

Structural bioinformatics Project – Assignment 3: Structural analysis.

- Get a set of 4-6 structures from the PDB that belong to the family of your protein of interest. Try to get a set that is not biased, so avoid pairs of proteins that are identical or very similar. How would you do that? What programs would you use? What are the PDB IDs of the structures you have selected?**

Make an unbiased search on uniprot, using the psiblast command:

```
psiblast -query 6xg6.fa -db ~/Documents/databases/uniprot_sprot.fasta
-num_iterations 5 -out 6xg6_uniprot_5.out
```

Then, search in pdb, from our unbiased search to find the list of possible searches.

```
psiblast -query 6xg6.fa -num_iterations 5 -out_pssm 6xg6_sprot5.pssm -out
6xg6_uniprot_5.out -db ~/Documents/databases/uniprot_sprot.fasta

psiblast -db ~/Documents/databases/pdb_seq -in_pssm 6xg6_sprot5.pssm -out
6xg6_pdb_sprot5.out
```

These are the PDB structures we have selected from the double filtering we have performed:

Sequences producing significant alignments:	Score (bits)	E Value
1sf8_A mol:protein length:126 Chaperone protein htpG	153	2e-43
1qy5_A mol:protein length:269 Endoplasmin	275	6e-87
1y6z_A mol:protein length:263 heat shock protein, putative	279	7e-89
1yt2_A mol:protein length:273 Endoplasmin	276	2e-87
1qy8_A mol:protein length:269 Endoplasmin	275	6e-87

List of PDB IDs: 1SF8, 1QY5, 1Y6Z, 1YT2, 1QY8

List of programs used:

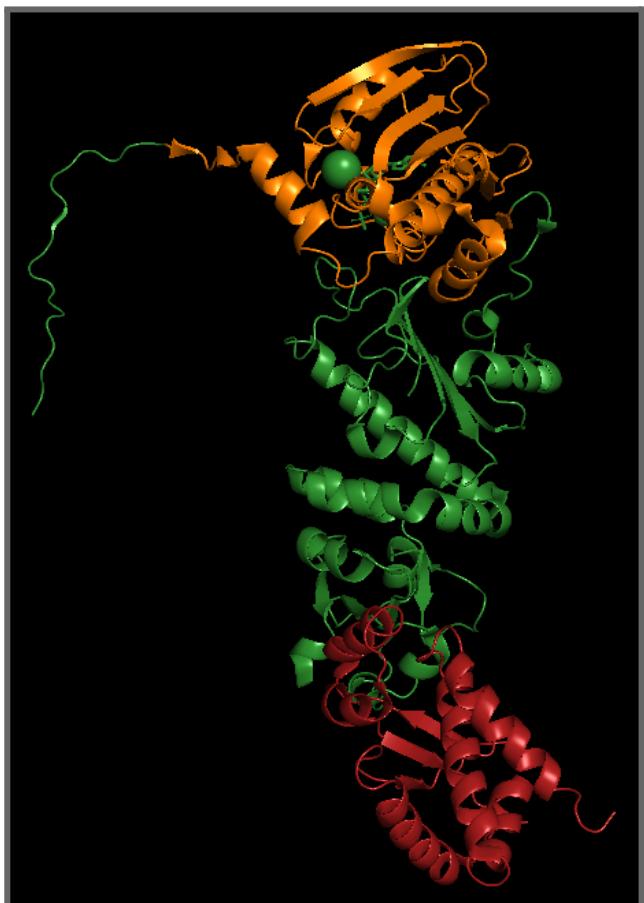
- **PSI-BLAST:** For initial sequence and structural searches.
- **Uniprot / AlphaFold:** For structural comparisons and obtaining the PBD structures.

2. Superimpose the structures you selected in question 1. Are they structurally similar? What is their RMSD? Can you identify some regions with higher variability? Why do you think these regions are more variable? What about the most conserved regions of your protein (the ones you described in assignment 1, question 6 and assignment 2, question 4), are they structurally variable or not? Can you relate this to the function of the protein? Include pymol images to support your explanation.

We tried to superimpose the structures obtained from the exercise 1, by focusing ocd n chain A:

```
remove resn hoh  
(select and rename all the A chains to superimpose them to the chain_A of  
the 6xg6)
```

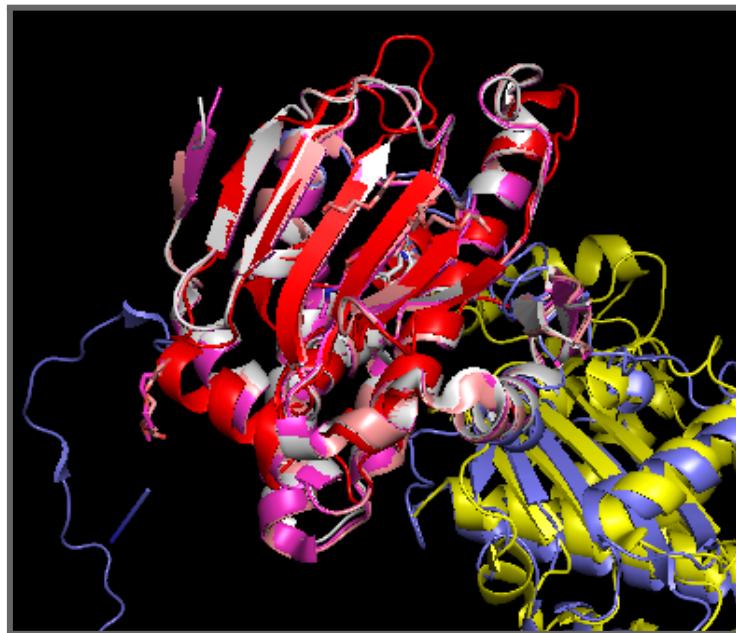
```
super 1sF8_A, chain_A, object = aln4, cutoff = 5.0  
>> RMSD = 2.957 (535 to 535 atoms)  
  
super 1qy5_A, chain_A, object = aln4, cutoff = 5.0  
>>> RMSD = 7.991 (1231 to 1231 atoms)  
  
super 1y6z_A, chain_A, object = aln4, cutoff = 5.0  
>>> RMSD = 2.824 (1468 to 1468 atoms)  
  
super 1yt2_A, chain_A, object = aln4, cutoff = 5.0  
>>> RMSD = 7.903 (1265 to 1265 atoms)  
  
super 1qy8_A, chain_A, object = aln4, cutoff = 5.0  
>>> RMSD = 7.629 (1252 to 1252 atoms)
```



On the left we have a colored representation of the 3 domains in our protein(6xg6).

On the left the studied protein, colored in purple, is superposed in some parts by the different proteins selected.

The N-terminal domain, what is mainly superposed in all the sequences, yet there is a lot of variability in the superposing of the different selected proteins:



This is the sequence with the aligned residues highlighted in red:

/1sf8	531	536	541	546	551	556	561	566	571	576	581	586	591	596	601	606	611	616	621	/B/B/
ERVKDVRLTHRLTDTPAIVSTDDEMSTQMAKLF	AAAGQKVPEVKYIFELNPDHVLVKRAADTEAKFSEW	VELLDQALLAERTGLEDPNLFIRRMNQLLV	S																	
/1qy5	101	106	111	116	121	126	131	136	141	146	151	156	161	166	171	176	181	186	191	196
NSLYKKNKEIFLRELISNASDALDKIRLISLTDENALAGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKNLGTIAKSGTSEFLNKMT	EAQEDGQSTSELIGQFGVGF	S																		
/1y6z	31	36	41	46	51	56	61	66	71	76	81	86	91	96	101	106	111	116	121	126
YKNTFKAYDDPLAYVHFNVEGQISNISLYIPGSLPWELS	KNMDEESRGIRLYVKRVIDKFSE	IPRJULTRLGIVDSENPLNVGREILQSKMLSIINKRIVLK	S																	
/1yt2	96	101	106	111	116	121	126	131	136	141	146	151	156	161	166	171	176	181	186	191
KLIINSLYKKNKEIFLRELISNASDALDKIRLISLTDENALAGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKNLGTIAKSGTSEFLNKMT	EAQEDGQSTSELIGQF	G																		
/1qy8	101	106	111	116	121	126	131	136	141	146	151	156	161	166	171	176	181	186	191	196
NSLYKKNKEIFLRELISNASDALDKIRLISLTDENALAGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKNLGTIAKSGTSEFLNKMT	EAQEDGQSTSELIGQFGVGF	S																		
/6xg6	86	91	96	101	106	111	116	121	126	131	136	141	146	151	156	161	166	171	176	181
SSTESVQGSTSKHEFQAETKKLLDIVARS	LYSEKEVFI	RELISNASDALEKLRLKLVS	DQALPEMEIHLQTNAEKGTITI	QDTGIGMTQEELVSNLGTIA	RSGSKAFL	S														

621	/B/B/499	506	511	516	521	526	531	536	541	546	551	556	561	566	571	576	581	586		
ILLVS	MRGSHHHHHHG	SFIDRVKALLGERVKD	VRLTHR	LTDTPAIVSTDDEMSTQMAKLF	AAAGQKVPEVKYIFELNPDHVLVKRAADTEAKFSEW	S														
191	196	201	206	211	216	221	226	231	236	241	246	251	256	261	266	271	276	281	286	291
LIGQFGVGFYS	YSAFLVADKVIVTSKHNNDTQHIWESDSNEFSVIA	DPRGNTLGRGTTITVLK	EEASDYLELDTIKLN	VKKYSQFINFPIYWSSKTETVE	PM															
121	126	131	136	141	146	151	156	161	166	171	176	181	186	191	196	201	206	211	216	221
IINKRIVLKSISMMKG	LKETGGDKWTKF	LNTFGKYLKIGVVEDKEN	QEEIASLVEFYSINS	SGDKKTDLDSYIENMKEDQKCIYYISGEN	KKTAQNNSPLEK	KLK														
86	191	196	201	206	211	216	221	226	231	236	241	246	251	256	261	266	271	276	281	286
ISTSELIGQFGVGFYS	SAFLVADKVIVTSKHNNDTQHIWESDSNEFSVIA	DPRGNTLGRGTTITVLK	EEASDYLELDTIKLN	VKKYSQFINFPIYWSSKTETVE	PM															
191	196	201	206	211	216	221	226	231	236	241	246	251	256	261	266	271	276	281	286	291
LIGQFGVGFYS	SAFLVADKVIVTSKHNNDTQHIWESDSNEFSVIA	DPRGNTLGRGTTITVLK	EEASDYLELDTIKLN	VKKYSQFINFPIYWSSKTETVE	PM															
176	181	186	191	196	201	206	211	216	221	226	231	236	241	246	251	256	261	266	271	276
ARSGSKAFL	DALQNQAEASS	KIIGQFGVGFYS	AFMVA	DRVEV	YRS	SAPGSL	GYQWLSDSGVFEIA	EASGVRTGKII	IHLKSDCKEFS	SEARVRDWTKY	S									

We tried to understand the structure of our protein better by superimposing some structures found in Uniprot that had an Alpha Fold profile where we could obtain the PDBs. This information was obtained from the first command in the previous step. ↴

```
psiblast -query 6xg6.fa -db ~/Documents/databases/uniprot_sprot.fasta
-num_iterations 5 -out 6xg6_uniprot_5.out
```

This is the list of selected structures we obtained the PDBs from Alphafold:

Sequences producing significant alignments:	Score (bits)	E Value
sp P24724 HSP90_THEPA Heat shock protein 90 OS=Theileria parva GN...	188	1e-49
sp P04809 HSP83_DROPS Heat shock protein 83 OS=Drosophila pseudoo...	183	8e-48
sp P06660 HSP85_TRYCR Heat shock-like 85 kDa protein OS=Trypanoso...	208	1e-56
sp Q8SSE8 HSP82_ENCCU Heat shock protein 90 OS=Encephalitozoon cu...	263	1e-76
sp P36181 HSP80_SOLLC Heat shock cognate protein 80 OS=Solanum ly...	332	8e-103
sp P0C938 HTPG_PORGI Chaperone protein htpG OS=Porphyromonas ging...	207	2e-56
Q12931 (target sequence Uniprot ID), P24724, P04809, P06660, Q8SSE8, P36181, P0C938		

Once we download the structures from the Alphafold, we upload them to Pymol. Then we superpose each of the structures specifically to the chain_A in our protein.

For each one of the superpositions we have also found the RMSD:

```
remove resn hoh

super p36181, chain_A, object = aln4, cutoff = 5.0
>>> Executive: RMSD =     8.518 (3452 to 3452 atoms)

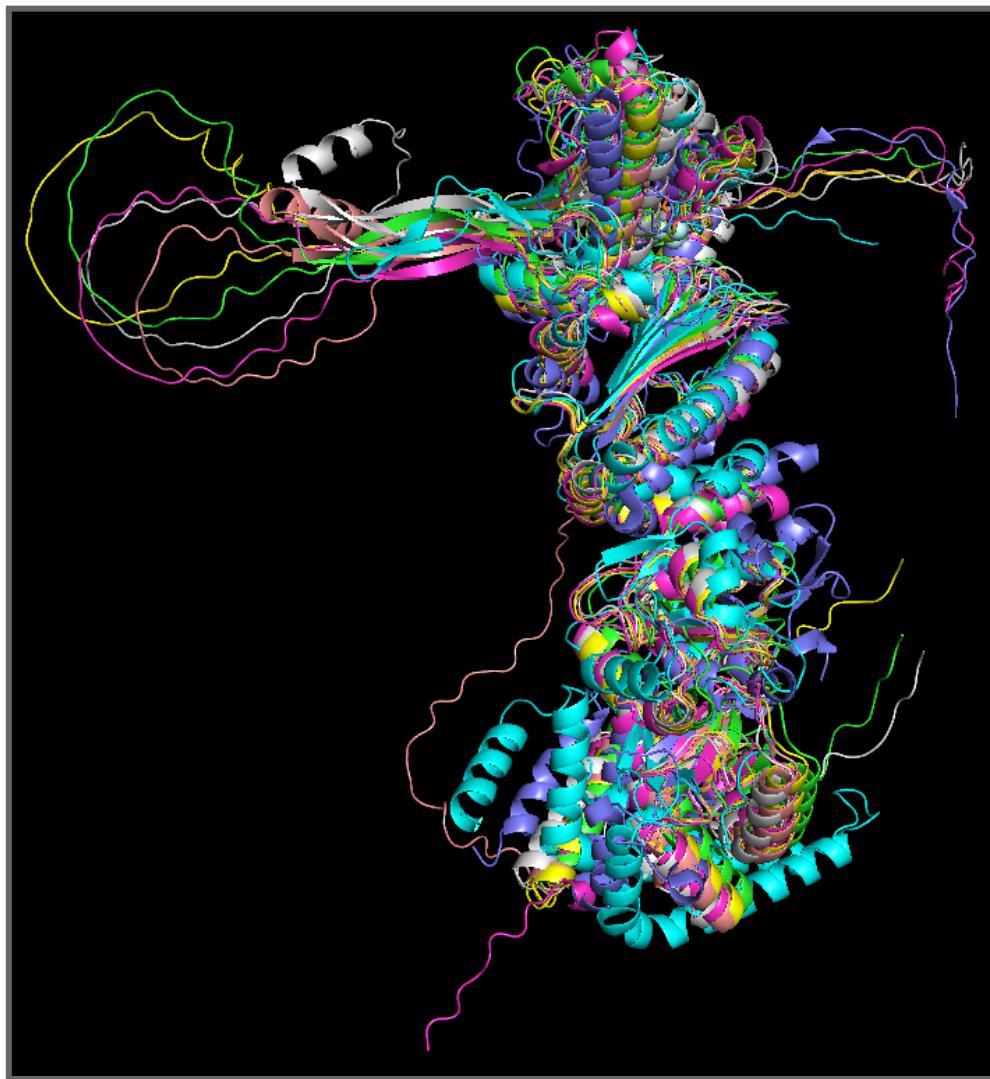
super p0c938, chain_A, object = aln4, cutoff = 5.0
>>> Executive: RMSD =    11.751 (3394 to 3394 atoms)

super q8sse8, chain_A, object = aln4, cutoff = 5.0
>>> Executive: RMSD =    6.836 (3559 to 3559 atoms)

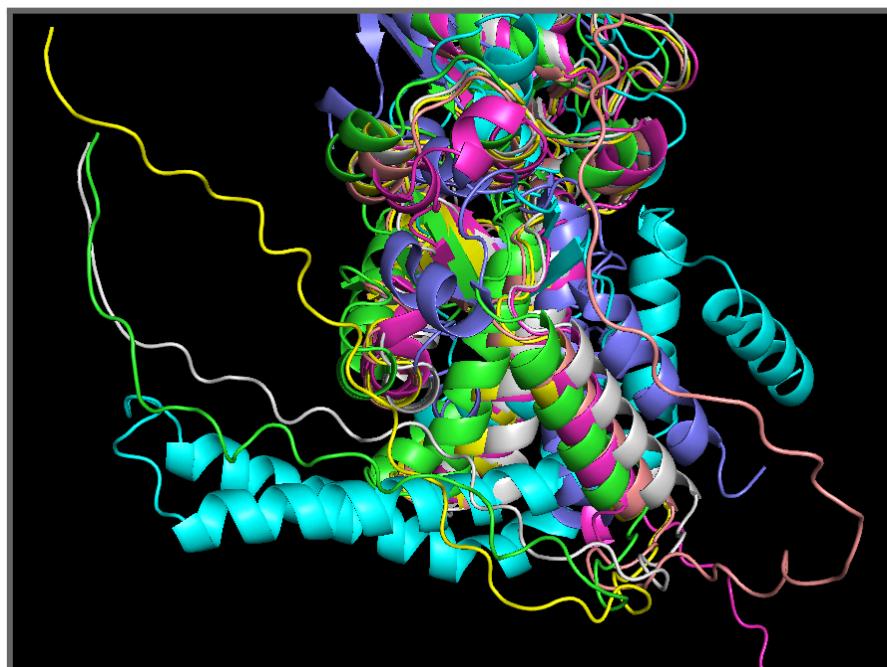
super p06660, chain_A, object = aln4, cutoff = 5.0
>>> Executive: RMSD =    7.095 (3601 to 3601 atoms)

super p04809, chain_A, object = aln4, cutoff = 5.0
>>> Executive: RMSD =    7.200 (3566 to 3566 atoms)

super p24724, chain_A, object = aln4, cutoff = 5.0
>>> Executive: RMSD =    9.760 (3346 to 3346 atoms)
```



On the top of the image we will find the N-terminal domain, then the middle domain and at the bottom the C-terminal. We can see a lot of variability specially on the loops of the proteins.



In the C-terminal domain, the turquoise blue (p0c938) looks like it's bigger than the rest of proteins since it has more alpha helices.

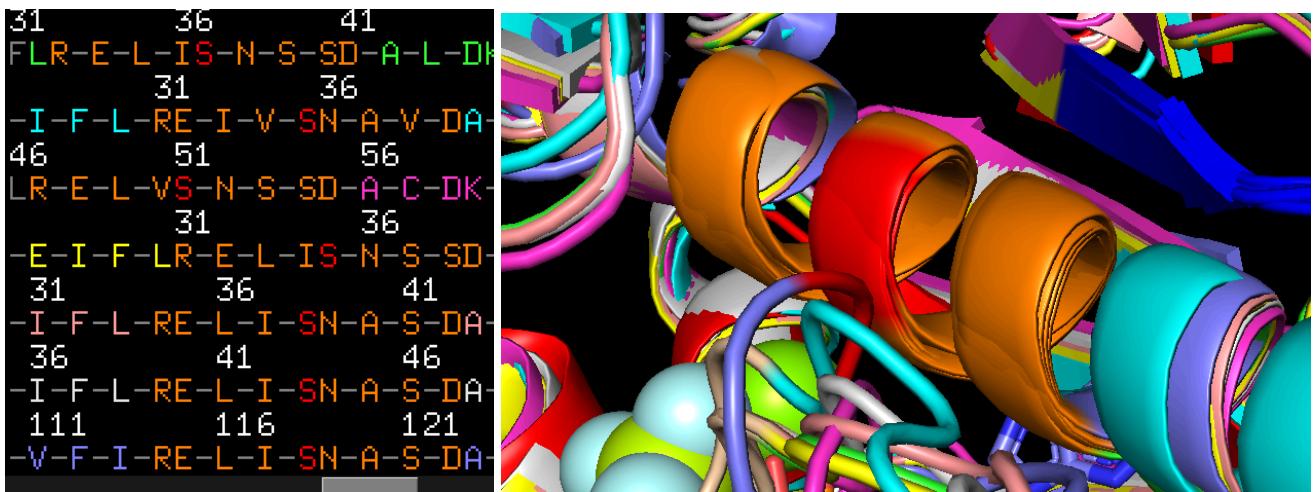


In the N-terminal domain, we can see variability between the types of beta sheets they have before the loops.

116	121
-SMIGQFGVG	
1	116
-AIIGHFGLG	
136	141
-SNLIGQFGLG	
116	12
-SMIGQFGVG	
16	121
-SMIGQFGVG	
21	126
-SMIGQFGVG	
196	201
:SKIIGQFGVG	



This section is found in the N-terminal domain and is structurally not variable but it is flexible since it is a loop, but they are conserved because it is related to the ATP binding site, which is important to for the function of the protein



This section is very conserved in general and not structurally variable, is not flexible since it is an alpha helix, the Serine is a ADP binding site, which can be found in a well conserved region.

3. **This is the most important part of the submission: Choose the region (or regions) that you think are the most important for the protein function. Then, describe this region, why can it carry out the function that it does? What are the weak interactions that allow this function to happen? Include PyMOL images to support your explanation. You can inspire yourselves with the works of students from previous years, find them in: <https://sbi.upf.edu/web/index.php/courses/undergraduatedprojects>. Here you have some examples of how to orient this question:**

Note: Alignment pictures from assignment 2

Features

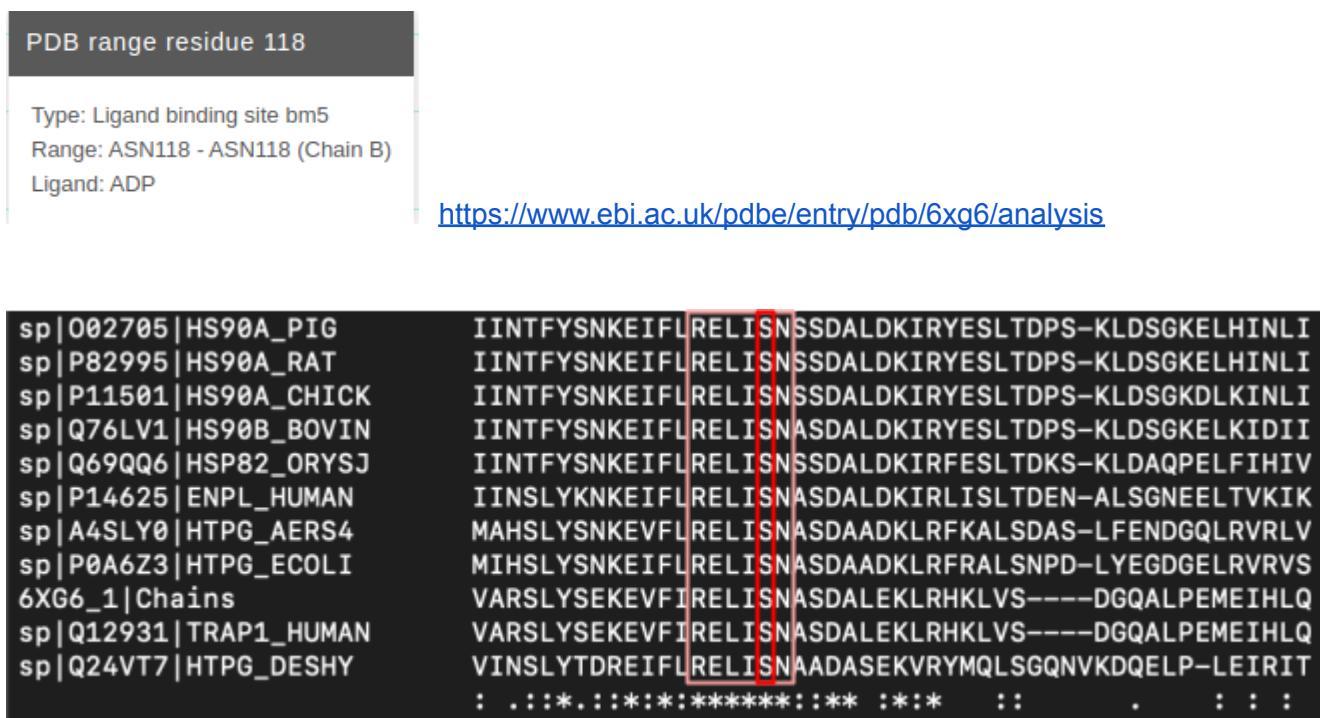
Showing features for binding siteⁱ.

Features			
Showing features for binding site ⁱ .			
			Download
1	50	100	150
	200	250	300
	350	400	450
Type	ID	Position(s)	Description
-- Select --			
▶ Binding site	119		ATP (UniProtKB ChEBI)
▶ Binding site	158		ATP (UniProtKB ChEBI)
▶ Binding site	171		ATP (UniProtKB ChEBI)
▶ Binding site	205		ATP (UniProtKB ChEBI)
▶ Binding site	402		ATP (UniProtKB ChEBI)

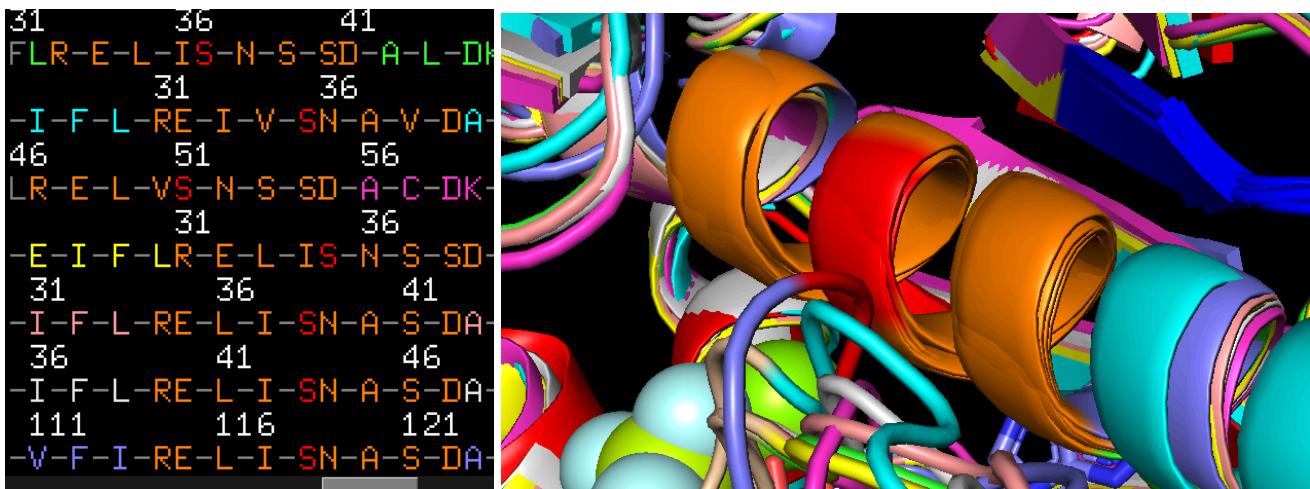


<https://www.rcsb.org/sequence/6XG6>

118 (CONSERVED REGION) ADP- Binding site [LISN]

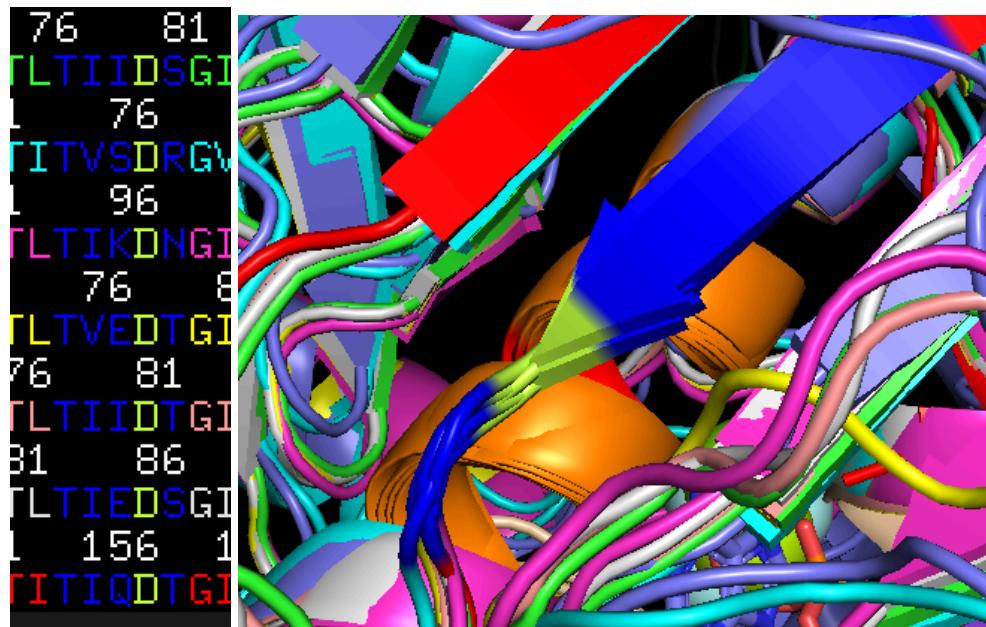


The Serine (in red) is the ADP binding site found in a conserved site (orange) among all the sequences, as seen in the picture above.



158 [D] atp binding site

sp 002705 HS90A_PIG	PNKQDR T LTIVDTGIGMTKADL NNL G T I A KSGTKAFMEALQAGAD-----
sp P82995 HS90A_RAT	PNKQDR T LTIVDTGIGMTKADL NNL G T I A KSGTKAFMEALQAGAD-----
sp P11501 HS90A_CHICK	PNKHDR T LTIVDTGIGMTKADLVNNL G T I A KSGTKAFMEALQAGAD-----
sp Q58FF7 H90B3_HUMAN	PNPQERT L ALVDTGIGMTKADL NNL R T I A KSGTKACMEALQA-----
sp Q76LV1 HS90B_BOVIN	PNPQERT L TVDTGIGMTKADLVNNL G T I A KSGTKAFMEALQAGAD-----
sp Q69QQ6 HSP82_ORYSJ	PDKASNTLSI I D S G V G M T K S D L V NNL G T I A RSGTKEFMEALAAGAD-----
sp P14625 ENPL_HUMAN	CDKEKNLLHV T DTGVGM T REELVKNL G T I A KSGTSEFLNK M TEAQEDGQS
sp A4SLY0 HTPG_AERS4	VDKENRT L TISDNGIGM T RDV Q VIEHL G T I A KSGTAEFFKNLSDQG---R
sp P0A6Z3 HTPG_ECOLI	FDKD K R T L T ISDNGVG M TRDEV I D H G T I A KSGTKSFLESLGSDQ---K
6XG6_1 Chains	TNAEK G TITI Q DTGIGMTQEELVNL G T I A RSGSKAFLDALQNQAEAS---
sp Q12931 TRAP1_HUMAN	TNAEK G TITI Q DTGIGMTQEELVNL G T I A RSGSKAFLDALQNQAEAS---
sp Q24VT7 HTPG_DESHY	PDENAK T L T I A DAGIGMTKEDLIENIGTIAHSGSKAFVQRLAEAGDKKD-----
	: : : * *:***: : : .: ***:***: . . :



198–205 → IGQFGVGF → SUPER CONSERVED REGION [ATP active site]

ATP active site → N-Terminal

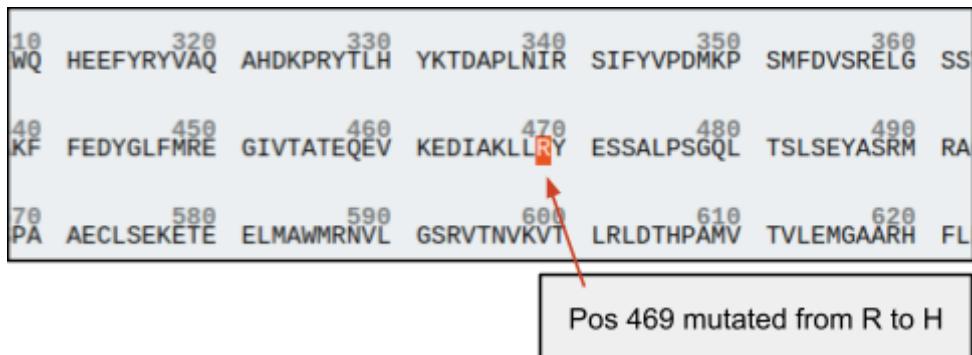
sp 002705 HS90A_PIG	DISM I GQFGVGFYSAYLVAEKVTVITKHND---DEQYAWESSAGGSFTVR
sp P82995 HS90A_RAT	DISM I GQFGVGFYSAYLVAEKVTVITKHND---DEQYAWESSAGGSFTVR
sp P11501 HS90A_CHICK	DISM I GQFGVG S YSAYLVAEKVTVITKHND---DEQYAWESSAGGSFTVR
sp Q76LV1 HS90B_BOVIN	DISM I GQFGVG F YSAYLVAEKVVVITKHND---DEQYAWESSAGGSFTVR
sp Q69QQ6 HSP82_ORYSJ	DVSM I GQFGVG F YSAYLVAERVVTTKHND---DEQYVWESQAGGSFTVT
sp P14625 ENPL_HUMAN	TSEL I GQFGVG F YSAFLVADKVIVTSKHNN---DTQHIWESDSN-EFSVI
sp A4SLY0 HTPG_AERS4	DSQL I GQFGVG F YSAFIVADKTVVSRAAGTAPEQGVQWESEGEGSFTVA
sp P0A6Z3 HTPG_ECOLI	DSQL I GQFGVG F YSAFIVADKTVVRTRAAGEKPENGVFWESEGEGEYTV
6XG6_1 Chains	-SKI I GQFGVG F YSAFMVADRVEVYRSAAPG-SLGYQWLSDGSGVFEIA
sp Q12931 TRAP1_HUMAN	-SKI I GQFGVG F YSAFMVADRVEVYRSAAPG-SLGYQWLSDGSGVFEIA
sp Q24VT7 HTPG_DESHY	DVN I GQFGVG F YSAFMVADKVSLSRSYEPD-AQGYRWESDGRGSYSIS
	. :***** * ***:***:*** : :: * * . : :

116	121
-SMIGQFGVG	
1	116
-AIIGHFGLG	
136	141
-SNLIGQFGLG	
116	12
-SMIGQFGVG	
16	121
-SMIGQFGVG	
21	126
-SMIGQFGVG	
196	201
:SKI I IGQFGVG	

402: Arginine (R) significant in ATP hydrolysis → Middle Domain

sp 002705 HS90A_PIG	IMDNCEELIPEYLNFIRGVVDSEDLPLNISREMLQQSKILKVIRKNLVKK
sp P82995 HS90A_RAT	IMDNCEELIPEYLNFIRGVVDSEDLPLNISREMLQQSKILKVIRKNLVKK
sp P11501 HS90A_CHICK	IMDNCEELIPEYLNFMRGVVDSEDLPLNISREMLQQSKILKVIRKNLVKK
sp Q76LV1 HS90B_BOVIN	IMDSCDELIPEYLNFIRGVVDSEDLPLNISREMLQQSKILKVIRKNIVKK
sp Q69QQ6 HSP82_ORYSJ	IMDNCEELIPEWLSFKVKGIVDSEDLPLNISREMLQQNKILKVIRKNLVKK
sp P14625 ENPL_HUMAN	ITDDFHDMMPKYLNFKVKGVVDSDDLPLNVSRETLQQHKLLKVIRKKLVRK
sp A4SLY0 HTPG_AERS4	IMDDAEQFMPTYLRFVKGVLDSDLPLNVSREILQDNKVTVSLRKACSKR
sp P0A6Z3 HTPG_ECOLI	IMDDAEQFMPNYLRFVRGLIDSSDLPLNVSREILQDSTVTRNLRNALTKR
6XG6_1 Chains	IQTAKATDILPKWLRFIRGVVDSEDIPLNLISRELLQESALIRKL RDVLQQR
sp Q12931 TRAP1_HUMAN	IQTAKATDILPKWLRFIRGVVDSEDIPLNLISRELLQESALIRKL RDVLQQR
sp Q24VT7 HTPG_DESHY	IQEKA KDIVPEWLRFARGVV DSEELPLNISRETMQDSALIAKLNKV VTSR
	* . :*** :* * :***:***.:***:*** :*: : : .. :

4. Use MODELLER to create a model of your protein of interest that includes the mutation you chose in the first assignment (assignment 1, question 7). Show pymol images comparing the wild type structure of your protein and the structure of the mutant you just modeled. By comparing the structures hypothesize why the mutation has an effect in the protein function.



We've modified the sequence to obtain a fasta with the two sequences the wild one and the mutated one.

```
perl ~/Documents/perl_scripts/PDBtoSplitChain.pl -i 6xg6.pdb -o 6xg6
```

```
cat 6xg6_mutated.fa > target_template.fa
```

```
cat 6xg6A.fa >> target_template.fa
```

```
clustalw target_template.fa
```

```
perl ~/Documents/perl_scripts/aconvertMod2.pl -in c -out p
<target_template.aln>target_template.pir
```

```
mod10.5 modeling.py
```

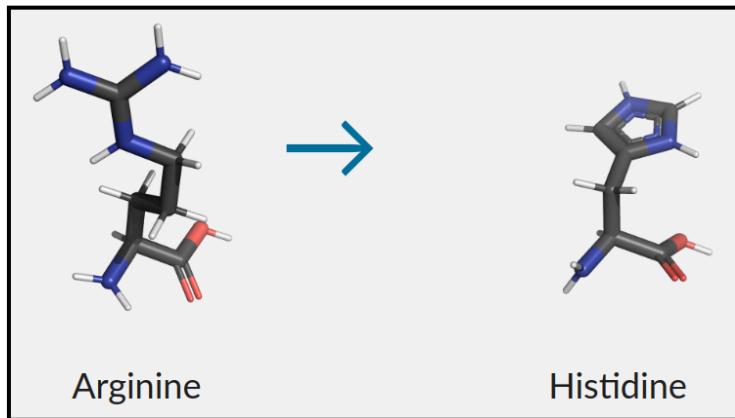
```
# Homology modeling with multiple templates
from modeller import *          # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class

log.verbose()      # request verbose output
env = environ()   # create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = [ '.', '../atom_files' ]

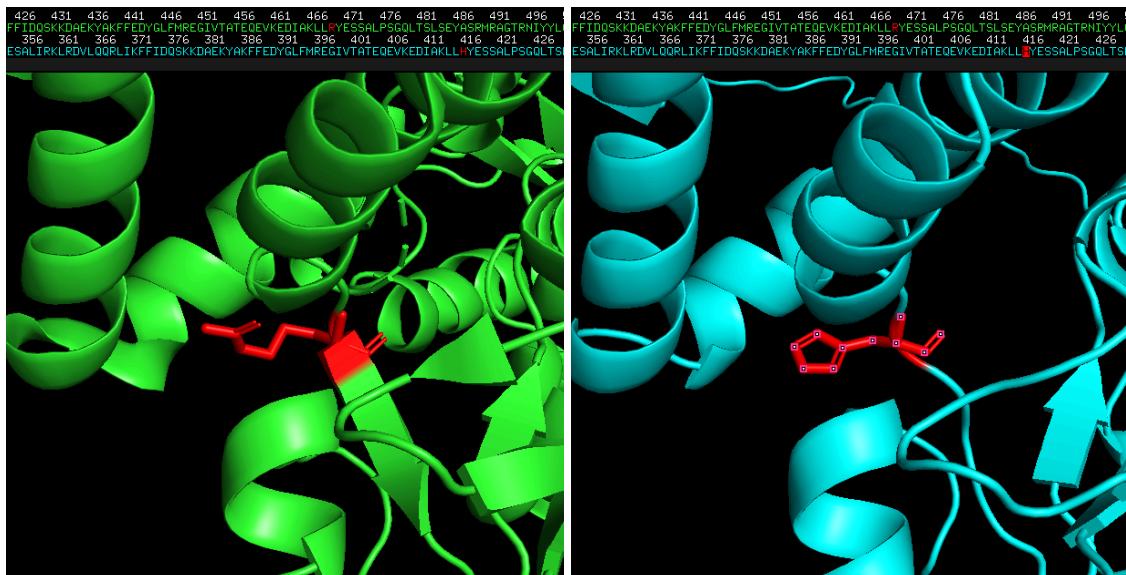
a = automodel(env,
              alnfile = 'target_template.pir', # alignment filename
              knowns = ('6xg6A'),        # codes of the templates
              sequence = '6xg6_mutated') # code of the target
a.starting_model= 1             # index of the first model
a.ending_model  = 2             # index of the last model
                               # (determines how many models to calculate)
a.make()                      # do the actual homology modeling
```

Why mutation has an effect on the protein function?



Both amino acids are positive but we can observe that it is going from a big molecule to a smaller one which also has an aromatic ring. This mutation alters an Arginine that is significant in ATP hydrolysis.

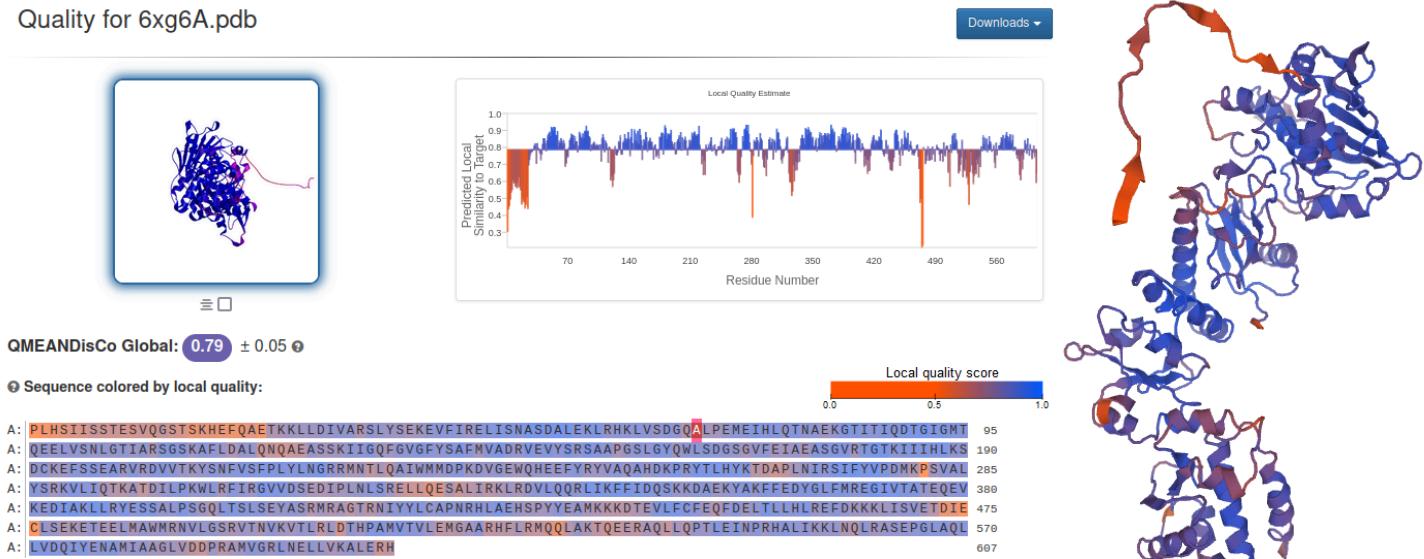
sp P02705 HS90A_PIG	CLELFTELAED-KENYKKFYEEQFSKNIKLG--IHEDSQNRKKLSELRLRYH
sp P82995 HS90A_RAT	CLELFTELAED-KENYKKFYEEQFSKNIKLG--IHEDSQNRKKLSELRLRYH
sp P11501 HS90A_CHICK	CLELFTELAED-KENYKKFYEEQFSKNIKLG--IHEDSQNRKKLSELRLRYH
sp Q76LV1 HS90B_BOVIN	CLELFSELAED-KENYKKFYEAFSKNLKLG--IHEDSTNRRRLSELRLRYH
sp Q69QQ6 HSP82_ORYSJ	CVELFFEIAEN-KEDYNKFYEAFSKNLKLG--IHEDSTNRTKIAELLRHYH
sp P14625 ENPL_HUMAN	TLDIMKKIADD-KYN-DTFWKEFGTNKLKG--VIEDHSNRTRLAKLLRIFQ
sp A4SLY0 HTPG_AERS4	VLTMLAKLAKDDAEKYAKFWSEFGNVLKEG--PAEDYANREEIAKLLRFA
sp P0A6Z3 HTPG_ECOLI	VLQMЛЕЛКЛАКДДАЕКҮҚТФWQQFGLVLKEG--PAEDFANQEAIAKLLRFA
6XG6_1 Chains	LIKFFIDQSKKDAEKYAKFFEDYGLMREGIVTATEQEVKEDIAKLLRYE
sp Q12931 TRAP1_HUMAN	LIKFFIDQSKKDAEKYAKFFEDYGLMREGIVTATEQEVKEDIAKLLRYE
sp Q24VT7 HTPG_DESHY	FLKFLLDDQAKNEPEIFKEFWNEFSIFLKEG--AANDFTHRQEILKLLRFE



5. Use Qmeans to compare the energy profiles of the wild type protein with the structure of the mutant you modeled in the previous question. Is this mutation improving or worsening the energies of your protein? Make sure that the two proteins that you are comparing have similar lengths.

Model of the Wild type

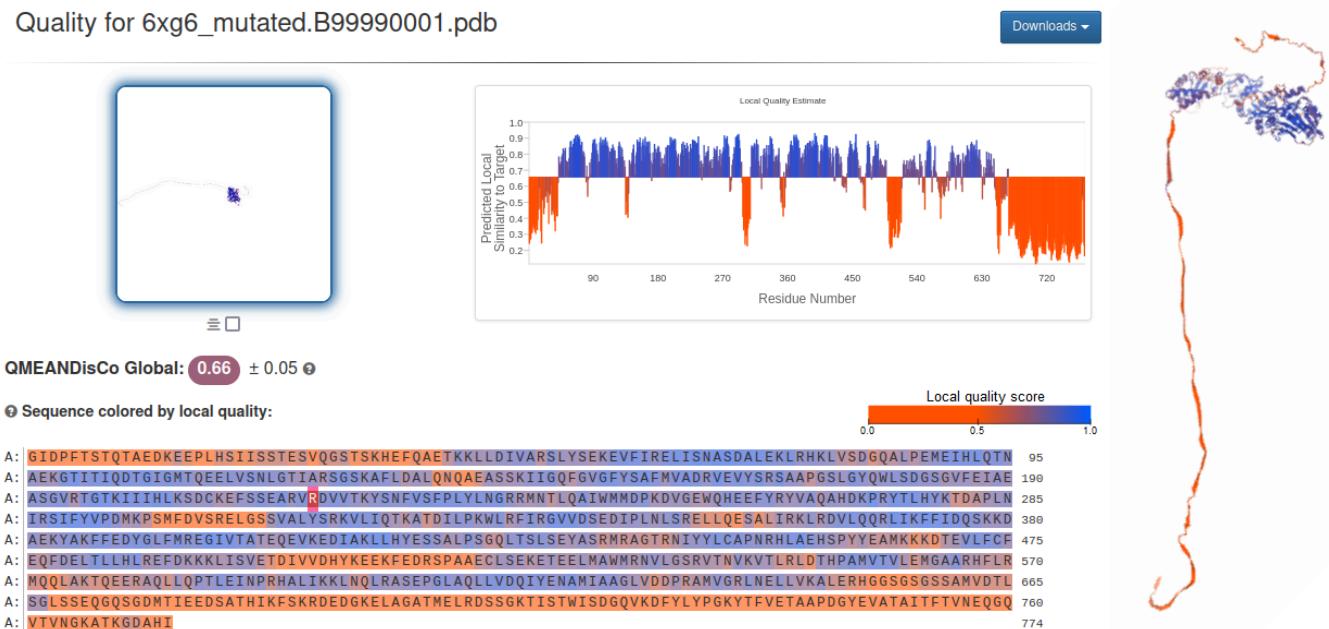
Quality for 6xg6A.pdb



The model of the wild type has a good quality, with a length sequence of 607, this is only the chain A of the 6xg6, PDB structure. The beginning of the sequence has a little chunk of bad quality with 4 beta sheets.

Model 1: Mutated

Quality for 6xg6_mutated.B99990001.pdb



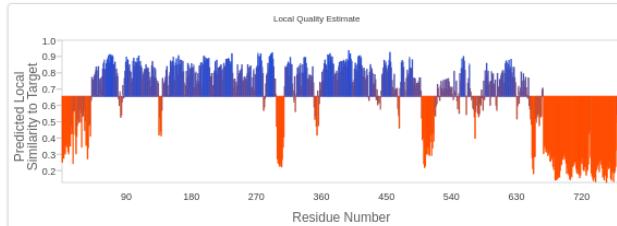
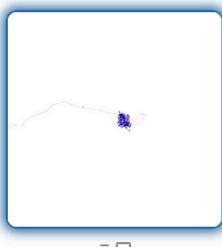
This is one of the models we have produced with modeler, it maintains the bad quality in the beginning but also has a large region that cannot be properly modeled since the sequence is larger than the target sequence, having 774 aminoacids.

Not taking that into account this mutation does not really affect the protein energetically speaking since it has more or less the same high peaks and low peaks.

Model 2: Mutated

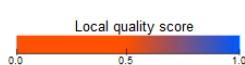
Quality for 6xg6_mutated.B99990002.pdb

Downloads ▾



QMEANDisCo Global: 0.66 ± 0.05 ⚡

Sequence colored by local quality:



```

A: GIDPFTSTQTAEDKEEPLHSIISSTESVQGSTSKHEFQAETKKLLDIVARSLYEKEVFIRELINASDALEKLRHKLVSDGQALPEEIHLQTN 95
A: AEKGTTIIDTGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEASSKIIGQFGVGFSAFMVADRVEVYRSAAPGLGYOWLSDGSGVFEIAE 199
A: ASGVRTGTKIIIHLKSDCKEFSSEARVRDVVTTKYSNEVSFPLYLNGRRMNTLQAIWMMDPKDVGEWFQHEEFYRYVAQAHDKPRYTLHYKTDAPLN 285
A: IRSIFYVPDNKPSNFDSRELGSSVALYSRKVLIQTKATDILPKWLRFIRGVVDSEIDPLNLSRELLQESALIRKLRDVLQORLIKFFFDQSKDD 380
A: AEKYAKFEDYGLFMREGIVTATEQEVKEDIAKLLHYESSSALPS6QLTSLEYASRMRAGTRNIYYLCAPNRHLAEHSPYYEAMKKKDTEVLFC 475
A: EOFDELTLLHLREFDKKKLISVETDIVDHYKEEKFDRSPAAECLSEKETELMAWMWRNVLGSRTVNVKVTLRLDTHPAMVTVLEMGAARHFLR 570
A: MQQLAKTQEERAQLLQPTLEINPRHALIKLNQLRASEPGLAQLLVDQIYENAMIAAGLVDDPRAMVGRNELLVKALERHGGSGSSAMVDTL 665
A: SGLSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEOG 768
A: VTVNGKATKGDAH1 774

```

The second model does not differ that much from the previous one, so the same can be said. It is true that we could have taken the amino acids that cannot be modeled out. We will take that out for future deliveries.