

BINF-F405: Biophysics and structural bioinformatics II

Practical course : Task1

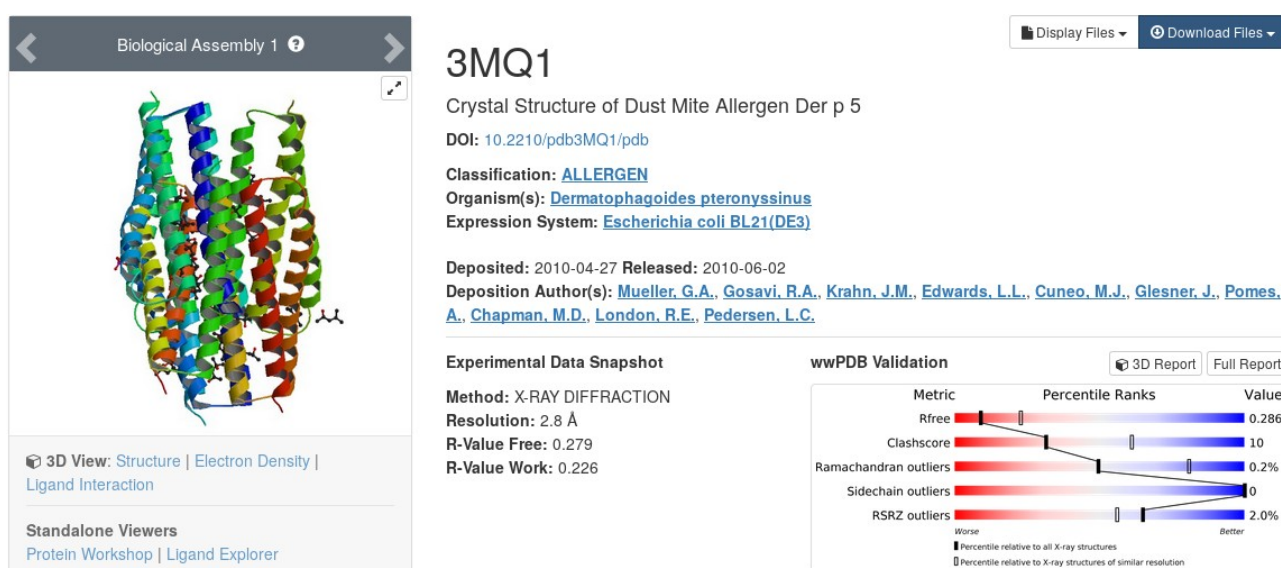
A) Download and save the sequences of Der p 5 (Uniprot code P14004), Blo t 5 (Uniprot code O96870) and Der f 5 (Uniprot code Q3LHF8).

```
>sp|P14004|ALL5_DERPT Mite allergen Der p 5 OS=Dermatophagoides pteronyssinus
OX=6956 GN=DERP5 PE=1 SV=2
MKFIIAFFVATLAVMTVSGEDKKHDYQNEFDLFLMERIHEQIKKGELALFYLQEQINHFE
EKPTKEMKDKIVAEMDTI IAMIDGVRGVLDRMLQRKDLDFEQYNLEMAKKSGDILERDL
KKEEARVKKIEV
```

```
>sp|O96870|ALL5_BLOTA Mite allergen Blo t 5 OS=Blomia tropicalis OX=40697
GN=BLOT5 PE=1 SV=1
MKFAIVLIACFAASVLAQEHKPKKDDFRNEFDHLLIEQANHAIEKGEHQLLYLQHQLDLDEL
NENKSKELQEKIIRELDVVCAMIEGAQGALERELKRTDLNILERFNYEEAQTLSKILLKD
LKETEQKVVDIQTQ
```

```
>tr|Q3LHF8|Q3LHF8_DERFA Group 5 allergen OS=Dermatophagoides farinae OX=6954
GN=Der f 5 PE=2 SV=1
MKFIIAIAVCTLAVVCVSGEPKKHDYQNEFDLFLMQRIHEQMRKGEEALLHLHHQINTFE
ENPTKEMKEQILGEMDTI IALIDGVRGVLNRLMKRTDLDFERYNVEIALKSNEILERDL
KKEEQRVKKIEV
```

B) Analyze the quality of the 6 chains that are in the PDB file 3MQ1. Use for that purpose the PDB website (<http://www.rcsb.org>). Which chain has the best quality?

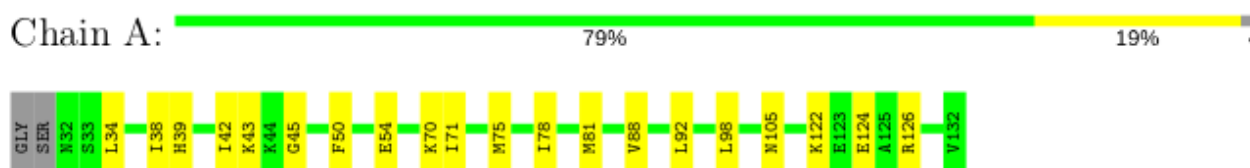


On peut voir via le lien suivant (<https://www.rcsb.org/structure/3MQ1>) que la méthode expérimentale qui a permis d'obtenir la structure de Der p 5 est la cristallographie aux rayons X. La résolution est de 2,8 angströms associée à une R-Value Work de 0.226. On dit en général que la résolution est bonne en dessous de 2,5 angströms, on peut donc la qualifier de moyenne ici. Quant à la R-Value, elle décrit la différence entre les observations expérimentales et les valeurs idéales calculées à partir du modèle cristallographique. Le minimum de la R-value est 0 correspondant à un accord parfait entre les valeurs prédites à partir d'un modèle et les valeurs observées expérimentalement. Pour les grosses molécules, la R-value se situe entre 0,2 et 0,6 et elle est inférieure à 0,2 pour les plus petites molécules. La R-value est donc correcte ici. Toutefois, bien qu'elle ait une signification, celle-ci peut être améliorée en utilisant les résultats expérimentaux et le

modèle construit afin d'améliorer celui-ci. C'est pour cette raison qu'il est préférable d'y associer la R-Value Free (0,279 ici) qui est calculée avec la même formule mais sur un petit échantillon stochastique de données mis de côté dans ce but et jamais inclus dans le raffinement que subit la R-value work. On peut ensuite creuser les investigation grâce au « Full Validation Report » (https://files.rcsb.org/pub/pdb/validation_reports/mq/3mq1/3mq1_full_validation.pdf) afin de choisir le meilleur dimère en terme de qualité. On peut par exemple comparer la qualité des chaînes entre elles :

Mol	Chain	Length	Quality of chain
1	A	103	
1	B	103	
1	C	103	
1	D	103	
1	E	103	
1	F	103	

- Molecule 1: Mite allergen Der p 5



On voit que seule la chaîne A ne contient pas de valeurs aberrantes (en rouge ci-dessus). Concernant les angles de torsion du squelette protéique, on peut également noter (tableau ci-dessous) qu'un pourcentage de 1 % de valeurs aberrantes dans le diagramme de Ramachandran apparaît pour la chaîne C, raison de plus de l'éliminer. Si on s'attarde sur les chaînes A, B et F pour lesquelles 97 % des résidus ont été analysés en ce qui concerne les angles de torsion du squelette protéique, c'est B qui a le plus grand pourcentage de résidus favorisés, toutefois, B et F présentent des valeurs aberrantes (tableau ci-dessus), c'est donc définitivement la chaîne A qui se place en première position.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	100/103 (97%)	97 (97%)	3 (3%)	0	100	100
1	B	100/103 (97%)	99 (99%)	1 (1%)	0	100	100
1	C	100/103 (97%)	97 (97%)	2 (2%)	1 (1%)	17	48
1	D	99/103 (96%)	95 (96%)	4 (4%)	0	100	100
1	E	98/103 (95%)	96 (98%)	2 (2%)	0	100	100
1	F	100/103 (97%)	97 (97%)	3 (3%)	0	100	100
All	All	597/618 (97%)	581 (97%)	15 (2%)	1 (0%)	49	81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	63	PRO

C) Use the PDBeFold server (<http://www.ebi.ac.uk/msd-srv/ssm/>) to superimpose the 3D structure of 2JMH and chain A of 3MQ1. Visualize the structure superimposition in Pymol and analyze the result. Which structural feature of 3MQ1 allows the formation of the dimer?

Afin de superposer 2JMH avec la chaîne A de 3MQ1, on utilise PDBeFold et on obtient ceci :

Submission Form
pairwise

3D alignment
multiple

Query

Source: Coordinate file

Upload file: Parcourir... 2jmh.pdb

Select chains: ▼ Find chains

Chains: *(all)

Target

Source: Coordinate file

Upload file: Parcourir... 3mq1.pdb

Select chains: ▼ Find chains

Chains: A

Structure Alignment Results.

Query: 2jmh.pdb:A, 117 residues

NMR SOLUTION STRUCTURE OF BLO T 5, A MAJOR MITE ALLERGEN FROM BLOMIA TROPICALIS

Examined 1 entry, (1 chain). Displaying Match 1 of 1.

Back to query Sort by Q-score

#	Scoring				RMSD	N _{align}	N _d	N _{seq}	Query	Target structure					Title
	Q	P	Z	%						Match	%	N _{res}	*		
1	0.35	1.5	3.5	2.26	80	1	40	100	3mq1.pdb: A	100	101				CRYSTAL STRUCTURE OF DUST MITE ALLERGEN DER P 5

Examined 1 entry, (1 chain). Displaying Match 1 of 1.

Back to query Sort by Q-score

STRUCTURAL ALIGNMENT																								
query.pdb	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24
1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1	0.013	1
	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2	0.013	2
	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3	0.021	3
	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4	0.010	4
	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5	0.000	5
	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6	0.010	6
	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7	0.010	7
	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8	0.000	8
	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9	0.000	9
	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10	0.000	10
	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11	0.000	11
	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12	0.010	12
	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13	0.010	13
	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14	0.000	14
	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15	0.000	15
	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16	0.010	16
	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17	0.010	17
	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18	0.010	18
	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19	0.010	19
	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20	0.010	20
2	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21	0.010	21
	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22	0.010	22
	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23	0.010	23
	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24	0.010	24
	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25	0.010	25
	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26	0.010	26
	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27	0.010	27
	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28	0.010	28
	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29	0.010	29
	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30	0.010	30
	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31	0.010	31
	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32	0.010	32
	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33	0.010	33
	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34	0.010	34
	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35	0.010	35
	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36	0.010	36
	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37	0.010	37
	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38	0.010	38
	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39	0.010	39
	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40	0.010	40
0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	0.010	41	
0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	0.010	42	
0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	0.010	43	
0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	0.010	44	
0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	0.010	45	
0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	0.010	46	
0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	0.010	47	
0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	0.010	48	
0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	0.010	49	
0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	0.010	50	
0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	0.010	51	
0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	0.010	52	
0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	0.010	53	
0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	0.010	54	
0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	0.010	55	
0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	0.010	56	
0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	0.010	57	
0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	0.010	58	
0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.010	59	0.0		

Le Q-score représente la qualité de l'alignement, la RMSD et la longueur de l'alignement sont pris en compte pour son calcul. Sa valeur atteint 1 seulement dans le cas de structures identiques. On voit ici que le Q-score est faible du fait qu'une partie n'est pas alignée.

P-score = $-\log(\text{P-value})$ sachant que la P-value mesure la probabilité d'arriver à la même qualité de match par chance. Plus grand est le P-score, plus le match est significatif. Les P-scores inférieurs à 3 indiquent des matchs insignifiants. La P-value est surtout utilisée pour classer les matchs et ici il n'y a pas de comparaison (on a réalisé qu'un match) donc difficile de donner une signification au P-score qui dépend directement de la P-value.

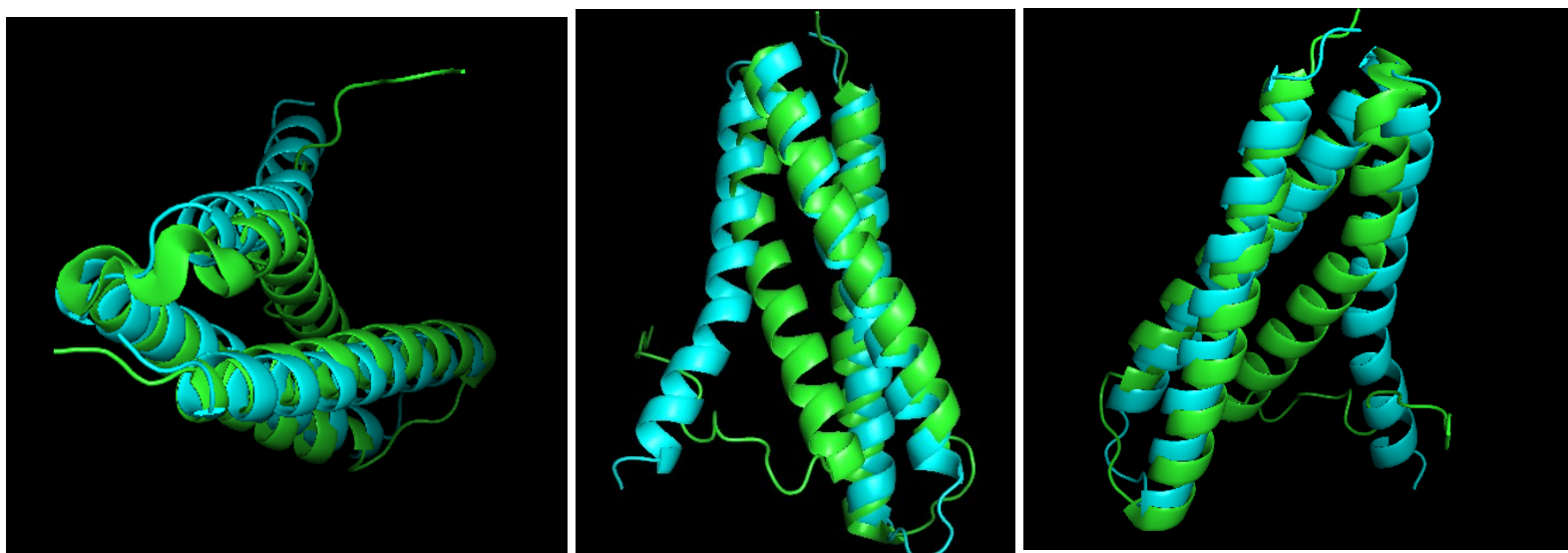
Le Z-score est calculé à partir de l'hypothèse selon laquelle la qualité de l'alignement 3D, en choisissant au hasard les structures dans la base de données, obéit aux statistiques gaussiennes. Le score Z est ensuite associé à la valeur P par le type de relation suivant: $P = \text{erfc}(Z / 2)$. Plus le Z-score est haut, plus le match est significatif. Le Z-score est surtout utilisé pour classer les matchs et ici il n'y a pas de comparaison donc difficile de lui donner une signification.

On peut également noter qu'on a un pourcentage d'identité ($\%_{\text{seq}}$) de 40 % ($\%_{\text{seq}} = N_{\text{ident}} / N_{\text{align}}$). Il faut noter que le pourcentage d'identité est calculé à partir de la structure 3D et non à partir de la séquence 2D, 2 séquences presque identiques peuvent donc avoir un faible pourcentage d'identité si leur repliement est légèrement différent.

On peut également noter que 100% des structures secondaires de 2JMH sont retrouvées dans 3MQ1.

La RMSD (root-mean-square deviation) mesure la distance en angströms entre les atomes de carbone alpha qui correspondent entre eux/se superposent. Donc plus la RMSD est grande, plus les structures correspondantes sont distantes. Notre RMSD est ici égale à 2.26 Å ce qui n'est pas si mal car on sait que la résolution de la structure de la chaîne A de 3MQ1 est de 2,8 Å (X-Ray). Cela signifie que notre RMSD est en dessous de la résolution donc nos structures présentent une superposition comprise entre 0 et $2.8 + 2.26 = 5.06$ Å.

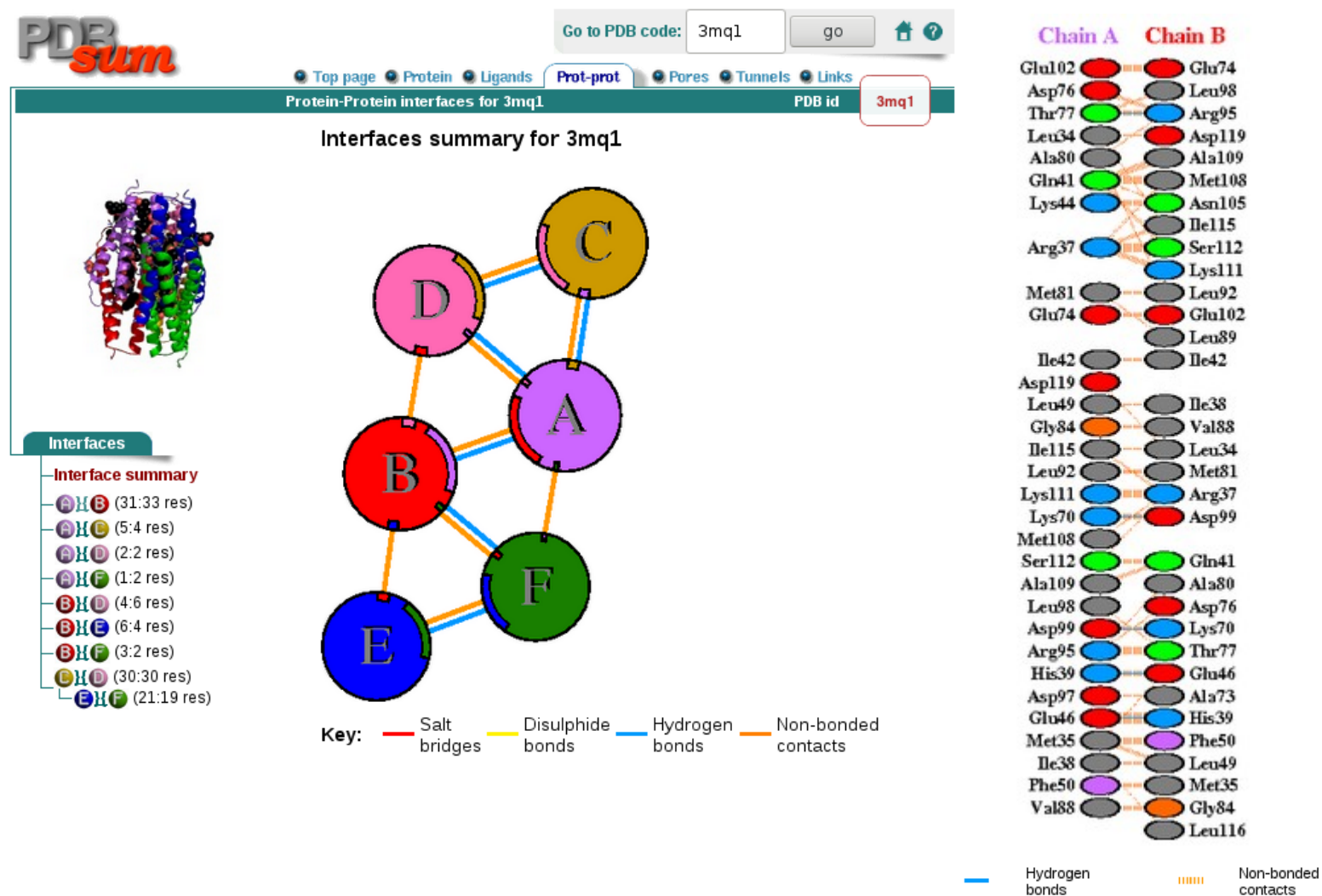
On peut visualiser la superposition via Pymol :



On voit qu'il y a une superposition globale, toutefois on voit bien qu'une grosse partie d'une des 3 hélices (image du milieu) n'est pas superposée comme déjà vu précédemment avec PDBFold où l'importante partie en gris représentait la partie non superposée. Il faut également noter qu'il n'y a

pas de superposition de structure au niveau des jonctions entre les hélices (petit bout gris dans PDBeFold).

Concernant la dimérisation de 3MQ1, elle a lieu au niveau des hélices alpha via des liaisons hydrogène et Van der Waals, on peut le voir via PDBsum, j'ai sélectionné les chaînes A et B comme exemple (<https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=3mq1&template=interfaces.html&c=999>) :



On voit clairement dans le tableau ci-dessous que la surface d'interaction est particulièrement importante entre les hélices A et B puis entre C et D et enfin entre E et F, de même pour le nombre de liaisons hydrogène des 3 dimères qui est respectivement de 5, 9 et 4.

Interface statistics

Chains	No. of interface residues	Interface area (Å ²)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
AHB	31 : 33	1715:1671	-	-	5	117
AHE	5 : 4	256:241	-	-	1	24
AHD	2 : 2	123:120	-	-	3	9
AHF	1 : 2	98:93	-	-	-	4
BHD	4 : 6	310:282	-	-	-	17
BHE	6 : 4	231:268	-	-	-	19
BHF	3 : 2	142:139	-	-	2	7
EHD	30 : 30	1663:1637	-	-	9	117
& EHF	21 : 19	1335:1356	-	-	4	66

D] Read the paper of Mueller et al. Which residues are considered as important for the dimerization of the protein? Use the sequence alignment provided on the VU to analyze the sequence conservation at these positions.

D'après l'article de Mueller et al. , les résidus ayant une importance dans la dimérisation sont les résidus 84 à 88 (GVRGV pour Der p 5). Der p 5 et Blo t 5 sont semblables, les résidus de leurs hélices sont conservés mis à part au niveau des résidus 84 à 88 ce qui fait de Blo t 5 un monomère contrairement à Der p 5 qui se dimérise comme vu précédemment. Le résidu Val 88 d'une chaîne s'insère dans une encoche créée par Gly 84 de l'autre chaîne, le Val 85 contribuant à des interactions hydrophobes pour créer une « valine zipper », analogue aux « leucine zippers ».

Concernant la conservation de séquence dans cette région, Der p 5 et Der f 5 ont exactement la même séquence, on peut en conclure que Der f 5 se dimérise. Blo t 5 reste sous forme de monomère comme vu précédemment. On remarque ensuite que Der p 21 et Der f 21 ont la même séquence dans cette région. Si on fait l'hypothèse que Der p 21 se dimérise, il en ira donc de même pour Der f 21. Enfin, on ne peut pas vraiment conclure quant à une dimérisation ou pas de Sui m 5 et Lep d 5 car les séquences de cette région varient beaucoup, on peut toutefois remarquer qu'en position 85, il n'y aura pas de « valine zippers » comme pour Der p 5 mais ça n'apporte pas d'information qui nous permettrait de prédire ou pas une dimérisation.