BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 20 2010, pages 2624–2625
doi:10.1093/bioinformatics/btq480

Structural bioinformatics

Advance Access publication August 24, 2010

2Struc: the secondary structure server

D. P. Klose[1], B. A. Wallace[2] and Robert W. Janes[1,*]

[1]School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS and [2]Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, University of London, London WC1E 7HX, UK

Associate Editor: Anna Tramontano

ABSTRACT

Summary: The defined secondary structure of proteins method is often considered the gold standard for assignment of secondary structure from three-dimensional coordinates. However, there are alternative methods. '2Struc: The Secondary Structure Server' has been created as a single point of access for eight different secondary structure assignment methods. It has been designed to enable comparisons between methods for analyzing the secondary structure content for a single protein. It also includes a second functionality, 'Compare-the-Protein' to enable comparisons of the secondary structure features from any one method to be made within a collection of nuclear magnetic resonance models, or between the crystal structures of two different proteins.

Availability: http://2struc.cryst.bbk.ac.uk

Contact: r.w.janes@qmul.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 29, 2010; revised on August 13, 2010; accepted on August 14, 2010

1 INTRODUCTION

In 1983, Kabsch and Sander (1983) developed what has become the *de facto* approach when defining secondary structure, the defined secondary structure of proteins (DSSP). DSSP uses a single hydrogen bond energy term to assign eight states of secondary structure elements (SSEs) to three-dimensional coordinates and is arguably the golden standard for this task. Although extensively used, DSSP is not the only tool available and other approaches have been developed which incorporate different means of defining SSEs. It has been suggested that DSSP can miss the elucidation of edge strands (Martin *et al*., 2005) and has difficulties dealing with low resolution structures (Martin *et al*., 2005) and with structures where only the Cα trace atoms are reported (Labesse *et al*., 1997). Further methods have been developed including DSSPcont (Anderson *et al.*, 2002), STRIDE (Frishman and Argos, 1995), PALSSE (Majumdar *et al.*, 2005), P-SEA (Labesse *et al.*, 1997), KAKSI (Martin *et al.*, 2005), STICK (Taylor, 2001) and XTLSSTR (King and Johnson, 1999), the latter created for deriving secondary structure from three-dimensional coordinates to match comparable data determined from circular dichroism spectra. These other methods meet specific needs that are not optimally addressed by DSSP; all have merit, and as such broaden the means by which SSEs can be identified.

The aim of the 2Struc server is to provide easy access to the information produced by a variety of these SSE assignment tools. Currently, eight methods are available on the server, covering all atom approaches such as DSSP, DSSPcont (an extension of DSSP using more detailed hydrogen bonding energies) and STRIDE, vector/line segment methods such as STICK and PALSSE and distance/angle methods such as P-SEA, KAKSI and XTLSSTR. Citation references to each of the methods are provided on the server and we request all users to cite these original articles. The output of each method is presented numerically in the form of total secondary structure content as well as for individual residues in a protein's sequence and on a graphical display of the protein's three-dimensional structure using a Jmol (http://www.jmol.org/) representation. In addition to method comparisons, differences within nuclear magnetic resonance (NMR) models and between two *x*-ray structures are possible using the 'Compare-the-Protein' feature.

2 2Struc FEATURES

The 2Struc server comprises two functionalities: 2Struc and Compare-the-Protein. A screenshot of the front page is available as Supplementary Material.

2.1 2Struc analyses

2Struc analyses are compatible with protein atomic coordinates either by specifying the PDB (Berman *et al*., 2007) code or by uploading a coordinates file in PDB format containing ATOM specifiers with or without SEQRES data. Once uploaded, the file is presented to the 'control panel' the point from which the user can select which structures are to be analyzed, the tools to be used and the analyses to run. The five types of analyses available are as follows:

Protein Structure Summary provides a list of composite chains along with their lengths. 2Struc calculates the percentage SSE either according to the original method's classifications or a reduced three-state representation: helix (encompassing H, I, G), sheet (E) and 'other' (-) (all remaining elements), where H, I, G and E are from their DSSP definitions (Kabsch and Sander, 1983). The percentage of each structure element is presented in a tabular form allowing easy comparisons to be drawn. The three-state SSE assignments for each method are compared to DSSP using several commonly levied metrics including the percentage similarity, the more stringent Matthews correlation coefficient (MCC) (Matthews, 1975) and the MCC squared. This summary information can be downloaded as a plain text file.

Structure Summary by Chain is comparable to the previous analysis, but is now on a per chain basis. For each of the chains analyzed, the original method's classifications and three-state percentages are displayed (as above); however, rather than

*To whom correspondence should be addressed.

having a statistical comparison, the user can view the different assignments on a three-dimensional representation of the structure using an embedded Jmol viewer.

Multiple Structure Alignment presents all the reduced three-state assignments for each individual residue aligned in parallel to the protein sequence. For clarity, residues that are in helices are background-colored red and those in beta strands green. A majority vote consensus is also included that is similarly colored.

Original Structure Alignment presents SSE alignments for each method that produces more than a three-state output. The majority consensus is not included here as the algorithms share few commonly defined elements; however, the structures are colored in a similar way to those of the previous analysis approach.

Sequence Structure Alignment presents three state and, where available, original SSE output assignments aligned to the sequence for each method, and similarly colored as before.

2.2 Compare-the-Protein analyses

Compare-the-Protein is the second functionality. It allows the user to examine, for each method, structural variation either within a collection of NMR models or between two *x*-ray crystallography structures. Comparisons can be either in reduced three-state format or in original output assignments where determined. For comparisons within NMR models, the SSE assignments for each model are aligned to the sequence and colored in the same manner as was used in 2Struc. X-ray structure comparisons are between chains. The user can either use the default of the entire chain or specify the start and end position of each chain. The chains can be aligned by SSE using one of two different methods: (i) the tm-align algorithm (Zang and Skolnick, 2005) or (ii) minimizing the number of differences between the assigned SSEs retaining the chains as contiguous components. The SSE assignments are aligned to each of the protein sequences and a 'D' is indicated in the comparison output line where a difference exists between them.

3 APPLICATIONS

2Struc offers a means of comparing specific SSEs for a given three-dimensional protein structure as defined by eight different methods. Compare-the-Protein enables comparisons of SSE assignments to be made between two or more crystal/NMR/model structures for any chosen method. Envisioned applications for the 2Struc and Compare-the-Protein tools are as follows:

As a means of determining the 'consensus' SSEs within a protein as defined from a variety of methods.

For comparison between multiple chains, either for identical subunits within a protein quaternary structure or identical components within the asymmetric unit of a crystal structure, to determine possible areas of differences and the cause of such differences.

As a tool for developing new methods for determining secondary structure based on different structural biology techniques such as circular dichroism, Raman and infrared spectroscopies.

For visual display of differences between homologous proteins or wild-type and mutant protein structures.

As a means of identifying proteins that have unusual SSE features that might be of interest for further structural examination.

To identify differences in SSE assignments between apo- and ligand-bound structures of a protein near to an active site.

For visualization of differences in SSEs that arise from binding of an allosteric effector molecule at a site distal to the active site.

For comparing the SSEs of protein structures prepared under different crystallization conditions.

For comparing the SSE assignments of a solution (NMR) and crystal structure of the same protein.

For comparing the SSEs of a model structure with an experimentally determined cognate or homologous structure.

To enable comparisons of structurally homologous proteins that are not sequence homologous to determine how closely their SSE assignments align.

4 CONCLUSION

The 2Struc server provides two functionalities; 2Struc, which generates SSE assignments for protein structures for up to eight different methods enabling easy analyses of similarities and differences between them, and Compare-the-Protein, which allows comparisons and highlighting of differences between the SSEs generated within a series of NMR models or between two *x*-ray structures. 2Struc is unique in providing summaries of percentage secondary structure content for reduced three-state data and original SSE output assignments where generated. The server is freely available to all users at http://2struc.cryst.bbk.ac.uk.

ACKNOWLEDGEMENTS

We thank Dr Lee Whitmore for testing and improvements, Dr Andrew Miles and Ben Woollett for comments on the manuscript and Dr David Houldershaw for computing technical support, all from Birkbeck College. We also thank Prof. William R. Taylor from the National Institute of Medical Research for supplying the source code for STICK.

Funding: U.K. Biotechnology and Biological Sciences Research Council (BBSRC) [BB/F010362 to R.W.J. and BB/F010346 to B.A.W.].

Conflict of Interest: none declared.

REFERENCES

Anderson,C.A.F. *et al.* (2002) Continuum secondary structure captures protein flexibility. *Structure*, **10**, 175–184.

Berman,H. *et al.* (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D313.

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

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

King,S.M. and Johnson,W.C. (1999) Assigning secondary structure from protein coordinate data. *Proteins Struct. Funct. Genet.*, **35**, 313–320.

Labesse,G. *et al.* (1997) P-SEA: a new efficient assignment of secondary structure from C alpha trace of proteins. *Comput. Appl. Biosci.*, **13**, 291–295.

Majumdar,I. *et al.* (2005) PALSSE: a program to delineate linear secondary structural elements from protein structures. *BMC Bioinformatics*, **6**, 202.

Martin,J. *et al.* (2005) Protein secondary structure assignment revisited: a detailed analysis of different assignment methods. *BMC Struct. Biol.*, **5**, 17.

Matthews,B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta*, **405**, 442–451.

Taylor,W.R. (2001) Defining linear segments in protein structure. *J. Mol. Biol.*, **310**, 1135–1150.

Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.