OXFORD

Structural bioinformatics

ProteinFishing: a protein complex generator within the ModelX toolsuite

Damiano Cianferoni [1], Leandro G. Radusky[1], Sarah A. Head[1], Luis Serrano[1,2,3,*] and Javier Delgado [1,*]

[1]Systems Biology, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona 08003, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona 08002, Spain and [3]ICREA, Barcelona 08010, Spain

*To whom correspondence should be addressed.
Associate Editor: Arne Elofsson

Abstract

Summary: Accurate 3D modelling of protein–protein interactions (PPI) is essential to compensate for the absence of experimentally determined complex structures. Here, we present a new set of commands within the ModelX tool-suite capable of generating atomic-level protein complexes suitable for interface design. Among these commands, the new tool *ProteinFishing* proposes known and/or putative alternative 3D PPI for a given protein complex. The algorithm exploits backbone compatibility of protein fragments to generate mutually exclusive protein interfaces that are quickly evaluated with a knowledge-based statistical force field. Using interleukin-10-R2 co-crystalized with interferon-lambda-3, and a database of X-ray structures containing interleukin-10, this algorithm was able to generate interleukin-10-R2/interleukin-10 structural models in agreement with experimental data.

Availability and implementation: *ProteinFishing* is a portable command-line tool included in the ModelX toolsuite, written in C++, that makes use of an SQL (tested for MySQL and MariaDB) relational database delivered with a template SQL dump called *FishXDB*. *FishXDB* contains the empty tables of ModelX fragments and the data used by the embedded statistical force field. *ProteinFishing* is compiled for Linux-64bit, MacOS-64bit and Windows-32bit operating systems. This software is a proprietary license and is distributed as an executable with its correspondent database dumps. It can be downloaded publicly at http://modelx.crg.es/. Licenses are freely available for academic users after registration on the website and are available under commercial license for for-profit organizations or companies.

Contact: javier.delgado@crg.eu or luis.serrano@crg.eu

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

1 Introduction

The ModelX toolsuite (Delgado Blanco *et al.*, 2019) has been developed, among other purposes, for modelling biomolecular interactions. ModelX uses fragment libraries generated by *in silico* digestion of Protein Data Bank (PDB) structures (Berman *et al.*, 2000) and stored in SQL databases. This strategy has proven successful when applied to the design of DNA–protein and RNA–protein interfaces (Blanco *et al.*, 2018; Delgado Blanco *et al.*, 2019). In the protein–protein interactions (PPI) prediction field, few examples of tools performing fast large-scale docking exist. MEGADOCK 4.0 (Ohue *et al.*, 2014) is one, but it requires sophisticated heterogeneous supercomputing environments equipped with hardware accelerators such as GPUs. Another example is InterPred (Mirabello *et al.*, 2017), which uses homology modelling of binding partners

and whole protein superimposition to gather interaction templates. Here, we present *ProteinFishing*, a tool based on the ModelX philosophy that enables the fast generation of 3D interaction models from observed protein–protein interfaces while fulfilling the requirements for local backbone compatibility.

2 New ModelX tools

In addition to *ProteinFishing*, the latest ModelX release contains two more commands: *GeneratePeptides*, which is needed to populate the *FishXDB* database, and *FishingLure*, an automatized version of *ProteinFishing*. The three mentioned commands can be used with any type of PDB file containing standard amino acids and/or nucleotides, including X-ray, nuclear magnetic resonance (NMR), homology models or any other PDB model created by users.
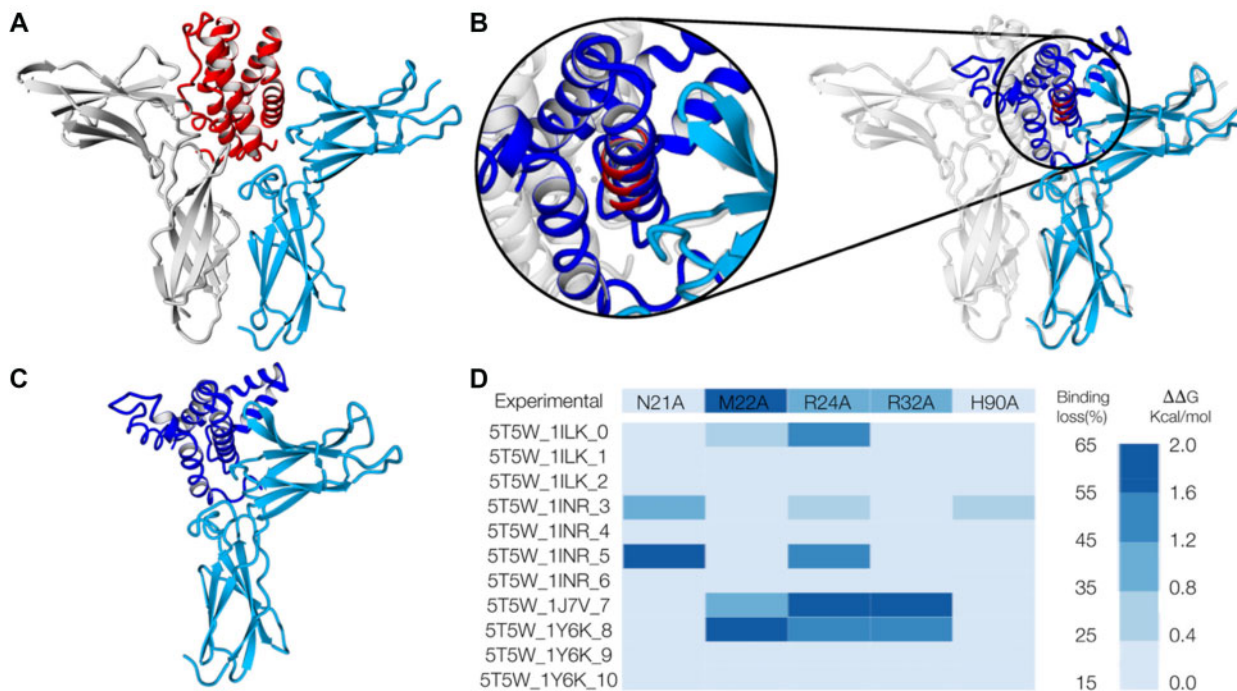
Fig. 1. Algorithm description. (**A**) The IFN-lambda-R1/IFN-lambda-3/IL-10-R2 complex (PDB: 5T5W) containing the 'Hook' (IFN-lambda-3: red), the 'Fisher' (IL-10-R2: light blue) and IFN-lambda-R1 (grey); (**B**) The IFN-lambda-R1/IL-10/IL-10-R2 virtual complex superimposed with the 'Hook' window (red); (**C**) The IL-10/IL-10-R2 or 'Fish/Fisher' complex (IL-10: dark blue; IL-10-R2: light blue); (**D**) A comparison between the reported binding levels (first row) and the $\Delta\Delta G$ of interaction as calculated by FoldX (rows 2–12). A unique colour scale has been used to make energies and percentages comparable. The binding loss (%) numerical scale corresponds to 100%—'binding levels' for experimentally measured point mutations (Yoon *et al.*, 2006) and the $\Delta\Delta G$ (kcal/mol) numerical scale corresponds to FoldX interaction energy

2.1 *GeneratePeptides* command

The *GeneratePeptides* command allows ModelX users to customize their fragment library. It takes PDB structures as input and digests them into protein fragments of user-defined length in an overlapping sliding-window fashion. These fragments are stored in *FishXDB* and are therefore available for the *ProteinFishing* algorithm.

2.2 *ProteinFishing* command

ProteinFishing uses protein complex structures as input, and requires the user to select one molecule to be part of the output complex ('Fisher', Fig. 1A, light blue) and another molecule to be used as the structural template for the retrieval of new docking partners ('Hook', Fig. 1A, red). The algorithm requires the user to define an amino acid window from the 'Hook' molecule to query the *FishXDB* protein fragment database with fragment windows interacting with the 'Fisher'. When the geometrical backbone compatibility and sequence similarity—according to user-configurable options—matches the peptide window with a *FishXDB* fragment, the full PDB model ('Fish', Fig. 1A) from which the fragment was obtained is placed over the 'Hook' fragment by local fitting (Fig. 1B). In this way, complexes containing both the 'Fisher' and the 'Fish' molecules are built (Fig. 1C). Finally, the generated complexes go through two energy filters: the first filter evaluates the presence of atomic clashes between the backbones of the two molecules, and the second filter uses a customizable threshold for free energy values calculated over the generated models. Free energies (representing backbone compatibility) are obtained using a statistical force field embedded in ModelX. The force field is based on a Boltzmann device (Sippl, 1990) with the Kono modification (Kono *et al.*, 1999) of the Sippl method. The models that pass these filters are later returned as PDB files, together with a summary file showing the number of intermolecular contacts, backbone clashes and energy values.

2.3 *FishingLure* command

The *FishingLure* command represents a fully automated, multi-thread version of *ProteinFishing* in which the algorithm itself determines all possible overlapping sliding windows around the 'Hook' residues contacting the 'Fisher'. The *FishingLure* command allows the use of *ProteinFishing* over multiple scanning windows computed in parallel.

3 Demonstration

To test the utility of our tool, we focused on the complexes of interleukin-10 (IL-10) with its two receptors (IL-10-R1 and IL-10-R2). While structures are available for the IL-10/IL-10-R1 complex (PDB: 1J7V, 1Y6K), the structure of the IL-10/IL-10-R2 complex has not been elucidated. We chose the crystallographic structure of IL-10-R2 complexed with interferon-lambda-3 and interferon-lambda-receptor-1 (PDB: 5T5W) as input. IL-10-R2 was used as the 'Fisher' molecule and interferon-lambda-3 was used as the 'Hook' (Fig. 1A, red). Defining the scanning window between residues 89–94 of the 'Hook', *ProteinFishing* yielded 11 models that were then energetically minimized. Next, using the *BuildModel* command of FoldX (Delgado *et al.*, 2019), five point mutations experimentally reported to modify 'binding levels' between IL-10 and IL-10-R2 (Yoon *et al.*, 2006) were modelled. For each mutation, we computed the FoldX free energy variations ($\Delta\Delta G$ [kcal/mol] of interaction) between the 'Fisher' and the mutated 'Fish'. Finally, we compared the variations between the 'binding levels' of the five IL-10 mutants with IL-10-R2, as reported in literature, with those predicted by FoldX in each of the 11 models (Fig. 1D). The two best-fitting models, as ranked by the statistical force field of ModelX (Supplementary Table S1; 5T5W_1Y6K_8 and 5T5W_1J7V_7), were found to have the best agreement between FoldX energy values and the experimental results (Yoon *et al.*, 2006) (Fig. 1D and Supplementary Table S3). Complete details of the entire process, including the specific parameters used, can be found in Supplementary Appendix: User Tutorial.

4 Conclusions

The tools presented here enable the fast structural modelling of PPI suitable for protein design. The *ProteinFishing* algorithm can be applied in two types of scenarios. The first scenario, described

above, allows the user to model a protein complex for which there is no structure available. Depending on the structures with which the user populates the *FishXDB*, the possible interactors 'fished' can be restricted to specific desired targets, or can be exploratory, using all structures from the PDB. In a second scenario, the tool could be used to model different possible conformations between two members of a complex for which a structure already exists. This second scenario could be useful for performing energetic filtering of different conformations, redesigning interfaces by mutagenesis or identifying putative small-molecule binding pockets in the interface between complex members, for example.

References

Berman,H. M. *et al.* (2000) The protein data bank. *Nucleic Acids Res*, **28**, 235–242.

Blanco,J.D. *et al.* (2018) FoldX accurate structural protein–DNA binding prediction using PADA1 (Protein Assisted DNA Assembly 1). *Nucleic Acids Res.*, **46**, 3852–3863.

Delgado Blanco,J. *et al.* (2019) Protein-assisted RNA fragment docking (RnaX) for modeling RNA-protein interactions using ModelX. *Proc. Natl. Acad. Sci. USA*, **116**, 24568–24573.

Delgado,J. *et al.* (2019) FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*, **35**, 4168–4169.

Kono,H. *et al.* (1999) Structure-based prediction of DNA target sites by regulatory proteins. *Proteins Struct. Funct. Genet.*, **35**, 114–131.

Mirabello,C. *et al.* (2017) InterPred: a pipeline to identify and model protein–protein interactions. *Proteins*, **85**, 1159–1170.

Ohue,M. *et al.* (2014) MEGADOCK 4.0: an ultra-high-performance protein–protein docking software for heterogeneous supercomputers. *Bioinformatics*, **30**, 3281–3283.

Sippl,M.J. (1990) Calculation of conformational ensembles from potentials of mena force. *J. Mol. Biol.*, **213**, 859–883.

Yoon,S.I. *et al.* (2006) Conformational changes mediate interleukin-10 receptor 2 (IL-10R2) binding to IL-10 and assembly of the signaling complex. *J. Biol. Chem.*, **281**, 35088–35096.