

Dihedral angle and secondary structure database of short amino acid fragments

Saravanan Dayalan*, Nalaka Dilshan Gooneratne, Savitri Bevinakoppa, Heiko Schroder

School of Computer Science and Information Technology, RMIT University,
GPO Box 2474V, Melbourne 3001, Australia; Saravanan Dayalan* - Email: sdayalan@cs.rmit.edu.au;

* Corresponding author

received November 25, 2005; revised December 18, 2005; accepted December 23, 2005;

published online January 1, 2006

Abstract:

Dihedral angles of amino acids are of considerable importance in protein tertiary structure prediction as they define the backbone of a protein and hence almost define the protein's entire conformation. Most *ab initio* protein structure prediction methods predict the secondary structure of a protein before predicting the tertiary structure because three-dimensional fold consists of repeating units of secondary structures. Hence, both dihedral angles and secondary structures are important in tertiary structure prediction of proteins. Here we describe a database called DASSD (Dihedral Angle and Secondary Structure Database of Short Amino acid Fragments) that contains dihedral angle values and secondary structure details of short amino acid fragments of lengths 1, 3 and 5. Information stored in this database was extracted from a set of 5,227 non-redundant high resolution (less than 2-angstroms) protein structures. In total, DASSD stores details for about 733,000 fragments. This database finds application in the development of *ab initio* protein structure prediction methods using fragment libraries and fragment assembly techniques. It is also useful in protein secondary structure prediction.

Availability: DASSD can be accessed and downloaded from <http://www.cs.rmit.edu.au/dassd/>

Keywords: proteins; dihedral angles; secondary structure; tertiary structure; fragments

Background:

Proteins are macromolecules that play a very important role in the functioning of living organisms. Proteins are made of 20 different types of amino acids formed in different combinations. Proteins are active only in their native or folded state and prediction of protein structure from sequence is challenging. [1] Each amino acid residue in a protein constitutes the phi and psi dihedral angles. Dihedral angles are of considerable importance in protein structure prediction as they define the backbone of a protein, which together with side chains define the entire protein conformation. [2] The three-dimensional tertiary structure of a protein consists of repeating units of secondary structures. The two major types of secondary structures are alpha helix and beta strands. Most *ab initio* protein structure prediction methods predict secondary structural

elements as a starting point to predict the tertiary structure of proteins. [3-5] Hence, both dihedral angles and secondary structure details of proteins are important in *ab initio* protein tertiary structure prediction.

The calculation of dihedral angle values from each known structure downloaded from PDB (protein databank) is trivial. However, calculation of all occurring dihedral angle values of a short fragment of amino acids for all known structures available at the PDB is computationally intensive. A protein structure prediction program using fragment assembly techniques would initiate hundreds of queries to obtain dihedral angle values of different short fragments of amino acids. Thus, a database of dihedral angles of short fragments is critical.

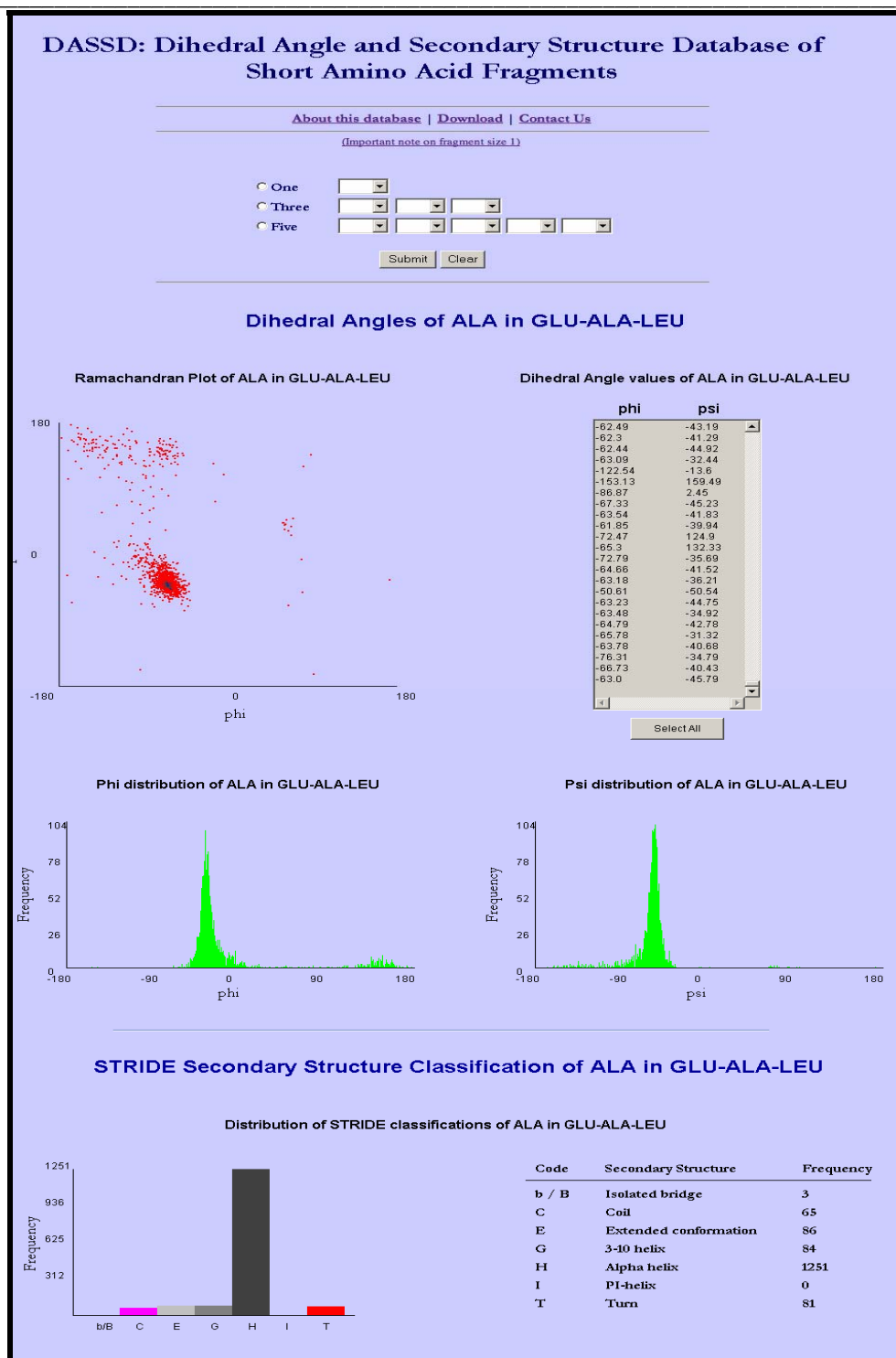


Figure 1: Screen shot of DASSD showing the dihedral angles distribution and the STRIDE secondary structure classifications for the amino acid fragment GLU-ALA-LEU

Methodology of development:

This article describes the dihedral angle and secondary structure database (DASSD) of short amino acid peptide fragments of lengths 1, 3 and 5 derived from PDB (protein databank). For each fragment, all occurring dihedral angles of the middle amino acid residue and its STRIDE secondary structure classification [6] were extracted from a set of 5,227 non-redundant high resolution (less than 2-angstroms) protein structures. The non-redundant dataset (cullpdb_pc90_res2.0_R0.25) was obtained from the PISCES protein sequence culling server for removing redundancy. [8] Even though there are many protein secondary structure assignment programs as described elsewhere [7], we chose the most widely used STRIDE for this analysis [7] and it is shown to perform better than DSSP. [6]

DASSD contains information for dihedral angles (phi and psi), phi distribution, psi distribution, STRIDE secondary structure classification and Ramachandran plot of the middle residues in about 733,000 different amino acid peptide fragments. Information stored in this database would assist in *ab initio* protein tertiary structure prediction methods using fragment libraries and fragment assembly techniques. It is also useful in protein secondary structure prediction. Information in DASSD is stored as flat files. PHP is used to extract information for each query from the database. The extracted results from DASSD are displayed using a Java Applet as shown in Figure 1.

Utility to the biological community:

One of the methods used in many *ab initio* structure prediction methods is to use secondary structure prediction as a start point to predict the tertiary structure and then use fragment assembly techniques where a library of fragments is generated from which the protein's tertiary structure is built. [3,4,5,9,10] The proposed database of dihedral angles and secondary structures of short amino acid fragments would assist such *ab initio* methods as fragment libraries or in building new fragment libraries. The database can also be used to analyze the dihedral angles and secondary structure properties of an amino acid in relation to other amino acids. Such analysis would help in assigning

structures for amino acids in tertiary structure prediction and also in protein loop structure prediction. The secondary structures of fragments in DASSD can also be used in secondary structure prediction.

Future development:

The current database contains information for dihedral and secondary structure details of peptide fragment length ≤ 5 residues. We plan to update and improve the database using a more representative dataset of structures for peptide fragment length > 5 residues. Visual comparison of dihedral angles distribution and secondary structure of different fragments will be facilitated in future developments.

Acknowledgement:

We thank Victorian Partnership for Advanced Computing (VPAC), Melbourne, Australia for providing access to their parallel computing facility. We also thank Nathan Hall, The Ludwig Institute for Cancer Research, Melbourne, Australia, for his valuable advice.

References:

- [1] K. Ginalski, *et al.*, *Nucleic Acids Res.*, 33:1874 (2005) [PMID: 15805122]
- [2] L. Esposito, *et al.*, *J Mol Biol.*, 347:483 (2005) [PMID: 15755444]
- [3] C. A. Rohl, *et al.*, *Methods Enzymol.*, 383:66 (2004) [PMID: 15063647]
- [4] J. Lee, *et al.*, *Biophys Chem.*, 115:209 (2005) [PMID: 15752606]
- [5] J. Lee, *et al.*, *Proteins*, 56:704 (2004) [PMID: 15281124]
- [6] D. Frishman & P. Argos, *Proteins*, 23:566 (1995) [PMID: 8749853]
- [7] J. Martin, *et al.*, *BMC Struct Biol.*, 15:17 (2005) [PMID: 16164759]
- [8] G. Wang & R. L. Dunbrack, *Bioinformatics*, 19:1589 (2003) [PMID: 12912846]
- [9] D. T. Jones, *et al.*, *Proteins*, [Epub ahead of print] (2005) [PMID: 16187356]
- [10] K. Karplus, *et al.*, *Proteins*, [Epub ahead of print] (2005) [PMID: 16187355]

Edited by Darren R. Flower

Citation: Dayalan *et al.*, *Bioinformatics* 1(3): 78-80 (2006)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.