BMC Systems Biology

Poster presentation

Open Access

PFF – an integrated database of residues and fragments critical for protein folding

Manuel Corpas*, James Sinnott, Dave Thorne, Steve Pettifer, Terri Attwood and the PFF consortium

Address: Faculty of Life Sciences and Computer Science, University of Manchester, Manchester, UK

Email: Manuel Corpas* - corpas@bioinf.man.ac.uk

* Corresponding author

from BioSysBio 2007: Systems Biology, Bioinformatics and Synthetic Biology
Manchester, UK. 11–13 January 2007

Published: 8 May 2007

BMC Systems Biology 2007, **1**(Suppl 1):P48 doi:10.1186/1752-0509-1-S1-P48

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

© 2007 Corpas et al; licensee BioMed Central Ltd.

Background

Despite decades of work, understanding how proteins fold remains a major research challenge. The fruits of this massive research effort have been: development of (i) methods for predicting the likely structures that protein sequences will adopt, or for simulating the folding process itself; and (ii) databases of structural information (e.g., containing 3D coordinates, fold classifications, structure summary data, and so on). As part of the ongoing endeavour to understand the principles of protein folding, we have been involved in the development of a new, integrated structure information resource, based on a small subset of the PDB [1]. The resource contains information derived from a combination of sequence analysis tools, structure analysis software and fold simulation algorithms; to make the contents more accessible to the wider community, we have also developed a user-friendly front-end for visualising the integrated data. The motivation for combining data from these various approaches is to offer insights into the role of particular types of residues and fragments in protein folding, and hence to improve our understanding of factors that are critical to the folding process in general.

Results

A structural annotated database has been generated derived from several unrelated algorithms and data sources for an integrated analysis of critical fragments in protein folding. From an initial analysis of the data, we found, not surprisingly, that certain results were strongly correlated: e.g., residue accessibility values (denoting the degree of internal constraint on flexibility), Fold-X [2] scores (denoting the stabilising contributions to the fold), Popmusic [3] values (denoting destabilising contributions), and lattice simulations [4] (denoting the number of close neighbours or interaction partners within the fold). We used these values to synthesise a 'folding score'.

Conclusion

The PFF database collected analyses and their companion resources to be made publicly available. A goal of the PFF consortium was to create a consensus "prediction" tool combining the strengths of different methods. We found that integration of different methods has indeed added value over individual ones. Coupled with the degree of conservation of residues, a folding score was created to delineate regions that are likely to contribute to (i) the stability of the fold (and hence may contribute to the folding nucleus), and (ii) the function of the protein. This offers a means of automatic motif detection, which can be used for protein family characterisation and functional/structural annotation of evolutionarily conserved regions. We present here a simple case-study to illustrate how the combined data can be used to pinpoint such motifs with potential structural and functional roles.

Availability

Version 1.0 of the PFF dataset is accessible in a DSSP-flat-file format from http://babylone.ulb.ac.be/LIFE/; it is also available in an XML format through the UTOPIA toolkit, together with the UTOPIA visualisation tools for OS X, Windows and Linux at http://utopia.cs.manchester.ac.uk. The Web resource for calculating combined folding scores is accessible at http://umber.sbs.man.ac.uk/~corpas/db/. For a more detailed explanation on the meaning and biological implications of the folding score please refer to http://umber.sbs.man.ac.uk/corpas/db/method_doc.html.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28(1):**235-242.
2. Guerois R, Nielsen JE, Serrano L: **Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations.** *J Mol Biol* 2002, **320(2):**369-387.
3. Kwasigroch JM, Gilis D, Dehouck Y, Rooman M: **PoPMuSiC, rationally designing point mutations in protein structures.** *Bioinformatics* 2002, **18(12):**1701-1702.
4. Papandreou N, Berezovsky IN, Lopes A, Eliopoulos E, Chomilier J: **Universal positions in globular proteins.** *Eur J Biochem* 2004, **271(23–24):**4762-4768.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral