

Structural bioinformatics Project - Assignment 3:Structure analysis.

G-PROTEINS

Protein Name: [Ras GTPase-activating protein 1](#)

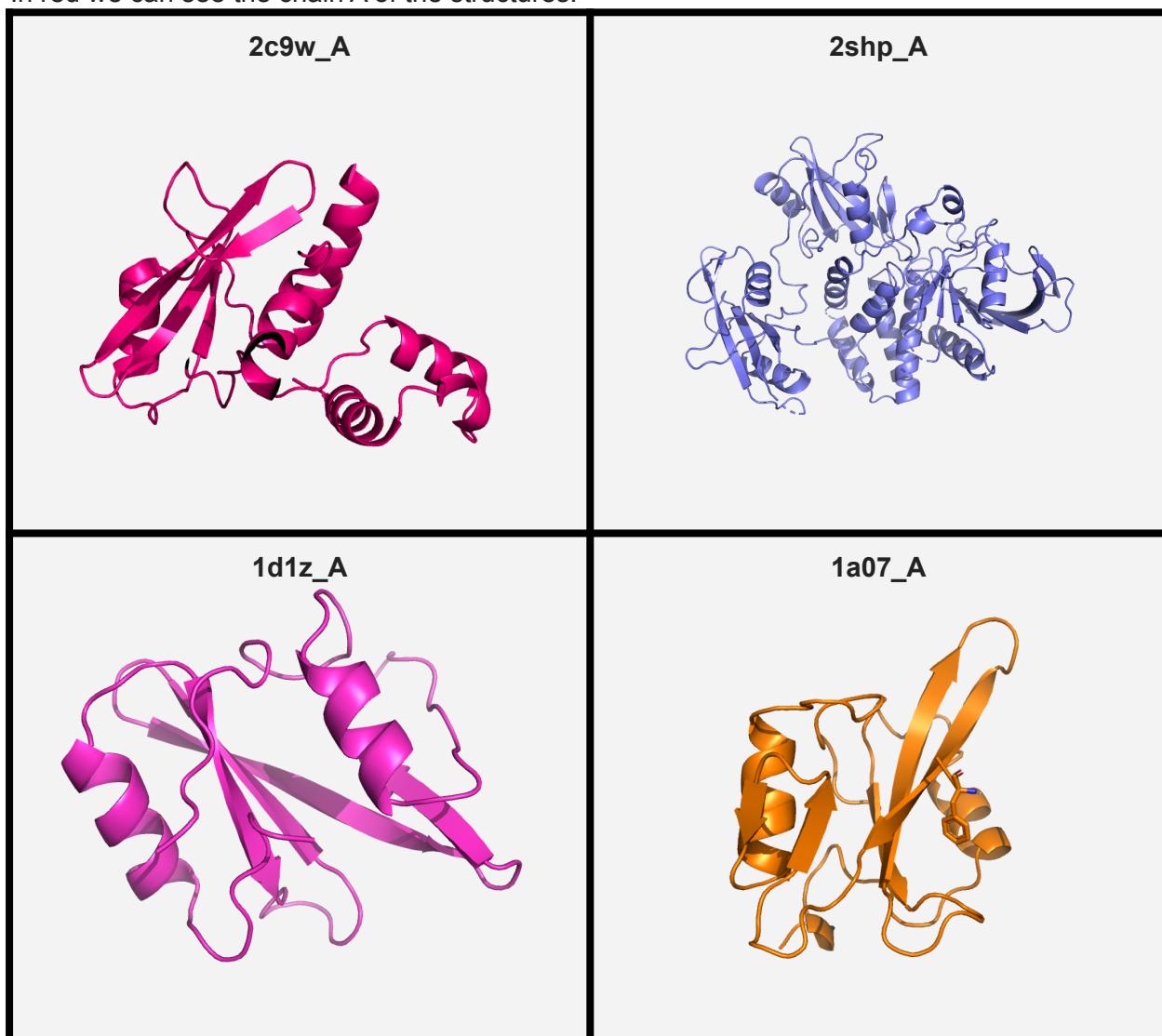
RASA1 gene provides instructions for making a protein called p120-RasGAP which helps in regulation of RAS/MAPK signaling pathway. This protein is a negative regulator of this pathway. This pathway transmits signals from outside the cell to the nucleus. It helps to regulate cellular functions such as growth, division of cells and cell movement. As it is a negative regulator so dysfunctioning can cause excessive cell proliferation which can give rise to cancer because it would not be able to turn off the signals when they are not needed allowing RAS/MAPK signaling pathway to proceed in an uncontrolled way.

1. 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?

To get the structures of the PDB we used a template. This template is the same we worked on the last assignment **6pxc**.

The obtained structures are:

In red we can see the chain A of the structures.



How do we obtained this structures:

There are two ways to obtain the structures, the one we used is using phmmer or jackhmmer. Using them we can get the structures from the PDB database. The other one can be using BLAST.

To avoid pairs that are **identical** we will build a hmm not based on our sequence but using the sequence from the domain family (SH2). And to avoid **bias** we will build hmm using sequences from UniProt by jackhmmer and then using a modified hmm from 5th iteration to search in the PDB database using hmmssearch.

```
jackhmmer --chkhmm sh2_markov target.fa  
/shared/databases/blastdat/uniprot_sprot > target_jack.out
```

```
hmmsearch sh2_markov_last.hmm /shared/databases/blastdat/pdb_seq >
          sh2_pdb_by_hmm.out
```

2. Analyze the secondary structure of these structures. What program would you use for that? Represent the secondary structure of each structure as a sequence of characters, where each character corresponds with the secondary structure of one amino acid. Then, compare the secondary structures of the structures you selected. Are they similar?

We can get the secondary structures of these models using dssp. Then we can concatenate the dssp secondary structures and run clustal alignment to see variability and similarity. (we are not showing shp2 in table below because its size was much larger than others)

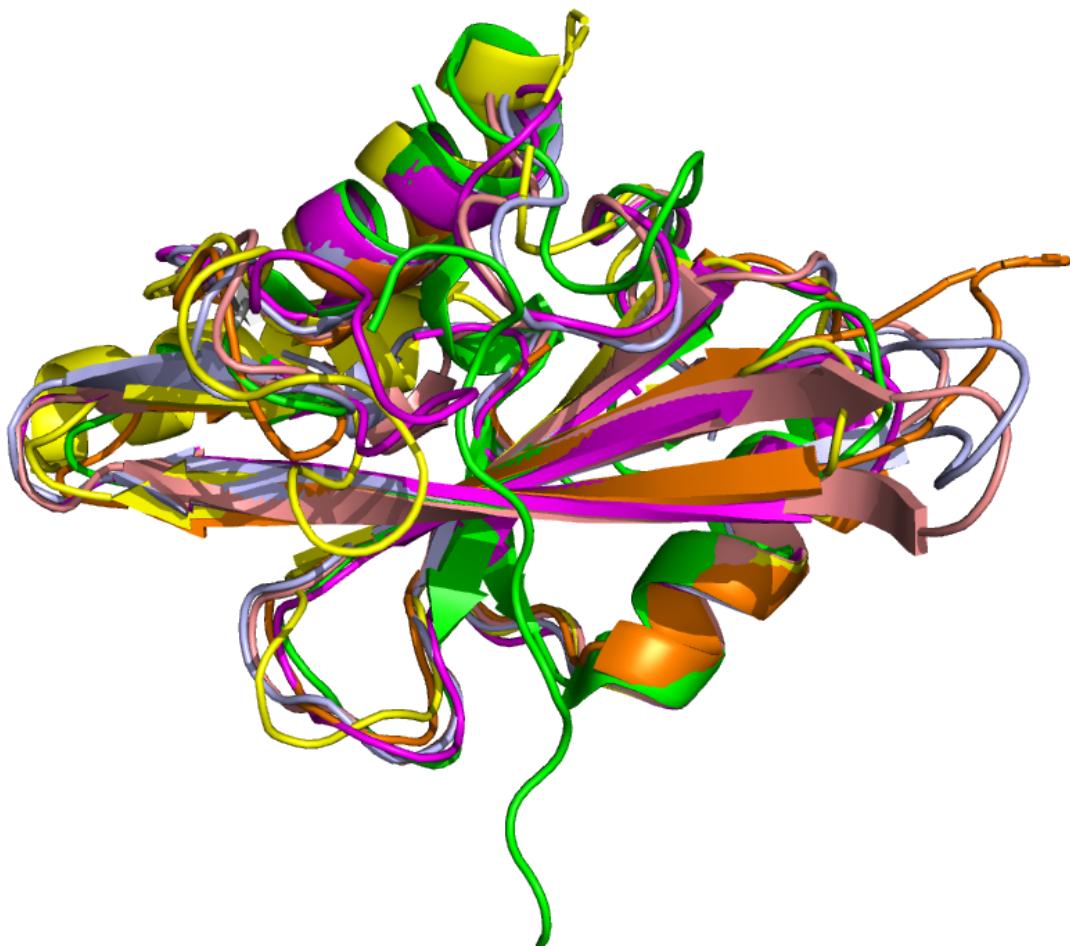
Here we can see a conserved block of residues. If we want to analyze blocks which are not conserved we can use psipred to predict secondary structures and compare them with the above obtained structures by aligning both.

3. 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? Include pymol images to support your explanation.

To superimpose the structures we are going to use super in pymol with a cutoff of 5.0.

Once we have superimposed the different sequences we have obtained that some positions, especially loops are no longer conserved.

The obtained structure after superimposition is:



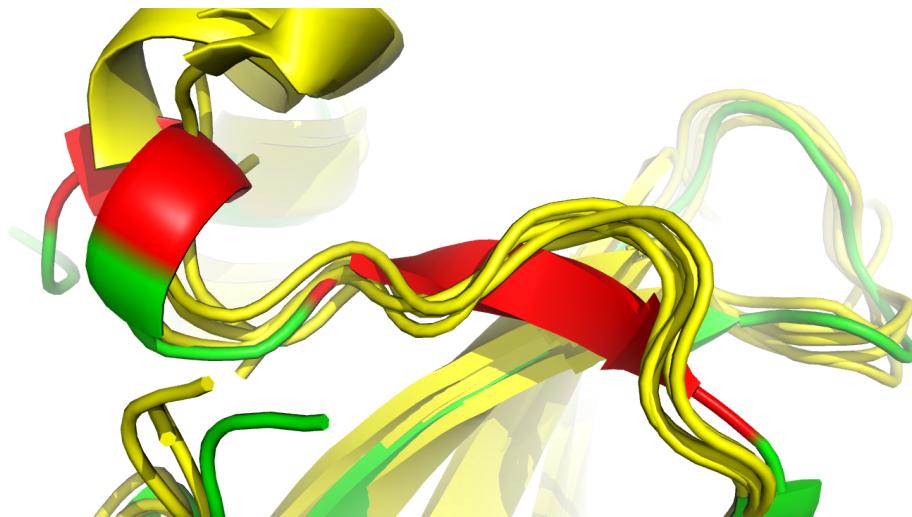
Along the superimpositions we have found different RMSD:

SEQ 1	SEQ 2 (template)	RMSD
2c9wA	6pxc_RASA1_sh2	2.083
2shpA	6pxc_RASA1_sh2	1.670
1d1zA	6pxc_RASA1_sh2	2.059
1a07A	6pxc_RASA1_sh2	1.675
1o41A	6pxc_RASA1_sh2	2.195

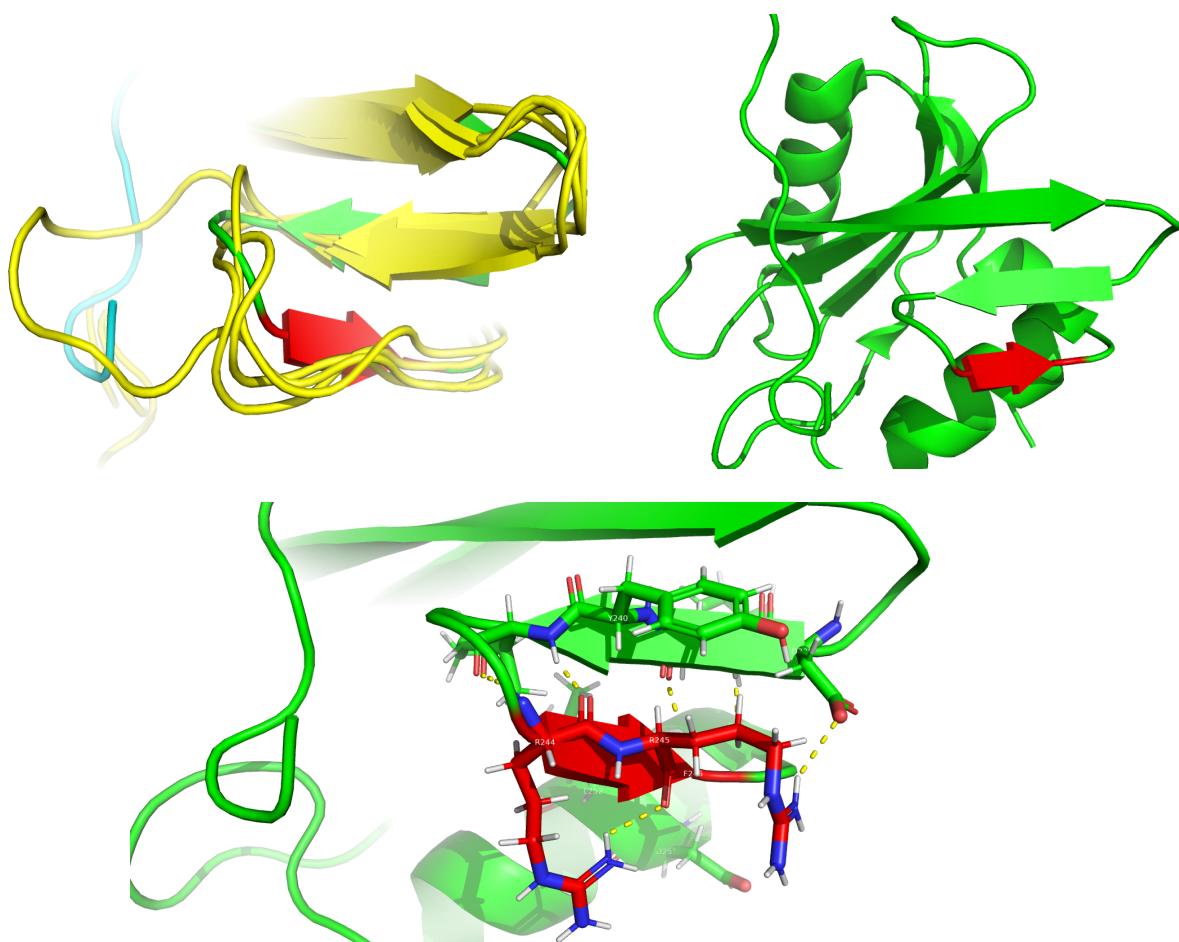
We can see higher variability in loop regions as loops are more flexible. There is also higher variability at starts and ends of helices and beta sheets. These variabilities also depend on how the structure was obtained in the lab and its quality. The regions more important to keep the structure similar are the ones that are conserved.

Regions of conflict:

We can see that in the region 182-186 of Rasa1_human the Sh2 domain is the only that has a small beta sheet and a small helix, while all other structures have loops. **To use the Sh2 domain of Rasa1 for modelling mutations we modified this region to convert it to a loop.**



Another region is in position 244-245-246 where the sh2 domain of Rasa1(green) has a small beta sheet connected to a nearby beta sheet , while other sequences have a loop.



4. Work with the conserved regions of your protein (the ones you described in assignment 1, question 6 and assignment 2, question 4). How do these regions look in the superimposition? are they variable or not? From these conserved regions, choose the one that you think is 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:

- If your protein is an enzyme, you should describe its active site. How is this active site interacting with its substrates? What contacts are made between substrate and enzyme? What amino acids are essential in this active site? How these amino acids contribute to catalyze a chemical reaction?
 - If your protein needs to interact with another protein to carry out its function, describe the interaction between the two proteins. How are the two proteins interacting? What are the interactions that make the two proteins have chemical affinity for each other? Can you find any amino acid that is essential for that interaction to happen?

After superimposition some regions are not conserved anymore.

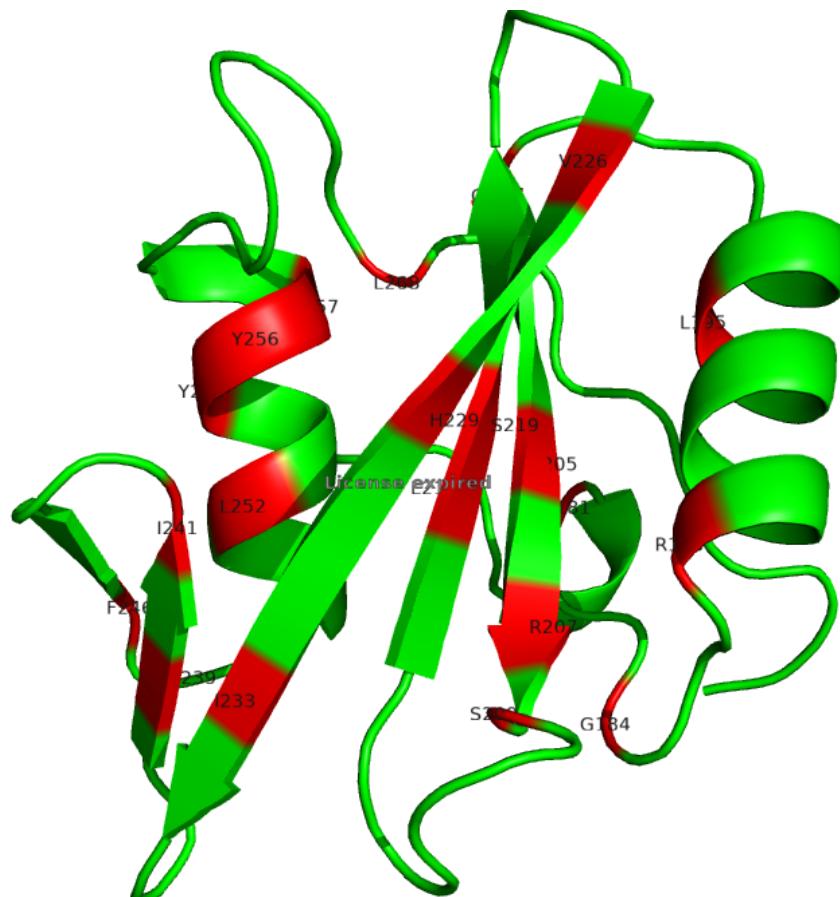
In the next image we can observe the differences in the sequence obtained in the first assignment and the differences after superimposition from this assignment.

In the top we can see the alignment obtained in the first assignment and in the other figure the alignment obtained after superimposition.

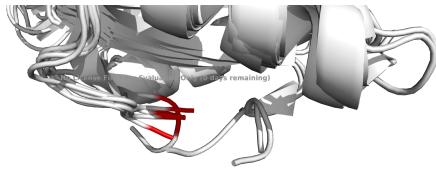
	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211					
6pxc_RASA1	w	Y	h	G	K	I	d	r	t	i	a	e	e	r	L	r	q	-	g	g	K	S	G	s	y	L	i	R	-	-	-	e	S	d	r	
1o41A	w	y	f	G	k	i	t	r	r	r	e	s	e	r	I	L	I	n	a	n	n	p	r	G	t	f	L	v	R	-	-	-	e	S	e	t
1a07A	w	y	f	G	k	i	t	r	r	r	e	s	e	r	I	L	I	-	-	n	n	p	r	G	t	f	L	v	R	-	-	-	e	S	e	t
2c9wA	w	y	w	G	s	m	t	v	n	e	a	k	e	k	L	-	k	-	a	a	p	e	G	t	f	L	i	R	-	-	-	s	S	-	h	
1d1zA	v	y	h	G	k	i	s	r	e	t	g	e	k	I	L	I	a	-	g	g	l	d	G	s	y	L	I	R	-	-	-	d	S	e	s	
2shpA	w	f	h	G	h	l	s	g	k	e	a	e	k	I	L	t	e	-	g	g	k	h	G	s	f	L	v	R	-	-	-	e	S	q	s	
*	:	:	:	*	:	:	*	:	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	:	*	*			
	212	213	214	215	216	217	218	219	220	221	222	223	224						225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240		
6pxc_RASA1	r	p	g	s	f	v	l	s	f	l	s	q	m	-	-	-	-	n	v	v	n	h	f	r	I	I	l	a	m	s	g	d	-	-	y	y
1o41A	t	k	g	a	y	c	I	s	f	d	n	a	k	-	-	-	-	g	v	n	k	h	y	k	i	r	k	l	d	s	g	-	-	-	g	f
1a07A	t	k	g	a	-	y	I	s	v	s	d	f	t	-	-	-	-	n	n	v	k	h	y	k	i	r	k	l	s	g	g	-	-	-	f	y
2c9wA	s	d	y	l	I	t	I	s	v	k	t	s	-	-	-	-	-	g	p	t	n	l	r	i	e	y	q	d	-	-	-	-	-	-	-	
1d1zA	v	p	g	v	-	y	c	I	c	v	l	y	h	-	-	-	-	g	y	i	y	t	y	r	v	s	q	t	e	t	g	-	-	s	w	
2shpA	h	p	g	d	-	f	I	s	v	r	t	g	-	-	-	-	-	k	v	t	h	m	m	i	r	v	q	e	l	k	-	-	-	y	d	
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	:	:	:	:	:	
	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272				
6pxc_RASA1	I	-	g	-	r	r	F	s	s	l	s	d	L	g	y	y	s	h	v	s	s	l	l	k	g	e	k	I	l	y	p	-	-	-		
1o41A	y	-	i	r	t	q	F	n	s	l	q	q	L	v	a	y	y	s	-	k	h	a	l	c	-	h	r	I	t	t	v	c	-	-	-	
1a07A	i	-	t	-	t	q	F	n	s	l	q	q	L	v	a	y	y	s	-	k	h	a	l	c	-	h	r	I	t	t	v	c	-	-	-	
2c9wA	-	-	-	-	-	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
1d1zA	s	-	a	-	r	y	F	r	k	i	k	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
2shpA	v	g	g	-	e	r	f	d	s	l	d	-	L	v	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
*	:	:	:	*	*	*	*	*	:	:	:	:	*	*	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

We have seen the sequence differences between the two alignments. Now we are going to look for the structural differences along the conserved regions, some of which are not conserved anymore. because the sequences we selected now are maybe more distant then the previous ones.

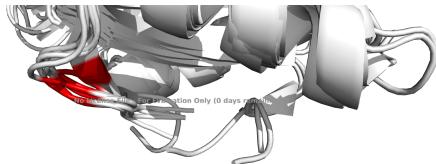
First we mark the conserved regions of the model sequence:



Pos 181: all loops



Pos 184: loop and sheet



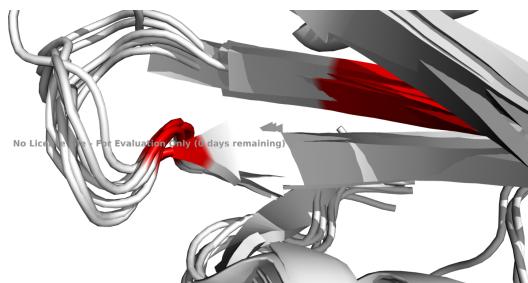
Pos 195: all helix



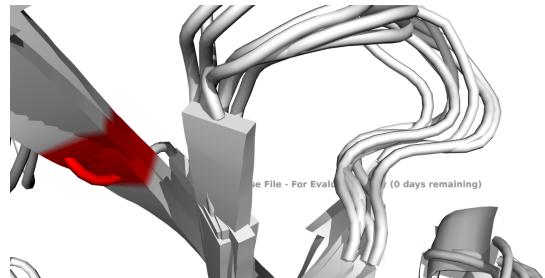
Pos 202, 205, 207: loop and sheet in one position, sheet in the rest



Pos 209, 218, 219: in position 209 we see loops and sheets, in position 218, 219 all sheets



Pos 232: sheets and loop



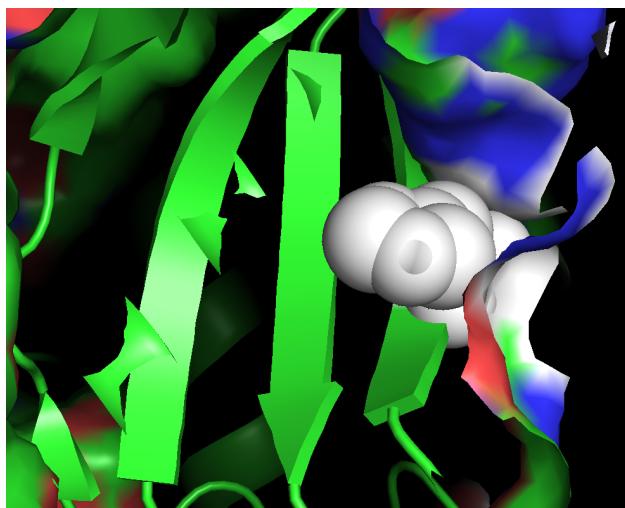
Pos 246, 252: position 246 loops and a sheet, position 252 all helices



Pos 255,256: All helices



We will focus on the position 229 because it is one of the ones involved in the active site.

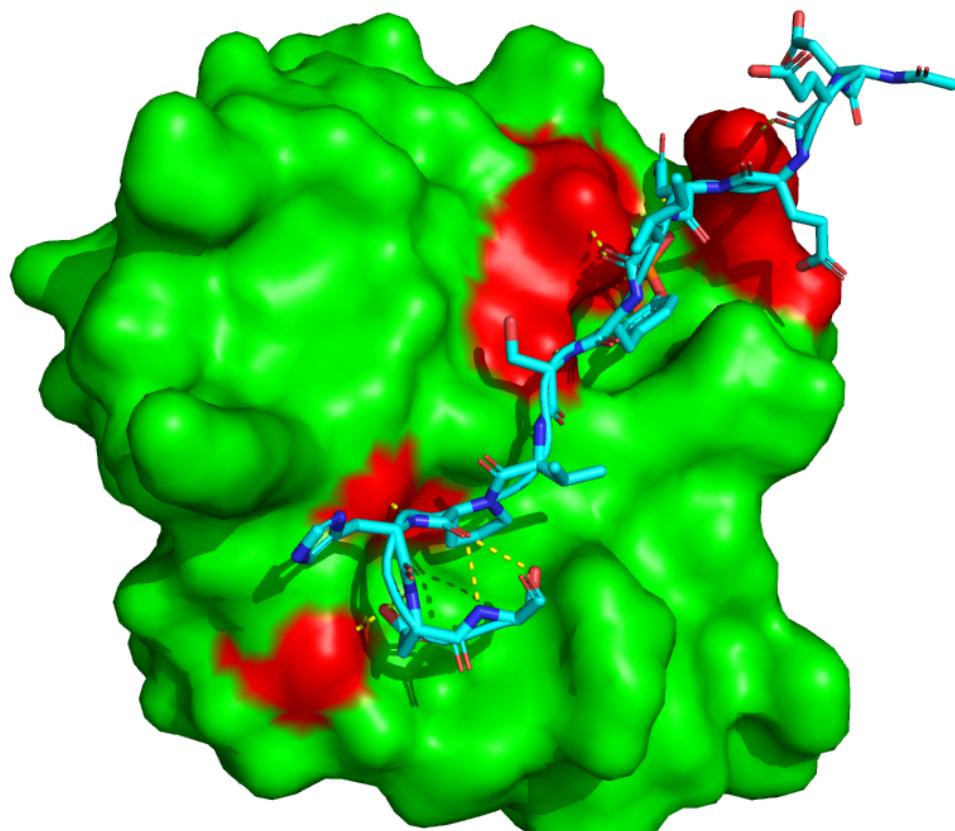


As we can see at this position, in the first alignment this region is conserved, but in the second alignment it is not conserved for 2c9wA and 1d1zA, where we got an Asparagine and a Threonine instead of a Histidine.

Position 229 is in the active site, that is present on a beta sheet, when the ligand enters it binds to the G-protein coupled receptor. When this happens there is a change in conformation of the GPCR and also in the G-protein that it hydrolyzes and changes the GDP for a GTP, now the G-protein is activated.

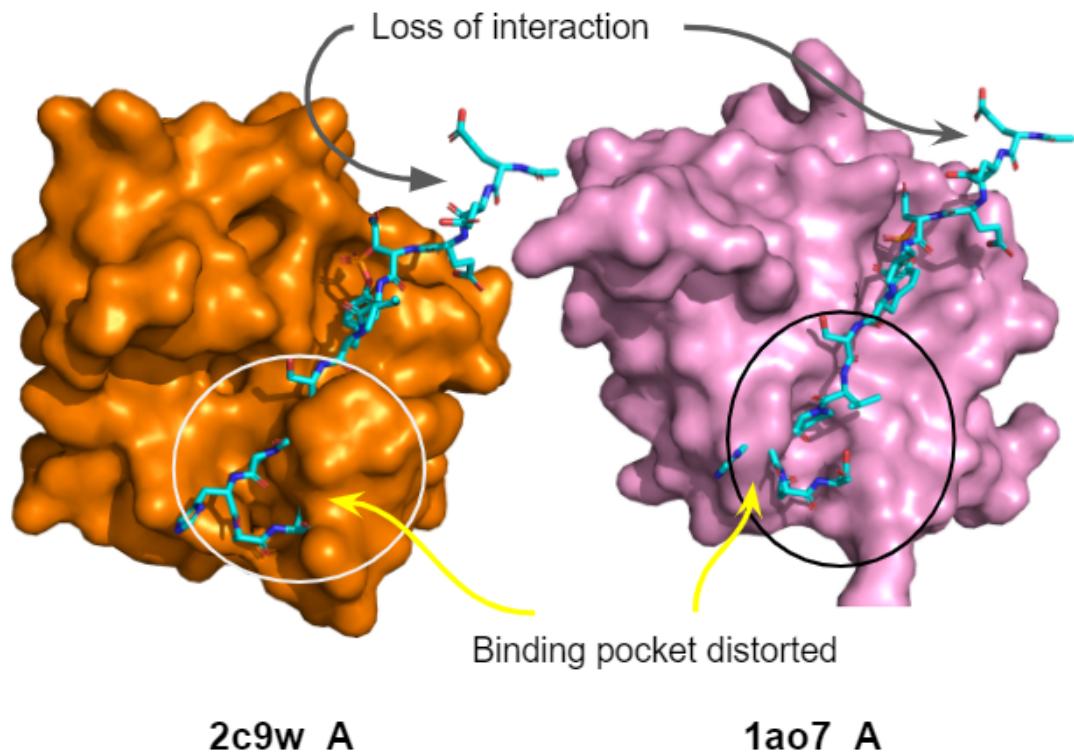
Alpha is dissociated from the beta and gamma subunit, now the alpha subunit is able to regulate an enzyme which can act like a messenger and can give a biological response.

Moreover we can see that in some of these sequences the binding site of ligand is affected as shown below:



Regions important for interaction are shown in red

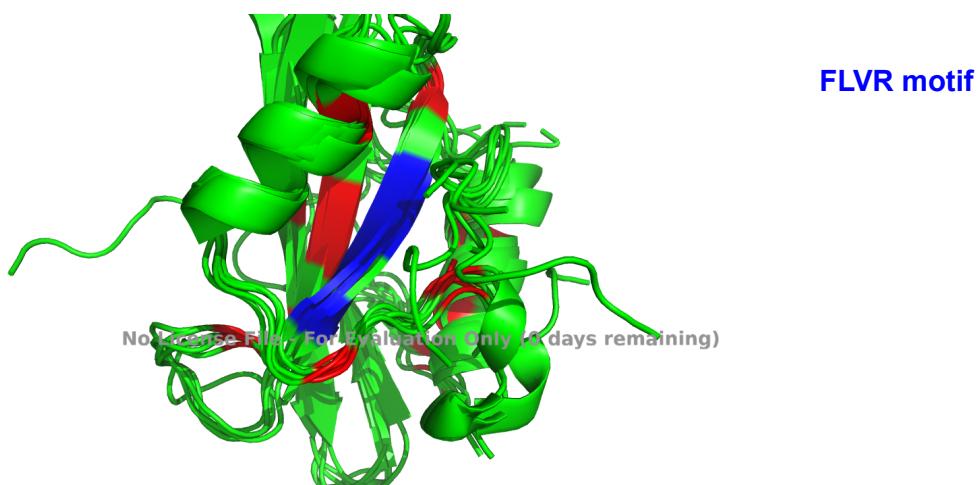
Template 6pxc



Focusing on the active site:

The SH2 domain comprises a central antiparallel seven-stranded β -sheet located between two alpha-helices. This creates two binding sites; a deep pocket and a shallow cleft. The deep pocket binds the pTyr residue, which is captured by electrostatic interactions with an arginine residue invariant in all pTyr-binding SH2 domains. In contrast, the shallow cleft binds residues C-terminal to the pTyr and varies considerably between SH2 domains allowing specificity determination.

In addition, the deep pocket arginine residue is part of the conserved FLVR motif (Phe-Leu-Val-Arg207) which is required for SH2-pTyr binding as it has direct interactions with pTyr. Consequently, mutations of FLVR arginine are invariably used to generate a “dead” SH2 domain. Then, we can consider this motif as an essential region for the protein function and Arg207 as an essential amino acid in the active site.



5. 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.

Green → wild type

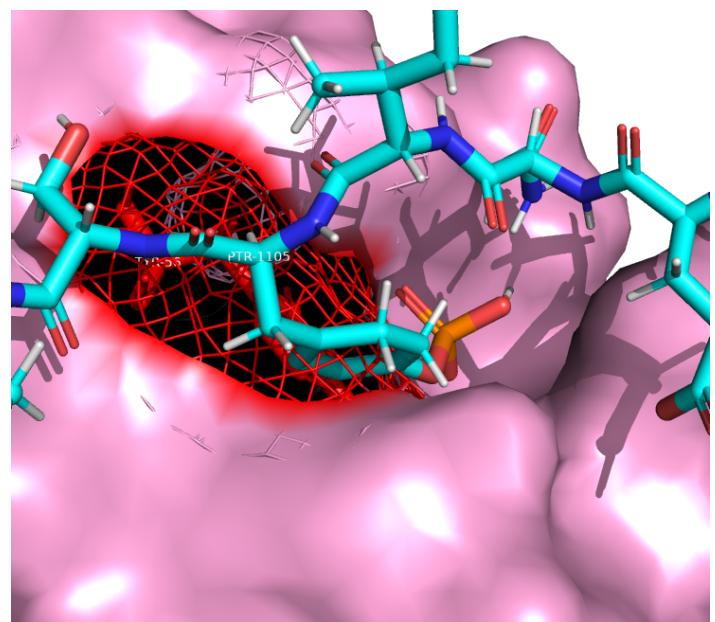
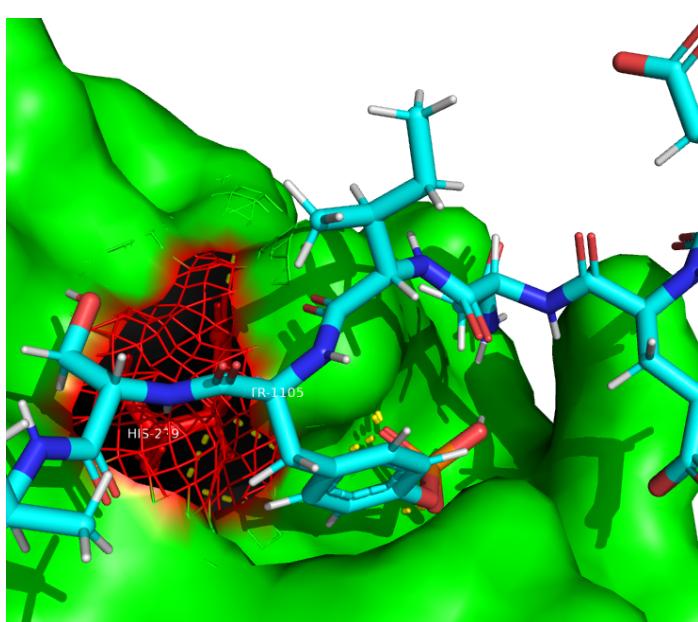
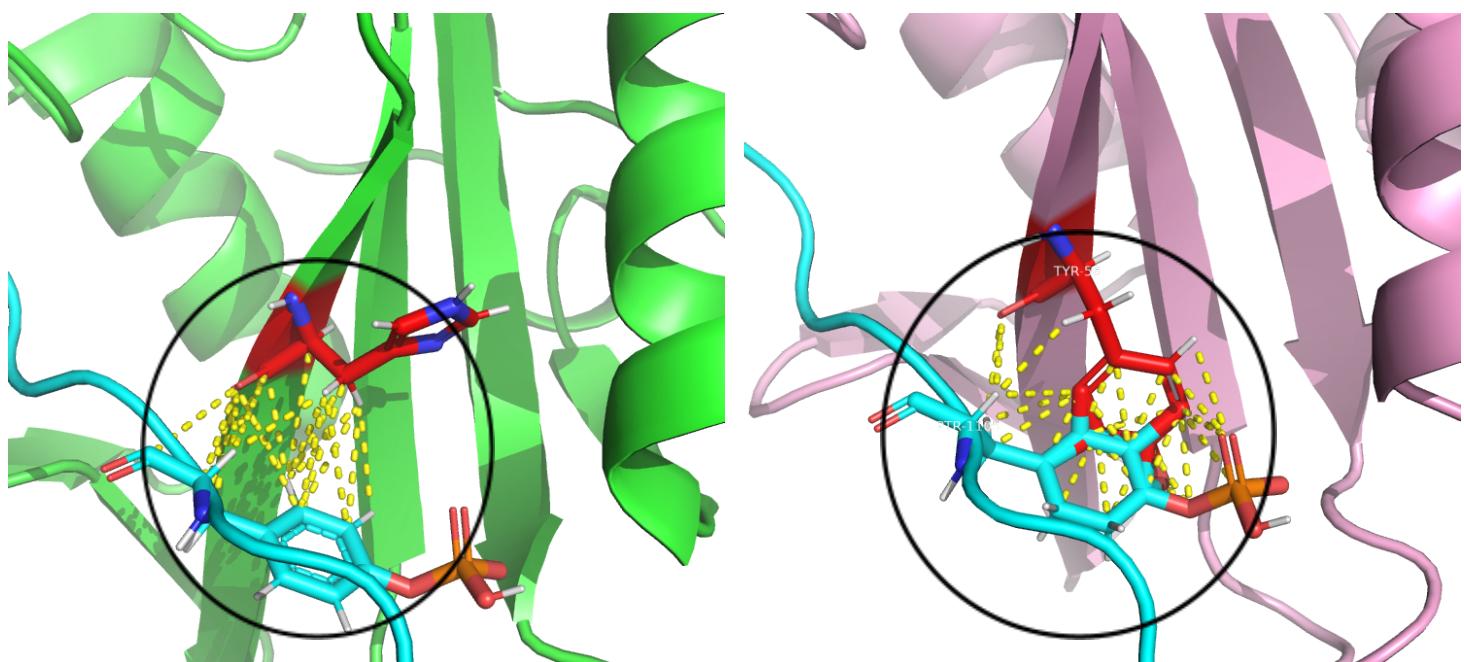
Pink → mutated

Red → residue under study

Cyan → ligand

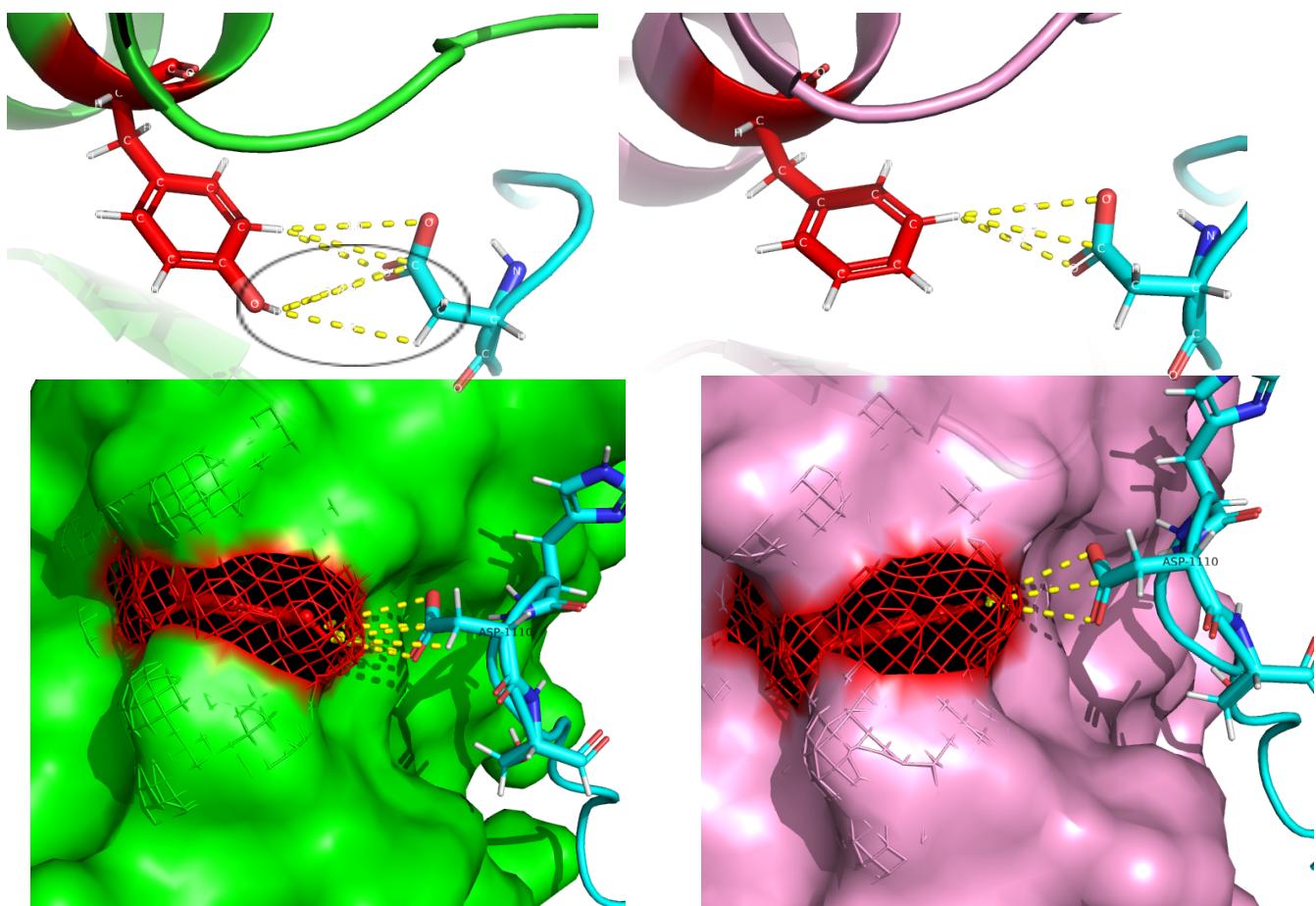
Position 229 Histidine (+) → Tyrosine (neutral polar)

Histidine at position 229 of SH-2 domain is present at the binding site of the ligand. It interacts with PTR-1105 of p190RhoGAP phosphotyrosine peptide but when it is mutated to Tyrosine it distorts the binding pocket and PTR-1105 cannot fit there maybe it would not bind because of this mutation,



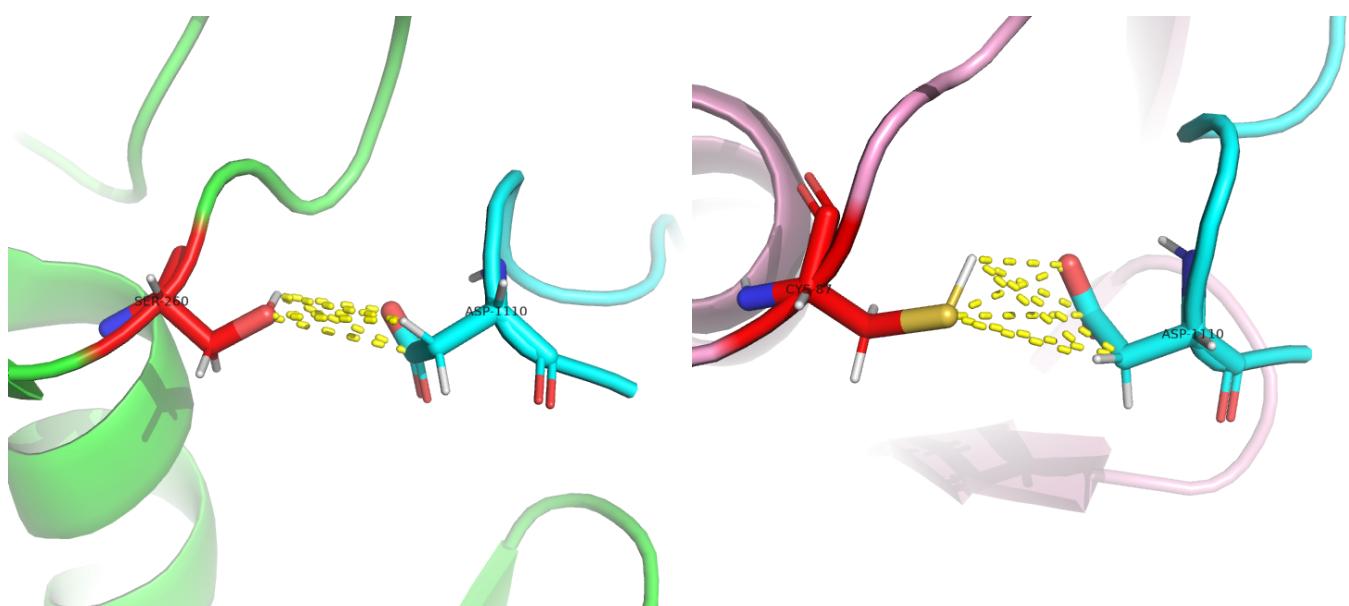
Position 255 Tyrosine (Polar) → Phenylalanine (non polar):

Tyr-255 is also present at the binding region of ligand and when it is mutated to phenylalanine some of the interactions with ligand are lost.



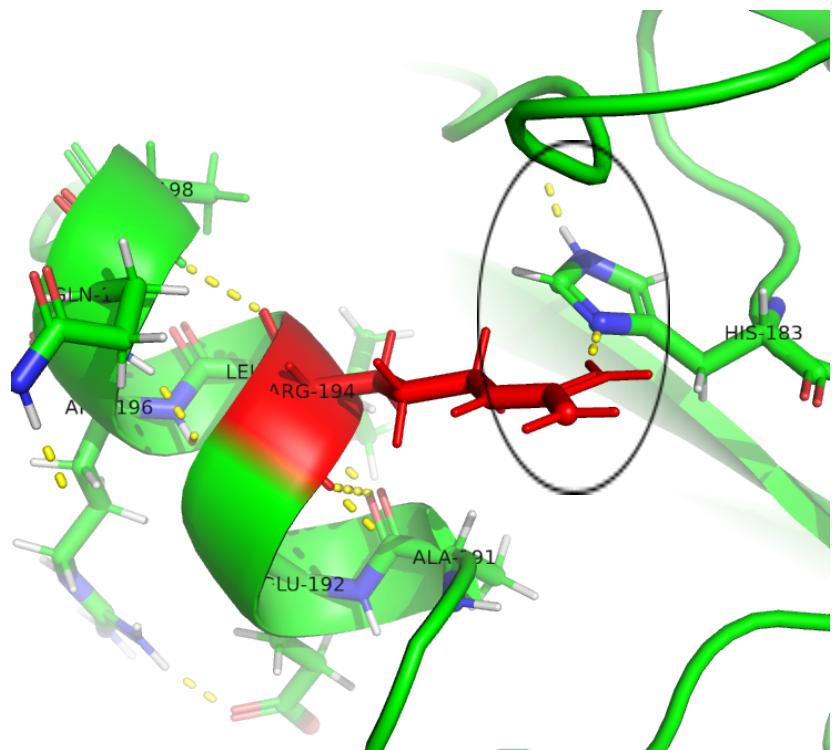
Position 260: (S → C) :

This mutation does not seem to be producing much difference in interaction. But if we see z-scores given at the end then this mutation seems to be decreasing the quality more than other mutations

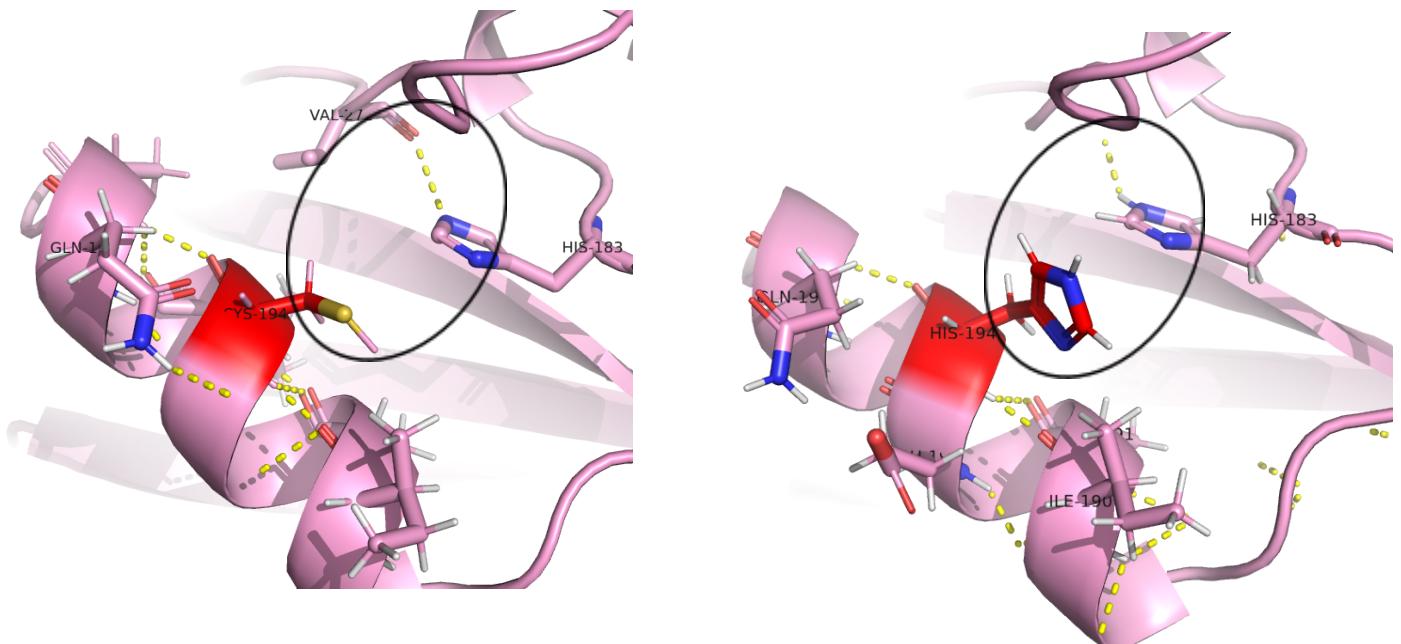


Position 194:

arginine at pos 194 when mutated to Histidine or cystein causes loss of interaction with loops outside the helix because of their small size. This may cause a small relaxation of the structure which can change the binding pockets structure a bit.

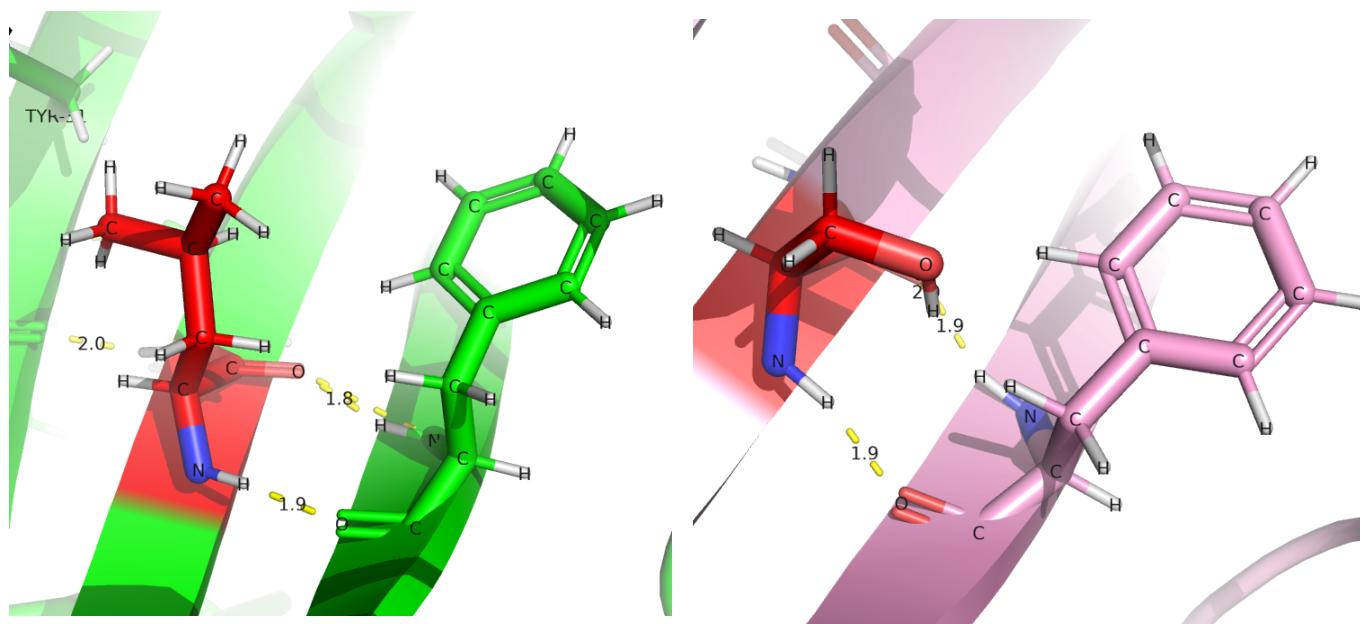


Arg194 when mutated to Cysteine (left fig) or Histidine (right fig) causes loss of interactions with nearby loops. It is not directly involved in binding site but may have an effect on structure.

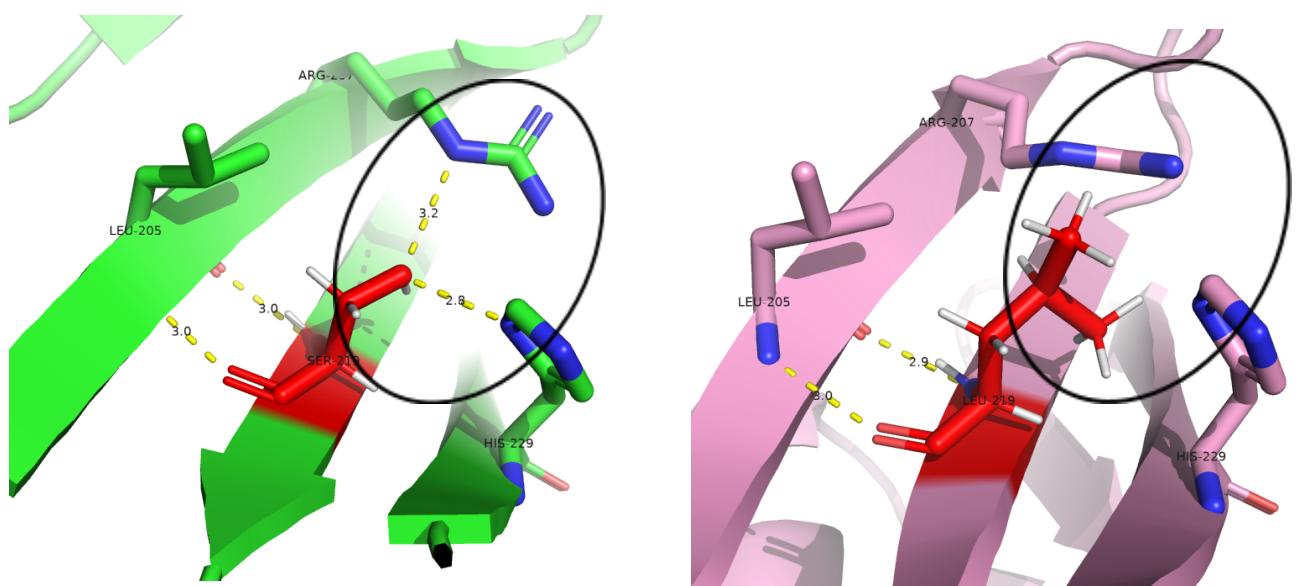


Position 218: 45 (L → S)

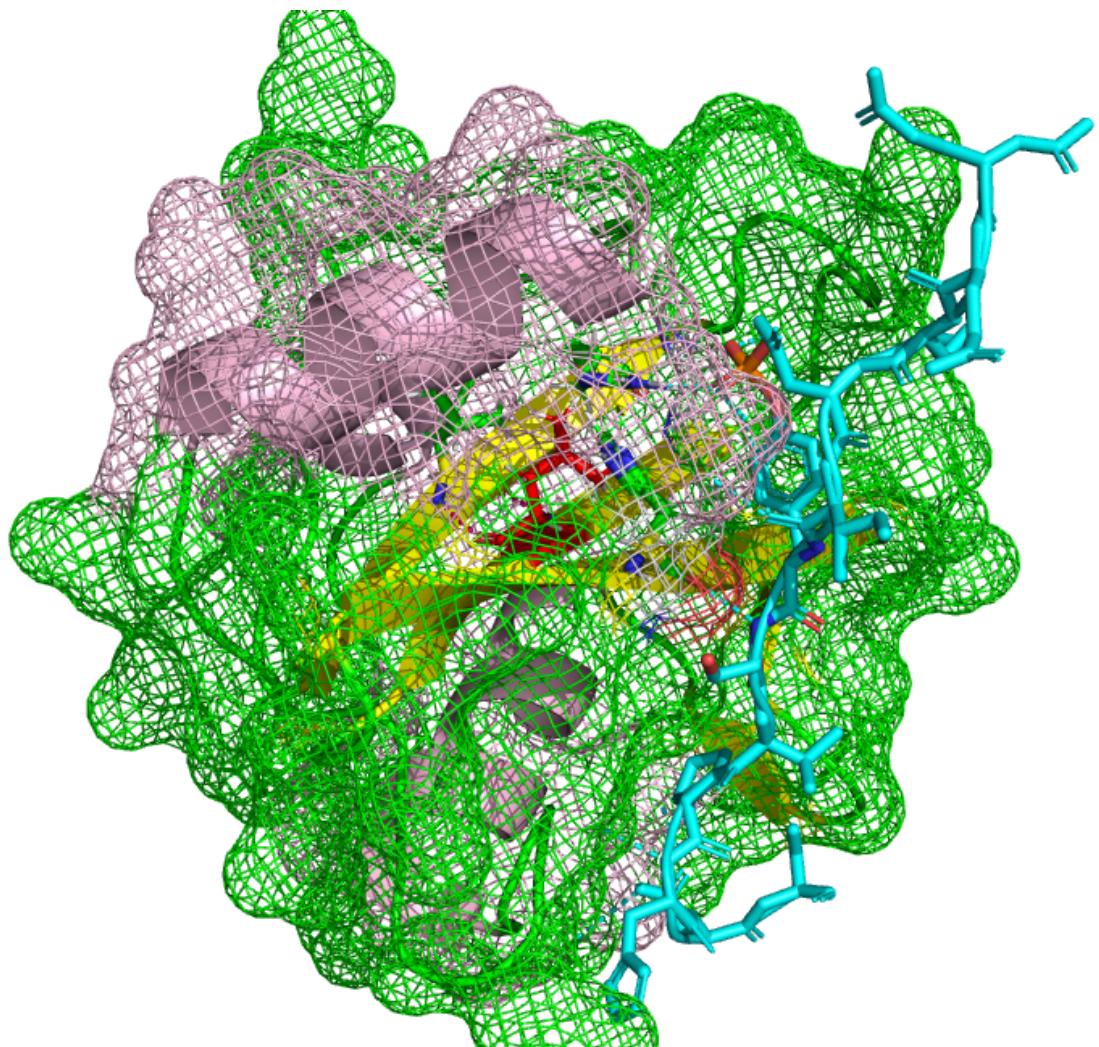
At this position the loss of interactions seems to be very less

**Position 219: 46 (S → L)**

At this position we can see that because of mutation there is loss of interaction with the nearby beta strand residues which can cause relaxation of this betasheet. This relaxation may affect the shape of ligand binding pocket.



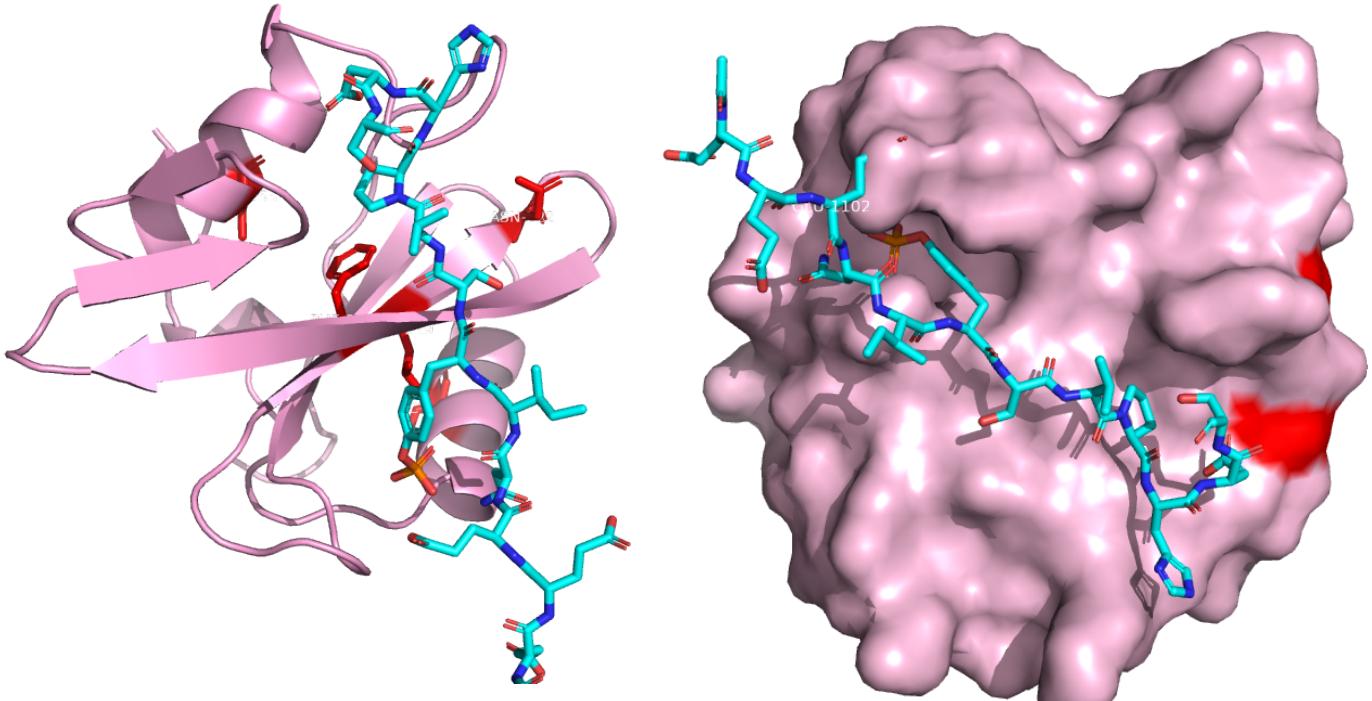
We can see in this figure that due to relaxation of yellow beta sheet in middle the binding pocket for the ligand can get distorted.



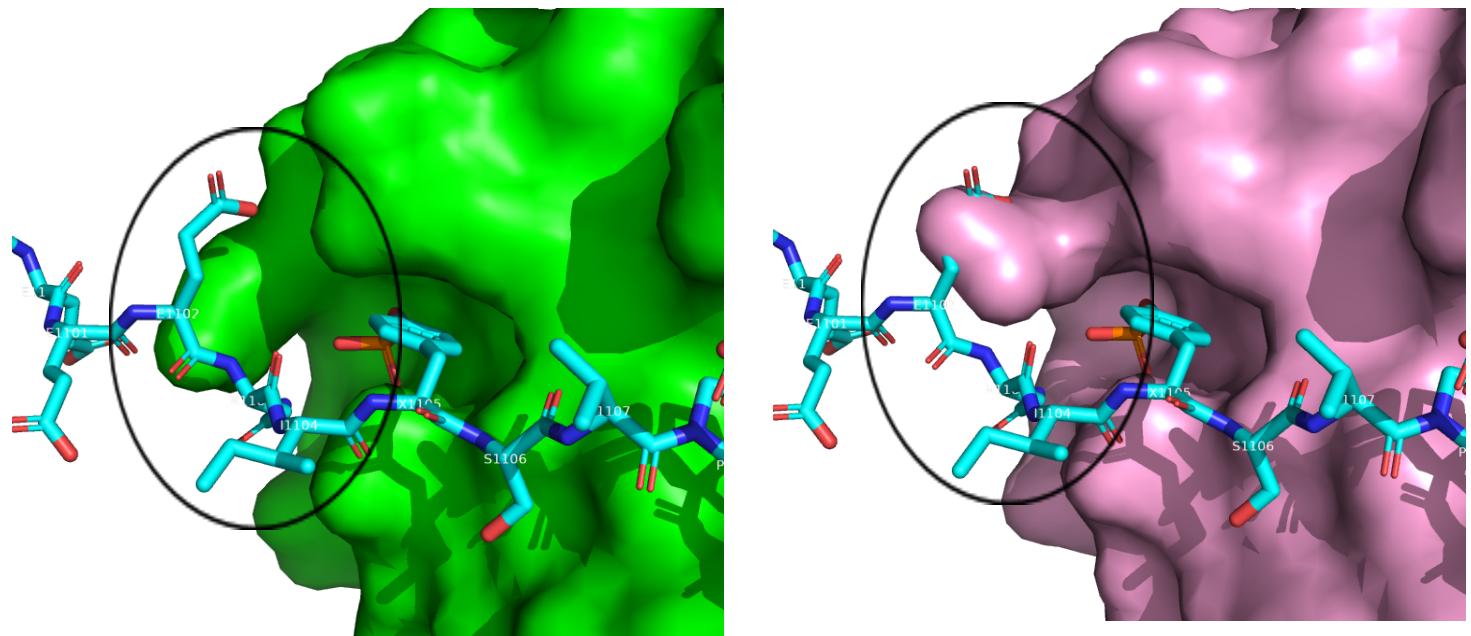
Mutation involved in melanoma

green → wild type

194(R → C) 218(L → F) 219(S → L) 222(S → N) 253(I → T) 267(K → N)
These mutations are shown in red.

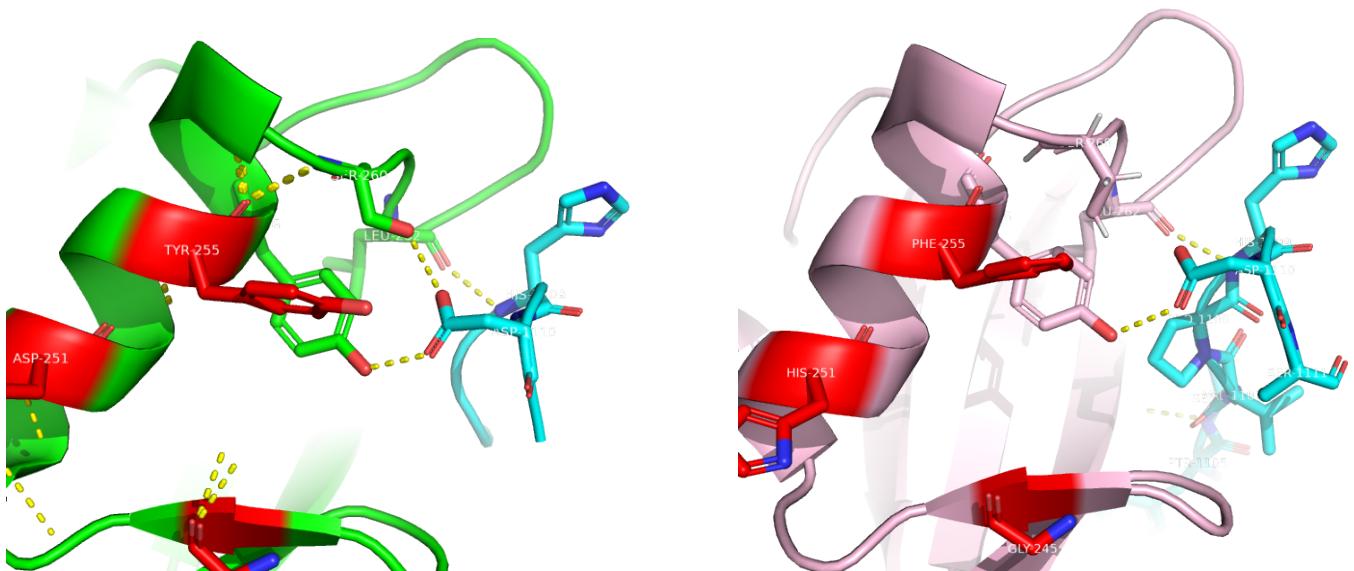


These mutations are not directly linked to binding site but are distorting the binding pocket for the ligand.

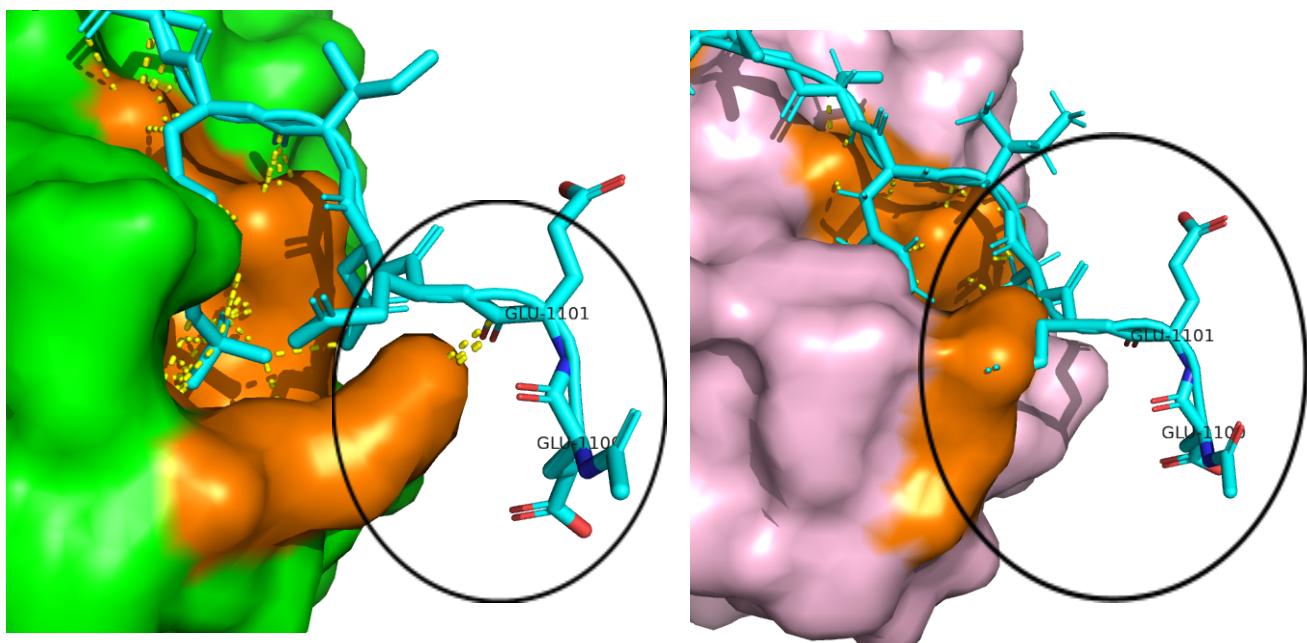


Mutations involved in breast cancer

245(R → G) 251(D → H) 255(Y → F)



Because of these mutations we lose only one bond at one site interacting with the ligand as shown above but on the other end it affects more the binding of Glu 1101 and 1100 of p120-rho-gap to the interface.



For the mutations causing the above two diseases we cannot see much change in energy or z-score(quality). But z-scores increased a bit indicating better stability of our template .

	template	melanoma	breast cancer
Z-scores	-6.31	-6.16	-6.02

6. Use ProSa 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? Would it make sense to analyze with ProSa only the wild type structure?

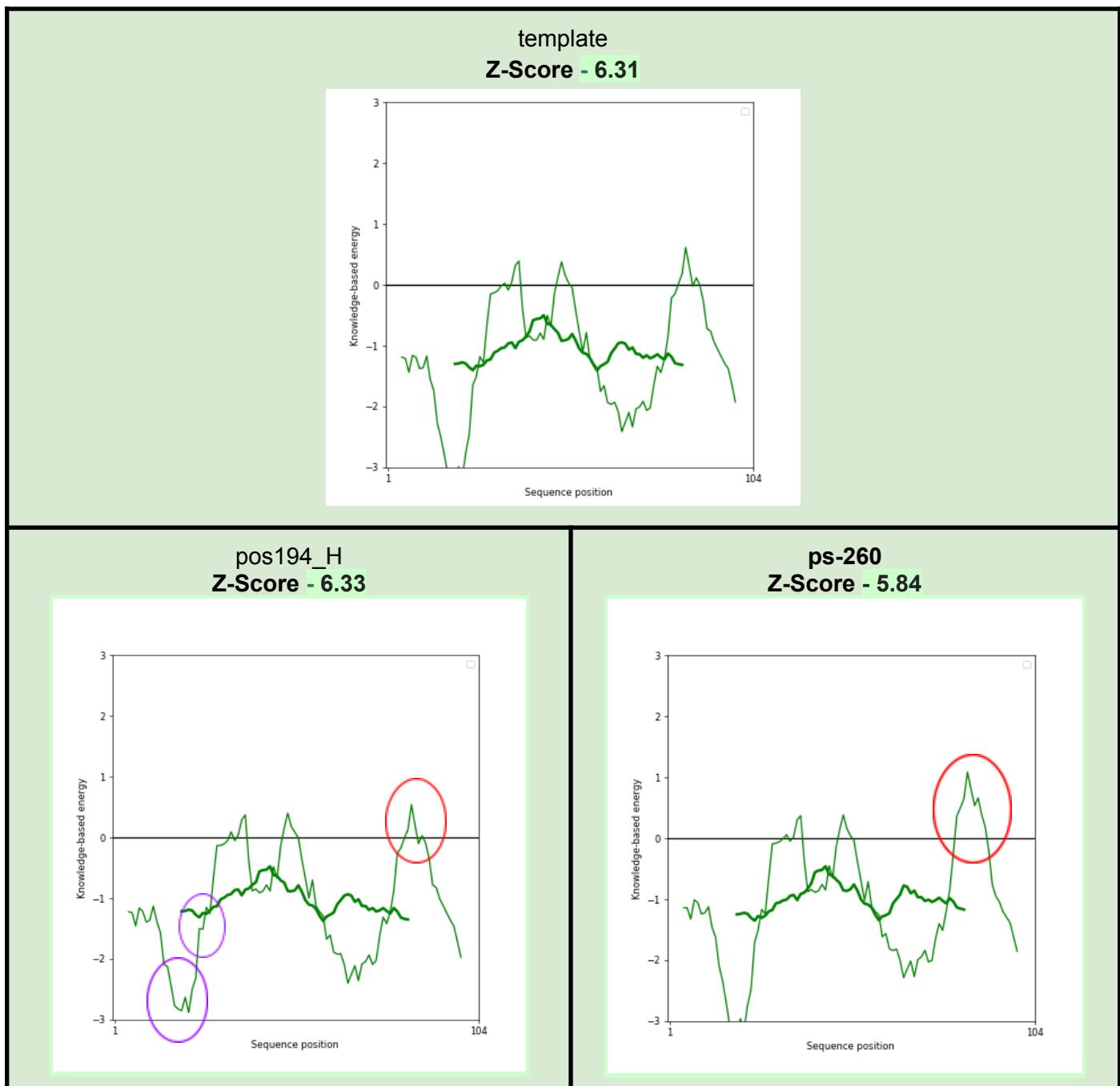
Prosa energy profiles are used to compare mutant or different models with wild type structure. Just checking wild type energetic profiles can give an idea of regions which are wrongly structured having high energies and then we may use a different template for that region to model our targets. Once we modeled our targets we can evaluate their correctness using Prosa.

Z-score tells overall model quality. Lower the z-score better is the quality, we can see that only one mutation (p194_H) decreases it a bit while all others are slightly increasing the z-score.

If we only analyze the wild type it would not make sense as we won't have any model which to compare it with.

Model	Z-Score
template	-6.31
pos260	-5.84
pos255	-6.23
pos 229	-6.23
pos 219	-6.27
pos218	-6.19
pos194_H	-6.33
pos194_C	-6.2

We have taken the energy profiles of the mutants with the highest and lowest Z-scores to compare with the template:



Generally, the three energy profiles are very similar, but we can see little differences at the beginning and the end of pos194_H mutant and at the end of pos260 mutant. In the first case, at the beginning the energy is higher but the curve is wider from the top, so it compensates a little. And at the end, the last peak is a little lower than in the template. In the second case the last peak the energy is much higher (so then, less stable) than in the template and this makes the Z-score increase.