Knitr Template

Zachary Foster

May 14, 2014

I decided to explore the structure of 354aa protein of unknown function from *Pelagibacter* strain 7214 using a variety of structural prediction methods. I submitted the unknown sequence to the Phyre2 server using the normal mode. 62 residues (18% of the total sequence) hit a heptosyltransferase with 85.4% confidence (Figure 3). The two next best hits were to a glycosyl transferase and a glycosidase. I also submitted the sequence to SWISS-MODEL using the `Build Model` function. The top hit was to a light-independent protochlorophyllide reductase subunit B, but was not estimated to be significant (Figure 4). Next, the query was submitted to iTASSER. The best hits were to PDB proteins with a hydroxycinnamoyltransferase function. All 10 top hits were some kind of transferase. The best hit had 98% coverage, but only 8% sequence identity (Figure 1). Function Prediction using COFACTOR and binding site prediction suggested similarity to acetyltransferases.

I then searched ProSite for hits to the query using the `ScanProsite` tool, but none were found. PFAM was also searched and only one hit to a protein of unknown function was returned (Figure 5). I attempted to used SSCP (Secondary Structural Content Prediction), but I was unable to access the webpage; all links to it did not work. I searched PDB using the query sequence and no matches were returned. Finally I submitted the PDB file returned by Phyre2 to the Dali server. The top matches were of 13% identity to a lipopolysaccharide heptosyltransferase-1 and the structure seemed to overlap well for a portion of the target (Figure 2)

Overall, the results weakly suggest that the protein is some kind of transferase, but there is little confidence in the prediction. Most of the programs used aligned only 20-40% of the query sequences and had very low sequence identity. There were multiple iTASSER alignments of nearly full length and of acceptable quality, typically to various types of transerases.

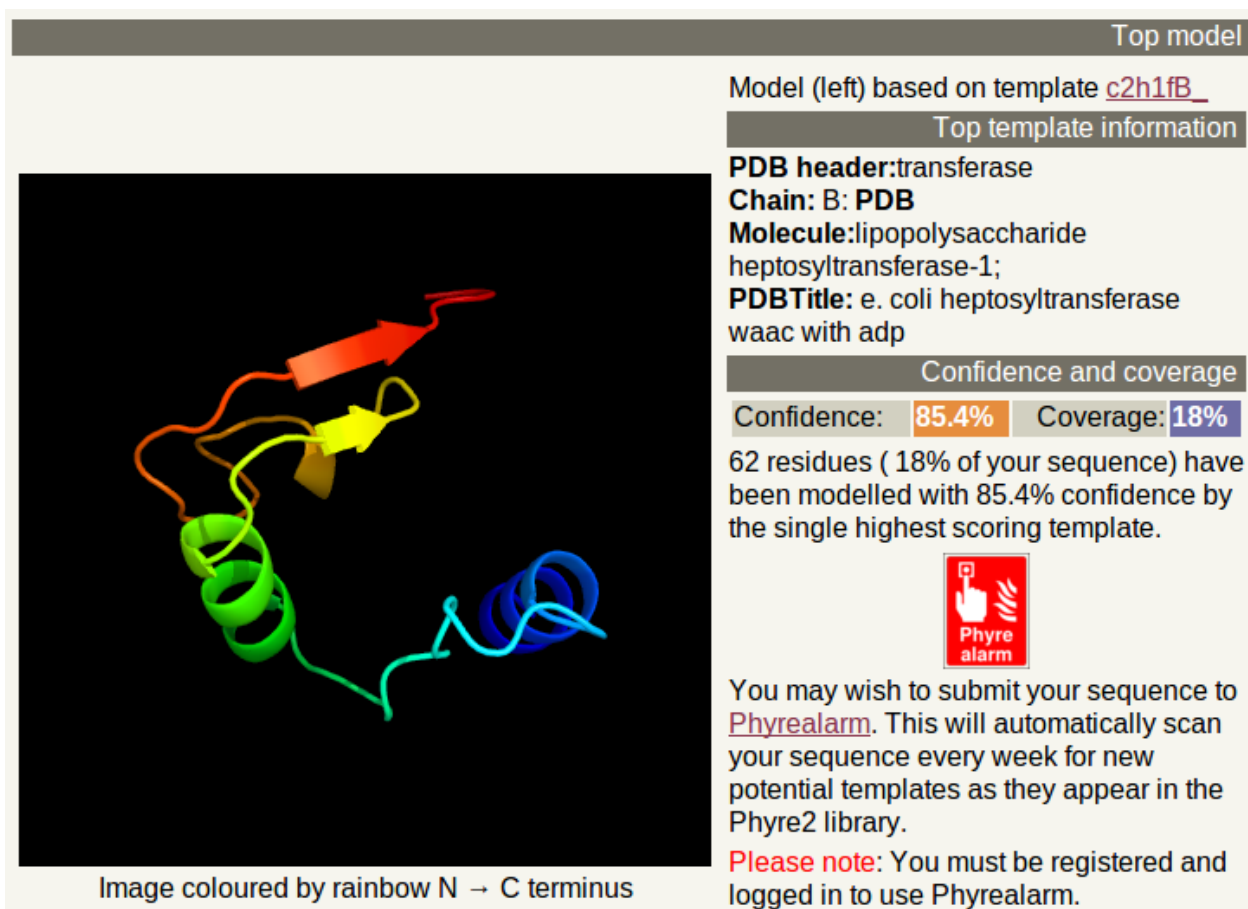Figure 1: Top iTASSER result 3D alignment

Figure 2: Dali Server result

Top model

Model (left) based on template c2h1fB_

Top template information

PDB header:transferase
Chain: B: PDB
Molecule:lipopolysaccharide heptosyltransferase-1;
PDBTitle: e. coli heptosyltransferase waac with adp

Confidence and coverage

Confidence: 85.4%   Coverage: 18%

62 residues (18% of your sequence) have been modelled with 85.4% confidence by the single highest scoring template.

You may wish to submit your sequence to Phyrealarm. This will automatically scan your sequence every week for new potential templates as they appear in the Phyre2 library.

Please note: You must be registered and logged in to use Phyrealarm.

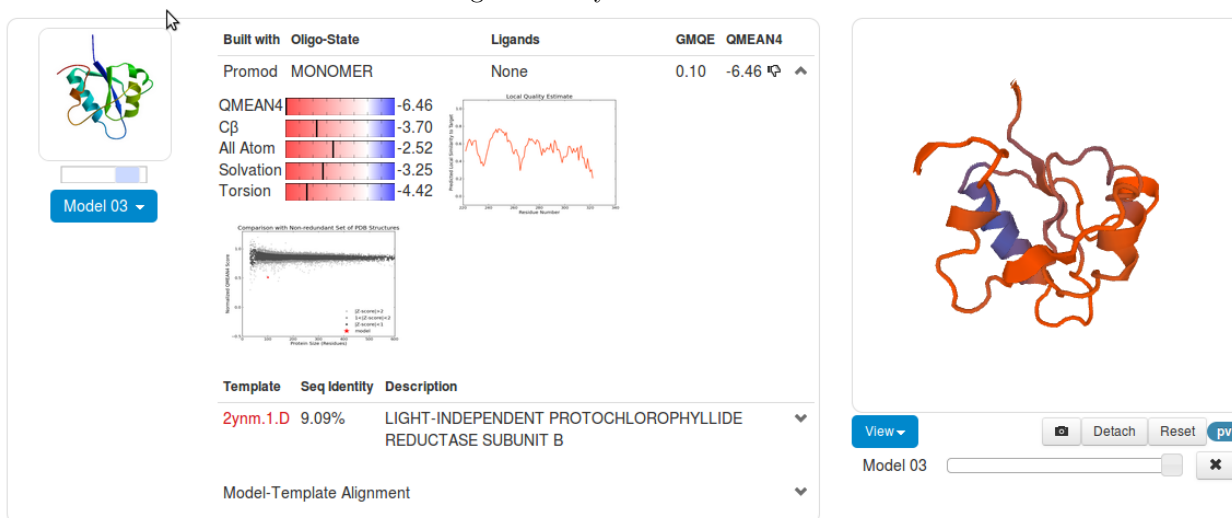Image coloured by rainbow N → C terminus

Figure 3: Phyre2 results

Figure 4: SWISS-MODEL results

Sequence search results

Show the detailed description of this results page.

We found **1** Pfam-A match to your search sequence (**all** significant). You did not choose to search for Pfam-B matches.

DUF563

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
DUF563	Protein of unknown function (DUF563)	Family	n/a	113	329	114	323	**2**	**201**	207	117.0	9.7e-34	n/a	Show

Figure 5: PFAM result