**MINI PROJECT**

**ANALYSIS OF TWO PROTEINS**

**PROBLEM STATEMENT:**

*To analyze and align two different proteins in terms of their sequence, structure, and function.*

**METHODOLOGY:**

To solve bioinformatics problems, sequence data retrieval is crucial. The primary goal of this process is to determine the evolutionary relationship between functional regions in a sequence and sequences from different organisms. Protein Data Bank (PDB) has been the centralized repository for 3D structures of proteins, nucleic acids, and complex assemblies since 1971. The proteins chosen are Prothrombin and Death-effected domain protein. Both proteins' FASTA sequences were retrieved. Both proteins were analyzed to see if they had any similar functions.

Using Pairwise Sequence Alignment, you can identify regions of similarity between two biological sequences (protein or nucleic acid) that can determine functional or structural relationships. A global alignment tool aligns sequences from end to end. By using the Needleman-Wunsch algorithm, EMBOSS Needle aligns two sequences optimally.

In RasMol, proteins, nucleic acids, and small molecules are visualized using molecular graphics. With PredictProtein (PP), you can search public sequence databases, create alignments, and predict aspects of protein structure and function. Upon sending a protein sequence, users receive a single file containing database comparisons and predictions.

A protein domain analysis can help classify and separate proteins based on their functional motifs. During evolution, secondary structures are much more conserved than sequences. PyMOL is a multiplatform molecular graphic tool widely used to visualize macromolecules in 3D. A variety of plugins have been developed for PyMOL, which include macromolecular analysis, homology modeling, docking of proteins and ligands, pharmacophore modeling, VS, and MD simulations. The Swiss-PDBViewer (aka DeepView) provides a user-friendly interface for analyzing multiple proteins simultaneously. The proteins can be superimposed in order to derive structural alignments and compare their active sites.

**RESULTS:**

***Prothrombin ( PDB ID-6C2W):***



*FASTA Sequence:*

>6C2W\_1|Chains A, B|Prothrombin|Homo sapiens (9606)

ANTFLEEVRKGNLERECVEETCSYEEAFEALESSTATDVFWAKYTACETARTPRDKLAACLEGNCAEGLGTNYRGHVNITRSGIECQLWRSRYPHKPEINCTTHPGADLQENFCRNPDSSTMGPWCYTTDPTVRRQECSIPVCGQDQVTVAMTPRSEGSSVNLSPPLEQCVPDRGQQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPDGDEEGVWCYVAGKPGDFGYCDLNYCEEAVEEETGDGLDEDSDRAIEGRTATSEYQTFFNPRTFGSGEADCGLRPLFEKKSLEDKTERELLESYIDGRIVEGSDAEIGMSPWQVMLFRKSPQELLCGASLISDRWVLTAAHCLLYPPWDKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIHPRYNWRENLDRDIALMKLKKPVAFSDYIHPVCLPDRETAASLLQAGYKGRVTGWGNLKETWTCNVGKGQPSVLQVVNLPIVERPVCKDSTRIRITDNMFCAGYKPDEGKRGDACEGDSGGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYGFYTHVFRLKKWIQKVIDQFGEYLE

***Human TNFRSF25 death domain [DED] (PDB ID-5YGS):***

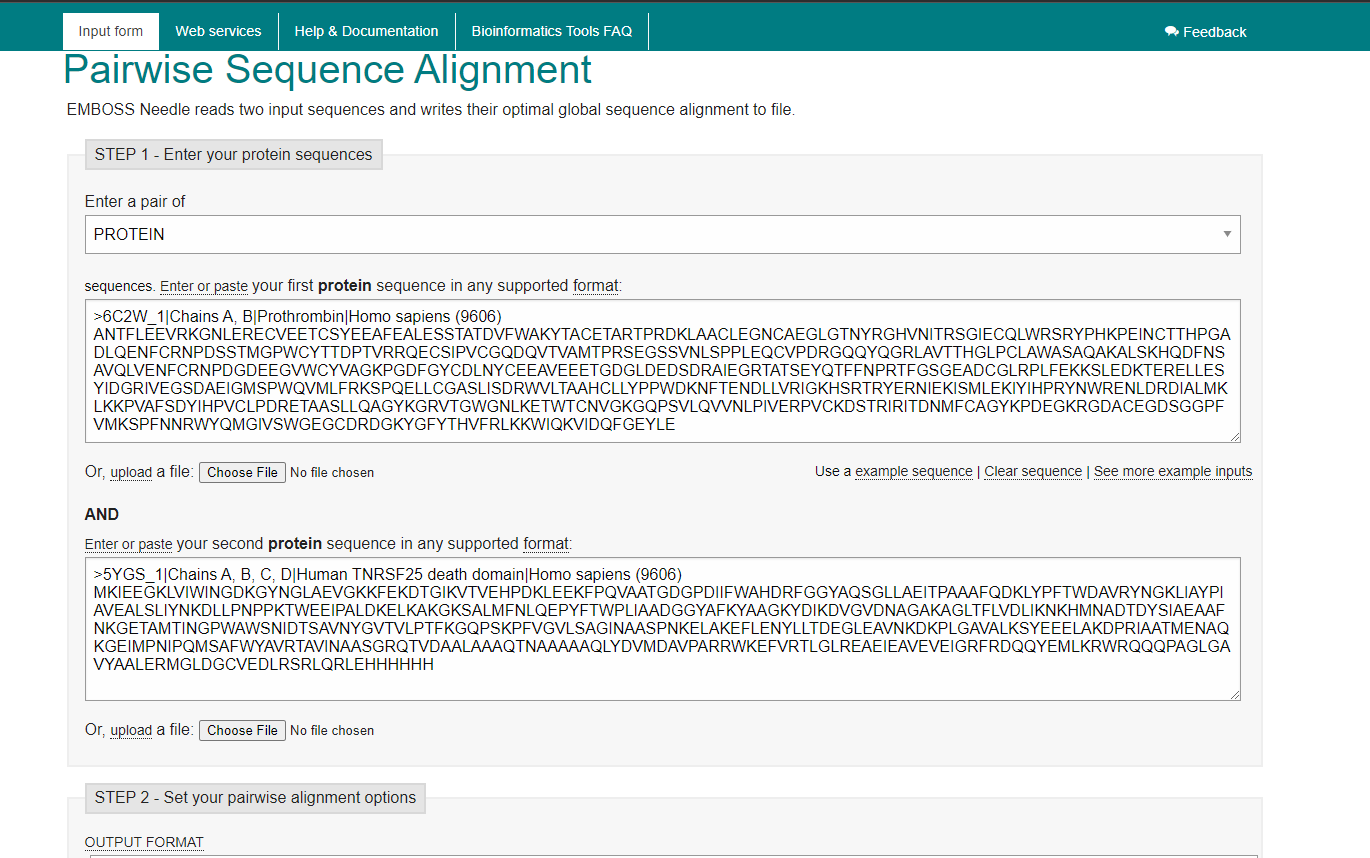


*FASTA Sequence:*

>5YGS\_1|Chains A, B, C, D|Human TNRSF25 death domain|Homo sapiens (9606)

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPAAAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYAAGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSAVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDAALAAAQTNAAAAAQLYDVMDAVPARRWKEFVRTLGLREAEIEAVEVEIGRFRDQQYEMLKRWRQQQPAGLGAVYAALERMGLDGCVEDLRSRLQRLEHHHHHH

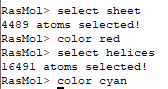
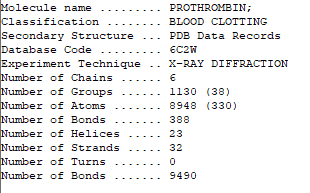
1. ***Pairwise Alignment***



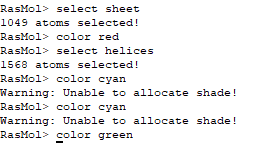
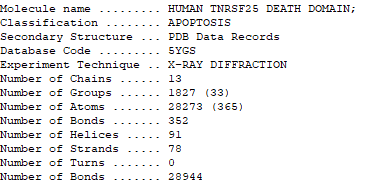
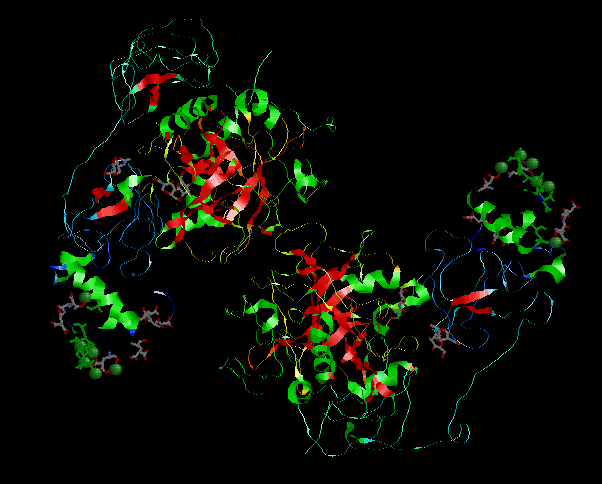


1. ***Structure Visualization through Rasmol***

***Prothrombin:***

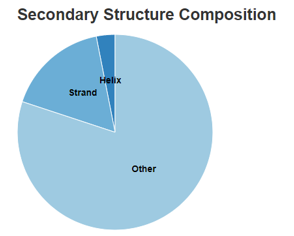
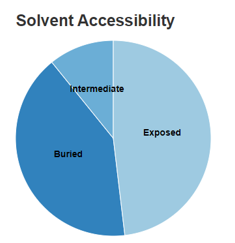
 

***Human TNFRSF25 death domain:***

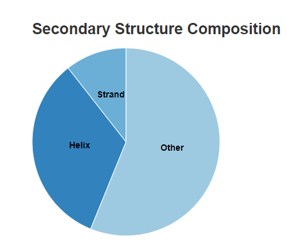
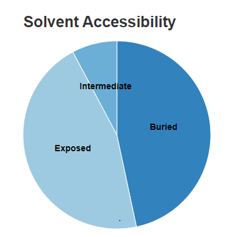
 

1. ***Predict using the secondary structure composition and solvent accessibility by Predict Protein***

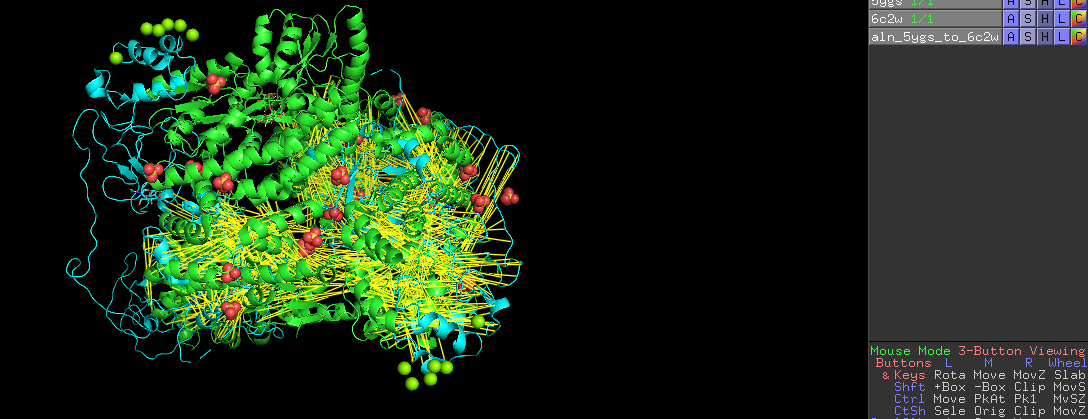
***Prothrombin:***

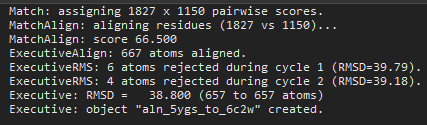
 

***Human TNFRSF25 death domain:***

1. ***Structural Alignment by Pymol***

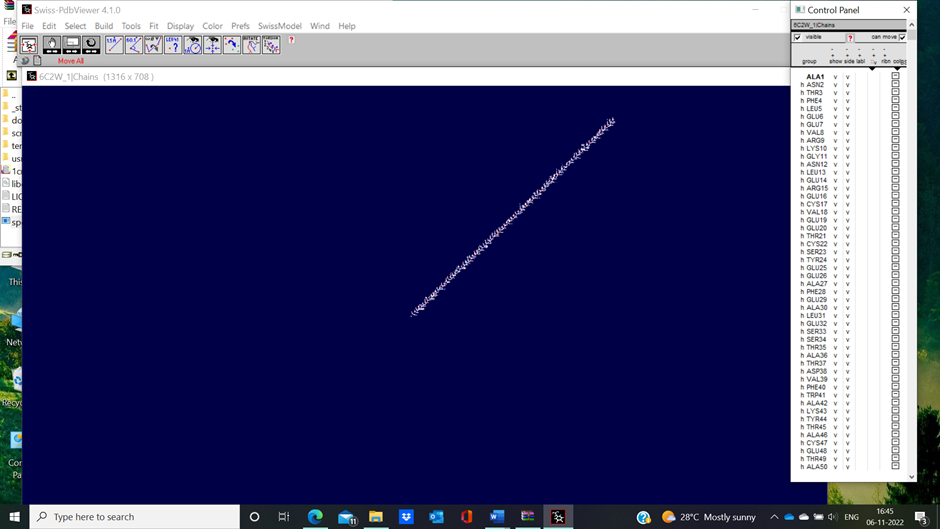




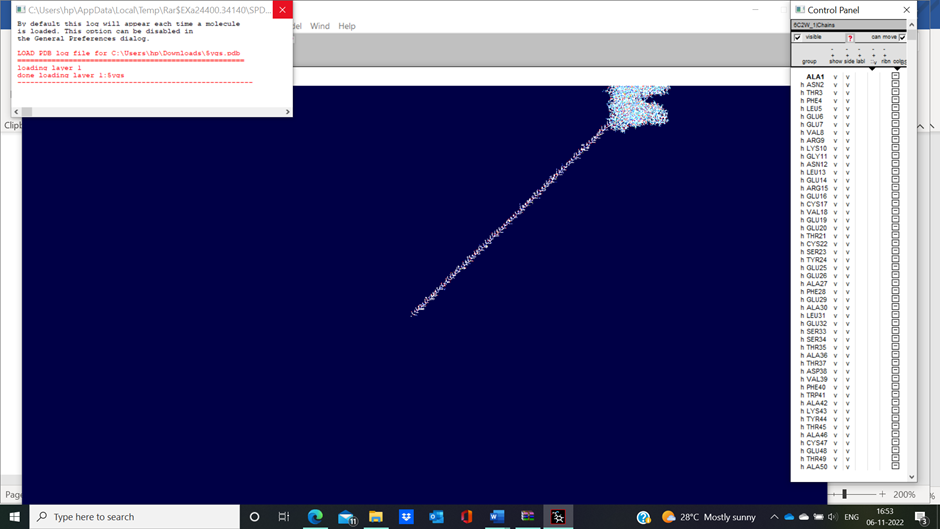
RMSD value is 38.8

1. ***Constructing a model based on the homology and similarity of the two proteins using Swiss PDB Software***

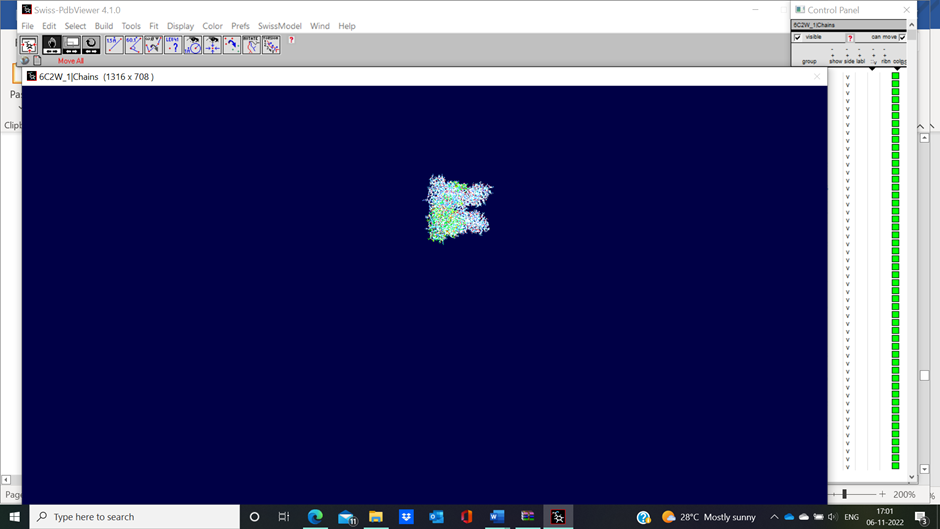
* Prothrombin sequence (PDB ID-6C2W):

****

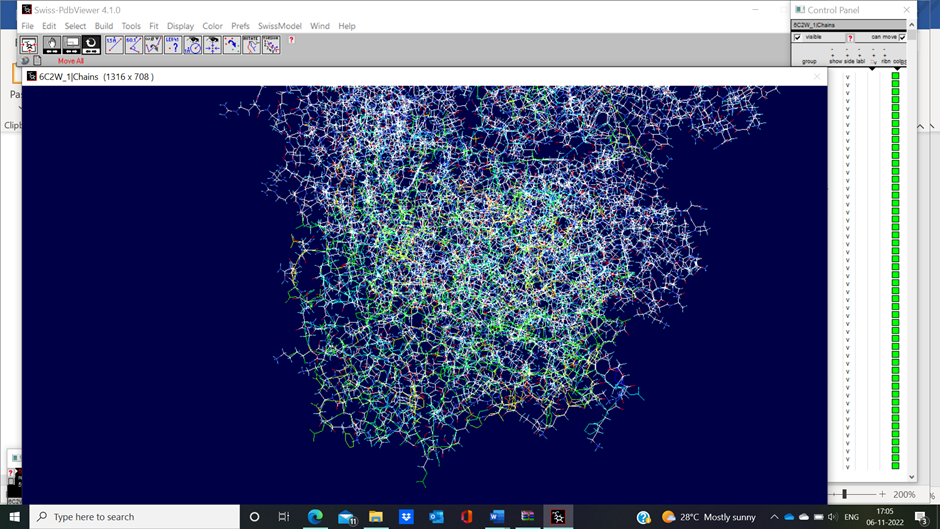
* Open 5YGS Sequence (Death effector domain-containing protein),
* Align the 6C2W against the structural alignment:



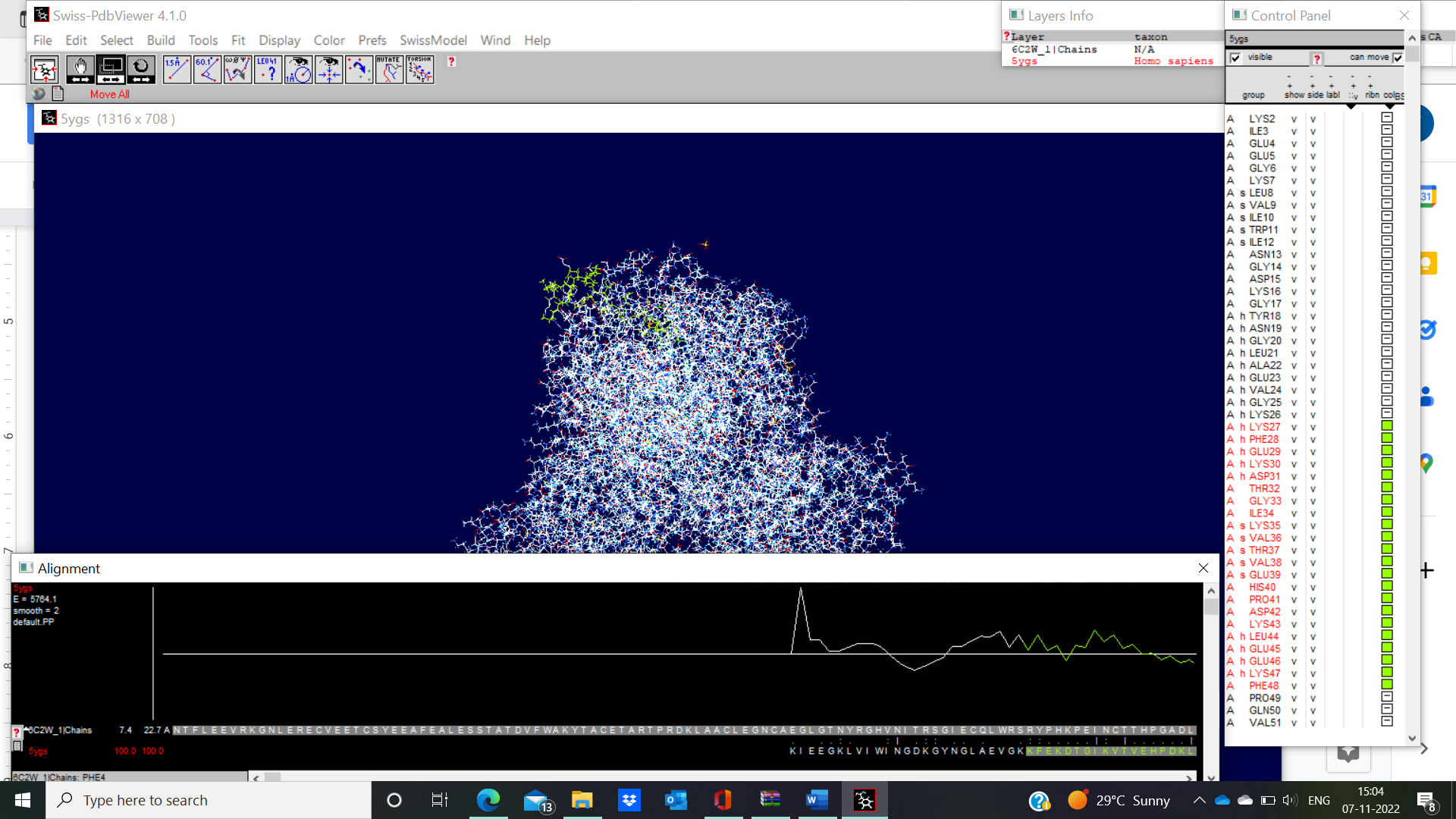
* Fit raw sequence:



* 5YGS sequence is colored:



* The small white arrow next to red question mark was selected to display the individual threading energy for every residue in the 6C2W sequence.
* The total threading energy is given in the top left corner of the alignment window.
* Smoothing factor=2



The two proteins prothrombin(PDB ID: 6C2W) and death effector domain-containing protein (PDB ID:5YGS) exhibit very little similarity with each other.

**INTERPRETATION/ CONCLUSION:**

* **Human prothrombin** (S10C/A470C) is an α-globulin protein present in blood plasma and encoded by the F2 gene. It is a mutation of the original human prothrombin molecule which consists of **622 amino acids.** It has a sequence length of 582 and a molecular weight of 134.12 kDa. P Its structure consists of 2 main chains and 2 attached ligands namely 2-acetamido-2-deoxy-beta-D-glucopyranose and magnesium ion. Being a glycoprotein it also has an attached carbohydrate which is beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose. Prothrombin is produced in the liver and is distributed throughout the body through the bloodstream. It is one of several factors known together as clotting factors. Prothrombin plays an important role in the blood clotting process. It exists in equilibrium between closed and open conformations, but the physiological role of these forms remains unclear. The open form of prothrombin has been characterized structurally, but little is known about the architecture of the closed form that predominates in solution under physiological conditions. This protein is proteolytically cleaved in multiple steps to form the activated serine protease thrombin. Thrombin, in turn, converts soluble fibrinogen into insoluble fibrin strands which in combination with platelets form a clot as well as catalyzing many other coagulation-related reactions. Prothrombin thus helps in blood hemostasis, inflammation, and wound healing.
* **Human TNFRSF25** ( tumor necrosis factor receptor superfamily 25) protein has a sequence length of **463 amino acids** and is encoded by the gene TNFRSF25. It has a molecular weight of 208.61 kDa. It has 4 main chains and a single sulfate ion as a ligand. This protein is a cell surface receptor of the TNF receptor family which mediates apoptosis ( programmed cell death ) and cell differentiation. The TNF receptor superfamily is a protein superfamily of cytokine receptors that have the ability to bind TNFs via an extracellular cysteine-rich domain. This receptor is expressed primarily by activated T lymphocytes and antigen-experienced T lymphocytes and plays a role in regulating lymphocyte homeostasis. This protein stimulates NF-kappa B ( a protein complex that controls cytokine production and cell survival) activity and regulates cell apoptosis. Since the activation of this protein is T cell receptor-dependent, the activity of TNFRSF25 in vivo is specific to those T cells that are encountering an antigen.
* There is no functional similarity between the two proteins studied here.
* After Pairwise Alignment,

|  |  |
| --- | --- |
| **Length** | 979 |
| **Identity** | 19/979 ( 1.9%) |
| **Similarity** | 34/979 ( 3.5%) |
| **Gaps** | 913/979 (93.3%) |
| **SCORE** | 33.0 |

* From Predict Protein, we came to know that both the proteins belong to mixed classes and does not contain any transmembrane helices.
* We used Pymol to get the structural alignment and we got an RMSD score of 38.8.