DATE – A Database of TIM Barrel Enzymes

2.1 Introduction ........................................................................................................... 25
2.2 Objective and salient features of the database .............................................. 26
   2.2.1 Choice of the dataset ......................................................................................... 26
2.3 Statistical information on the database ............................................................... 27
2.4 Features .................................................................................................................. 28
   2.4.1 Structure and the Topology diagram of the protein ........................................ 28
   2.4.2 Other informations ......................................................................................... 29
   2.4.3 Conformational information ........................................................................... 29
   2.4.4 BLAST output for the particular entry .......................................................... 30
2.5 Availability, Update and the structure of the website ........................................ 31
2.6 References ............................................................................................................. 32
2.1 Introduction:

The eight-stranded alpha/beta barrel (TIM barrel) is by far the most common tertiary fold observed in high-resolution protein crystal structures. It is estimated that 10% of all known enzymes have this domain. The members of this large family of proteins catalyze very different reactions. Such diversity in function has made this family an attractive target for protein engineering. Moreover, the evolutionary history of this protein family has been the subject of rigorous debate. Arguments have been made in favor of both convergent and divergent evolution. Because of the lack of sequence homology, the ancestry of this molecule is still a mystery. In this study, an analysis has been attempted on proteins which were found to have the alpha/beta structural motif to study their structural features like conformational preferences, functional significance, topological features such as solvent accessibility, residue preference, salt bridges, sequence similarity etc. This database is a collection of sequence, structural, functional and conformational information about all enzymes that have the TIM barrel (alpha8/beta8) topology (where the beta strand order is parallel:1,2,3,4,5,6,7,8). The TIM database currently contains 84 different enzymes.

A lot of work has been done on the TIM barrel proteins. People have made attempts to classify TIM barrel protein based on evolutionary studies, and have also come up with some interesting results like the ‘Backward evolution model’ is not preferred and that modern proteins have arisen from common ancestors that bound key metabolites. This has been studied extensively by Peer Bork (2000) by mapping the known TIM barrel folds to the pathways of central metabolism. Another interesting observation that needs to be mentioned is that all the TIM barrel proteins are made up of approximately 250 residues, with a deviation of 10 residues. Though some examples may have 500 residues in this dataset, when one considers only the ‘curated’ protein, the number of residues falls down to approximately 250 residues.

![Fig. 1: Snapshot of the DATE website available at http://144.16.74.43/~skumar/date](http://144.16.74.43/~skumar/date)
2.2 Objective and salient features of the database

\textbf{DATE} is an acronym for \textbf{D}atabase of \textbf{T}IM barrel \textbf{e}nzymes. The database was created with the objective to help people working with TIM barrel proteins, get quick and comprehensive information about the protein of interest. This is different from the TIM-DB which is maintained by Dr. Pujadas (1999), where he has given links to the PDB website, SCOP and CATH. He has also classified the known TIM barrel proteins according to the E.C number. It is interesting to note from our analysis and their analysis that there are no examples for the class Ligase. This database is not a general database but a specially database as described by Helen Berman (1999). She had defined a speciality database as follows:

\textquote{Another type of database that has proven invaluable in research has been the speciality database. These databases are curated by people working experts in the field and provide information beyond the structure themselves. These may include derived structural data, sequence information, and other biochemical information.”}

\subsection*{2.2.1 Choice of the dataset}

The information available in the database is briefly discussed in this section. When the analysis on TIM barrel protein was carried out, some of the observations were very striking. Most of the helices terminated with either glycines or with asparagines in the left handed helical conformation. The dataset was characterized by looking at the phi-psi values of the proteins taken. The set of proteins, which were actually taken for the study, was selected as follows. The homology between the proteins is less than 25%. The resolution at which the protein was used was given as another criterion. Here, we had used a resolution cut-off of 2.5 angstrom. The residue frequency to occur at the helix region and to occur at the strand region was calculated and it was found that Ala and Leu had the maximum frequency to occur in helices. The residues with the maximum frequency to occur in strands include Val and Leu. The residue, which terminated the helix in the left-handed helical conformation, was Glycine, followed on by Asn.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig2.png}
\caption{Choice of the dataset}
\end{figure}
2.3 Statistical information on the database

There are 85 enzymes in the database with 22,062 residues. The propensity for the helix to terminate in the left handed helical conformation was the maximum for Glycine, followed by Asparagine and histidine. Some of the statistics of the database like, residue preference in the helix, strand, helix termination with positive phi and psi, etc are shown in the figures below.

Fig. 3: Conformational characterization of the residues (non-gly) in the dataset and the composition of the dataset.

Fig. 4: Residue frequency to occur in the Helix region and the strand region for the residues in the database.
2.4 Features

Some of the salient features in this database are discussed below.

2.4.1 Structure and Topology diagram of the Protein

The structure of each of the protein in the database is represented in a ribbon diagram generated by Molscript written by Kraulis. The topology diagram of each of the enzymes is also depicted as a cartoon, with the cylinders representing the helix and the arrowmarks representing the strand. The arrowhead points towards the C-terminus of the protein. Incase where there are large extra domains in the proteins, they are not represented by cylinders and arrowmarks, instead, they are marked to be extra domains in the protein. The crescent represents the loops, which connect the helix and the strand.

**Triosephosphate Isomerase**

![Triosephosphate Isomerase](image)

**Fig. 5**: One half of the webpage for the entry Triosephosphate Isomerase
2.4 Features

2.4.2 Other informations

Other than the structure and the topology diagram, information on the function, which this enzyme performs, multimeric unit of this protein, number of residues in this protein and the pathway in which this is involved is also given. Other than these, the link to the PDB entry is also given. All these are shown in figure 5.

2.4.3 Conformational information

Another link by the name “Ramachandran Phi Psi Values” is provided for each of the entry in the database. This links to a file, which contains complete information on the phi psi values of the protein. The representation and the conventions used in making the file are also defined in each page. For reference, pls see figure 6.

Fig. 6: This shows the file for which the link labelled “Ramachandran phi psi values” is given. This file gives the backbone dihedral values for the residues in the protein. The format of the file and the conventions used in explaining is also described in the header of each of the file.
2.4.4 BLAST Output for the particular entry

Other than these structural informations, sequence information is also provided. For each of the entry, the BLAST (Basic Local Alignment Search Tool) result is displayed. This gives the global sequence alignment of the proteins from different organisms. See figure 7.

Fig. 7: shows the page containing the BLAST output for that particular entry.
2.5 Availability, update and the structure of the website

DATE is currently updated once every 3 months. Researchers who are interested to use DATE may directly download DATE as a zip file at http://144.16.74.43/~skumar/date/date.zip or as a tar file at http://144.16.74.43/~skumar/date/date.tar. Researchers who are interested to browse the contents of the database can visit DATE at http://www.144.16.74.43/~skumar/date. Another mirror site is also hosted at http://www.geocities.com/date and at http://www.geocities.com/madanm2/date. Incase of any mistakes, please email to M. Madan Babu (mmadanbabu@rediffmail.com) or S. Kumar Singh (skumar@mbu.iisc.ernet.in).

Fig. 8: Structure of the website
2.6 References


