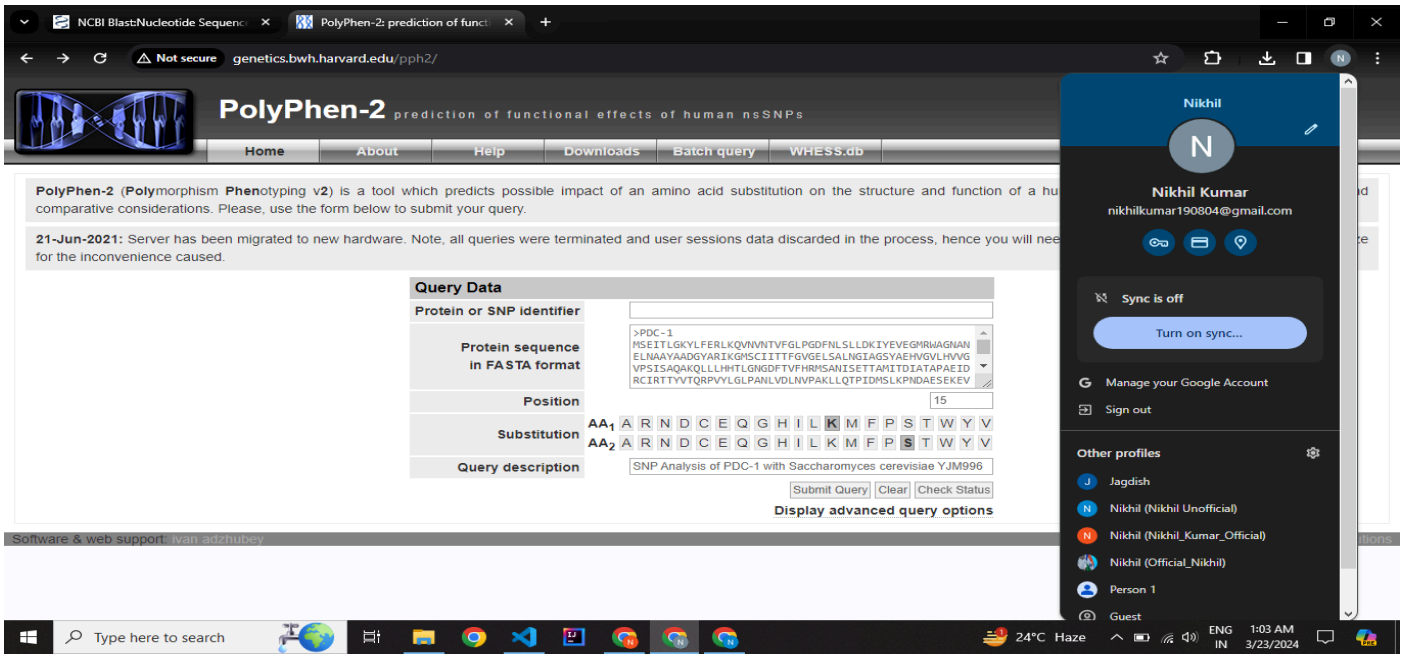
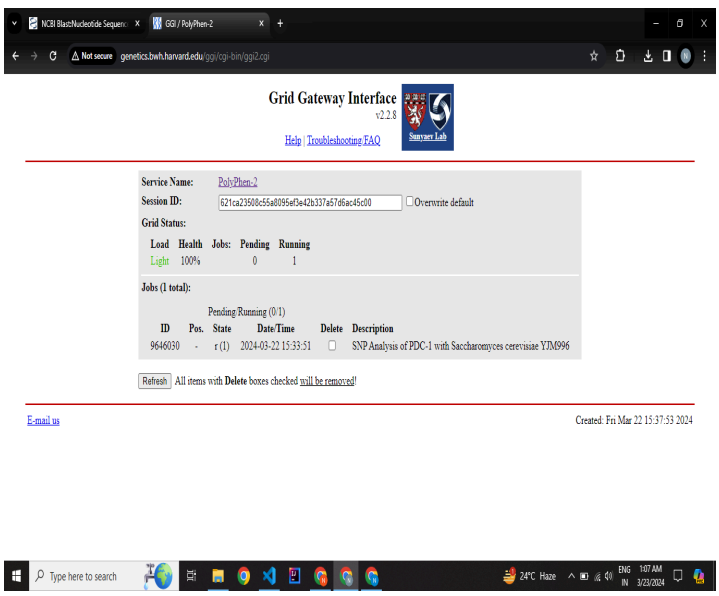
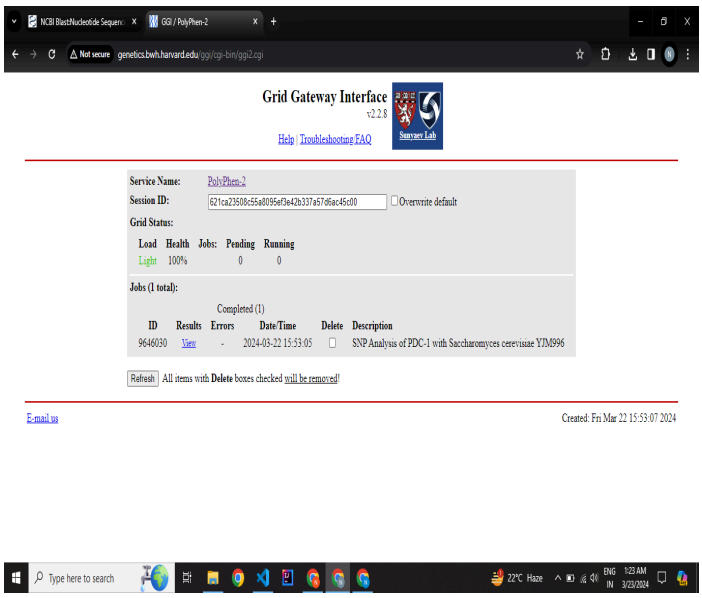
Using PolyPhen To Predict the Impact of the SNP



This is the PolyPhen website interface, which can be accessed via LINK. It takes input a protein sequence in fasta format or Protein/SNP identifier. Then we have to give it the position number (starting from 1) where the amino acid have changed. Along with that we have to select the substitution i.e- which amino acid has changed to which one. And then we have to submit our query and wait for the results.
Now, here i am giving amino acid sequence of PDC-1 as the input sequence in fasta format. Now from the python script : utility.py i found the position of substitution between this PDC gene and Saccharomyces Cerevisiae YJM996 strain which gave me that on position 15 there's a change of K to S between the PDC-1 amino acid sequence and that strain amino acid sequence. And hence i gave the position value as 15 and selected the appropriate substitution i.e- K to S. After that i gave the query description and then submitted the query.





After waiting for some time and refreshing this several times, the results were ready to be analyzed.
Left image shows the submitted query. Right image shows the page when results were ready to be displayed.

Results of the PolyPhen:-

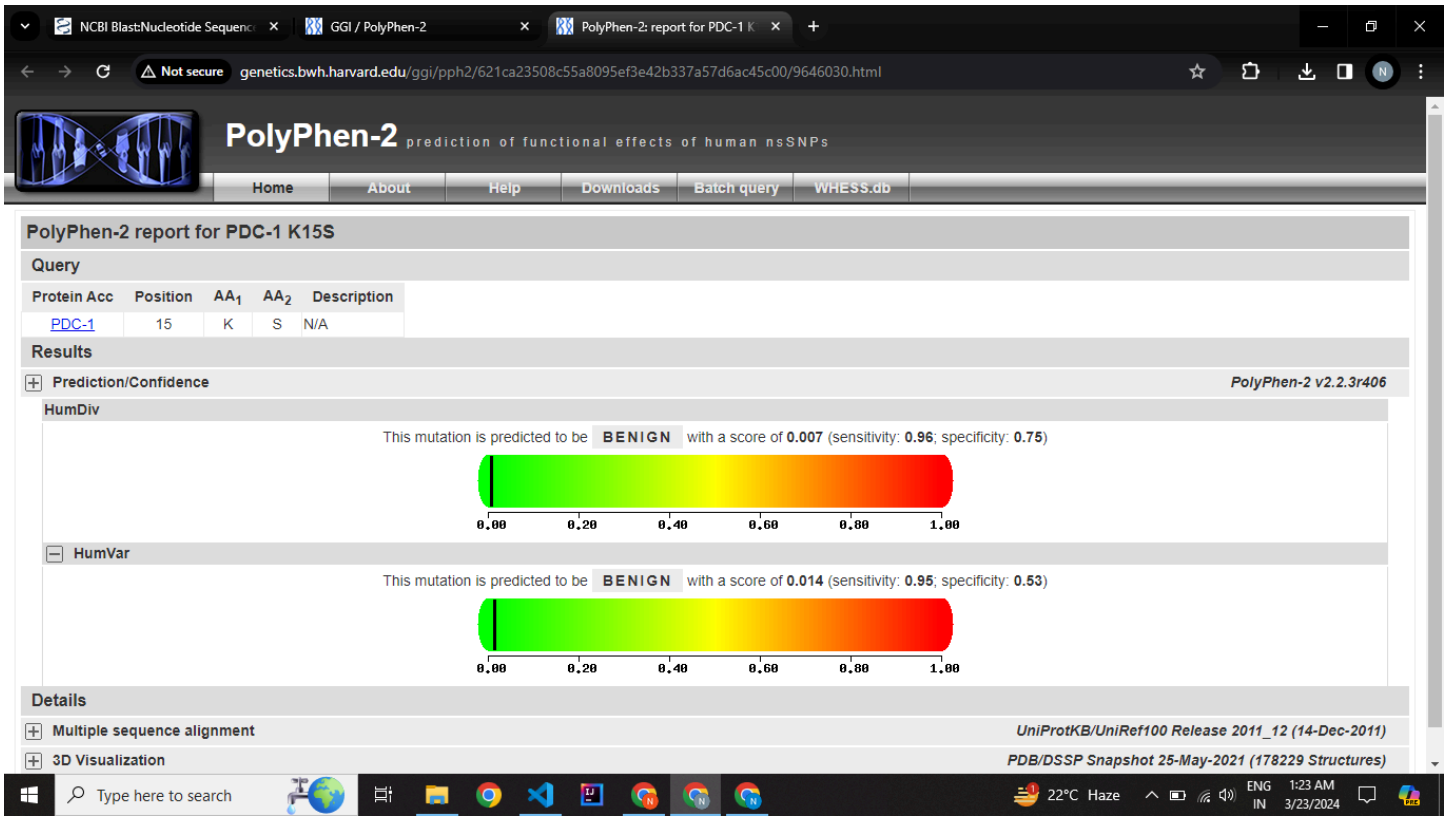


Image1: Displaying the Prediction/Confidence

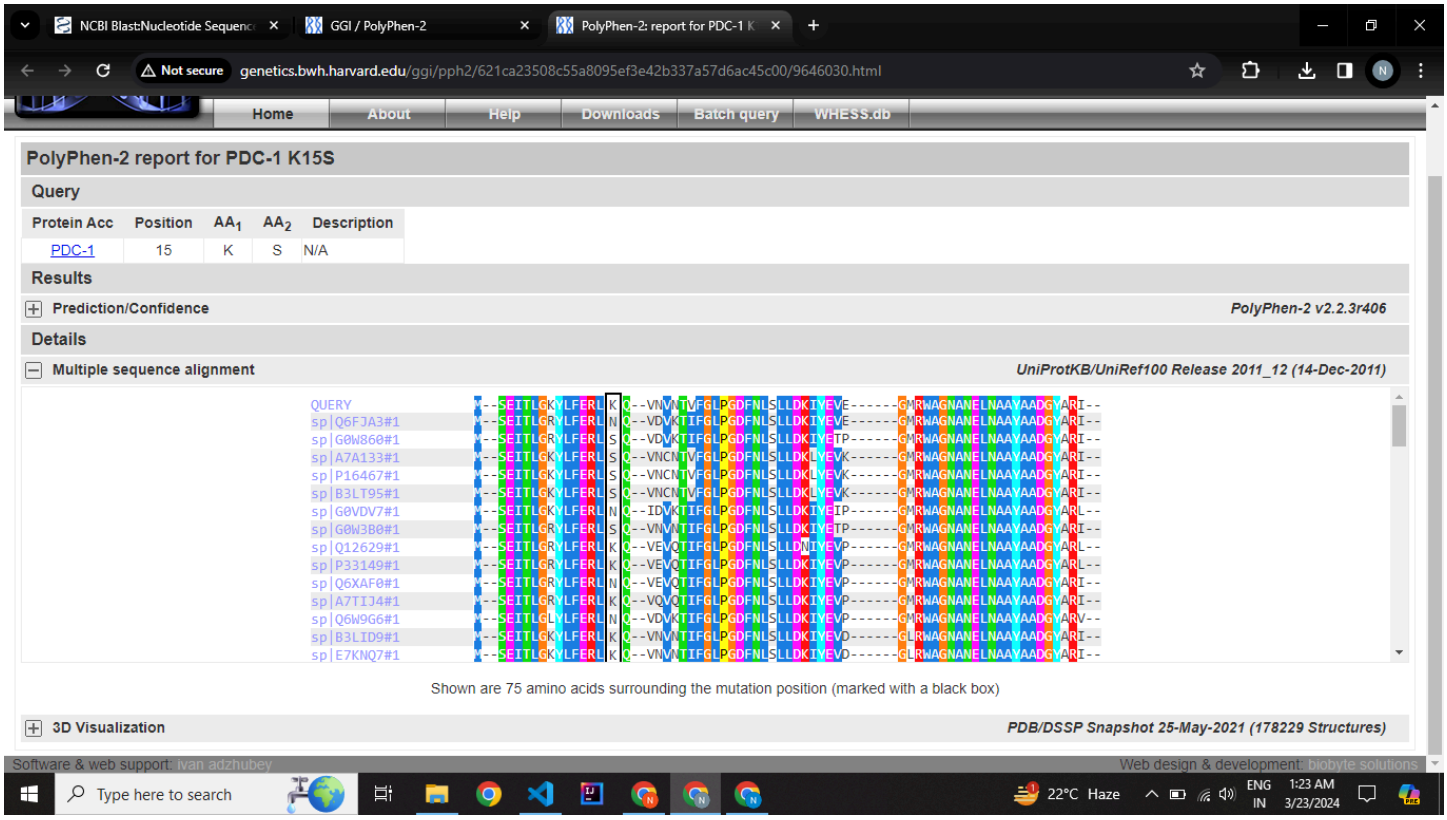


Image2: Displaying the Multiple Sequence Alignment

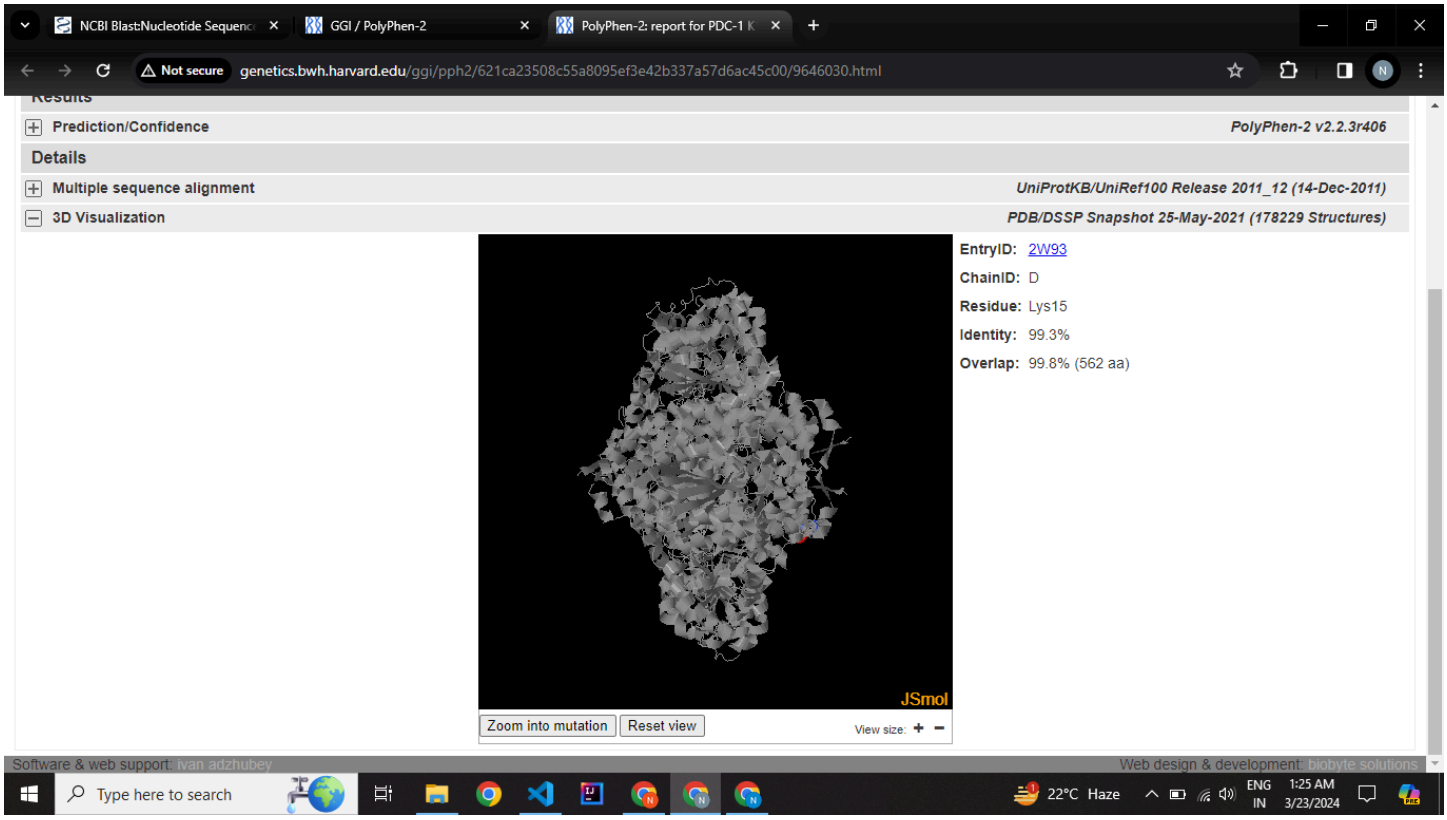


Image3: Displaying the 3D Visualization of the Protien

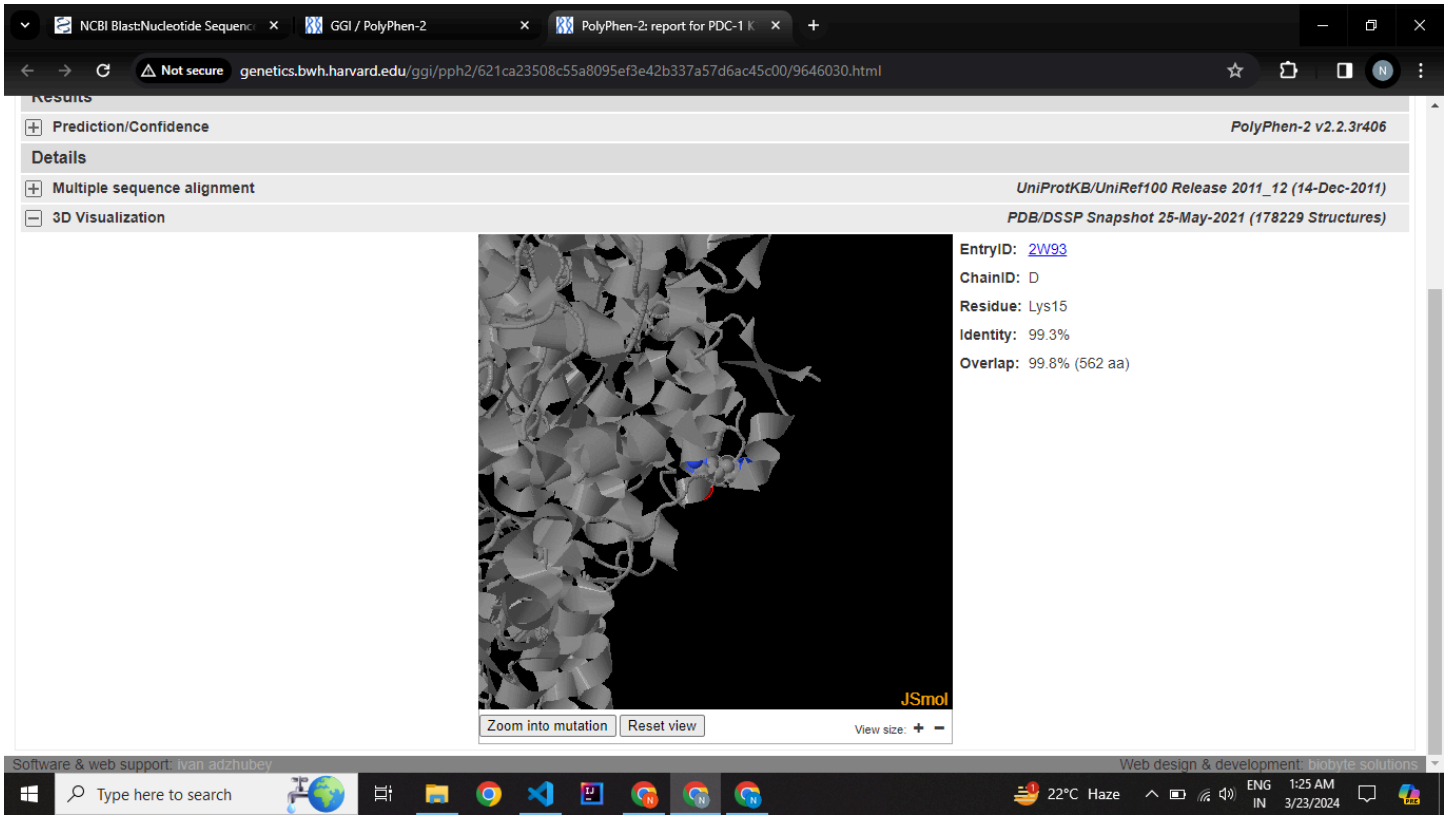


Image4: Displaying the zoomed-in view of the Mutation

Interpretation of the Images

Image1 displays the Prediction of the mutation, which in this case is a “BENIGN”. This indicates that the substitution of K to S will have less impact on the protein function and structure.

Also the score of “BENIGN” is 0.007 (HumanDiv). As this value is closer to 0, it indicates that this substitution is likely to have no effect on the protein. The sensitivity was 0.96, which clearly indicates that PolyPhen was able to detect this mutation in the “BENIGN” category.
There are also some other types of classification of the mutations, such as “Possibly Damaging” and “Probably Damaging”. Mutations under “Possibly Damaging” have a moderate chance of changing the structure and function of the Protein. Mutations under “Probably Damaging” have a high chance of affecting the protein and these can even contribute to various diseases.

Image2 displays the Multiple Sequence Alignment of the amino acid sequences.
And it displays the 75 Amino acids which surround the positon where the mutation has occurred.

Image3 and Image4 display the 3d Model of the protein.

The EntryID shown refers to a unique ID for the protein in the database. And it helps in identifying the protein structure to be analyzed.

The ChainID refers to the chain of amino acids which is being referred here.
Here it is chain D.

Residue indicates the position within the protein sequence where the variant or mutation is located. Here it is Lys15, which indicates that lysine is in position 15 of the amino acid sequence where the mutation has occurred.

Overlap indicates the extent of overlap between the original and the mutated amino acid sequence. Here it is 99.8% which indicates a very good overlap between the two.