Prediction of binding energies upon mutation in 3D-structure-known complexes through PyDock scoring functions

Bruno Cuevas, Miguel Romero, and Juan Fernandez Recio
Barcelona Supercomputing Centre, Nexus II Building c/Jordi Girona, 29 08034 Barcelona (Spain)

***Abstract*—The combination of emerging technologies in biology and the exponential increase of available biological data is leading to the new paradigm of personalized medicine where each patient will be diagnosed and treated by integrating all known personal features such as its genetic background.**

In this context, one of the main challenges is the development of tools that can help to characterize pathological mutations. A pathological mutation can act in different ways. Some mutations can modify the overall structure of the protein, leading to incorrectly unfolded proteins that cannot play their functions, or that are targeted by the mechanisms of protein turnover. Other groups of mutations can play crucial roles in the interaction between proteins. In those cases, proteins could remain structurally unchanged, yet the formation of protein-protein complexes is disturbed, inhibiting different paths in which protein-protein interactions (PPIs) are important.

Different strategies have been employed to deal with this challenge, most of them using structural information about protein interactions. In this regard, it is possible to distinguish between two groups of methods: those that make use of machine learning methods, such as random forests or neural networks, applied directly at structure level, like PoPMuSiC [1], ELASPIC [2], BindProf ([3]), ZEMu [4], mCSM [5]; and those that make use of energy scoring functions [6].

The main objective of this work is to build a simple but robust predictor of binding energy changes upon mutation, once structural information is provided. Three different tools are being employed: First, given the wild type structure of a complex, mutations are modelled with Modeller, a powerful program extensively used in protein homology modeling. Using the workframe provided by the tool, 50 different models are created for each mutation. This model diversity helps to take into account protein flexibility and explore, more efficiently, the conformational space of interactions.Then, models are evaluated using Modeller DOPE assessment tool and pyDock scoring function. DOPE [7] is statistical-potential-based tool that evaluates the quality of a model. The pyDock scoring function [8] is formed by different energy terms (electrostatic, desolvation and van der Waals), and was originally designed to deal with protein-protein docking problems. However, due to its energetic basis, we wanted to test their ability to evaluate changes in binding energies.

The correlation between pyDock energy and experimental $\Delta\Delta G$ values is being tested on SKEMPI ([9]), a dataset that provides a large set of mutations in protein complexes for which there are both structural information and experimental measures of the changes in binding energy ($\Delta\Delta G$) upon mutation. A careful curation of the experimental data, coming from different experimental techniques, environmental conditions (e.g., temperature, pH) has been performed. Optimization of parameters and selection of unbiased experimental data are expected to lead to a high correlation coefficient between theoretical values and experimental ones.

Compared with other state-of-the-art methods based on machine-learning approaches, the method we here proposed, offers a much simpler workframe, keeping a scientific-knowledge base.

bruno.cuevas@bsc.es
miguel.romero@bsc.es
juanf@bsc.es

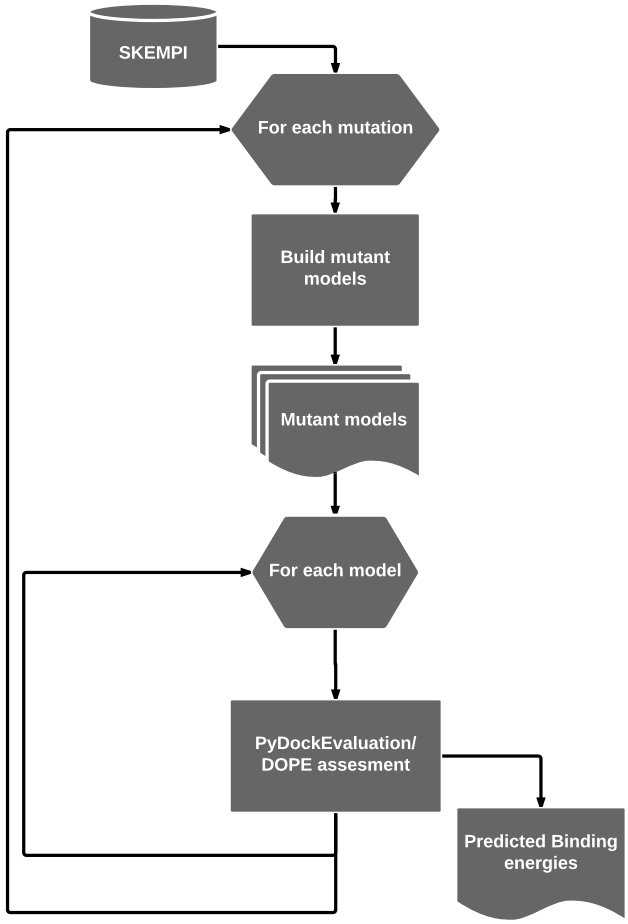

Fig. 1. Binding energy prediction workframe

***Index Terms*—Docking, Personalized Medicine, Structural Biology, PPI**

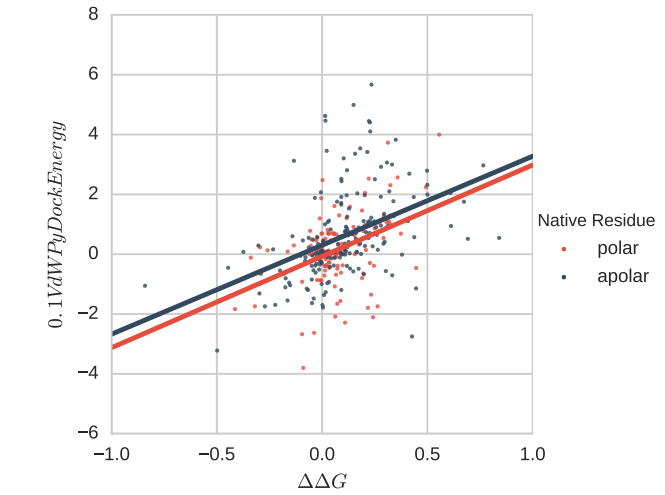

Fig. 2. Real experimental values against predicted ones using PyDock scoring functions when mutations and original residues are polar or apolar

REFERENCES

[1] Y. Dehouck, A. Grosfils, B. Folch, D. Gilis, P. Bogaerts, and M. Rooman, "Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0," *Bioinformatics*, vol. 25, no. 19, pp. 2537–2543, 2009.

[2] N. Berliner, J. Teyra, R. Çolak, S. G. Lopez, and P. M. Kim, "Combining structural modeling with ensemble machine learning to accurately predict protein fold stability and binding affinity effects upon mutation," *PLoS ONE*, vol. 9, no. 9, 2014.

[3] J. R. Brender and Y. Zhang, "Predicting the Effect of Mutations on Protein-Protein Binding Interactions through Structure-Based Interface Profiles," *PLoS Computational Biology*, vol. 11, no. 10, pp. 1–25, 2015.

[4] D. F. A. R. Dourado and S. C. Flores, "A multiscale approach to predicting affinity changes in proteinprotein interfaces," no. June, pp. 2681–2690, 2014.

[5] D. E. V. Pires, D. B. Ascher, and T. L. Blundell, "MCSM: Predicting the effects of mutations in proteins using graph-based signatures," *Bioinformatics*, vol. 30, no. 3, pp. 335–342, 2014.

[6] C. Geng, A. Vangone, and A. M. J. J. Bonvin, "Exploring the interplay between experimental methods and the performance of predictors of binding af fi nity change upon mutations in protein complexes," vol. 29, no. 8, pp. 291–299, 2016.

[7] A. Sali and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Science*, pp. 2507–2524, 2006.

[8] T. M.-K. Cheng, T. L. Blundell, and J. Fernandez-Recio, "pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking," *Proteins: Structure, Function, and Bioinformatics*, vol. 68, no. 2, pp. 503–515, apr 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17444519 http://doi.wiley.com/10.1002/prot.21419

[9] I. H. Moal and J. Fernández-Recio, "SKEMPI: A Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models," *Bioinformatics*, vol. 28, no. 20, pp. 2600–2607, 2012.

Bruno Cuevas Zuviria Graduated in Biotechnology by Universidad Politecnica de Madrid in 2016, and current student of master in Modelling For Science and Engineering in Universidad Autonoma de Barcelona. He works at the protein-protein interaction and docking group in BSC, under direction of Juan Fernandez Recio. He got specialized in Computational Biology during the degree, and in Complex Systems during the master. His main interests are the systems biology, the structural biology and the epidemiology, field in which he used to work before.