pyDockDNA: A new approach for protein-DNA docking

[1]Luis Ángel Rodríguez-Lumbreras, [2]Brian Jiménez-García, [3]Juan Fernández-Recio

Life Sciences - Protein Interactions and Docking, Carrer de Jordi Girona, 29-31, 08034 Barcelona (Spain)

[1]luis.rodriguez@bsc.es, [2]brian.jimenez@bsc.es, [3]juanf@bsc.es

***Keywords*— *ab-initio* docking, protein-DNA interaction, docking scoring function.**

ABSTRACT

Structural prediction of protein-DNA interactions can contribute to the understanding of essential cell processes at molecular level, especially those related to gene expression and regulation. While some of the existing protein-protein docking methods, such as HADDOCK [1], can handle nucleic acids, very few algorithms specifically developed for protein-DNA docking have been reported so far [2].

In this context, our pyDock docking scheme, which has been successfully applied to model a significant number of protein-protein cases [3], can be also applied to model protein-RNA complexes [4]. However, the modeling of protein-DNA interactions with pyDock had not been fully explored so far.

Here we present pyDockDNA, which is based on the pyDock program, with a new module for reading and parsing DNA molecules. The protocol is composed of two major steps: sampling and scoring. The first sampling step consists in the generation of 10,000 protein-DNA docking models by FTDock [5]. This program takes a protein and a nucleic acid coordinate file, discretizes the molecules into corresponding 3D grids, and computes their geometric correlation by using Fast Fourier Transform algorithms to speed up the translations between the two molecules.

For the second scoring step, we have used a modification of the default scoring function implemented in pyDock [6] in order to reevaluate the protein-DNA docking models generated during the sampling step. In the new scoring function within pyDockDNA, the binding energy of each docking model is based on van der Waals (VDW), desolvation and electrostatics energy, using the charges for the DNA molecule as defined in amber 94 [7] force-field.

$$U_{pyDock} = \Delta U_{electrostatic} + \Delta U_{desolvation} + \Delta U_{vanderWaals}$$
$$= \frac{1}{4\pi\epsilon}\sum_{i}^{N_{rec}}\sum_{j}^{N_{lig}}\frac{q_i \cdot q_j}{r_{ij}^2} +$$
$$+ \sum_{i}^{N_{rec}}\Delta ASA_i \cdot \sigma_i + \sum_{j}^{N_{lig}}\Delta ASA_j \cdot \sigma_j +$$
$$+ \sum_{i}^{N_{rec}}\sum_{j}^{N_{lig}}\epsilon(i,j)\left[\left(\frac{r_m(i,j)}{r_{ij}}\right)^{12} - 2\left(\frac{r_m(i,j)}{r_{ij}}\right)^{6}\right]$$

Fig.1: pyDock scoring function, with details of each energetic term.

We have thoroughly tested this protocol on an available protein-DNA docking benchmark [8], comparing several versions of the scoring function, and different combinations of the electrostatics, VDW, solvation energy and the SCscore (FTDock geometric score). The preliminary results show that a combination of electrostatics and VDW might be the best option to re-score the different protein-DNA decoys.

References

[1] Dominguez, C., et al., *HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information.* J. Am. Chem. Soc., 2003. **125**(7): p. 1731-1737.

[2] Banitt, I. and H.J. Wolfson, *ParaDock: a flexible non-specific DNA--rigid protein docking algorithm.* Nucleic Acids Res, 2011. **39**(20): p. e135.

[3] Pallara, C., et al., *Expanding the frontiers of protein-protein modeling: from docking and scoring to binding affinity predictions and other challenges.* Proteins, 2013. **81**(12): p. 2192-200.

[4] Pérez-Cano, L. and J. Fernández-Recio, *Optimal protein-RNA area, OPRA: A propensity-based method to identify RNA-binding sites on proteins.* Proteins: Structure, Function and Bioinformatics, 2010. **78**(1): p. 25-35.

[5] Gabb, H.A., R.M. Jackson, and M.J. Sternberg, *Modelling protein docking using shape complementarity, electrostatics and biochemical information.* J. Mol. Biol., 1997. **272**(1): p. 106-120.

[6] Cheng, T.M.-K., T.L. Blundell, and J. Fernandez-Recio, *pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking.* Proteins, 2007. **68**(2): p. 503-515.

[7] Wendy D. Cornell, P.C., Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, Peter A. Kollman, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* J. Am. Chem. Soc., 1995. **119**(19): p. 5179–5197.

[8] van Dijk, M. and A.M.J.J. Bonvin, *A protein-DNA docking benchmark.* Nucleic Acids Research, 2008. **36**(14): p. e88-e88.

Presenting author biography

Luis Ángel was born in Cadiz, Spain, in 1985. He received a BSc degree in biochemistry from the Universidad de Sevilla, Spain, in 2013, and a Master in Bioinformatics for Health Sciences from the Universitat Pompeu Fabra, Barcelona, Spain, in 2015.

In October 2014, he joined the Protein Interactions and Docking group at the Life Sciences department of the Barcelona Supercomputing Center, first as a Visitor Student, and a year later as a PhD candidate. His current research includes software development and optimization for protein-protein and protein-DNA interactions.