Utilizing Grand Canonical Monte Carlo methods in drug discovery

Michael Steven Bodnarchuk*, Martin J. Packer and Alexe Haywood

Computational Chemistry, R&D Oncology, AstraZeneca, Cambridge, United Kingdom.

* michael.bodnarchuk@astrazeneca.com

Supplementary Information

Challenge 1: Displacing multiple waters in a binding site – a proxy for ligandability?

Preparation

An in-house structure of **1** crystallised within ATAD2 (resolution 1.95 Å) was prepared using the 'Protein Preparation Wizard' tool in Maestro. Protein and ligand protonation states assigned using PROPKA and Epik respectively at pH 7.0 ± 1. The hydrogen-bonding network was assigned for the complex, followed by a restrained minimisation using the OPLS3 forcefield to a RMSD of 0.3 Å.

GCMC simulations

GCMC simulations were performed using the ProtoMS software package (v3.4)

The protein, ligands, and water were modeled using Amber14SB, gaff14, and TIP4P force-fields, respectively. A scoop of 20 Å around the ligand was taken to reduce the system size, with side chain and backbone sampling in the inner 16 Å and side chain only in the remaining 4 Å. The complex was solvated with TIP4P water using a half-harmonically restrained sphere of radius of 30 Å. Water molecules within the GCMC region were removed prior to the simulation, whilst other crystallographic waters were retained.

A simulation box, measuring 10 Å x 9 Å x 12 Å and centered around the crystallographic ligand, was used for the GCMC runs. An initial GCMC equilibration of 5 M MC moves was performed, using a 1:1:1 ratio of insertion, deletion, and GC water sampling moves. A further 5 M equilibration moves were performed, followed by 40 M production MC moves across the entire system using the sampling rations detailed in Table **S1**.

Bulk solvent is prohibited from entering the GCMC region whilst ligand and protein atoms can fully sample the region. GCMC was performed using 59 equally-spaced B values spanning B=0.0 to B=-29.0. Data was collected over the last 20 M MC moves.

Challenge 2: Understanding selectivity through bound waters

Preparation

PDB structures of cKIT (6GQL) and KDR (6GQQ) crystallised with **2** were prepared using the 'Protein Preparation Wizard' tool in Maestro. Protein and ligand protonation states assigned using PROPKA and Epik respectively at pH 7.0 ± 1. The hydrogen-bonding network was assigned for each complex, followed by a restrained minimisation using the OPLS3 forcefield to a RMSD of 0.3 Å.

GCMC simulations

GCMC simulations were performed using the ProtoMS software package (v3.4)

The protein, ligands, and water were modeled using Amber14SB, gaff14, and TIP4P force-fields, respectively. A scoop of 20 Å around the ligand was taken to reduce the system size, with side chain and backbone sampling in the inner 16 Å and side chain only in the remaining 4 Å. The complex was solvated with TIP4P water using a half-harmonically restrained sphere of radius of 30 Å. Water molecules within the GCMC region were removed prior to the simulation, whilst other crystallographic waters were retained.

Two different simulation boxes were used for the cKIT and KDR GCMC runs. The first was a 3 Å x 3 Å x 3 Å box centered around the gatekeeper water of cKIT, whilst the second was a 6 Å x 6 Å x 6 Å box placed in proximity to the triazole portion of **2**. An initial GCMC equilibration of 5 M MC moves was performed, using a 1:1:1 ratio of insertion, deletion, and GC water sampling moves. A further 5 M equilibration moves were performed, followed by 40 M production MC moves across the entire system using the sampling rations detailed in Table **S1**. Bulk solvent is prohibited from entering the GCMC region whilst ligand and protein atoms can fully sample the region. GCMC was performed using 59 equally-spaced B values spanning B=0.0 to B=-29.0. Data was collected over the last 20 M MC moves.

Challenge 3: Predicting the effects of solvent displacement

Preparation

A PDB structure of DPP1 (6RNI) crystallised with **4** was prepared using the 'Protein Preparation Wizard' tool in Maestro. Protein and ligand protonation states assigned using PROPKA and Epik respectively at pH 7.0 ± 1. The hydrogen-bonding network was assigned

for each complex, followed by a restrained minimisation using the OPLS3 forcefield to a RMSD of 0.3 Å.

Grand Canonical Alchemical Perturbation (GCAP) simulations

GCMC simulations were performed using the ProtoMS software package (v3.4)

The protein, ligands, and water were modeled using Amber14SB, gaff14, and TIP4P force-fields, respectively. A scoop of 20 Å around the ligand was taken to reduce the system size, with side chain and backbone sampling in the inner 16 Å and side chain only in the remaining 4 Å. The complex was solvated with TIP4P water using a half-harmonically restrained sphere of radius of 30 Å. Water molecules within the GCMC region were removed prior to the simulation, whilst other crystallographic waters were retained.

The GCAP transformation from **4** to **3** was performed in two stages; the electrostatic parameters were first perturbed, followed by the van der Waals (vdW) interactions. Each simulation was split across 16 equally spaced λ windows, with perturbations performed both in the bound state and for the ligand in bulk solvent. 5 M MC equilibration steps are performed, followed by 40 M production steps. The bound state simulations were performed at a B value of -10, using a box size of 3 Å x 6 Å x 3 Å centered around the S2 pocket. Sampling ratios are shown in Table **S1**, with resultant free-energies calculated using Multistate BAR (MBAR).

Table S1: MC move rations for each simulation type performed

Simulation	Solvent	Protein	Ligand	Insertion	Deletion	GC solvent
GCMC	349	145	7	167	167	167
GCAP	349	144	7	167	167	167