ASSIGNMENT 1: UNDERSTANDING PROTEIN FUNCTION.

STRUCTURAL BIOINFORMATICS PROJECT

ADRIÀ NAVARRO PAU
JOAN LLORET CASTELLVÍ
SERGI OCAÑA ALAMO
JANA MORENO RUESTES

INDEX

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.

INTRODUCTION

Understanding the roles that proteins play in diverse biological processes is critical to comprehending the complexity of living beings. Our focus in this structural bioinformatics study is on α-amylase 1A, a member of the α-amylase family that catalyzes the hydrolysis of (1->4)-α-D-glucosidic linkages and breaks down starch molecules. This enzyme's function starts in the mouth cavities.

All enzymes known as the α-amylase family have the same objective, catalyzing the hydrolysis of α-1,4 glycosidic linkages. Our project tries to show the important aspects that contribute to activity and the structural and functional characteristics of α-amylase 1A, including how it carries out its enzymatic function.

During the project, we use the knowledge from pertinent scientific publications, databases like UniProt and the Protein Data Bank (PDB), structural analyses to give a thorough understanding of α-amylase 1A with PymoL and also alignment programs like blast.

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

The protein member we selected is α-amylase 1A (The structure available in the protein data bank is 1Z32). Its function is to degrade starch molecules and to catalyze the hydrolysis of *(1->4)-α-D-glucosidic* bonds. Meaning that it acts as a calcium-binding enzyme that initiates starch digestion in the oral cavity.

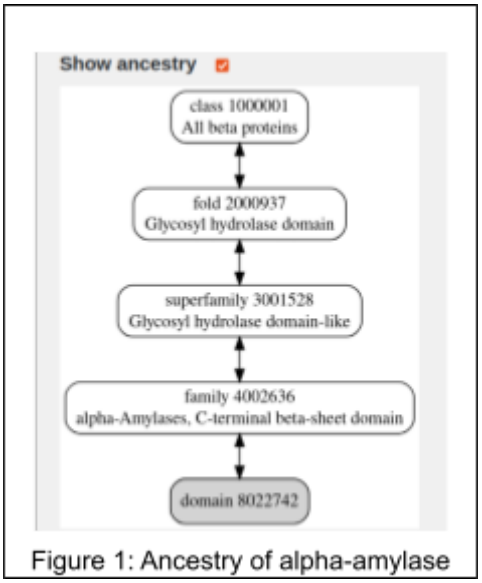


Figure 1: Ancestry of alpha-amylase

It's an endo-enzyme that belongs to the α-Amylase family, whose vast majority of integrants main function is to Catalyze the hydrolysis of α-1,4 glycosidic bonds glycosidic linkages on oligosaccharides similar to glycogen.

The α-Amylase Family is also a part of a more extensive family tree going from wider and more generic groups to our specific protein as seen in the image below, all of which main function is to act as enzymes.

2. How is able to carry out this function?

The catalytic process involves the active site of the enzyme where particular amino acid residues engage with the substrate, starch, and promote the hydrolysis reaction. α-amylases are endo-enzymes, which means they split the polysaccharide chain's internal glycosidic linkages to produce a variety of shorter oligosaccharides.
For the case of α-amylases 1A it binds to calcium, which plays an important role in stabilizing the enzyme structure and in increasing the reaction rate.

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

Yes, since its purpose as an enzyme is to facilitate a reaction it needs to interact with surrounding molecules in order to accomplish its function.

First, the starch molecules, as said previously α-amylase 1A is an endo-enzyme, it can interact with starch molecules on its own to catalyze the hydrolysis of α-1,4 glycosidic linkages inside the polysaccharide complex of starch. The enzyme can break down starch into smaller oligosaccharides and dextrins thanks to its independent catalytic activity.

On the other hand, every α-amylase 1A subunit is capable of binding a single calcium ion. The stability and appropriate folding of the enzyme depend on the binding of calcium ions. It is possible that certain amino acid residues in the active site of the enzyme interact with calcium ions, preserving the structural integrity of the enzyme and its catalytic activity.

And finally, it has been observed that α-amylase 1A also binds a single chloride ion per subunit. These monomer-monomer interactions involve the binding of small molecules such as calcium and chloride ions. These interactions contribute to the overall structure and function of the enzyme.

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)

The structure of our protein contains α-helix and parallel β-strands, which in the center of our protein the β-strands form a β-barrel which contains 8 β-strands (Figure 2.). That is defined by a cylindrical arrangement of β strands.

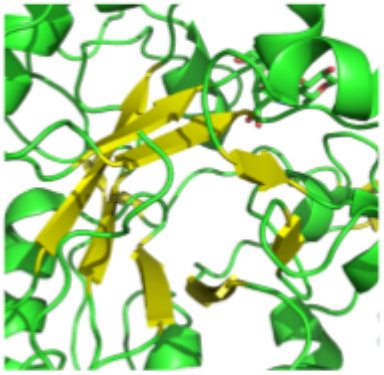


Figure 2: β-barrel

When we ran BLAST analysis, we found that the protein family has a high level of structural conservation. The E-value close to zero from the first three rounds of BLAST indicates significant similarity. Subsequent rounds also demonstrated consistent structural features, further supporting the notion that the family members share the same structure.

The enzyme also demonstrates particular catalytic activity. In polysaccharides with three or more (1->4)-α-linked D-glucose units, it catalyzes the endohydrolysis of these links. The generic name for this enzyme activity is "4-α-D-glucan glucanohydrolase." The enzyme is also referred to by a number of other names, such as "Taka-amylase A," "α amylase," "Endoamylase," and "Glycogenase." These other names are indicative of its function in hydrolyzing particular glucosidic connections to break down complex carbs like starch or glycogen.

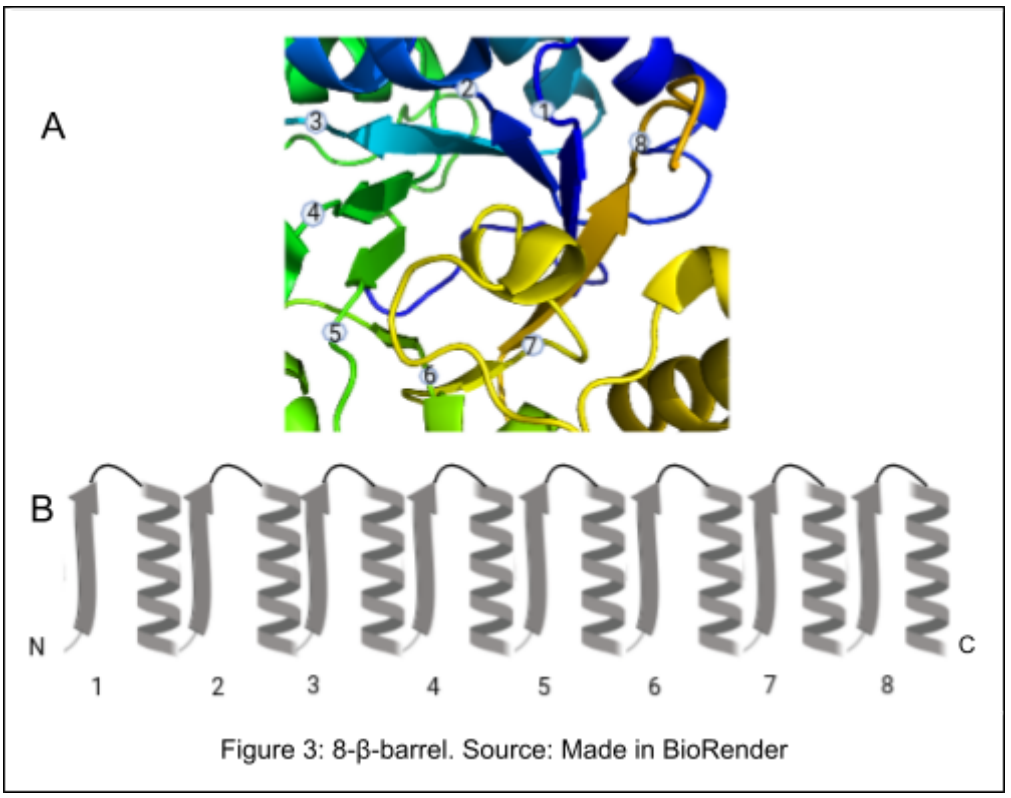


Figure 3: 8-β-barrel. Source: Made in BioRender

The (βα) 8-barrel Fold. (A) View of the β sheets, colored blue, green, and red, which form a cylindrical parallel β-sheet. (B) Topological diagram of (βα) 8-barrel Fold.

In summary, the family members share a common β-barrel fold, as supported by both structural analysis and the consistent results obtained through BLAST.

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

For the α amylase in humans we have found some structures in the databases. In PDB the IDs are 1SMD, 1HNY, among others.

If we dive into the PDB IDs and explore their structure, we can observe that all three α-amylases will have the same structure (Cl- and Ca+2). We can also see that in 1HNY, the α-amylase is found in the pancreas.

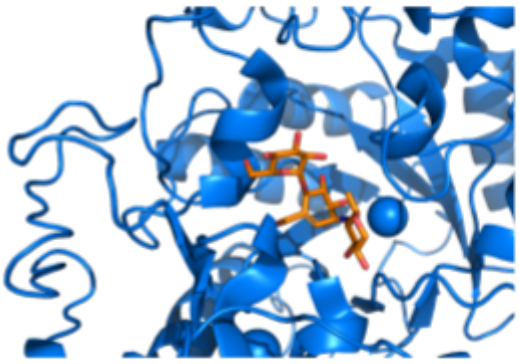
6. Does your protein have a region that is essential for its function? What is this region? Why is it essential to its function? Is this region also essential for the other proteins of the family?

Our protein has 2 regions that are essential for its function.

The most intuitive one is its active site.

Is the part of the enzyme where substrate molecules are bound and undergo a chemical reaction. This active site is proton donor (sequence D) and nucleophile (sequence E). In our structure, we find ligands that can identify this region (GLC, α-D-glucopyranose; HMC, 5-hydroxymethyl-chonduritol; and AGL, 4-amino-4,6-dideoxy-α-D-glucopyranose).

The second essential region is the bindings with calcium and chlorine.
α-amylase has a characteristic in which it contains calcium and chloride ions. The calcium ion acts as a calcium-binding ion, and as a result, the enzyme may initiate starch digestion by binding calcium.
Most of the other proteins in the family also contain the region for calcium and chloride ions. This is because as we said before, it is a characteristic specific to α-amylase.



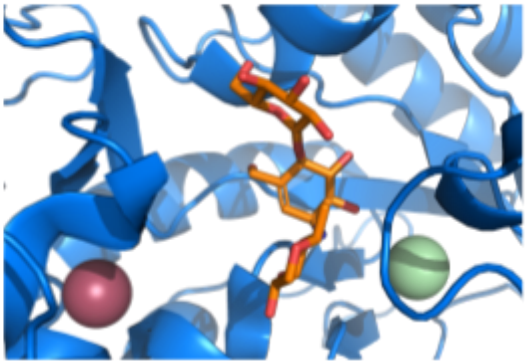


Figure 4 and 5: In the image of the left, we can observe only the active site, with the 3 ligands. In the image of the right, we can observe the active site, the calcium ion (colored in red raspberry) and the chloride ion (colored in cyan palecyan)

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.

The mutation we take is a TCGA novel which is located in position 296. The variant is assessed as Somatic and has a MODERATE impact. (NCI-TCGA) from which there is not enough information, but what we found is from the UniProt database:

Molecular Level:

- **Mutation Type:** A missense mutation is a single nucleotide alteration at position 103660367, where an adenine (A) takes the place of a guanine (G).
- **Genomic Location:** 1p21.1, at position NC_000001.11:g.103660367G>A.
- **Transcript Impact:** The NCI-TCGA transcript ENST00000370083 is mutated at c.886, which results in an amino acid change at p.G296R.

Phenotypic Level:

- **Amino Acid Change:** At position 296, arginine (R) replaces glycine (G) as a result of a missense mutation.
- **Consequence Type:** A single amino acid in the encoded protein sequence can be changed by missense mutations, which can change the structure and function of the protein.
- **Potential Disease Association:** Since functional changes in proteins can result in a variety of medical disorders, the detrimental predictions made by PolyPhen and SIFT raise the possibility of a connection with a disease.

BIBLIOGRAPHY

1. *AMY1A (amylase, α 1A) | Center for Academic Research and Training in Anthropogeny (CARTA)*. (s. f.). https://carta.anthropogeny.org/moca/topics/amy1a-amylase-α-1a

2. Bank, R. P. D. (s. f.). *RCSB PDB - 1Z32: Structure-function relationships in human salivary α-amylase: Role of Aromatic Residues*. https://www.rcsb.org/structure/1Z32

3. *Conserved Protein Domain Family AmyAc_family*. (s. f.). NCBI. https://www.ncbi.nlm.nih.gov/Structure/cdd/cl38930

4. *InterPro*. (s. f.). InterPro. https://www.ebi.ac.uk/interpro/entry/cdd/CD11319/

5. MMDB. (s. f.). *1Z32: Structure-function relationships in human salivary α-amylase: Role of Aromatic Residues*. https://www.ncbi.nlm.nih.gov/Structure/pdb/1Z32

6. Protein Data Bank in Europe. (s. f.). *PDB 1Z32 Function and Biology ‹ Protein Data Bank in Europe (PDBE) ‹ EMBL-EBI*. https://www.ebi.ac.uk/pdbe/entry/pdb/1z32/biology

7. Ragunath, C., Manuel, S. G. A., Kasinathan, C., & Ramasubbu, N. (2008). Structure-function relationships in human salivary A-amylase: role of aromatic residues in a secondary binding site. *Biologia*, *63*(6), 1028-1034. https://doi.org/10.2478/s11756-008-0163-3

8. *UniProt*. (s. f.). https://www.uniprot.org/uniprotkb/P0DUB6/entry

9. Interpro. https://www.ebi.ac.uk/interpro/entry/cdd/CD11319/