Coppin Georges
MA1 BioInformatique

Rapport des TP de « Biophysics and Structural Bioinformatics I »

CHIM-F-451

Professeurs responsables : Marianne Rooman et Dimitri Gilis

1. Secondary structure prediction and 3D structure prediction

Etude de la structure de la leptine de poulet (Uniprot code LEP_CHICK).

a) Predict the secondary structure of this protein.

Pour prédire la structure secondaire d'une protéine, plusieurs possibilités s'offrent à nous, dont GOR IV, HNN et Sopma. Ces trois méthodes utilisent des critères différents pour déterminer sur base de la séquence d'acides aminés, la structure secondaire d'une protéine. Ces méthodes ne caractérisent que les trois états de structure secondaire : « Helix », « beta strand » et « coil ». Pour chacune des méthodes, les résultats sont affichés de la même façon : où H = Helix, E = beta strand, C = coil

La méthode GOR IV se base sur les fréquences d'appariement possible entre 17 acides aminés (J. Garnier. J.F. Gibrat et B.Robson dans « Methods en Enzymology », vol 266, p540-553 (1996)). Après exécution du programme, le programme renvoie la séquence prédite. L'exactitude des prédictions par le modèle GOR IV est de l'ordre de 64,4% pour ces trois états cités.

```
         10         20         30         40         50         60         70
          |          |          |          |          |          |          |
MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
ccccccccccceeeeeccccceecccccchhhheeeeeeccccccchhhhhccccccccccccchhhhhc
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccchhhhhhhccchhh
VALSRLQGSLQDILQQLDISPEC
hhhhhhhchhhhhhhhhceecee
Sequence length :   163
```

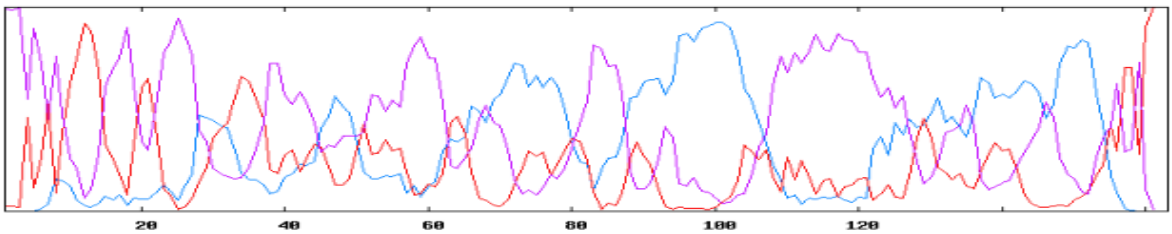


Figure 1 : Graphique des valeurs de probabilités de structure secondaire pour chaque acide aminé avec GOR

La méthode de prédiction de structure secondaire HNN se base sur deux critères pour évaluer les probabilités de structure pour chaque acide aminé. Sur base de la séquence primaire brute, et sur base de structure en elle-même. Pour cela, la méthode HNN utilise des données physico-chimiques liées aux structures incorporées dans le modèle de prédiction. Ces nouveaux paramètres augmentent les risques d'erreurs. Ce risque pourrait être diminué en travaillant avec des profiles d'alignements multiples.

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
cchhhhhhhhhheeeeecccceeecccccchheehheeeecccccccccccccceecccccchhhhhhhccccc
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhcccccccccccccccccccchhhhhhhhhhhh
VALSRLQGSLQDILQQLDISPEC
Hhhhhhhchhhhhhhhccccccc
```

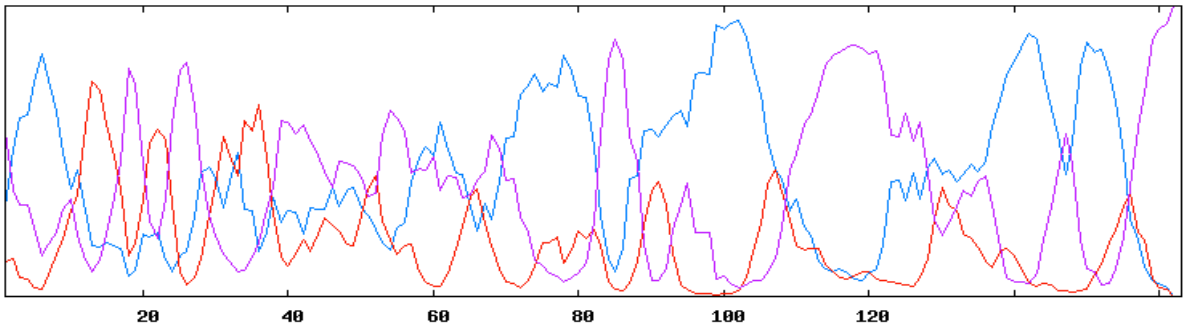


Figure 2 : Graphique des valeurs de probabilités de structure secondaire pour chaque acide aminé avec HNN

La méthode de prédiction de structure secondaire SOPMA se base sur une base de données de séquences de protéines alignées appartenant à la même famille. Cette base de données contient 126 séquences d'acides aminés de protéines dites non homologues (présentant moins de 25% d'identité de séquence). L'exactitude des prédiction est de l'ordre de 69.5% pour les trois états de structures secondaires pris en compte.

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
ecccchhhhheeeeecccccchhhccchhhhhhhhheehhhccccccccccccceettteeecccccchhhhhh
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhhtccccchhhhhhhhhhhhhhhhhhhhhtccccccccchhhhhhhhhhhhcccchhh
VALSRLQGSLQDILQQLDISPEC
hhhhhhhhhhhhhhhhhhhcccc
```

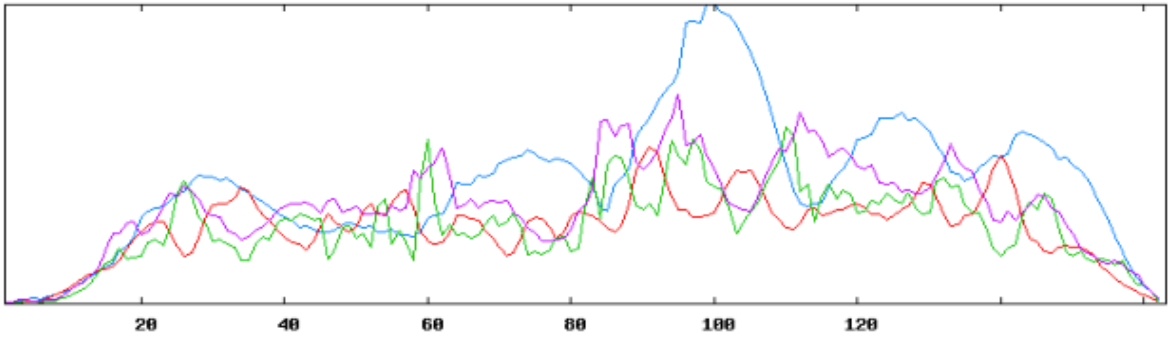


Figure 3 : Graphique des valeurs de probabilités de structure secondaire pour chaque acide aminé avec SOPMA

b) Search for a possible template to predict the 3D structure of the chicken leptin and analyze in detail the quality of the template structure.

Il est possible de modéliser une structure 3D de protéine sur base d'une autre protéine, dont la structure a déjà été résolue expérimentalement. On parle alors de modélisation comparative, où le template est le modèle de comparaison.

Un template possible pour prédire la structure 3D de la protéine LEP_CHICK est par exemple celui de la protéine 1AX8A, identifiée avec l'outil BLAST. En effet, cette protéine présente 81,7% d'identité de séquence avec la leptine de poulet. Avec l'outil « mobyle » qui permet d'aligner nos deux séquences, on observe ces résultats :

```
# Aligned_sequences: 2
# 1: 1AX8A
# 2: LEP_CHICK
# Matrix: EBLOSUM62
# Gap_penalty: 12
# Extend_penalty: 2
#
# Length: 164
# Identity:     120/164 (73.2%)
# Similarity:   134/164 (81.7%)
# Gaps:          19/164 (11.6%)
# Score: 543
```

On constate que les deux protéines partagent 120 acides aminés en commun (81,7% d'identité de séquence), présente 11,6% de GAPS, et ont un score de 543. Sachant que l'on constate que 30 à 40% d'identité de séquence entre le template et la séquence cible suffit pour considérer que la structure 3D de ces deux protéines sera équivalente, il faut néanmoins vérifier la fiabilité de la structure du template fourni par la littérature.

Le choix du template est donc très important dans la prédiction de la structure tertiaire d'une protéine. Une étude sur la qualité expérimentale d'un éventuel template est donc indispensable pour évaluer la fiabilité de nos prédictions. Pour cela, nous pouvons utiliser l'outil « PROCHECK » par exemple. Cet outil, disponible sur la page PDB des protéines étudiées expérimentalement, évalue la qualité stéréochimique de la structure, résidu par résidu.

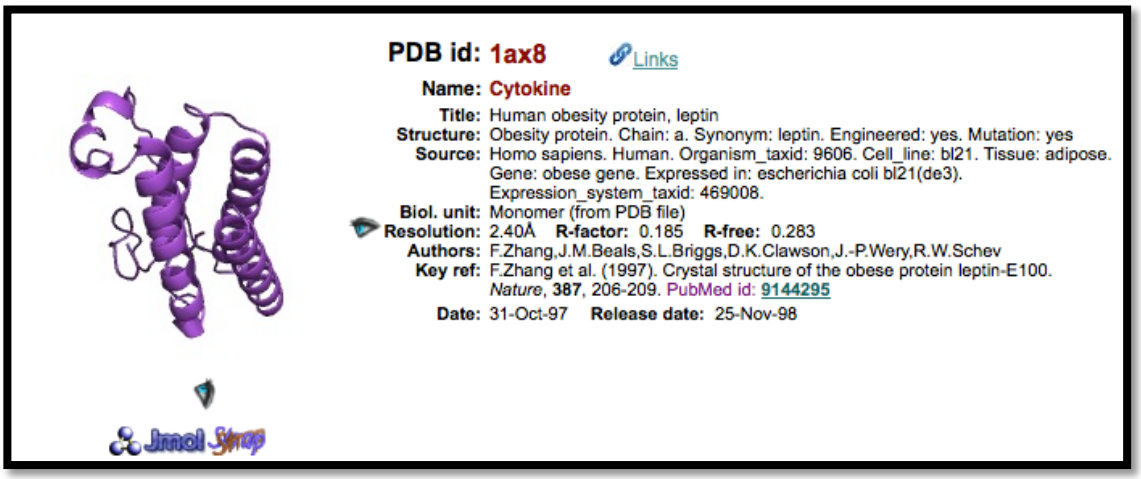


Figure 4 : Informations fournies par PROCHECK

Sur la figure 4, on peut voir que la structure de la protéine a été déterminée par cristallographie aux rayons-X. PROCHECK donne une évaluation sur la qualité de structure 3D de 1AX8A, notamment à l'aide de calculs de paramètres. Parmi ceux-ci, on distingue la résolution, le R-factor, et R-free.

La résolution est la mesure du niveau de détail qui sera visible lorsque la carte de densité électronique sera calculée. Au plus la résolution est précise, au plus le niveau de détail est important. A partir de 3Å ou plus, seuls les contours basiques de la protéine sont distinguable. Dans notre cas, la résolution de la cristallographie donnée par PROCHECK vaut 2.40 Å, qui permet donc une interprétation correcte possible. Sur la figure 5, on peut voir la densité électronique d'une autre protéine (en bleu et jaune) pour différentes résolutions.

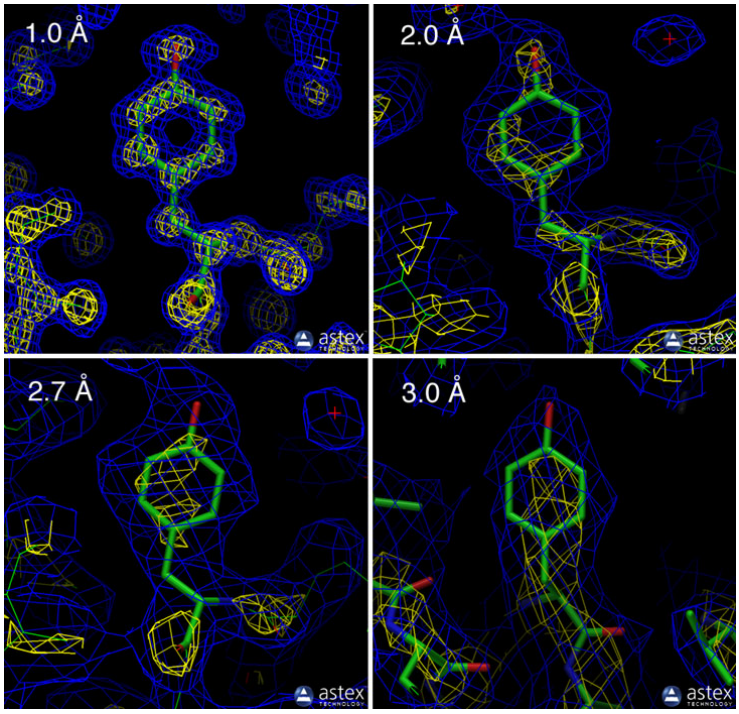


Figure 5 : carte de densité électronique pour différente résolution en cristallographie de rayon-X.

R-factor (ou R-value) donne des informations sur la qualité atomique du modèle. Après résolution de la structure atomique, on peut simuler le spectre de diffraction sur base du modèle. Le R-factor indique la différence entre les mesures de diffraction expérimentales et simulées. Un jeu d'atomes aléatoires donnent un R-factor d'environ 0.63, tandis qu'un ajustement parfait donne un R-factor de zéro. On estime qu'un R-factor approchant 0.20 est un bon indicateur. Dans notre cas, le R-factor vaut 0.185, signe d'une modélisation fiable. Néanmoins, un processus de raffinement est souvent utilisé le modèle atomique et l'adapter aux données expérimentales, ce qui peut induire un biais. On peut utiliser un autre paramètre, la valeur R-free, qui permet une estimation moins biaisée. La valeur R-free est calculée comme celle du R-factor, mais avant le raffinement, environ 10% des observations expérimentales sont enlevées du jeu de données. On peut alors comparer le R-factor et R-free, pour un modèle idéal, R-factor et R-free seront similaires, avec une valeur typique d'environ 0.26. C'est ce qu'on observe pour 1AX8A, où R-free présente une valeur de 0.283.
Réf : http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/Rvalue.html

PROCHECK permet aussi de vérifier les valeurs d'angles mesurés expérimentalement entre acides aminés. Notamment les valeurs d'angles de torsion phi, psi et omega repris dans le diagramme de Ramachandran. Les critères d'évaluations ne sont plus énergétiques dans ce cas-ci, mais physico-chimiques. Le diagramme de Ramachandran réalisée pour 1AX8A est visible sur la figure 6.

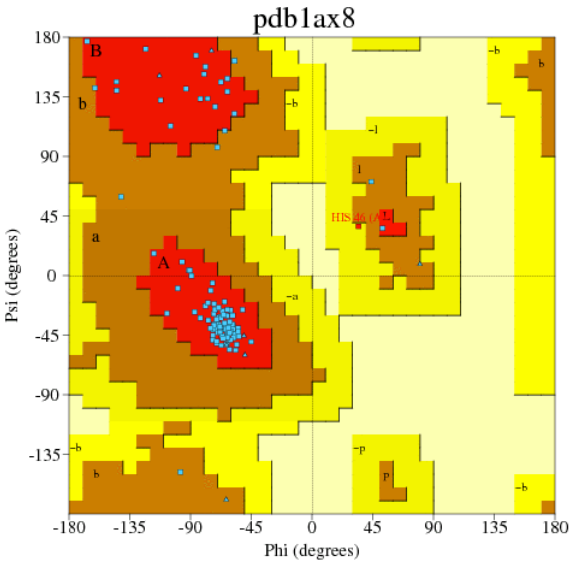


Figure 6 : diagramme de Ramachandran de la protéine 1AX8A

		residues	%-tage
Most favoured regions	[A,B,L]	108	94.7%
Additional allowed regions	[a,b,l,p]	5	4.4%
Generously allowed regions	[~a,~b,~l,~p]	1	0.9%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		114	100.0%
End-residues (excl. Gly and Pro)		4	
Glycine residues		7	
Proline residues		5	
Total number of residues		130	

Au dessus, on peut observer une série de résultats statistiques relatifs au diagramme de Ramachandran pour 1AX8A. On observe notamment que 94.7% des résidus se retrouvent dans les régions les plus favorables du diagramme. Qu'aucun résidu ne se retrouve dans une région non permise.

Avec toutes ces données, on peut dire que le modèle structural de 1AX8A déterminé expérimentalement est fiable et peut donc servir de template en vue de modéliser la structure 3D de la protéine d'intérêt LEP_CHICK.

c) *Predict the 3D structure of the chicken leptin. Which method did you use to predict the structure of this protein? Analyse and discuss the quality of the model that you obtain.*

Pour prédire la structure 3D de la leptine du poulet, plusieurs possibilités s'offre à nous. On peut notamment utiliser la modélisation comparative, en utilisant le template 1AX8A trouver plus tôt. En effet, comme dit précédemment, lorsque deux protéines partages plus de 30 à 40% d'identité de séquence, celles-ci présenteront la même structure 3D. Plusieurs outils en ligne s'offrent à nous pour

réaliser une modélisation comparative, 3DJigsaw, HHPRED (couplé à Modeller) qui réalisent eux même l'alignement de séquence. On peut également utiliser Modeller, mais l'alignement de séquence doit être établi au préalable. Ces différents outils disponibles en ligne ne renvoient qu'un seul modèle de structure 3D. Le programme Modeller peut être installé sur ordinateur et donc être utiliser localement avec davantage de possibilités. On peut, entres autres fournir une population de modèles plutôt qu'un seul. Comme Modeller fonctionne par satisfaction de contrainte, certaines régions présentent un nombre de contraintes plus faibles et parfois plusieurs versions de modèles peuvent donc être compatibles. Une fois la population de modèle obtenu, on retient les modèles qui présentent les structures les plus favorables.

Dans mon cas, j'ai utilisé HHPRED couplé avec Modeller. Je n'ai pas réussi à installer le programme Modeller sur mon ordinateur. On obtient un fichier .pdb téléchargeable, avec la structure 3D observable avec le programme pymol (figure 7).

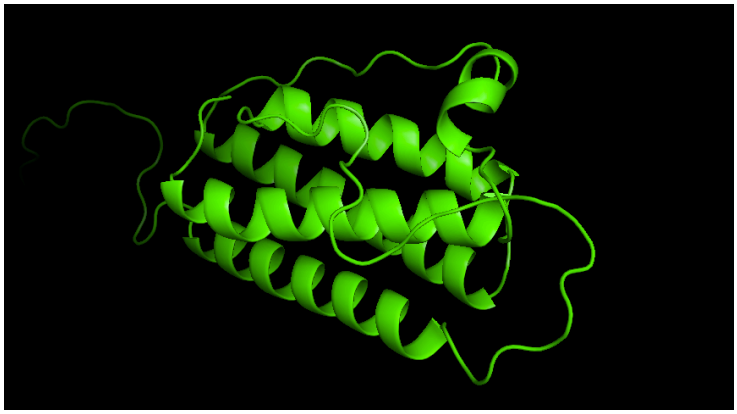
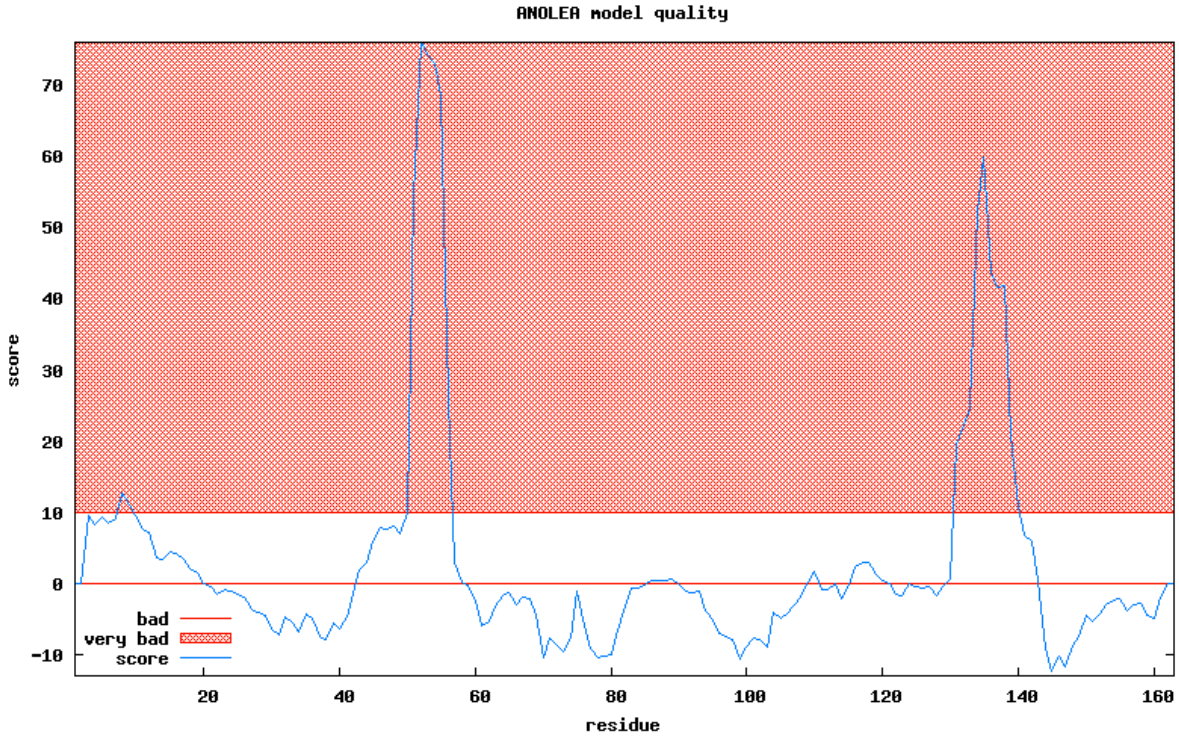


Figure 7 : Prédiction du modèle structurale 3D de LEP_CHICK obtenu avec HHPRED-Modeller

Une fois le modèle obtenu, il faut tester la qualité de la structure modélisée. Pour cela, il existe plusieurs méthodes d'évaluation : via PROCHECK, en mesurant le profil énergétique de notre structure, ou encore en croisant notre modèle structurale avec des données expérimentales trouvées dans la littérature.
Pour mesurer le profil énergétique d'un modèle, on peut utiliser deux outils : Anolea, et Prosaweb. L'énergie de la structure est calculée pour chaque acide aminé localement ou non. Au plus bas l'énergie est mesurée, au plus stable sera la structure.

ANOLEA (Atomic Non-Local Environment Assessment) est un serveur qui effectue des calculs d'énergie sur la chaine étudiée de manière non locale. Ce calcul d'énergie est basée sur la distance relative des atomes, et sur une base de données dont l'énergie a déjà été mesurée pour 147 protéines non- redondantes, avec une identité de séquence inférieure à 25%, résolue par cristallographie par rayon X avec résolution inférieure à 3 Å.
Référence : http://melolab.org/anolea/

Sur la figure 8, on observe que la majeure partie de la séquence présente un bon score. Néanmoins, par trois fois, celui-ci se trouvera dans la zone « très mauvaise » et mauvaise. La première zone est peut être due au Gaps qu'il y a entre 1AX8A et LEP_CHICK au niveau des 20 premiers acides aminés. Le z-score (défini plus loin) affiché par ANOLEA est de 8.06, ce qui indique une bonne valeur statistique du modèle.



F. Melo and E. Feytmans (1998) Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 277, 1141 - 1152.

List of amino acids with high energy: 3-19; 43-58; 86-89; 109-110; 116-120; 130-143;
Total amino acids with high energy = 58 Percentage = 35.58
Total number of aminoacids = 163 Total number of atoms = 1275
Total number of non-local atomic interactions = 13324
Total non-local energy of the protein (E/kT units) = 487
Non-local normalized energy Z-score = 8.06

Figure 8 : Graphique ANOLEA du SCORE obtenu après modélisation de la structure 3D de LEP_CHICK

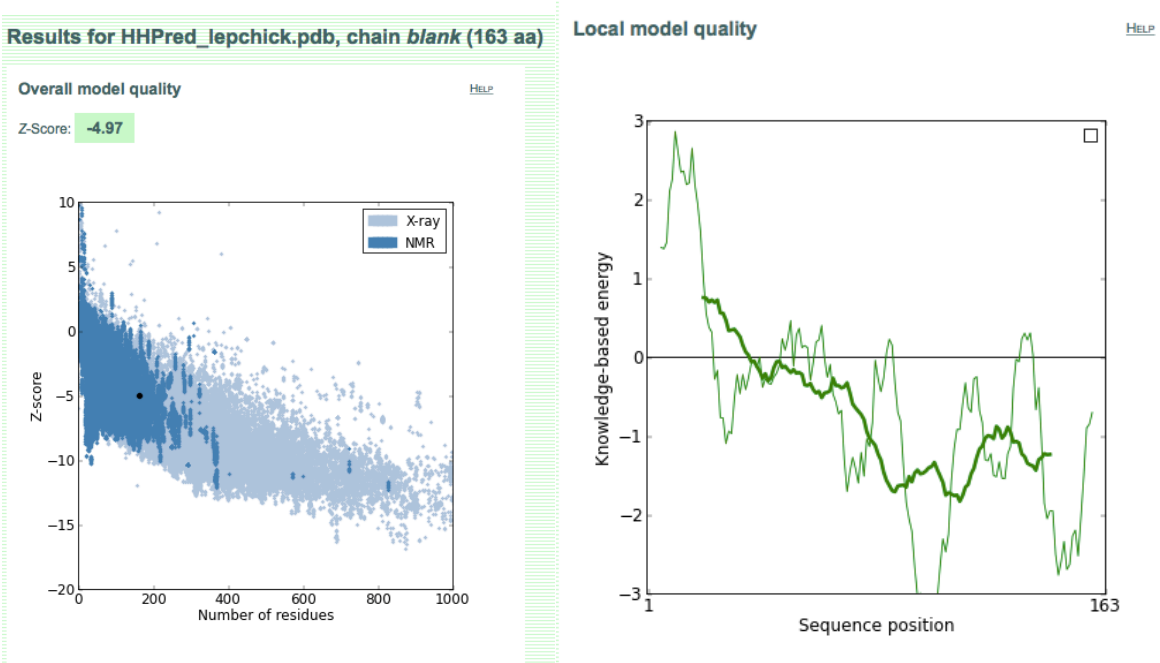


Figure 9 : Graphiques PROSAWEB obtenu après modélisation de la structure 3D de LEP_CHICK.

La figure 9 correspond aux résultats obtenus avec Prosaweb. Le premier graphique compare notre modèle avec sa base de données. Celle-ci répertorie une série de protéines résolues expérimentalement (via NMR ou X-ray). Un modèle qui se situe dans les zones « permises » du graphique aura plus de

probabilité d'être juste. Notre modèle, le point noir sur le graphique, semble présenter une conformation permise.

Sur le second graphique fourni par Prosaweb, on constate également que l'énergie passe parfois en positif. Ici, le calcul est fait localement en fonction des angles mesurés entre chaque acide aminé. L'énergie la moins favorable à notre modèle se retrouve en début de séquence, tout comme pour ANOLEA, au niveau du GAP entre 1AX8A et LEP_CHICK. Pour le reste de la séquence, l'énergie semble favorable au modèle.

Pour les deux outils de vérification, on constate que la grande partie des zones où le score est négatif, celles-ci correspondent aux boucles du modèles. En effet, elles ne correspondent généralement pas aux hélices alpha modélisées. Ceci s'explique sans doute par le fait qu'il est plus difficile de modéliser ce genre de structure, et augmente ainsi le risque d'erreurs de modélisation dans ces zones.

Ces deux outils se basent donc sur différents calculs d'énergie, il est donc normal d'y trouver des différences. Par exemple, le voisinage d'un acide aminé peut être favorable énergétiquement, mais la position relative de cette acide aminé vis-à-vis d'une structure présente sur la séquence peut ne pas l'être.

d) *Compare the secondary structures predicted in section (1a) to the secondary structures that you find in the model obtained in section (1b).*

La structure secondaire de la protéine 1AX8A est disponible sur la base de données UniProt (dont le lien est disponible sur la page pdb de la protéine). Cette structure a été déterminée sur base de preuves expérimentales et informatiques. *« Manually validated information inferred from a combination of experimental and computational evidence. »*

Figure 10 : Structure secondaire de la protéine 1AX8A de la base de données UniProt

On distingue parmi la séquence, 7 hélices alpha :

Feature key	Position(s)	Length	Description	Graphical view
Helix	25 – 44	20	Combined sources	
Helix	72 – 87	16	Combined sources	
Helix	92 – 114	23	Combined sources	
Helix	128 – 131	4	Combined sources	
Helix	132 – 135	4	Combined sources	
Helix	142 – 160	19	Combined sources	
Helix	161 – 163	3	Combined sources	

Tableau 11 : Positions des différentes structures secondaires le long de la séquence de la protéine 1AX8A

Comparaison des structures secondaires de LEP_CHICK avec celle de 1AX8A :

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
cchhhhhhhhhheeeeecccceeecccccchheeheeeeccccccccccccceecccccchhhhhhhccccc
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhhcccccccccccccccccccchhhhhhhhhhhh
VALSRLQGSLQDILQQLDISPEC
hhhhhhhchhhhhhhhccccccc                    GOR IV LEP_CHICK

MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
cchhhhhhhhhheeeeecccceeecccccchheeheeeeccccccccccccceecccccchhhhhhhccccc
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhhcccccccccccccccccccchhhhhhhhhhhh
VALSRLQGSLQDILQQLDISPEC
hhhhhhhchhhhhhhhccccccc                    HNN LEP_CHICK

MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
ecccchhhhheeeeecccccchhhccchhhhhhhhhheehhhccccccccccccceettteeecccccchhhhhh
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhhtccccchhhhhhhhhhhhhhhhhhhhhtccccccccchhhhhhhhhhhhhccccchhh
VALSRLQGSLQDILQQLDISPEC
hhhhhhhhhhhhhhhhhhhccccc                   SOPMA LEP_CHICK

MCWRPLCRLWSYLVYVQAVPCQIFQDDTKTLIKTIVTRINDISHTSVSAKQRVTGLDFIPGLHPILSLSK
ccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccchh
MDQTLAVYQQVLTSLPSQNVLQIANDLENLRDLLHLLAFSKSCSLPQTSGLQKPESLDGVLEASLYSTEV
hhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhcccccccccccchhhhhhhhccccccchhh
VALSRLQGSLQDILQQLDISPEC
hhhhhhhhhhhhhhhhhhhccccc                   HHPRED-Modeller LEP_CHICK

------------------VPIQKVQDDTKTLIKTIVTRINDISHTQSVSSKQKVTGLDFIPGLHPILTLS
                  ccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccc
KMDQTLAVYQQILTSMPSRNVIQISNDLENLRDLLHVLAFSKSCHLPEASGLETLDSLGGVLEASGYSTE
cchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccchhhhhhhhccccc
VALSRLQGSLQDMLWQLDLSPGC
chhhhhhhhhhhhhhhhhhhhhh                   UniProt 1AX8A aligné

         |         |         |         |         |         |         |
         10        20        30        40        50        60        70
```

On constate que toutes ces structures présentent des similitudes et des différences. Les hélices alpha sont situées relativement similairement. On observe également qu'on ne retrouve aucun feuillet beta dans la structure des protéines 1AX8A et LEP_CHICK avec HHPRED tandis qu'on en observe pour tous les autres modèles de la protéine LEP_CHICK. En terme de comparaison, la méthode SOPMA semble être la plus proche du modèle LEP_CHICK déterminé avec le template 1AX8A.

2. Protein structure superimposition

Describe the function of human thioredoxin (PDB code: 3TRX) and of E. coli glutaredoxin (PDB code 3GRX).

La protéine 3TRX fait partie de la famille des thiorédoxines. Cette classe de protéines est présente dans tous les organismes connus et est essentielle pour la vie des mammifères. En effet, chez l'humain, une perte de fonction de cette protéine est létale dès le stade 4 cellules lors du développement embryonnaire. Les thioredoxines sont des protéines antioxydants en facilitant la réduction d'autres protéines par échange de cystéine thiol-disulfide (wikipedia).

La protéine Glutaredoxines de E. coli (3GRX) catalyse de manière réversible l'oxydation ou la réduction de protéines disulfides pouvant comporter du glutathion. La structure de 3GRX est similaire à celle des thiorédoxines. La famille des glutaredoxines a également la particularité d'être très spécifique.

Which experimental technique has been used to resolve the structure of E. coli glutaredoxin (3GRX)? What are the limits of the secondary structures (beta strands and helices) of this protein? Which ligand is present in the structure of 3GRX? Which residues are in interaction with this ligand? What is the CATH classification of 3GRX?

La séquence 3D de la protéine 3GRX a été résolue expérimentalement par résonnance magnétique nucléaire (NMR).

Sur le tableau 2, on peut observer la disposition des différentes structures secondaires de la protéine 3GRX avec leur position relative sur la séquence.

Structure	Position	Longueur (a.a)
Brin Beta	4-8	5
Hélice Alpha	13-24	12
Brin Beta	29-33	5
Hélice Alpha	39-48	10
Brin Beta	55-58	4
Brin Beta	61-62	2
Hélice Alpha	66-74	9
Hélice Alpha	79-82	4

Tableau 2 : Disposition des structures secondaires de la protéine 3GRX

Le ligand présent sur la structure de 3GRX est le Glutathion, $C_{10}H_{17}N_3O_6S$.

Figure 12 : Disposition des résidus fixant le ligand le long de la séquence de la protéine 3GRX (pdb)

Sur la figure 12, on peut observer les 3 résidus de 3GRX capables de se lier avec le ligand:

1. 3GRX.A : site de liaison pour résidu GSH A83 – AC1 (position 12)
2. 3GRX.A : site de liaison pour résidu GSH A83 – AC1 (position 51)
3. 3GRX.A : site de liaison pour résidu GSH A83 – AC1 (position 52)

La classification « CATH » est une classification de protéine de structure téléchargée depuis la base de données « Protein Data Bank » où les protéines sont groupées en superfamilles lorsqu'il existe suffisamment de preuves d'une divergence avec un ancêtre commun.

Superimpose both structures and align the sequence of both proteins. Analyse and discuss the results.

L'alignement structurel de deux protéines peut être réalisée via l'outil « PDBe Fold ». Cet outil fourni, en plus de la superposition, une série de valeurs évaluant les qualités de ces superpositions. Parmi ces valeurs, il y a notamment le rmsd et le *z-score*. Le rmsd, pour *« Root Mean Square Deviation »,*

correspond à la mesure des distances entre deux atomes de structures alignées. Il permet d'évaluer la différence entre deux modèles alignés. Au plus les séquences présenteront des identités de séquences, au plus bas la valeur du rmsd sera. Après superposition de 3GRX avec 3TRX, la valeur du rmsd vaut 3.074. Le *z-score* est une valeur statistique qui évalue l'alignement de structure. Cette valeur correspond à la distance, en déviation standard, entre l'alignement observé et un alignement aléatoire de pairs. Dans notre cas, le *z-score* est de 3.38, on estime qu'en dessous de 2, la valeur statistique du modèle n'est pas pertinente.

Query 3GRX.pdb:A				Alignment (1 of 1)				Target 3TRX.pdb:A			
N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$	Q	P	RMSD	N_{align}	N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$
82	88	7	71	0.294	0.82	3.074	72	105	69	8	63
NMR STRUCTURE OF ESCHERICHIA COLI GLUTAREDOXIN 3-GLUTATHIONE MIXED DISULFIDE COMPLEX, 20 STRUCTURES				$\%_{seq}$	Z	N_{SSE}	N_{gaps}	HIGH-RESOLUTION THREE-DIMENSIONAL STRUCTURE OF REDUCED RECOMBINANT HUMAN THIOREDOXIN IN SOLUTION			
				13.9	3.38	5	7				

Figure 13 : résultat de la superposition de 3GRX avec 3TRX avec l'outil PDBe Fold

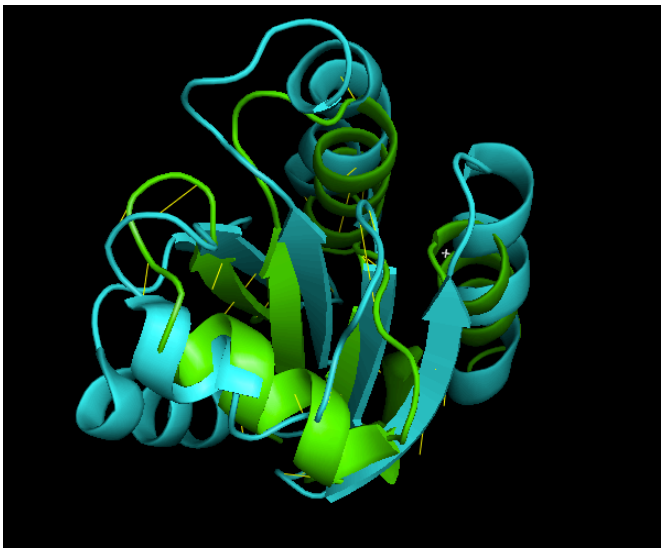


Figure 14 : Modélisation de la superposition de 3GRX (vert) et 3TRX (bleu) avec Pymol

Sur la figure 14, on constate que les deux structures présentent les mêmes structures avec des positions relatives équivalentes. A noter que la protéine 3TRX présente une hélice alpha en 7-18 qui n'a pas d'équivalent sur l'autre protéines. Ces similitudes étaient attendues étant donné que ces protéines sont relativement bien conservées au travers de l'évolution. Néanmoins, l'identité de séquence entre ces deux protéines n'est « que » de 13.9%. Ce qui est relativement peu lorsque l'on sait que ce n'est qu'à partir de 30 à 40% d'identité de séquence que l'on estime que deux protéines présentent des structures équivalentes. Ceci est sans doute dû au fait que la fonction de ces protéines doit être conservée. Ainsi, certains acides aminés, qui n'altèrent pas la structure des protéines peuvent différer d'une protéine à l'autre. En général, là où l'on observe des différences d'alignement, c'est généralement au niveau des régions de boucle, au niveau des régions n et c terminale qui sont plus difficile à modéliser. Aussi au niveau des Gaps s'il y en a. Ce qui n'est pas observé ici malgré de nombreux Gaps (26,9% avec Mobyle).

Références :

Ce rapport se base sur ce qui a été dit lors des séances de TP, ainsi que sur les informations relatives aux outils utilisés disponibles sur leur site respectif (pdb, uniprot, prosaweb…). Lorsque d'autres sites ont été consultés, ceux-ci sont référencés dans le texte.