**ABSTRACT**

**Motivation** The remodeling of short fragments(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 reusing existing fragments extracted from protein structures i.e. knowledge-based. We present Frag’r’Us, a web server designed to sample alternative backbone conformations in loop regions. The method relies on a database of super secondary structural motifs called smotifs. Thus, sampling of conformations reflects structurally feasible fragments compiled from existing protein structures.

**Availability and implementation** Frag’r’Us has been implemented as a 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 Fleshman, 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 smotifs 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 database 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 smotifs is fully sampled (Fernandez-Fuentes et al., 2010).

Here we extend the use of smotifs 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 smotifs (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 smotifs 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 smotifs and geometry features

Smotifs 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 smotif 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 smotifs, every geometrical property of the smotifs, the loop length and the structural fitting. Finally, it can perform three different tasks, namely, identify smotifs in the query coordinates, search for geometrically compatible smotifs and structurally align them. The search of candidate smotifs 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 A) 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 Smotifs sampling as strategy to generate useful initial conformations for backbone remodeling

The sampling of backbone conformations of loop regions is done using smotifs 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 (β8) barrel scaffolds (Clairen 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 smotif 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 smotifs. 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.

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 EU12009-04018) and the Institute of Biological, Environmental and Rural Sciences.

Conflict of Interest: none declared.

REFERENCES