STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins

Matthias Heinig and Dmitrij Frishman*

Department of Genome Oriented Bioinformatics, Technical University of Munich, Wissenschaftszentrum Weihenstephan, 85354 Freising, Germany

ABSTRACT

STRIDE is a software tool for secondary structure assignment from atomic resolution protein structures. It implements a knowledge-based algorithm that makes combined use of hydrogen bond energy and statistically derived backbone torsional angle information and is optimized to return resulting assignments in maximal agreement with crystallographers' designations. The STRIDE web server provides access to this tool and allows visualization of the secondary structure, as well as contact and Ramachandran maps for any file uploaded by the user with atomic coordinates in the Protein Data Bank (PDB) format. A searchable database of STRIDE assignments for the latest PDB release is also provided. The STRIDE server is accessible from http://webclu.bio.wzw.tum.de/stride/.

INTRODUCTION

Identification of secondary structure elements is a major step in the characterization of a newly determined protein structure. It serves as a basis for virtually all subsequent analyses, including visualization, structure comparison and classification, homology modelling, threading and sequence alignment. To a large extent, our visual notion of proteins is based on cartoon diagrams showing α-helices and β-strands as cylinders and arrows, respectively.

Several automatic tools for secondary structure assignment from known atomic coordinates are available [reviewed in (1)]. The most widely used method, DSSP (2), defines secondary structure elements as repeating elementary hydrogen bonded patterns. Hydrogen bonds between peptide units are assigned if the electrostatic interaction energy between C=O of one residue and N–H of another residue is <-0.5 kcal/mole. The DEFINE algorithm (3) compares inter-atomic distance matrices of structural fragments to idealized reference distance masks typical for a particular secondary structure type, while P-Curve (4) is based on quantification of backbone curvature using differential geometry. More recently, an improved version of DSSP, called DSSPcont, has been developed which takes into account the structural variations in proteins (5).

Our method, STRIDE (6), was developed with a specific goal to accurately reproduce secondary structure designations created by human experts. It is thus a knowledge-based approach which uses, as training data, a carefully verified set of secondary structural elements defined by crystallographers who have deposited structures in the Protein Data Bank (PDB) (7). The main difference between STRIDE and DSSP is that STRIDE considers both hydrogen bonding patterns and backbone geometry. The hydrogen bond energy is calculated using an empirical energy function which takes into account the distance between the donor and the acceptor and the deviations from linearity of the bond angles (8,9). A weighted product of hydrogen bond energy and torsion angle probabilities for α-helix and β-sheet is used to determine the start and stop positions of secondary structure elements based on empirically optimized recognition thresholds.

The source code of STRIDE has been freely accessible from the FTP server of the European Bioinformatics Institute since 1995 (ftp://ftp.ebi.ac.uk/pub/software/unix/stride/src). It is also available as part of several molecular graphics packages and websites [e.g. Visual Molecular Dynamics (VMD) (10)]. Here, we report a dedicated STRIDE web server and a database of secondary structure assignments.

STRIDE WEB SERVER AND DATABASE

The STRIDE web server, written in the python programming language, makes accessible all functions implemented in the STRIDE software and also provides several additional visualization tools (Figure 1). It accepts as input atomic coordinates in standard PDB format, which can be either uploaded or

*To whom correspondence should be addressed. Tel: +49 8161 712134; Fax: +49 8161 712186; Email: d.frishman@wzw.tum.de
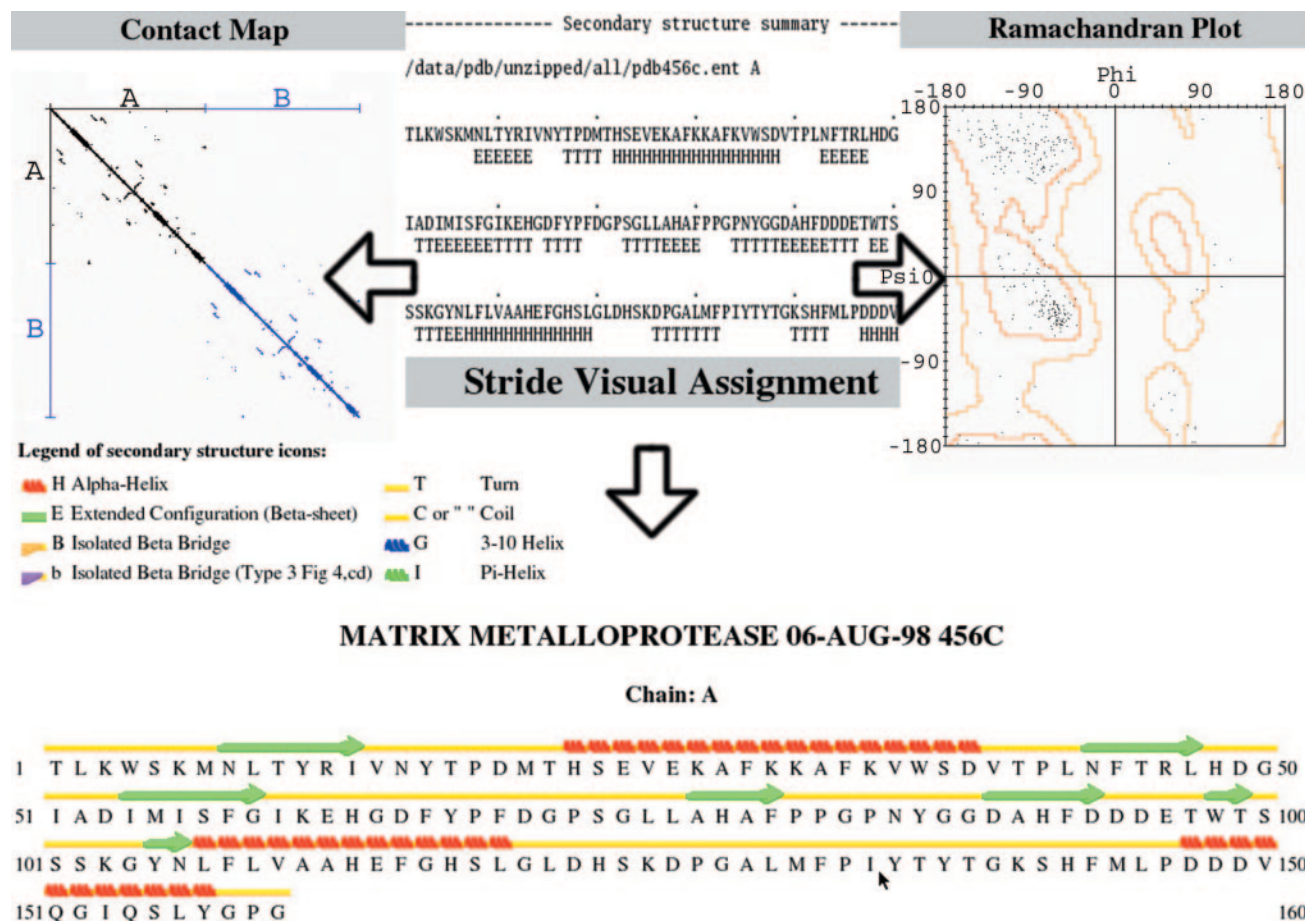
Figure 1. Available views of the secondary structure assignment created by STRIDE from the sample structure 456c. Center top, original STRIDE output; left, contact map; right, Ramachandran plot; bottom, cartoon representation of secondary structure.

pasted into a web form. The STRIDE home page offers the following options.

Basic secondary structure assignment. A secondary structure assignment in text form is produced. Its header section gives general information about the structure (author, compound, etc.) as well as a secondary structure summary which lists locations of identified secondary structure elements. Then follows a detailed per residue assignment of secondary structure states complemented by information on backbone dihedral angles and solvent accessible area computed according to Eisenhaber *et al.* (11,12).

Graphical representation of secondary structure assignment. A cartoon representation similar to the 'wiring diagram' of the PDBsum web server (13) is produced. Technically, this representation is generated not as a rendered image or postscript file, but rather in the form of an html table incorporating both individual graphical items and the protein sequence. The table is constructed by parsing the output of STRIDE and assigning an image to each structural state. An interactive point-and-click interface reveals detailed per residue information.

Contact map. A contact map is derived from a symmetric square matrix of distances between all C-α atoms in a given protein. An interactive mouse-sensitive image indicates distances below a certain threshold defined by the user, typically 6 Å.

Ramachandran plot. No secondary structure web server is complete without a Ramachandran plot (14) representing the distribution of φ and ψ torsion angles in a given protein. The allowed areas for α-helix and β-sheet are shown in the background of the plot. They are taken from a recent update of the Ramachandran map by Lovell *et al.* (15). The orange line is the border for the core region of favourable angles. Gold limits the region of disfavoured but allowed angles. Like the contact map, this image is also mouse sensitive and gives additional information on the residue that is pointed at.

Database of STRIDE assignments. A complete database of STRIDE secondary structure assignments is calculated from each weekly update of PDB (7). Individual entries can be accessed either by PDB code or through a text search interface allowing for the construction of logically structured queries.

REFERENCES

1. Andersen,C.A.F. and Rost,B. (2003) Secondary structure assignment. *Methods Biochem. Anal.*, **44**, 341–363.
2. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

3. Richards,F.M. and Kundrot,L.E. (1988) Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins*, **3**, 71–84.

4. Sklenar,H., Etchebest,C. and Lavery,R. (1989) Describing protein structure: a general algorithm yielding complete helicoidal parameters and a unique overall axis. *Proteins*, **6**, 46–60.

5. Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.

6. Andersen,C.A.F., Palmer,A.G., Brunak,S. and Rost,B. (2002) Continuum secondary structure captures protein flexibility. *Structure*, **10**, 175–185.

7. Bourne,P.E., Addess,K.J., Bluhm,W.F., Chen,L., Deshpande,N., Feng,Z., Fleri,W., Green,R., Merino-Ott,J.C., Townsend-Merino,W. *et al.* (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.*, **32** (Database issue), D223–D225.

8. Boobbyer,D.N.A., Goodford,P.J., McWhinnie,P.M. and Wade,R. (1989) New hydrogen-bond potentials for use in determining energetically favorable binding sites in molecules of known structure. *J. Med. Chem.*, **32**, 1083–1094.

9. Wade,R.C., Clark,K.J. and Googford,P.J. (1993) Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.*, **36**, 140–156.

10. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD—Visual Molecular Dynamics. *J. Molec. Graphics*, **14**, 33–38.

11. Eisenhaber,F. and Argos,P. (1993) Improved strategy in analytic surface calculation for molecular systems: handling of singularities and computational efficiency. *J. Comput. Chem.*, **14**, 1272–1280.

12. Eisenhaber,F., Lijnzaad,P., Argos,P., Sander,C. and Scharf,M. (1995) The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.*, **16**, 273–284.

13. Laskowski,R.A., Hutchinson,E.G., Michie,A.D., Wallace,A.C., Jones,M.L. and Thornton,J.M. (1997) PDBsum: A web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.*, **22** , 488–490.

14. Ramachandran,G.N. and Sasisekaran,V.V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.*, **23**, 283–438.

15. Lovell,S.C., Davis,I.W., Arendall,W.B., III, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S. and Richardson,D.C. (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins*, **50**, 437–450.