

Master 2 - Biologie Informatique

2021 - 2022

Projet Long Affimer

Théo Jamay



| | |
|---|-----------|
| I - Introduction | 2 |
| 1- Contexte | 2 |
| 2- Objectif | 2 |
| II - Matériel et Méthodes | 3 |
| 1- Démarche | 3 |
| 2- Antibody data ressource (AbDb) | 3 |
| 3- Binding Sites et surface NS1 | 4 |
| 4- PatchSearch | 6 |
| 5- Affimer | 6 |
| 6- Script d'automatisation | 7 |
| III - Résultats et discussion | 8 |
| 1- Binding sites et surface NS1 | 8 |
| 2- PatchSearch | 9 |
| 3- Sélection des binding sites les plus similaire à NS1 | 10 |
| 4- Sélection des boucles en contact | 10 |
| 5- Affimer | 11 |
| IV - Conclusion | 14 |
| V - Bibliographie | 14 |
| Annexe | 15 |

I - Introduction

1- Contexte

Les virus sont responsables d'un grand nombre de maladies plus ou moins graves et l'épidémie de Covid-19 a montré à quel point ils pouvaient avoir des conséquences dramatiques sur le plan sanitaire et économique [1]. Notre système immunitaire nous défend contre ces agents extérieurs, appelés antigènes, grâce à un ensemble complexe d'éléments de reconnaissance et de défense. Les anticorps sont un acteur essentiel de ce système, leur rôle est de reconnaître les antigènes afin de les neutraliser. Ils sont formés de deux chaînes lourdes (H) et deux chaînes légères (L) qui forment une structure en Y caractéristique. Ils possèdent également une partie variable, le paratope, qui reconnaît une zone bien spécifique de l'antigène, l'épitope. Les anticorps monoclonaux sont des anticorps ne reconnaissant qu'un seul type d'épitope sur un antigène donné.

Les anticorps monoclonaux sont largement utilisés comme outils de diagnostic et comme moyens thérapeutiques dans toute une série de maladies allant de l'infection au cancer [2]. On les retrouve par exemple dans les tests de grosses ou encore dans des techniques d'identification de protéines virales. Cependant ils présentent des défauts, par exemple malgré leur importante spécificité ils peuvent être amenés à interagir avec des composés indésirables et provoquer des erreurs de diagnostics. Leur production peut également s'avérer coûteuse, lente et soulever des problèmes d'éthique animale.

Ainsi les chercheurs ont développé une alternative aux anticorps, les aptamères peptidiques ou Affimer. A la manière d'un anticorps ils sont composés d'une partie constante et d'une partie variable. La partie variable est composée de courte séquence d'acides aminés incluse sous forme de boucle et ces boucles sont conçues pour interagir de manière spécifique avec la surface d'une protéine cible. Ils présentent certains avantages comparés aux anticorps, d'abord ils sont caractérisés par une haute spécificité mais également par une stabilité plus élevée [3], [4]. En termes de production, la création d'Affimer est moins chère, plus simple, plus rapide et ne demande pas d'animaux.

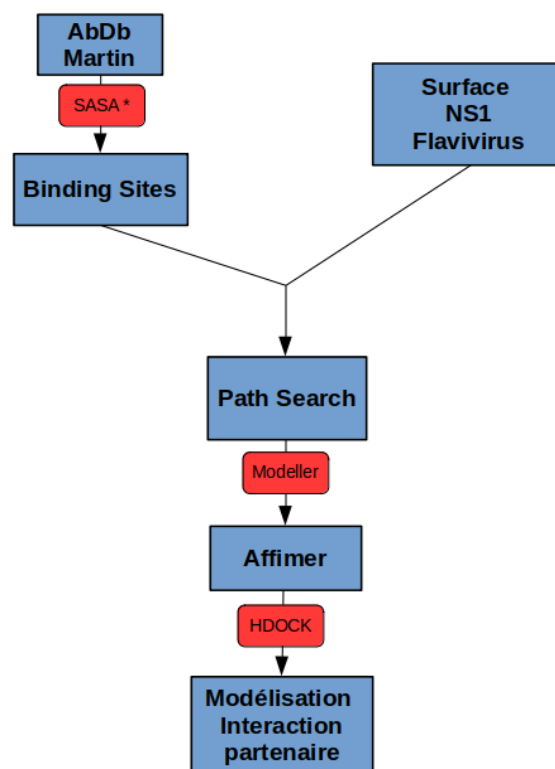
2- Objectif

Nous allons chercher à modéliser des Affimers dans le but de reconnaître spécifiquement des protéines de Flavivirus. Les Flavivirus, sont un groupe de virus qui comprend entre autres les virus de la Dengue, du Zika et Nil occidental (west line). Nous nous intéressons à ce groupe car la similitude de séquence entre ces différentes souches est étonnamment élevée rendant la caractérisation par anticorps du type d'infection difficile [5], [6]. Or du fait qu'il existe une grande diversité de virulence entre les souches, les conséquences d'une erreur diagnostique peuvent être lourdes. Les protéines de type non structural (NS) sont importantes pour le traitement et l'identification des flavivirus [7], [8].

Elles sont de par le schéma de virulence probablement les plus pertinentes pour l'identification du type de virus. L'objectif est donc de modéliser un Affimer capable d'interagir avec la protéine non structurale NS1. Notre objectif est d'également chercher à automatiser certaines étapes du processus de modélisation d'un Affimer grâce à des scripts Python.

II - Matériel et Méthodes

1- Démarche



SASA : Solvent-Accessible Surface Area

Figure 1 : Schéma explicatif de la démarche

Notre démarche globale, représentée schématiquement sur la figure 1, comprend la détermination des sites de liaison (binding sites) côté antigène pour l'ensemble des complexes anticorps-antigène (Ac/Ag) de l'AbDb. La recherche de binding sites similaire sur la surface de la protéine de flavivirus NS1 se fait grâce à PatchSearch. L'extraction des meilleurs binding sites afin de récupérer les boucles d'anticorps en interaction avec l'antigène permettra de modéliser des Affimers spécifiques de NS1.

2- Antibody data ressource (AbDb)

L'AbDb [9] est une base de données contenant des structures d'anticorps issus de la Protéine Data Bank (PDB) [10]. Les structures d'anticorps sont simplifiées pour ne conserver que les régions

variables (Fv). Les chercheurs se sont également basés sur la base de données SACS [11] qui contient une liste des structures d'anticorps issus de la PDB avec quelques annotations de base pour construire l'ABdB.

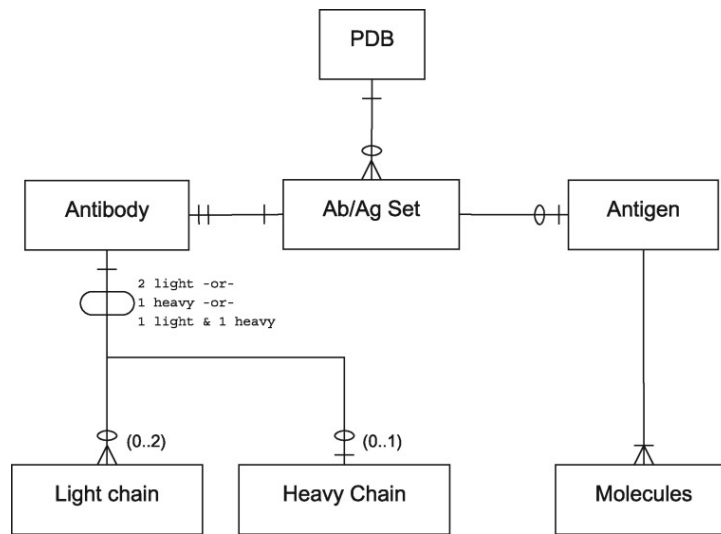


Figure 2 : Schéma organisation AbDb

On constate sur ce schéma qu'il existe globalement 3 jeux de données, un premier contenant les complexes Ac/Ag, ensuite les anticorps uniquement et enfin les antigènes. Il existe également 3 schéma de numérotation structurelle des régions variables d'anticorps disponibles pour l'AbDb : Kabat [12], Chothia [13] et enfin Martin [14]. Dans le cadre de notre projet, nous nous intéressons aux complexes Ac/Ag afin d'en extraire les binding sites côté antigène. Nous avons également choisi d'utiliser le schéma de Martin car il s'agit de la publication la plus récente et elle est présentée comme une amélioration des schéma précédent (Kabat).

3- Binding Sites et surface NS1

Lorsque nous parlons pour ce projet de binding site ou site de liaison / fixation ou encore de zone de contact côté antigène, cela représente les atomes de l'antigène en interaction avec l'anticorps. Afin de déterminer cette zone d'interaction ou zone de contact, nous passons par le calcul de l'accessibilité au solvant. On commence par calculer l'accessibilité au solvant du complexe puis ensuite de chaque chaîne du complexe de manière séparée. Les atomes dont l'accessibilité a changé entre les 2 étapes constituent notre zone de contact ou solvent-accessible surface area (SASA).



Figure 3 : Binding site de l'antigène avec de la chaîne lourde de l'anticorps pour le complexe 6DFJ

La figure 3 illustre le site de fixation côté antigène, on retrouve bien les atomes de l'antigène qui sont en interaction avec la chaîne lourde de l'anticorps. Nous avons choisi d'utiliser le module biopython Bio.PDB.SASA [15] afin de calculer l'accessibilité au solvant. Ce module utilise l'algorithme "rolling ball" développé par Shrake & Rupley [16]. Cet algorithme simule une sphère avec un rayon correspondant par défaut à une molécule d'eau (1.4 Å) pour scanner la surface d'une protéine. Nous avons choisi d'utiliser ce module pour sa complémentarité avec le module biopython Bio.PDB.Parser et nous avons vérifié sur le GitHub si des problèmes n'étaient pas signalés.

Nous l'avons vu précédemment en introduction, la protéine non structurale NS1 est une protéine clé pour la reconnaissance des flavivirus par le système immunitaire. Il s'agit d'une protéine fortement exprimée dans les cellules infectées par les flavivirus qui agit notamment en bloquant la synthèse cellulaire. Nous allons chercher dans un premier temps, s'il existe des structures de NS1 sans anticorps disponible dans la PDB. L'objectif est ensuite de rechercher sur la surface de NS1 une zone fortement similaire à un binding Ac/Ag pour un autre virus.

Nous allons également explorer l'AbDb pour voir les complexes Ag/Ac existant pour les flavivirus. L'objectif sera ensuite de récupérer la surface des antigènes de ces complexes pour chercher à voir si avec un Affimer nous pouvons avoir une interaction aussi spécifique que pour un anticorps.

4- PatchSearch

PatchSearch permet d'identifier des sites de fixation structurellement proches à la surface de protéines candidates. Ainsi nous allons utiliser PatchSearch [17] afin d'identifier des sites de fixation similaires à la banque de binding sites que nous avons créé à partir de l'AbDb sur la surface d'antigène de flavivirus. Cela va nous permettre d'identifier des anticorps potentiellement capables de se fixer sur la surface d'antigène de flavivirus. En effet, comme montré sur la figure 3, les binding sites similaires sur la surface de NS1 identifié par PatchSearch correspondent initialement à un anticorps.

Ainsi à partir du fichier de sortie PatchSearch, nous allons conserver uniquement les binding sites issus de l'AbDb fortement similaire à un site de fixation sur la surface de l'antigène de flavivirus cible. Pour cela nous allons nous baser sur un score de conservation des distances entre paires d'atomes correspondant appelé Distorsion score (D-score). Nous sélectionnons donc les anticorps correspondant aux binding sites avec un D-score élevé. Ensuite nous conservons uniquement les boucles des anticorps qui sont en contact direct avec l'antigène. L'objectif final étant d'incorporer ces boucles au sein de la région variable d'un Affimer.

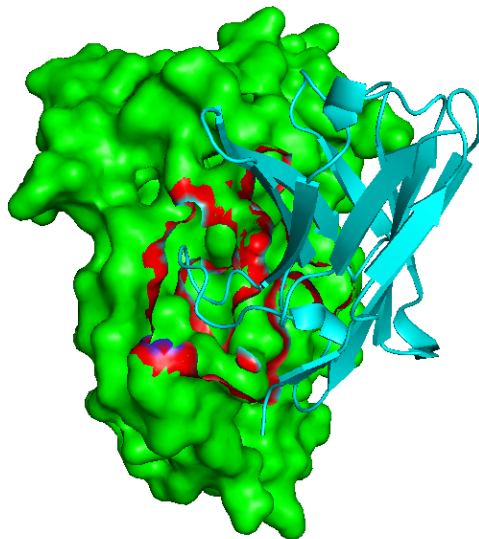


Figure 4 : En vert surface antigène, en rouge binding site (BS) côté antigène et en bleu la chaîne lourde de l'anticorps correspondant au BS

5- Affimer

Nous avons brièvement évoqué les Affimers ou aptamères peptidiques en introduction, il s'agit de protéine de petite taille, très stable capable de se lier spécifiquement à une protéine cible. Il existe des structures cristallographique de complexe Affimer / protéine cible sur la PDB.

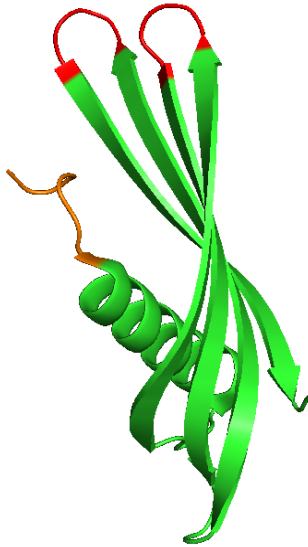


Figure 5 : Affimer récupérer dans le complexe 7NY8

Tout comme un anticorps, un affimer est composé d'une partie constante et d'une partie variable. La partie constante est composée d'une hélice alpha et d'un feuillet bêta de 4 brins représenté en vert sur la figure 5. La partie variable est composée de deux coudes (Bêta turn) représenté en rouge sur la figure 5 et de la séquence N-terminal représenté en orange sur la figure 5. Généralement l'interface d'interaction entre un Affimer et une protéine est de 12 et 36 acides aminés. Il peut ainsi s'agir d'une grande surface d'interaction et en conséquence un Affimer est potentiellement capable d'établir une interaction hautement spécifique et précise, tout à fait comparable avec celle d'un anticorps.

Nous allons chercher à modéliser un Affimer spécifique à NS1 en utilisant à l'outil Modeller [18]. Il s'agit d'un programme de modélisation par homologie pour produire des structures 3D à partir de support connu. Nous allons chercher à modéliser l'interaction avec l'Affimer modifié et la protéine NS1 en utilisant le serveur web HDock [19]. HDock est basé sur un algorithme hybride de modélisation (basée sur des supports) et d'interaction ab initio free. HDock utilise un Docking score qui est un score d'énergie, plus le Docking score est négatif plus l'interaction est "optimal" selon HDock.

6- Script d'automatisation

Nous allons chercher à automatiser certaines étapes de ce projet, tout d'abord nous allons développer un script qui convertit les complexes Ac/Ag de l'AbDb en binding sites côté antigène. Nous allons également automatiser le processus de conversion d'une protéine en surface protéique. Ainsi qu'écrire des scripts pour la sélection des binding sites hautement similaires à notre surface protéique d'intérêt NS1. Enfin nous allons développer un script qui retourne les séquences d'acides aminés consécutives

(+/- 5) de l'anticorps (chaîne H ou L) d'une taille supérieure à 4 acides aminés en contact avec l'antigène pour les meilleurs binding sites.

III - Résultats et discussion

1- Binding sites et surface NS1

Nous avons téléchargé les 3000 complexes Ac/Ag avec le schéma Martin disponible dans l'AbDb. Nous avons choisi de prendre la version redondante du jeu de données de complexe. L'AbDb compare chaque paire d'anticorps sur la base des résidus ATOM présents dans les deux séquences afin de savoir s'ils sont redondant. Nous avons fait ce choix pour permettre une vérification des résultats de match similaire fourni par PathSearch et également parce que s'il représente un problème il est toujours possible de les filtrer dans l'output.

Les complexes PDB peuvent contenir des structures anticorps multiples, par exemple pour les protéines homodimériques, on peut retrouver deux anticorps dans le complexe avec un anticorps liés à chaque sous-unité. Dans ce cas de figure, l'AbDb divise le complexe pour ne conserver qu'une structure d'anticorps est associée à son antigène. Par exemple, le complexe 3IYW qui est un homodimers lié à 2 anticorps (1 pour chaque sous-unité) est découpé en 2 complexes : 3IYW_1 et 3IYW_2.

Comme évoqué précédemment, un anticorps est constitué d'une chaîne lourde et d'une chaîne légère. Les chaînes sont certes liées entre elles, mais chaque chaîne possède une zone de contact spécifique et différente par rapport à l'autre. Séparer la zone de contact de la chaîne lourde et de la chaîne légère en deux fichier distincts augmente la probabilité de trouver une similarité forte avec PatchSearch. En effet, cela permet de rechercher une zone de contact spécifique au lieu de 2 zones de contact spécifique lié entre elles. Ainsi nous avons choisi de séparer les chaînes lourdes et légères, le résultat est que nous avons une zone de contact côté antigène pour chaque chaîne (soit deux au total). Par exemple, le complexe 5KVF_1 devient 5KVF_1E:L pour la zone de contact de l'antigène E avec la chaîne légère L et 5KVF_1E:H pour la zone de contact de l'antigène E avec la chaîne lourde H.

Il existe également des complexes Ac/Ag où l'antigène est divisé en plusieurs chaînes, dans ce cas de figure nous avons choisi de séparer les 2 chaînes. Il peut y avoir des complexes où la chaîne lourde interagit avec une sous-unité (chaîne) et la chaîne légère interagit avec une autre. Séparer les 2 chaînes permet de savoir précisément qui interagit avec qui et permet également de voir si les sous-unités qui n'ont pas d'interaction, de zone de contact. En reprenant l'antigène homodimérique 3IYW composé

d'une chaîne A et d'une chaîne B, on se retrouve avec 4 fichiers de sortie : 3IYW_1A:L, 3IYW_1A:H, 3IYW_1B:L et enfin 3IYW_1A:L.

Ainsi en résumé nous séparons les chaînes lourdes et légères pour l'anticorps et nous séparons les chaînes des antigènes multimériques. Ensuite nous avons cherché à faire des vérifications pour voir que notre script en python détermine correctement les zones de contact côté antigène. Pour cela, nous avons pour certaines protéines comparé la zone de contact basée sur la distance obtenue avec pymol via l'option "find any contact between chains within 4.0 Å". Nous avons également comparé avec les résultats d'accessibilité au solvant obtenus avec DSSP avec les résultats obtenus avec notre script. L'algorithme DSSP permet d'assigner des structures secondaires à partir d'un fichier PDB mais également de déterminer l'accessibilité au solvant. On passe finalement d'un jeu de données de 3000 complexes à un jeu de données de 6197 zones de contact ou binding sites côté antigène.

Nous avons après avoir recherché les structures de NS1 sans anticorps disponible dans la PDB sélectionné deux candidats. La structure 5GS6 qui est un variant NS1 de Zika identifié au Brésil et la structure 5K6K qui est également un variant NS1 de Zika mais identifié en Ouganda. Nous avons finalement sélectionné la structure 5GS6 car malgré qu'elle est une moins bonne résolution cristallographique que 5K6K, elle possède moins de résidus manquants dans son fichier PDB.

2- PatchSearch

Tout d'abord avant de lancer PatchSearch, il est nécessaire de "typer" les fichiers d'entrée. "Typer" signifie que chaque atome se voit attribuer une étiquette correspondant à son type : N, O, S, C et Ca (pour les carbones α). Les cycles aromatiques sont remplacés par un ou deux pseudo-atomes correspondant à un centroïde pour la phénylalanine ou la tyrosine et deux centroïdes pour le tryptophane. Les fichiers "typer" sont renommés avec un "t", par exemple 5KVF_1E:L.pdb devient 5KVF_1E:Lt.pdb. Nous avons donc commencé par typer l'ensemble de notre jeu de 6197 zones de contact ou binding sites issus de l'AbDb. Enfin nous avons "typé" la surface NS1.

PatchSearch nous fournit en sortie deux fichiers output, un fichier qui va contenir les informations comme le nombre d'atomes similaires retrouvés sur la totalité des atomes du binding site. On retrouve également la valeur de RMSD, le recouvrement ainsi que le D-score évoqué précédemment. Un deuxième fichier output qui est un fichier d'alignement où on retrouve les atomes qui ont été mis en relation (alignés) entre le support et la cible.

3- Sélection des binding sites les plus similaires à NS1

Nous l'avons dit précédemment nous allons utiliser le premier output contenant les informations générales afin de ne conserver que les binding sites les plus similaires à la surface protéique. Nous avons fait un petit script pour classer les binding sites par ordre croissant en fonction du D-score. Comme évoqué précédemment il s'agit d'un score de conservation des distances entre paires d'atomes correspondants. Dans le cas de la structure NS1 5GS6, nous avons choisi de sélectionner les binding sites avec un scores supérieur à 13 pour la chaîne A ou pour la chaîne B.

Un score supérieur à 13 est un très bon score de similarité au vu de la distribution des scores dans l'output. Généralement du fait qu'on scanne un antigène issu de l'AbDb contre 6197 binding sites issus également de l'AbDb, le score le plus élevé correspond au binding sites de l'antigène cible. Par exemple, le scan de l'antigène du complexe 4OII_1A va retrouver un match parfait avec les binding sites 4OII_1A:H et 4OII_1A:L.

Ensuite nous allons retrouver des scores de match similaire pour les anticorps redondant évoqué précédemment. Ces matchs permettent seulement une vérification, on ne conserve qu'une version de l'anticorps les autres versions sont écartées. Ainsi en il existe une version redondante de 4OII_1A qui est 4OII_2B seule 4OII_1A sera finalement conservée.

Une fois ces 2 cas particulier traité nous avons généralement autour de 25 % des binding sites qui ont un score de 0, cela signifie que PathSearch n'a pas trouvé de match sur la surface. Ensuite on retrouve autour de 72 % des binding sites qui ont un score compris entre 2 et 10 donc cela nous laisse généralement autour de 3 % (~150) de binding sites avec un score supérieur à 10. En prenant en compte les redondances d'anticorps on passe finalement d'environ 150 à environ 100 binding site avec un score supérieur à 10.

4- Sélection des boucles en contact

Nous avons dans un premier temps fait correspondre les binding sites avec les meilleurs scores avec leur complexe AbDb initiaux. En effet, pour rappel le jeu de données de 6197 binding sites est issu des 3000 complexe AbDb. Le fait de repartir des complexes initiaux nous permet de récupérer les boucles des anticorps qui sont en contact direct avec l'antigène. Nous avons pour cela écrit un script en python, "best_boucle.py", qui retourne les séquences d'au moins 4 acides aminés consécutif avec une tolérance de +/- 5 acide aminés. L'objectif est d'ensuite remplacer les régions variables de l'affirmer que sont les 2 Bêta-turn et la région N-terminal par ces séquences/boucles isolées par notre script.

5- Affimer

Nous sommes passés par l'outil Modeller pour modéliser un nouvel Affimer à partir d'une structure cristallographique d'Affimer existante. Notre support est une structure cristallographique PDB d'un Affimer donné et la cible est ce même Affimer mais avec les régions variables modifiées. Nous réalisons un alignement global par paire Needleman-Wunsch [20] entre la séquence support et la séquence cible.

Nous avons construit cet Affimer modifié à partir du complexe 7NY8.pdb. Ce complexe est composé de deux Affimer en interaction avec la GTPase KRas. Nous avons choisi ce complexe car l'Affimer correspondant à la chaîne C dans la structure 7NY8.pdb (figure 6 A) interagit au niveau des 2 Bêta-turn et au niveau de la queue N-terminal. Ainsi la théorie qui dit qu'un Affimère interagit spécifiquement au niveau des régions variables est confirmé par l'expérience pour cette structure.

Pour remplacer les régions variables de l'Affimer 7NY8, nous avons pris les boucles de la chaîne lourde de l'anticorps du complexe Ac/Ag 4OI1.pdb. Nous avons fait ce choix le binding sites 4OI1_1A:Ht.pdb est le meilleur match trouvé par PatchSearch avec un D-score de 20.48 sur la chaîne A et de 21.36 sur la chaîne B de la protéine NS1, 5GS6.pdb.

Ensuite ce choix est également motivé par le fait que notre script best_boucle.py nous renvoie 3 séquences consécutives : "GYTTSYWH", "WDNNS" et "DDGYRY". Cela permet de remplacer les 3 régions variables par ces séquences 3 identifiées comme en contact et consécutives. Nous allons remplacer la queue N-terminal de 7NY8 : ASNSLE → GYTTSY et les 2 Bêta-turn : DYQQ → WDNN et RMGS → DDGYRY.

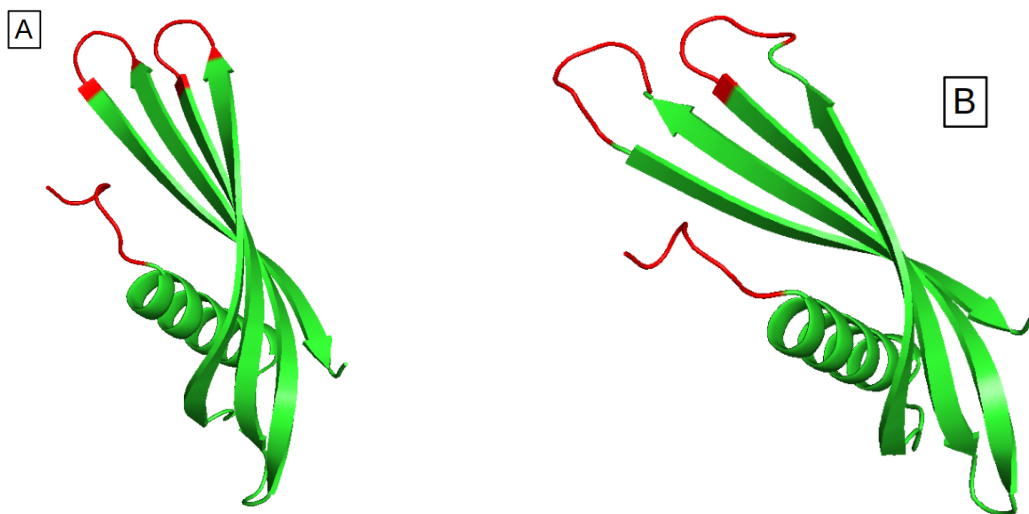


Figure 6 (A) : Structure Affimer initial Figure 6 (B) : Modélisation Affimer modifié
Les zones colorié en rouge sont les zones variable remplacé

On constate que la partie constante de l’Affimer est conservée lors de modélisation puisqu’on retrouve bien l’hélice alpha et le feuillet Bêta de 4 brins. On constate également que les boucles insérés représentés en rouge ont été correctement modélisés. En effet, on les retrouve bien au niveau des 2 Bêta turn ainsi que de la queue N-terminale.

Nous avons essayé d’évaluer la capacité de cet Affimer modifié à interagir spécifiquement avec la protéine NS1, 5GS6. Pour cela nous avons utilisé le serveur web HDock, la cible est la protéine NS1, 5GS6 et le ligand l’Affimer modifié.

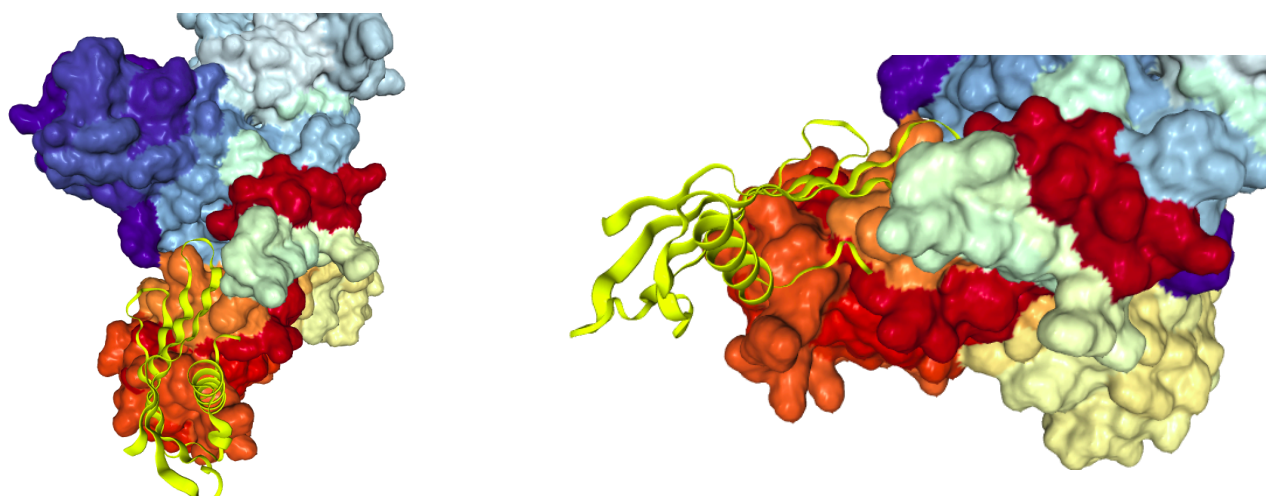


Figure 7 : Modèle N°1 pour l'interaction NS1-Affimer. Docking score de -301.1

On constate que HDock trouve bien une interaction entre NS1 et l’Affimer modifié au niveau des régions variables. Cependant on constate que l’interaction n’est pas spécifique au région variable, la partie constante de l’Affimer interagit également. Il existe des structures cristallographiques de complexe protéine-Affimer où la partie constante interagit mais je vais quand même essayer d’améliorer la modélisation de mon Affimer pour que seule les régions variables interagissent.

Je modélise un Affimer en utilisant cette fois des boucles issues de différents binding site. Je reprend la boucle “WDNNS” issus de 4OII_1A:Ht.pdb, ensuite je prend la boucle “FWSTR” issus de 4OII_2B:Lt.pdb et enfin la boucle “GYTTSY” issus de 2QHR_1P:Ht.pdb. Ce choix est de nouveau guidé par le D-score, c’est 3 binding sites ayant obtenus les meilleurs D-scores. Ce choix est également motivé par le fait que je voulais remplacer les Bêta-turn par une boucle de chaîne lourde pour l’un et une boucle de chaîne légère pour l’autre. Cela permet de reproduire la configuration d’un anticorps et j’espère ainsi augmenter la spécificité. Ainsi je remplace les 2 Bêta-turn : DYQQ → WDNNS et RMGS → FWSTR et je remplace la queue N-terminale : ASNSLE → YSRGGY

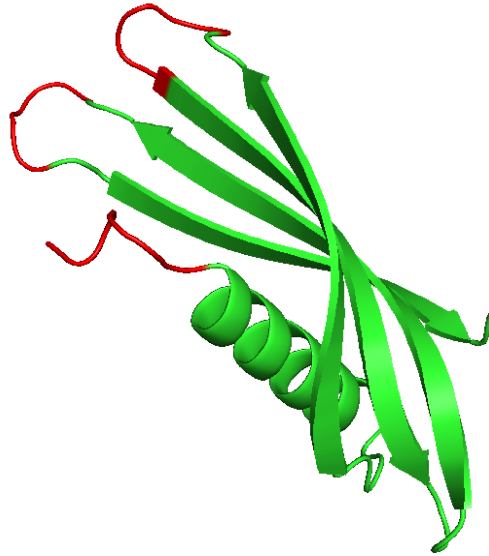


Figure 8 : Nouvelle modélisation Affimer modifié

On constate qu'à nouveau la partie constante de l'Affimer et les boucles insérés représentés en rouge ont été correctement modélisés. On lance HDock pour rechercher les interactions possibles.

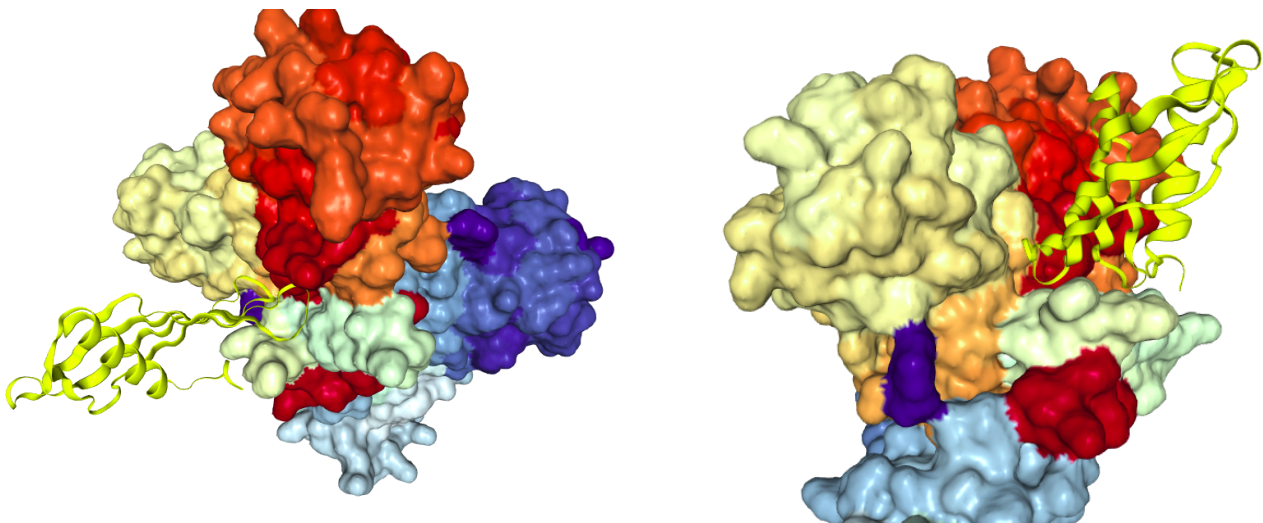


Figure 9 : Modèle N°1 pour l'interaction NS1-Affimer. Docking score de -265.74

On constate que cette fois seule les régions variables de l'Affimer modifié sont en interaction avec la protéine NS1. En effet, on constate que l'Affimer interagit au niveau des Bêta-turn modifié ainsi que au niveau de la queue N-terminale. Le Docking score est moins bon car une moins grande partie de l'Affimer est en interaction avec NS1. En interagissant uniquement par le biais de ces régions variables cet Affimer peut potentiellement être aussi spécifique qu'un anticorps classique.

IV - Conclusion

Nous avons réussi à modéliser un Affimer théoriquement spécifique de la protéine NS1, 5GS6. Nous sommes donc partie d'une protéine (5GS6) pour laquelle il n'existe pas d'anticorps. Nous avons cherché sur la surface de cette protéine les binding sites côté antigène de l'AbDb similaire. Nous avons ensuite récupéré les acides aminés consécutif des d'anticorps (correspondant au meilleur binding site) en contact direct avec l'antigène. Enfin nous avons remplacé les régions variables de l'Affimer 7NY8 par ces séquences, appelées boucles et testé sa spécificité avec HDock.

Il s'agit d'un résultat encourageant qu'en à la question de départ qui était de savoir si on pouvait améliorer le diagnostic/dépistage des flavivirus actuellement réalisé par le biais d'anticorps. Ce n'est pas une preuve tangible mais un indice que le concept est prometteur.

V - Bibliographie

- [1] « Coronavirus disease (COVID-19) – World Health Organization ». <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (consulté le janv. 10, 2022).
- [2] O. Leavy, « Therapeutic antibodies: past, present and future », *Nat. Rev. Immunol.*, vol. 10, n° 5, p. 297, mai 2010, doi: 10.1038/nri2763.
- [3] R. Woodman, J. T.-H. Yeh, S. Laurenson, et P. K. Ferrigno, « Design and Validation of a Neutral Protein Scaffold for the Presentation of Peptide Aptamers », *J. Mol. Biol.*, vol. 352, n° 5, p. 1118-1133, oct. 2005, doi: 10.1016/j.jmb.2005.08.001.
- [4] P. Colas, B. Cohen, T. Jessen, I. Grishina, J. McCoy, et R. Brent, « Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2 », *Nature*, vol. 380, n° 6574, p. 548-550, avr. 1996, doi: 10.1038/380548a0.
- [5] L. Priyamvada *et al.*, « Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus », *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, n° 28, p. 7852-7857, juill. 2016, doi: 10.1073/pnas.1607931113.
- [6] W. Dejnirattisai *et al.*, « Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with Zika virus », *Nat. Immunol.*, vol. 17, n° 9, p. 1102-1108, sept. 2016, doi: 10.1038/ni.3515.
- [7] S. A. Poveda-Cuevas, C. Etchebest, et F. L. Barroso da Silva, « Insights into the ZIKV NS1 Virology from Different Strains through a Fine Analysis of Physicochemical Properties », *ACS Omega*, vol. 3, n° 11, p. 16212-16229, nov. 2018, doi: 10.1021/acsomega.8b02081.
- [8] S. A. Poveda-Cuevas, F. L. Barroso da Silva, et C. Etchebest, « How the Strain Origin of Zika Virus NS1 Protein Impacts Its Dynamics and Implications to Their Differential Virulence », *J. Chem. Inf. Model.*, vol. 61, n° 3, p. 1516-1530, mars 2021, doi: 10.1021/acs.jcim.0c01377.
- [9] S. Ferdous et A. C. R. Martin, « AbDb: antibody structure database—a database of PDB-derived antibody structures », *Database J. Biol. Databases Curation*, vol. 2018, p. bay040, avr. 2018, doi: 10.1093/database/bay040.
- [10] S. K. Burley, H. M. Berman, G. J. Kleywegt, J. L. Markley, H. Nakamura, et S. Velankar, « Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive », *Methods Mol. Biol. Clifton NJ*, vol. 1607, p. 627-641, 2017, doi: 10.1007/978-1-4939-7000-1_26.
- [11] L. C. Allcorn et A. C. R. Martin, « SACS--self-maintaining database of antibody crystal structure information », *Bioinforma. Oxf. Engl.*, vol. 18, n° 1, p. 175-181, janv. 2002, doi: 10.1093/bioinformatics/18.1.175.
- [12] E. A. Kabat, T. T. Wu, H. M. Perry, C. Foeller, et K. S. Gottesman, *Sequences of Proteins of Immunological Interest*. DIANE Publishing, 1992.

- [13] C. Chothia *et al.*, « Conformations of immunoglobulin hypervariable regions », *Nature*, vol. 342, n° 6252, p. 877-883, déc. 1989, doi: 10.1038/342877a0.
- [14] K. R. Abhinandan et A. C. R. Martin, « Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains », *Mol. Immunol.*, vol. 45, n° 14, p. 3832-3839, août 2008, doi: 10.1016/j.molimm.2008.05.022.
- [15] « Bio.PDB.SASA module — Biopython 1.80.dev0 documentation ». <https://biopython.org/docs/dev/api/Bio.PDB.SASA.html> (consulté le janv. 15, 2022).
- [16] « Environment and exposure to solvent of protein atoms. Lysozyme and insulin - ScienceDirect ». <https://www.sciencedirect.com/science/article/abs/pii/0022283673900119?via%3Dihub> (consulté le janv. 15, 2022).
- [17] J. Rey, I. Rasolohery, P. Tufféry, F. Guyon, et G. Moroy, « PatchSearch: a web server for off-target protein identification », *Nucleic Acids Res.*, vol. 47, n° W1, p. W365-W372, juill. 2019, doi: 10.1093/nar/gkz478.
- [18] B. Webb et A. Sali, « Comparative Protein Structure Modeling Using MODELLER », *Curr. Protoc. Bioinforma.*, vol. 54, p. 5.6.1-5.6.37, juin 2016, doi: 10.1002/cpbi.3.
- [19] Y. Yan, D. Zhang, P. Zhou, B. Li, et S.-Y. Huang, « HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy », *Nucleic Acids Res.*, vol. 45, n° Web Server issue, p. W365-W373, juill. 2017, doi: 10.1093/nar/gkx407.
- [20] F. Madeira *et al.*, « The EMBL-EBI search and sequence analysis tools APIs in 2019 », *Nucleic Acids Res.*, vol. 47, n° W1, p. W636-W641, juill. 2019, doi: 10.1093/nar/gkz268.

Annexe

Mon GitHub pour ce projet long affimer :

<https://github.com/TheoJamay/AFFIMER.git>