

## CADB-2.0: Conformation Angles Database

K. Samaya Mohan,<sup>a</sup> S. S. Sheik,<sup>a</sup>  
J. Ramesh,<sup>a</sup> B. Balamurugan,<sup>a</sup>  
M. Jeyasimhan,<sup>a</sup> C. Mayilarasi<sup>‡</sup>  
and K. Sekar<sup>a,b\*</sup>

<sup>a</sup>Bioinformatics Centre, Indian Institute of  
Science, Bangalore 560 012, India, and

<sup>b</sup>Supercomputer Education and Research  
Centre, Indian Institute of Science,  
Bangalore 560 012, India

<sup>‡</sup> Summer trainee.

Correspondence e-mail:  
sekar@physics.iisc.ernet.in

Received 10 December 2004

Accepted 23 February 2005

The Conformation Angles Data Base (CADB) is a comprehensive, authoritative and timely knowledge base with a powerful query engine developed to facilitate the retrieval of information related to the conformational angles (main chain and side chain) of the amino-acid residues present in non-redundant (both 25 and 90%) data sets. The updated version has improved options for determining the dependency of the conformation angles of a particular residue upon the flanking residues, doublet analysis, triplet analysis and analysis of a particular protein structure. It is worth mentioning that for all options a user-friendly and convenient Java graphical user interface (GUI) has been provided to display the output on the client machine. The database is accessible at the URL <http://cluster.physics.iisc.ernet.in/cadb/> or <http://144.16.71.148/cadb/>.

## 1. Introduction

The conformation angles taken up by various naturally occurring amino acids in the polypeptide chain of a protein molecule differ significantly from one another for stereochemical reasons (Kleywegt & Jones, 1996). The conformation angles that the main-chain amino acids make with their neighbouring residues play an important role in determining the contribution of amino-acid residues towards the folding of the polypeptide chain. Additionally, the conformation angle of the side-chain atoms seems to influence the tertiary structure of the protein molecules to a great extent. For many years, the Ramachandran plot (Ramachandran *et al.*, 1963) has been a unique representative to infer the behaviour of the main-chain conformation angles of the various amino-acid residues present in a protein molecule. The proposed knowledge base, CADB, acts as a comprehensive source of the main-chain and side-chain conformation angles of all 20 amino-acid residues. To provide a unique knowledge base, the conformation angles are computed for all the protein chains that fall into the 25 and 90% non-redundant data sets (Hobohm & Sander, 1994)

## 2. New features

As a major resource for conformation angles, the primary objective of the database is to provide complete information about all the protein structures classified into the 25 and 90% non-redundant data sets. The present knowledge base contains the conformation angles of about 2.28 million amino-acid residues. The earlier version (Sheik *et al.*, 2003) had options for the calculation of conformation angles (main chain and side chain) and the interrelationship between the main-chain conformation angles. The updated version includes additional features such as (i) main-chain conformation (with the preceding and succeeding residues), (ii) main-chain conformation angle within specified limits, (iii) side-chain conformation angles within specified limits, (iv) doublet analysis, (v) triplet analysis and (vi) analysis of a specific protein structure available in the 25 or 90% non-redundant data set (see Figs. 1 and 2 for details).

The first option (i) is meant to analyze the effect of neighbouring (on either side of the middle residue) residues over the main-chain conformation angle of a particular middle residue opted by the user. This analysis is essential because  $\varphi$  is dependent on the preceding residue and  $\psi$  is dependent on the succeeding residue. However,

## short communications

since the effect is inductive, it is not enough to consider the immediate predecessor and successor alone; the group of residues in either direction of the polypeptide chain must also be considered. The next options (ii) and (iii) allow the user to retrieve the conformation angles (main chain and side chain) of a particular residue whose angles fall within the user-specified range. This enhances the

search for a particular amino acid based on its stereochemical possibilities, which in turn depend on its properties.

The doublet analysis (iv) selects the doublets and plots the  $\varphi$  and  $\psi$  angles that the pair makes with the preceding and the succeeding residues. For example, let  $X_1-X_2$  be the chosen doublet; this option retrieves the  $\varphi$  value of  $X_1$  and the  $\psi$  value of  $X_2$ . This helps in

analyzing the effect of the flanking residues on the specified doublet (the effect of  $X_1$  on  $\varphi$  and  $X_2$  on  $\psi$ ). The triplet analysis (v) option is also included with the same notion but for a set of three amino acids, to study the effect of the flanking residues over the middle residue. So far, all the described options display the necessary angles for a particular amino-acid residue in the non-redundant protein chains. In the new version, we provide an option to view the conformation angles (both main chain and side chain) of a particular protein structure available in the non-redundant data set. The user has the option to see the conformation angles of an entire structure or a particular chain or part of the polypeptide chain. The user again has the freedom to select well resolved and well refined high-quality structures by properly fine-tuning the parameters such as the crystallographic  $R$  factor, resolution and temperature factors.

As pointed out above, several useful options have been provided in the updated version. In addition, some of the options are improved or removed for better robustness and compatibility to make this particular query engine unique. For example, the present version enables the user to distinguish cysteines based on their oxidized or reduced states. In the first version, there were two options for the selection of protein structures solved using the nuclear magnetic resonance (NMR) technique; namely, first model and average structure. Since the average structure did not prove to be reliable enough, it has been removed in the updated version. The current version of the query engine and the knowledge base, CADB (v.2.0), has 8595 chains from 7836 protein structures in the 90% non-redundant data set and 2485 chains from 2379 structures in the 25% non-redundant data set.

### 3. Case study

A sample output of a typical search looking for the influence of the neighbouring residues on the conformation angles (Ala-Ala-Ala) is shown in Fig. 1, which shows the effect of the flanking residues over the middle residue Ala. The triplet Ala-Ala-Ala is taken from 90% non-homologous protein chains (resolution better than or equal to 1.5 Å and crystallographic  $R$  factor better

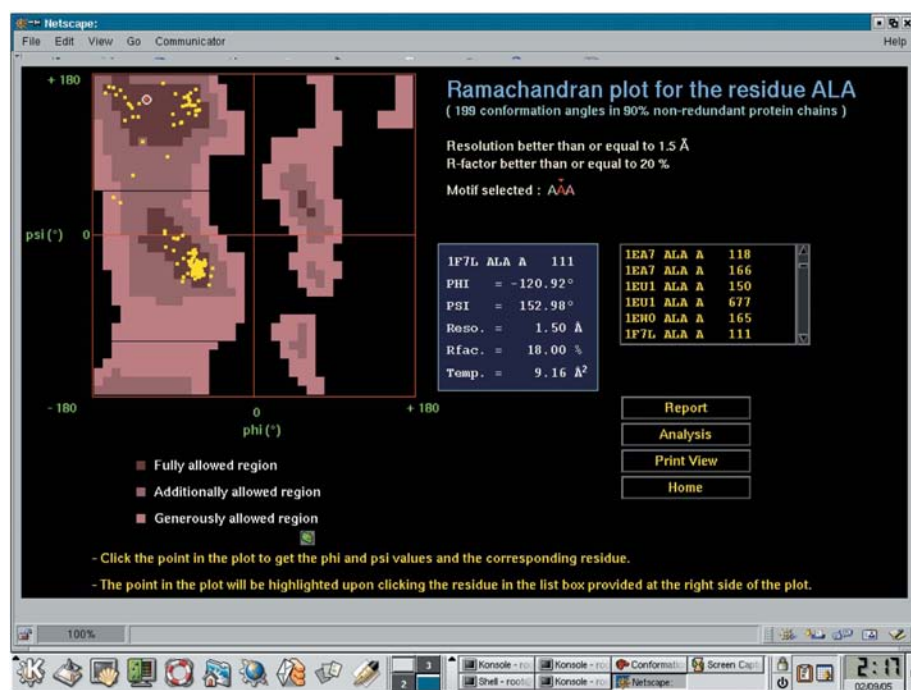


Figure 1

An output frame depicting the effect of the flanking residues on the middle residue of the triplet Ala-Ala-Ala in all protein chains falling into the 90% redundant data set. The highlighted point in the plot corresponds to the residue AlaA111 in PDB code 1f7l. The cutoff criterion used is also shown in the plot.

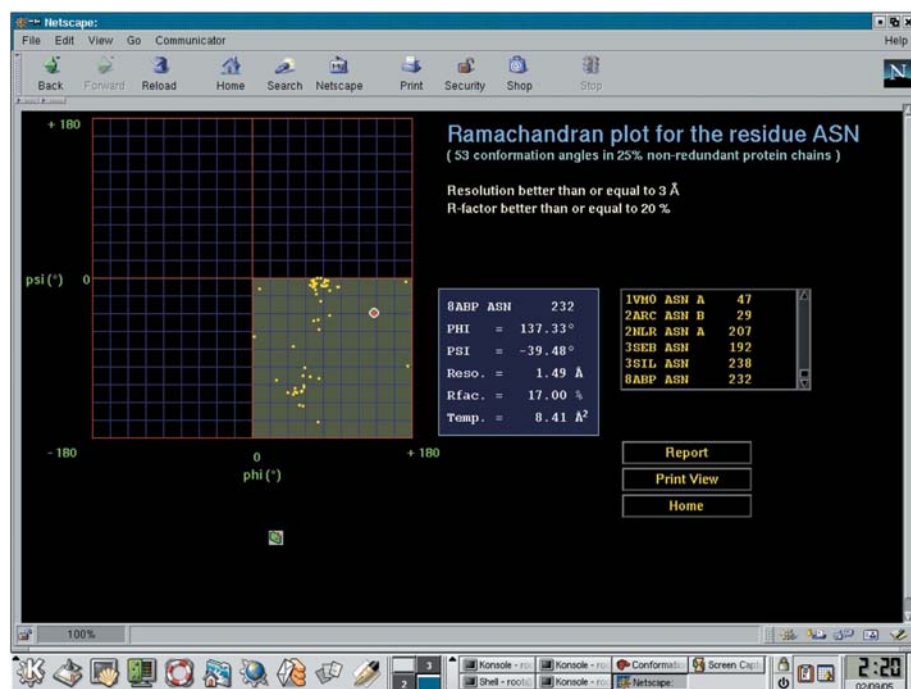


Figure 2

A sample output showing the  $(\varphi, \psi)$  plot (for a particular region of the plot) for the residue asparagine in all protein chains from the 25% non-redundant data set. The highlighted point in the plot corresponds to the residue Asn232 in PDB code 8abp.

than or equal to 20%). Based on the above criteria, a total of 199 conformation angles are plotted in the Ramachandran diagram (Fig. 1). Of these, 156 conformation angles (78%) are in the right-handed  $\alpha$ -helix region and the remaining (22%) angles are in the  $\beta$ -sheet region. This study is performed by clicking the link 'main-chain with preceding and succeeding residues'. Fig. 2 shows the conformation angles of the residue Asn in a particular region of the Ramachandran plot. The limits used here are 0–180° in the case of the angle  $\varphi$  and 0 to –180° in the case of  $\psi$ . As a result, 53 main-chain conformation angles (resolution better than or equal to 3.0 Å and crystallographic  $R$  factor better than or equal to 20%) in 25% non-redundant protein chains lie completely outside the allowed Ramachandran plot. It is interesting to see that most of the outliers are in a narrow range ( $\varphi = 40$ –80° and  $\psi = 0$  to –180°).

#### 4. Conclusions

The updated CADB knowledge base provides unique information about the main-chain and the side-chain conformation angles of all amino-acid residues present in non-redundant protein chains (25 and

90%). The enriched knowledge base will be very useful for those interested in molecular modelling and structural bioinformatics.

The knowledge base and the query engine were developed and are maintained at the Indian Institute of Science, Bangalore 560 012, India with financial support from the Department of Biotechnology, Government of India. The authors gratefully acknowledge the use of the Bioinformatics Centre, the Interactive Graphics-based Molecular Modelling and the Supercomputer Education and Research Centre. We are grateful for individual project support from the DBT (KS). One of the authors (CM) thanks KS for providing an opportunity to undergo summer internship during 2004.

#### References

- Hobohm, U. & Sander, C. (1994). *Protein Sci.* **3**, 522–524.
- Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
- Sheik, S. S., Anathlakshmi, P., Ramya Bhargavi, G. & Sekar, K. (2003). *Nucleic Acids Res.* **31**, 448–451.