Evaluation and Characterization of Potential SMYD3 Inhibitors

Shunzhou Wan[1], Agastya Bhati[1], David Wright[1], Peter Coveney[1], Ian Wall[2], Alan Graves[2], Darren Green[2]

[1]Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom

[2]GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, United Kingdom

In the last few years, we have developed two ensemble-based protocols for free energy calculations, termed "enhanced sampling of molecular dynamics with approximation of continuum solvent" (ESMACS) and "thermodynamic integration with enhanced sampling" (TIES). The protocols have been used to investigate drug-like small molecules bound to various therapeutic targets, including G protein-coupled receptors (GPCRs), kinases, major histocompatibility complex, etc. In this study, we investigate small molecules binding to SMYD3, a versatile lysine methyltransferase. SMYD3 is associated with multiple types of cancer, including colorectal, liver, and breast cancer.

The predicted binding free energies from the 1-trajectory ESMACS approach exhibit a high correlation with the experimental data, with a Pearson correlation coefficient of 0.84. The calculations correctly distinguish the charged compounds from the neutral ones. The charged compounds have favourable binding free energies because the charged R2 group forms favourable electrostatic interactions with a hydrophilic pocket. In TIES calculations, the overall mean unsigned error (MUE) is 1.21 kcal/mol for the entire dataset, and 0.68 kcal/mol when pairs involving one of the compounds, S10, are excluded. The mean signed errors (MSEs) are 0.62 kcal/mol and -0.06 kcal/mol for the dataset with and without S10, respectively. Both the ESMACS and TIES results show that there is a systematic deviation in the binding free energy for S10 between calculations and experimental measurements. As there are no satisfactory explanations for the disagreement between the experiments and the calculations, S10 remains as an unexplained outlier. Such unexplained outliers are not unusual in drug discovery and development projects.

We also investigate the distributions of experimentally measured free energies, and find that the distributions are highly skewed. The practical implications of this discovery are important to apprehend. Non-normal distributions imply the occurrence of more 'outliers', making it essential to perform multiple measurements to pit down average behaviour. It is also a call to exercise caution in the use of statistical methods for the comparison of experimental data and computational predictions, as the assumption of normal distributions may not or should not be applied.