Development of corrections for the absolute free binding energy prediction

Ignasi Puch-Giner*†, Marti Municoy†‡, Victor Guallar*‡
*Barcelona Supercomputing Center, Barcelona, Spain
†Universitat Politècnica de Catalunya, Barcelona, Spain
‡Nostrum Biodiscovery, Barcelona, Spain
E-mail: ignasi.puchginer@bsc.es, mmunicoy@nostrumbiodiscovery, victor.guallar@bsc.es

Keywords - Binding energy, forcefield, SBDD, Monte Carlo

I. Extended Abstract

The early stages of drug design rely on hit discovery programs, where initial possible inhibitors' binding affinities are assessed when bound to their biological target. It is an expensive and time-consuming process, requiring multiple iterations of trial and error designs. This sets the perfect ground for computer simulations.

Structure-based drug design has been in the past decade a widely used computational methodology to speed up the drug discovery process for resolved protein-ligand systems[1]. However, providing a fast and reliable answer to the protein-ligand affinity problem can be an arduous task. In this context, the capacity of the software to score the binding affinity of the inhibitors will be crucial to determine possible drug leads that will be later on optimized.

Hence, the main goal of this research is to add physically justified corrections as well as Machine Learning models to the energetic predictions to obtain absolute binding free energies that match the experimental results. To do it we will need to review the physics involved in the forcefields used in the simulations done with the software used in the group: PELE[2].

PELE stands for Protein Energy Landscape Exploration and it is a self-contained Monte Carlo software to model protein-ligand interactions. The reachable conformations by the protein and ligand are explored and energetically assessed with the forcefield. The forcefield is the parameterized functional (eq. 1) that enables a Monte Carlo or a Molecular dynamics simulation to calculate the potential energies involved[3].

$$\begin{aligned} E_{\text{total}} &= E_{\text{bonded}} + E_{\text{nonbonded}} \\ E_{\text{bonded}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} \\ E_{\text{nonbonded}} &= E_{\text{electrostatic}} + E_{\text{van der Waals}}. \end{aligned} \tag{1}$$

This functional form does not take into account different energetic contributions that should be addressed. Right now we have considered adding correction terms regarding the strain and the conformational entropy loss of the ligand upon binding, as in eq. 2.

$$\Delta G = \Delta G_{\text{be}} + \Delta H_{\text{strain}} - T\Delta S_{\text{conf}} \tag{2}$$

A. Results

1) PELE's binding energy: First of all, we have addressed the problem of calculating the binding energy of a ligand. This affinity represents the amount of energy needed to separate the ligand from the protein.

In a PELE simulation, thousands of positions are reached each of which has a binding energy assigned with eq. 3.

$$\Delta G_{\text{PELE}} = G_{\text{protein-ligand}} - (G_{\text{protein}} + G_{\text{ligand}}). \tag{3}$$

The binding energy has been calculated for eight different protein-ligand systems with eq. 4

$$\Delta G_{\text{be}} = \frac{\sum_{i=1}^{N} \Delta G^{i}_{\text{PELE}} e^{-\beta G^{i}_{\text{tot}}}}{\sum_{i=1}^{N} e^{-\beta G^{i}_{\text{tot}}}} \tag{4}$$

where $\beta = RT$ which has units (kcal/mol)$^{-1}$ and G^{i}_{tot} is equal to the total energy of the i's conformation in kcal/mol making the exponent dimensionless.

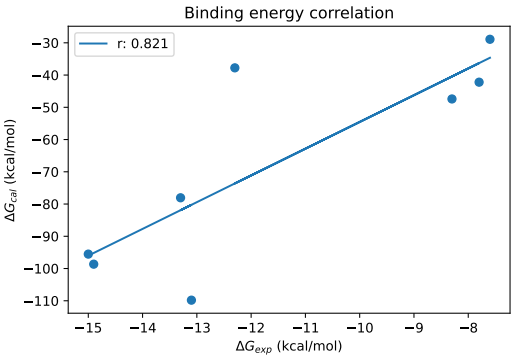

Fig. 1. Correlation between calculated and experimental binding energies. Calculated results are computed with the Boltzmann average of the binding energy of eight different PELE simulations with OPLS. Each simulation corresponds to a different protein-ligand system. (PDB codes: 1CB0[4], 1K27[5], 6QGE[6], 6QGF[6], 1HPV[7], 1HSG[8], 1MSM[9] and 1T3R)[10].

With these results, we have been able to check how well the calculated affinities correlate with the experimental data as shown in Fig. 1.

2) Ligand strain: In the second place, we have considered the strain a ligand undergoes upon its binding in the target's binding site. This energetic term is related to the change in the ligand conformation when solvated in water and its conformation when bound to the protein. In our case, we take it into account by performing two simulations: one with the isolated ligand and the other with the protein-ligand system. From the first simulation, we can obtain the minimum energy conformation of the ligand which will be its solvent conformation energy (H_{sol}). From the second we obtain the ligand energies associated with the conformations adopted inside the binding site of the protein, called bioactive conformations (H_{bc}). With eq. 5 we can associate a ligand strain to a simulation and see how well the results correlate to QM/MM calculated strains, given that ligand strains cannot be measured experimentally.

$$\Delta H_{\text{strain}} = \frac{\sum_{i=1}^{N} H_{\text{bc}}^{i} - H_{\text{sol}}}{N} > 0 \tag{5}$$

3) Ligand conformational entropy: The last correction implemented for now has been the entropic conformational loss of the ligand upon binding. Since the conformational space available is reduced, there is a toll on the free energy.

To obtain estimations of the entropic loss we have considered the dihedral angles containing the N rotatable bonds of the ligand. With this, we have been able to track the angles reached by the different dihedral angles. Then, we can perform a binning to see the occurrences of each different interval of angles and with that, assign a probability p to each of the m bins. We can associate an entropic term to each rotatable bond (rb) and, consequently, an entropic toll to the protein-ligand simulation and the ligand-in-solvent simulation as shown in eq. 6

$$\begin{aligned} S &= -R \sum_{i=1}^{N} \sum_{j=1}^{m_i} p_j \ln\left(p_j\right) \\ -T\Delta S_{\text{conf}} &= -T(S_{\text{in}} - S_{\text{s}}) \end{aligned} \tag{6}$$

where S_{in} and S_{s} are the ligand conformational entropy inside the protein and in the solvent respectively.

4) Further work: First and foremost we need good experimental datasets to be able to validate how good the predictions are. For now, the small amount (and lack) of data impedes good statistics and with that good or definitive results.

Further corrections could involve, for example, taking into account the entropic loss associated with the change in the explorable conformational space of the residues located in the protein's binding site. Another approach could be developing Machine Learning models to fit the calculated results to the experimental data. In the end, we could end up with mixed physically and Machine Learning based corrections to have a good binding free energy estimation.

References

[1] Śledź, P., & Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. Current opinion in structural biology, 48, 93-102.

[2] Borrelli, K. W., Vitalis, A., Alcantara, R., & Guallar, V. (2005). PELE: protein energy landscape exploration. A novel Monte Carlo based technique. Journal of chemical theory and computation, 1(6), 1304-1311.

[3] Frenkel, D., Smit, B., & Ratner, M. A. (1996). Understanding molecular simulation: from algorithms to applications (Vol. 2). San Diego: Academic press.

[4] Appleby, T. C., Erion, M. D., & Ealick, S. E. (1999). The structure of human 5-deoxy-5-methylthioadenosine phosphorylase at 1.7 Å resolution provides insights into substrate binding and catalysis. Structure, 7(6), 629-641.

[5] Singh, V., Shi, W., Evans, G. B., Tyler, P. C., Furneaux, R. H., Almo, S. C., & Schramm, V. L. (2004). Picomolar transition state analogue inhibitors of human 5 -methylthioadenosine phosphorylase and X-ray structure with MT-Immucillin-A. Biochemistry, 43(1), 9-18.

[6] Verteramo, M. L., Stenström, O., Ignjatovic, M. M., Caldararu, O., Olsson, M. A., Manzoni, F., ... & Akke, M. (2019). Interplay between conformational entropy and solvation entropy in protein–ligand binding. Journal of the American Chemical Society, 141(5), 2012-2026.

[7] Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., & Navia, M. A. (1995). Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. Journal of the American Chemical Society, 117(3), 1181-1182.

[8] Chen, Z., Li, Y., Chen, E., Hall, D. L., Darke, P. L., Culberson, C., ... & Kuo, L. C. (1994). Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. Journal of Biological Chemistry, 269(42), 26344-26348.

[9] Vega, S., Kang, L. W., Velazquez-Campoy, A., Kiso, Y., Amzel, L. M., & Freire, E. (2004). A structural and thermodynamic escape mechanism from a drug resistant mutation of the HIV-1 protease. Proteins: Structure, Function, and Bioinformatics, 55(3), 594-602.

[10] Surleraux, D. L., Tahri, A., Verschueren, W. G., Pille, G. M., De Kock, H. A., Jonckers, T. H., ... & Wigerinck, P. B. (2005). Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. Journal of medicinal chemistry, 48(6), 1813-1822.

Ignasi Puch-Giner received his BSc degree in Physics at the Universitat de Barcelona (UB). He holds a MSc on Computational Modeling for Physics, Chemistry and Biochemistry from UB and Universitat Politècnica de Catalunya (UPC). Since 2021 he has been in the Electronic and Atomic Protein Modeling group led by Victor Guallar in the Life Sciences department at the Barcelona Supercomputing Center. He is currently enrolled at the UPC's Computational and Applied Physics PhD programme.