Entropy estimation methods in ensemble end-point binding free energy simulations

David W. Wright[1], Fouad Husseini[1], Shunzhou Wan[1], Christophe Meyer[2], Herman van Vlijmen[2], Gary Tresadern[2], and Peter V. Coveney[1,3]

[1]Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom
[2]Janssen Research & Development, Turnhoutseweg 30, B-2340 Beerse, Belgium
[3]Informatics Institute, University of Amsterdam, Netherlands

Introduction

Fragment-based lead generation (FBLG) involves scanning a library of low molecular weight compounds (fragments) to see if they bind to the target of interest and using those that do as building blocks to create create higher affinity molecules. A frequent strategy is to link multiple fragments binding different regions of the protein. FBLG represents an attractive application for *in silico* binding affinity calculations, but the need to obtain comparable values from different binding modes represents a considerable challenge for many computational techniques. We evaluate the performance of our range of ensemble simulation based binding free energy calculation protocols, called ESMACS (enhanced sampling of molecular dynamics with approximation of continuum solvent). In studies of drugs binding to a single site these protocols have been shown to produce results which correlate well with experiment (correlation coefficients $>$ 0.7) and provide robust uncertainty estimates.

ESMACS is based on the molecular mechanics Poisson–Boltzmann surface area (MMPBSA) binding affinity calculation methodology. The speed and ease of setup (compared to alchemical free energy calculations) makes MMPBSA an attractive candidate for use throughout the drug discovery pipeline. Our previous work has demonstrated that MMPBSA analysis of individual simulation runs is highly unreliable with calculations initiated from the same structures varying by up to 12 kcal mol^{-1} for small molecules and even more for flexible ligands binding to proteins. Here we investigate both the use of independent replica simulations to account for ligand and receptor flexibility, and multiple approaches to incorporating entropic contributions.

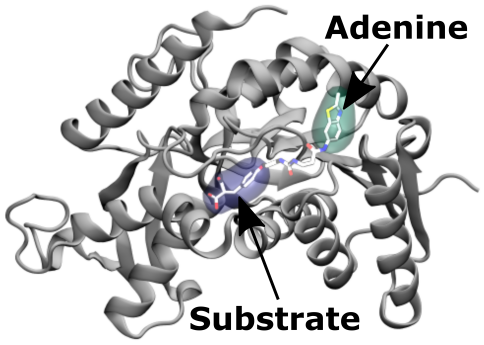

Figure 1: A representative bridging ligand bound to the LDHA protein. The adenine and substrate pockets are highlighted in green and blue. respectively.

LDHA is one component of the lactate dehydogenase (LDH) tetramer which is upregulated in clinical tumours (high expression being linked to poor prognosis). Ward *et al.* [4] reported the use of X-ray crystallography alongside surface plasmon resonance (SPR) and nuclear magnetic resonance (NMR) based screening to develop LDHA fragment hits into lead compounds. These fragments bind to two distinct locations in the protein, known as the substrate and adenine sites Figure 1). We use this dataset to evaluate the performance of ESMACS protocols when targeting multiple binding sites. It contains 22 ligands; 4 of which bind to the substrate pocket, 9 to the adenine pocket and 9 bridge the two sites. Binding strengths ranged between -11.0 and -3.1 kcal mol^{-1}.

Methods

Simulation models were built (based on PDB: 4AJP) using protocols automated by the BAC tool. AmberTools 17 was used to create neutralized water boxes with protein parameters taken from the Amber ff14SB forcefield. Drug parameterization was conducted via the RESP procedure within the Amber package and general Amber force field (GAFF). For each drug an ensemble of 25 replica MD simulations were conducted using the package NAMD 2.11 for the complex and ligand alone. Production simulations were conducted for 4 ns in the NPT ensemble with (T=300 K, P=1 bar). MMPBSA analysis was performed using AmberTools 17. We employed four protocols to investigate the influence of ligand and receptor flexibility [5]; single trajectory (1traj), single trajectory with averaged receptor contribution (1traj-ar), two trajectory with flexible ligand (2traj-fl) and two trajectory with averaged receptor (2traj-ar). Entropic contributions were calculated using three different methods - normal mode analysis (NMODE) [2], weighted solvent accessible surface area (WSAS) [3], and variational (var) entropy [1] their influence on rankings being compared. The WSAS and variational entropy approaches are computationally trivial, whereas normal mode analysis (as implemented in AmberTools) can require tens of hours of

computation.

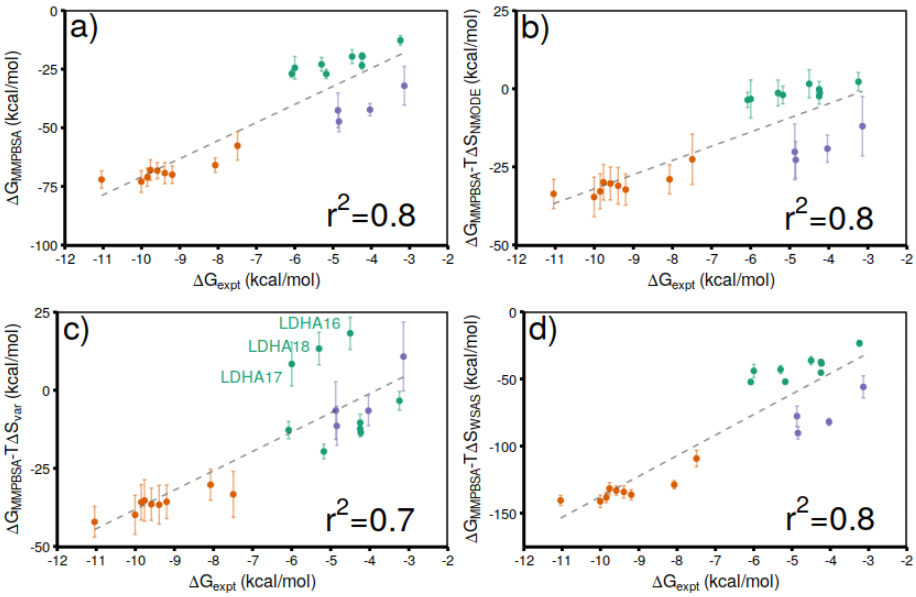

Figure 2: Comparison of ESMACS rankings combining MMPBSA (ΔG_{MMPBSA}) and different entropic contributions with experimental binding free energies for the 22 ligands in the dataset. The combinations shown are: a) MMPBSA alone, b) including normal mode entropy, c) with variational entropy, d) using WSAS. Ligands are coloured according to binding site; green for adenine pocket, purple for substrate and orange for bridging ligands that bind to both.

Results

Table 1 shows there was no difference in the statistical performance of MMPBSA alone ESMACS protocols which accounted for ligand and/or receptor flexibility compared to the standard 1traj protocol. Consequently, we combine the entropic contributions with 1traj MMPBSA derived changes in free energy.

Rankings using MMPBSA and the three different entropic contributions are shown in Figure 2. We find that MMPBSA calculations alone could separate bridging ligands that occupy two binding sites from those that bind to only a single pocket. However, the substrate and adenine pocket ligands are artificially separated in all rankings except the one including variational entropy. None of the entropic contributions improve the correlation, with r^2 of 0.80, 0.80 and 0.71 for normal modes, WSAS and variational entropy respectively. The WSAS calculations are strongly correlated with the normal mode values (r^2 of 0.93),

Table 1: Performance of different MMPBSA based ESMACS protocols in reproducing experimental binding free energies, measured by mean unsigned error (MUE), Pearson correlation coefficient (r^2) and Spearman's rank coefficient (r_s). Bootstrapped error provided in brackets where appropriate.

Protocol	**MUE***	$\mathbf{r^2}$		$\mathbf{r_s}$	
1traj	17.82	0.81	(0.04)	0.82	(0.11)
1traj-ar	22.73	0.83	(0.03)	0.81	(0.09)
2traj-fl	16.40	0.79	(0.04)	0.83	(0.11)
2traj-ar	21.21	0.81	(0.04)	0.82	(0.09)

* In kcal mol^{-1} and corrected for mean signed error.

providing the same information at a fraction of the computational cost. In the ranking including variational entropy the reduced correlation is due to three outliers, LDHA16-18, all of which have transient interactions outside the adenine pocket. Other than these ligands the variational entropy brings the values for ligands bound to the two sites into line with one another.

Conclusions

These results show that ESMACS protocols have potential to be used in FBLG applications. Additionally, they suggest that with further development variational entropy offers an effective way of combining ligands binding to multiple sites in a single ranking.

References

[1] L. Duan et al. "Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy". In: *J. Am. Chem. Soc.* 138.17 (2016). DOI: 10.1021/jacs.6b02682.

[2] S. M. Schwarzl et al. "Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction?" In: *J. Comput. Chem.* 23.12 (2002). DOI: 10.1002/jcc.10112.

[3] Junmei Wang et al. "Develop and test a solvent accessible surface area-based model in conformational entropy calculations". eng. In: *J. Chem. Inf. Model.* 52.5 (2012). DOI: 10.1021/ci300064d.

[4] R. A. Ward et al. "Design and synthesis of novel lactate dehydrogenase A inhibitors by fragment-based lead generation". In: *J. Med. Chem.* 55.7 (2012).

[5] D. W. Wright et al. "Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4". In: *Scientific reports* 9.1 (2019). DOI: 10.1038/s41598-019-41758-1.