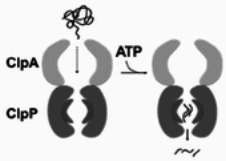
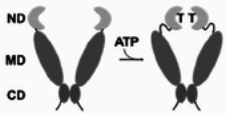
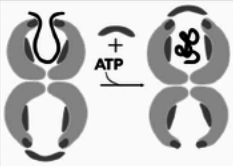
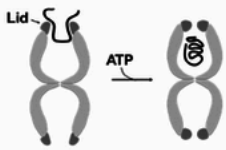
ExampleSpeakerStructural bioinformatics Project – Assignment 1: Understanding protein function.

Chaperones are a group of proteins that play a crucial role in protein folding, stability, and function within cells. The High Shock Proteins (HSP) are part of the chaperone protein family, they play a key role in making sure other proteins fold properly, especially when cells are under stress, like in high temperatures. These proteins are similar across different species, highlighting their importance. They work using ATP (a kind of cellular energy currency) to bind and release other proteins, ensuring these proteins maintain their correct 3D shape. This is crucial because misfolded proteins can lead to a variety of diseases. To sum up, HSP are like a quality control manager in a cell, keeping proteins in shape and the cell healthy.

Purple

Chaperone Family	Topology of Binding	Co-chaperone	Known Function
Hsp100		ClpP, SspB, Hsp70, Hsp40	• Works with DnaK in ATP-dependent disaggregation and proteolysis • Prevents aggregation, degradation and turnover of unassembled mitochondrial proteins • Reactivates heat-damaged proteins • Establishes and maintains prion phenotype in yeast
Hsp90		Hop, Hip, Hsp70, Immunophilins, Grp78	• Refolds proteins in stressed cells. Probable secretory chaperone in prokaryotes • Major cytosolic chaperone in eukaryotes. • Cytoprotection and intracellular signaling • In ER, controls protein homeostasis, folding and assembly of secretory proteins
Hsp70		Hsp40, GrpE	• Ubiquitous Principal folding chaperone • Works with ClpB as disaggregase • Folding of newly synthesized proteins • Protein transport into ER and mitochondria
Hsp60 (Group I)		Hsp10	• Major chaperone for protein folding in prokaryotes • Stabilizes proteins during heat stress • Promotes folding of over-produced proteins • Major chaperone in mitochondria and chloroplast
Hsp60 (Group II)		Prefoldin/GimC	• Promotes folding of a cytosolic proteins in eukaryotes • Refolding of unfolded polypeptides in vitro
sHsps		-	• Stabilizes unfolded polypeptides • Prevents aggregation • Works with Hsp70 in protein refolding • Structural protein of eye lens

Once

1. What is the function of your protein? Is the same for the whole protein family?

HSP90, an essential member of the chaperone protein family, goes beyond just assisting in protein folding under stress like high temperatures. It's integral to various cellular processes, including cell cycle control, signal transduction, and mitochondrial import, maintaining the stability and function of diverse client proteins. Working through a dynamic cycle with ATP, HSP90 interacts with co-chaperones and client proteins, ensuring they achieve their correct 3D structure. This role is vital in preventing diseases caused by misfolded proteins and is conserved across species. Essentially, HSP90 acts as a versatile quality control manager in the cell, adapting to different needs and conditions to keep the cellular machinery running smoothly.

Each HSP90 protein may have unique co-chaperones and client proteins, leading to different roles in cellular processes like signal transduction, protein trafficking, immune response, and stress response.

Thus, while the core function of assisting in protein folding is conserved across the family, the details of their activity can vary significantly.

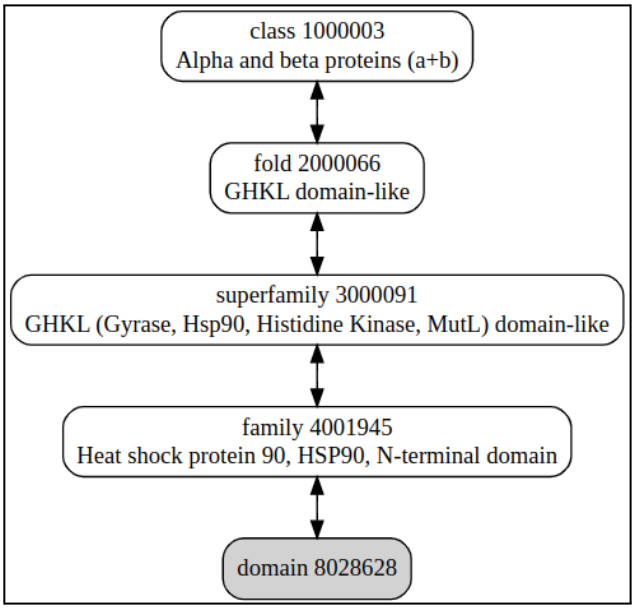
2. How is able to carry out this function?

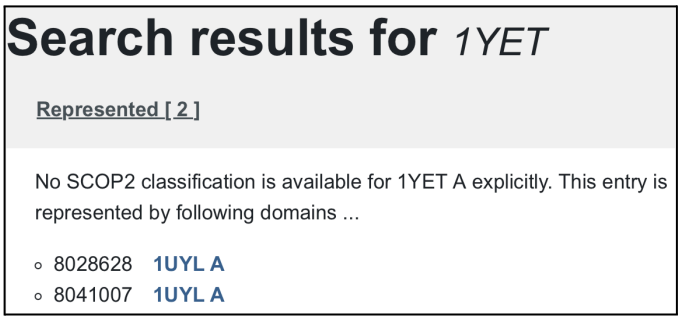
HSP90, a molecular chaperone, carries out its function through an ATP-dependent process, essential for its chaperoning activity. It binds to partially folded or unfolded protein substrates, preventing aggregation and guiding them towards their native state. This is achieved by cycling between ATP-bound and ADP-bound states, which induces conformational changes in HSP90. These changes are crucial for the folding, stabilization, and activation of its client proteins, ensuring they achieve their correct functional forms and contribute to cellular health, especially under stress conditions.

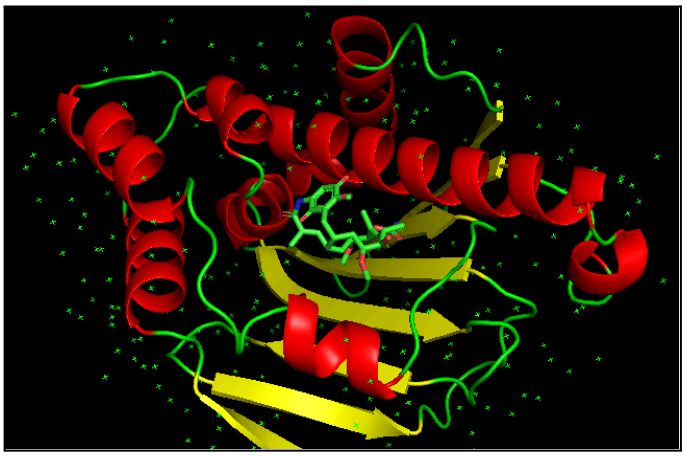
3. Does your protein require the interaction with other proteins or molecules to carry out this function?

HSP90's function crucially depends on interactions with co-chaperones and various proteins, including Hsp70, p23, and Hop, which modulate its activity and substrate specificity. These co-chaperones, along with other interacting proteins, play a pivotal role in regulating HSP90's ATPase cycle and determining the specific client proteins it assists. The client proteins, which are the direct beneficiaries of HSP90's chaperoning activity, further underscore its versatility and importance in cellular processes, from maintaining protein stability to facilitating stress responses.

4. What is the fold of your protein? Is this the same fold for the other proteins of the family?







We chose **1YET** as a representative structure for our protein HSP90 family. They belong to the superfamily GHKL, which share the domain 8028628.

```
Effective search space used: 14956560520
Results from round 2


Query= 1YET_1|Chain A|HEAT SHOCK PROTEIN 90|Homo sapiens (9606)

Length=228
                                                                      Score     E
Sequences producing significant alignments:                          (Bits)  Value
Sequences used in model and found again:

sp|P30946|HS90A_RABIT Heat shock protein HSP 90-alpha OS=Oryctola...  411     2e-140
sp|P07900|HS90A_HUMAN Heat shock protein HSP 90-alpha OS=Homo sap...  405     1e-137
sp|Q76LV2|HS90A_BOVIN Heat shock protein HSP 90-alpha OS=Bos taur...  405     1e-137
sp|Q9GKX7|HS90A_HORSE Heat shock protein HSP 90-alpha OS=Equus ca...  405     1e-137
sp|Q4R4P1|HS90A_MACFA Heat shock protein HSP 90-alpha OS=Macaca f...  405     1e-137
sp|O02705|HS90A_PIG Heat shock protein HSP 90-alpha OS=Sus scrofa...  405     2e-137
sp|P82995|HS90A_RAT Heat shock protein HSP 90-alpha OS=Rattus nor...  405     2e-137
sp|P46633|HS90A_CRIGR Heat shock protein HSP 90-alpha OS=Cricetul...  403     1e-136
sp|P07901|HS90A_MOUSE Heat shock protein HSP 90-alpha OS=Mus musc...  402     2e-136
sp|A5A6K9|HS90A_PANTR Heat shock protein HSP 90-alpha OS=Pan trog...  402     4e-136
```

```
Reference for composition-based statistics starting in round 2:
Alejandro A. Schaffer, L. Aravind, Thomas L. Madden, Sergei
Shavirin, John L. Spouge, Yuri I. Wolf, Eugene V. Koonin, and
Stephen F. Altschul (2001), "Improving the accuracy of PSI-BLAST
protein database searches with composition-based statistics and
other refinements", Nucleic Acids Res. 29:2994-3005.



Database: pdb_seq
           133,266 sequences; 30,839,592 total letters

Results from round 1


Query= 1YET_1|Chain A|HEAT SHOCK PROTEIN 90|Homo sapiens (9606)

Length=228
                                                                      Score     E
Sequences producing significant alignments:                          (Bits)  Value

2h55_A mol:protein length:256 Heat shock protein HSP 90-alpha 4       395     6e-141
2fwz_A mol:protein length:256 Heat shock protein HSP 90-alpha         395     6e-141
2fwy_A mol:protein length:256 Heat shock protein HSP 90-alpha         395     6e-141
1yc4_A mol:protein length:264 Heat shock protein HSP 90-alpha         393     5e-140
1yc3_A mol:protein length:264 Heat shock protein HSP 90-alpha         393     5e-140
1yc1_A mol:protein length:264 Heat shock protein HSP 90-alpha         393     5e-140
1uyi_A mol:protein length:236 HEAT SHOCK PROTEIN HSP 90-ALPHA         391     8e-140
2vcj_A mol:protein length:236 HEAT SHOCK PROTEIN HSP 90-ALPHA         389     5e-139
2vci_A mol:protein length:236 HEAT SHOCK PROTEIN HSP 90-ALPHA         389     5e-139
2uwd_A mol:protein length:236 HEAT SHOCK PROTEIN HSP 90-ALPHA         389     5e-139
2cdd_B mol:protein length:236 HEAT SHOCK PROTEIN HSP-90 ALPHA         389     5e-139
```

These are two proteins similar to our query, based on score and E-value:

Search results for *2UWD*

Represented [2]

No SCOP2 classification is available for 2UWD A explicitly. This entry is represented by following domains ...

- 8028628 **1UYL A**
- 8041007 **1UYL A**

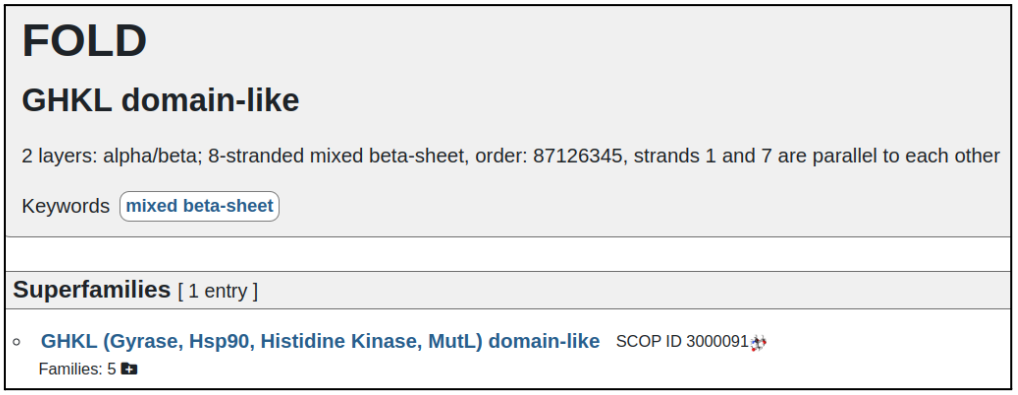
Search results for *2H55*

Represented [2]

No SCOP2 classification is available for 2H55 A explicitly. This entry is represented by following domains ...

- 8028628 **1UYL A**
- 8041007 **1UYL A**

The fold of this proteins is:



5. Are there available structures for your protein family? What are their PDB IDs?
There are 19 Family domains and their PDB IDs are:

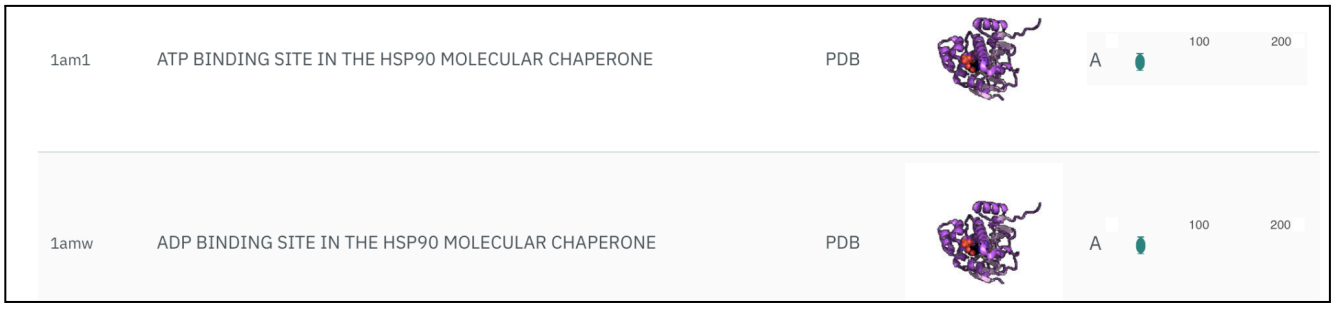
- 1UYL
- 1UYM
- 4XKA
- 3PEJ
- 2O1U
- 7K9S
- 3Q5K
- 3OPD
- 2IWX
- 2IOR
- 2JKI
- 4GQT
- 5F3K
- 6CJI
- 3O6O
- 3K60
- 4NH9
- 3IED
- 4X9L

6. Does your protein have a region that is essential for its function? What is this region? Why is it essential to its function? Is this region also essential for the other proteins of the family?
https://www.rcsb.org/annotations/6xg6

The sequences of hsp90s show a distinctive domain structure, with a highly-conserved N-terminal domain separated from a conserved, acidic C-terminal domain by a highly-acidic, flexible linker region.

N-terminal
This entry represents the histidine kinase-like ATPase (HATPase) domain Hsp90 from eukaryotes and its bacterial homologues, known as HtpG (High temperature protein G).

C-terminal
~~This entry represents the C-terminal domain of HPS90.~~ Structurally, this domain has an α-β(3)-α(3) fold arranged in two layers (α/β) with mixed β-sheet and crossing loops.



Conserved region

Biological Process	Molecular Function	Cellular Component
• protein folding (GO:0006457)	• ATP binding (GO:0005524) • unfolded protein binding (GO:0051082)	None

The signature pattern for the hsp90 family of proteins is located in a highly conserved region found in the N-terminal part of these proteins.

As we can see it is relevant for the family since these domains are shared between them.

InterPro: Protein Family Classification — InterPro: Classification of protein families

Chains	Accession	Name	Type
A	IPR003594	Histidine kinase/HSP90-like ATPase	Domain
A	IPR020568	Ribosomal protein uS5 domain 2-type superfamily	Homologous Superfamily
A	IPR001404	Heat shock protein Hsp90 family	Family
A	IPR020575	Heat shock protein Hsp90, N-terminal	Domain
A	IPR019805	Heat shock protein Hsp90, conserved site	Conserved Site
A	IPR036890	Histidine kinase/HSP90-like ATPase superfamily	Homologous Superfamily
A	IPR037196	HSP90, C-terminal domain	Homologous Superfamily

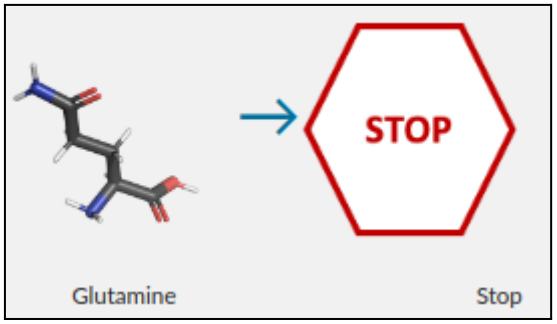
7. Use the UniProt database to choose a mutation that affects your protein. Try to find an interesting case, for example a mutation that causes a disease. Describe the effects of this mutation at a molecular and phenotypical level.

▶ COSM3793477	23	Q>*	Variant assessed as Somatic; HIGH impact. (NCI-TCGA)	NCI-TCGA Cosmic

https://www.ebi.ac.uk/ProtVar/query?search=P07900%20K314E

Glutamine → Stop

Conservation score: 0.891



(Microbial infection) Seems to interfere with N.meningitidis NadA-mediated invasion of human cells. Decreasing HSP90 levels increases adhesion and entry of E.coli expressing NadA into human Chang cells; increasing its levels leads to decreased adhesion and invasion

These are some more diseases associated with the HSP 90-alpha as a drug target:

Pharos: Disease Associations

Pharos: Illuminating the Druggable Genome

Chains	Drug Target	Associated Disease
A	Pharos : P07900	• anaplastic large cell lymphoma • contact dermatitis • HIV infectious disease • breast cancer • Neoplasm Metastasis • breast carcinoma • osteosarcoma • plasma cell myeloma • Sjogren syndrome • breast disease • exocrine pancreatic carcinoma • lung adenocarcinoma • lymphoplasmacytic lymphoma • malignant pancreatic neoplasm • ovarian cancer

CORRECT DOMAIN FOR 6xg6

Search results for *6xg6*

Represented [4]

No SCOP2 classification is available for 6xg6 A explicitly. This entry is represented by following domains ...

- 8070969 **5F3K A**
- 8070970 **5F3K A**

No SCOP2 classification is available for 6xg6 B explicitly. This entry is represented by following domains ...

- 8070969 **5F3K A**
- 8070970 **5F3K A**

FAMILY Heat shock protein 90, HSP90, N-terminal domain

DOMAIN

5F3K A:82-294

SCOP ID: 8070969

Heat shock protein 75 kDa, mitochondrial

Species *Homo sapiens*

Show ancestry ☐

Representative sequence | Other domains for this sequence | External links

UniProtQ12931
This domain

```
MARELRALLL WGRRLRPLLR APALAAVPGG KPILCPRRTT AQLGPRRNPA WSLQAGRLFS TQTAEDKEEP LHSIISSTES VQGSTSKHEP QAETKKLLDI VARSLYSEKE VPIRELISNA SDALEKLRHK LVSDQQALPE MEIHLQTNAE KGTITIQDTG IGMTQEELVS NLGTIARSGS KAFLDALQNQ
AEASSKIIGQ FGVGFYSAFM VADRVEVYSR SAAPGSLGYQ WLSDGSGVFE IAEASGVRTG TKIIIHLKSD CKEFSSEARV RDVVTKYSNF VSFPLYLNGR RMNTLQAIWM MDPKDVREWQ HEEFYRYVAQ AHDKPRYTLH YKTDAPLNIR SIFYVPDMKP SMFDVSRELG SSVALYSRKV LIQTKATDIL
PKWNLRFIRGV VDSEDIPLNL SRELLQESAL IRKLRDVLQQ RLIKFFIDQS KKDAEKYAKF FEDYGLFMRE GIVTATEQEV KEDIAKLLRY ESSALPSGQL TSLSEYASRM RAGTRNIYYL CAPNRHLAEH SPYYEAMKKK DTEVLFCFEQ FDELTLLHLR EFDKKKLISV ETDIVVDHYK EEKFEDRSPA
AECLSEKETE ELMAWMRNVL GSRVTNVKVT LRLDTHPAMV TVLEMGAARH FLRMQQLAKT QEERAQLLQP TLEINPRHAL IKKLNQLRAS EPGLAQLLVD QIYENAMIAA GLVDDPRAMV GRLNELLVKA LERH
```

SUPERFAMILY GHKL (Gyrase, Hsp90, Histidine Kinase, MutL) domain-like

DOMAIN

5F3K A:82-294

SCOP ID: 8070970

Heat shock protein 75 kDa, mitochondrial

Species *Homo sapiens*

Show ancestry ☐

Representative sequence | Other domains for this sequence | External links

UniProtQ12931
This domain

```
MARELRALLL WGRRLRPLLR APALAAVPGG KPILCPRRTT AQLGPRRNPA WSLQAGRLFS TQTAEDKEEP LHSIISSTES VQGSTSKHEF QAETKKLLDI VARSLYSEKE VFIRELISNA SDALEKLRHK LVSDGQALPE MEIHLQTNAE KGTITIQDTG IGMTQEELVS NLGTIARSGS KAFLDALQNQ
AEASSKIIGQ FGVGFYSAFM VADRVEVYSR SAAPGSLGYQ WLSDGSGVFE IAEASGVRTG TKIIIHLKSD CKEFSSEARV RDVVTKYSNF VSFPLYLNGR RMNTLQAIWM MDPKDVREWQ HEEFYRYVAQ AHDKPRYTLH YKTDAPLNIR SIFYVPDMKP SMFDVSRELG SSVALYSRKV LIQTKATDIL
PKWLRFIRGV VDSEDIPLNL SRELLQESAL IRKLRDVLQQ RLIKFFIDQS KKDAEKYAKF FEDYGLFMRE GIVTATEQEV KEDIAKLLRY ESSALPSGQL TSLSEYASRM RAGTRNIYYL CAPNRHLAEH SPYYEAMKKK DTEVLFCFEQ FDELTLLHLR EFDKKKLISV ETDIVVDHYK EEKFEDRSPA
AECLSEKETE ELMAWMRNVL GSRVTNVKVT LRLDTHPAMV TVLEMGAARH FLRMQQLAKT QEERAQLLQP TLEINPRHAL IKKLNQLRAS EPGLAQLLVD QIYENAMIAA GLVDDPRAMV GRLNELLVKA LERH
```