On the way to real time protein-ligand sampling

Daniel Lecina-Casas[†], Ryoji Takahashi[†], Victor Guallar[†‡]
[†]Joint BSC-IRB Research Program in Computational Biology, Barcelona Supercomputing Center,
[‡]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
daniel.lecina@bsc.es

Abstract *- Protein-ligand binding free energy is one of the keystones of drug design, and developing a fast method to calculate it would have great impact in personalized medicine. However, it is a daunting task for computational methods, since the conformational space is rugged, having a lot of metastable states that hinder the exploration. Using PELE and an adaptive sampling scheme, one can quickly get thermodynamic properties by traversing the conformational space on a simulation time scale (24h). We show the performance on a new benchmark of a series of different families of proteins and ligands with a large range of binding free energy differences (about 8 kcal/mol).*

I. INTRODUCTION

Protein-ligand binding free energy is one of the keystones of drug design, since it is related to the binding affinity, and developing a fast method to calculate it would have great impact in personalized medicine. Experimental approaches are a standard way to measure thermodynamic properties, and computational methods complement well with them, since they are cheaper, easier to prepare, and give the experimenter an atomistic detail of the process of interest.

Two of the most popular computational methods are Molecular Dynamics (MD) and Monte Carlo (MC). In the first one, we explore the conformational space by numerically integrating Newton equations of motion, whereas in the second by making proposals that are accepted or rejected according to the Metropolis criterion. The protein energy landscape exploration software (PELE)[1][2], our own MC sampling technique, uses small random moves mixed with protein structure prediction algorithms to make proposals, and has been proven to accurately describe protein-ligand interactions and thermodynamics[3][4].

Typically, good estimates of the binding free energy rely on a good sampling of the relevant states. However, both MD and MC often face the problem of getting trapped in metastable minima: in biomolecule simulations there are a lot of competing interactions, which yield a rugged energy landscape that hinders the conformational exploration, oversampling some metastable states whereas undersampling others. To overcome this problem, we developed an adaptive sampling scheme that enhances the traversal of the conformational space.

In this work we will first show the results of applying standard long PELE trajectories on a set of a different proteins and ligands with a large range of binding free energy differences (about 8 kcal/mol). In the second place, we will see the sampling improvements when adaptive sampling is used. This new methodology opens a way in adding PELE and adaptive sampling in pharmaceutical drug design.

II. METHODS

1)System setup

We initialize our system with the closest protein – ligand distance being greater than 15Å, ensuring a sufficient solvent exploration. Instead of exploring the whole protein surface, the ligand is constrained to a sphere of radius ~20Å, that contains the main pocket. We use OPLS2005 with derived QM/MM charges for the ligand and implicit solvent OBC. All explicit waters are removed.

2)Simulations

Unbiased simulations are run with PELE. It combines a stochastic approach, usually called perturbation, with protein prediction algorithms, usually called relaxation. In the perturbation, the ligand is randomly moved, followed by protein backbone displacements using Cartesian coordinate anisotropic network model proposals. PELE will soon incorporate an internal coordinates normal mode analysis, which will allow smoother backbone moves. The resulting structure is relaxed by means of a side chain prediction and a global minimization with constraints on alpha carbons and the center of mass of the ligand. At the end of each iteration, the step is accepted or rejected according to the Metropolis criterion. Simulations are performed for 24h on 128-512 processors, depending on the system.

3)Analysis

In order to analyze the results, we build Markov State Models (MSM) with the ligand center of mass using EMMA[5]. MSM[6] are a methodology based on discrete master equations that allows us to calculate ΔG and obtain a coarse-grain description of the simulations, facilitating the understanding of the molecular mechanisms. We compute the binding free energy, ΔG, using:

$$\Delta G = -k_b T \ln(V_b/V_o) + \Delta W,$$

where ΔW stands for the difference of population in bulk and binding site, V_b is the binding volume and $V_o = 1661Å^3$.

4)Adaptive sampling

Adaptive sampling[7] is an iterative procedure that aims to balance the sampling of the different metastable states. We perform rounds of short simulations (e.g. 15 minutes), and clusterize all visited conformations according to RMSD. This allows us to redistribute simulations taking into account the exploration time and interest of each cluster (exploration-exploitation problem).

III.RESULTS

These are preliminary results since it is work in progress, so they may vary in the final publication.

First, we show the performance on a benchmark of a series of different families of protein and ligands with a large range of binding free energy differences (about 8 kcal/mol). As we can see in Figure 1, we can predict the tendency of binding free energies.

System (PDB id)	ΔG_{comp} (kcal/mol)	ΔG_{exp} (kcal/mol)
3ptb	-7.7	-6.7
1ecv	-9.3	-6.6
1vfn	-8.6	-7.74
1q5k	-9.6	-10.1
1b80	-12.4	-14.7

Table 1: Experimental and computational results for a benchmark of different ligands. Errors were not computed, but are estimated to be in the order of 1kcal/mol.

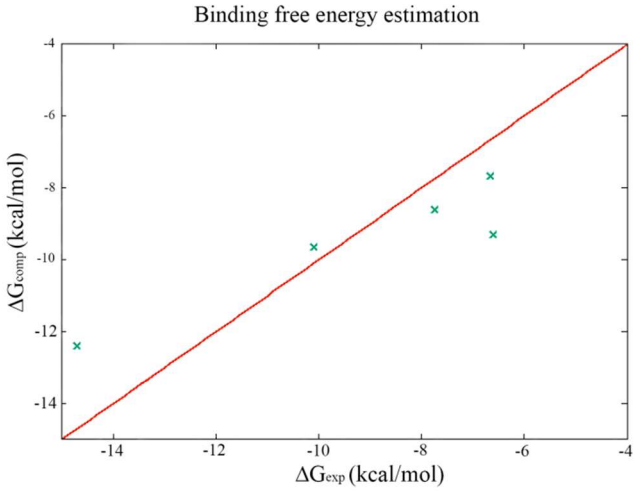
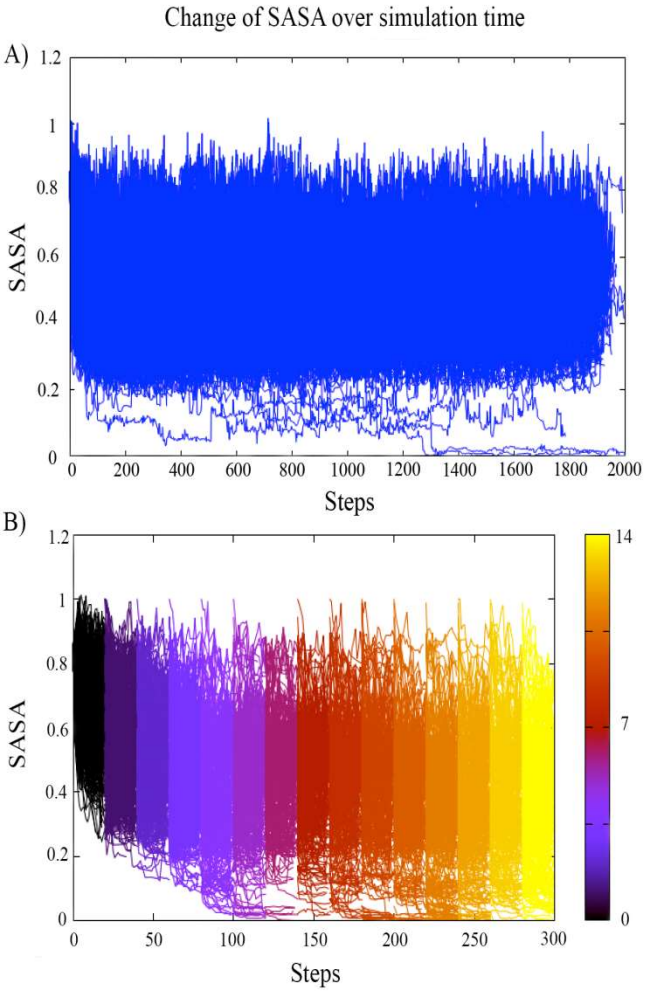


Figure 1: Ligand binding free energy computation with experimental results. Red line serves as guide to compare with exact results.

Estrogen receptors are interesting systems to study the efficiency of adaptive sampling, since we need to reproduce conformational changes in the protein in order to simulate the ligand binding. We will serve of the non-adaptive scheme as reference, and in both scenarios we will run 400 simulations. As we can see in Figure 2 panel A, in the non-adaptive scheme we sample only two binding events, the first being produced at step ~1300, while in the adaptive sampling, the first one is being produced at around step 160. The region of SASA<0.2 is much better sampled in the latter at a



fraction of the cost.

Figure 2: A) Evolution of SASA for the ligand, where 0 means totally buried and 1 means completely solvent exposed, for 400 simulations in 24h of simulation time (around 2000 steps). B) Evolution of SASA for the ligand over different 20-step epochs (color code at the right). A total amount of 15 different epochs where produced. Notice the different scales in the x-axis.

IV.CONCLUSIONS

A benchmark of different systems has been presented: from more rigid to more flexible proteins, and from small to medium sized ligands;

where PELE is able to reproduce a correct ΔG estimation using long trajectories.

Also, we showed preliminary results of applying adaptive sampling to a difficult case, hormone receptors. In this case, the sampling of the binding site improves considerably, showing binding events in an order of magnitude less of computational time.

We still need to test the adaptive sampling scheme in the ΔG benchmark, even though the first results (not shown) seem in good agreement.

The new procedure, mixing PELE sampling with adaptive sampling, seems a promising alternative to be used in a near future in pharmaceutical drug design.

ACKNOWLEDGMENT

D.L-C thanks SEV-2011-00067, awarded by the Spanish government.

REFERENCES

[1] K. W. Borrelli, A.Vitalis, R. Alcantara & V. Guallar , "PELE: Protein Energy Landscape Exploration. A novel Monte Carlo Based Technique" J. Chem. Theory Comput. 1, 1304-1311 (2005).

[2] A. Maddadkar-Sobhani & V. Guallar. "PELE web server: atomistic study of biom." Nucl. Acids Res., 41, W322-W328 (2013).

[3] R. Takahashi, V. A. Gil & V. Guallar. "Monte Carlo free ligand diffusion with Markov State Model analysis and absolute binding free energy calculations." J. Chem. Theory Comput. 10, 282-288 (2014)

[4] K. Edman, A. Hosseini, M. K Bjursell, A. Aagaard, L. Wissler, A. Gunnarsson, T. Kaminski, C. Köhler, S. Bäckström, T. J Jensen, A. Cavallin, U. Karlsson, E. Nilsson, D. Lecina, R. Takahashi, C. Grebner, S. Geschwindner, M. Lepistö, A. C Hogner, V. Guallar. "Ligand Binding Mechanism in Steroid Receptors: From Conserved Plasticity to Differential Evolutionary Constraints" Structure, 23, 2280-2290 (2015)

[5] M. Senne, B. Trendelkamp-Schroer, A. S.J.S. Mey, C. Schütte, and F. Noé. "EMMA: A Software Package for Markov Model Building and Analysis." J. Chem. Theory and Comput. 8, 2223−2238 (2012).

[6] G. R. Bowman, V. S. Pande, F. Noé, Eds. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Advances in Experimental Medicine and Biology, Springer: Heidelberg, Germany, 2014; Vol. 797.

[7] N. Hinrichs, V. Pande. Calculation of the Distribution of Eigenvalues and Eigenvectors in Markovian State Models for Molecular Dynamics. J. Chem. Phys. 126, 244101 (2007)