
RAPPORT DE STAGE

Ciblage de l'interface IL-3/IL-3R α avec des méthodologies de *drug design* et de modélisation moléculaire dans le cadre de maladies atopiques

ROUAUD Lucas – 22 117 185
Master 1^{re} année mention Bio-Informatique ;
parcours BI / IPFB

Responsable de stage :
M. DIHARCE J.

Directeur d'équipe :
M. de BREVERN A.

UNIVERSITÉ PARIS CITÉ
UFR des sciences du vivant
Département des Sciences du Vivant

85 boulevard Saint-Germain – 75 006 PARIS
+33 (0) 157 279 000
<https://u-paris.fr/sdv/>

INSERM
Unité mixte de recherche n°1134
DSIMB

6 rue Alexandre Cabanel – 75 739 PARIS
+33 (0) 144 493 058
<https://www.dsimb.inserm.fr>

Remerciements

Je remercie Thomas pour son oreille toujours attentive, ses conseils avisés sur mes figures ou mes scripts.

Je remercie Carla (et son égo) de m'avoir aidé à bien démarrer mes dynamiques et pour ses avis sur le rendu de mes différentes figures.

Je tiens à sous- et surligner que ces deux personnes précédemment citées m'ont été fort sympathique, m'ont très vite intégré et m'ont permis de passer un très bon moment au cours de ce stage ! Mille merci (mais un de moins pour Carla). Je les remercie aussi pour la rédaction de ce rapport.

Je remercie aussi Mohamed qui, malgré sa courte présence, à grandement et majoritairement participé à la bonne ambiance dans la salle des stagiaires !

Je remercie l'équipe DSIMB de m'avoir accueilli et pour la superbe ambiance qu'il y a au sein de l'équipe. Je remercie tout particulièrement Gabriel, sans qui nous serions tous morts au premier bug informatique rencontré.

Je remercie plus généralement les stagiaires pour l'ambiance très conviviale (même si on dirait tout le contraire) ! Je remercie tout particulièrement Carolynn qui m'a aidé à la rédaction de ce rapport.

Je remercie Alexandre pour ses conseils avisés (et oui, un script est plus efficace qu'une vérification à la main).

Je remercie ma famille qui donne un peu de sens à tout ceci.

Et enfin, je remercie (le meilleur pour la fin) Julien, mon maître de stage, qui a été patient, attentif et fort sympathique avec moi ! Grâce à vous j'ai beaucoup appris, alors merci !

Je terminerai par dire que franchement, un *cheesecake* sans coulis... Non, vraiment ça m'énerve trop...

***« Hope for the best,
plan for the worst »***

– Une personne avisée qui utilise GROMACS.

Table des matières

Remerciements	i
Table des matières	iii
<i>Corpus</i>	1
1 Introduction	1
1.1 Atopies	1
1.2 Cytokines	1
1.3 IL-3 et son récepteur	2
1.3.1 IL-3	2
1.3.2 IL-3R	3
1.4 Problématique et thérapies	3
2 Matériels et méthodes	4
2.1 Simulation de dynamique moléculaire	4
2.1.1 Obtention de la structure d'IL-3	4
2.1.2 Principe de la simulation de dynamique moléculaire	4
2.1.3 Protocole de simulation	5
2.2 Amarrage moléculaire	5
2.3 Logiciels utilisés	6
3 Résultats	6
3.1 Simulation de dynamique moléculaire	6
3.2 Amarrage moléculaire	8
4 Conclusion	10
5 Bibliographie	11
Annexes	I
Résumé/Abstract	III

1. Introduction

1.1. Atopies

Les atopies sont une sensibilité d'origine génétique à une substance *a priori* inoffensive. Il s'agit plus précisément d'allergies. Il en découle une réponse immunitaire forte et exagérée de l'organisme. Une allergie se déroule en deux phases : la première dite d'exposition, où l'individu rentre en contact avec l'allergène (substance provoquant l'allergie) puis une deuxième phase dite de ré-exposition à ce dernier, provoquant les symptômes.

Cette « hyper-sensibilité » provoque des maladies dites atopiques, avec des manifestations physiques comme de l'eczéma, de l'asthme ou une conjonctivite. Biologiquement, il y a aussi une surproduction de médiateurs de l'inflammation comme des cytokines.

1.2. Cytokines

Les cytokines sont une grande famille de protéines. On y dénombre par exemple plus de 33 interleukines ou plus de 20 facteurs de nécrose tumorale. Elles sont produites par un large panel de cellules (neurones par exemple), mais surtout des cellules immunitaires (comme les lymphocytes T, les macrophages...). Groupe très hétérogène impliqué dans des réponses immunitaires (inflammation) et du cycle cellulaire, elles possèdent plusieurs caractéristiques communes [1] :

- Ce sont des protéines à faible poids moléculaire.
- Elles sont impliquées dans des réponses inflammatoires.
- Elles sont produites localement et ont une action auto- ou paracrine.
- Elles sont très actives, c'est-à-dire qu'elles peuvent agir à des concentrations très faibles (ordre du picomolaire).

La sous-famille des β commons fait partie de la famille des cytokines. Produites par des cellules T ou des macrophages par exemple, on y retrouve l'interleukine 3 (IL-3), l'interleukine 5 (IL-5) et le facteur de stimulation des colonies de granulocytes et de macrophages (GM-CSF). La particularité de cette famille est qu'elle se fixe sur un récepteur composé de deux sous-unités : une spécifique à l'une des trois cytokines et une commune aux trois. Ce dernier est appelé récepteur β .

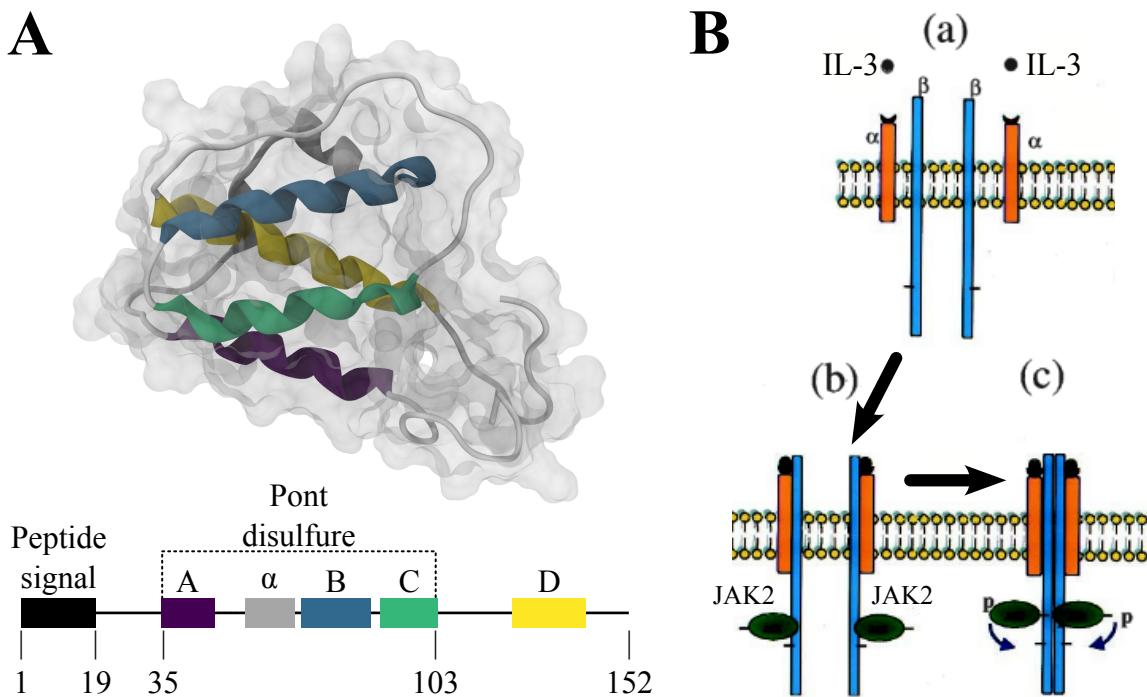


Fig. 1 – A = Structure 3D et schéma de la séquence d'IL-3. Boîte noire = peptide signal non considéré pour la suite de l'étude ; autres boîtes = hélices présentes dans la structure 3D d'IL-3. Leurs noms sont indiqués au-dessus des boîtes. / B = Schéma de l'activation de la voie IL-3/IL-3R adapté de l'article de Reddy E. P. et col. [2]. Orange = sous-unité IL-3R α ; bleu = sous-unité IL-3R β ; (a) = Le ligand se fixe sur la sous-unité IL-3R α ; (b) = la sous-unité IL-3R β est recrutée ; (c) = une fois les dimérisations faites, le récepteur peut transduire un signal cellulaire, ici via JAK2.

1.3. IL-3 et son récepteur

1.3.1. IL-3

L'IL-3 est une glycoprotéine humaine codée par le gène du même nom [3,4] [fig. 1 : A]. Constituée d'un peptide-signal [0 – 19], elle a 152 acides aminés et quatre grandes hélices α ainsi qu'une hélice plus petite. Elle est excrétée dans le compartiment extracellulaire où elle joue des rôles dans l'inflammation, dans la transduction du signal en tant qu'hormone et dans la signalisation en tant que facteur stimulant de colonies hématopoïétiques [5]. Elle est produite par des lymphocytes T et sert aussi à la croissance et à la différenciation des mastocytes impliqués dans des réactions allergiques [6]. Cette protéine jouant un rôle dans l'inflammation [7] peut interagir avec d'autres protéines [I]. Par exemple, l'interaction avec son récepteur peut activer un facteur de transcription, Runx1, impliqué dans l'hématopoïèse [8]. Sa séquence ainsi que d'autres informations sont condensées dans la base de données UNIPROT : Swissprot via le code d'accèsion P08700 [9].

Des études ont montré que IL-3 et sa voie d'activation sont impliquées et même surexprimées dans le cas des maladies atopiques. Ainsi, moduler son activité pourrait être d'un fort intérêt thérapeutique pour contrer ces pathologies [10].

1.3.2. IL-3R

Comme expliqué auparavant, le récepteur d'IL-3 est composé d'une sous-unité spécifique à la protéine, IL-3R α , ainsi qu'une partie commune à la famille des β commons, IL-3R β (ou CSF2RB). Ce récepteur est transmembranaire et permet la transduction du signal lorsqu'il y a interaction avec IL-3 [7]. Pour interagir avec sa protéine et transduire un signal, il y a quatre étapes [11] [fig. 1 : B] :

1. Fixation d'IL-3 sur la sous-unité IL-3R α .
2. Recrutement de la sous-unité IL-3R β c.
3. Dimérisation de deux unités.
4. Transduction du signal.

1.4. Problématique et thérapies

Ainsi, la cytokine IL-3 et son récepteur sont impliqués dans des pathologies atopiques. Il serait donc intéressant d'étudier IL-3 ou son récepteur dans le cadre d'un traitement contre ces maladies. Il est à noter qu'IL-3R α est déjà ciblé pour le traitement de la leucémie aiguë récidivante ou réfractaire par une molécule appelée SGN-CD123A, actuellement en cours d'essai (phase 1). Il s'agit d'un anticorps dirigé contre le récepteur CD123 ou IL-3R α , conjugué à un médicament pour faire de la chimiothérapie [12,13].

Ce stage se place dans le cadre d'un projet ANR ayant pour objectif d'étudier en profondeur la protéine et son récepteur, ainsi que de concevoir des molécules pouvant interagir avec ces protéines et d'inhiber leur interaction. Le but principal étant de trouver sur le long terme un traitement efficace contre des maladies atopiques.

Ici, ce stage s'ancre dans le projet en réalisant un criblage virtuel pour médier *in fine* la voie d'IL-3. Ceci afin de permettre l'identification de molécules pouvant interagir avec le complexe IL-3/IL-3R α . Pour le criblage, IL-3 est choisie car :

- Il s'agit d'un modèle plus simple à modéliser que son récepteur transmembranaire.
- Viser le récepteur avec un ligand pourrait provoquer une activation d'autres récepteurs de manière anarchique s'il pouvait interagir avec.

La protéine va tout d'abord être ciblée à l'aveugle. Le but étant de déterminer quels ligands vont interagir avec la zone d'interaction IL-3/IL-3R α en surface et lesquels ne le font pas. Ces derniers pourront alors être supprimés. Afin de réaliser un criblage correct, il est important de déterminer les potentielles conformations de la protéine étudiée. Si celle-ci en a plusieurs, les résultats du criblage ne seront pas les mêmes. Plus précisément pour ici, il faut déterminer les conformations sensiblement différentes dans la zone d'interaction.

2. Matériels et méthodes

2.1. Simulation de dynamique moléculaire

2.1.1. Obtention de la structure d'IL-3

Plusieurs structures d'IL-3 sont actuellement disponibles sur le site de la *Protein Data Bank* (PDB). Parmi celles-ci, la structure PDB 5UWC [14] a été retenue comme point de départ, car elle a une meilleure résolution (2,40 Å) et le moins de défauts structuraux. Cette structure contient IL-3 en interaction avec son récepteur IL-3R α , donc la zone d'interaction entre les deux protéines est connue de manière structurale, comme indiqué sur la [fig. 2]. On notera particulièrement les résidus d'intérêts PHE-56 et LYS-135 (mutée en TRP). Le logiciel MODELLER [15–18] (9.16) a été utilisé pour compléter les fragments structuraux manquants sur la protéine au niveau du N-ter (11 résidus). Ensuite, pour protoner la structure, le logiciel H++ [19–21] (4.0) a été utilisé pour un pH de 7,0.

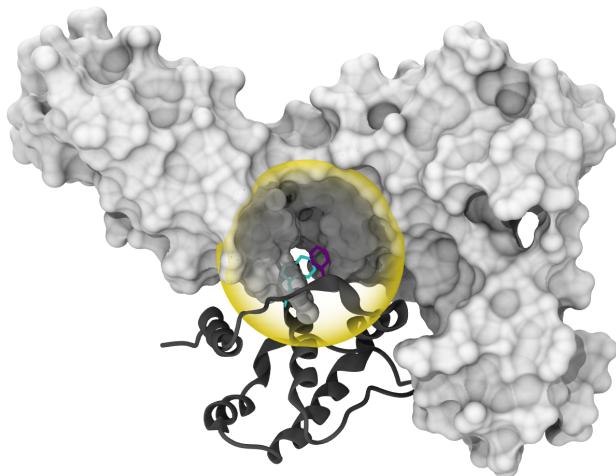


Fig. 2 – Représentation du récepteur IL-3 et de son récepteur IL-3R α , issu du code PDB 5UWC. Jaune = zone d'interaction IL-3/IL-3R α ; surface blanche = récepteur; cartoon noir = IL-3; violet = PHE-35 annoté ainsi dans le PDB, équivaut ici à la PHE-56; vert = TRP-116, résidu muté et annoté ainsi dans le PDB, équivaut ici à la LYS-135.

2.1.2. Principe de la simulation de dynamique moléculaire

Le but d'une simulation de dynamique moléculaire est de déterminer le mouvement d'une protéine à partir des équations du mouvement. Pour cela, on se base sur la seconde loi de Newton :

$$\vec{F} = \vec{a} \times m$$

C'est à partir de cette équation qu'il est possible de réaliser la simulation d'une dynamique moléculaire d'une protéine. Pour qu'elle donne des résultats corrects et interprétables, il faut que la structure d'origine ainsi que le système soient bien préparés. Pour ce faire, il faut obtenir une structure 3D sûre, protoner correctement la protéine puis minimiser et thermaliser le système afin d'initier la simulation.

Une fois celle-ci terminée, des conformations de la protéine sont trouvées et extraites, le cas échéant, si elle concerne la zone d’interaction au récepteur.

2.1.3. Protocole de simulation

Pour réaliser les simulations de dynamique moléculaire, le logiciel GROMACS [22] (2020.4) a été utilisé. Le champ de force charmm36 [23] et le modèle d’eau TIP3P [24] ont été utilisés lors de la création de la topologie. La protéine a été mise dans une boîte cubique d’eau avec une distance protéine/boîte de 14 Å et un ion Cl afin de neutraliser les charges. Le protocole général pour la simulation est donné dans l’annexe [III]. L’algorithme de Berendsen est utilisé pour stabiliser la température et la pression. Les étapes d’équilibration, d’abord en NVT puis en NPT, se font sous contrainte appliquée aux liaisons entre atomes lourds. Ces dernières sont diminuées au cours de l’équilibration en NPT, jusqu’à être totalement retirées. L’algorithme LINCS [25] est utilisé pendant toute la simulation pour contraindre toutes les liaisons atomes lourds/hydrogène. Trois répliques ont été fait en tout, d’une durée de 400 ns chacun.

2.2. Amarrage moléculaire

Une fois la simulation réalisée et l’obtention ou non de conformations, un amarrage moléculaire est réalisé sur la protéine, à l’aveugle. L’objectif est de chercher où un potentiel ligand peut se fixer sur la protéine ; en plus de la zone d’interaction d’IL-3 avec son récepteur. Pour cela, des logiciels employant des algorithmes de minimisations et de recherche sont utilisés. La protéine est placée dans une grille dans laquelle un logiciel va tenter de placer des ligands en prenant la meilleure énergie d’interaction protéine/ligand.

Une chimiothèque est à l’origine une collection de composés chimiques afin de les conserver et de les répertorier de manière claire. Il existe une version informatique de celles-ci, jouant le rôle de base de données. Pour ces dernières, on peut noter la ZINC avec 1 825 113 472 molécules [26–28], la DUD-e avec 22 886 decoys [29,30], la CSD avec 995 907 molécules [31,32] ou encore la ChEMBL avec 2 157 379 ligands [33,34].

Ici, c’est la Chimiothèque Nationale (CN) française qui sera utilisée (80 000 ligands) [35–42]. Plus précisément, une « sous-base de données », la Chimiothèque Nationale Essentielle (CNE) qui contient 1 040 ligands. Même si elle est plus petite que la Chimiothèque Nationale, elle représente sa diversité structurale et son espace chimique. Le but est de regarder sur un petit échantillon représentatif les résultats de l’amarrage afin de réaliser des filtres pour accélérer les procédures sur des plus grosses bases de données. Également, elle servira pour réaliser plusieurs tests sur différents logiciels.

Avant l’amarrage, GYPSUM-DL [43] est utilisé pour préparer les ligands (en retirant les artefacts, en protonant les ligands, en générant différents tautomères, en formant des centres chiraux et des isomères CIS-TRANS et en générant les conformations des cycles). À la fin, 4 378 conformations des ligands ont été obtenues, sachant que trois ligands aberrants ont été supprimés [III]. Ils ont été repérés

par un script personnel et une inspection visuelle.

Pour faire l'amarrage moléculaire, plusieurs logiciels sont employés :

1. Autodock VINA 2009 [44].
2. Smina [45], qui est basé sur le précédent. Il optimise certains paramètres et poids (ceux-ci ont été optimisés par des algorithmes d'apprentissages).
3. PLANTS [46], qui utilise un algorithme de minimisation basé sur l'organisation des colonies de fourmis.
4. Autodock VINA 2021 [47] (1.2.0), une nouvelle version de ce premier avec un *forks* Python. Il est utilisé avec la fonction de score VINARDO [48].
5. rXDOCK [49], qui détermine lui-même les cavités où placer les ligands.

Ceux-ci ont différents algorithmes et méthodes de recherche et de minimisation. De part ces différences, nous avons voulu les tester sur notre système d'étude. Ceci afin d'étudier l'influence de leur fonction de score sur le classement du jeu de ligands. Ceci sera utile afin d'harmoniser les classements de chacun de ces logiciels via un *rescoring* final.

2.3. Logiciels utilisés

Pour faire les graphiques et les jeux de données, R [50,51] (4.0.3) a été utilisé. Pymol [52–54] (2.3.0) et VMD [55,56] (1.9.3) permettent de visualiser les molécules. Python [57] (3) est utilisé pour réaliser les scripts. Pour convertir les molécules en différents formats, Open Babel [58] a été utilisé.

3. Résultats

3.1. Simulation de dynamique moléculaire

En visualisant les trois répliques des simulations, la forme globale de la protéine reste plutôt stable, surtout au niveau des structures secondaires. En les analysant, le RMSD entre la première et le reste des *frame* de la production est assez élevé ($> 3 \text{ \AA}$) [fig. 3 : A]. Toutefois, en analysant les RMSF, cela est du aux parties N- et C-terminales [fig. 3 : B].

Au niveau de la zone d'interaction d'IL-3 avec IL-3R α , une des simulations montre un changement particulier au niveau de cette zone. En effet, la LYS-135 interagit avec la PHE-56 [fig. 4 : B–C]. Ensuite, il y a un état transitoire qui mène ensuite à une conformation ouverte [fig. 4 : D–E][fig. 3 : C]. Mis à part ce cas, il n'y a pas d'autres changements conformationnels assez importants au niveau de la cavité ou de la protéine pour en extraire d'autres. En effet, les autres répliques ne présentent pas de changement de la zone d'interaction [V]. Il y a donc deux conformations au niveau de la zone d'interaction qui sont extraites et utilisées pour l'amarrage moléculaire.

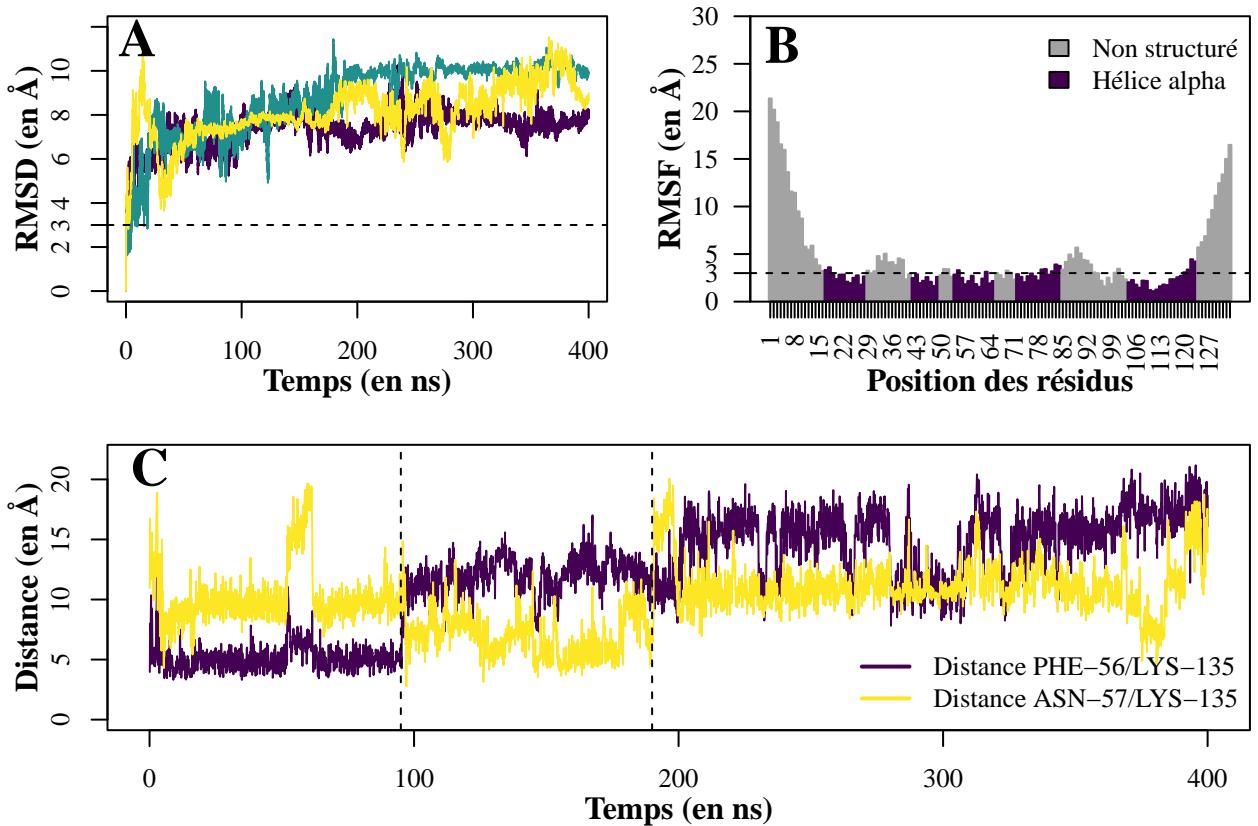


Fig. 3 – A = RMSD des différents répliques. / B = RMSF par résidus des trois répliques concaténés. / Distance PHE-56/LYS-135 et ASN-57/LYS-135 par rapport au temps, pour la 1^{re} simulation. Chaque trait vertical sépare un type de conformation (respectivement à gauche puis à droite : fermée puis ouverte).

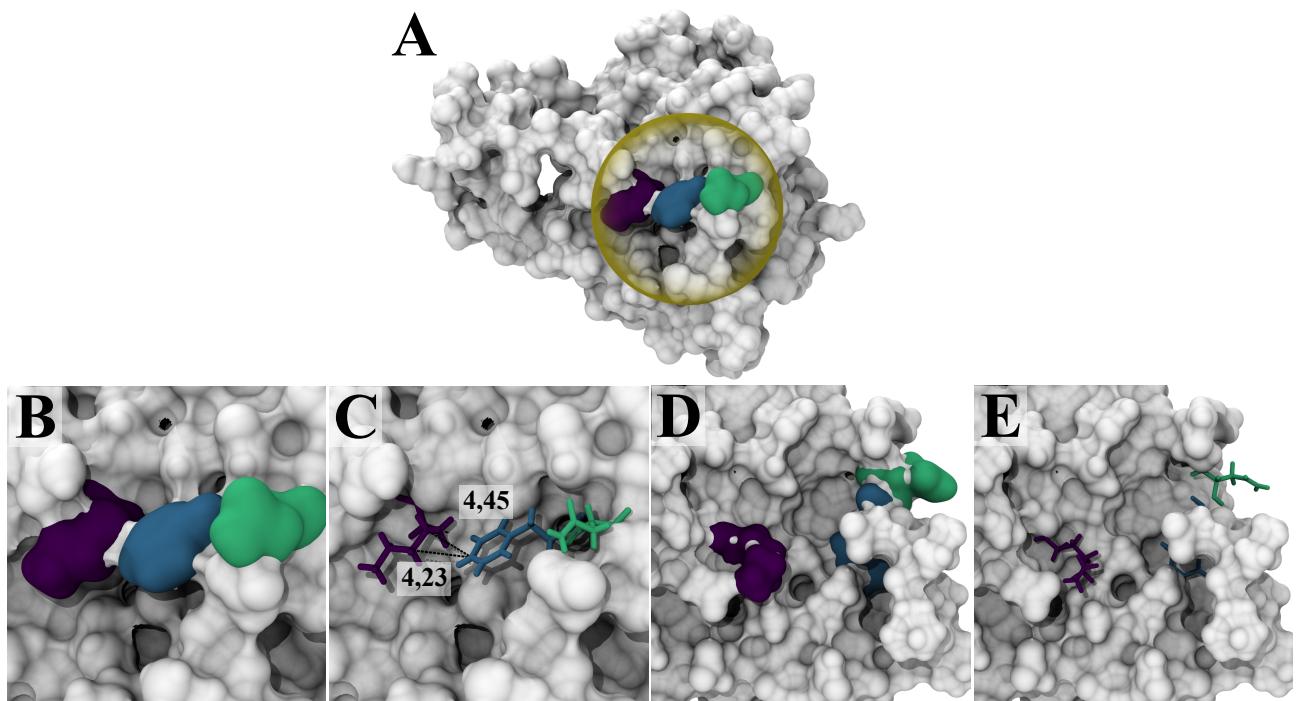


Fig. 4 – Représentation des deux conformations échantillonées de la zone d’interaction IL-3/IL-3R α . Bleu = PHE-56; vert = ASN-57; violet = LYS-135. / A = La protéine globale, conformation fermée. La zone encerclée en jaune est celle sur laquelle le zoom est effectué : il s’agit de la zone d’interaction IL-3/IL-3R α . / B = Conformation fermée représentée en surface. Ici, la PHE-56 est proche de la LYS-135. / C = Conformation fermée représentée en surface et licorice. / D = Conformation ouverte représentée en surface. Ici, la LYS-135 est éloignée du résidu précédemment cité, provoquant une ouverture de la cavité. / E = Conformation ouverte représentée en surface et licorice.

3.2. Amarrage moléculaire

Tout d'abord, certains ligands n'ont pas pu être amarrés correctement pour certains logiciels, car tous les atomes ne sont pas présents dans leur champ de force. C'est le cas, entre autres, des ligands avec des atomes de brome (B) ou d'étain (Sn) qui n'ont pas pu être amarrés avec Autodock VINA 2021.

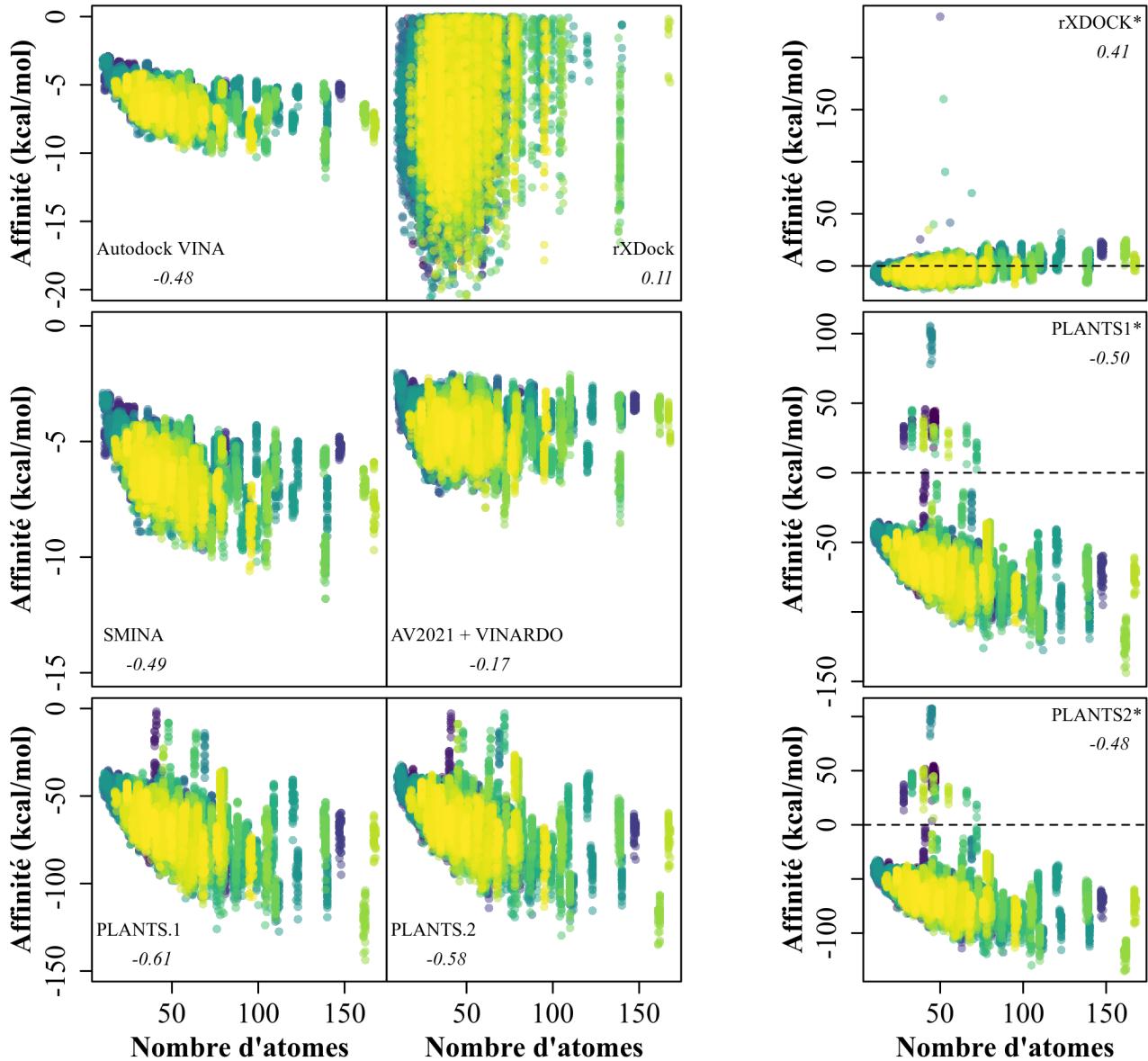


Fig. 5 – Graphiques des scores obtenus, pour différents logiciels d'amarrage moléculaire, en fonction de la taille des ligands. Le coefficient de corrélation r a été calculé sur toutes les valeurs de score inférieures ou égales à 0, SAUF pour les méthodes marquées avec une * (colonne de droite). La coloration a été faite selon l'ID des ligands.

Suite à un premier amarrage sur la conformation fermée (structure initiale), des résultats d'énergies d'interactions sont obtenus pour différents logiciels. Afin de sélectionner et de comparer les différents logiciels, une première analyse a été faite : la relation du score en fonction de la taille de la molécule en nombre d'atomes [fig. 5]. Le but est de vérifier s'il y a une corrélation entre ces deux paramètres, mais également de trouver une bonne fonction pour faire le *rescoring*. Il s'avère qu'à part pour Autodock VINA 2021 avec la fonction de score VINARDO, tous les autres résultats

montrent une corrélation score/taille plus ou moins forte. Ainsi, un *rescoring* a été effectué avec la fonction de score VINARDO pour se débarrasser de cette corrélation et avoir des résultats comparables. En comparant ces derniers, seuls les logiciels PLANTS et Autodock VINA 2021 donnent des résultats avec les meilleurs scores et des ligands dans la zone d'interaction IL-3/IL-3R α . Toutefois, PLANTS donne aussi des résultats d'énergies d'interactions positifs à cause d'encombrements stériques [IV]. Aucune explication sur la génération de ces poses n'a été trouvée.

Suite à ces résultats, le 2^e amarrage a été fait, en utilisant seulement PLANTS avec un *rescoring* via VINARDO et Autodock VINA 2021, sur la conformation ouverte obtenue par simulation de dynamique moléculaire.

Après avoir fait ces deux amarrages, les résultats les plus intéressants sont ceux donnés par Autodock VINA 2021 pour la conformation fermée, puisque ce sont des ligands petits [fig. 6]. En effet, même si PLANTS donne de bons résultats de ligands amarrés dans la cavité, ceux-ci sont gros (type hème) ou longs (type chaîne aliphatique). En plus de tout cela, grâce à l'amarrage à l'aveugle réalisé, un jeu de données de ligands amarrés sur toute la protéine a été obtenu [tab. 1].

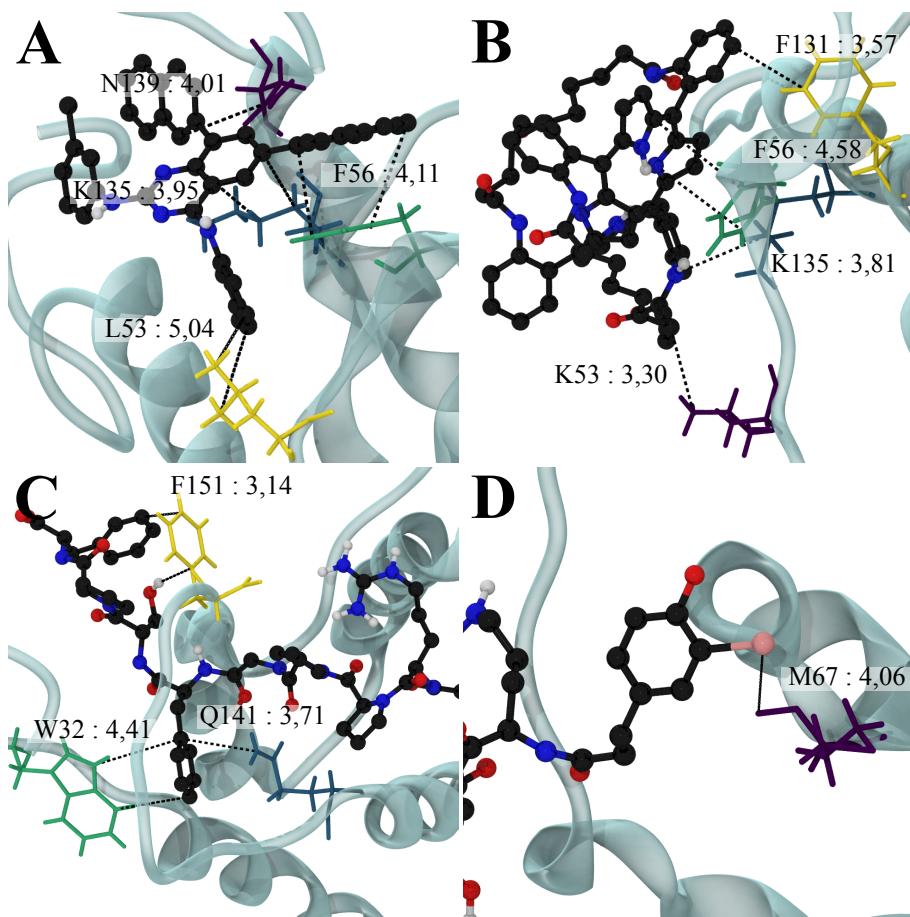


Fig. 6 – Représentation de quelques résultats de l'amarrage, le top des poses. Noter que les étiquettes indiquent : NOM DU RÉSIDU-position dans la séquence: distance (en Å) entre un atome du résidu et un atome du ligand. / **A = Ligand 3 647.** Cette pose a été obtenue via Autodock VINA 2021 et est dans la zone d'interaction IL-3/IL-3R α . / **B = Ligand 3 198.** Cette pose a été obtenue via Autodock VINA 2021 et est dans la zone d'interaction IL-3/IL-3R α . Ce ligand étant beaucoup trop gros, celui-ci pourrait ne pas être un bon ligand. / **C-D = Ligand 2 279.** Cette pose a été obtenue par Autodock VINA 2009, mais n'est pas dans la cavité. Ils s'agit d'une top pose après *rescoring*. Ce ligand est trop gros et n'est pas dans la zone d'interaction IL-3/IL-3R α . De plus, il interagit avec les N- et C-ter qui sont très instable. Ainsi, cette pose est plutôt mauvaise.

Par exemple, le ligand 3 647 interagit dans la zone d’interaction IL-3/IL-3R α [fig. 6 : A]. Il interagit entre autre avec la LYS-135 et la PHE-56 qui sont responsables dans cette interaction, comme montrée dans une figure précédente [fig. 2]. Ainsi, cela montre que ce ligand fait des interactions plutôt intéressantes avec notre cible. Il réalise surtout des interactions de type hydrophobe avec IL-3.

Ce faisant, à partir des résultats obtenus, il sera possible pour la suite de faire des filtres pour garder ou supprimer des ligands en fonction de leurs propriétés physico-chimiques (taille, masse molaire, charges...) par rapport à leur zone d’interaction. Ceci est actuellement en cours de développement via C++11 et sera grandement utile pour le criblage virtuel sur de grandes chimiothèques, telle que la ZINC.

À noter que certains ligands partent interagir avec les domaines N- et C-ter qui sont très flexibles, comme vu auparavant avec le RMSF. Ainsi, ces interactions sont à négliger puisqu'il est très peu probable que ces ligands puissent interagir avec ces domaines.

Tab. 1 – Tableau du jeu de données produit lors de l’amarrage moléculaire. Il s’agit d’un extrait qui illustre le jeu. Ici, huit exemples de ligands dans la zone d’interaction IL-3/IL-3R α pour la conformation fermée sont donnés.

**RÉSULTAT AVANT RESCORING, DANS LA ZONE
D’INTERACTION IL-3/IL-3R α**

ID	Nombre d’atomes	Pose	Score	Rang	Logiciel d’AM*
3647	72	1	-8.599	1	Autodock VINA 2021
3647	72	3	-8.143	5	Autodock VINA 2021
3647	72	5	-8.084	7	Autodock VINA 2021
3647	72	6	-8.068	8	Autodock VINA 2021
3647	72	7	-7.865	11	Autodock VINA 2021
3198	139	2	-7.692	16	Autodock VINA 2021
3651	73	2	-7.594	17	Autodock VINA 2021
3648	73	1	-7.529	18	Autodock VINA 2021

**RÉSULTAT APRÈS RESCORING, HORS DE LA ZONE
D’INTERACTION IL-3/IL-3R α**

2279	140	1	-11.631	1	Autodock VINA 2009
3686	161	2	-11.386	2	Autodock VINA 2009
4275	96	1	-11.361	3	SMINA
3688	161	2	-11.288	4	Autodock VINA 2009
3687	162	8	-11.049	5	Autodock VINA 2009
2278	140	1	-11.014	6	SMINA
4277	96	2	-10.000	7	SMINA
3131	104	1	-10.969	8	Autodock VINA 2021

*AM = amarrage moléculaire.

4. Conclusion

Pendant ce stage, nous avons pu étudier et bien comprendre la protéine IL-3, au travers d’expériences de bio-informatique structurale. Suite à des simulations de dynamique moléculaire, une nouvelle conformation de la protéine a pu être obtenue, mettant en jeu un réarrangement de la région d’in-

teraction avec IL-3R α . Ensuite, les différents logiciels ainsi que leur fonction de score ont été étudiés pour trouver celui avec la meilleure efficacité, ici Autodock VINA 2021 avec la fonction de score VINARDO. Enfin, des premiers résultats de l'amarrage moléculaire ont pu être analysés et s'avèrent intéressants pour la suite du projet.

Tout d'abord, il faut obtenir et compiler les informations de l'amarrage pour pouvoir réaliser les dits filtres pour exploiter de plus grosses chimiothèques. Par exemple, un programme en C++11 est en train d'être écrit pour obtenir des informations sur les types d'interactions des ligands selon la zone d'interaction mise en jeux sur la protéine. Ensuite, il faut ré-amarrer les ligands de manière plus précise (top 50 dans la conformation) puis réaliser des simulations de dynamique moléculaire afin de voir s'ils se maintiennent dans la zone d'interaction IL-3/IL-3R α . Enfin, si des ligands ne sont finalement pas stables dans leur interaction, il faudra refaire ces manipulations à partir de plus grosses chimiothèques.

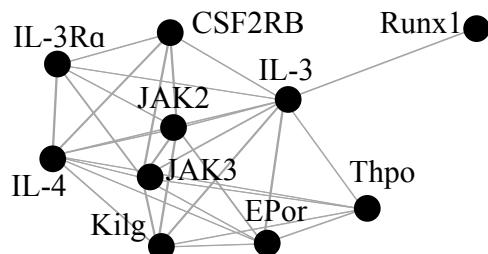
Les résultats obtenus lors de ce stage sont néanmoins prometteurs.

5. Bibliographie

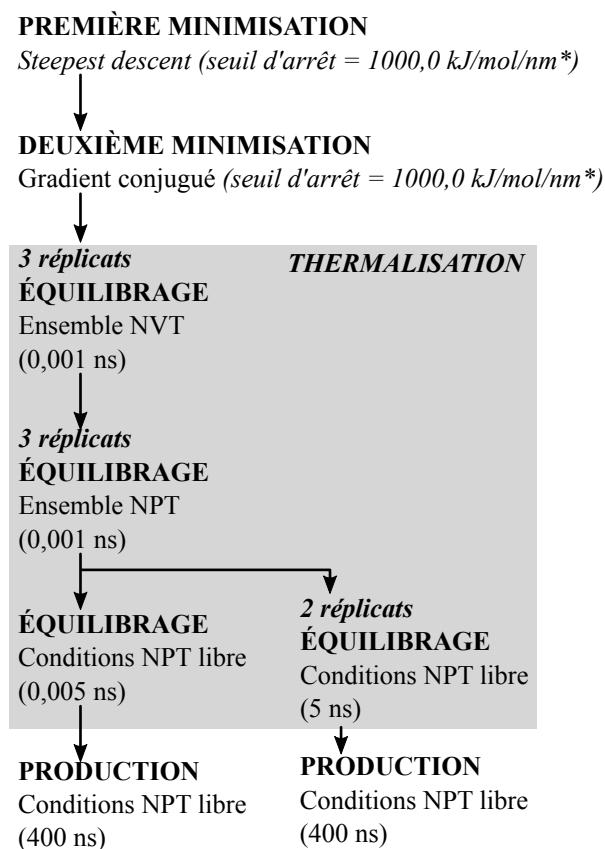
- [1] Balkwill, F. & Burke, F. The cytokine network **10**, 299–304. URL <https://linkinghub.elsevier.com/retrieve/pii/0167569989900856>.
- [2] Reddy, E. P., Korapati, A., Chaturvedi, P. & Rane, S. IL-3 signaling and the role of Src kinases, JAKs and STATs : A covert liaison unveiled **19**, 2532–2547. URL <https://www.nature.com/articles/1203594>.
- [3] IL3 interleukin 3 [Homo sapiens (human)] - Gene - NCBI. URL <https://www.ncbi.nlm.nih.gov/gene/3562>.
- [4] Mizel, S. B. The interleukins **1** **3**, 2379–2388. URL <https://onlinelibrary.wiley.com/doi/10.1096/fasebj.3.12.2676681>.
- [5] Le Beau, M. M. *et al.* The interleukin 3 gene is located on human chromosome 5 and is deleted in myeloid leukemias with a deletion of 5q. **84**, 5913–5917. URL <https://pnas.org/doi/full/10.1073/pnas.84.16.5913>.
- [6] Schrader, J. W. (ed.) *Interleukin 3 : The Panspecific Hemopoietin*. No. 15 in Lymphokines (Academic Press).
- [7] Reactome | IL3 [extracellular region]. URL <https://reactome.org/content/detail/R-HSA-450059>.
- [8] Robin, C. *et al.* An Unexpected Role for IL-3 in the Embryonic Development of Hematopoietic Stem Cells **11**, 171–180. URL <https://linkinghub.elsevier.com/retrieve/pii/S1534580706003030>.
- [9] The UniProt Consortium *et al.* UniProt : The universal protein knowledgebase in 2021 **49**, D480–D489. URL <https://academic.oup.com/nar/article/49/D1/D480/6006196>.
- [10] Esnault, S. *et al.* IL-3 Maintains Activation of the p90S6K/RPS6 Pathway and Increases Translation in Human Eosinophils **195**, 2529–2539. URL <http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1500871>.
- [11] IL3RA interleukin 3 receptor subunit alpha [Homo sapiens (human)] - Gene - NCBI. URL https://www.ncbi.nlm.nih.gov/gene?cmd=retrieve&dopt=default&list_uids=3563&rn=1.
- [12] Seattle Genetics Initiates Phase 1 Trial of SGN-CD123A for Patients with Relapsed or Refractory Acute Myeloid Leukemia. URL <https://www.businesswire.com/news/home/20160919005140/en/Seattle-Genetics-Initiates-Phase-1-Trial-of-SGN-CD123A-for-Patients-with-Relapsed-or-Refractory-Acute-Myeloid-Leukemia>.
- [13] SGN-CD123A Drug Description » ADC Review. URL <https://www.adcreview.com/sgn-cd123a-drug-description/>.
- [14] Broughton, S. E. *et al.* A dual role for the N-terminal domain of the IL-3 receptor in cell signalling **9**, 386. URL <http://www.nature.com/articles/s41467-017-02633-7>.
- [15] Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER **54**. URL <https://onlinelibrary.wiley.com/doi/10.1002/cpb.3>.
- [16] Martí-Renom, M. A. *et al.* Comparative Protein Structure Modeling of Genes and Genomes **29**, 291–325. URL <https://www.annualreviews.org/doi/10.1146/annurev.biophys.29.1.291>.
- [17] Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints **234**, 779–815. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022283683716268>.
- [18] Fiser, A., Do, R. K. G. & Šali, A. Modeling of loops in protein structures **9**, 1753–1773. URL <https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.9.9.1753>.
- [19] Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0 : Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations **40**, W537–W541. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks375>.
- [20] Myers, J., Grothaus, G., Narayanan, S. & Onufriev, A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules **63**, 928–938. URL <https://onlinelibrary.wiley.com/doi/10.1002/prot.20922>.
- [21] Gordon, J. C. *et al.* H++ : A server for estimating pKas and adding missing hydrogens to macromolecules **33**, W368–W371. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki464>.
- [22] Bauer, P., Hess, B. & Lindahl, E. GROMACS 2022.1 Manual URL <https://zenodo.org/record/6451567>.
- [23] Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field : Validation based on comparison to NMR data **34**, 2135–2145. URL <https://onlinelibrary.wiley.com/doi/10.1002/jcc.23354>.
- [24] Boonstra, S., Onck, P. R. & van der Giessen, E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State **120**, 3692–3698. URL <https://pubs.acs.org/doi/10.1021/acs.jpcb.6b01316>.
- [25] Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS : A linear constraint solver for molecular simulations **18**, 1463–1472. URL [https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- [26] ZINC. URL <https://zinc15.docking.org/>.
- [27] Welcome to ZINC Is Not Commercial - A database of commercially-available compounds. URL <https://zinc12.docking.org/>.

- [28] Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC : A Free Tool to Discover Chemistry for Biology **52**, 1757–1768. URL <https://pubs.acs.org/doi/10.1021/ci3001277>.
- [29] DUD-E : A Database of Useful (Docking) Decoys — Enhanced. URL <http://dude.docking.org/>.
- [30] Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E) : Better Ligands and Decoys for Better Benchmarking **55**, 6582–6594. URL <https://pubs.acs.org/doi/10.1021/jm300687e>.
- [31] Sevryugina, Y. Research Guides : Cambridge Structural Database : CSD : Access Cambridge Structural Database. URL <https://guides.lib.umich.edu/c.php?g=283163&p=8352036>.
- [32] Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database **72**, 171–179. URL <http://scripts.iucr.org/cgi-bin/paper?S2052520616003954>.
- [33] ChEMBL Database. URL <https://www.ebi.ac.uk/chembl/>.
- [34] Gaulton, A. *et al.* ChEMBL : A large-scale bioactivity database for drug discovery **40**, D1100–D1107. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr777>.
- [35] ChemBioFrance - Infrastructure de recherche. URL <https://chembiofrance.cn.cnrs.fr/fr/composante/chimiotheque>.
- [36] Douguet, D. Data Sets Representative of the Structures and Experimental Properties of FDA-Approved Drugs **9**, 204–209. [29541361](#).
- [37] Carles, F., Bourg, S., Meyer, C. & Bonnet, P. PKIDB : A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials **23**, 908. URL <http://www.mdpi.com/1420-3049/23/4/908>.
- [38] Desaphy, J., Bret, G., Rognan, D. & Kellenberger, E. Sc-PDB : A 3D-database of ligandable binding sites—10 years on **43**, D399–D404. URL <http://academic.oup.com/nar/article/43/D1/D399/2439494/scPDB-a-3Ddatabase-of-ligandable-binding-sites10>.
- [39] Basse, M.-J., Betzi, S., Morelli, X. & Roche, P. 2P2Idb v2 : Update of a structural database dedicated to orthosteric modulation of protein–protein interactions **2016**, baw007. URL <https://academic.oup.com/database/article-lookup/doi/10.1093/database/baw007>.
- [40] Torchet, R. *et al.* The iPPi-DB initiative : A community-centered database of protein–protein interaction modulators **37**, 89–96. URL <https://academic.oup.com/bioinformatics/article/37/1/89/6069541>.
- [41] Pupin, M. *et al.* Norine : A powerful resource for novel nonribosomal peptide discovery **1**, 89–94. URL <https://linkinghub.elsevier.com/retrieve/pii/S2405805X15300144>.
- [42] Rey, J., Deschavanne, P. & Tuffery, P. BactPepDB : A database of predicted peptides from a exhaustive survey of complete prokaryote genomes **2014**, bau106–bau106. URL <https://academic.oup.com/atabase/article-lookup/doi/10.1093/database/bau106>.
- [43] Ropp, P. J. *et al.* Gypsum-DL : An open-source program for preparing small-molecule libraries for structure-based virtual screening **11**, 34. URL <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-019-0358-3>.
- [44] Trott, O. & Olson, A. J. AutoDock Vina : Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading NA–NA. URL <https://onlinelibrary.wiley.com/doi/10.1002/jcc.21334>.
- [45] Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise **53**, 1893–1904. URL <https://pubs.acs.org/doi/10.1021/ci300604z>.
- [46] Korb, O., Stützle, T. & Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS **49**, 84–96. URL <https://pubs.acs.org/doi/10.1021/ci800298z>.
- [47] Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0 : New Docking Methods, Expanded Force Field, and Python Bindings **61**, 3891–3898. URL <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00203>.
- [48] Quiroga, R. & Villarreal, M. A. Vinardo : A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening **11**, e0155183. URL <https://dx.plos.org/10.1371/journal.pone.0155183>.
- [49] Li, L., Chen, R. & Weng, Z. RDOCK : Refinement of rigid-body protein docking predictions **53**, 693–707. URL <https://onlinelibrary.wiley.com/doi/10.1002/prot.10460>.
- [50] R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. URL <https://www.R-project.org/>.
- [51] Osorio, D., Rondon-Villarreal, P. & Torres, R. Peptides : A package for data mining of antimicrobial peptides **7**, 4–14.
- [52] Schrödinger, LLC. The AxPyMOL molecular graphics plugin for Microsoft PowerPoint, version 1.8.
- [53] Schrödinger, LLC. The JyMOL molecular graphics development component, version 1.8.
- [54] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8.
- [55] Humphrey, W., Dalke, A. & Schulten, K. VMD : Visual molecular dynamics **14**, 33–38. URL <https://linkinghub.elsevier.com/retrieve/pii/0263785596000185>.
- [56] Decherchi, S. & Rocchia, W. A general and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale **8**, e59744. URL <https://dx.plos.org/10.1371/journal.pone.0059744>.
- [57] Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual (CreateSpace)*.
- [58] O’Boyle, N. M. *et al.* Open Babel : An open chemical toolbox **3**, 33. URL <https://jcheminf.biomedcentral.com/articles/10.1186/1758-2946-3-33>.
- [59] Il3 protein (*Rattus norvegicus*) - STRING interaction network. URL <https://string-db.org/network/10116.ENSRNOP00000030704>.

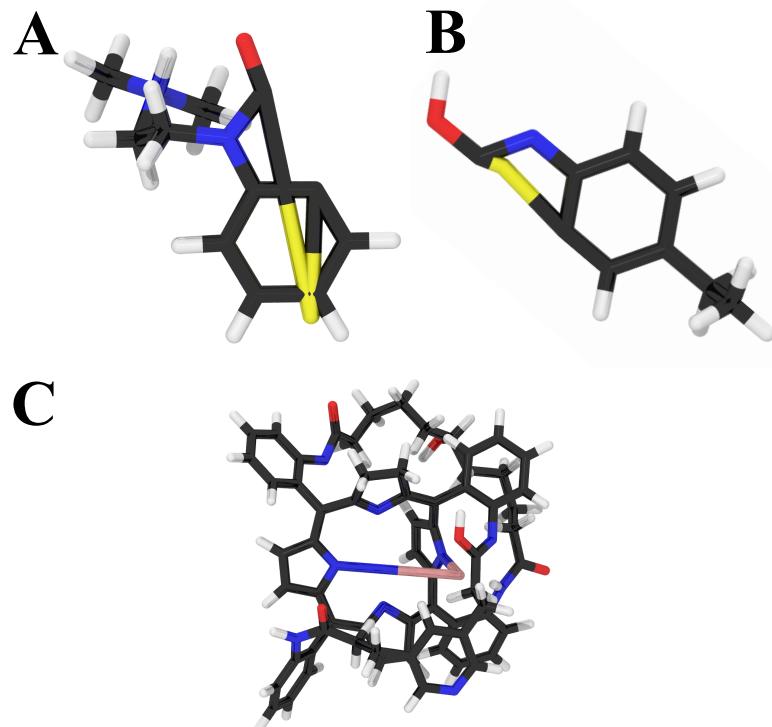
Annexes



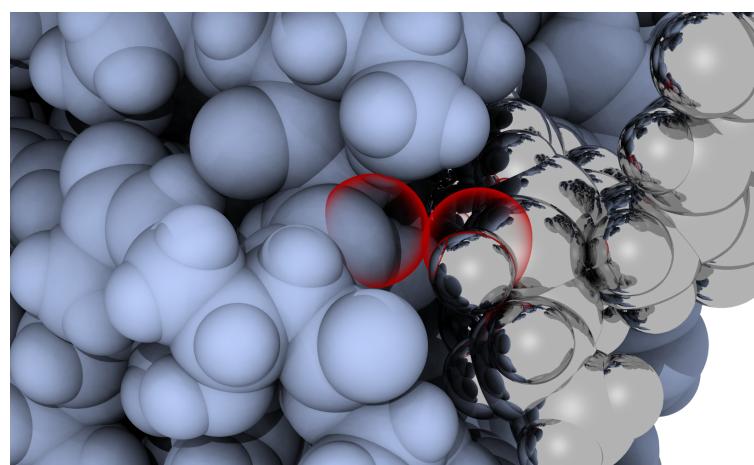
Annexe I – Réseau d’interactions ou *interactome* de la protéine IL-3. Le réseau a été réalisé via STRING [59]. Le score minimum d’interaction est défini sur « moyenne de confiance égale à 0,400 » ; seul une 1^{re} couche d’interaction, limitée à 10, a été définie.



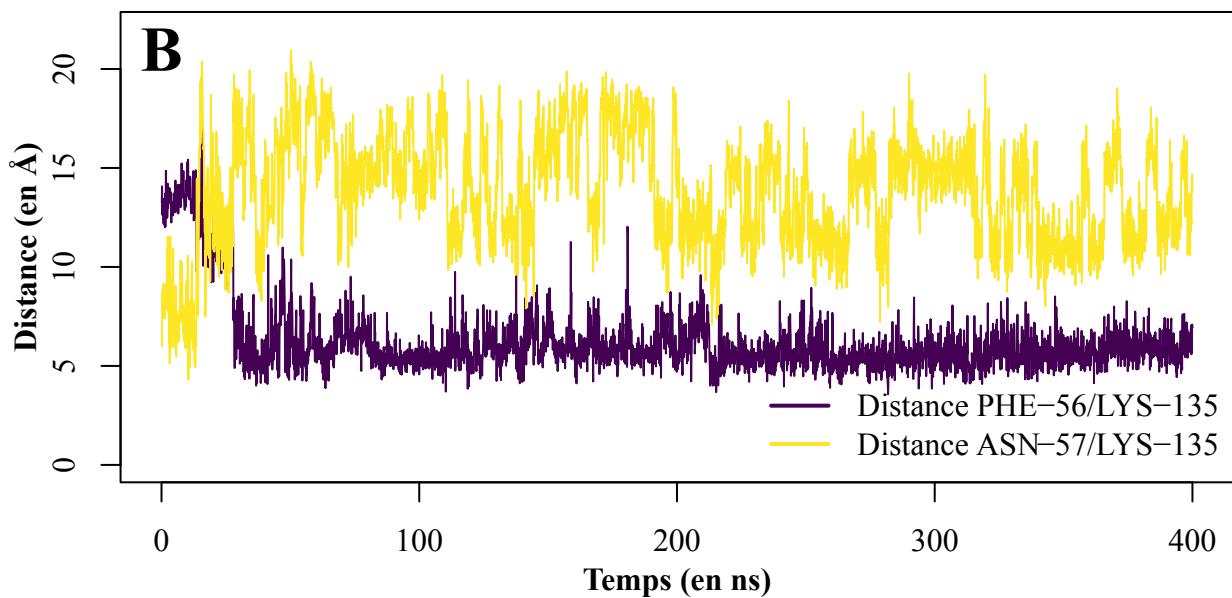
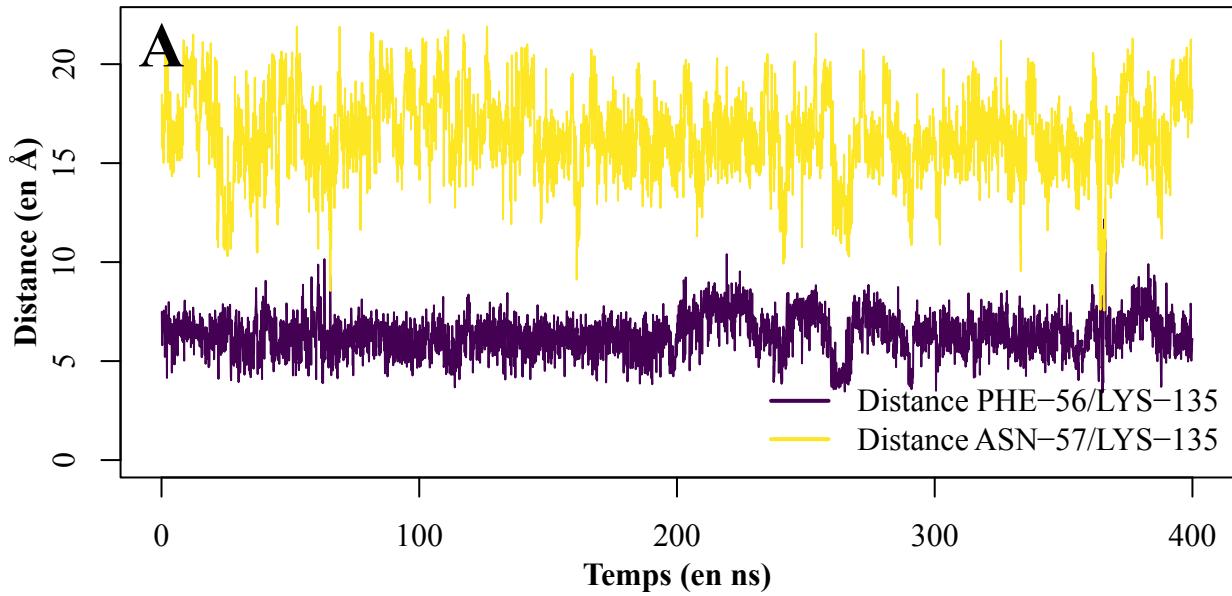
Annexe II – Logigramme du protocole appliqué pour la simulation de dynamique moléculaire. * = Seuil d’arrêt définit pour la minimisation qui correspond à la force maximale exercée sur un atome.



Annexe III – Aberrations générées par GYPSUM DL. A = Un atome Si a une liaison trop grande avec un clash stérique ; B = un atome Si a une liaison trop courte avec un clash stérique ; C = un atome de Fe a une liaison trop grande.



Annexe IV – Clash stérique entre deux atomes oxygène, après amarrage par PLANTS. rouge = oxygène ; métalique = ligand amarré ; bleu = IL-3. Ici, deux atomes oxygène ont leurs rayons atomiques qui s'interpénètrent, provoquant une augmentation de l'énergie d'interaction sortie par PLANTS.



Annexe V – A = Distance dans le cas du second réplica. / B = Distance dans le cas du troisième réplica. Dans chacun de ces réplicas, seule la conformation fermée est observable.

Résumé/Abstract

Les maladies atopiques sont des maladies particulières, provenant d'une très forte sensibilité de l'organisme à des allergènes extérieurs. Il a été montré que l'interleukine-3, une cytokine associée aux réponses inflammatoires, est dans ces cas précis surexprimée. Ainsi, dans le cadre de ce stage, nous nous intéressons à étudier en profondeur IL-3, ceci afin de cibler et inhiber cette protéine dans le cadre des maladies atopiques. Pour ce faire, deux méthodes de bio-informatique structurale seront mises en oeuvre afin d'atteindre notre objectif : la simulation de dynamique moléculaire et le criblage virtuel appliquée sur une chimiothèque précise, la Chimiothèque Nationale Essentielle (CNE). Les simulations ont mis en exergue deux conformations de la zone d'interaction d'intérêt sur IL-3 et le criblage virtuel a donc été réalisé sur ces deux conformations, tandis que les expériences d'amarrage moléculaire ont permis d'obtenir de premiers résultats intéressant pour poursuivre cette étude.

Atopics diseases are specific diseases induced by high sensitivity of the organism to external allergens. Studies have shown that INterleukin-3, a cytokin involved in inflammatory responses, is sur-expressed in those cases. By consequence, during this internship, we are interested to deeply study IL-3, in order to target and inhibit this protein in the case of atopic diseases. To do so, two structural bio-informatics methods are going to be carry on in order to achieve this objective : molecular dynamics simulations and virtual screening apply on a specific chemical library, the Chimiothèque Nationale Essentielle (CNE). Simulations have shed light on two conformations of the interaction area of IL-3 with its receptor, so virtual screening have been done on these two conformations, while docking experiments have permit to obtain first interesting results to continue further this study.