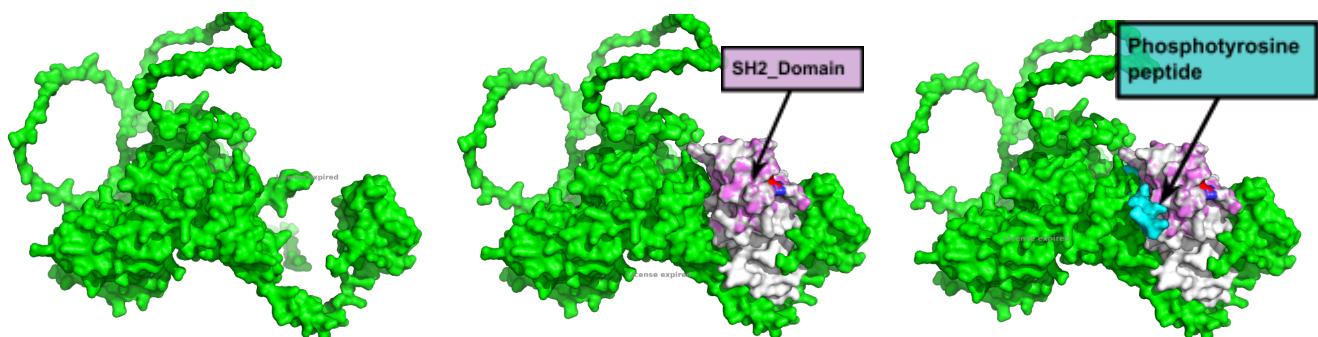
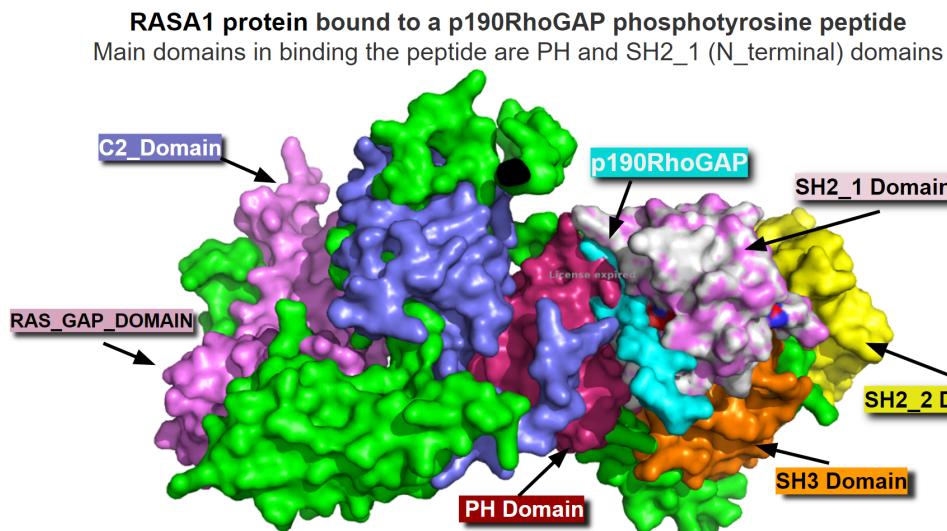


Structural bioinformatics Project - Assignment 2: Sequence analysis.

G-PROTEINS

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

Domain under study: SH2-domain



This binding takes place for the activation of rho ras pathways.

The Ras pathway is involved in cell proliferation, differentiation, migration and apoptosis

The Rho GTPases are essential in cell adhesion, protrusion, polarity, migration and cell motility.

Therefore dysregulation in this pathway is mainly involved in different cancer types.

1. Does your protein have an HMM available in the PFAM database?

To find if there are HMM available in the PFAM database for our protein we will use hmmscan.

```
hmmscan/shared/databases/pfam-3/Pfam-A.hmm P20936.fa > P20936_db.out
```

If we inspect the output :

Query:	sp P20936 RASA1_HUMAN [L=1047]								
Description:	Ras GTPase-activating protein 1 OS=Homo sapiens OX=9606 GN=RASA1 PE=1 SV=1								
Scores for complete sequence (score includes all domains):									
--- full sequence --- --- best 1 domain --- -#dom-									
E-value	score	bias	E-value	score	bias	exp	N	Model	Description
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1.6e-43	146.2	0.1	7.3e-21	73.6	0.0	2.6	2	SH2	SH2 domain
7.4e-40	136.6	1.9	1.4e-39	135.7	1.9	1.5	1	RasGAP	GTPase-activator protein for Ras-like GTPase
3.5e-15	55.9	0.1	1.3e-14	54.1	0.0	2.1	2	PH	PH domain
3.7e-10	38.9	0.0	1.1e-09	37.4	0.0	1.9	1	SH3_1	SH3 domain
3.7e-09	36.1	0.1	7.8e-09	35.1	0.1	1.5	1	C2	C2 domain
0.0071	15.7	0.0	0.014	14.7	0.0	1.5	1	SH3_2	Variant SH3 domain
----- inclusion threshold -----									
0.034	13.5	0.1	6	6.3	0.0	2.5	2	zf-U1	U1 zinc finger

According to this search our protein has 6 PFAM hits, top 5 corresponding to each domain shown in protein figure above. The 6th domain is a hit for variant of SH3 domain but it has a very high E-value and very low score indicating that its not a good hit.

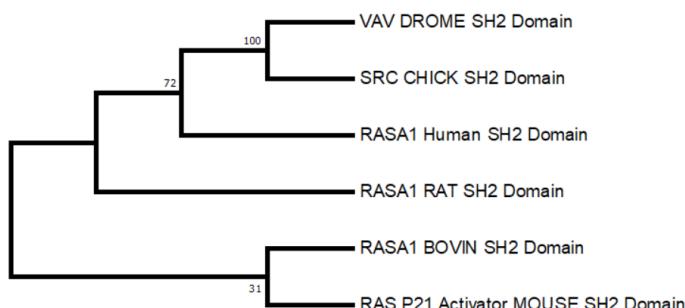
2. Choose a set of 6 to 8 amino acid sequences that belong to the protein family you are studying. These sequences should represent the evolutionary history of your protein family, so you want them to have some diversity between them and avoid redundant or highly similar pairs of sequences. You will use these sequences to build a multiple sequence alignment. From what database should you retrieve these sequences? Why?

We have chosen to work with the SH2 domain of same or different proteins from different organisms, then we used the Uniprot database to obtain the sequences because uniprot contains way more sequences than any other database like PDB which have only sequences that have structures. We extracted these structures by fetching SH2 domain using hmmfetch command and then hmmsearch in uniprot database.

HUMAN - P20936	BOVIN - P09851
MOUSE - E9PYG6	RAT - P50904
FRUIT FLY - Q9NHV9	CHICKEN - P00523

We have performed the alignment and then created a **phylogeny trees** of our sequences using [MEGA7 software](#).

Minimum evolution tree

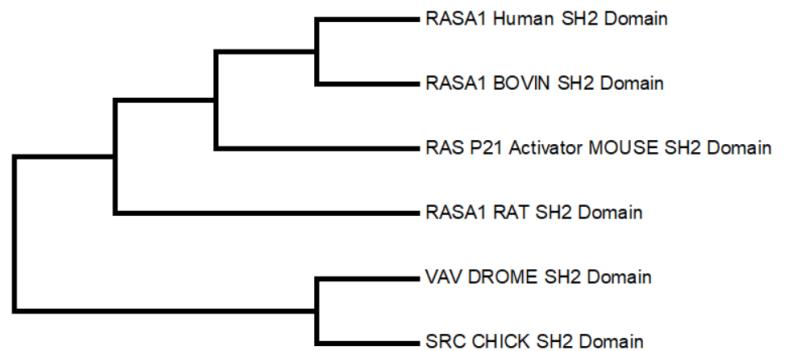


This method tries to **minimize branch lengths by minimizing the distance** so according to this tree

SH2 domain in RASA1 of human is more closer to SH2 domains of SRC and VAV protein as compared to other RAS proteins.

Maximum Parsimony tree :

This method tries to **minimize branch lengths by minimizing the number of mutations** so according to this tree SH2 domain in RAS proteins are more closely related as compared to SH2 domains from vav and SRC protein.



3. Make a sequence alignment with the sequences you just obtained in the previous step. To create this alignment, use the HMM you found in PFAM and the programs from the HMMer package.

To do the multiple alignment we will use the program `hmmaalign`

1. save the above sequences in fasta file → sequences.fa

- ## 2. Use hmmfetch to get sh2 domain HMM in pfam database

```
hmmfetch /shared/databases/pfam-3/Pfam-A.hmm SH2 > sh2.hmm
```

```
[u174453@ip-10-49-0-235 ~]$ vi sh2.hmm
 42  2.61350  4.29064  2.25598  2.40645  2.88382  3.77092  3.18022  3.4150
 2.65087  3.00156  3.86817  3.00227  4.16065  3.00942  2.55252  2.81063  2.613
 0  3.03834  3.95574  3.48767  56 d - E
 2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.2935
 2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.775
 9  2.98518  4.58477  3.61503
 0.00883  5.12972  5.85207  0.61958  0.77255  0.48576  0.95510
 43  2.93146  5.43447  2.09727  2.39945  3.49435  2.67509  3.89378  3.8944
 2.12743  3.73838  4.47598  2.21905  3.89699  2.51815  2.83898  2.62758  3.162
 1  3.82338  4.21641  4.18868  57 d - E
 2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.2935
 2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.775
 9  2.98518  4.58477  3.61503
 0.00883  5.12972  5.85207  0.61958  0.77255  0.48576  0.95510
 44  2.65970  5.42673  2.11142  2.47536  4.19620  1.93093  3.89587  4.2464
 2.37504  3.72917  4.09865  2.71889  4.10138  2.77174  2.90335  2.39525  3.093
 2  3.18690  5.86342  4.46468  58 g - T
 2.68675  4.42234  2.77477  2.73120  3.46445  2.40460  3.72557  3.2929
 2.67698  2.69400  4.24755  2.90349  2.73687  3.18022  2.89855  2.37872  2.775
 8  2.98579  4.58568  3.61574
 0.50776  0.97296  3.90237  1.60359  0.22461  0.48576  0.95510
@
```

3. Use hmmpalign to align sequences in fasta file from step 1 to hmm found in step 2

```
hmmpalign sh2.hmm sequences.fa > sh2_hmm.sto
```

```
■ STOCKHOLM 1.0
sp|P20936|181-272          WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQM....
nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lgek1lypv...
#=GR sp|P20936|181-272 PP *****9*****9*****9*****9*****9*****9...
89*****8....7777.89*****9*****9*****9*****9...
sp|P09851|178-269          WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQT...
nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lgek1lypv...
#=GR sp|P09851|178-269 PP *****9*****9*****9*****9*****9...
89*****8....7777.89*****9*****9*****9*****9...
sp|P50904|172-263          WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQT...
nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lgek1lypv...
#=GR sp|P50904|172-263 PP *****9*****9*****9*****9*****9...
89*****8....7777.89*****9*****9*****9*****9...
sp|Q9NHV9|622-726          WFAGNMDRETAAHRLEN.-RRIGTYLLRvpqgPSTAHE TMYALSLKTDD...
nVIKHMKINQENSGdsmLYCLSSRRHFKTIVELVSYYerndlgenfaglnqslqwp...
#=GR sp|Q9NHV9|622-726 PP *****9*****9*****9*****9*****9...
89*****9*****9*****9*****9*****9*****9*****9...
sp|P00523|148-245          WFQKIKTRRESERLLLpENPRGTFLVR....ESETTKGAYCLSVDFdna...
1INVKYKIRLDSG..GFYITSRTQFSSLQQLVAYYskhadglchr1tnvc....
#=GR sp|P00523|148-245 PP *****9*****9*****9*****9*****9...
*****9*****9*****9*****9*****9*****9...
@ "sh2_hmm.sto" 15L, 1676C
```

2. Change the file format from .sto to .fa

```
perl /shared/PERL/sto2fasta.pl -g sh2_hmm.sto > sh2_hmm.fa
```

```
>sp|P50904|172-263
WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQT
....nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lkg
ek1lypv...
>sp|P00523|148-245
WFAGNMDRETAAHRLEN.-RRIGTYLLRvpqgPSTAHE TMYALSLKTDD
....nVIKHMKINQENSGdsmLYCLSSRRHFKTIVELVSYYerndlgenf
aglnqslqwp...
>tr|E9PYG6|172-263
WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQT
....nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lkg
ek1lypv...
>sp|P20936|181-272
WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQM
....nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lkg
ek1lypv...
>sp|P09851|178-269
WYHGKLDRTIAEERLRQ.AGKGSYLI R....ESDRRPGSFVLSFLSQT
....nVVNHFRIIAMCGD...-YYIGG-RRFSSLSDLIGYYshvsc1lkg
```

We have also obtained an alignment with clustal.

clustalw2 sequences.fa

```
CLUSTAL 2.1 multiple sequence alignment

sp|P09851|178-269      WYHGKLDRTIAEERLRQAGKS-GSYLIRESDRRPGS-----FVLSFLSQTN----VVNHF
sp|P50904|172-263      WYHGKLDRTIAEERLRQAGKS-GSYLIRESDRRPGS-----FVLSFLSQTN----VVNHF
sp|P20936|181-272      WYHGKLDRTIAEERLRQAGKS-GSYLIRESDRRPGS-----FVLSFLSQMN----VVNHF
sp|P00523|148-245      WYFGKITRRESERLLLNPNPRGTFLVRESETTKGA-----YCLSVSDFDnakglNVKHY
sp|Q9NHV9|622-726      WFAGNMDRetaahrlen--RRIGTYLLRVRPQGPSTAHEtmyalslktddn---VIKHM
*: *:: * : . * : . *::*::* .: : **. * ::* ~

sp|P09851|178-269      RIIAM-CGD---YYIGGR-RFSSLSDLIGYYYS--HVSCLLKGE--KLLYPV
sp|P50904|172-263      RIIAM-CGD---YYIGGR-RFSSLSDLIGYYYS--HVSCLLKGE--KLLYPV
sp|P20936|181-272      RIIAM-CGD---YYIGGR-RFSSLSDLIGYYYS--HVSCLLKGE--KLLYPV
sp|P00523|148-245      KIRKLDSGG---FYITSRTQFSSLQQLVAYYSK-HADGLCHRL--TNVC--
sp|Q9NHV9|622-726      KINQENSGDsmLYCLSSRRHFkTIVelvsvyyerndlgenfaglnqslqwpv
*: *.* : : *.* :*::*.* . . ~
```

4. Search for conserved regions in your alignment. Do these regions correspond with the essential regions you described in the previous assignment (question 6)? Why do you think this is happening? Provide images of your alignment to support your explanation. In these images, the alignments should be in clustalw format, use the perl script we learnt in practice 2 to change the format of the alignments produced by hmmer programs.

Since we are working with the domain SH2 we can say that is an essential region of the protein. As we can see at the alignment there are some conserved regions among the different species. This domain is important in the protein-protein interactions to regulate intracellular signalling.

```
hmmpalign sh2.hmm sequences.fa > sequences_hmm.sto
perl/shared/PERL/aconvertMod2.pl -in h -out c <sequences_hmm.sto>sequences_hmm.clu
```

```
CLUSTAL W(1.60) multiple sequence alignment

sp|P20936|181-272      WYHGKLDRTIAEERLRQ-AGKGSYLIr-----ESDRPGSFVLSFLSQM----nVVNHF
sp|P09851|178-269      WYHGKLDRTIAEERLRQ-AGKGSYLIr-----ESDRPGSFVLSFLSQT----nVVNHF
tr|E9PYG6|172-263      WYHGKLDRTIAEERLRQ-AGKGSYLIr-----ESDRPGSFVLSFLSQT----nVVNHF
sp|P50904|172-263      WYHGKLDRTIAEERLRQ-AGKGSYLIr-----ESDRPGSFVLSFLSQT----nVVNHF
sp|Q9NHV9|622-726      WFAGNMDRetaahrlen--RRIGTYLLRvrpqgPSTAHEtmyalslktddn---VIKHM
sp|P00523|148-245      WYFGKITRRESERLLNpENPRGTFLVR-----ESETTKGAYCLSVSDFDnakglNVKHY

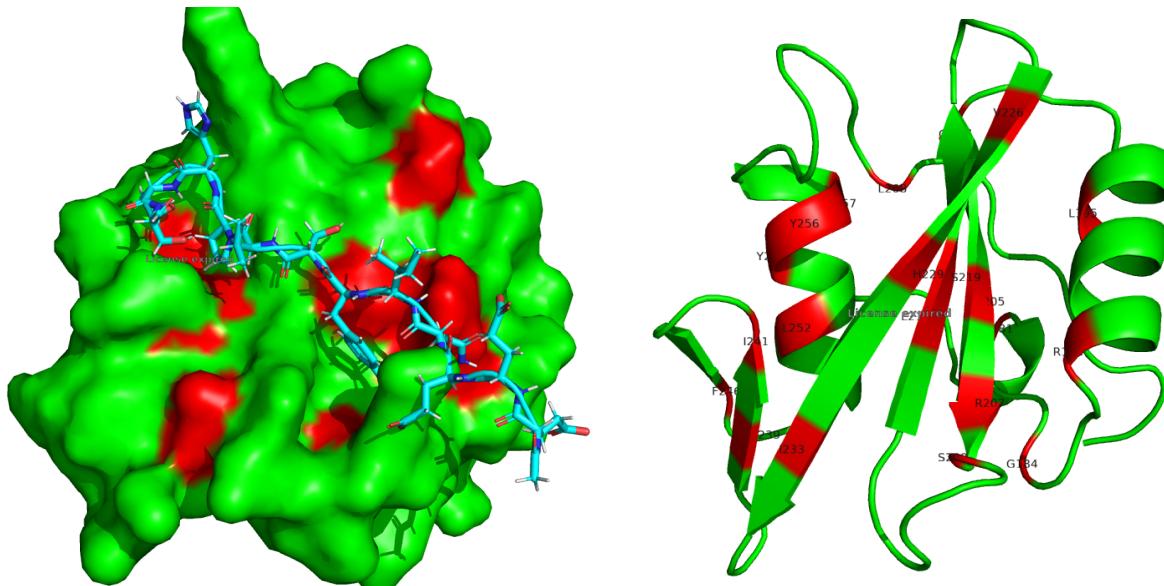
sp|P20936|181-272      RIIAMCGD---YYIGG-RRFSSLSDLIGYYshvscllkgekklypv---
sp|P09851|178-269      RIIAMCGD---YYIGG-RRFSSLSDLIGYYshvscllkgekklypv---
tr|E9PYG6|172-263      RIIAMCGD---YYIGG-RRFSSLSDLIGYYshvscllkgekklypv---
sp|P50904|172-263      RIIAMCGD---YYIGG-RRFSSLSDLIGYYshvscllkgekklypv---
sp|Q9NHV9|622-726      KINQENSGDsmLYCLSSRRHFkTIVelvsvyyerndlgenfaglnqslqwpv
sp|P00523|148-245      KIRKLDSG---GFYITSRTQFSSLQQLVAYYskhadglchrlnvc---
```

We can observe better the MSA with the following coloured image:

		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						
HUMAN_RASA1_SH2	P20936	W	Y	H	G	K	L	D	R	T	I	A	E	E	R	R	R	-	A	G	K	S	G	S	Y	L	I	R	-	-	-	E	S	D	R			
RAT_RASA1_SH2	P50904	W	Y	H	G	K	L	D	R	T	I	A	E	E	R	R	R	-	A	G	K	S	G	S	Y	L	I	R	-	-	-	E	S	D	R			
BOVIN_RASA1_SH2	P09851	W	Y	H	G	K	L	D	R	T	I	A	E	E	R	R	R	-	A	G	K	S	G	S	Y	L	I	R	-	-	-	E	S	D	R			
MOUSE_RASA1_SH2	E9PYG6	W	Y	H	G	K	L	D	R	T	I	A	E	E	R	R	R	-	A	G	K	S	G	S	Y	L	I	R	-	-	-	E	S	D	R			
CHICK_SRC_SH2	P00523	W	F	G	K	I	T	R	R	E	S	E	R	R	L	L	L	N	p	E	N	P	R	G	T	F	L	V	R	-	-	-	E	S	E	T		
FRUIT_FLY_VAV_SH2	Q9NHV9	W	F	A	G	N	M	D	R	E	T	A	A	H	R	R	E	N	-	R	R	I	G	T	Y	L	L	R	v	r	p	q	g	P	S	T	A	
		*	:	*	:	*	:	*	:	:	:	*	:	*	*	*	*	*	*	:	*	*	*	*	*	*	*	*	:	*	:	*						
		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							
HUMAN_RASA1_SH2	P20936	R	P	G	S	F	V	V	L	S	F	L	S	Q	M	-	-	-	n	V	V	N	H	F	R	I	I	A	M	C	G	D	-	-	-	Y	Y	
RAT_RASA1_SH2	P50904	R	P	G	S	F	V	V	L	S	F	L	S	Q	T	-	-	-	n	V	V	N	H	F	R	I	I	A	M	C	G	D	-	-	-	Y	Y	
BOVIN_RASA1_SH2	P09851	R	P	G	S	F	V	V	L	S	F	L	S	Q	T	-	-	-	n	V	V	N	H	F	R	I	I	A	M	C	G	D	-	-	-	Y	Y	
MOUSE_RASA1_SH2	E9PYG6	R	P	G	S	F	V	V	L	S	F	L	S	Q	T	-	-	-	n	V	V	N	H	F	R	I	I	A	M	C	G	D	-	-	-	Y	Y	
CHICK_SRC_SH2	P00523	T	K	G	A	Y	C	C	L	S	V	S	D	F	D	n	a	k	g	I	N	V	K	H	Y	K	I	R	K	L	D	S	G	-	-	G	F	Y
FRUIT_FLY_VAV_SH2	Q9NHV9	H	E	T	M	Y	A	L	S	L	K	T	D	D	-	-	-	n	V	I	K	H	M	K	I	N	Q	E	N	S	G	d	s	m	L	Y	C	
		*	*	:	:	*	:	:	*	*	*	*	*	:	*	*	*	*	*	:	*	*	*	:	:	:	:	:	:	:	:	:	:	:	:			
		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					
HUMAN_RASA1_SH2	P20936	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	s	c	l	I	k	g	e	k	I	l	y	p	V	-	-	-	
RAT_RASA1_SH2	P50904	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	s	c	l	I	k	g	e	k	I	l	y	p	V	-	-	-	
BOVIN_RASA1_SH2	P09851	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	s	c	l	I	k	g	e	k	I	l	y	p	V	-	-	-	
MOUSE_RASA1_SH2	E9PYG6	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	s	c	l	I	k	g	e	k	I	l	y	p	V	-	-	-	
CHICK_SRC_SH2	P00523	I	T	S	R	T	Q	F	S	S	L	Q	Q	L	V	A	Y	Y	s	k	h	a	d	g	I	c	h	r	l	t	n	v	c	-	-	-		
FRUIT_FLY_VAV_SH2	Q9NHV9	L	S	S	R	R	H	F	K	T	I	V	E	L	V	S	Y	Y	e	r	n	d	l	g	e	f	a	g	i	n	q	s	l	q	w	p	Y	C
		:	:	*	:	:	:	*	*	*	*	*	*	:	*	*	*	:	:	*	*	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		

From this MSA we can observe that all positions of SH2 domain in RASA are conserved

In figures below conserved regions are shown in red. Some of them seem to be interacting with the p190RhoGAP phosphotyrosine peptide. while others are stabilizing the structure. mutations in these regions can affect the structure of the domain that's why they are conserved.

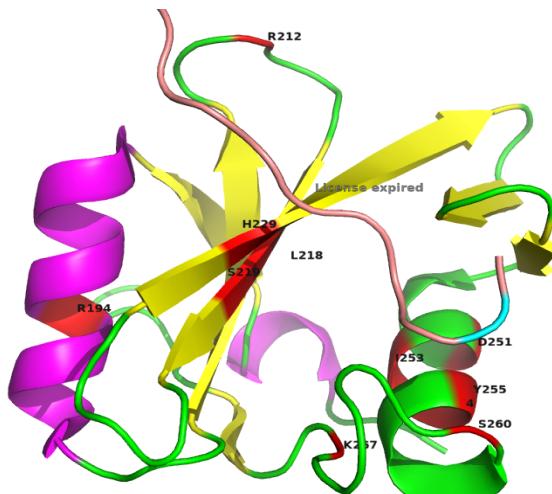


5. Work with the mutation you choose in the previous assignment (assignment 1, question 7). Find where this mutation would happen in the alignment you created in question 3. Compare the mutated amino acid with the amino acids that you find at that position in your alignment, do they share similar properties or not? Make a hypothesis of how this mutation is affecting the function of the protein. Provide images of your alignment to support your explanation.

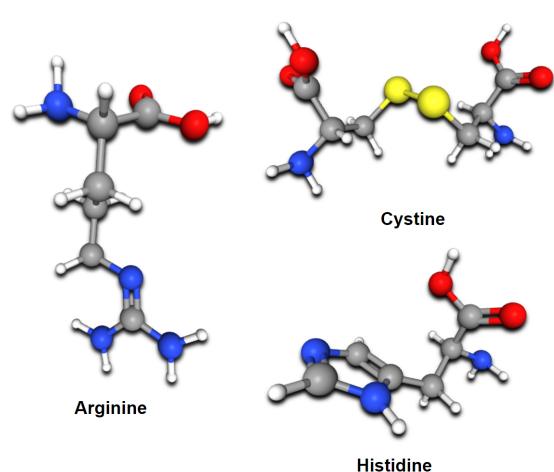
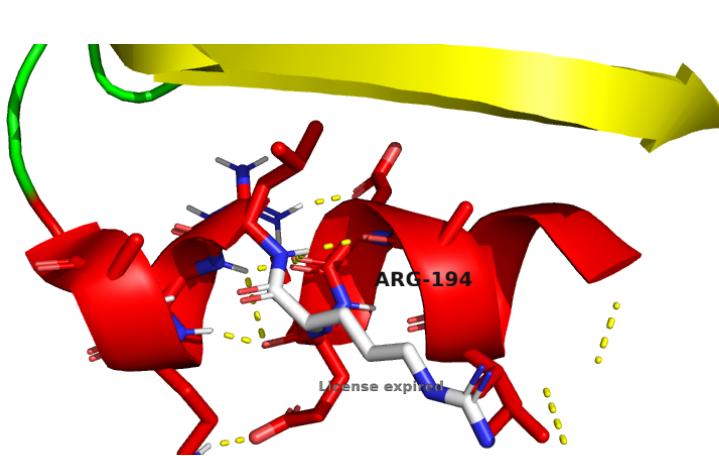
Some of the regions of RASA1 involved in cancer are shown as bold in figure below:

	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	#	200	201	202	203	204	205	206	207	208	209	210	211							
HUMAN_RASA1_SH2	P20936	W	Y	H	G	K	L	D	R	T	I	A	E	R	L	R	Q	-	A	G	K	S	G	S	Y	R	-	-	-	E	S	D	R					
RAT_RASA1_SH2	P50904	W	Y	H	G	K	L	D	R	T	I	A	E	R	L	R	Q	-	A	G	K	S	G	S	Y	L	-	-	-	E	S	D	R					
BOVIN_RASA1_SH2	P09851	W	Y	H	G	K	L	D	R	T	I	A	E	R	L	R	Q	-	A	G	K	S	G	S	Y	L	-	-	-	E	S	D	R					
MOUSE_RASA1_SH2	E9PYG6	W	Y	H	G	K	L	D	R	T	I	A	E	R	L	R	Q	-	A	G	K	S	G	S	Y	L	-	-	-	E	S	D	R					
CHICK_SRC_SH2	P00523	W	Y	F	G	K	I	T	R	R	E	S	E	R	L	L	N	p	E	N	P	R	G	T	F	L	V	R	-	-	-	E	S	E	T			
FRUIT_FLY_VAV_SH2	Q9NHV9	W	F	A	G	N	M	D	R	E	T	A	H	R	L	E	N	-	-	R	R	I	G	T	Y	L	L	R	v	r	p	q	P	S	T	A		
		*	:	*	:	*	:	*	:	:	:	:	*	:	*	*	*	*	:	*	:	*	*	*	*	*	*	:	*	:	*	:	*	:				
	212	213	214	215	216	217	218	219	220	221	222	223	224		225	226	#	228	229	230	231	232	233	234	235	236	237	238		239	240							
HUMAN_RASA1_SH2	P20936	R	P	G	S	F	V	L	S	F	L	S	Q	M	-	-	-	n	V	V	N	H	F	R	Y	I	A	M	C	G	D	-	-	-	Y	Y		
RAT_RASA1_SH2	P50904	R	P	G	S	F	V	L	S	F	L	S	Q	T	-	-	-	n	V	V	N	H	F	R	Y	I	A	M	C	G	D	-	-	-	Y	Y		
BOVIN_RASA1_SH2	P09851	R	P	G	S	F	V	L	S	F	L	S	Q	T	-	-	-	n	V	V	N	H	F	R	Y	I	I	A	M	C	G	D	-	-	-	Y	Y	
MOUSE_RASA1_SH2	E9PYG6	R	P	G	S	F	V	L	S	F	L	S	Q	T	-	-	-	n	V	V	N	H	F	R	Y	I	I	A	M	C	G	D	-	-	-	Y	Y	
CHICK_SRC_SH2	P00523	T	K	G	A	Y	C	L	S	V	S	D	F	D	D	n	a	k	g	I	N	V	K	H	Y	K	J	R	K	L	D	S	G	-	-	G	F	Y
FRUIT_FLY_VAV_SH2	Q9NHV9	H	E	T	M	Y	A	L	S	L	K	T	D	D	-	-	-	n	V	I	K	H	M	K	J	N	Q	E	N	S	G	d	s	m	L	Y	C	
		:	:	*	:	:	:	*	:	*	:	*	:	*	:	*	*	:	*	:	*	*	*	*	*	*	*	:	:	:	*	:	:	:				
	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	#	260	261	262	263	264	265	266	267	268	269	270	271	272		239	240			
HUMAN_RASA1_SH2	P20936	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	S	c	l	l	k	g	e	k	l	l	y	p	v	-	-	-	-
RAT_RASA1_SH2	P50904	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	S	c	l	l	k	g	e	k	l	l	y	p	v	-	-	-	-
BOVIN_RASA1_SH2	P09851	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	S	c	l	l	k	g	e	k	l	l	y	p	v	-	-	-	-
MOUSE_RASA1_SH2	E9PYG6	I	G	G	-	R	R	F	S	S	L	S	D	L	I	G	Y	Y	s	h	v	S	c	l	l	k	g	e	k	l	l	y	p	v	-	-	-	-
CHICK_SRC_SH2	P00523	I	T	S	R	T	Q	F	S	S	L	Q	Q	L	V	A	Y	Y	s	k	h	a	d	g	l	c	h	r	l	t	n	v	c	-	-	-	-	
FRUIT_FLY_VAV_SH2	Q9NHV9	L	S	S	R	R	H	F	K	T	I	V	E	L	V	S	V	V	e	r	n	d	l	g	e	n	f	a	g	l	n	q	s	l	q	w	p	y
	:	:	*	:	:	:	*	:	*	*	*	*	*	*	*	:	:	*	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			

Cancer causing mutations are shown in red in figure below:



Position 194: (ARG_194 shown in white)



Mutations at position 194 are present in the alpha helix region of SH2 domain of RASA1, where we have an Arginine as reference residue. Two different mutations are observed to be involved in cancer in this position.

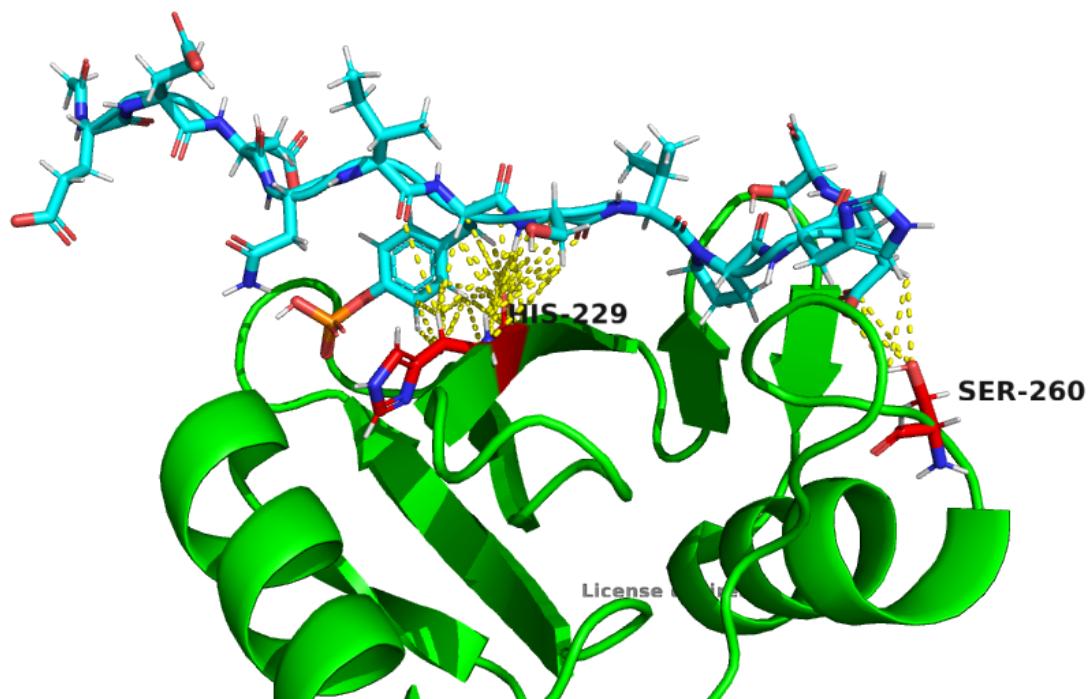
1. Altered residue C instead of R. Arginine is positively charged and Cystine is non-polar. So there will be:
 - a. Loss of electrostatics
 - b. Loss of H-bonds
 - c. Loss of vdw interactions because of small size of cystine as compared to arginine so some vdw stabilizing the helix will be lost
 - d. Cysteine can form disulphide bond

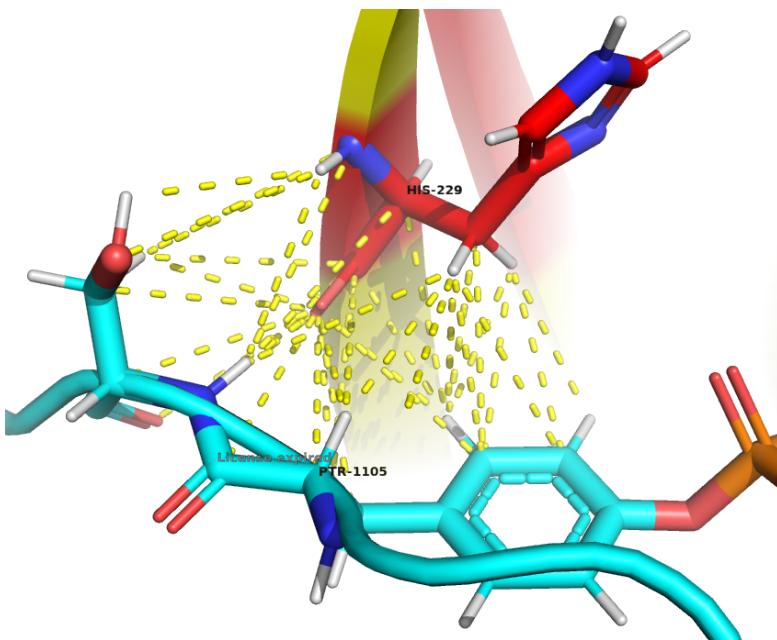
2. The second one occurs when we have an altered residue H instead of R. Histidine and arginine are both positively charged, but histidine is smaller than arginine so there will be:
 - a. Loss of electrostatics
 - b. Loss of H-bonds
 - c. Loss of vdw

In position 194 in our sequences we can find in all organisms the R residue, but in the chicken sequence instead of an R there is a L. R is positively charged and L is not.

Position 229,255,260:

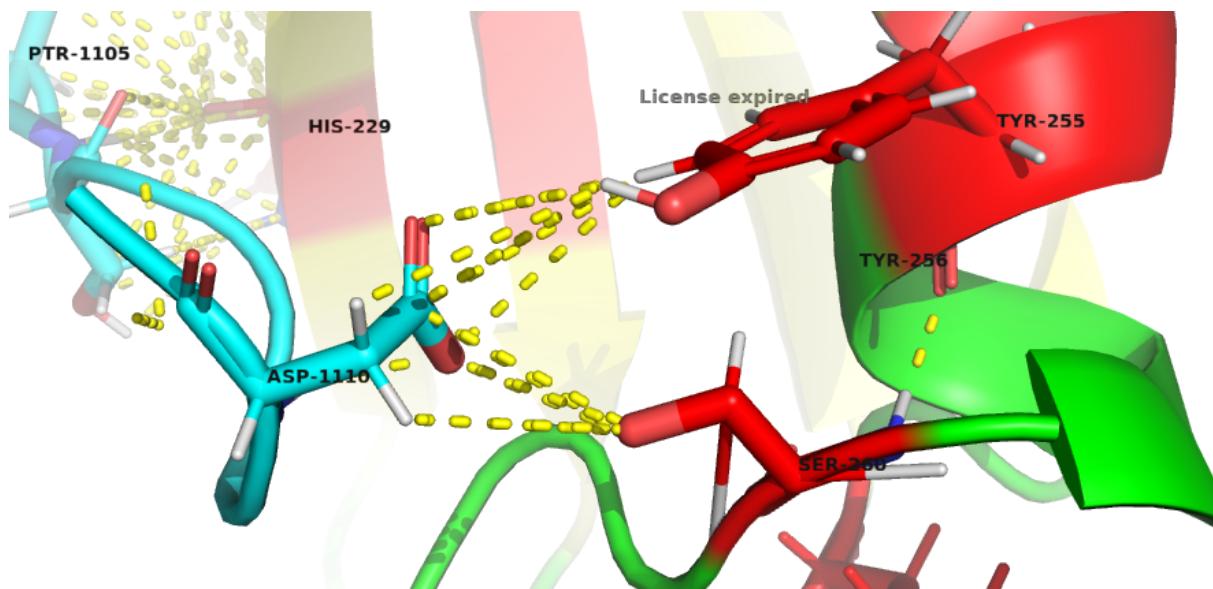
Histidine 229 , serine 260, TYR-255 seems to be involved in rho_ras activation pathway (p190RhoGAP phosphotyrosine peptide is shown in cyan colour), so maybe mutations at these positions can prevent binding and blocking the rho-ras pathways (plays an important role in cell growth, division and mortality) when required or result in overactivation of the pathway.



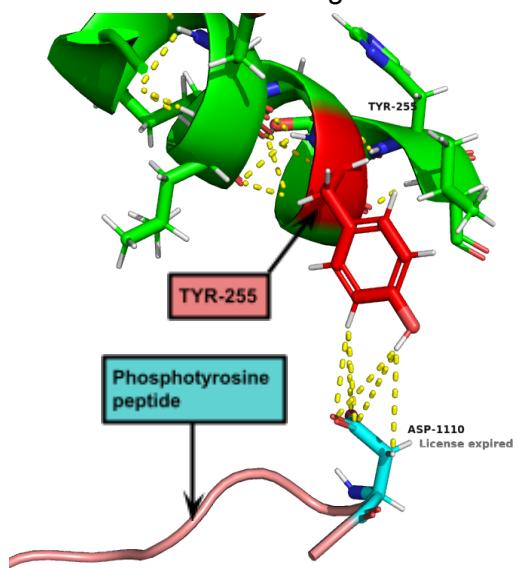


Histidine-229 (+) → Tyrosine (Neutral Polar)

Amino acids having full charges can have more stronger electrostatics than dipolar amino acids so maybe Tyrosine mutant will have less electrostatic connection with p190RhoGAP phosphotyrosine peptide



TYR-255 and TYR-256 are conserved in selected sequences and seems to be interacting with SER-260 and binding to ASP-1110 of Rho_peptide.



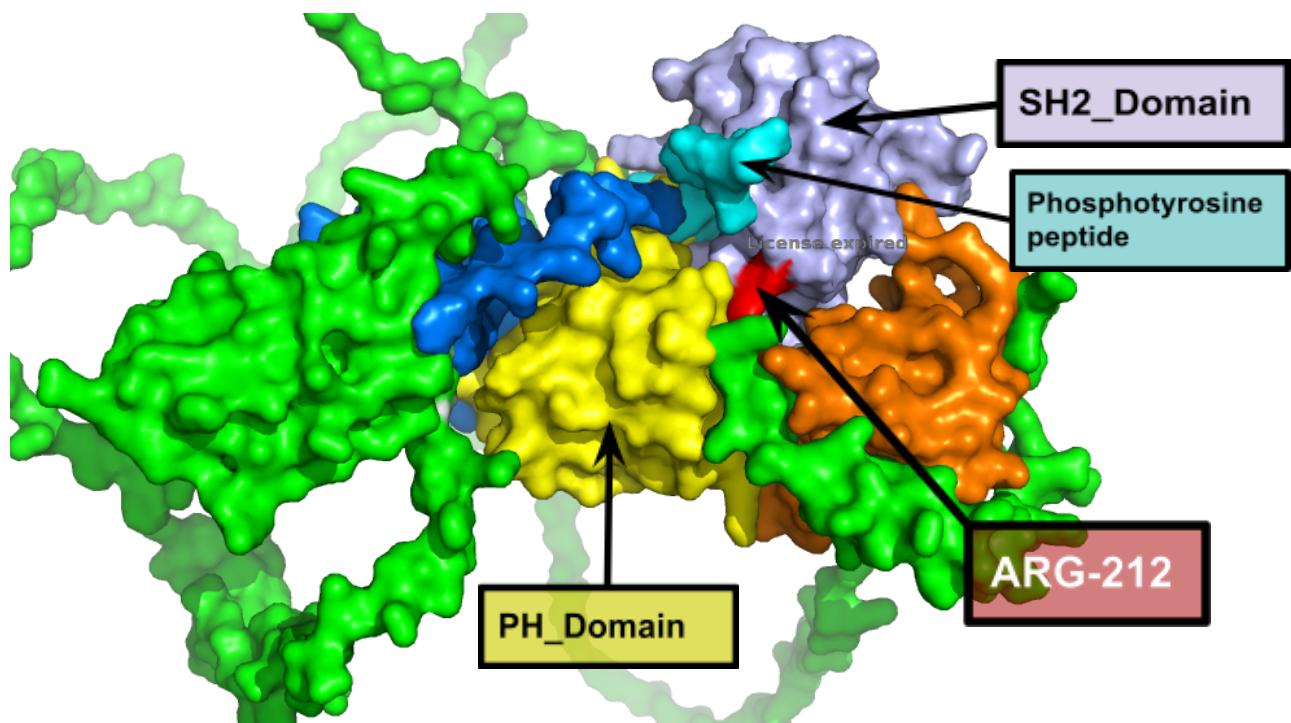
Mutation of Tyrosine 255 into phenylalanine is involved in breast cancer. Tyrosine 255 is also stabilizing the helix by H-bonds. Tyrosine is polar and phenylalanine is non polar so there will be :

- loss of H Bonds
- loss of electrostatics

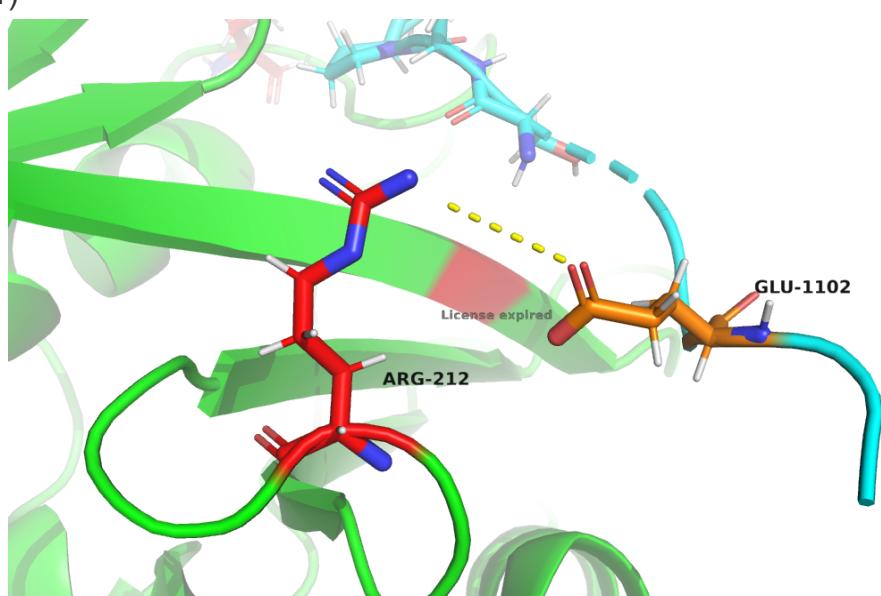
Position 212:

Arginine (+) → Methionine (Non Polar)

- loss of H-bonds
- This position seems to be on the surface of protein interacting with the solvent and non polar residue in contact with the environment can cause destabilization.

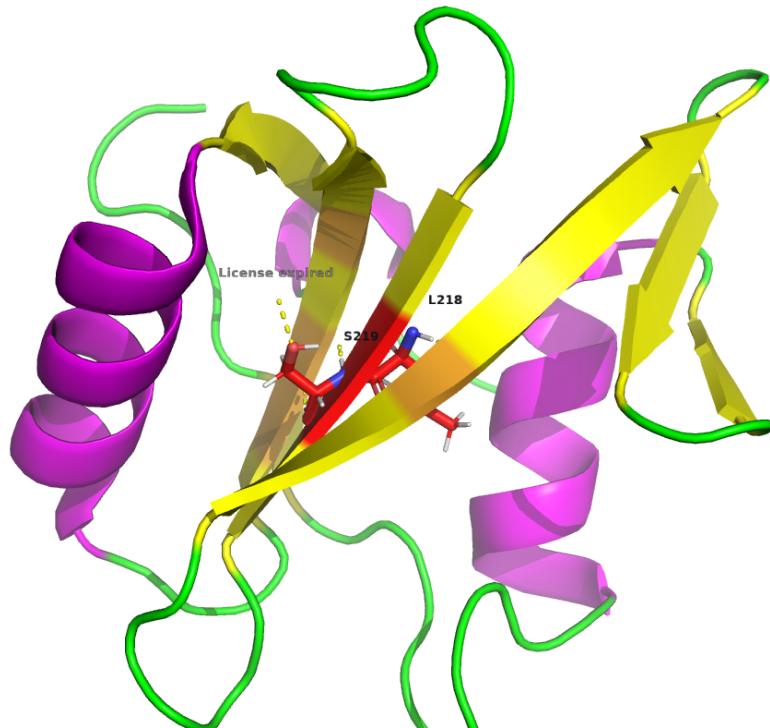


It also seems to be involved in p190RhoGAP phosphotyrosine peptide binding so mutation at this place can affect the binding. (phosphotyrosine peptide shown in cyan color)



Position 218, 219:

Region 218 and 219 are also conserved and mutations in them are involved in disease. They are present in beta sheet in center of two helices and are involved in stabilization of beta sheets



Position 218 : L → S : involved in melanoma

Position 219 : S → L : involved in urinary bladder cancer

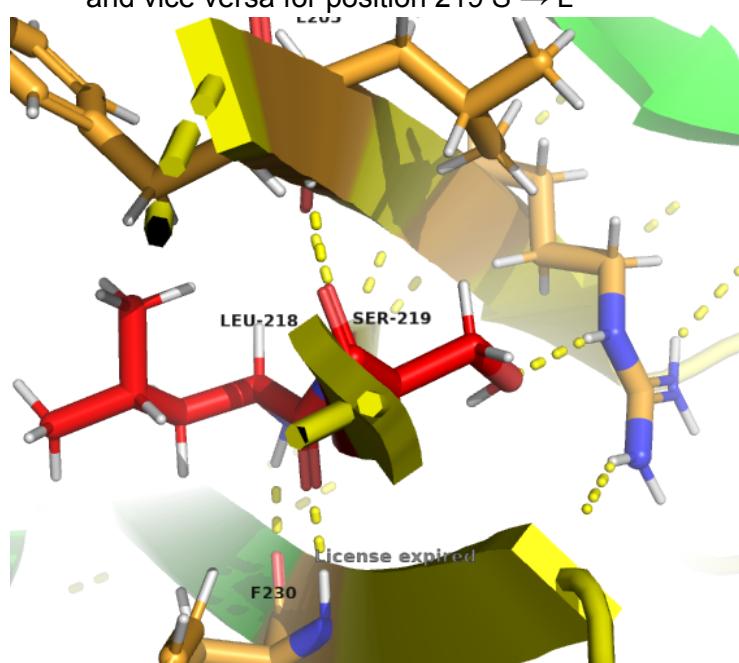
Leucine (Non Polar) ↔ Serine (Polar) :

- These regions are in a hydrophobic environment so having a polar residue may destabilize the beta sheets or it may be neutralizing the charges of other polar residues in the hydrophobic region to stabilize the structure.

Leucine (non polar) → serine (polar)

- loss of hydrophobic effect
- gain of electrostatics
- gain of H Bonds
- loss of vdw due to small size of serine

and vice versa for position 219 S → L



These two positions are on the same region on the beta sheet but on opposite sides. If the location of beta sheet was such that one side is in contact with the environment (polar) and other is not (non polar) then we could have said that

- Leucine-218 being non polar and on non polar side stabilize the structure and its mutation into Serine (non polar) would have distorted the beta sheet and vice versa for Serine 219 → Leucine mutation. But this region doesn't seem to be in contact with the surface.