FrAG-PELE: Novel Fragment-based Growing Tool for hit-to-lead in Early Drug Discovery

Carles Perez Lopez*, Victor Guallar*†
*Barcelona Supercomputing Center (BSC), Barcelona, Spain
†Institucio Catalana de Recerca i Estudis Avanats (ICREA), Barcelona, Spain
E-mail: {carles.perez, victor.guallar}@bsc.es

Keywords—Drug Discovery, Fragment-based Growing, Scoring, Hit-to-lead, Free Energy, PK properties, ADMET, Ligand.

I. Extended Abstract

A. Introduction

The pharmaceutical industry has a clear need for improvement in drug design techniques due to the incremental research cost for each new drug delivered to the market. To address this aspect, computer tools play an important role in the reduction of expenses, specially in early drug discovery (EDD).

EED process can be summarized in three main steps: (1) target identification and validation, (2) hit finding and (3) hit-to-lead and lead optimization. In this last step, the main goal is to refine each hit trying to improve their efficacy, selectivity, and adequate their ADMET (Administration, Distribution, Metabolism, Excretion, and Transport) and PK properties[1].

This preclinical research is typically done using repositories of existent molecules. However, in many cases, there is a need for completely novel approaches to cure diseases where current chemical compounds have repeatedly failed. Here is when computational methods come into play.

There are several strategies to construct new molecules. One of those is the growing method, which can be split in two different approaches depending on the size of our ”bricks”: Atom-based and fragment-based.

Fragment-based have become more popular than the atom based counterpart, as indicated by the recently developed programs[2][3]. In fact, the current state of the art for this kind of software focuses on the growing part, while the evaluation of the resultant molecules is lagging behind. This is obviously due to the associated difficulties of computationally predicting accurate protein-ligand binding free energies. For this reason, an improvement of the scoring functions would be highly beneficial for the drug design community.

B. Our approach

In our lab, the Electronic and Atomic Protein Modeling lab at Barcelona Supercomputing Center, a state of the art technique for molecular simulations had been developed. The method, called Protein Energy Landscape Exploration (PELE)[4], combines a perturbation step, trough a random translation and rotation of the ligand, with a relaxation step, via protein structure prediction techniques and energy minimization. The final result of these steps is accepted or rejected according to the Metropolis algorithm. The combination of Monte Carlo sampling with protein structure prediction techniques represents a breakthrough in modeling (sampling) protein-ligand interactions, in fact, it was recently highlighted in the latest CSAR challenge (a blind benchmark for docking and scoring methods) [5] as a remarkable achievement in drug design.

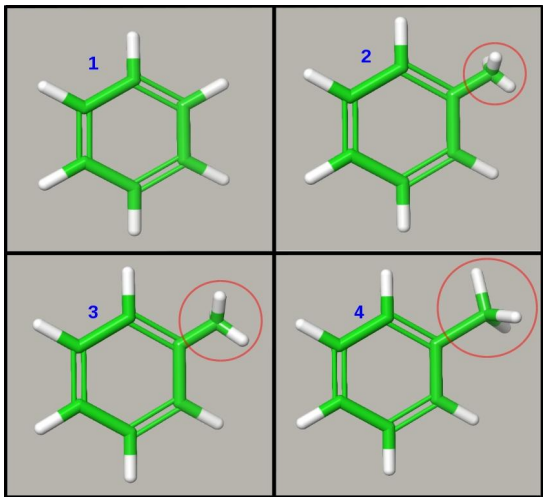

Fig. 1. A simple representation of the slow-growing schema. From a phenyl molecule, we are building a methyl-phenyl after the addition of a methyl in four steps (1-4).

In relation with the growing part, given an initial protein-ligand structure, a fragment in PDB format and the linking site, FrAG (Fragment-based Automatic Growing) is able to set up all necessary files to run PELE and use it to grow.

The whole process follows a slow-growing scheme as shown in the **Figure 1**. Our strategy is based on running successive PELE simulations at the same time that the growth is taking place. Before each simulation, FrAG computes a linear increase of Van Der Waals radius, bond length and charges modification for all atoms of the fragment to avoid major alterations in free energy. Then, once PELE has finished, FrAG analyses the results and choose the best structure that will be used as input for the next simulation.

Through this method, thanks to PELE's protein structure prediction, it is expected to find new spaces to place the molecule which would be difficult to obtain running ordinary simulations.

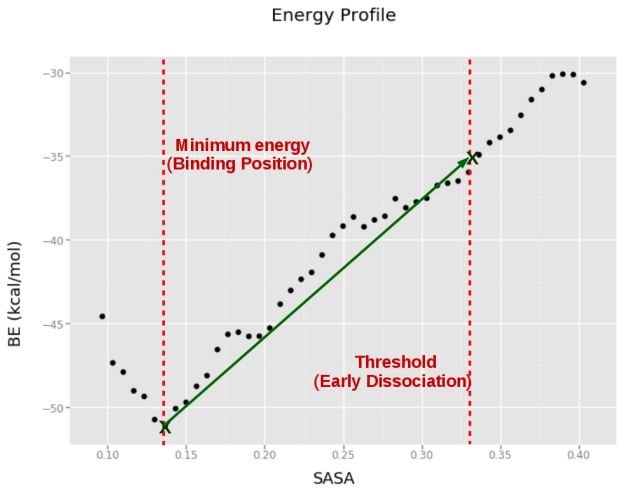

Fig. 2. Graphical representation of the scoring method. The energy profile has been generated from the data obtained during PELE simulation. The first point is identified using the minimum energy value that is considered as the binding position. The second point is obtained after defining a threshold (treated as early dissociation state) of SASA and interpolate the value of BE from the energy profile. Then, the slope between this two points is computed in order to get the score.

When the growing part has finished it is required to score the results. In accordance, it is being developed an unbinding scoring protocol with PELE based on recent work called DUck by Barril's lab[6]. In this method, we will use PELE to lead the ligand outside the binding pocket. During the unbinding, different measurements of Binding Energy (BE) and Solvent Accessible Surface Area (SASA) will be computed, and then, after analyzing the collected data a profile of SASA vs BE can be done. You can see an example in the **Figure 2**. Afterward, our score is the slope computed by the following way:

$$\frac{BEBindingPosition - BEEarlyDissociation}{SASABindingPosition - SASAEarlyDissociation} \quad (1)$$

C. Preliminary results

FrAG-PELE has been tested in a simple case of a T4 lysozyme with a benzene bonded (PDB ID: 181L), where we performed the growing until reaching a phenylethane. This result was compared with the crystallized structure (PDB ID: 1NHB) and we obtained an RMSD between ligands of 0'956.

The scoring method has been proved in a target where high-quality binding data for a series of fragment-sized ligands and the high-quality crystal structure was available[7]. Eleven simulations were performed in DNA Ligase (PDB ID: 4CC5), one for each fragment-sized ligand that we wanted to score. Finally, we set the threshold in a SASA value of 0'4 and we computed the score for the eleven ligands. A regression analysis was performed to compare experimental data (Gibbs free energy calculated) with the score obtained and we got an R-squared of 0'875.

D. Conclusions

Although that the first results are satisfactory, it is needed to perform more tests in order to further evaluate our software. In future updates, we would like to implement a pre or post-filtering method in order to take into account the synthesizability and the ADMET properties of the resultant molecules. Furthermore, we would like to automatize growing process, finding automatically which is the best position to grow the new fragment without user's intervention. When all these would be implemented, we think that FrAG-PELE could be a useful tool for hit-to-lead in EDD.

II. Acknowledgment

The authors would like to thank the cooperation in the project to all the members of Electronic and Atomic Protein Modeling group in the Barcelona Supercomputing Center and the assistance provided by some members of Nostrum Biodiscovery.

References

[1] J. P. Hughes, S. S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery," *British Journal of Pharmacology*, vol. 162, no. 6, pp. 1239–1249, 2011.

[2] Y. Yuan, J. Pei, and L. Lai, "LigBuilder 2: A Practical de Novo Drug Design Approach," *Journal of Chemical Information and Modeling*, vol. 51, no. 5, pp. 1083–1091, 2011. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ci100350u

[3] N. Chéron, N. Jasty, and E. I. Shakhnovich, "OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands," *Journal of Medicinal Chemistry*, vol. 59, no. 9, pp. 4171–4188, 2016.

[4] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, "PELE: Protein energy landscape exploration. A novel Monte Carlo based technique," *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1304–1311, 2005.

[5] H. A. Carlson, R. D. Smith, K. L. Damm-Ganamet, J. A. Stuckey, A. Ahmed, M. A. Convery, D. O. Somers, M. Kranz, P. A. Elkins, G. Cui, C. E. Peishoff, M. H. Lambert, and J. B. Dunbar, "CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma," *Journal of Chemical Information and Modeling*, vol. 56, no. 6, pp. 1063–1077, 2016.

[6] S. Ruiz-Carmona, P. Schmidtke, F. J. Luque, L. Baker, N. Matassova, B. Davis, S. Roughley, J. Murray, R. Hubbard, and X. Barril, "Dynamic undocking and the quasi-bound state as tools for drug discovery," *Nature Chemistry*, no. November, pp. 2–7, 2016. [Online]. Available: http://dx.doi.org/10.1038/nchem.2660

[7] T. B. Steinbrecher, M. Dahlgren, D. Cappel, T. Lin, L. Wang, G. Krilov, R. Abel, R. Friesner, and W. Sherman, "Accurate Binding Free Energy Predictions in Fragment Optimization," *Journal of Chemical Information and Modeling*, vol. 55, no. 11, pp. 2411–2420, 2015.

Carles Perez Lopez has his Bachelor's Degree in Biomedical Sciences in the Autonomous University of Barcelona (UAB). He is currently enrolled in his second year of Master Degree of Bioinformatics for Health Sciences in Pompeu Fabra University (UPF) and at the same time, he is doing his Master Thesis in Electronic and Atomic Protein Modeling (EAPM) group in Barcelona Supercomputing Center (BSC).