Bruno Álvarez, Jan Izquierdo, Jaume Jurado & Marc Trujillo

Structural bioinformatics Project – Assignment 1: Understanding protein function.

To answer the following questions, choose one member of the protein family you are working with. You will focus part of your project in this particular protein. Please, check that this protein has an available structure in the protein data bank.

https://www.uniprot.org/uniprotkb/P00846/entry

https://www.uniprot.org/uniprotkb/P03928/entry

https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P03928

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

Generates ATP from ADP by using the protons from the electrochemical gradient produced by the electron transport chain. ATP is used to store energy for later use in the cell.

The function of this protein is the same across all the family. But you can find variations in the structure and some subunits of ATP synthase between different organisms.

2. How is able to carry out this function?

The ATP synthase(F_1F_0 ATP synthase or Complex V) generates ATP from ADP, using the protons that are generated by the electron transport complexes of the respiratory chain, which generate a proton gradient across the mitochondrial membrane.

The ATP synthase is composed by:

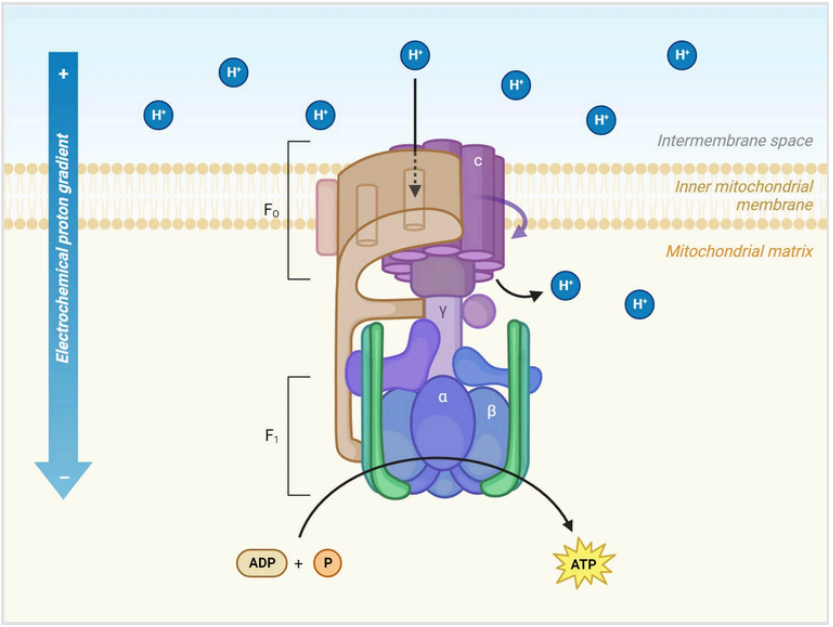
- F1: containing the extramembranous catalytic core
- F0: containing the membrane proton channel, linked together by a central stalk and a peripheral stalk.

In F1 there are three active sites capable of binding ADP and Pi, these binding sites have 3 different conformations, open, bind, and release. When the central stalk subunits are in movement due to the flow of protons through F0 the conformations in F1 changes. When F1 is in bind conformation the ADP and Pi bind to it. Which creates a High-energy bond between them thanks to the pass of a phosphate group to ADP from the Pi.

(ADP + Pi ⟶ ATP)

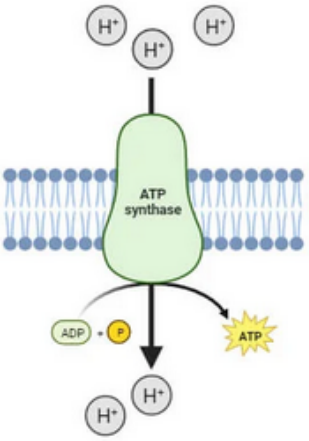
ADP + Pi + 2H ⇌ ATP + H2O + 2H

(phosphorylation)



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

Yes it does, the most important interactions of ATP synthase are with ADP, phosphates and protons. This interaction is the one that completes the main function of ATP synthase, it transforms ADP into ATP.



4. What is the fold of your protein? Is this the same fold for the other proteins of the family? (You need to do practice 1, the one about BLAST, to answer this question)

I downloaded the sequence from uniprot (P00846). Afterwards, in order to know which structure correlates more to it I performed the following commands:

psiblast -query target.fa -db /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq -num_iterations 5 -out target_pdb_5.out

psiblast -query target.fa -num_iterations 5 -out_pssm target_sprot5.pssm -out target_sprot_5.out -db /mnt/NFS_UPF/soft/databases/blastdat/uniprot_sprot.fasta

psiblast -db /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq -in_pssm target_sprot5.pssm -out target_pdb_sprot5.out

Then we can see that our structure has the ID 1c17. To see its fold and if it has the same fold as the other proteins of the family we will use the scop database.

Finally we can observe it has the following fold

FOLD

F1F0 ATP synthase subunit A

SCOP ID: 2000351

core: up-and-down bundle of four transmembrane helices

Keywords bundle up-and-down

So we can conclude that the protein has a characteristic arrangement that has a core structure of a bundle formed by 4 transmembrane helices that show an “up-and-down” orientation This “up-and-down” means that the helices in the bundle are alternate in their directionality, this creates a structural pattern in the fold of the protein.

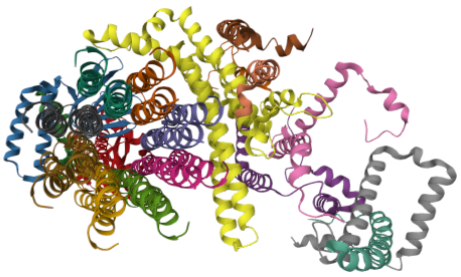
Regarding protein of the family we could not find much, however from what we saw, the fold might be the same for the other proteins of the family.

5. Are there available structures for your protein family? What are their PDB IDs?

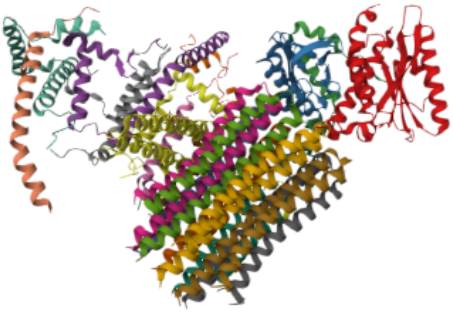
Yes, there are structures available for our protein family. Here are the available structure for our chosen protein:

PDB	8H9F
PDB	8H9J
PDB	8H9M
PDB	8H9Q
PDB	8H9S
PDB	8H9T
PDB	8H9U
PDB	8H9V

As an example, here are the structure of the first two:



3D Structure: 8H9J

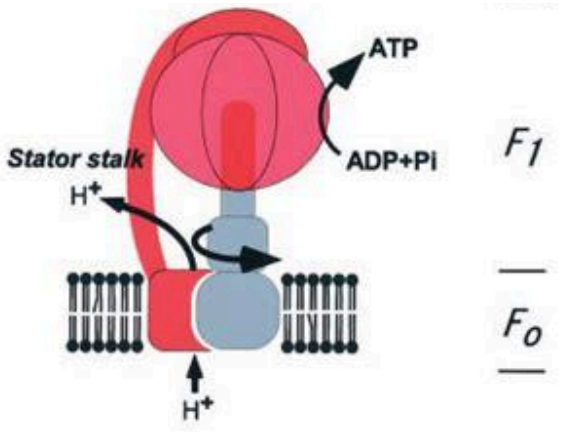


3D Structure: 8H9M

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?

The most important region of the protein would be the F1 region, as it is the one that synthesizes the ATP, the F0 part is also important and necessary, but it's not the part that

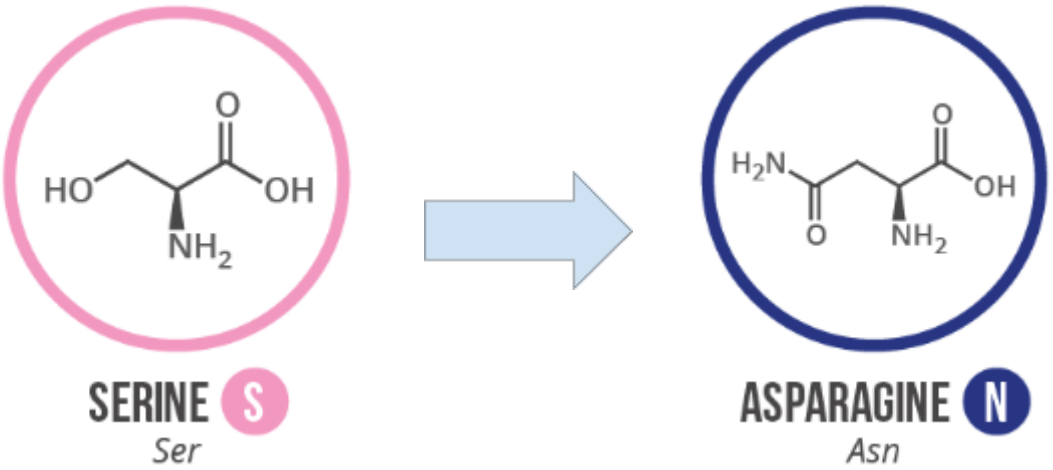
carries out the main function of the protein.
This region is essential to the synthesis of ATP because it uses the energy created by the flow of protons through F0 to transform ADP and P into ATP, we can also observe that F0 is also essential because it generates the energy F1 needs to synthesize ATP using the proton gradient (The proton gradient is generated by the flow of electrons in the electron transport chain, but that is not part of the system we are observing).
This region is essential in all proteins of the family, as it is needed to accomplish the main function of this protein. The proteins in this family are defined by their function, so if a protein of this family has a non-essential F1 part it wouldn't be accomplishing its function, therefore it wouldn't be part of the family.



However, despite having a region that does the main function of the protein, any mutation that affects the structure on any part of the protein will cause the protein to malfunction.
Therefore, despite being one region that is considered to be more relevant we could also consider that all parts of the protein are essential.

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 phenotypic level.

An interesting case of a mutation that causes a disease from UniProt database is the one with variant ID VAR_073699 (at subunit a). The mutation occurs in position 148 and it changes from Serine to Aspargine(S>N).



Name of the disease: myopathy, lactic acidosis and sideroblastic anemia 3 (MLASA3).

PHENOTYPE:

Sideroblastic anemia, muscle weakness, exercise intolerance associated with persistent lactic acidemia. Additional features: failure to thrive, hearing loss, epilepsy, stroke-like episodes, severe developmental delay.

EFFECTS ON A MOLECULAR LEVEL:

Serine is a polar amino acid that has hydroxyl group in its side chain while asparagine is also polar but it has an amide group in its side chain. However, there is a change in polarity because serine is more polar than asparagine meaning that we pass from a more polar amino acid to a one that is less polar. It is something that will change the 3d structure of the protein because of the difference of interactions between nucleotides caused by this change of polarity.

The change of amino acids also affects the formation of hydrogen bonds. Despite both amino acids being able to form hydrogen bonds (seranine through its oxygen atoms and asparagine through the amide group), serine is able to form stronger hydrogen bonds due to oxygen being more electrostatic than nitrogen. Thus we pass from a single very strong hydrogen bond donor to 2 weaker hydrogen bond donors. It surely will affect the structure (but not a lot).

In addition, we pass from a smaller amino acid to a bigger one, which also changes the 3d structure due to its different size. There is also the fact that the difference in size affects Van Der Waals interaction (bigger = more interaction) but in our case the difference is not very big so it shouldn't affect as much as the other conditions we have previously discussed.

Overall, the change, from a molecular point of view, doesn't seem to be super big. The change is basically passing from polar to less polar, from 1 strong hydrogen bond donor to 2 weaker hydrogen bond donors and from a small amino acid to a slightly bigger one.

CONCLUSION:

Despite the fact that from a molecular level the change doesn't seem to be super big, it causes a great and serious amount of problems. We think this is due to the fact that the mutation, despite not being a big change, it negatively affects an extremely important protein. ATP synthetase is in charge of producing energy for the cells (a relatively small mutation affects a vital protein causing a great impact).

If we think about it, the synthoms make sense and can be explained by the lack of ATP production:

- Anemia could be caused by the lack of energy in red blood cells making them die or making mother cells to replicate at a slower rate and, as a result, they produce an insufficient amount of red blood cells which ends up causing anemia.
- Muscle weakness due the lack of muscle cells (lack of ATP causes development problems) or the low performance caused by the lack of ATP
- Exercise intolerance because there is not enough energy for the body to function normally when the consumption is increased.
- Persistent lactic acidemia: the ATP production through normal pathways is insufficient so the body uses other metabolic pathways to get energy that in normal conditions

would take place only under a very high demanding physical activity. The lactic acid is a product of the alternative pathway and because it is produced at a higher rate the body cannot eliminate it fast enough causing problems.