BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 13 2014, pages 1935–1936
doi:10.1093/bioinformatics/btu129

Structural bioinformatics

Advance Access publication March 6, 2014

Frag'r'Us: knowledge-based sampling of protein backbone conformations for *de novo* structure-based protein design

Jaume Bonet[1], Joan Segura[2], Joan Planas-Iglesias[1], Baldomero Oliva[1,*] and Narcis Fernandez-Fuentes[3,*]

[1]Structural Bioinformatics Lab (GRIB), Universitat Pompeu Fabra, Barcelona, Catalonia 08003, Spain, [2]Department of Structure of Macromolecules, CNB-CSIC, Cantoblanco, 28049 Madrid, Spain and [3]Institute of Biological, Environmental and Rural Science, Aberystwyth University, Aberystwyth, SY23 3DA Ceredigion, UK

Associate Editor: Anna Tramontano

ABSTRACT

Motivation The remodeling of short fragment(s) of the protein backbone to accommodate new function(s), fine-tune binding specificities or change/create novel protein interactions is a common task in structure-based computational design. Alternative backbone conformations can be generated *de novo* or by redeploying existing fragments extracted from protein structures i.e. knowledge-based. We present Frag'r'Us, a web server designed to sample alternative protein backbone conformations in loop regions. The method relies on a database of super secondary structural motifs called *s*motifs. Thus, sampling of conformations reflects structurally feasible fragments compiled from existing protein structures.

Availability and implementation Frag'r'Us has been implemented as web application and is available at http://www.bioinsilico.org/FRAGRUS.

Contact narcis.fernandez@gmail.com or baldo.oliva@upf.edu

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

Received on November 12, 2013; revised on February 7, 2014; accepted on March 1, 2014

1 INTRODUCTION

Structural-based computational design aims to alter and change the properties of proteins by using structural information. These include improving protein stability, change and/or optimization of catalytic activities, grafting of novel functionalities or altering the binding to small molecules, nucleic acids or proteins [see reviews (Khare and Fleishman, 2013; Kiss *et al.*, 2013) and references therein]. Pioneering approaches in the field were focused in single amino acid changes with null or limited backbone flexibility (Lees *et al.*, 2012). However, the remodeling and redesign of short fragments of the protein backbone, i.e. *de novo* design, overcomes the intrinsic limitations of fixed backbone design. Recent works have shown the potential of *de novo* design (Eiben *et al.*, 2012; Murphy *et al.*, 2009).

The remodeling of the protein backbone usually occurs in the loop regions, given its intrinsic flexibility and variable nature. Usually, remodeling the loop regions seeks to accommodate specific interactions and restraints, e.g. catalytic residues (Murphy *et al.*, 2009), or to create new local segments (Hu *et al.*, 2007). The major limitation of *de novo* approaches is the inherent combinatorial complexity that makes almost impossible the systematic sampling of large insertions.

In early works we have described the use of *s*motifs in loop structure prediction [including protein kinases (Fernandez-Fuentes *et al.*, 2004)], functional annotation (Espadaler *et al.*, 2006) and prediction of protein–protein interactions (Planas-Iglesias *et al.*, 2013). Furthermore, we described that the dramatic increase of structures in the protein databank has resulted in the saturation of sampling of loop structures, even in the case of long loops (Bonet *et al.*, 2014; Fernandez-Fuentes and Fiser, 2006) subsequently confirmed (Choi and Deane, 2010), and that the geometry of *s*motifs is fully sampled (Fernandez-Fuentes *et al.*, 2010).

Here we extend the use of *s*motifs and we present Frag'r'Us, a knowledge-based approach designed to sample the conformation of loop regions between fixed flanking regular secondary structures. The sampling is based in the comparison of the geometry between query and candidate *s*motifs (see Server Usage in supplementary data), is length independent (the final loop length is not predetermined) and is very fast: a search among thousand of potential conformations is done in a matter of seconds. Candidate *s*motifs matching the geometrical constraints can then be grafted into the protein scaffold to be used as starting conformations for different *de novo* design manipulations.

2 APPLICATION

2.1 Definition of *s*motifs and geometry features

*S*motifs are super secondary elements composed of a loop region flanked by two regular secondary structures, i.e. α-helix or β-strand. The local structural arrangement of the two flanking secondary structures, i.e. Nt and Ct, defines the geometry of the *s*motif through four internal variables, a distance D and three angles, i.e. delta, theta and rho, as described in our earlier work (Fernandez-Fuentes *et al.*, 2006).

2.2 Server usage

The method has been implemented as a step-by-step web server. Frag'r'Us' basic inputs are the 3D coordinates and secondary structure definitions. In both cases, those inputs can be provided by the user (Supplementary Materials) or the coordinates can be automatically gathered from a Protein Data Bank (PDB) identification code (Berman *et al.*, 2000) and

*To whom correspondence should be addressed.

the secondary structure calculated using DSSP (Kabsch and Sander, 1983). The chain of interest has to be defined. Additionally, the user can tune the tolerance for every search parameter exploited by the server. This includes the source database of *s*motifs, every geometrical property of the *s*motifs, the loop length and the structural fitting. Finally, it can perform three different tasks, namely, identify *s*motifs in the query coordinates, search for geometrically compatible *s*motifs and structurally align them. The search of candidate *s*motifs can be done in two different libraries: a redundant and non-redundant derived from the high-quality protein structures (i.e. X-ray resolution better than 2.5 Å) deposited in the PDB (Berman *et al.*, 2000) using PISCES (Wang and Dunbrack, 2003) at a 95 and 40% sequence identity cutoff respectively. See the Supplementary Material for a detailed explanation on both inputs and outputs.

2.3 *S*motifs sampling as strategy to generate useful initial conformations for backbone remodeling

The sampling of backbone conformations of loop regions is done using *s*motifs definitions. The sampling generates a comprehensive set of realistic conformations to be used as a starting point for further manipulation and optimization using protein design techniques. To further demonstrate that the sampled conformations are suitable, we have compiled a list of cases where the 3D structures of the scaffold and computationally engineered proteins are known. The set includes a wide range of examples in *de novo* computation design including a Diels-Alderase (Eiben *et al.*, 2012), a human guanine deaminase (Murphy *et al.*, 2009), the catalytic loops in ($\beta\alpha$)8-barrel scaffolds (Claren *et al.*, 2009), two retro-aldol enzymes (Jiang *et al.*, 2008; Wang *et al.*, 2012), a triosephosphate isomerase (Saab-Rincon *et al.*, 2012) and the motifs to target antibody b12 (Azoitei *et al.*, 2011). A detailed analysis of each case can be found in the Supplementary Material.

Of the cases presented above, the remodeling of a 24-residue long insertion that included a helix-turn-helix motif is perhaps the most challenging (Eiben *et al.*, 2012). The remodeling of this region was carried by combining computational structure-based design: Rosetta (Leaver-Fay *et al.*, 2011), and crowdsourcing using the game-driven protein folding Foldit (http://fold.it). As shown in Figure 1, the loop region of and *s*motif extracted from an uncharacterized protein from *Porphyromonas gingivalis* W83 (PG_1388; PDB code 2p3p) closely resembles the structure of the remodeled region.

3 CONCLUSION

In this article, we present a server, Frag'r'Us, designed to sample loop conformations using the geometrical constrains of *s*motifs. Frag'r'Us has clear applications in *de novo* computational design by seeding the search of alternative backbone conformations using geometrically suitable backbone regions extracted from known protein structures. Frag'r'Us can be also used in protein–protein, protein–nucleic acids and protein-small ligand interface design by generating a set of alternative conformation of interface loops. Finally, Frag'r'Us can complement loop structure prediction methods and X-ray/NMR refinement protocols by generating loop conformations that are length independent.

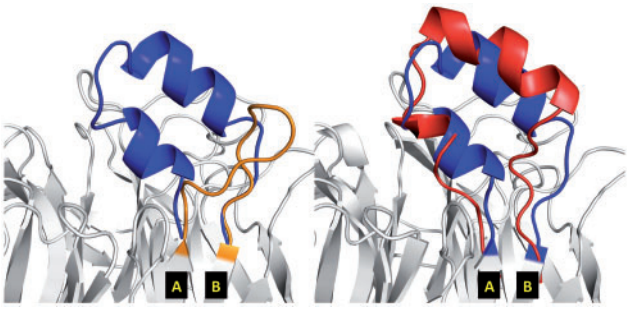

Fig. 1. Ribbon representation of the superposed structure of scaffold (orange), engineered (blue) proteins and the best matching candidate *s*motif (red). See details in Supplementary Figure S2

ACKNOWLEDGEMENTS

The authors thank Dr Gendra for insightful comments to the manuscript.

Funding: This work has been partially funded by the Spanish Government (Grants FEDER BIO2008-0205, BIO2011-22568 and EUI2009-04018] and the Institute of Biological, Environmental and Rural Sciences.

Conflict of Interest: none declared.

REFERENCES

Azoitei,M.L. *et al.* (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science*, **334**, 373–376.
Berman,H.M. *et al.* (2000) The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, **7** (**Suppl**), 957–959.
Bonet,J. *et al.* (2014) ArchDB 2014: structural classification of loops in proteins. *Nucleic Acids Res.*, **42**, D315–D319.
Choi,Y. and Deane,C.M. (2010) FREAD revisited: accurate loop structure prediction using a database search algorithm. *Proteins*, **78**, 1431–1440.
Claren,J. *et al.* (2009) Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc. Natl Acad. Sci. USA*, **106**, 3704–3709.
Eiben,C.B. *et al.* (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat. Biotechnol.*, **30**, 190–192.
Espadaler,J. *et al.* (2006) Identification of function-associated loop motifs and application to protein function prediction. *Bioinformatics*, **22**, 2237–2243.
Fernandez-Fuentes,N. *et al.* (2010) Structural characteristics of novel protein folds, *PLoS Comput. Biol.*, **6**, e1000750.
Fernandez-Fuentes,N. and Fiser,A. (2006) Saturating representation of loop conformational fragments in structure databanks, *BMC Struct. Biol.*, **6**, 15.
Fernandez-Fuentes,N. *et al.* (2004) Classification of common functional loops of kinase super-families. *Proteins*, **56**, 539–555.
Hu,X. *et al.* (2007) High-resolution design of a protein loop. *Proc. Natl Acad. Sci. USA*, **104**, 17668–17673.
Jiang,L. *et al.* (2008) *De novo* computational design of retro-aldol enzymes. *Science*, **319**, 1387–1391.
Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
Khare,S.D. and Fleishman,S.J. (2013) Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett.*, **587**, 1147–1154.
Kiss,G. *et al.* (2013) Computational enzyme design. *Angew. Chem. Int. Ed. Engl.*, **52**, 5700–5725.
Leaver-Fay,A. *et al.* (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.*, **487**, 545–574.
Lees,J.P. *et al.* (2012) A designed point mutant in Fis1 disrupts dimerization and mitochondrial fission. *J. Mol. Biol.*, **423**, 143–158.
Murphy,P.M. *et al.* (2009) Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl Acad. Sci. USA*, **106**, 9215–9220.
Planas-Iglesias,J. *et al.* (2013) Understanding protein-protein interactions using local structural features. *J. Mol. Biol.*, **425**, 1210–1224.
Saab-Rincon,G. *et al.* (2012) Evolutionary walk between (beta/alpha)(8) barrels: catalytic migration from triosephosphate isomerase to thiamin phosphate synthase. *J. Mol. Biol.*, **416**, 255–270.
Wang,G. and Dunbrack,R.L. Jr (2003) PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589.
Wang,L. *et al.* (2012) Structural analyses of covalent enzyme-substrate analog complexes reveal strengths and limitations of de novo enzyme design. *J. Mol. Biol.*, **415**, 615–625.