INTERFACE FOCUS

royalsocietypublishing.org/journal/rsfs

Research

Cite this article: Wan S, Potterton A, Husseini FS, Wright DW, Heifetz A, Malawski M, Townsend-Nicholson A, Coveney PV. 2020 Hit-to-lead and lead optimization binding free energy calculations for G protein-coupled receptors. *Interface Focus* **10**: 20190128. http://dx.doi.org/10.1098/rsfs.2019.0128

Accepted: 6 August 2020

One contribution of 9 to a theme issue 'Computational biomedicine. Part I: molecular medicine'.

Subject Areas:
computational biology

Keywords:
molecular dynamics, free energy, binding affinity prediction, G protein-coupled receptors, adenosine receptors, ensemble simulations

Author for correspondence:
Peter V. Coveney
e-mail: p.v.coveney@ucl.ac.uk

†Should be considered joint first authors.

THE ROYAL SOCIETY PUBLISHING

Hit-to-lead and lead optimization binding free energy calculations for G protein-coupled receptors

Shunzhou Wan[1,†], Andrew Potterton[2,†], Fouad S. Husseini[1,†], David W. Wright[1], Alexander Heifetz[2,3], Maciej Malawski[4], Andrea Townsend-Nicholson[2] and Peter V. Coveney[1,5]

[1]Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, UK
[2]Institute of Structural and Molecular Biology, Research Department of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, UK
[3]Evotec (UK) Ltd, 114 Innovation Drive, Milton Park, Abingdon OX14 4RZ, UK
[4]ACK Cyfronet, AGH University of Science and Technology, Nawojki 11, 30-950, Kraków, Poland
[5]Computational Science Laboratory, Institute for Informatics, Faculty of Science, University of Amsterdam, 1098XH Amsterdam, The Netherlands

SW, 0000-0001-7192-1999; AP, 0000-0003-1001-8952; FSH, 0000-0001-8974-2622; DWW, 0000-0002-5124-8044; AT, 0000-0002-7250-2208; PVC, 0000-0002-8787-7256

We apply the hit-to-lead ESMACS (enhanced sampling of molecular dynamics with approximation of continuum solvent) and lead-optimization TIES (thermodynamic integration with enhanced sampling) methods to compute the binding free energies of a series of ligands at the A_1 and A_{2A} adenosine receptors, members of a subclass of the GPCR (G protein-coupled receptor) superfamily. Our predicted binding free energies, calculated using ESMACS, show a good correlation with previously reported experimental values of the ligands studied. Relative binding free energies, calculated using TIES, accurately predict experimentally determined values within a mean absolute error of approximately 1 kcal mol^{-1}. Our methodology may be applied widely within the GPCR superfamily and to other small molecule–receptor protein systems.

1. Introduction

There is an urgent need for approaches and tools that permit the prediction of rapid, accurate and reliable properties of systems across science as a whole. We have a longstanding interest in the development of *in silico* methodologies able to predict values computationally that agree with and therefore may replace experimental measurements [1–4]. Here, we focus our efforts on a subject of global importance in computational biomedicine: the accurate prediction of protein–small molecule binding affinities. The calculation of accurate binding affinities will provide substantial insight into ligand–receptor interactions for scientists, significantly impact the drug discovery process in industry and expedite the implementation of personalized medicine, making it more commercially attractive and facilitating the development of bespoke, individualized, pharmaceuticals to significantly improve patient prospects and bring about economic savings for healthcare programmes. The prediction of binding free energies is a computationally tractable task that, with iteration, can provide a self-reinforcing loop between experimental data and theoretical calculations.

Binding free energy can be calculated using pathway or endpoint methods. A pathway can be either a physical binding path or an alchemical path. The former is usually defined by a suitable collective variable with which simulation is driven and free energy change is derived. Large conformational space needs to be sampled along the binding path, which usually requires enhanced sampling approaches. Recently a metadynamics protocol has been

applied for ligand binding free energies to G protein-coupled receptors (GPCRs) [5,6], generating encouraging results. The approach is valuable to explore the pathways of ligand binding, to find the binding site(s), to predict the binding poses and to estimate binding free energies. However, such simulations require a timescale of microseconds and need to run for days if not weeks on high performance supercomputers. The rapid development of computational power may make it possible for these techniques to deliver actionable predictions within ten years. The metadynamics protocol, however, is unable to satisfy the requirement for pharmaceutical drug development today. Alchemical and endpoint approaches, on the other hand, are being increasingly promoted by pharmaceutical companies collectively [7] as they can be implemented in a rapid, accurate and reliable manner [8].

To accurately and reliably compute these values, it is necessary to appreciate that macromolecular biological systems are capable of adopting various conformations depending on how they are studied by simulation and that results obtained from single trajectory (one-off) simulations—particularly long ones—lack the accuracy and reproducibility needed for convergence with experimentally determined values [9]. This is only now becoming fully understood by many practitioners of molecular simulation. Despite this, single trajectory approaches that use methods such as MMPBSA (molecular mechanics Poisson Boltzmann surface area) [10], WaterMap (WM) [11], and the semi-empirical, linear interaction method (LIE) [12,13] have been used extensively. Although the benefit of using ensembles comprised of multiple simulations, or replicas, has been demonstrated [14], and applied to the calculation of free energies [15] it is only recently that new methods have been introduced which improve sampling and accessibility of the conformational space in molecular dynamics (MD)-produced trajectories [3,16–18]. One of these, ESMACS [2,14,16,17,19] (enhanced sampling of molecular dynamics with approximation of continuum solvent), uses ensembles of multiple and typically relatively short duration simulations to calculate absolute binding free energies with high precision. There is a wealth of evidence in the literature and in unpublished work that, under equilibrium conditions, multiple short MD simulations sample better than a single long MD simulation and provide a meaningful uncertainty of the results [8,9,20–22]. Under general non-equilibrium conditions, ensembles are essential since there is then no meaning to time averaging. Here, we use an ensemble approach for precise sampling of restricted regions of conformational space which are important for the calculation of the properties of interest. Many other properties, such as kinetics and transition rates, require sampling of a much larger conformational space. Long time scale simulations will be needed, usually with accelerated methods such as metadynamics. It should be noted that single long time simulations are inaccurate, as we have explained before (e.g. [22]). Ensembles are required for all MD simulations [20,22], as precise predictions, along with their uncertainties, can be obtained only when the most relevant conformations have been extensively sampled. A particular benefit of ESMACS is the freedom to choose multiple trajectory versions to enhance predictions and provide qualitative information about and insight into the associated binding mechanisms. Most publications of MMPBSA studies use 1-trajectory approach in which conformations of the complexes, proteins and ligands are all extracted from single simulations of the complexes. The multiple trajectory versions of the approach require separate simulations for complexes, proteins and ligands, and take into account the flexibility and conformational changes of the proteins and ligands upon binding. Such multiple trajectory versions of ESMACS can significantly improve the predictions compared with those from the 1-trajectory approach when 'induced fit' of a ligand is a key feature of the recognition mechanism [18,19]. Relative binding free energies in the alchemical free energy domain have also attracted significant interest, particularly for drug design and drug discovery programmes. Methods for the calculation of these values include, but are not limited to, free energy perturbation (FEP)-based approaches [23,24], which have shown some potential for predicting binding affinities at the accelerated time frames needed for drug discovery, although their accuracy remains inconclusive. We have introduced a method, TIES [25] (thermodynamic integration with enhanced sampling), that makes use of ensemble techniques to ensure reproducibility, accuracy and precision in the calculation of relative binding free energies and to control the errors associated with alchemical predictions; it compares favourably with commercially offered solutions based on FEP [8]. ESMACS can be applied to highly diverse sets of ligands [2,14,16,17,19] whereas TIES is applicable to pairs of ligands of similar chemical structure. Hence ESMACS is suitable for hit-to-lead structure identification in drug discovery, while TIES has a key role in lead optimization [2,8,25]. We emphasize that both TIES and ESMACS have the advantage of being more reproducible and reliable because of the ensemble-based approaches that both protocols use [9,20]. While this certainly increases computational cost, running ensemble simulations in parallel on powerful computers reduces the wall clock time to one relevant to timescales for drug discovery and personalized medicine [18,25,26].

Both ESMACS and TIES have been employed on a variety of globular proteins, including kinase domains of different proteins, HIV proteases, peptide-MHC, bromodomains and so on [3,8,18,19,21,25–28]. To test the accuracy of computational predictions of binding affinities on membrane proteins, we have elected to use GPCRs because of their importance to the academic community and to the pharmaceutical industry. GPCRs comprise the single biggest drug target [29] with many unexploited receptors remaining to be used for drug discovery, making the calculation of accurate binding affinities an important means by which to improve the number of drugs that successfully progress from the development pipeline to the clinic. Interestingly, and despite the wealth of published experimental data that exist for this receptor superfamily, there have been relatively few studies that report the computational prediction of the binding affinity of a ligand to its target GPCR [30–34], providing us with the opportunity to compare computational calculations of binding affinities using ESMACS and TIES with published experimental results.

Proteins, generally, and GPCRs, in particular, are dynamic and function in a complex energy landscape, possessing different conformational states and interconverting between these in response to the available free energy of the system [35,36]. They can be considered to exist 'mainly as a group of structures not too different from one another in free energy, but frequently differing considerably in energy and entropy' [37]. The conformational changes that GPCRs undergo are elicited in response to interactions between the receptor and the ligands that bind specifically to it and to interactions between the receptor and additional proteins involved in the signalling process, including G proteins. Binding affinities and receptor conformations are inextricably intertwined. The advent of

high-resolution X-ray crystal structures for GPCRs in active and inactive physiological states [38] has provided an unprecedented opportunity to examine the structural coverage of binding sites and receptor–ligand interactions [39] and affords a means by which to explore the conformational states and sub-states of these receptors, a number of which can be correlated with receptor activity. It is somewhat ironic that the experimental confirmation of the multiple active states predicted from the quantitative mathematical models of GPCRs has been provided by X-ray crystallography, a technique that emphasizes a single static macromolecule. However, the available active state GPCR X-ray crystal structures vary between crystal structures of the same GPCR and between those of different GPCRs in a manner that may be attributed either to the formation of an intermediate state which precedes the existence of a fully active state, or to the detection of one of several active structures of the GPCR [40].

We have explicitly chosen to interrogate the A_1 and A_{2A} adenosine receptors for this work as high-resolution X-ray crystal structures of both the active and inactive forms [38] are available, substantial amounts of kinetic binding data exist and these are GPCRs with which we are familiar experimentally [41–43]. Our findings are broadly applicable to other GPCRs and to other, different cell surface receptors. The automation of the ESMACS and TIES protocols within our binding affinity calculator (BAC) [44] allows the rapid generation of binding affinities for GPCRs and other receptor protein families of interest.

2. Methods

In this section, we first describe the set-up of the simulations before explaining the two methodologies used to predict binding affinity values.

2.1. Creation of receptor models

The computation of accurate binding affinities depends upon having both an accurate model of the target protein and accurately predicted poses for the ligands. GPCR structures are highly plastic, frequently adopting different conformations depending on the type of ligand to which they are bound. Three different states have been identified experimentally for the A_{2A} receptor depending on its binding partners, inactive (antagonist bound), active (also referred to as partially active, agonist bound) and fully active (in complex with a G-protein) [45,46]. The closely related A_1 receptor is believed to explore similar states but as yet structures only exist for the inactive and fully active states [47]. The G-protein complexed (fully active) form of the receptors is extremely large and we excluded it from investigation on the grounds of computational cost. In the absence of a suitable active state structure on which to base our models, we chose to simulate A_1 receptor agonists in the inactive state model (figure 1).

All available structures of the A_1 and A_{2A} adenosine receptors are incomplete; all structures contain unresolved loop regions and incorporate mutations designed to facilitate crystallization. In order to obtain complete and wild-type structures for simulation, we employed the homology modelling functionality of the Molecular Operating Environment (MOE) package. The wild-type sequences of both receptors were taken from the GPCR database (gpcrdb.org) [48]. The following PDB structures were used as templates for the modelling of the different receptor states: 4UHR (CGS21680 bound) [45], 5IU4 (ZM-241,385 bound) [49], for the active and inactive forms of the A_{2A}, and 5UEN [50] (co-crystallized with a ligand with no kinetic binding data) for the inactive form of the A_1. The complete models of the two receptors, used in the study, are shown in figure 1.

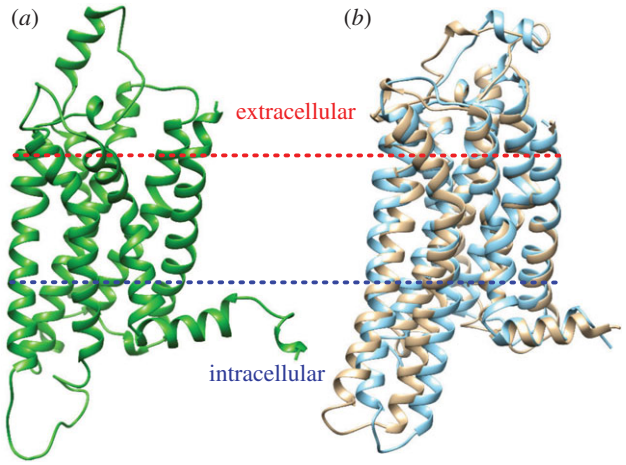

Figure 1. Structures of the (*a*) inactive A_1 (PDB accession number: 5UEN) and (*b*) inactive (beige) and active (blue) A_{2A} receptors (PDB accession numbers: 5IU4 and 4UHR, respectively) in cartoon representation.

GPCRs are membrane proteins. To ensure physiologically relevant simulations the models we have generated must be inserted into appropriate ligand membranes. Coordinate models of the membrane bound protein were generated within CHARMM [51] using a temporary CGENFF [52] parametrization for the ligands. A 100% DPPC lipid bilayer was generated around each receptor using the replacement method based on scripts adapted from the CHARMM-GUI membrane builder [53]. Each protein–membrane model was solvated in a tetragonal box containing TIP3P water molecules [54]. The ParmEd tool from AmberTools 16 [55] was then used to convert the systems to use the Amber FF14SB [56] forcefield for the protein and Lipid 14 [57] for the membrane (water remained parametrized using TIP3P). Histidines were assigned standard AMBER protonation states for a pH 7 environment. Final box dimensions for the A_1 receptor were $86 \times 86 \times 138$ Å, with 100 and 97 lipid molecules in the top and bottom layers of the membrane, respectively. Box dimensions for the inactive and active A_{2A} receptor models were $76 \times 76 \times 132$ Å and $77 \times 77 \times 132$ Å, respectively, with 76 lipid molecules in the top layer and 77 in the bottom layer of the membrane. All systems were examined to identify any water molecules trapped in the centre of the bilayer which were then removed. Counter ions were added to neutralize the simulation boxes, with 12 Cl^- and 9 Cl^- for the A_1 and A_{2A} receptors, respectively.

2.2. Molecular dynamics simulations

MD simulations were then performed using the NAMD 2.10 package [58]. Periodic boundary conditions were applied, the Particle Mesh Ewald method [59] was applied for long range electrostatics and a Lennard-Jones cut-off of 12 Å employed. The Langevin thermostat [60] was used with a low damping coefficient of 1 ps^{-1} to keep the fluctuations between the current temperature and the target temperature to a minimum. Simulations were run at 310.15 K, the human physiological temperature, to mimic the behaviour of these receptors *in vivo*. Langevin piston control [61] was used with a damping period set to 50 fs and a time-decay period of 20 fs to maintain the pressure at 1 atm. Snapshots were saved every 1 ps for all simulations. Furthermore, all covalent hydrogen bonds were constrained using the SHAKE algorithm [62].

30 000 steps of energy minimization were performed for each system using the default conjugate gradient-coupled line search algorithm. An equilibration protocol (table 1) was followed for a total of 12.5 ns during which the first 3 ns were performed in the NVT (constant number of atoms, volume and temperature) ensemble to let the lipid molecules adjust to the volume space,

Table 1. Description of the equilibration protocol and the harmonic constraints applied per step in the simulation set-up.

				harmonic constraints (kcal mol^{-1} Å^{-2})				
step	time step (fs)	ensemble	equilibration time (ns)	backbone	sidechains	lipid heads	lipid tails	ions
1	1	NVT	1	10	5	2.5	2.5	5
2	1	NVT	1	5	2.5	2.5	2.5	0
3	1	NVT	1	2.5	1	1	1	0
4	2	NPT	2	1	0.5	0.5	0.5	0
5	2	NPT	2	0.5	0.1	0.1	0.1	0
6	2	NPT	1.5	0.1	0	0	0	0
7	2	NPT	4	0	0	0	0	0

and the remaining 9.5 ns were performed in the NPT (constant number of atoms, pressure and temperature) ensemble to mimic the biological experimental setting. Velocity rescaling was performed every 500 steps while applying constraints to the backbone, side-chains, and the heavy atoms of membrane lipid heads and tails. Constraints were slowly released towards the end of equilibration as described in table 1. The last 4 ns of equilibration for each system was performed in an NPT ensemble to allow sufficient time for the complexes to relax, adjust and adopt initial stable configurations in the absence of restraints. The final frame of the simulation was used in all docking and subsequent simulations.

2.3. Ligand dataset

The ligands in the dataset (table 2 and figure 2) were chosen as they all had binding affinities determined using kinetic radioligand binding assays. The dataset is highly diverse comprising both agonists, antagonists and inverse agonists. In addition, eight of the ligands have experimental binding affinity values determined for both the A_1 and A_{2A} receptor, enabling experimentally determined receptor-relative selectivity binding free energies to be calculated.

In order to accurately compare the computational predictions of binding values to the experimental results, the experimentally determined equilibrium dissociation constants, K_D, were converted into Gibbs free energies using the equation

$$\Delta G = -RT \ln K_D. \tag{2.1}$$

The binding free energies calculated from the experimentally determined K_D for all ligands are shown in table 2. Where the same ligand had its K_D value measured in multiple publications, the average value is shown.

2.4. Ligand parametrization and docking

Crystal structures of 6 ligands in our datasets are available in complex with the A_{2A} receptor (table 2). In these cases, we use those conformations in our modelling. For the remaining ligands, the structures were manually produced and optimized in IQMol [71].

All ligands were parametrized using the Antechamber component of AmberTools 16 [72,73]. For ESMACS ligands charges were obtained using the AM1BCC approach and parametrized using the general AMBER force field (GAFF) [73]. Individual ligand topologies employed in the creation of hybrid topologies used in TIES were also parametrized with GAFF but partial charges were derived by invoking the RESP algorithm following geometry optimization in Gaussian09 [74] (employing a Hartree–Fock wavefunction with a 6 – 31 + G* basis set).

For the six ligands bound to A_{2A} for which crystal binding poses were available (table 2), the experimental crystal docking poses were retained by aligning the experimental structure with the appropriate target structure. All other ligands were docked into the binding pocket of their respective receptors using the AutoDock Vina [75] plugin in the UCSF Chimera [76] package. Single binding poses were chosen on the basis of the best docking score obtained, which showed that all of the ligands (including the antagonists) bind to the orthosteric binding site. This agrees with the experimental observation from a high number of co-crystallized structures and site-directed mutagenesis binding data. We would like to point out that in cases where a ligand may target multiple binding sites, or target an allosteric pocket, the metadynamics protocol [5,6] may be useful. In the remainder of this work, we refer to a complete parametrized model containing protein, equilibrated membrane and docked ligand as a 'starting structure'.

2.5. Binding free energy protocols

Here, we use two computational techniques to gain information about ligand binding strengths: ESMACS, which ranks absolute binding free energies (ΔG) directly, and TIES, which computes differences in Gibbs free energies between two related systems (ΔΔG). The set-up of the simulations for each protocol is substantially different and is described in detail below. Simulation execution and analysis for both protocols were automated via our BAC [44] workflow tool.

2.6. ESMACS

ESMACS protocols are designed to provide converged binding free energies calculated using the MMPBSA methodology from ensembles of relatively short duration simulations for diverse ligand datasets. They include a range of methodologies to compute the entropic contribution to binding usually neglected in standard MMPBSA approaches and may also account for ligand and receptor flexibility using multiple trajectories, including not only that of the complex but those of the unbound ligand and apo protein.

The binding free energy associated with the binding of a ligand to its target protein is calculated as follows:

$$\Delta G = \langle G^{\mathrm{Complex}} \rangle - \langle G^{\mathrm{receptor}} \rangle - \langle G^{\mathrm{ligand}} \rangle , \tag{2.2}$$

where $\langle G^{\mathrm{Complex}} \rangle$, $\langle G^{\mathrm{receptor}} \rangle$ and $\langle G^{\mathrm{ligand}} \rangle$ are the average values of the free energy contribution from complex, receptor (protein) and ligand respectively. Traditionally, in MMPBSA calculations, sampling is conducted using simulations of the complex alone.

Table 2. Table of ligands used in this study including associated experimental binding affinity data. The PDB column contains the PDB accession number of A_{2A} receptor structures from which we extract three-dimensional ligand binding poses.

				experimental binding free energies (kcal mol^{-1}) [62–69]	
ligand name	abbreviation	ligand type	PDB	A_{2A}	A_1
CGS15943	CGS	antagonist	—	−12.70 ± 0.06	−12.49 ± 0.10
LUF5834	LUF34	agonist	—	−9.77 ± 0.25	−11.53 ± 0.10
LUF5963	LUF3	antagonist	—	−8.70 ± 0.15	−10.96 ± 0.05
LUF5964	LUF4	antagonist	—	−9.28 ± 0.42	−12.59 ± 0.09
LUF5967	LUF7	antagonist	—	−8.54 ± 0.33	−12.03 ± 0.10
NECA	NECA	agonist	2YDV	−9.52 ± 0.13	−8.69 ± 0.13
theophylline	Theo	antagonist	5MZJ	−7.16 ± 0.09	−7.68 ± 0.11
XAC	XAC	antagonist	3REY	−10.11 ± 0.15	−10.86 ± 0.06
CGS21680	NGI	agonist	4UHR	−8.14 ± 0.09	—
LUF5448	LUF8	agonist	—	−8.49 ± 0.15	—
LUF5549	LUF9	agonist	—	−9.90 ± 0.14	—
LUF5550	LUF0	agonist	—	−8.84 ± 0.15	—
LUF5631	LUF1	agonist	—	−9.17 ± 0.20	—
LUF5833	LUF33	agonist	—	−9.83 ± 0.27	—
LUF5835	LUF35	agonist	—	−9.85 ± 0.26	—
UK-432,097	UK	agonist	3QAK	−10.31 ± 0.07	—
ZM-241,385	ZMA	inverse agonist	5IU4	−11.71 ± 0.09	—
LUF6057	7	agonist	—	—	−11.19 ± 0.15
CCPA	CCPA	agonist	—	—	−9.59 ± 0.10
CHEMBL3613119	119	agonist	—	—	−11.64 ± 0.10
CHEMBL3613120	120	agonist	—	—	−11.17 ± 0.22
DPCPX	DPX	inverse agonist	—	—	−12.11 ± 0.07
FSCPX	FPX	antagonist	—	—	−11.91 ± 0.14

The estimate of the component free energy provided by ESMACS can be decomposed as follows:

$$\left.\begin{aligned} G^i_{\mathrm{ESMACS}} &= G^i_{\mathrm{MMPBSA}} - TS^i_{\mathrm{conf}} \\ &= E^i_{\mathrm{MM}} + G^i_{\mathrm{solvent}} - TS_{\mathrm{conf}} \\ &= E^i_{\mathrm{int}} + E^i_{\mathrm{vdW}} + E^i_{\mathrm{ele}} + G^i_{\mathrm{PB}} + G^i_{\mathrm{SA}} - TS_{\mathrm{conf}}, \end{aligned}\right\} \quad (2.3)$$

where E^i_{MM} is the molecular mechanical energy contribution of the species i, in a *complex*, free *receptor* or unbound *ligand* in the gas phase. S_{conf} is the configurational entropy. This comprises internal bonded energies (E^i_{int}), van der Waals (E^i_{vdW}), and electrostatic interactions (E^i_{ele}). G^i_{solvent} is the solvent free energy term estimated from the sum of the Poisson–Boltzmann (G^i_{PB}) and the non-polar solvation free energy terms (G^i_{SA}). G^i_{SA} is calculated from the solvent accessibility surface area (SASA) using

$$G^i_{\mathrm{SA}} = \gamma \times \mathrm{SASA} + b, \quad (2.4)$$

where γ is the surface tension, and b the offset (we use the default values of 0.00542 kcal mol^{-1} Å^{-2} and 0.92 kcal mol^{-1} respectively [77]). The entropic term is introduced as the product of the temperature (T) and configurational entropy. The most common method of computing the configurational entropy is through normal mode analysis [78,79]. However, converging these calculations is computationally demanding for large systems [18] (potentially using as many computational resources as the original molecular dynamics calculations). This motivated the creation of an alternative solution: the weighted solvent accessible surface area (WSAS) model [19,80]. This method was parametrized to reproduce normal mode analysis results from computationally cheap atomistic surface area calculations. In this approach the solvent accessible surface area (SAS) and buried surface area (BSAS) are weighted according to atom type and the sum of the contributions of each atom is used to estimate S_{conf} as per the following relationship:

$$S^{\mathrm{WSAS}}_{\mathrm{conf}} = \sum_{i=1}^{N} w_i(\mathrm{SAS}_i - k\mathrm{BSAS}_i), \quad (2.5)$$

where w_i is the atom-type specific weighting of the atom i and k is a scaling parameter of BSAS. BSAS for atom i is computed using:

$$\mathrm{BSAS}_i = 4\pi(r_i + r_{\mathrm{prob}})^2 - \mathrm{SAS}_i, \quad (2.6)$$

where r_i is the radius of the atom i, and r_{prob} the probe radius of a water molecule. Here, we compute the surface areas using the Lee and Richards algorithm [81] as implemented in freesasa [80].

The starting structure generated for each protein–ligand was used to initiate ESMACS runs using a protocol modified from that used for previous work on globular proteins [3,16,18]. In each run 25 replica simulations were executed varying only by initial velocities, which were randomly drawn from the Maxwell–Boltzmann distribution. Each run was initialized with weak harmonic constraints (of up to 3 kcal mol^{-1} Å^{-2}) applied to the receptors' backbone and ligand's heavy atoms, which

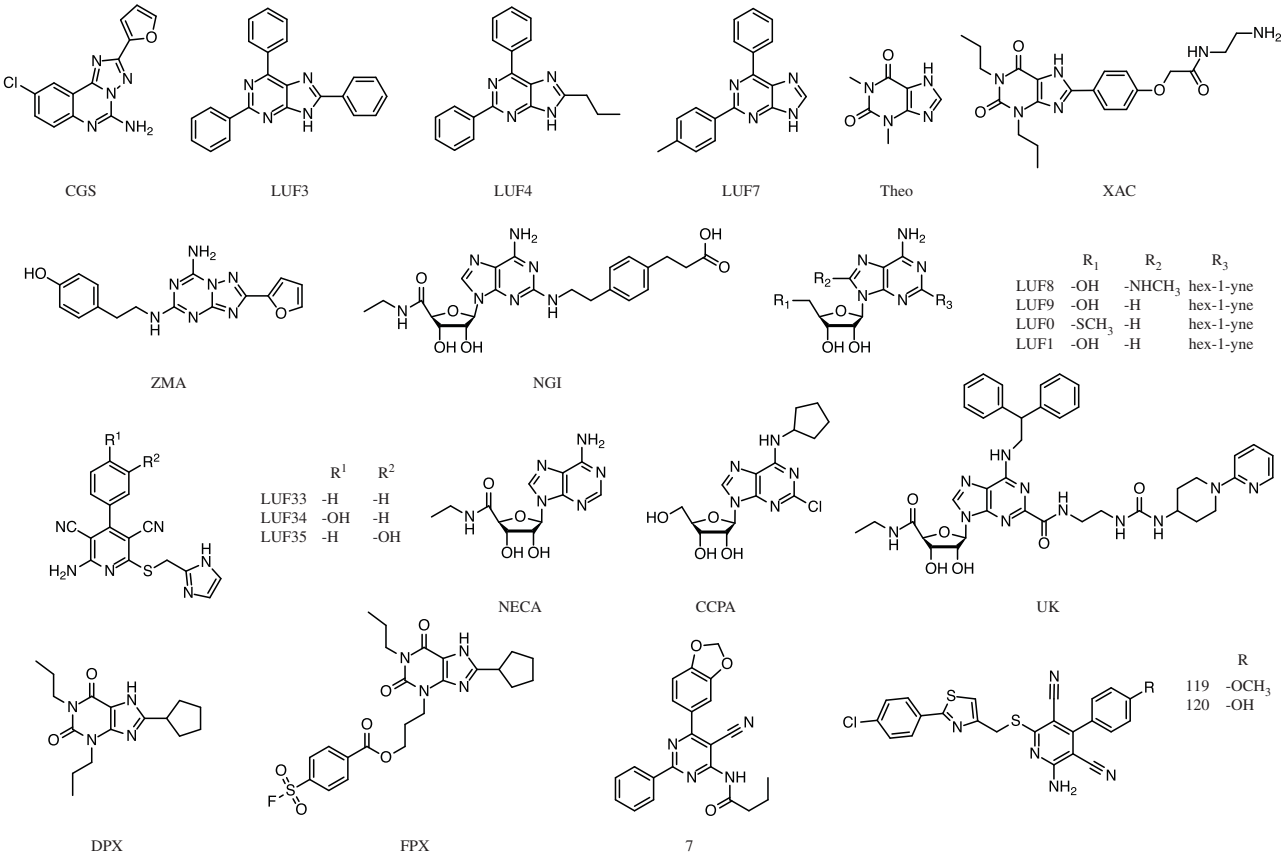

Figure 2. The structures of the ligands used in our study with the shortened names corresponding to table 2.

were slowly released during 0.5 ns of equilibration. Following this, production simulations were instigated. The same NAMD settings were used in production simulations as for the NPT steps of the membrane equilibration protocol.

Between 30 and 45 ns were required for convergence for models containing docked ligands, with slightly more rapid convergence for the ligands using poses copied from crystal structures (up to 28 ns). This is around 10 times longer than the protocol used in previous ESMACS studies and is due to the complex nature of GPCRs. The binding free energies predicted from ESMACS were based on 50 uniformly distributed frames across the last 10 ns of each of the 25 replicas and then averaged.

2.7. TIES

TIES is based on thermodynamic integration (TI), a well-established example of so-called alchemical binding free energy methods [82–85]. Alchemical free energy calculations employ unphysical (alchemical) intermediates to calculate changes in free energies between two physically real systems. It is common in these methods to refer to a variable, λ, which describes the path taken to transform one ligand into another. The parameter varies between 0 and 1, with 0 representing the initial ligand, L1, and 1 the final ligand, L2. The potential between these endpoints is given by

$$V(\lambda,x) = (1-\lambda)V_1(\lambda,x) + \lambda V_2(\lambda,x)\,, \tag{2.7}$$

where V_1 and V_2 are the potential energies of L1 and L2, respectively, and x represents the coordinates of the system. The derivative of the hybrid potential energy with respect to λ, $\partial V(\lambda,x)/\partial\lambda$, is used to compute the free energy difference using

$$\Delta G_{TI} = \int_0^1 \left\langle \frac{\partial V(\lambda,x)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda, \tag{2.8}$$

where $\langle \ldots \rangle_\lambda$ denotes the ensemble average at the chosen λ. In practice, the integral is calculated numerically, with MD sampling used in the computation of the ensemble averages at a set of discrete points (the so-called λ-windows). In TIES, multiple replica MD simulations are performed at each λ-window.

We employ a thermodynamic cycle approach to calculate relative free energy difference ($\Delta\Delta G_{\mathrm{TIES}}$) between two ligands:

$$\Delta\Delta G_{\mathrm{TIES}} = \Delta G_1 - \Delta G_2 = \Delta G_{TI}^{\mathrm{aqueous}} - \Delta G_{TI}^{\mathrm{bound}}, \tag{2.9}$$

where ΔG_1 and ΔG_2 are the binding energies for L1 and L2, respectively. $\Delta G_{TI}^{\mathrm{aqueous}}$ and $\Delta G_{TI}^{\mathrm{bound}}$ are the free-energy components resulting from the alchemical transformation of L1 to L2 in the unbound and bound states.

As described in previous work [25], treating the integrals in equation (2.9) through the lens of stochastic calculus provides a robust method to estimate uncertainties. The ensemble average of the potential derivative is calculated as the average of its values from all replica simulations in our ensemble simulation, where the individual value for each replica is taken to be the average potential derivative over the whole simulation length. The error associated with each λ-window is computed as a bootstrapped standard error of the mean of the λ derivatives from all sampled replicas using

$$\sigma^2_{x=\mathrm{aqueous,bound}} = \sum_\lambda \sigma^2_\lambda (\Delta\lambda)^2, \tag{2.10}$$

where σ^2_λ is the variance associated with the relevant λ-window in the aqueous or bound calculation, as appropriate. This error is the convolution of the individual errors for each λ window. The overall error, σ, then is computed using

$$\sigma^2 = \sigma^2_{\mathrm{aqueous}} + \sigma^2_{\mathrm{bound}}. \tag{2.11}$$

The domain of validity of TIES targets resides in determining differences in binding free energies between closely chemically related (for example congeneric) ligands between which there are no charge differences. A list of ligand pairs which meet these criteria in our experimental dataset is provided in table 3.

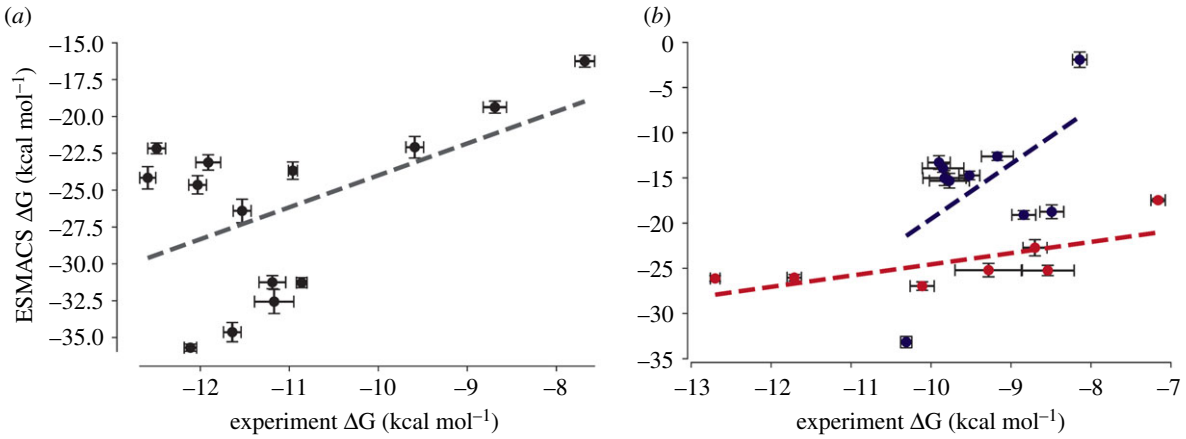

Figure 3. Summary of results obtained using ESMACS for ligands of the (*a*) A_1 and (*b*) A_{2A} receptors. The grey dashed line in (*a*) is the linear correlation line for all the ligands. For the A_{2A} receptor (*b*), agonists and antagonists are coloured blue and red, respectively. The two dashed lines in (*b*) show the linear correlation for agonists and antagonists, also coloured blue and red, respectively.

Table 3. The ligand pairs (L1 and L2) for which TIES calculations were performed in this study and their associated experimentally determined relative binding affinities (ΔΔG).

transformation		$\Delta\Delta G_{Experiment}$ (kcal mol^{-1})	
L1	L2	A_{2A}	A_1
LUF3	LUF4	−0.58 ± 0.45	−1.63 ± 0.10
LUF3	LUF7	0.16 ± 0.36	−1.07 ± 0.11
LUF4	LUF7	0.74 ± 0.53	0.56 ± 0.13
Theo	XAC	−2.95 ± 0.17	−3.18 ± 0.13
LUF8	LUF1	−0.68 ± 0.25	—
LUF34	LUF35	−0.08 ± 0.36	—
LUF33	LUF34	0.06 ± 0.37	—
LUF33	LUF35	−0.02 ± 0.37	—
NECA	CCPA	—	−0.90 ± 0.16
119	120	—	0.47 ± 0.24
XAC	DPX	—	−1.25 ± 0.09

A hybrid ligand topology must be created based on the chemically common region, a disappearing domain comprising the atoms only present in L1 and an appearing domain containing atoms unique to L2. We generated initial common region for each pair of ligands using FESetup [86] and used these to align the two ligands. Atoms were removed from the common region if their charge differed by more than 0.1*e*. The partial atomic charges for the hybrid ligand were obtained from the RESP derived partial atomic charges on the individual ligands such that the common atoms had identical charges, taken to be the average of their charges in the individual ligands. The charges on disappearing and appearing parts were then adapted by reparametrizing the ligands after constraining the charges on the common atoms to their new values.

We deployed the same basic TIES protocol set out by Bhati *et al.* [25], using 5 replica simulations in each of 13 non-interacting λ-windows, placed at: 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 1.0. We employed a soft-core potential [87,88] for the van der Waals interactions to prevent divergent potential energies, which may arise when there are sudden appearances or disappearances of atoms close to the endpoints of the alchemical transformations. The electrostatic interactions of the disappearing atoms were linearly decoupled from the simulations between λ values of 0 and 0.55 and then turned off, while those of the appearing atoms were linearly coupled from λ values of 0.45 to 1 and then fully activated.

The hybrid ligands were docked into the target structures using the same approaches employed for the single ligands within the ESMACS calculations. Each replica simulation was instigated from the starting structure and varied only by the initially randomized velocities [25]. To ensure accurate and precise values of $\Delta\Delta G_{TIES}$, a three-step equilibration process lasting 2 ns was performed at each λ window. The integration time steps over the three equilibration periods were 0.5, 1 and 2 fs, respectively. The electrostatic interaction energies were computed at every step to ensure integrator stability. As the TIES protocol only computes the difference in binding free energy between two similar ligands, 4 ns production runs were performed, as done for other systems within the TIES protocol [2,25]. All molecular dynamics simulations were performed using NAMD with the same thermostat and barostat settings as applied in ESMACS and the membrane equilibration protocol.

3. Results

In this section, we present binding affinity predictions from both ESMACS and TIES.

3.1. ESMACS

Predictions of binding free energy (ΔG), using the ESMACS protocol, were carried out on 14 and 17 ligands of the A_1 and A_{2A} receptors, respectively (structures and experimentally determined data shown in figure 2 and table 2). The results of the predicted ΔG compared to experimentally determined data are shown in figure 3. As there is an approximately 1 kcal mol^{-1} range of experimentally determined ΔG values available from different publications, it is likely that the associated errors are underreported. For the 14 A_1 receptor ligands, the correlation obtained is reasonable ($R_P = 0.53$). The overall correlation for the A_{2A} receptor ligands is slightly weaker ($R_P = 0.43$). As 'active' and 'inactive' starting structures were used to calculate ΔG values for agonists and antagonists (and inverse agonists), respectively, use of two separate correlation lines is more appropriate for the A_{2A} receptor. This results in stronger correlations of $R_P = 0.73$ for the antagonists (and the inverse agonist) and $R_P = 0.55$ for the agonists. If one includes the

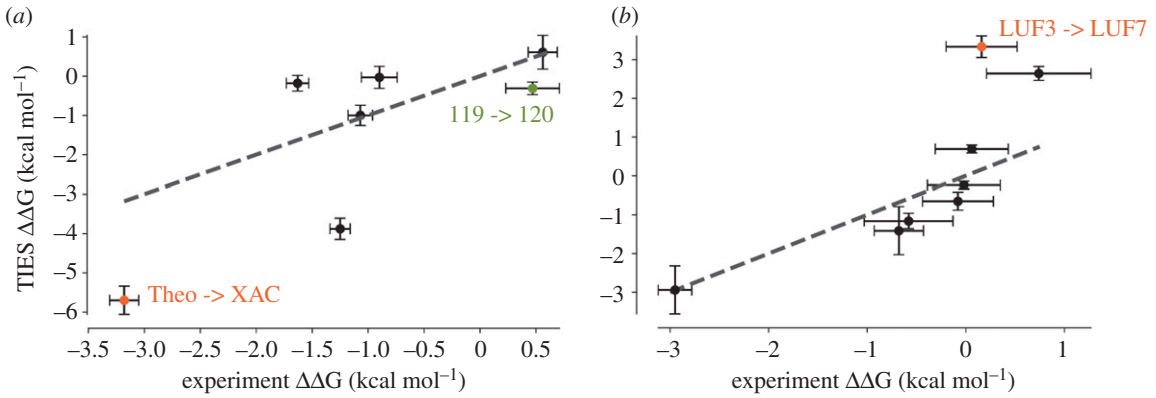

Figure 4. Plot of computed relative binding free energies using TIES (TIES ΔΔG) against experimentally determined relative binding free energies for ligand pairs of the (*a*) A_1 and (*b*) A_{2A} receptor. All reported ligand transformations are listed in table 3. Values were classified as outliers when their Cook's distance [89] was larger than $4/n$, where n represents the total number of points used in the regression. These outliers are labelled and plotted in orange. The dotted line plots $y = x$, representing correct prediction values.

configurational entropy in the ESMACS ΔG binding free energies predictions, the A_1 and A_{2A} correlations become weaker, with the exception of the A_{2A} agonists which retains the same strength in correlation ($R_P = 0.55$). Overall ESMACS performs similarly well for structurally diverse A_1 ligands, A_{2A} agonists and A_{2A} antagonists and better without inclusion of the configurational entropy.

3.2. TIES

Relative binding free energies (ΔΔG) were calculated using TIES for 7 and 8 pairs of ligands for the A_1 and A_{2A} receptor, respectively (table 3). These 15 ligand pairs were the only ligands structurally similar enough in the data (table 2) that we were able to calculate hybrid topologies. To our knowledge, this is the first published use of an alchemical binding free energy prediction method on ligands of the A_1 receptor. The results of these relative binding free energy calculations against experimentally determined data are plotted in figure 4. As TIES aims to calculate relative binding free energies, the correlation line was set as $y = x$. The overall mean absolute error was calculated as 1.2 and 0.98 kcal mol^{-1} for all ligands of the A_1 and A_{2A} receptors, respectively. This is similar to values reported previously using TIES on much simpler, non-membrane proteins [25]. All but one of the predicted ΔΔG values are directionally correct.

Furthermore, only one transformation pair's predicted ΔΔG values, from each receptor subtype, was classified as an outlier from its true value when its Cook's distance [89] was greater than $4/n$ (n being the number of points used in the regression line). These pairs are highlighted in orange in figure 4. One of the outliers, the Theo -> XAC ligand pair in the A_1 receptor, is the largest alchemical transformation performed in our TIES calculations, as these ligands are the most structurally divergent among all the ligand pairs and therefore the least reliable calculation. Excluding these outliers, the mean absolute error improves to 0.98 and 0.66 kcal mol^{-1} for remaining A_1 and A_{2A} ligand pairs, respectively. This is similar to the weighted mean absolute errors achieved using alchemical free energy calculations on two ligand series of the A_{2A} receptor [31].

4. Conclusion

Using the TIES and ESMACS protocols, we have computed the binding free energies of a series of ligands at the A_1 and A_{2A} adenosine receptors, two GPCRs for which substantial quantities of structural and functional data exist. Our rankings for binding free energies determined by both ESMACS and TIES are in line with previous experiments, confirming our ability to use these protocols on GPCRs, which are much larger protein targets than used previously with these protocols. ESMACS predicts values that correlate well with experimentally determined binding free energy values for structurally diverse sets of ligands, confirming its value for the hit-to-lead phase of drug discovery. TIES is again found to be a powerful protocol for the accurate calculation of relative binding free energies between structurally similar ligands, and is superior to methods that do not use equivalent ensemble-based sampling techniques, such as FEP+ [8]. TIES is thus of considerable value in lead optimization; here, we demonstrated this in a relatively extreme case involving a complex protein class with large alchemical transformations and correspondingly small common substructures. Although drug design has numerous constraints, we conclude that this methodology provides a useful tool with which to inform structure-based drug discovery workflows for the development of novel GPCR therapeutics.

Data accessibility. This article has no additional data.

Authors' contributions. All authors participated in the design of the study and contributed to the writing of the paper. S.W., A.P. and F.H. carried out the simulation work and performed analysis. M.M. provided assistance with simulation work performed on Prometheus. A.T.N. and P.V.C. conceived and directed the project. All authors gave final approval for publication and agree to be held accountable for the work performed herein.

Competing interests. We declare we have no competing interests.

Funding. We thank the European Commision for Horizon 2020 funding for the Computational Biomedicine Centre of Excellence (grant nos. 675451 and 823712) and VECMA (grant no. 800925), the MRC Medical Bioinformatics project (MR/L016311/1), and the UCL Provost for special funding to P.V.C. A.P. was supported by the Biotechnology and Biological Sciences Research Council (grant number BB/M009513/1) and the London Interdisciplinary Bioscience PhD Consortium (LIDo).

Acknowledgements. We thank the University College London High Performance Computing infrastructure for providing access to their GRACE and Legion platforms. We acknowledge the Leibniz Supercomputing Centre (LRZ) for providing access to SuperMUC, the Poznan Super Computing and Network Centre (PSNC) for providing access to Eagle, Cyfronet in Kraków for providing access to Prometheus, and the US National Centre for Supercomputing Applications for providing access to BlueWaters (NSF grant no. 1713749). These computers were all used to produce different parts of the computational work presented.

References

1. Altwaijry NA, Baron M, Wright DW, Coveney PV, Townsend-Nicholson A. 2017 An ensemble-based protocol for the computational prediction of helix-helix interactions in g protein-coupled receptors using coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **13**, 2254–2270. (doi:10.1021/acs.jctc.6b01246)
2. Wan S, Bhati AP, Zasada SJ, Wall I, Green D. 2017 Rapid and reliable binding affinity prediction for analysis of Bromodomain inhibitors : a computational study. *J. Chem. Theory Comput.* **13**, 784–795. (doi:10.1021/acs.jctc.6b00794)
3. Wright DW, Hall BA, Kenway OA, Jha S, Coveney PV. 2014 Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *J. Chem. Theory Comput.* **10**, 1228–1241. (doi:10.1021/ct4007037)
4. Potterton A, Husseini FS, Southey MWY, Bodkin MJ, Heifetz A, Coveney PV, Townsend-Nicholson A. 2019 Ensemble-based steered molecular dynamics predicts relative residence time of A2A receptor binders. *J. Chem. Theory Comput.* **15**, 3316–3330. (doi:10.1021/acs.jctc.8b01270)
5. Saleh N, Ibrahim P, Saladino G, Gervasio FL, Clark T. 2017 An efficient metadynamics-based protocol to model the binding affinity and the transition state ensemble of G-protein-coupled receptor ligands. *J. Chem. Inf. Model.* **57**, 1210–1217. (doi:10.1021/acs.jcim.6b00772)
6. Saleh N *et al.* 2018 Multiple binding sites contribute to the mechanism of mixed agonistic and positive allosteric modulators of the cannabinoid CB1 receptor. *Angew. Chem. Int. Ed.* **57**, 2580–2585. (doi:10.1002/anie.201708764)
7. Sherborne B *et al.* 2016 Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. *J. Comput. Aided Mol. Des.* **30**, 1139–1141. (doi:10.1007/s10822-016-9996-y)
8. Wan S, Tresadern G, Pérez-Benito L, Vlijmen H, Coveney PV. 2020 Accuracy and precision of alchemical relative free energy predictions with and without replica-exchange. *Adv. Theory Simul.* **3**, 1900195. (doi:10.1002/adts.201900195)
9. Wan S, Bhati AP, Zasada SJ, Coveney PV. 2020 Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. *Interface Focus* **10**, 20200007. (doi:10.1098/rsfs.2020.0007)
10. Kollman PA *et al.* 2000 Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897. (doi:10.1021/ar000033j)
11. Abel R, Young T, Farid R, Berne BJ, Friesner RA. 2008 Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **130**, 2817–2831. (doi:10.1021/ja0771033)
12. Hansson T, Åqvist J. 1995 Estimation of binding free energies for HIV proprotein inhibitors by molecular dynamics simulations. *Protein Eng. Des.* **8**, 1137. (doi:10.1093/protein/8.11.1137)
13. Åqvist J. 1996 Calculation of absolute binding free energies for charged ligands and effect of long-range electrostatic interactions. *J. Comp. Chem.* **17**, 1587. (doi:10.1002/(SICI)1096-987X(19961115)17:14<1587::AID-JCC1>3.0.CO;2-H)
14. Elofsson L, Nilsson L. 1993 How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized Escherichia coli thioredoxin. *J. Mol. Biol.* **233**, 766–780. (doi:10.1006/jmbi.1993.1551)
15. Mongan J, Case DA, McCammon JA. 2004 Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048. (doi:10.1002/jcc.20139)
16. Sadiq SK, Wright DW, Kenway OA, Coveney PV. 2010 Accurate ensemble molecular dynamics binding free energy of multidrug-resistance HIV-1 protease. *J. Chem. Inf. Model.* **50**, 890–905. (doi:10.1021/ci100007w)
17. Wan S, Coveney PV. 2011 Rapid and accurate ranking of binding affinities of epidermal growth factor receptor sequences with selected lung cancer drugs. *J. R. Soc. Interface* **8**, 1114–1127. (doi:10.1098/rsif.2010.0609)
18. Wan S, Knapp B, Wright DW, Deane CM, Coveney PV. 2015 Rapid, precise, and reproducible prediction of peptide – MHC binding affiinities from molecular dynamics that correlate well with experiment. *J. Chem. Theory Comput.* **11**, 3346–3356. (doi:10.1021/acs.jctc.5b00179)
19. Wan S, Bhati AP, Skerratt S, Omoto K, Shanmugasundaram V, Bagal SK, Coveney PV. 2017 Evaluation and characterization of Trk kinase inhibitors for the treatment of pain: reliable binding affinity predictions from theory and computation. *J. Chem. Inf. Model.* **57**, 897–909. (doi:10.1021/acs.jcim.6b00780)
20. Coveney PV, Wan S. 2016 On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys. Chem. Chem. Phys.* **18**, 30 236–30 240. (doi:10.1039/C6CP02349E)
21. Bhati AP, Wan S, Hu Y, Sherborne B, Coveney PV. 2018 Uncertainty quantification in alchemical free energy methods. *J. Chem. Theory Comput.* **14**, 2867–2880. (doi:10.1021/acs.jctc.7b01143)
22. Wan S, Sinclair RC, Coveney PV. 2020 Uncertainty quantification in classical molecular dynamics. (https://arxiv.org/abs/2006.07104)
23. Zwanzig RW. 1954 High-temperature equation of state by a perturbation method in non polar gases. *J. Chem. Phys.* **22**, 1420. (doi:10.1063/1.1740409)
24. Wang L, Friesner RA, Berne BJ. 2011 Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B.* **115**, 9431–9438. (doi:10.1021/jp204407d)
25. Bhati AP, Wan S, Wright DW, Coveney PV. 2017 Rapid, accurate, precise and reliable relative free energy prediction using ensemble based thermodynamic integration. *J. Chem. Theory Comput.* **13**, 210–222. (doi:10.1021/acs.jctc.6b00979)
26. Bhati AP, Wan S, Coveney PV. 2019 Ensemble-based replica exchange alchemical free energy methods: the effect of protein mutations on inhibitor binding. *J. Chem. Theory Comput.* **15**, 1265–1277. (doi:10.1021/acs.jctc.8b01118)
27. Wright DW, Husseini F, Wan S, Meyer C, van Vlijmen H, Tresadern G, Coveney PV. 2020 Application of the ESMACS binding free energy protocol to a multi-binding site lactate dehydogenase A ligand dataset. *Adv Theory Simul.* **3**, 1900194. (doi:10.1002/adts.201900194)
28. Wright DW, Wan S, Meyer C, van Vlijmen H, Tresadern G, Coveney PV. 2019 Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4. *Sci. Rep.* **9**, 6017. (doi:10.1038/s41598-019-41758-1)
29. Sriram K, Insel PA. 2018 G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs? *Mol. Pharmacol.* **93**, 251–258. (doi:10.1124/mol.117.111062)
30. Cappel D, Hall ML, Lenselink EB, Beuming T, Qi J, Bradner J, Sherman W. 2016 Relative binding free energy calculations applied to protein homology models. *J. Chem. Inf. Model.* **56**, 2388–2400. (doi:10.1021/acs.jcim.6b00362)
31. Lenselink EB *et al.* 2016 Predicting binding affinities for GPCR ligands using free-energy perturbation. *ACS Omega* **1**, 293–304. (doi:10.1021/acsomega.6b00086)
32. Mobley DL, Klimovich PV. 2012 Perspective: alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **137**, 230901. (doi:10.1063/1.4769292)
33. Boukharta L, Gutierrez-de-Teran H, Åqvist J. 2014 Computational prediction of alanine scanning and ligand binding energetics in G-protein coupled receptors. *PLoS Comput. Biol.* **10**, e1003585. (doi:10.1371/journal.pcbi.1003585)
34. Lee HS, Seok C, Im W. 2015 Potential application of alchemical free energy simulations to discriminate GPCR ligand efficacy. *J. Chem. Theory Comput.* **11**, 1255–1266. (doi:10.1021/ct5008907)
35. Frauenfelder H, Parak F, Young RD. 1988 Conformational substates in proteins. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 451–479. (doi:10.1146/annurev.bb.17.060188.002315)
36. Frauenfelder H, Sligar SG, Wolynes PG 1991 The energy landscapes and motions of proteins. *Science* **254**, 1598–1603. (doi:10.1126/science.1749933)
37. Linderstrom-Lang KU, Shellmann J. 1959 *The enzymes*. New York, NY: Academic Press.
38. Robertson N, Jazayeri A, Errey J, Baig A, Hurrell E, Zhukov A, Langmead CJ, Weir M, Marshall FH. 2011 The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* **60**, 36–44. (doi:10.1016/j.neuropharm.2010.07.001)

39. Vass M, Kooistra AJ, Yang D, Stevens RC, Wang M-W, Graaf C. 2018 Chemical diversity in the G protein-coupled receptor superfamily. *Trends Pharmacol. Sci.* **39**, 494–510. (doi:10.1016/j.tips.2018.02.004)
40. Park PS. 2012 Ensemble of G protein-coupled receptor active states. *Curr. Med. Chem.* **19**, 1146–1154. (doi:10.2174/092986712799320619)
41. Townsend-Nicholson A, Shine J. 1992 Molecular cloning and characterisation of a human brain A1 adenosine receptor cDNA. *Brain Res. Mol. Brain Res.* **16**, 365–370. (doi:10.1016/0169-328X(92)90248-A)
42. Townsend-Nicholson A, Schofield PR. 1994 A threonine residue in the seventh transmembrane domain of the human A1 adenosine receptor mediates specific agonist binding. *J. Biol. Chem.* **269**, 2373–2376.
43. Townsend-Nicholson A, Baker E, Schofield PR, Sutherland GR. 1995 Localization of the adenosine A1 receptor subtype gene (ADORA1) to chromosome 1q32.1. *Genomics.* **26**, 423–425. (doi:10.1016/0888-7543(95)80236-F)
44. Sadiq S, Wright DW, Watson SJ, Zasada S, Stoica I, Coveney PV. 2008 Automated molecular simulation based binding affinity calculator for ligand-bound HIV-1 protease. *J. Chem. Inf. Model.* **48**, 1909. (doi:10.1021/ci8000937)
45. Lebon G, Edwards PC, Leslie AGW, Tate CG. 2015 Molecular determinants of CGS21680 binding to the human adenosine A2A receptor. *Mol. Pharmacol.* **87**, 907–915. (doi:10.1124/mol.114.097360)
46. Carpenter B, Nehmé R, Warne T, Leslie AGW, Tate CG. 2016 Structure of the adenosine A2A receptor bound to an engineered G protein. *Nature* **536**, 104–107. (doi:10.1038/nature18966)
47. Draper-Joyce CJ *et al.* 2018 Structure of the adenosine-bound human adenosine A1 receptor–Gi complex. *Nature* **558**, 559–563. (doi:10.1038/s41586-018-0236-6)
48. Munk C *et al.* 2019 An online resource for GPCR structure determination and analysis. *Nat. Methods* **16**, 151–162. (doi:10.1038/s41592-018-0302-x)
49. Segala E *et al.* 2016 Controlling the dissociation of ligands from the adenosine A2A receptor through modulation of salt bridge strength. *J. Med. Chem.* **59**, 6470–6479. (doi:10.1021/acs.jmedchem.6b00653)
50. Glukhova A, Thal DM, Nguyen AT, Vecchio EA, Jörg M, Scammells PJ, May LT, Sexton PM, Christopoulos A. 2017 Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* **168**, 867–877.e13. (doi:10.1016/j.cell.2017.01.042)
51. Brooks BR *et al.* 2009 CHARMM: the biomolecular simulation program. *J. Comp. Chem.* **30**, 1545–1614. (doi:10.1002/jcc.21287)
52. Vanommeslaeghe K *et al.* 2010 CHARMM general force field (CGenFF): a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690. (doi:10.1002/jcc.21367)
53. Jo S, Lim JB, Klauda JB, Im W. 2009 CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58. (doi:10.1016/j.bpj.2009.04.013)
54. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983 Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935. (doi:10.1063/1.445869)
55. Salomon-Ferrer R, Case DA, Walker RC. 2013 An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210. (doi:10.1002/wcms.1121)
56. Maier J, Martinez C, Kasavajhala J, Wickstrom L, Hauser K, Simmerling C. 2015 ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713. (doi:10.1021/acs.jctc.5b00255)
57. Dickson CJ, Madej BD, Skjevik ÅA, Betz RM, Teigen K, Gould IR, Walker RC. 2014 Lipid14: the Amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879. (doi:10.1021/ct4010307)
58. Phillips JC *et al.* 2005 Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **26**, 1781–1802. (doi:10.1002/jcc.20289)
59. Darden T, York D, Pedersen L. 1993 The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **99**, 8345–8348. (doi:10.1063/1.465608)
60. Davidchack RL, Handel R, Tretyakov MV. 2009 Langevin thermostat for rigid body dynamics. *J. Chem. Phys.* **130**, 234101. (doi:10.1063/1.3149788)
61. Feller SE, Zhang YH, Pastor RW, Brooks BR. 1995 Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621. (doi:10.1063/1.470648)
62. Ryckaert JP, Ciccotti G, Berendsen HJC. 1977 Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341. (doi:10.1016/0021-9991(77)90098-5)
63. Guo D, Xia L, Van Veldhoven JPD, Hazeu M, Mocking T, Brussee J, Ijzerman AP, Heitman LH. 2014 Binding kinetics of ZM241385 derivatives at the human adenosine A2A receptor. *ChemMedChem.* **9**, 752–761. (doi:10.1002/cmdc.201300474)
64. Guo D, Van Dorp EJH, Mulder-Krieger T, Van Veldhoven JPD, Brussee J, IJzerman AP, Heitman LH. 2013 Dual-point competition association assay: a fast and high-throughput kinetic screening method for assessing ligand-receptor binding kinetics. *J. Biomol. Screen.* **18**, 309–320. (doi:10.1177/1087057112464776)
65. Guo D, Mulder-Krieger T, IJzerman AP, Heitman LH. 2012 Functional efficacy of adenosine A2A receptor agonists is positively correlated to their receptor residence time. *Br. J. Pharmacol.* **166**, 1846–1859. (doi:10.1111/j.1476-5381.2012.01897.x)
66. Guo D, Dijksteel GS, Van Duijl T, Heezen M, Heitman LH, Ijzerman AP. 2016 Equilibrium and kinetic selectivity profiling on the human adenosine receptors. *Biochem. Pharmacol.* **105**, 34–41. (doi:10.1016/j.bcp.2016.02.018)
67. Xia L, de Vries H, IJzerman AP, Heitman LH. 2016 Scintillation proximity assay (SPA) as a new approach to determine a ligand's kinetic profile. A case in point for the adenosine A1 receptor. *Purinergic Signal.* **12**, 115–126. (doi:10.1007/s11302-015-9485-0)
68. Louvel J, Guo D, Soethoudt M, Mocking TAM, Lenselink EB, Mulder-Krieger T, Heitman LH, Ijzerman AP. 2015 Structure-kinetics relationships of Capadenoson derivatives as adenosine A1 receptor agonists. *Eur. J. Med. Chem.* **101**, 681–691. (doi:10.1016/j.ejmech.2015.07.023)
69. McNeely PM, Naranjo AN, Forsten-Williams K, Robinson AS. 2017 A2AR binding kinetics in the ligand depletion regime. *SLAS Discov.* **22**, 166–175. (doi:10.1177/1087057116667256)
70. Guo D, Venhorst SN, Massink A, Van Veldhoven JPD, Vauquelin G, Ijzerman AP, Heitman LH. 2014 Molecular mechanism of allosteric modulation at GPCRs: insight from a binding kinetics study at the human A1 adenosine receptor. *Br. J. Pharmacol.* **171**, 5295–5312. (doi:10.1111/bph.12836)
71. Shao Y *et al.* 2006 Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **8**, 3172–3191. (doi:10.1039/B517914A)
72. Wang J, Wang W, Kollman PA, Case DA. 2006 Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247260. (doi:10.1016/j.jmgm.2005.12.005)
73. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004 Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174. (doi:10.1002/jcc.20035)
74. Frisch M *et al.* 2016 *Gaussian09, revision C.01*. Wallingford, CT: Gaussian, Inc.
75. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009 AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791. (doi:10.1002/jcc.21256)
76. Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin TE. 2004 UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. (doi:10.1002/jcc.20084)
77. Massova L, Kollman PA. 1999 Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133–8143. (doi:10.1021/ja990935j)
78. Case DA. 1994 Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **4**, 285–290. (doi:10.1016/S0959-440X(94)90321-2)
79. Tidor B, Karplus M. 1993 The contribution of cross-links to protein stability: a normal mode analysis of the configurational entropy of the native state. *Proteins* **1**, 71–79. (doi:10.1002/prot.340150109)
80. Mitternacht S. 2016 FreeSASA: an open source C library for solvent accessible surface area calculations. *F1000Research* **5**, 189. (doi:10.12688/f1000research.7931.1)
81. Lee B, Richards FM. 1971 The interpretation of protein structures: estimation of static accessibility.

J. Mol. Biol. **55**, 379–400. (doi:10.1016/0022-2836(71)90324-X)

82. Straatsma TP, Berendsen HJC, Postma JPM. 1986 Free energy of hydrophobic hydration: a molecular dynamics study of nobel gases in water. *J. Chem. Phys.* **85**, 6720–6727. (doi:10.1063/1.451846)
83. Straatsma TP, Berendsen HJC. 1988 Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulation. *J. Chem. Phys.* **89**, 5876–5886. (doi:10.1063/1.455539)
84. Straatsma TP, McCammon JA. 1991 Multiconfigurational thermodynamic integration. *J. Chem. Phys.* **95**, 1175. (doi:10.1063/1.461148)
85. Straatsma TP, McCammon JA. 1992 Computational alchemy. *Annu. Rev. Phys. Chem.* **43**, 407–435. (doi:10.1146/annurev.pc.43.100192.002203)
86. Loeffler HH, Michel J, Woods C. 2015 FESetup: automating setup for alchemical free energy simulations. *J. Chem. Inf. Model.* **55**, 2485–2490. (doi:10.1021/acs.jcim.5b00368)
87. Zacharias M, Straatsma TP, McCammon JA. 1994 Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **100**, 9025–9031. (doi:10.1063/1.466707)
88. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF. 1994 Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529–539. (doi:10.1016/0009-2614(94)00397-1)
89. Cook RD. 1975 Detection of influential observation in linear regression. *Technometrics* **19**, 15–18. (doi:10.1080/00401706.1977.10489493)