Alfieri: a molecular modelling pipeline to score and rank protein-DNA interactions

Enrico Spiga (spigae)

June 26, 2019

Volli, e volli sempre, e fortissimamente volli.

Vittorio Alfieri

I have developed a molecular modeling pipeline of for consecutive steps in order to score and rank protein-DNA interactions. In what follows, we will describe every single step of the pipeline.

1. Creating all protein-DNA complexes. Firstly, it consists in creating all DNA sequences for which one intends to calculate the protein-DNA enthalpy of binding. Secondly, it consists in: a. creating, using the sequences generated in the step 2, the correspondent 3D structures of the protein-DNA complex using the program *mutate_bases* of the suite 3DNA [1–3]; b. generating the topologies of the structure using the Amber force field employing the Ambertools [4–9].

2. Minimization of all the protein-DNA complexes. It consists in running a short minimization of the protein-DNA complexes using the molecular dynamics package NAMD [10] and the visualization tool VMD [11] to store the minimized structures.

3. Calculating enthalpy of binding and its decomposition per interacting pairs. Using the minimized structures obtained in the previous step, it consists in evaluating the enthalpy of binding, and its decomposition per residue-nucleotide pairs, with the molecular mechanics score functions MMPB(GB)SA using MMPBSA.py [12] and the knowledge-based potential DDNA/DFIRE using ddna [13;14]. The decomposition is not done with ddna.

4. Ranking the DNA sequences. It consists in ranking the protein-DNA interactions, from the best to the worst DNA sequences, on the basis of the scores obtained in the step 3.

References

[1] Lu, X. J.; Olson, W. K. *Nucleic Acids Research* **2003**, *31*, 5108–5121.

[2] Lu, X. J.; Olson, W. K. *Nature Protocols* **2008**, *3*, 1213–1227.

[3] Colasanti, A. V.; Lu, X. J.; Olson, W. K. *Journal of visualized experiments* **2013**, *74*, e4401.

[4] Case, D. A. et al. *University of California, San Francisco* **2018**,

[5] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[6] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.

[7] Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

[8] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.

[9] Prez, A.; Marchn, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. I.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.

[10] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; ; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

[11] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

[12] Miller, B. R. I.; McGee, T. D. J.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *JCTC* **2012**, *8*, 3314–3321.

[13] Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. *Journal of Medicinal Chemistry* **2005**, *48*, 2325–2335.

[14] Zhao, H.; Yang, Y.; Zhou, Y. *Bioinformatics* **2010**, *26*, 1857–1863.