BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 22 2010, pages 2916–2917
doi:10.1093/bioinformatics/btq537

Structural bioinformatics

Advance Access publication September 21, 2010

MOBI: a web server to define and visualize structural mobility in NMR protein ensembles

Alberto J. M. Martin, Ian Walsh and Silvio C. E. Tosatto*
Department of Biology, University of Padova, viale G. Colombo 3, 35131 Padova, Italy

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: MOBI is a web server for the identification of structurally mobile regions in NMR protein ensembles. It provides a binary mobility definition that is analogous to the commonly used definition of intrinsic disorder in X-ray crystallographic structures. At least three different use cases can be envisaged: (i) visualization of NMR mobility for structural analysis; (ii) definition of regions for reliable comparative modelling in protein structure prediction and (iii) definition of mobility in analogy to intrinsic disorder. MOBI uses structural superposition and local conformational differences to derive a robust binary mobility definition that is in excellent agreement with the manually curated definition used in the CASP8 experiment for intrinsic disorder in NMR structure. The output includes mobility-coloured PDB files, mobility plots and a FASTA formatted sequence file summarizing the mobility results.

Availability: The MOBI server and supplementary methods are available for non-commercial use at URL: http://protein.bio.unipd.it/mobi/

Contact: silvio.tosatto@unipd.it

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

Received on July 5, 2010; revised on September 1, 2010; accepted on September 17, 2010

1 INTRODUCTION

NMR solved structures deposited in the Protein Data Bank (PDB; Berman *et al*., 2007) often contain an ensemble of structurally different conformers, where each model equally satisfies the experimentally obtained restrains. Among other reasons, differences between models are due to incompleteness or noise in the experimental data, as well as reflecting the dynamic behaviour of proteins in solution (Felli and Brutscher, 2009; Foster *et al*., 2007). To gain insight into protein dynamics, the structural ensemble can be further enriched (Laughton *et al*., 2009) or used to deduce quasi-rigid dynamical-domains (Aleksiev *et al*., 2009). In order to make sense of this structural variability, e.g. to identify reliable regions for protein structure prediction, it is commonly sieved for conserved regions (Kelley and Sutcliffe, 1997; Konagurthu *et al*., 2010; Theobald and Wuttke, 2006) or to select the best representative (Tosatto and Battistutta, 2007). An intuitive representation of NMR variability through colouring of the structure by RMSD values has been recently proposed in CARON (Sikic and Carugo, 2009).

Extreme structural variability in proteins is called intrinsic disorder and commonly has functional implications (Dyson and Wright, 2005). In the recent 8th round of the Critical Assessment of techniques for protein Structure Prediction experiment (CASP8), 19 structures solved by NMR were evaluated in the disordered regions prediction category (Noivirt-Brik *et al*., 2009). The 18 structures (one was fully disordered) with a PDB entry were visually and manually inspected by the CASP assessors to determine which amino acids do not have a fixed conformation. To the best of our knowledge, this was the first and fully manual attempt to define intrinsic disorder from NMR structural ensembles.

Here we present MOBI, a new web server to detect and visualize mobile regions in NMR solved structures in analogy to intrinsic disorder in X-ray crystallography. MOBI is based on a simple algorithm to find regions with different conformations among all the models in an ensemble optimized to replicate the ordered-disordered definition used in CASP8.

2 METHODS

2.1 Mobility definition

MOBI combines two basic measures of mobility, structural superposition and changes in local conformation (i.e. torsion angles), into a consensus definition. TM-Score (Zhang and Skolnick, 2004), a method designed to superimpose models with identical sequence (e.g. comparative models or NMR structures), is used to superimpose each of the models in a NMR PDB file on all the other models. Distances between the same Cαs in all superimpositions are computed and scaled according to the formula $SD = 1/((1 + (d/d_0)^2)$, where d is the distance between two Cαs and d_0 is a normalization factor. Average and standard deviation of the scaled distances are computed. Secondly, φ and ψ angles standard deviation for all the amino acids in all the models are also computed. These are used to assign mobility to neighbouring amino acids to those assigned as mobile according to SD average and standard deviation. DSSP (Kabsch and Sander, 1983) is then used to assign secondary structure to each of the models. If secondary structure is the same for an amino acid in all the models (no coils), the amino acid is not mobile, and if the secondary structure changes from model to model it is assigned as mobile independently of the other criteria. Finally, regular expressions are used to assign as mobile one or two ordered positions surrounded by mobile residues, and to filter out single and two consecutive mobile amino acids. A full description of the binary mobility definition can be found in the Online Methods page.

2.2 Agreement with CASP8 disorder definition

The thresholds for each of the mobility measurements are optimized by grid search to maximize the agreement with disorder definitions on the 18 NMR solved targets evaluated at CASP8 with a corresponding PDB structure file. Only amino acids with the same identity in PDB and the CASP8 sequence are considered in the optimization. Residues with no coordinates in the PDB file are also not considered. To optimize the agreement, d_0 is set to 4 Å. SD, its standard deviation and φ and ψ angles standard deviation thresholds are set to

*To whom correspondence should be addressed.

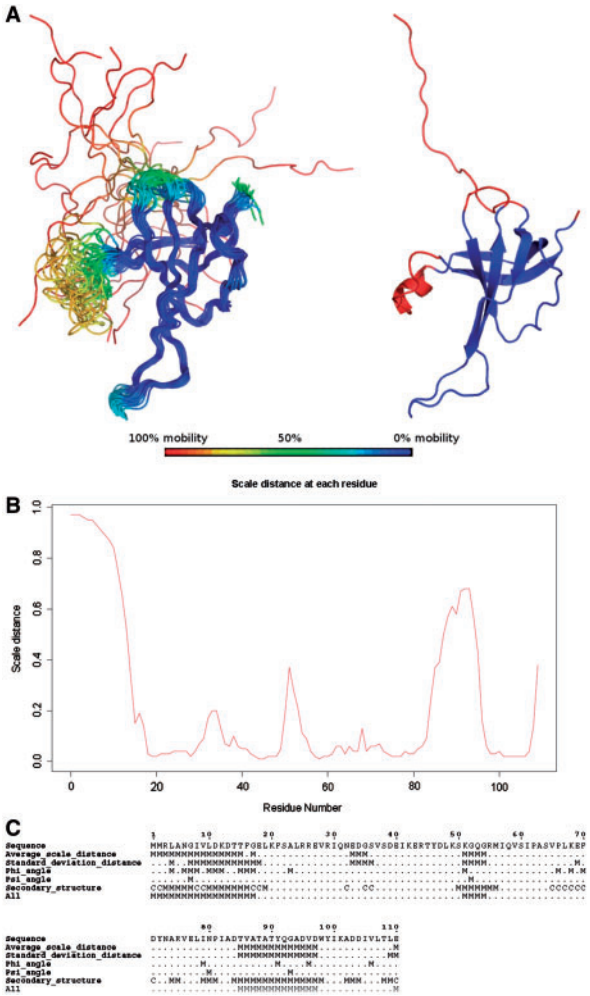

Fig. 1. MOBI results for the NMR structure with PDB code 2K5D. This structure, containing 110 residues and 20 different models, was used in CASP8 as target T0466. (**A**) The ensemble of 20 structures (left) coloured by scaled distance and the most representative structure (right) coloured by binary mobility. Notice how the colours match the structural variability. (**B**) The scaled distance graph describes the variability shown in (A) numerically. (**C**) Shows the correspondence between the protein sequence (top row) and mobility according to the single MOBI criteria, as well as the final MOBI consensus (bottom row).

0.85, 0.09, 20 ° and 20 ° respectively. The optimized binary definition has an F value (the average of precision and recall) of 93.9, indicating almost perfect agreement with the manual CASP8 assignment. See the Online Methods page for a full description of the optimization protocol.

3 IMPLEMENTATION

MOBI has been implemented as a web server which uses a set of Perl scripts and the Jmol (URL: http://www.jmol.org/) and Jalview (Waterhouse *et al*., 2009) helper applications. In the following, we will briefly describe the MOBI input and output pages. Detailed help pages and examples are available online.

3.1 Input

The server requires a PDB file which can be either selected by PDB identifier from the local PDB database or uploaded, but has to conform to the PDB NMR structure specification. Multiple chains are supported. A sequence name and the E-mail address are optional. A more complex interface allows the user to manipulate various parameters used in the mobility definition. A batch mode for processing multiple queries is also available.

3.2 Output

The output has been designed for usage to visualize structural variability and the corresponding mobility definition in various ways (Fig. 1) including binary mobility for each residue as a sequence alignment (optionally in Jalview). The mobility is visualized in Jmol from a modified PDB file (available for download) where B-factors and occupancy are replaced either by scaled distance or binary mobility, allowing easy visual interpretation. Several plots are available in PDF format, describing mobility in terms of structural superposition, torsion angle difference and consensus binary definition.

ACKNOWLEDGEMENTS

The authors are grateful to Joel Sussman and Orly Noivirt-Brik for providing the CASP8 disorder definitions and to members of the BioComputing UP lab for insightful discussions.

Funding: University of Padova grants (CPDA098382 to S.T. and CPDR097328 to A.M. and S.T.).

Conflict of Interest: none declared.

REFERENCES

Aleksiev,T. *et al.* (2009) PiSQRD: a web server for decomposing proteins into quasi-rigid dynamical domains. *Bioinformatics*, **25**, 2743–2744.

Berman,H. *et al.* (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.

Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.

Felli,I.C. and Brutscher,B. (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. *Chemphyschem.*, **10**, 1356–1368.

Foster,M.P. *et al*. (2007) Solution NMR of large molecules and assemblies. *Biochemistry*, **46**, 331–340.

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

Kelley,L.A. and Sutcliffe,M.J. (1997) OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And Domains. *Protein Sci.*, **6**, 2628–2630.

Konagurthu,A.S. *et al.* (2010) MUSTANG-MR structural sieving server: applications in protein structural analysis and crystallography. *PLoS One*, **5**, e10048.

Laughton,C.A. *et al*. (2009) COCO: a simple tool to enrich the representation of conformational variability in NMR structures. *Proteins*, **75**, 206–216.

Noivirt-Brik,O. *et al*. (2009) Assessment of disorder predictions in CASP8. *Proteins*, **77** (Suppl. 9), 210–216.

Sikic,K. and Carugo,O. (2009) CARON–average RMSD of NMR structure ensembles. *Bioinformation*, **4**, 132–133.

Theobald,D.L. and Wuttke,D.S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22**, 2171–2172.

Tosatto,S.C. and Battistutta,R. (2007) TAP score: torsion angle propensity normalization applied to local protein structure evaluation. *BMC Bioinformatics*, **8**, 155.

Waterhouse,A.M. *et al.* (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.

Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.