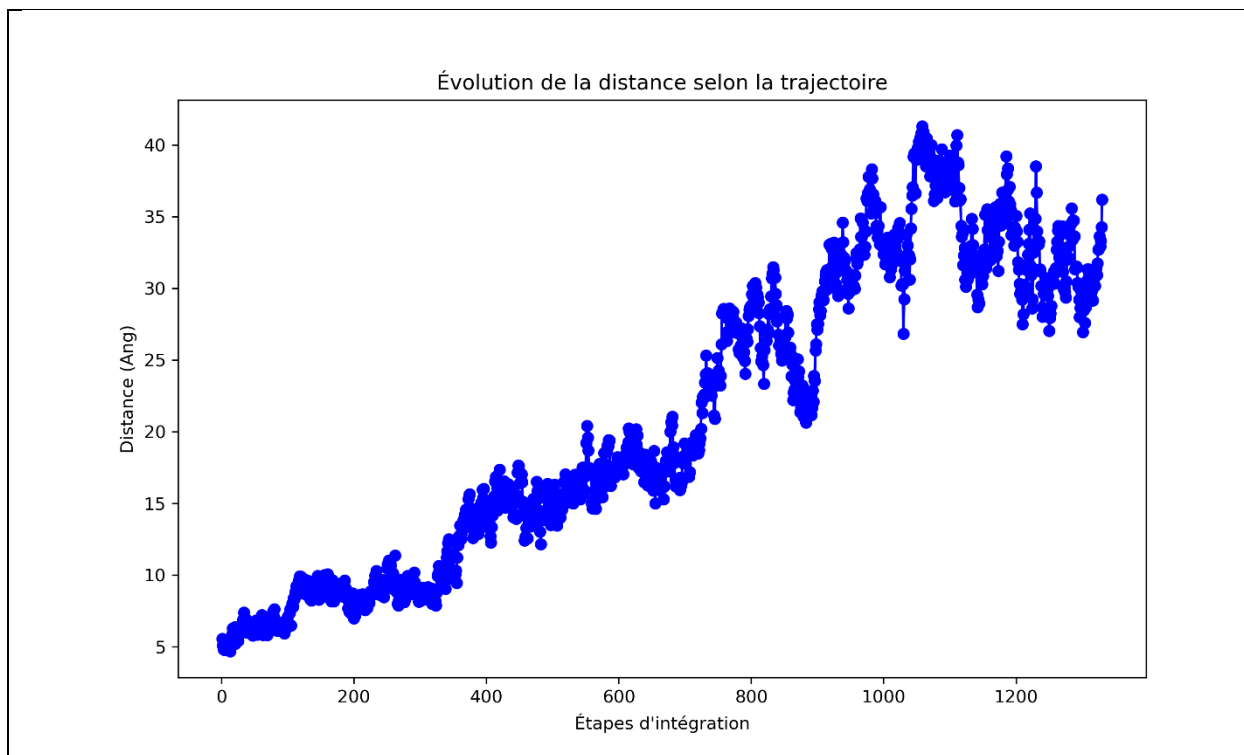
Cette stabilité accrue pourrait s'expliquer par la meilleure affinité du ligand TCH pour la protéine CmeR, comme indiqué par sa constante de dissociation plus faible ($K_d = 1,5 \mu\text{M}$ pour TCH contre $2,5 \mu\text{M}$ pour CHD). Le ligand TCH semble donc maintenir plus efficacement ses interactions avec le site actif, même en conditions de réchauffement.

Interprétation générale

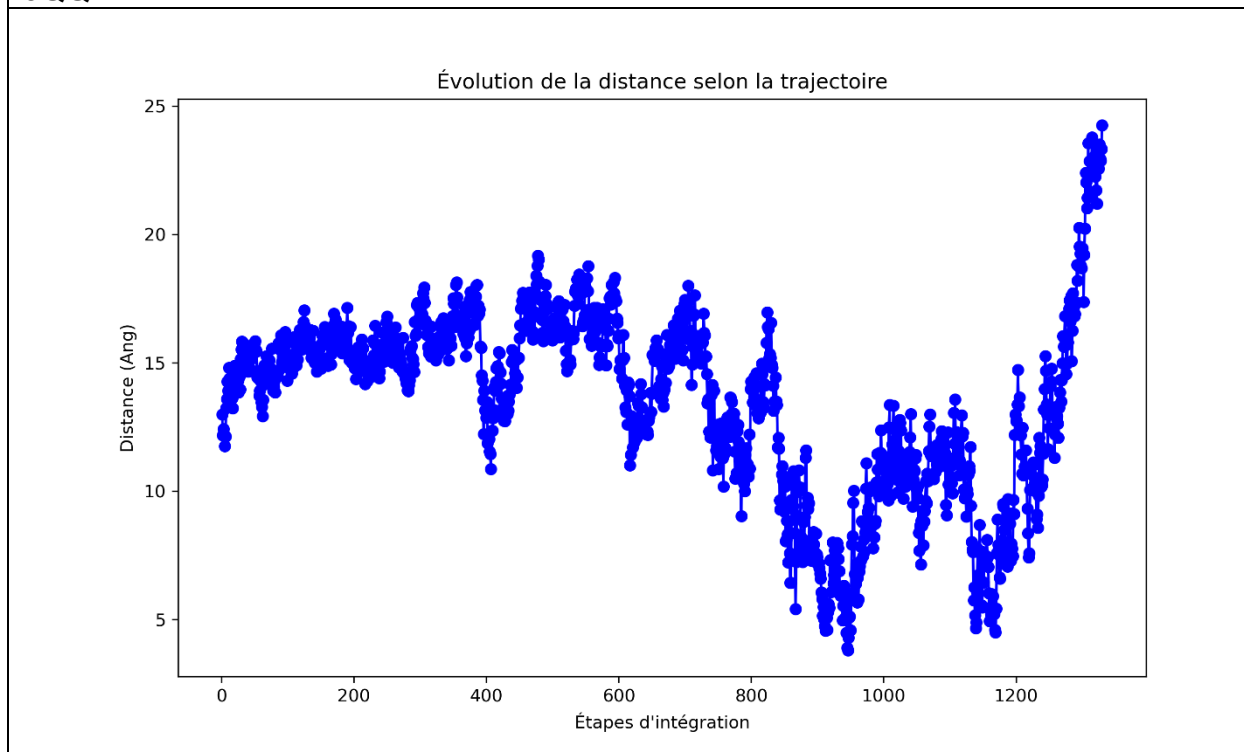
Les différences observées dans les profils RMSD des deux complexes suggèrent que le ligand TCH (3QQA) stabilise davantage le complexe protéine-ligand que le ligand CHD (3QPS). L'augmentation marquée de la RMSD dans certaines trajectoires du complexe 3QPS pourrait être associée à des événements de détachement du ligand, ce qui sera confirmé par l'analyse conjointe de la température et de la distance ligand-résidu clé.

3QPS





3QQA



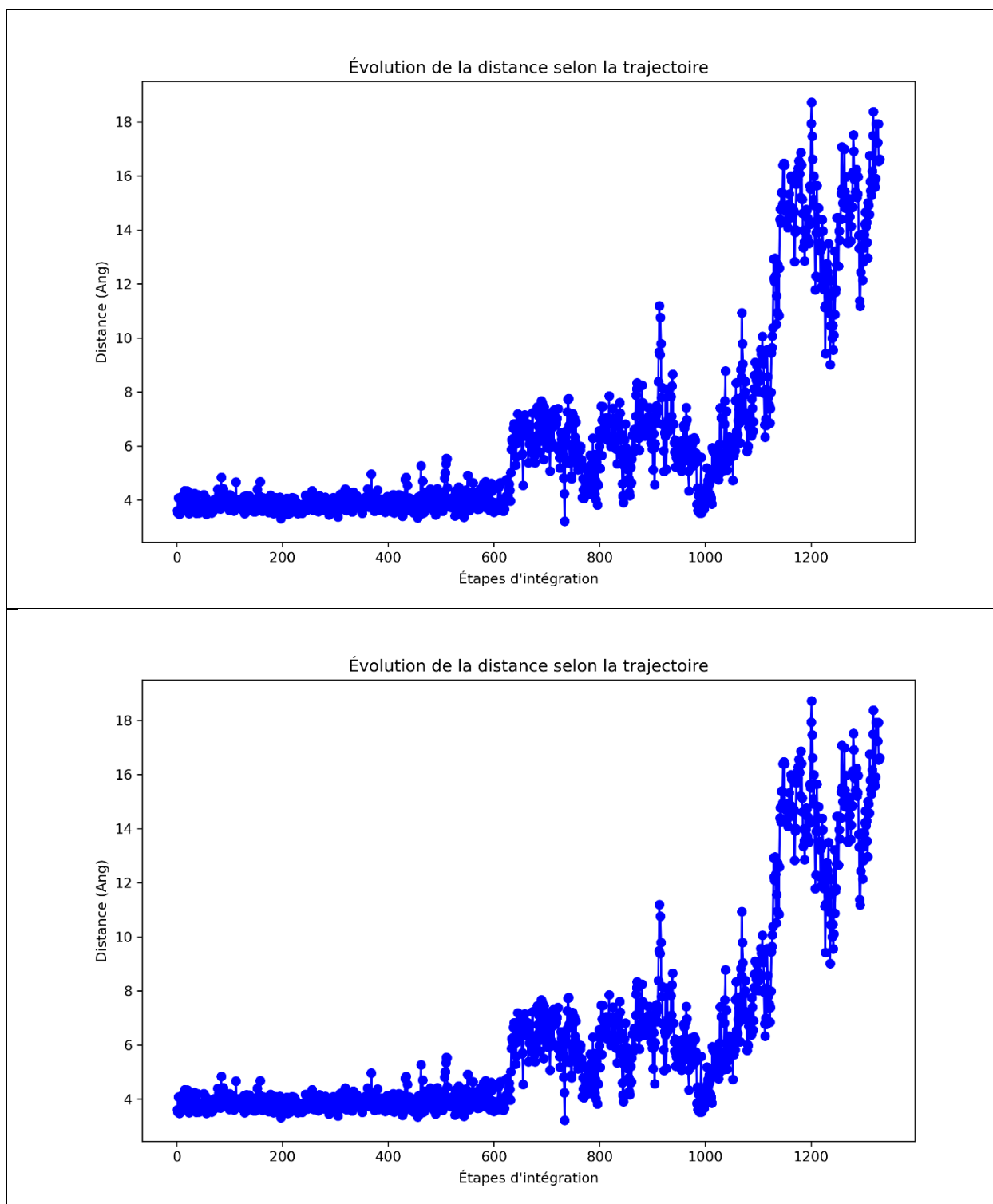


Figure 8 : Graphiques de l'évolution de la RMSD selon la trajectoire pour les deux ligand CHD et TCH avec 3 simulations pour chacun des deux ligands.

Distance entre un atome du ligand et un atome d'un acide aminé ayant une interaction

avec le ligand

Pour évaluer le comportement dynamique des ligands au sein du site actif de la protéine CmeR, la distance entre un atome représentatif du ligand et un atome d'un résidu clé du site de liaison a été mesurée tout au long des simulations GBIS. Ces atomes ont été choisis en fonction des interactions identifiées lors de l'analyse initiale via PoseView. Cette approche permet de visualiser les éventuels événements de dissociation ou les réarrangements du ligand au cours du réchauffement.

Complexe 3QPS (ligand CHD)

Les graphes de distance pour les trois trajectoires du complexe 3QPS montrent un comportement relativement instable. Dans chacune des simulations, la distance entre le ligand CHD et le résidu cible de la protéine débute à une valeur faible, indiquant un ancrage initial solide du ligand dans le site actif. Toutefois, à mesure que la simulation progresse et que la température augmente, des fluctuations significatives sont observées.

Dans certaines trajectoires, ces variations aboutissent à des pics marqués, suggérant des épisodes de détachement partiel, voire complet, du ligand. L'amplitude de ces fluctuations indique que CHD explore des conformations plus éloignées du site actif, ce qui reflète une interaction moins stable avec la protéine. Ce comportement pourrait indiquer une affinité plus faible ou une complémentarité structurale sous-optimale avec le site de liaison de CmeR.

Complexe 3QQA (ligand TCH)

En comparaison, les courbes de distance pour le complexe 3QQA présentent une stabilité beaucoup plus marquée. Dès le début des simulations, la distance entre l'atome du ligand TCH et le résidu de contact reste faible et relativement constante dans les trois trajectoires. Bien que de légères oscillations soient présentes, aucun événement de dissociation majeur n'est détecté.

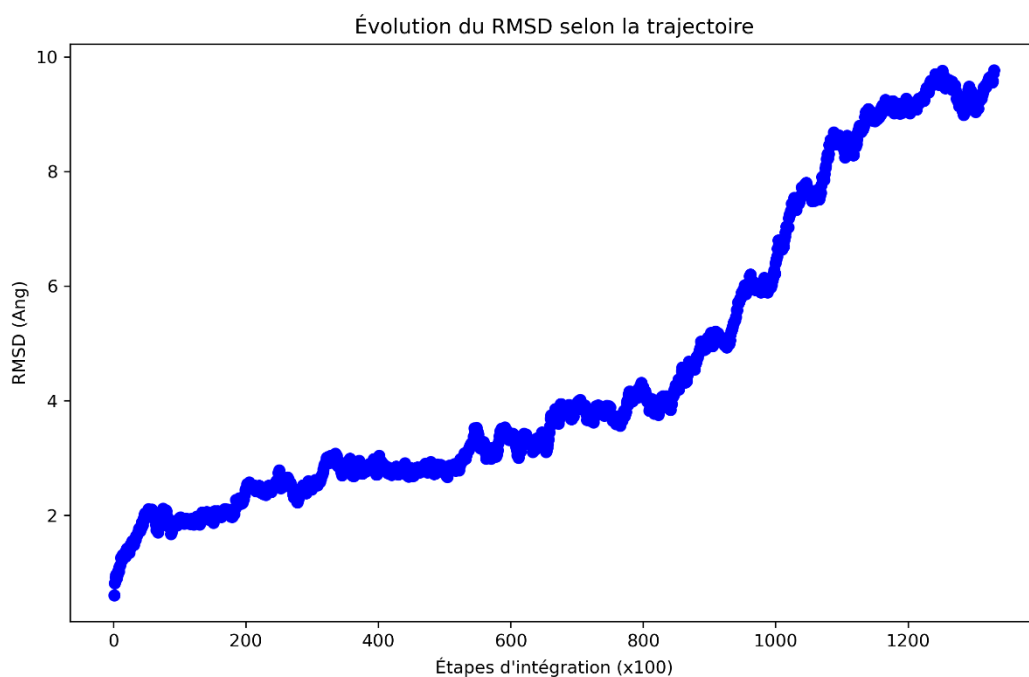
Ce profil indique que TCH maintient un contact étroit avec le site actif tout au long du réchauffement, suggérant une interaction robuste et durable. Cette stabilité accrue peut être corrélée à une meilleure affinité du ligand pour la protéine CmeR, comme en témoigne son plus faible K_d expérimental.

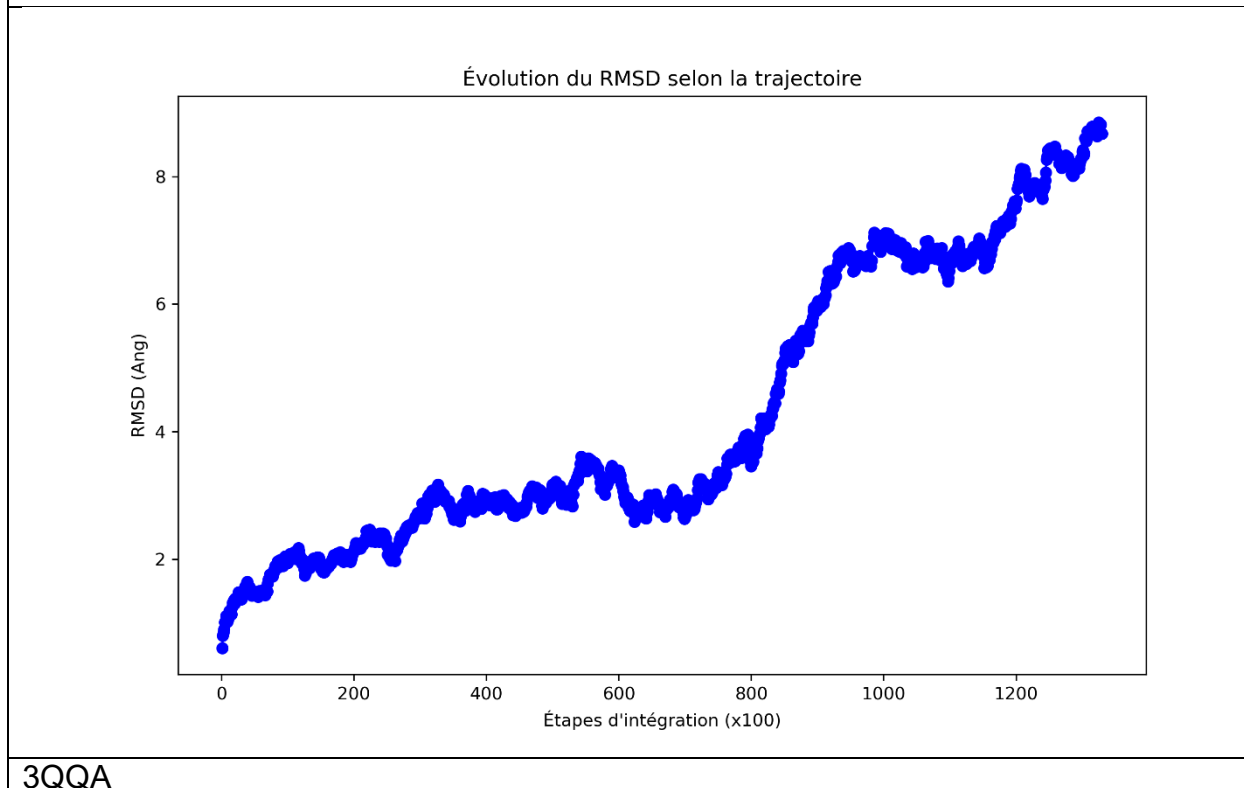
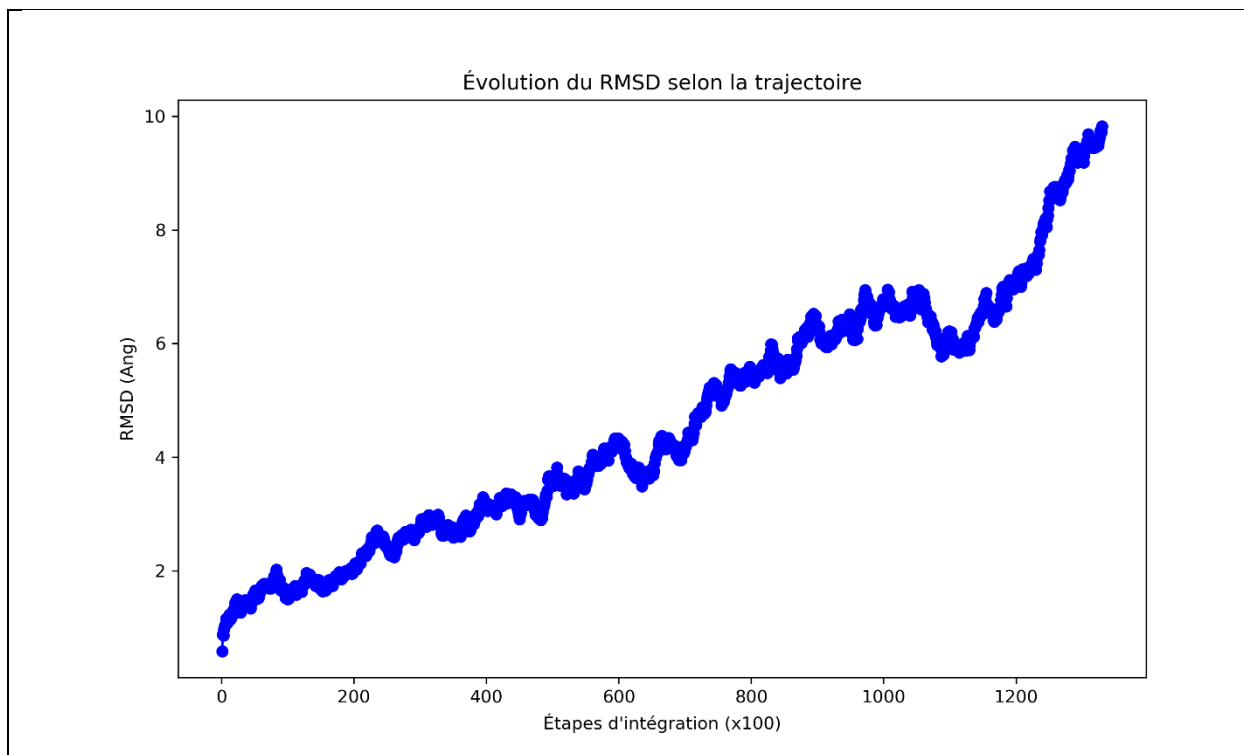
Comparaison et interprétation

L'ensemble des résultats révèle des différences claires dans le comportement des deux ligands. CHD montre une tendance à la dissociation partielle dans au moins deux des

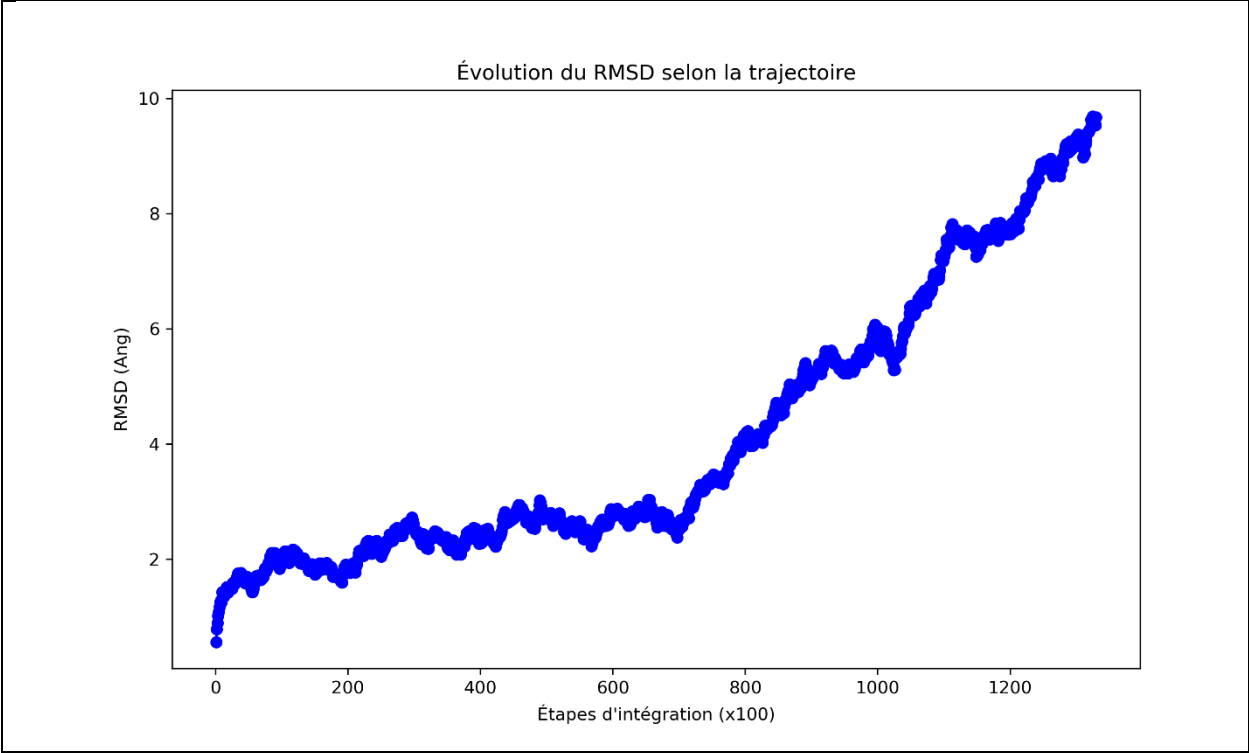
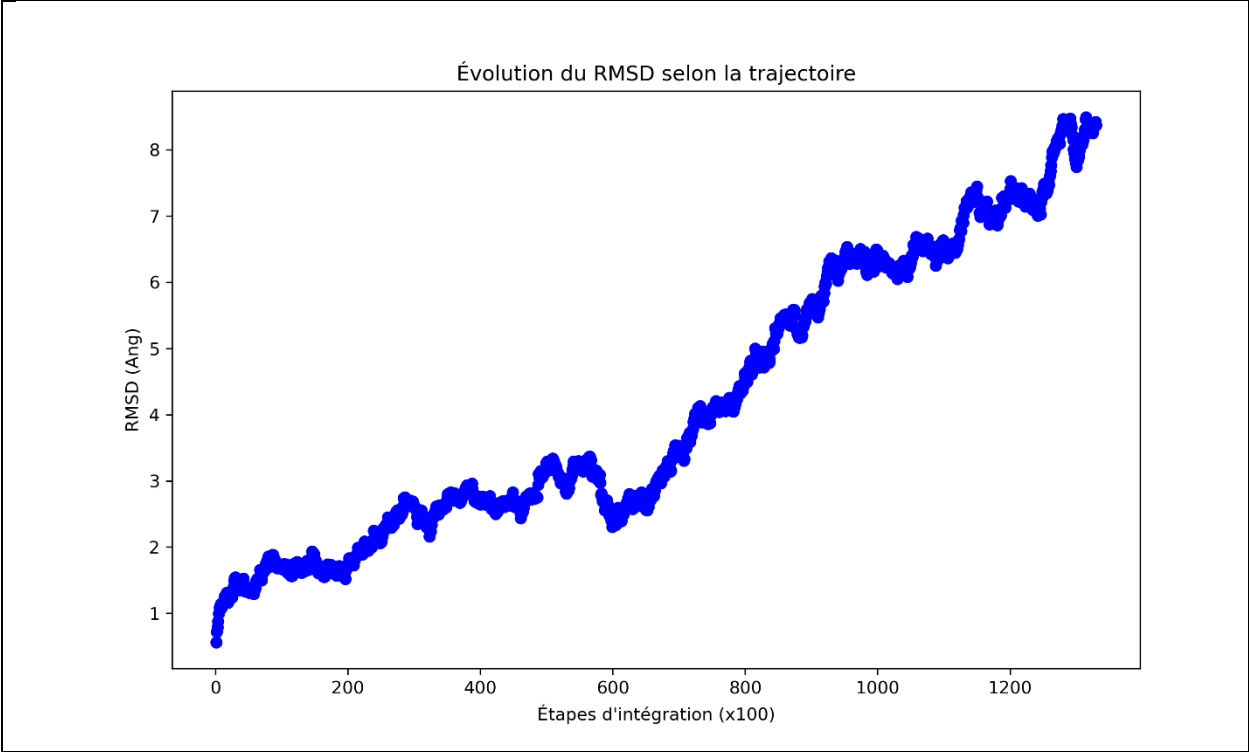
trois trajectoires, alors que TCH reste solidement ancré dans le site actif durant toutes les simulations. Ce contraste confirme que TCH présente une meilleure stabilité de liaison dans les conditions de simulation GBIS, ce qui est cohérent avec sa plus forte affinité expérimentale. Il est toutefois important de rappeler que ces résultats doivent être interprétés avec prudence, car les simulations ne reflètent qu'un échantillon du comportement réel du système.

3QPS





3QQA



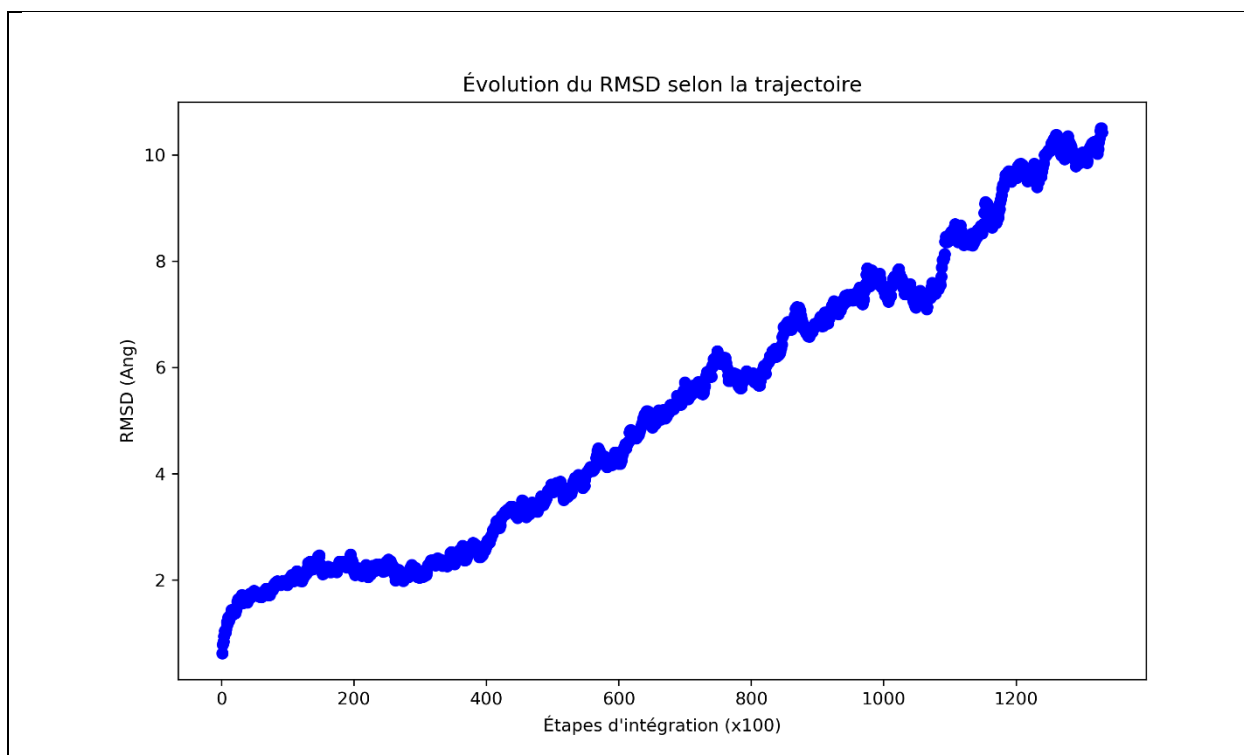


Figure 9 : Graphiques de l'évolution de la distance entre deux atomes, un situé sur le ligand et l'autre sur un acide aminé du site actif ayant une interaction avec le ligand, selon la trajectoire pour les deux ligand TCH et CHD avec 3 simulation pour chacun des deux ligand.

Température

Les courbes de température extraites des simulations GBIS avec réchauffement sont présentées pour les deux complexes protéine-ligand, 3QPS (CHD) et 3QQA (TCH). Ces graphes illustrent l'évolution de la température en fonction des étapes d'intégration pour chacune des trois trajectoires simulées.

Les profils de température observés montrent une augmentation régulière et linéaire dans toutes les trajectoires, confirmant que le protocole de réchauffement a été exécuté de manière stable et contrôlée. Aucun pic soudain, ni chute inattendue, n'a été détecté dans les courbes. Cela suggère que le processus de réchauffement ne provoque aucune instabilité thermique brutale, ce qui est essentiel pour une interprétation cohérente des données de dissociation.

Complexe 3QPS (ligand CHD)

Pour le ligand CHD, les températures d'éjection (c'est-à-dire, la température à laquelle le ligand quitte visiblement le site actif, d'après les distances analysées) sont les suivantes :

- **Trajectoire 1 :** 353 K

- **Trajectoire 2** : 366 K
- **Trajectoire 3** : 371 K

Cela donne une **température moyenne d'éjection de 363.3 K** pour le ligand CHD. On observe une certaine variabilité entre les trajectoires, mais les valeurs restent globalement proches, indiquant une cohérence dans le comportement de dissociation.

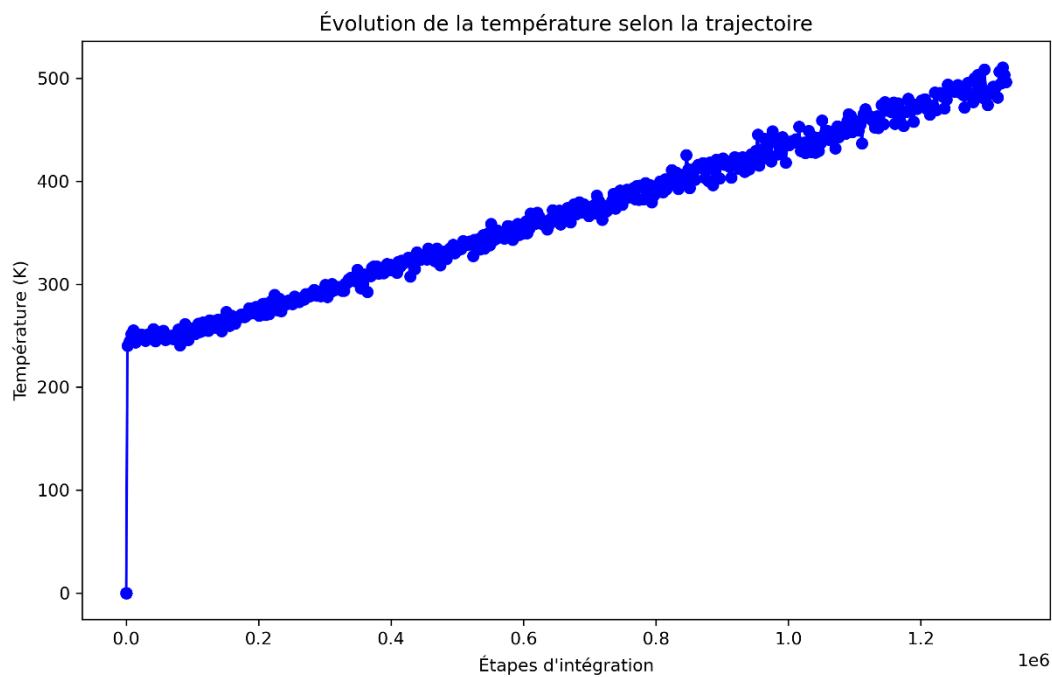
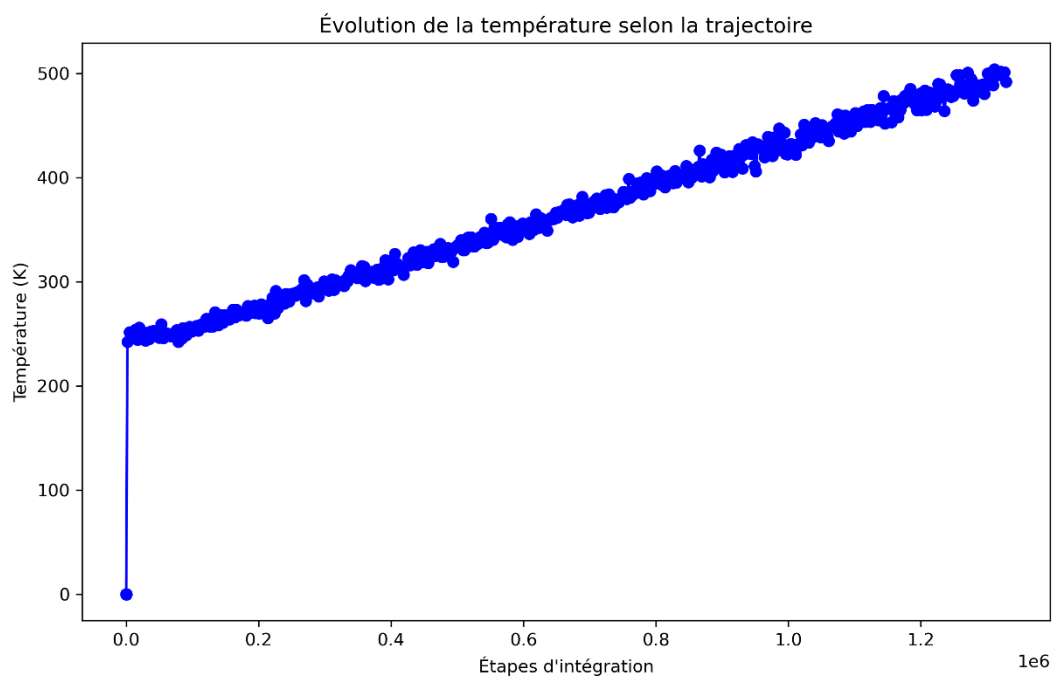
Complexe 3QQA (ligand TCH)

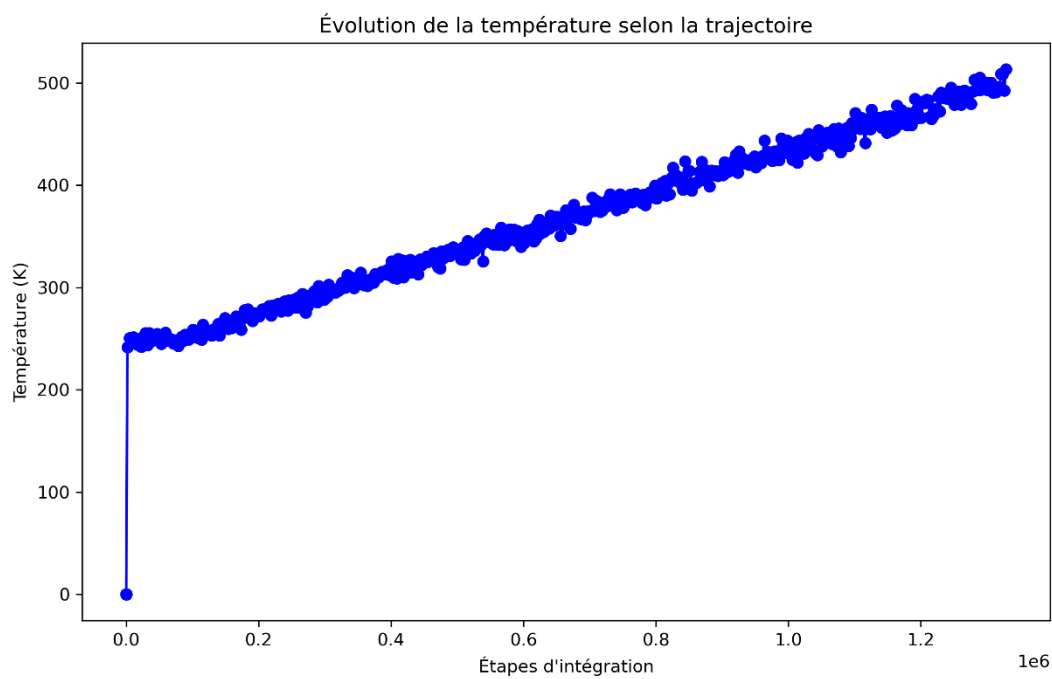
En ce qui concerne le ligand TCH, les courbes de distance indiquent que le ligand ne quitte pas clairement le site actif dans aucune des trois trajectoires. Ainsi, nous avons relevé la température maximale atteinte dans chaque simulation comme point de référence :

- **Trajectoire 1** : 375 K
- **Trajectoire 2** : 377 K
- **Trajectoire 3** : 379 K

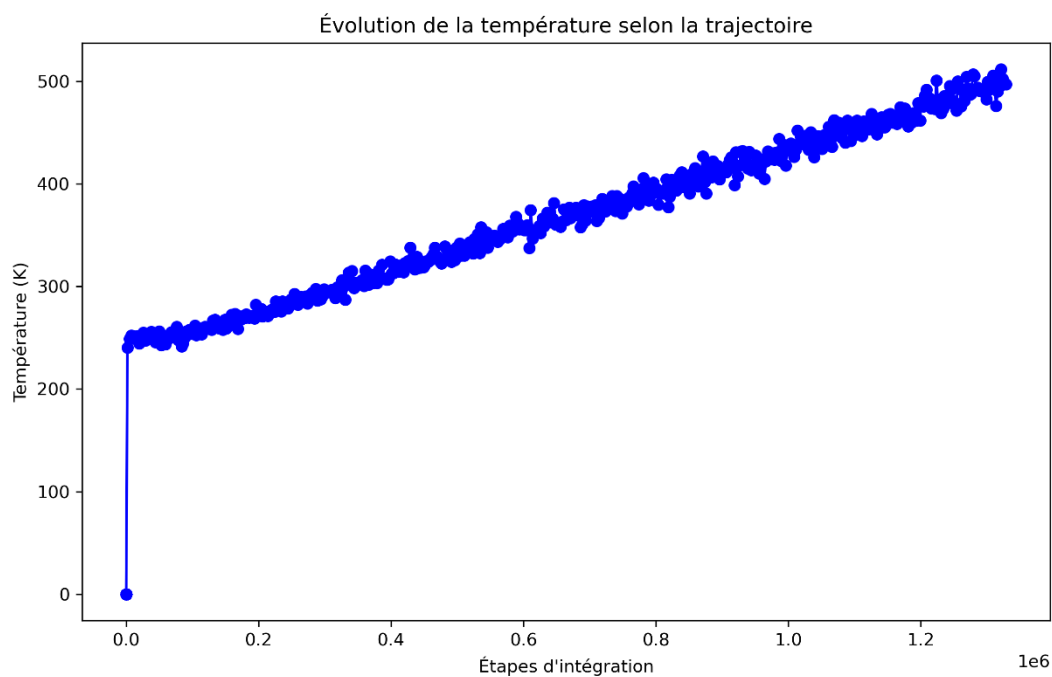
La température moyenne maximale est donc de **377.0 K** pour TCH.

Ces résultats confirment que le ligand TCH reste plus stable et mieux ancré dans le site actif comparé au ligand CHD, qui se dissocie à des températures inférieures. Cette différence est en accord avec les valeurs d'affinité expérimentale, où TCH présente une constante de dissociation (K_d) plus faible que CHD, indiquant une liaison plus forte à la protéine MTH1.





3QQA



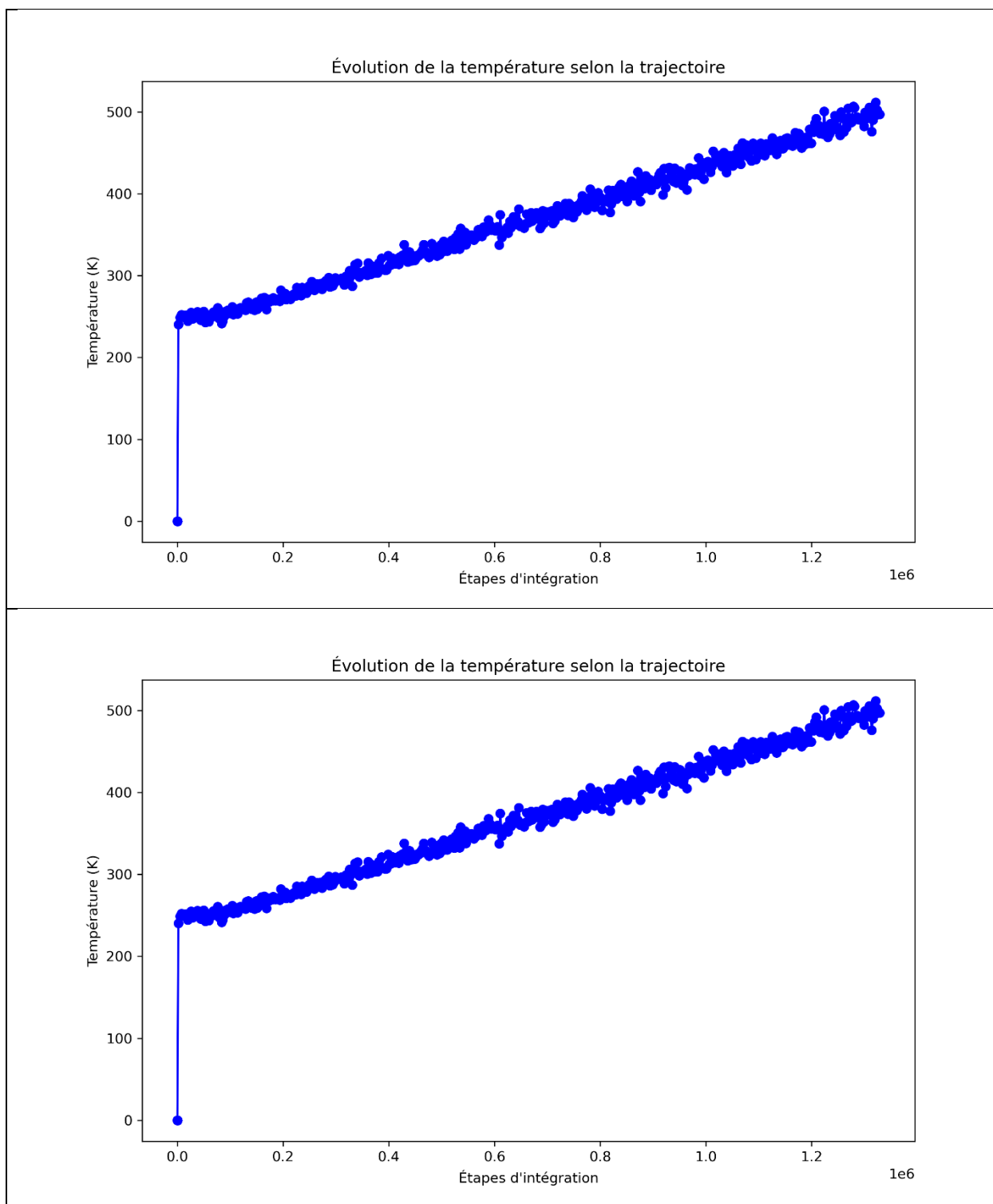


Figure 10 : Graphiques de l'évolution de la température selon la trajectoire pour les deux ligand TCH et CHD avec 3 simulations pour chacun des deux ligands.

Sortie du ligand du site actif et température de sortie

Afin de déterminer le moment approximatif où chaque ligand quitte son site de liaison sur la protéine CmeR, les données issues des graphiques de RMSD et de distance entre le ligand et un acide aminé clé ont été analysées. Les étapes d'intégration approximatives correspondant à la sortie des ligands ont été regroupées dans le tableau présenté ci-dessus.

Pour le ligand CHD (complexe 3QPS), les résultats montrent qu'un changement de conformation du complexe, visible par une augmentation significative du RMSD, débute autour de la 353^e à 371^e étape d'intégration selon les trajectoires. La distance entre le ligand et la protéine augmente également de manière marquée autour de l'étape 925, suggérant une dissociation du ligand de son site actif. Ces observations indiquent une stabilité modérée du ligand CHD au sein du complexe avec CmeR. La température moyenne de sortie du ligand CHD est estimée à environ 363 K, reflétant une affinité intermédiaire entre le ligand et la protéine.

En revanche, pour le ligand TCH (complexe 3QQA), aucun signal clair de dissociation n'a été observé dans les trois trajectoires, ni dans les profils de distance, ni dans ceux de RMSD. Le ligand reste solidement ancré dans le site actif jusqu'aux dernières étapes des simulations. Ainsi, les températures maximales atteintes (allant de 375 K à 379 K) ont été utilisées à défaut d'une température de sortie définie. La moyenne de ces températures est de 377 K, ce qui suggère que le ligand TCH possède une interaction beaucoup plus robuste et reste stable malgré l'augmentation thermique progressive.

En comparant les deux complexes, on constate que le ligand TCH reste lié plus longtemps et à des températures plus élevées que le ligand CHD, ce qui est cohérent avec les données expérimentales d'affinité. Le ligand TCH montre une plus faible constante de dissociation (K_d), indiquant une meilleure affinité pour la protéine CmeR. À l'inverse, le ligand CHD, qui se dissocie plus tôt, pourrait avoir une complémentarité structurale plus faible avec le site actif ou une sensibilité plus grande aux variations thermiques.

Ces résultats renforcent l'idée que la température de sortie peut être utilisée comme indicateur indirect de l'affinité ligand-protéine, et qu'une température de sortie plus élevée reflète potentiellement une liaison plus stable et une interaction plus forte dans le cadre de la simulation GBIS avec réchauffement.

Tableau 4 : Étape d'intégration approximative de sortie du ligand de son site de liaison sur la protéine CmeR, estimée par la RMSD et par la distance entre le ligand et la protéine, ainsi que la température de sortie du ligand (K).

	CHD (3QPS)		TCH (3QQA)	
--	------------	--	------------	--

Simulation	Selon RMSD (A)	Selon la distance (A)	Selon RMSD (A)	Selon la distance (A)
Simulation 1	353	900	375	1100
Simulation 2	366	900	380	1100
Simulation 3	371	975	385	1180
Moyenne approx.	363.3	925	380	1126.6
Température de sortie du ligand (K)	363		377	