J Mol Model (2007) 13:1081–1097
DOI 10.1007/s00894-007-0229-0

ORIGINAL PAPER

Calculation of binding affinities of HIV-1 RT and β-secretase inhibitors using the linear interaction energy method with explicit and continuum solvation approaches

Niall J. English

Received: 23 January 2007 / Accepted: 26 June 2007 / Published online: 10 August 2007

Abstract The linear interaction energy (LIE) approach has been applied to estimate the binding free energies of representative sets of HIV-1 RT and β-Secretase inhibitors, using both molecular dynamics (MD) and tethered energy minimization sampling protocols with the OPLS-AA potential, using a range of solvation methodologies. Generalized Born (GB), 'shell' and periodic boundary condition (PBC) solvation were used, the latter with reaction field (RF) electrostatics. Poisson-Boltzmann (PB) and GB continuum electrostatics schemes were applied to the simulation trajectories for each solvation type to estimate the electrostatic ligand-water interaction energy in both the free and bound states. Reasonable agreement of the LIE predictions was obtained with respect to experimental binding free energy estimates for both systems: for instance, 'PB' fits on MD trajectories carried out with PBC solvation and RF electrostatics led to models with standard errors of 1.11 and 1.03 kcal mol^{-1} and coefficients of determination, r^2 of 0.76 and 0.75 for the HIV-1 RT and β-Secretase sets. However, it was also found that results from MD sampling using PBC solvation provided only slightly better fits than from simulations using shell or Born solvation or tethered energy minimization sampling.

Keywords Binding energy · Drug design · Linear interaction energy · Molecular dynamics

Electronic supplementary material The online version of this article (doi:10.1007/s00894-007-0229-0) contains supplementary material, which is available to authorized users.

N. J. English (✉)
Chemical Computing Group, St. John's Innovation Centre,
Cambridge CB4 0WS, Great Britain
e-mail: niall.english@ucd.ie

Present address:
N. J. English
School of Chemical and Bioprocess Engineering,
The Centre for Synthesis and Chemical Biology,
Conway Institute of Biomolecular and Biomedical Research,
University College Dublin,
Belfield,
Dublin 4, Ireland

Introduction

A key challenge in structure-based drug design is the accurate estimation of ligand-receptor binding free energies. A variety of techniques have been proposed to tackle this elusive problem [1], often representing in a trade-off between computational resources and accuracy. On one hand, rapid QSAR-based scoring functions are derived from fits to simplified energy terms which seek to describe the principal contributions to binding [2–4], but usually contain many approximations and typically have large errors in binding affinity predictions. At the other end of the range, the more rigorous thermodynamic integration (TI) or free energy perturbation (FEP) approaches [5–8] are more accurate but are much more computationally demanding. Recently, however, improvements in sampling have been made with the use of replica exchange thermodynamic integration (RETI) [9] and highly parallelized sampling for FEP [10]. Even so, FEP usually requires little variation in the ligand structures, and neither FEP nor TI methods are well suited to application on today's large virtual or combinatorial screening libraries in commercial rational drug design.

Åqvist and co-workers developed the linear interaction energy (LIE) method [11–14], the essential philosophy of which is to view the binding process as a replacement of

the aqueous surroundings of a free ligand with a mixed protein/solvent environment. In this approach, molecular mechanics calculations of the two end points of ligand binding (*i.e.* the ligand free in solution and bound to a protein in the presence of solvent) are combined with experimental binding affinity data to develop a model scoring function for the estimation of binding free energies. As such, LIE-based techniques may be regarded as a compromise between rapid empirical scoring functions and rigorous FEP or TI methods. The LIE method has been applied to a variety of systems and has produced promising results, with predictive errors for binding affinity on the order of 1 kcal mol^{-1} [15–31]. LIE techniques have a number of advantages relative to more rigorous FEP simulations, chiefly the need for only simulations of the two ending windows and the ability to accommodate disparate ligands. In the molecular mechanics sampling of the energy terms in the free and the bound states — either by molecular dynamics (MD), Monte Carlo or energy minimization — only the interactions between the ligand and the environment (whether aqueous or protein/water) are required for the LIE fit, in contrast to protein-protein or protein-water interactions. This reduces the extent of noise or underlying uncertainties in the calculations, leading typically to more rapid convergence of the energy terms and hence inherently faster simulations.

In the original formulation of the LIE method by Åqvist et al. [11–14], the van der Waals and electrostatic interactions were averaged over MD simulations employing explicit water for each state to allow a fit for the binding free energy ΔG_{bind} to the following form:

$$\Delta G_{\text{bind}} = \alpha \Delta U_{\text{vdW}} + \beta|_{\beta=1/2} \Delta U_{\text{elec}} \quad (1)$$

where Δ denotes $\left(\left\langle U^b \right\rangle - \left\langle U^f \right\rangle\right)$, *i.e.* the difference in averaged van der Waals and electrostatic ligand-environment interaction energies between the bound and free states. In early LIE work by Åqvist et al. [11–14], it was assumed that the linear response approximation was reasonably valid and so the β term for electrostatic interactions was set to 1/2 for the fit. Although this linear response value may be recovered for simple systems, *e.g.* small molecule solvation in water [32, 33], some steps in the binding process are unlikely to be accounted for by a simple linear model, *e.g.* removal of water molecules from the binding site. In practice, Jorgensen et al. [16] have found that optimization of the β term in a wholly empirical fit leads to better level of accuracy for LIE fits; this approach is to be favored on rational drug design projects.

Jorgensen et al. introduced a more general three-term LIE fit [15, 16]

$$\Delta G_{\text{bind}} = \alpha \Delta U_{\text{vdW}} + \beta \Delta U_{\text{elec}} + \gamma \Delta (SASA) \quad (2)$$

where inclusion of the difference in the solvent accessible surface area (SASA) of the ligand in the free and bound states was found to improve the quality of the LIE fit, based on simulations with explicit solvation. The general approach of Eq. (2) has become more common in recent years in conjunction with the use of continuum solvation, often with modifications to the SASA term to represent non-polar interactions with the solvent. Zhou et al. [26] introduced an LIE approach based on Eq. (2), in which the SASA is replaced by a cavity energy term proportional to the SASA of the ligand in each state and generalized Born (GB) solvation is used for the sampling (MD, hybrid MC or energy minimization), with adjusted GB solvation free energy components added to LIE terms. This LIE approach has been validated for 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) analogues binding to HIV-1RT [26], oligopeptide inhibitors in β-Secretase (BACE) [28], and sulfonamide inhibitors in Matrix Metalloproteinase-2 [29], with reasonable results. The use of Poisson-Boltzmann (PB) electrostatic solvation free energy estimation has also been reported by Zhou and Madura [34] in conjunction with an SASA-based non-polar energy estimate for a linear response fit to binding free energies of TIBO inhibitors in HIV-1RT, using docked poses only. However, as Carlsson et al. [31] have indicated very recently, the direct addition of a ligand's GB or PB electrostatic solvation free energy to molecular mechanics electrostatic potential energy terms is inappropriate. Furthermore, Carlsson et al. [31] have shown that the GB-splitting approach for the calculation of bound state protein-ligand interactions used by Zhou et al. [26] is inconsistent.

This work is concerned with the development of LIE models, using either PB or GB approaches to estimate the ligand-water electrostatic interaction energy terms appropriately and with more accurate convergence than for with explicit water molecules, and on the systematic study of continuum electrostatics and sampling techniques on the resultant LIE formulations. The congeneric series of HEPT analogues and oligopeptides binding to HIV-1RT [26] and BACE [28], respectively, were used. These sets were used to validate the GB-based LIE technique of Zhou et al. [26], and two of the BACE inhibitor set are charged: these represent ideal systems to develop the LIE methodology in this study and to allow for comparison with this technique. For sampling, MD is carried out with explicit solvation using both periodic boundary conditions (PBC) and a shell of water molecules around the ligand and the complex and also with GB solvation. Energy minimization is also performed in conjunction with GB solvation. Åqvist [14] noted that the use of local reaction field electrostatics [35] led to improved LIE results for charged ligands. Reaction field electrostatics are used with PBC solvation for MD sampling for all ligands and complexes to model this

effectively, while 'standard' cut-off electrostatics are applied to shell solvation.

Methods

Structures of HEPT and BACE

The structure of the set of 20 HEPT analogues are specified in Fig. 1, with the R1, R2 and R3 substituents listed in Table 1. This is identical to the set studied by Zhou et al. [26], and the experimental binding affinities were taken from the studies of Tanaka et al. [36–39] and converted to ΔG_{bind} in kcal mol^{-1} [26]. The crystal structure of HIV1-RT complexed with H11 [40] (PDB code 1RT1) served as the starting structure, and all of the other ligands' pre-docking poses were built with reference to the H11 crystal structure.

The 12 structures of the BACE test set are given in Table 2 and Fig. 2. The experimental binding free energies are those of Ghosh et al. [41], converted to kcal mol^{-1} as by Tounge and Reynolds [28]. These ligands were designed based on the crystal structure of OM99-2 (cf. Fig. 2a) bound to BACE [42] (PDB code 1FKN), a potent BACE inhibitor. Ten of the other ligands' pre-docking poses were built with reference to the OM99-2 crystal structure, while that of OM00-3 (cf. Fig. 2b) complexed with BACE was taken from the crystal structure for that system [43] (PDB code 1M4H). OM00-3 and OM99-2 are charged, while the other ten ligands are neutral.

Forcefields and electrostatics

Unless stated otherwise, all calculations were performed using the MOE software package [44]. The OPLS-AA potential [45] was used, and water molecules were treated with the TIP3P model [46], where present, as the OPLS-AA forcefield has been parameterized with this water potential. The Onufriev-Bashford-Case implementation of Generalized Born solvation was used [47], in conjunction with pairwise descreening of solute charges as described by Hawkins et al. [48], and interior and solvent dielectric constants of 1 and 78, respectively. This GB model was used since it is also reasonably suitable for macromolecules. For Born solvation, all crystallographic water molecules were stripped from the protein. During MD, energy minimization, and docking, van der Waals interactions were treated using a twin-range method [49], with a short and long cut-off radius of 12 and 15 Å, respectively. In the case of Born and explicit shell solvation, the same switching function was applied to electrostatic interactions during MD, energy minimization and docking. However, during post-sampling LIE energy analysis of the trajectory, no cut-offs were used. All charges were set using the OPLS-AA model. Although two of the BACE ligands are charged, no quantum-derived charges were used, nor were any charge equilibration methods applied to the ligand; it was found in subsequent LIE analysis that use of the forcefield charges resulted in good predictions for the charged ligands (cf. Results). Åqvist [14] has noted the importance on the charge states in the convergence of energy terms, especially in the presence of charged ligands, and that the overall charges on the systems in the free and bound ligand states should be identical. Bearing this in mind, and that there is no solvent shielding beyond the cut-off in non-PBC simulations, the charge states in charged receptor residues further than 15 Å

Table 1 Structures of HIV-1RT inhibitors with details of R1, R2 and R3 groups in Fig. 1[a]

Ligand	R1	R2	R3	ΔG_{bind} kcal mol^{-1}
H01	Me	$CH_2OCH_2CH_2OH$	SPh	−7.32
H02	Me	$CH_2OCH_2CH_2CH_3$	SPh	−7.73
H03	Me	$CH_2OCH_2CH_3$	SPh	−9.20
H04	Me	CH_2OCH_3	SPh	−8.06
H05	Me	CH_2OCH_2Ph	SPh	−10.01
H06	*i*-Pr	CH_2OCH_2Ph	SPh	−12.16
H07	Me	Et	SPh	−8.03
H08	Me	Me	SPh	−5.43
H09	Et	$CH_2OCH_2CH_3$	SPh	−10.96
H10	*i*-Pr	$CH_2OCH_2CH_3$	SPh	−11.24
H11	*i*-Pr	$CH_2OCH_2CH_3$	CH_2Ph	−11.89
H12	*c*-Pr	$CH_2OCH_2CH_3$	SPh	−9.93
H13	Me	$CH_2OCH_2CH_2OH$	CH_2Ph	−6.52
H14	Me	$CH_2OCH_2CH_2OH$	OPh	−5.78
H15	Me	$CH_2OCH_2CH_2OH$	SPh-3,5 di-Me	−9.35
H16	Et	$CH_2OCH_2CH_2OH$	SPh-3,5 di-Me	−11.19
H17	*i*-Pr	$CH_2OCH_2CH_2OH$	SPh-3,5 di-Me	−12.16
H18	Et	CH_2OCH_2Ph	SPh	−11.68
H19	Me	H	SPh	−5.11
H20	Me	Bu	SPh	−8.40

[a] The experimental activities have been taken from Tanaka et al. [36–39], and converted to ΔG_{bind} in kcal mol^{-1}, as by Zhou et al. [26]. 'Me' denotes methyl, 'Et' ethyl, 'Pr' propyl, 'Bu' butyl and 'Ph' phenyl

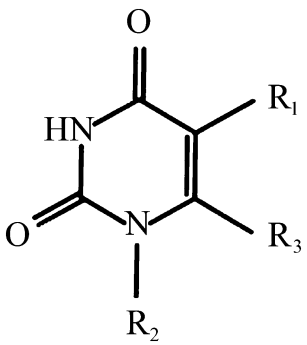
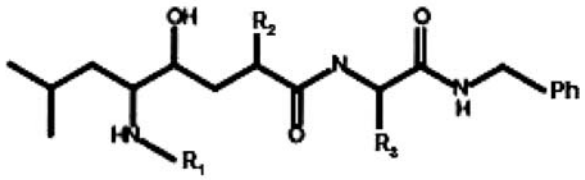
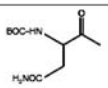
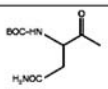
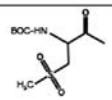
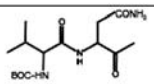
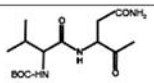
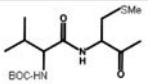
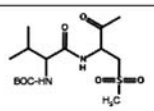
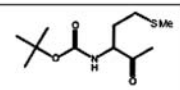
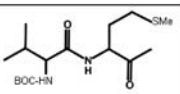
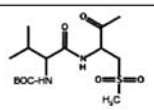


Fig. 1 Structures of HIV1-RT binders. See Table 1 for details of the R1, R2, R3 groups

Table 2 Structures of BACE inhibitors with details of R1, R2 and R3 groups in the motif below[a]



Ligand	R1	R2	R3	ΔG_{bind} kcal mol^{-1}
B01 (OM99-2)	Fig. 2a	Fig. 2a	Fig. 2a	-12.06
B02 (OM00-3)	Fig. 2b	Fig. 2b	Fig. 2b	-13.05
B03		Me	Me	-6.38
B04		Me	$CHMe_2$	-7.55
B05		Me	$CHMe_2$	-8.16
B06		Me	Me	-9.90
B07		Me	$CHMe_2$	-11.30
B08		Me	$CHMe_2$	-10.02
B09		Me	$CHMe_2$	-11.02
B10		Me	$CHMe_2$	-7.19
B11		Me	$CHMe_2$	-11.81
B12		Me	$CHMe_2$	-11.11

[a] The experimental activities have been taken from Ghosh et al. [41], and converted to ΔG_{bind} in kcal mol^{-1}, as by Tounge and Reynolds [28]. 'Ph' denotes the common fragment in OM99-2 and OM00-3 (Fig. 2a and b) by the R3 substituent and is present in all of the congeneric series, while 'Me' stands for methyl.

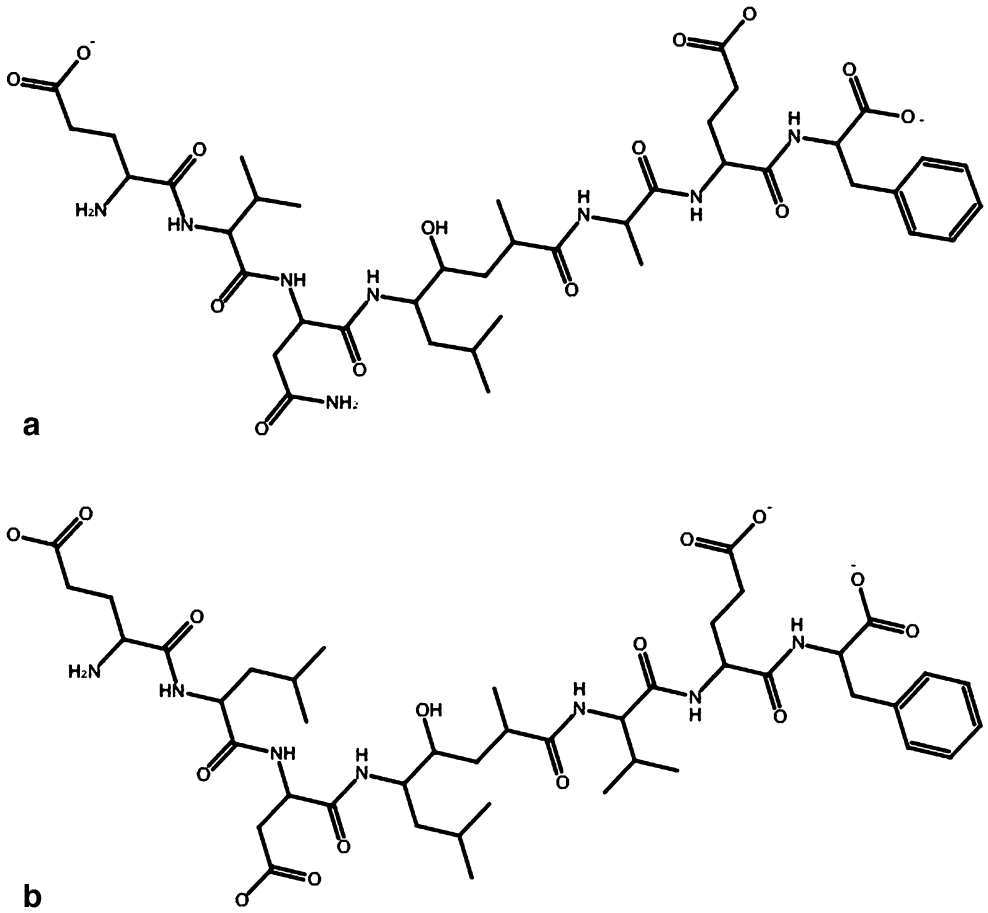


Fig. 2 Structures of BACE inhibitors: (**a**) OM99-2 and (**b**) OM00-3

from the ligand were adjusted so that the overall protein would be electroneutral.

Docking and preparation of bound state

In the initial preparation of the ligand-receptor complexes prior to docking, hydrogen atoms were added for the OPLS-AA potential and charges were set as described above. Since Born solvation was used in energy minimization and docking, all crystallographic water molecules were removed. The heavy atoms were fixed and the proton structure was relaxed by energy minimization using a composite protocol of steepest descent (SD), conjugate gradient (CG) and truncated Newton (TN) steps, with respective termination gradients of 1000, 100 and 10 kcal $mol^{-1}{\cdot}Å^{-1}$. From these structures, 25 runs of a multiple-start Monte Carlo docking procedure [50] were used to dock the HEPT analogues and BACE inhibitors (apart from the 'reference' H11 and OM99-2 crystal poses, which served to guide the initial placement of the other ligands in the congeneric series), with each run starting from the proton-relaxed configuration. Each run consisted of a sequence of 8 Monte Carlo cycles with a 'temperature' of 1000 K in the first cycle of each run. The energy function used for the MC cycles was the sum of the ligand's intramolecular energy and the van der Waals and electrostatic interaction energies. Following docking, the poses from each run with the lowest energy function value (and intramolecular energy) were chosen for further LIE analysis.

After docking, residues distant from the binding site were removed from both the proton-relaxed HIV-1RT and BACE receptors to reduce the computational cost of the simulations. This was accomplished by evaluating systematically for the bound state the ligand-environment interaction energy, using GB electrostatics, for subsets of the complex comprising of residues containing atoms within 15 to 20 Å from the ligand atoms. The Born electrostatic solvation free energy of the ligand in the bound state is

$$\Delta G^{b}_{ele,GB} = \Delta G^{b}_{comp,ele,GB} - \Delta G^{b}_{rec,ele,GB} \quad (3)$$

i.e. the Born electrostatic energy of the complex (with all charges and Born radii present) less that of the bound state receptor (*i.e.* the complex with all ligand charges turned off). Here, 'b' refers to the bound state, while 'ele' denotes electrostatic, 'comp' the entire complex and 'rec' the complex with the ligand ignored. Following the linear response approximation [31], the bound state ligand-water

interaction energy, $U^{b}_{l-w,ele,GB}$, is twice the electrostatic solvation free energy, *i.e.*

$$U^{b}_{l-w,ele,GB} = 2\Delta G^{b}_{ele,GB} \quad (4)$$

The ligand-surrounding electrostatic interaction energy, $U^{b}_{l-s,ele,GB}$, is obtained from the sum of the ligand-protein Coulombic and ligand-water electrostatic terms, *i.e.*

$$U^{b}_{l-s,ele,GB} = U^{b}_{l-p,ele} + U^{b}_{l-w,ele,GB} \quad (5)$$

In a similar manner, the bound state ligand-environment van der Waals term, $U^{b}_{l-s,vdW}$, is found from ligand-protein and ligand-water contributions, *i.e.*

$$U^{b}_{l-s,vdW} = U^{b}_{l-p,vdW} + U^{b}_{l-w,vdW} \quad (6)$$

In this case, $U^{b}_{l-w,vdW}$ is zero as no explicit water molecules are employed as residues are truncated. The total interaction energy $U^{b}_{l-s,tot,GB}$ is then found by

$$U^{b}_{l-s,tot,GB} = U^{b}_{l-s,vdW} + U^{b}_{l-s,ele,GB} \quad (7)$$

Applying the 12, 15 Å switching function, it was found that selection of HIV1-RT and BACE residues within 18 and 16 Å from the H11 and OM99-2 ligands led to interaction energies $U^{b}_{l-s,tot,GB}$ within 8 and 11 %, respectively, of those without cut-off function or residue truncation. This residue truncation with the 12, 15 Å switching function applied was deemed reasonable for subsequent MD and energy minimization calculations, and resulted in 168 HIV-1RT residues (2782 atoms) and 230 BACE residues (3611 atoms), in comparison with full receptors of 954 and 391 residues, respectively.

Explicit solvation

In the case of MD simulations with explicit water under PBC, the reaction field method [51] was used to handle long-range electrostatics, with a cut-off radius of 15 Å. Either the ligand or post-docking truncated complex was placed at the center of a rectangular periodic box of dimensions such that no heavy atoms were lying within less than approximately 8 Å from the wall in any Cartesian direction, to which water molecules were added based on the relaxed TIP3P liquid state at 298 K and 1 atm. Water molecules were placed such that their oxygen atoms were greater than 2.8 Å in distance from the heavy atoms. Prior to 'production' MD for LIE sampling, the dielectric constant in the reaction field method was optimized to minimize the RMS error in the calculation of electrostatic forces for relaxed systems of both HIV-1RT and BACE at 298 K with respect to the 'definitive' Lekner technique [52], implemented by an efficient interpolation scheme [53] using in-house Fortran code, as outlined by English et al. [54]. It was found that reaction field dielectric constants of 68.3 and 74.4 led to RMS force deviations of 24.5 and 18.3 pN for the HIV-1RT (H11) and BACE (OM99-2) complexes, respectively. In the free state, dielectric constants of 72.5 and 82.1 resulted in RMS deviations of 11.7 and 13.2 pN for H11 and OM99-2 ligands, respectively. These dielectric constants were used for the reaction field model in the respective congeneric series. To gauge if the (x, y, z) 'clearance' of around 8 Å relative to the box edge is sufficient, the average electrostatic interaction energy with the water was calculated using the optimized reaction field method for ten relaxed configurations of the HIV-1RT (H11) and BACE (OM99-2) systems at 298 K (*i.e.* with more solvation water molecules and larger 'clearances' of about 11 and 14 Å). The magnitudes of the electrostatic interaction energies were 7 and 10 % greater for HIV-1RT for the systems with clearances of 11 and 14 Å, respectively, with corresponding results of 9 and 13 % for BACE. Therefore, it was considered that the 8 Å clearance is reasonable.

In the case of explicit shell solvation for MD sampling, a layer of water molecules was added to envelop the free ligand and the complex. The layer was of constant thickness of approximately 15 Å, the water configuration based on the same relaxed TIP3P liquid state, using the 2.8 Å distance restriction for water oxygen atoms from the heavy atoms. The layer thickness was set so that this would be greater than the range of the switching function applied to the van der Waals and Coulombic interactions. To prevent water migration into the vacuum during MD, all water oxygen atoms in the layer's outer 3 Å were restrained with a harmonic restraint force of 50 kcal $mol^{-1}\cdot Å^{-2}$ for displacements greater than 0.1 Å from their initial position. Although the use of harmonic tethers is quite common [14, 21, 31], this results in a variable pressure, as do PBC simulations in the NVT ensemble; the rationale in this study is to evaluate the use of popular solvation schemes on LIE model performance.

Tethered energy minimization and dynamics

A composite protocol of energy minimization was carried out for both the free and bound states. In the case of energy minimization as a sampling tool for LIE, Born solvation was used, but the protocol below was also used for explicit shell and PBC solvation without restraints on the water molecules prior to MD sampling. In the free state, two-stage minimization was performed, in which SD, CG and TN methods were applied in sequence with termination gradients of 1,000, 100 and 1 kcal $mol^{-1}\cdot Å^{-1}$, respectively. In the first stage, a harmonic restraint force of 20 kcal $mol^{-1}\cdot Å^{-2}$ was applied to the heavy ligand atoms for displacements greater than 0.1 Å from their initial position, while restraints were removed in the second stage. In the bound state, the TN termination gradients were reduced

progressively in a four-stage process. This avoided unnecessary minimization to excessively low TN gradients (e.g. 1 to 5 kcal $mol^{-1}\cdot Å^{-1}$) when restraints were in place. Firstly, all heavy atoms were fixed and minimization applied to a TN gradient of 100 kcal $mol^{-1}\cdot Å^{-1}$. Secondly, a 500 kcal $mol^{-1}\cdot Å^{-2}$ harmonic restraint force beyond 0.1 Å was applied to receptor heavy atoms to a TN gradient of 50 kcal $mol^{-1}\cdot Å^{-1}$. Thirdly, all atoms in receptor residues containing atoms within 15 Å of the ligand were designated part of the 'binding region', and any heavy atoms therein were restrained with a 100 kcal $mol^{-1}\cdot Å^{-2}$ force beyond 0.1 Å with a 500 kcal $mol^{-1}\cdot Å^{-2}$ force for heavy atoms outside this region; a TN gradient of 10 kcal $mol^{-1}\cdot Å^{-1}$ was used in this case. Finally, the heavy atoms inside and outside the binding region were restrained with 20 and 100 kcal $mol^{-1}\cdot Å^{-2}$ forces beyond 0.1 Å, respectively, and a TN gradient of 1 kcal $mol^{-1}\cdot Å^{-1}$ used.

If MD sampling is used, a simulation was carried out in the NVT ensemble after the energy minimization described above and bond lengths were constrained with a relative tolerance of 10^{-8} [55]. A timestep of 1 fs was used. This was assessed for energy conservation in the NVE ensemble and was found to be satisfactory: the percentage relative drift in energy, defined as the ratio of the energy drift (expressed as a linear regression coefficient) to the average kinetic energy during the simulation [56], was less than 0.2 % over 50 ps for a relaxed PBC system HIV-1RT (H11) system. The period of the thermal reservoir [55] was set to 1 ps, to allow for a relatively weak coupling. Prior to production simulations, the system was 'heated' to 300 K in 25 K increments by MD in steps of 5 ps duration in the NVT ensemble, using velocity assignment from the Maxwell-Boltzmann distribution at the start of each step. For the bound state, the tether weights on the heavy atoms' restraint forces inside and outside the binding region were reduced at a constant rate at the start of each step, until they were zero for the 'production' simulation at 300 K. For explicit shell or PBC solvation, the system was then relaxed for a further 150 ps at 300 K, or for 30 ps in the case of Born solvation, during which the total system energy and the ligand-environment interaction energy were found to stabilize for both the free and bound cases.

Poisson-Boltzmann electrostatics

In analysis of the MD trajectories (with water molecules removed) or of the Born solvation post-minimization configuration, the Poisson-Boltzmann (PB) equation for the electrostatic potential field u was applied to n ions of charge q

$$\nabla\cdot[d(\mathbf{r})\nabla u(\mathbf{r})]+\sum_{i=1}^{n}q_iC_i\exp\left(-q_iu(\mathbf{r})/kT\right)+f(\mathbf{r})=0 \tag{8}$$

where C are the ion concentrations and f is the charge density, using the method of Grant et al. [57]. Here, the relative dielectric field d is treated as a sum of Gaussians to describe the variation between the solute and bulk solvent dielectric constants. The interior and solvent dielectric constants were set to 1 and 78, respectively. To reduce computational cost, especially for application to complexes and isolated receptor configurations, the linear form of Eq. (8), *i.e.* using $\exp x\approx(1+x)$, was solved with a multi-grid preconditioned gradient algorithm on a grid constructed with a spacing of 0.75 Å and an extent of 5 Å in each Cartesian direction from the extremities of the biomolecule. This allowed the PB estimation of the ligand's electrostatic solvation free energy in its free and bound states, $\Delta G^{\mathrm{f}}_{\mathrm{ele,PB}}$ and $\Delta G^{\mathrm{b}}_{\mathrm{ele,PB}}$, respectively, *e.g.* for the bound state it is computed exactly as in the Born case of Eq. (3):

$$\Delta G^{\mathrm{b}}_{\mathrm{ele,PB}}=\Delta G^{\mathrm{b}}_{\mathrm{comp,ele,PB}}-\Delta G^{\mathrm{b}}_{\mathrm{rec,ele,PB}} \tag{9}$$

i.e. the electrostatic solvation free energy of the complex ('comp' — with all charges present) less that of the bound state receptor (*i.e.* 'rec' — the complex with all ligand charges switched off).

Calculation of LIE terms

Following the linear response approximation [31], the ligand-water electrostatic interaction energy in either the free and bound states, $U^{\mathrm{f}}_{\mathrm{l-w,ele,M}}$ and $U^{\mathrm{b}}_{\mathrm{l-w,ele,M}}$, respectively (where 'f' denotes free, 'b' bound and 'M' the PB or GB method) is twice the continuum electrostatic solvation free energy, *i.e.*

$$U^{\mathrm{f}}_{\mathrm{l-w,ele,M}}=2\Delta G^{\mathrm{f}}_{\mathrm{ele,M}}\quad\text{or}\quad U^{\mathrm{b}}_{\mathrm{l-w,ele,M}}=2\Delta G^{\mathrm{b}}_{\mathrm{ele,M}} \tag{10}$$

In the case of explicit solvation, the water molecules are simply ignored in the MD trajectories for calculation of $\Delta G^{\mathrm{f}}_{\mathrm{ele,M}}$ and $\Delta G^{\mathrm{b}}_{\mathrm{ele,M}}$ with both PB and GB techniques. In the free state, the ligand-water van der Waals term $U^{\mathrm{f}}_{\mathrm{l-w,vdW}}$ is based on interactions with the water molecules without cut-off; for minimization sampling using GB electrostatics, this term will be zero due to the absence of any water molecules. In the free state, this ligand-water $U^{\mathrm{f}}_{\mathrm{l-w,vdW}}$ term coincides with the 'ligand-surrounding' ('l-s' or 'ligand-environment') $U^{\mathrm{f}}_{\mathrm{l-s,vdW}}$ term. In the bound state, the ligand-protein ('l-p') van der Waals and Coulombic interaction terms, $U^{\mathrm{b}}_{\mathrm{l-p,vdW}}$ and $U^{\mathrm{b}}_{\mathrm{l-p,ele}}$, respectively, are evaluated without cut-off (*i.e.* ignoring any water molecules). The ligand-water ('l-w') van der Waals term, $U^{\mathrm{b}}_{\mathrm{l-w,vdW}}$, is also computed without cut-off; this term is zero for GB minimization sampling where water molecules are absent.

As in Eq. (5), the ligand-surrounding ('l-s') electrostatic interaction energies in the free and bound states, $U^{f}_{l-s,ele,M}$ and $U^{b}_{l-s,ele,M}$, are obtained as

$$U^{f}_{l-s,ele,M} = U^{f}_{l-w,ele,M} \quad (11)$$

$$U^{b}_{l-s,ele,M} = U^{b}_{l-p,ele} + U^{b}_{l-w,ele,M} \quad (12)$$

from the sum of the ligand-protein Coulombic and ligand-water electrostatic terms for either continuum electrostatics method 'M'. In a similar manner to Eq. (6), the ligand-environment ('l-s') van der Waals term in each state, $U^{f}_{l-s,vdW}$ and $U^{b}_{l-s,vdW}$, is found as

$$U^{f}_{l-s,vdW} = U^{f}_{l-w,vdW} \quad (13)$$

$$U^{b}_{l-s,vdW} = U^{b}_{l-p,vdW} + U^{b}_{l-w,vdW} \quad (14)$$

These terms were computed every 0.5 ps over the production phase of MD runs, and canonical averages were taken, denoted by $\langle\ \rangle$. This allowed the LIE interaction energy terms, $\langle \Delta U_{vdW} \rangle$ and $\langle \Delta U^{M}_{ele} \rangle$, to be computed as the difference between the averaged ligand-surrounding interaction energies in the bound and free states:

$$\langle \Delta U_{vdW} \rangle = \left\langle U^{b}_{l-s,vdW} \right\rangle - \left\langle U^{f}_{l-s,vdW} \right\rangle \quad (15)$$

$$\langle \Delta U^{PB}_{ele} \rangle = \left\langle U^{b}_{l-s,ele,PB} \right\rangle - \left\langle U^{f}_{l-s,ele,PB} \right\rangle \quad (16)$$

$$\langle \Delta U^{GB}_{ele} \rangle = \left\langle U^{b}_{l-s,ele,GB} \right\rangle - \left\langle U^{f}_{l-s,ele,GB} \right\rangle \quad (17)$$

In the case of GB minimization sampling, there is no averaging over configurations and the post-minimization positions are used. The LIE fit is then carried out for both PB and GB cases according to

$$\Delta G_{bind} = \alpha_M \langle \Delta U_{vdW} \rangle + \beta_M \langle \Delta U^{M}_{ele} \rangle \quad (18)$$

A three-component fit was not used in Eq. (18), although this has been used by other workers [15, 16, 26, 28, 29]: it is considered that the size of the HIV-1RT and BACE series (20 and 12, respectively) is not sufficiently large to justify more than two LIE energy terms.

As a typical example, time requirements for the sampling and continuum electrostatics analysis calculations are given in Table 3 for the HIV1-RT inhibitor set, relative to the time needed for energy minimization sampling with Born solvation (the least computationally demanding sampling method).

Table 3 Time requirements for LIE sampling and continuum electrostatics analysis calculations on the HIV1-RT inhibitor set relative to time needed for energy minimization sampling with Born solvation

Method	Sampling	PB	GB
MD, PBC/RF (150 ps)	1426	205	12.7
MD, Shell (150 ps)	316	204	12.6
MD, Born (30 ps)	34.2	40.8	2.51
Min, Born	1	0.68	0.04

Sampling durations for MD are indicated. Continuum electrostatics energy analysis calculations were applied every 0.5 ps without cut-off to compute running averages of the LIE terms.

Table 4 Synopsis of the fits to Eq. (18) for the HIV1-RT inhibitor set using MD for sampling with reaction field (PBC), shell and Born solvation approaches and also for energy minimization with Born solvation

Method	α	β	SE kcal mol^{-1}	r^2
MD, PBC/RF PB	0.322	0.360	1.11	0.76
MD, PBC/RF GB	0.333	0.312	1.12	0.75
MD, Shell PB	0.286	0.356	1.15	0.74
MD, Shell GB	0.299	0.298	1.19	0.72
MD, Born PB	0.434	0.219	1.22	0.71
MD, Born GB	0.445	0.177	1.23	0.70
Min, Born PB	0.457	0.167	1.30	0.67
Min, Born GB	0.473	0.138	1.31	0.66

Results and discussion

In order to investigate systematically the influence of solvation on sampling and the subsequent use of continuum electrostatics for LIE model development, three sets of MD runs were carried out for both the BACE and HIV-1RT series, using GB, shell and PBC solvation. After checking

Table 5 Summary of the fits to Eq. (18) for the BACE inhibitor set using MD for sampling with reaction field (PBC), shell and Born solvation approaches and also for energy minimization with Born solvation

Method	α	β	SE kcal mol^{-1}	r^2
MD, PBC/RF PB	0.231	0.055	1.03	0.75
MD, PBC/RF GB	0.226	0.059	1.07	0.73
MD, Shell PB	0.218	0.055	1.10	0.72
MD, Shell GB	0.211	0.057	1.11	0.71
MD, Born PB	0.297	0.061	1.11	0.71
MD, Born GB	0.303	0.060	1.11	0.71
Min, Born PB	0.264	0.062	1.17	0.68
Min, Born GB	0.251	0.058	1.20	0.66

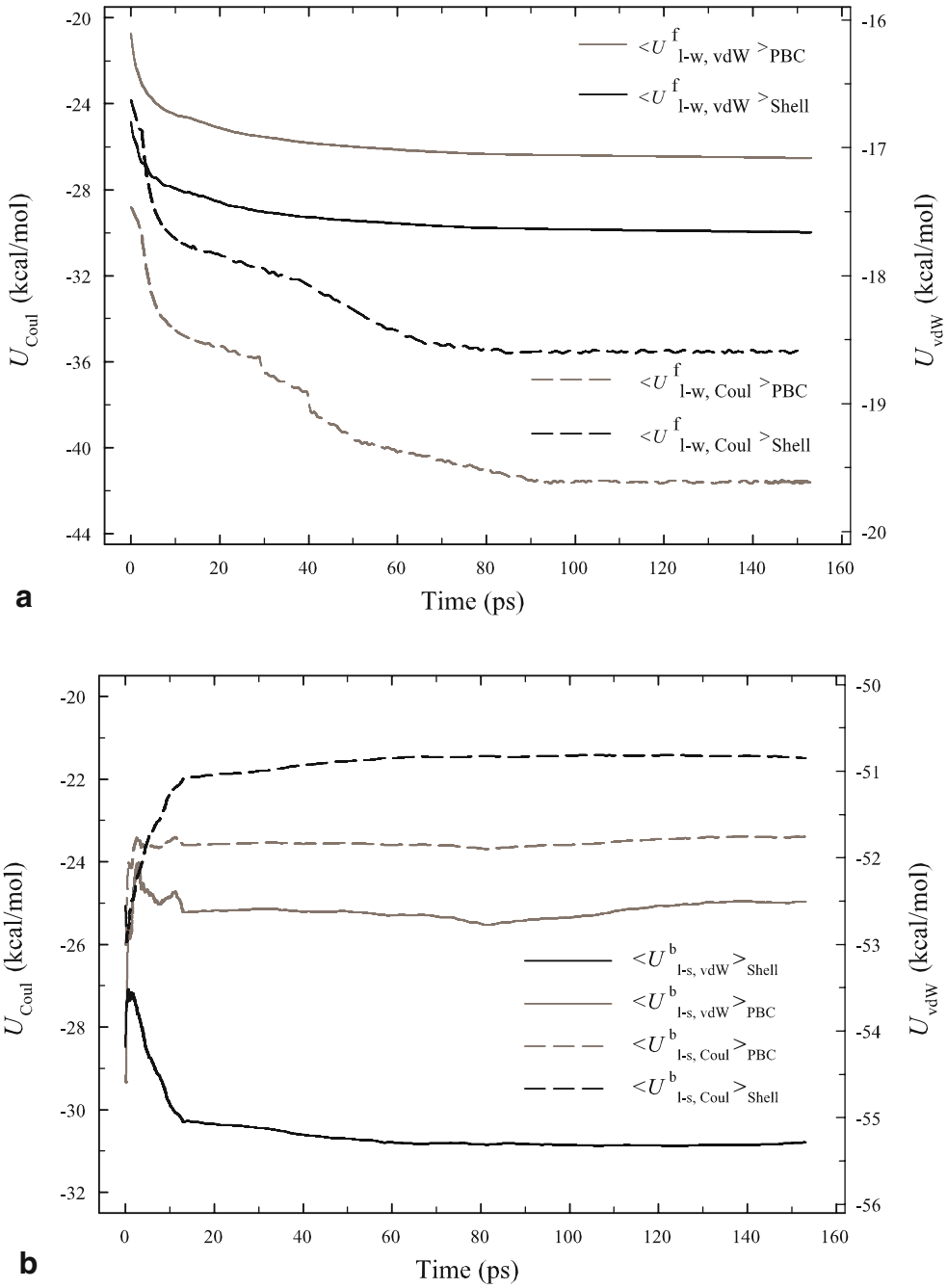


Fig. 3 Running averages for van der Waals and Coulombic interaction energies of the ligand with (**a**) explicit water in the free state and (**b**) with the receptor and explicit water in the bound state, for compound H11 (binding to HIV-1RT). Plots are shown after 50 ps of relaxation for both 15 Å 'shell' and PBC solvation (the latter with reaction field electrostatics). Note that these interaction energies have *not* been calculated using continuum solvation, and the cut-off is applied. Split axes are used in both plots with van der Waals energy terms on the right and Coulombic terms on the left

for convergence of the total system and ligand-environment energies in the production phase of the free and bound state simulations (with water molecules, if present, and cut-off applied), the ligand-environment energies of Eqs. (11) to (14) were evaluated for each sampled configuration. One set of energy minimization calculations was performed using GB solvation for each series for subsequent calculation of the overall LIE terms in Eqs. (15) to (17). The results of the fits are summarized in Tables 4 and 5 for HIV-1RT and BACE, respectively. In both cases, the α and β values are quite similar for sampling using explicit solvation (*i.e.* PBC and shell) and Born solvation (MD and energy minimization). For explicit solvation sampling, the respective α, β standard deviations are 0.021 and 0.031 for HIV-1RT and 0.008 and 0.002 for BACE, while the corresponding values for sampling with GB solvation are 0.017 and 0.034 for HIV-1RT and 0.025 and 0.002 for BACE; this contrasts with overall respective α, β standard deviations for all sampling/analysis methods of 0.078 and 0.089 for HIV-1RT and 0.035 and 0.003 for BACE. Generally, PB and GB continuum electrostatics analysis results in similar α, β coefficients for a given sampling and solvation approach. The Born MD (GB analysis) results of 0.445 and 0.177 for α and β, respectively, contrast to

respective values of 0.178 and 0.320 reported for the same set by Zhou et al. using MD with GB solvation [26]. However, it should be borne in mind that the coefficients are not directly comparable, since the non-polar solvation energy contribution was partitioned into a third parameter in the fit of Zhou et al. [26], in addition to an inconsistent treatment of the protein-ligand interaction in the bound state, as shown by Carlsson et al. [31]. In the case of the BACE set, the α coefficient tended to be larger than β value by a factor of 5 to 6, depending on the solvation and sampling strategy. This was also found by Tounge and Reynolds for the BACE set, although there was a larger variability in the α to β ratio for various cut-off and sampling strategies [28]. For a given solvation and

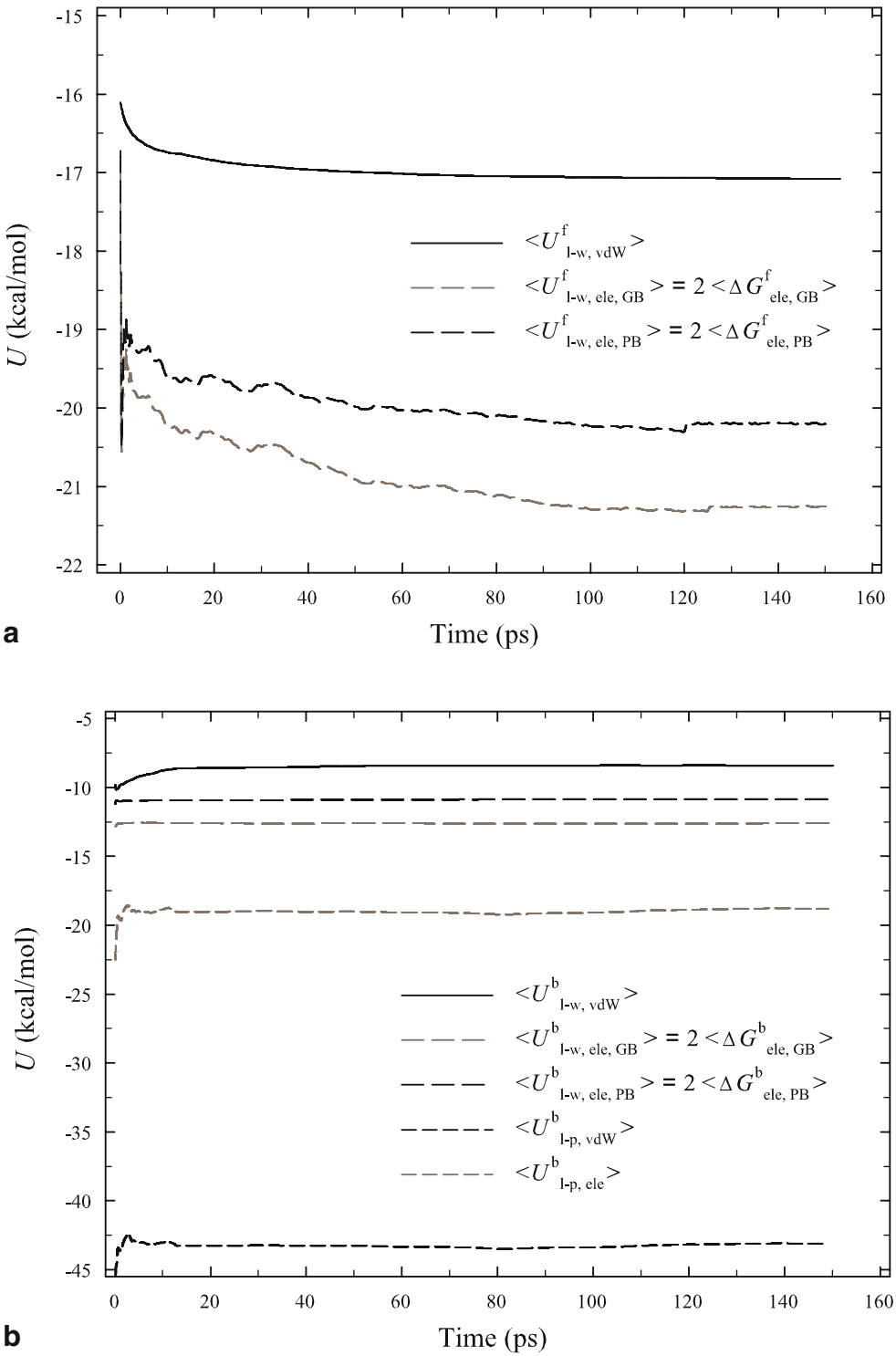


Fig. 4 Evolution of the running averages for compound H11 (binding to HIV-1RT) of the (**a**) free state ligand-water electrostatic and van der Waals interaction energy terms, $\langle U^{f}_{l-w,ele,M}\rangle$ and $\langle U^{f}_{l-w,vdW}\rangle$, respectively, and (**b**) bound state ligand-water and ligand-protein interaction energies. In each state, the ligand-water electrostatic terms, $\langle U^{f}_{l-w,ele,M}\rangle$ and $\langle U^{b}_{l-w,ele,M}\rangle$, are twice the corresponding GB and PB electrostatic solvation free energies. For the bound state, the ligand-receptor van der Waals $\langle U^{b}_{l-p,vdW}\rangle$ and Coulombic $\langle U^{b}_{l-p,ele}\rangle$ interaction energies are also shown, in addition to the ligand-water van der Waals interaction term $\langle U^{b}_{l-w,vdW}\rangle$. The terms were calculated (without application of a cut-off) from a trajectory sampled under PBC solvation with reaction field electrostatics

sampling strategy, comparison of statistical qualities of fit from PB and GB approaches reveals that PB analysis is usually only marginally better, reflecting the similar electrostatic solvation free energies predicted by both schemes. Indeed, Rizzo et al. [58] found for small molecules that the partial charge scheme is more influential on solvation energies than whether PB or GB schemes are used. Although MD sampling using PBC and reaction field electrostatics in conjunction with PB analysis yields the best fit in both cases (with r^2 values of 0.76 and 0.75 and standard errors (SE's) of 1.11 and 1.03 kcal mol^{-1} for HIV-1RT and BACE, respectively), use of the other sampling and solvation techniques results in r^2 and SE's of within a maximum (for the HIV-1RT set) of 16 and 15%, respectively, at a considerably lower computational expense (cf. Table 3).

Typical examples are depicted in Fig. 3a and b (using split axes) of the evolution of the 'running' average during the production stage of MD runs for HIV-1RT (H11) of the van der Waals and Coulombic ligand-environment interaction energies in the free and bound state (including with water molecules) for shell and PBC solvation. In these energy terms, the switching function is applied to the van der Waals interactions, and to the Coulombic interactions in the shell solvation case. In the PBC case, the 15 Å cut-off is used for reaction field electrostatics. These terms are used to assess if averaging during MD is satisfactory for subsequent application to the entire sampled trajectory of continuum electrostatics without any cut-off. In Fig. 3a, the van der Waals energies converge within 50 ps, while the Coulombic term requires 100 ps. The sensitivity of the Coulombic energy to shell or PBC solvation and long-range electrostatics is shown clearly by the lower PBC energy of around 6 kcal mol^{-1}. In the bound state (Fig. 3b), the energy terms converge more rapidly due to the greater inertia of the complex in the shell or box of water. In this case, the van der Waals interactions with the receptor (and the water molecules) are greater in magnitude than the Coulombic terms, subject to the cut-offs. The disparity between the shell and PBC Coulombic terms is lower in this case (about 2.5 kcal mol^{-1}) due to a smaller interaction energy with the water molecules.

To illustrate the impact of continuum electrostatics without any cut-off on the sampled configurations of the same example, the evolution of the individual interaction energy terms is shown for the free and bound states in Fig. 4a and b. These terms are tabulated for the entire HIV-1RT series in Table 6 for PB and GB electrostatics applied to MD sampling using PBC and reaction field electrostatics. The terms converge more rapidly, *e.g.* the free state electrostatic terms do so within around 50 ps in contrast to 100 ps. The more extensive ligand exposure to the solvent in the free state leads to a larger magnitude of electrostatic solvation free energy (and hence ligand-water electrostatic interaction energy, via linear response): $\left\langle \Delta G^{\mathrm{f}}_{\mathrm{ele,M}} \right\rangle$ is larger in magnitude and more negative than $\left\langle \Delta G^{\mathrm{b}}_{\mathrm{ele,M}} \right\rangle$, *e.g.* −10.63 versus −6.22 kcal mol^{-1}for GB electrostatics. GB solvation energies, $\left\langle \Delta G^{\mathrm{f}}_{\mathrm{ele,GB}} \right\rangle$ and $\left\langle \Delta G^{\mathrm{b}}_{\mathrm{ele,GB}} \right\rangle$, are typically up to 15% more negative than PB values. In Figs. 3a and 4a, the

Table 6 Averages of the energy terms used in the construction of the LIE components for the HIV-1RT set using PB analysis and sampling by MD under PBC and reaction field electrostatics

Ligand	$\left\langle \Delta G^{\mathrm{f}}_{\mathrm{ele,PB}} \right\rangle$	$\left\langle \Delta G^{\mathrm{f}}_{\mathrm{ele,GB}} \right\rangle$	$\left\langle U^{\mathrm{f}}_{\mathrm{l-s,vdW}} \right\rangle$	$\left\langle U^{\mathrm{b}}_{\mathrm{l-s,vdW}} \right\rangle$	$\left\langle U^{\mathrm{b}}_{\mathrm{l-p,ele}} \right\rangle$	$\left\langle \Delta G^{\mathrm{b}}_{\mathrm{ele,PB}} \right\rangle$	$\left\langle \Delta G^{\mathrm{b}}_{\mathrm{ele,GB}} \right\rangle$
H01	−15.08	−15.45	−17.81	−49.80	−28.56	−3.12	−3.56
H02	−10.14	−11.04	−16.49	−52.70	−17.34	−5.14	−5.92
H03	−10.18	−11.10	−15.82	−51.15	−19.56	−5.88	−6.74
H04	−10.65	−11.54	−15.33	−49.62	−16.45	−4.97	−5.66
H05	−10.70	−12.02	−15.29	−58.50	−20.26	−2.86	−3.28
H06	−10.57	−11.15	−16.37	−58.45	−20.69	−4.58	−5.37
H07	−10.17	−11.70	−14.79	−47.60	−17.24	−4.03	−4.43
H08	−10.23	−11.92	−13.83	−47.05	−16.30	−3.36	−4.57
H09	−10.12	−10.68	−16.43	−53.96	−21.46	−5.91	−6.83
H10	−10.03	−10.47	−16.99	−54.53	−18.96	−6.12	−7.09
H11	−10.12	−10.63	−17.08	−52.55	−18.80	−5.40	−6.22
H12	−10.44	−10.85	−16.54	−55.39	−15.95	−5.31	−6.14
H13	−14.98	−15.94	−17.77	−49.96	−24.83	−3.82	−4.49
H14	−16.59	−16.97	−17.81	−48.33	−22.58	−5.34	−6.08
H15	−15.31	−15.32	−20.34	−56.97	−23.91	−5.17	−5.91
H16	−14.72	−14.58	−21.35	−60.16	−26.83	−5.74	−6.54
H17	−14.18	−14.21	−21.27	−59.71	−22.43	−7.11	−5.53
H18	−10.51	−11.38	−16.17	−61.54	−17.56	−4.95	−5.76
H19	−10.58	−12.75	−12.81	−47.35	−17.37	−3.82	−5.51
H20	−10.04	−11.56	−15.38	−52.43	−18.61	−4.76	−5.46

discrepancy between the shell and PBC ligand-water Coulombic interaction energies (circa −35.5 and −41.5 kcal mol^{-1}, respectively) and the linear response-derived terms $\left\langle U^{f}_{l-w,ele,GB} \right\rangle$ and $\left\langle U^{f}_{l-w,ele,PB} \right\rangle$ (−21.26 and −20.24, respectively) is substantial. This is due in part to the neglect of entropy in the ligand-water Coulombic terms of Fig. 3a, unlike the PB/GB estimates. However, the continuum nature of the PB/GB approach is inherently more convergent than interaction energies with explicit water (subject to cut-offs), even when the reaction field approach is used (for the PBC case in Fig. 3a). Although separate ligand-explicit water interaction terms have not been shown for the bound state in Fig. 3b, the difference between $\left\langle U^{b}_{l-w,ele,GB} \right\rangle$ and $\left\langle U^{b}_{l-w,ele,PB} \right\rangle$ is around 0.8 kcal mol^{-1} (−12.44 and −10.8 kcal mol^{-1}, respectively), in comparison to about 2.5 kcal mol^{-1} between shell and PBC Coulombic terms (*i.e.* circa −21.5 and −24 kcal mol^{-1}, respectively, cf. Fig. 3b). Since $\left\langle U^{b}_{l-p,ele} \right\rangle$ is −18.8 kcal mol^{-1} (although without any cut-off, cf. Fig. 4b and Table 6), then it is clear that the ligand-explicit water interaction in the bound state is also subject to considerable convergence limitations. Therefore, this example shows the importance of continuum electrostatics to calculate converged energies, especially for electrostatic terms, without a cut-off applied for ligand-receptor terms.

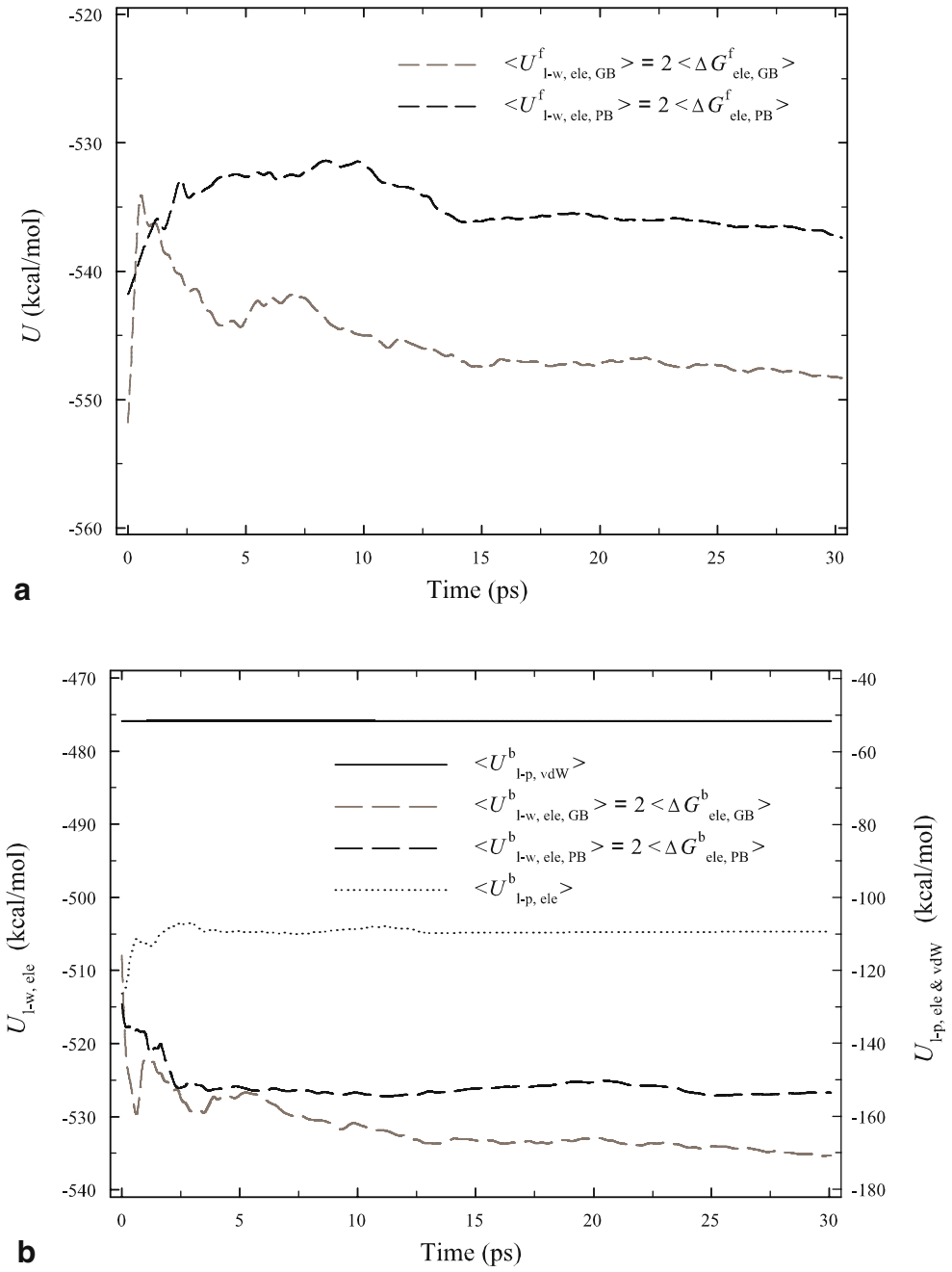


Fig. 5 Running average of interaction energy terms for compound B01 (*i.e.* OM99-2 binding to BACE). The terms have been calculated without cut-off from an MD trajectory sampled under Born solvation. In (**a**) the free state, the ligand-water electrostatic interaction energy $\langle U^{f}_{l-w,ele,M} \rangle$ is given for both PB and GB electrostatics (as twice the continuum solvation free energy). In the free state, the ligand-water van der Waals term $U^{f}_{l-w,vdW}$ is zero, as water molecules are absent during sampling. In (**b**) the bound state, the ligand-water electrostatic terms, $\langle U^{b}_{l-w,ele,M} \rangle$, and the ligand-receptor van der Waals $\langle U^{b}_{l-p,vdW} \rangle$ and Coulombic $\langle U^{b}_{l-p,ele} \rangle$ interaction energies are shown

An example is shown in Fig. 5a and b of the more rapid convergence of the interaction energy terms in each state using MD sampling with GB solvation. In this case, results are shown for BACE (OM99-2), where the ligand is charged. The 30 ps duration of the production phase of the MD simulation is more than sufficient for convergence in both states. Since the ligand is charged, the electrostatic ligand-water interaction energies are large and negative in both states: *e.g.* circa −548 and −536 kcal mol^{-1} for $\left\langle U^{f}_{l-w,ele,GB} \right\rangle$ and $\left\langle U^{f}_{l-w,ele,PB} \right\rangle$, respectively. The ligand-water term also dominates the ligand-receptor Coulombic energy in the overall bound state electrostatic interaction with the environment, *i.e.* $\left\langle U^{b}_{l-w,ele,GB} \right\rangle$ and $\left\langle U^{b}_{l-w,ele,PB} \right\rangle$ are −535 and −527 kcal mol^{-1}, respectively, in contrast to −110 kcal mol^{-1} for $\left\langle U^{b}_{l-p,ele} \right\rangle$. It is instructive to examine the ligand-environment interaction terms in each state prior to calculation of the overall LIE van der Waals and electrostatic terms by Eqs. (15) to 17. These overall LIE terms, $\langle \Delta U_{vdW} \rangle$, $\left\langle \Delta U^{PB}_{ele} \right\rangle$ and $\left\langle \Delta U^{GB}_{ele} \right\rangle$, are given in the Electronic Supplementary Material for all HIV-1RT and BACE ligands for each sampling technique.

In Tables 7 and 8, results are presented for LIE model predictions and jacknife cross-validation for both series using PB analysis on MD output with PBC and reaction field electrostatics. For HIV-1RT, the jacknife results exhibit closer agreement (r^2: 0.69, SE: 1.25 kcal mol^{-1}) with the original model (r^2: 0.76, SE: 1.11 kcal mol^{-1}) than for BACE (r^2: 0.61, SE: 1.30 kcal mol^{-1} versus r^2: 0.75, SE: 1.03 kcal mol^{-1}). This was found also for the other solvation and sampling cases. The best results of Zhou et al. [26] and of Tounge and Reynolds [28] for HIV-1RT and BACE have r^2 values of 0.74 and 0.71, and SE's of 1.1 and 1.1 kcal mol^{-1}, respectively, comparable to the results of this study. Plots of model versus experimental values for ΔG_{bind} are shown in Figs. 6, 7, 8 and 9. For the HIV-1RT series, PB analysis results from MD sampling are given in Fig. 6 and PB/GB analysis results for GB energy minimization are shown in Fig. 7. For both sampling techniques and different solvation strategies, the H11 ligand is predicted as being more weakly binding (*e.g.* around −9.6 to −10.0 kcal mol^{-1} versus an experimental value of −11.89 kcal mol^{-1}). This discrepancy also occurs for H17 for all sampling cases, and for H03, H06 and H12 with Born solvation MD sampling. For more weakly-binding ligands, MD sampling with Born solvation predicts stronger binding (cf. Fig. 6). For BACE, GB analysis results from MD are given in Fig. 8 and PB and GB results for energy minimization are shown in Fig. 9. For each solvation and sampling scheme, it is interesting to note that the charged ligands OM99-2 and OM00-3 (labeled B01 and B02, respectively) are predicted quite well, and is better than predictions for these charged ligands than those of Tounge and Reynolds [28], who used the approach of Zhou et al. [26]. However, B06 and B08 are predicted as more weakly and strongly bound, respectively, by around 1.5–2 kcal mol^{-1}. Inspection of Figs. 6 and 8 shows that there is less disparity between the various MD sampling methods for BACE than for HIV-1RT. As with HIV-1RT, PB and GB analysis of the GB-minimized structures for the BACE series leads to similar results (cf. Figs. 7 and 9).

The LIE fit is similar in quality for both the HIV-1RT and BACE sets with r^2 values of 0.70 to 0.76 for MD sampling and 0.66 to 0.67 for energy minimization, in

Table 7 Results for LIE fitting and jackknife cross-validation on the HIV-1RT set using PB analysis and sampling by MD using PBC and reaction field electrostatics

Ligand	Expt	LIE	Jackknife
H01	−7.32	−6.81	−6.72
H02	−7.73	−9.16	−9.25
H03	−9.20	−10.23	−10.48
H04	−8.06	−7.84	−7.83
H05	−10.01	−10.52	−10.74
H06	−12.16	−11.74	−11.66
H07	−8.03	−7.32	−7.24
H08	−5.43	−6.59	−6.77
H09	−10.96	−11.89	−12.26
H10	−11.24	−10.99	−10.95
H11	−11.89	−9.68	−9.39
H12	−9.93	−9.55	−9.51
H13	−6.52	−6.37	−6.34
H14	−5.78	−5.10	−4.84
H15	−9.35	−8.21	−8.08
H16	−11.19	−10.71	−10.66
H17	−12.16	−10.33	−10.19
H18	−11.68	−11.92	−12.07
H19	−5.11	−7.56	−7.78
H20	−8.40	−9.61	−9.69
		r^2: 0.76 SE: 1.11	r^2: 0.69 SE: 1.25

Table 8 Results for LIE fitting and jackknife cross-validation on the BACE set using PB analysis and sampling by MD using PBC and reaction field electrostatics

Ligand	Expt	LIE	Jacknife
B01 (OM99-2)	−12.06	−12.16	−12.21
B02 (OM00-3)	−13.05	−12.84	−12.50
B03	−6.38	−7.64	−8.05
B04	−7.55	−9.08	−9.26
B05	−8.16	−7.30	−6.97
B06	−9.90	−7.97	−7.46
B07	−11.30	−10.47	−10.38
B08	−10.02	−11.40	−11.82
B09	−11.02	−10.74	−10.67
B10	−7.19	−8.19	−8.39
B11	−11.81	−11.26	−11.06
B12	−11.11	−10.48	−10.27
		r^2: 0.75 SE: 1.03	r^2: 0.61 SE: 1.30

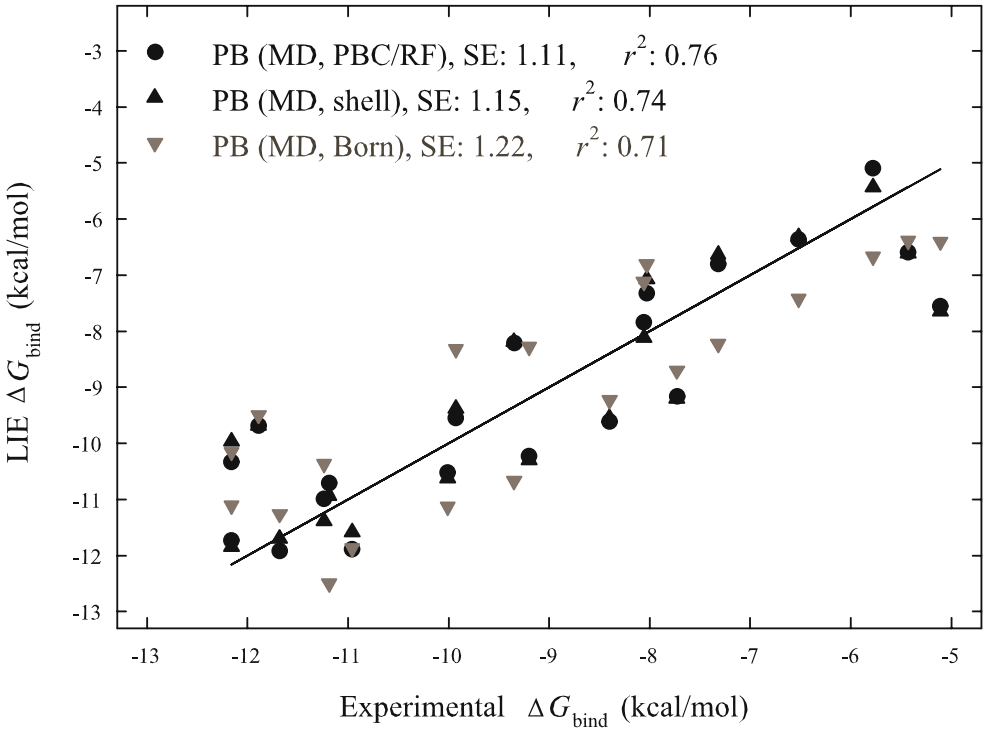


Fig. 6 LIE-predicted binding free energies using PB electrostatics analysis versus experimental values for the HIV-1RT set, with MD sampling under PBC, shell and Born solvation. The line is a visual aid representing ideal agreement between fit and experimental values

comparison to respective values of 0.71 to 0.75 and around 0.66 to 0.68 for BACE. In both cases, there is little difference between PB and GB analysis for a given solvation or sampling strategy, as shown by the similarity of the α and β values in Tables 4 and 5. Although PB analysis in conjunction with MD under PBC does produce the best fits, the use of energy minimization or shorter MD runs for sampling with Born solvation requires far less computational resources (cf. Table 3) and the statistical quality of fit is only slightly inferior, *e.g.* r^2 of 0.67, 0.71 and 0.76 for energy minimization, Born MD and PBC MD with PB analysis for the HIV-1RT case. The fit for the HIV-1RT set appears to be much more dependent on solvation and electrostatics than the BACE series, as evidenced by α

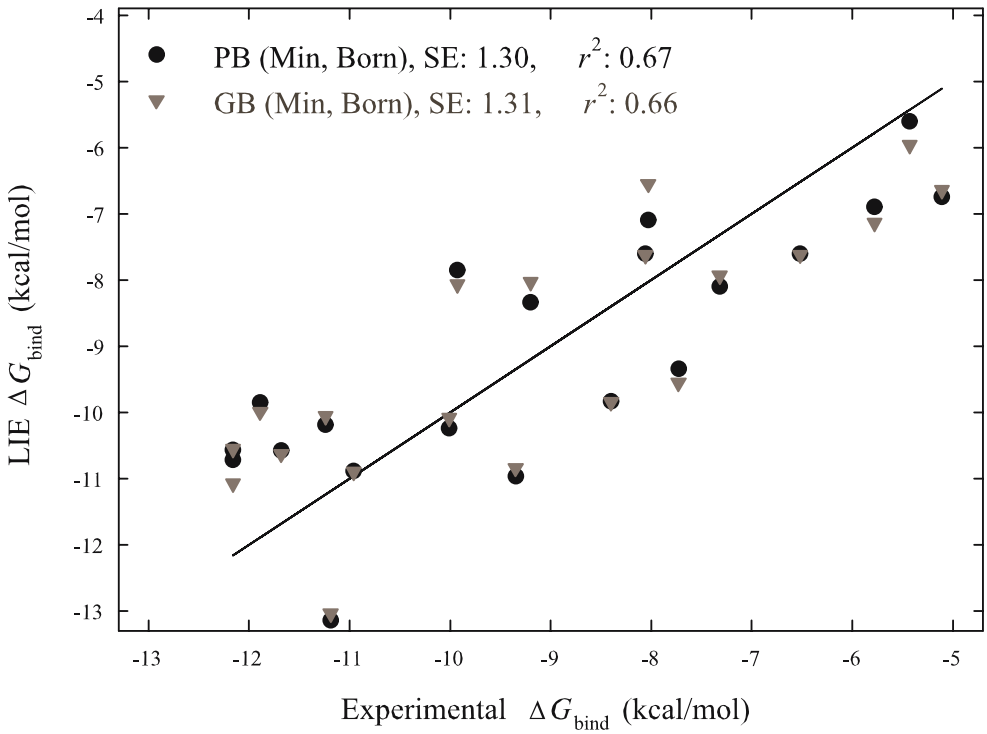


Fig. 7 LIE predictions of binding free energies using PB and GB electrostatics analysis versus experimental values for the HIV-1RT set, using energy minimization for sampling. The line is as in Fig. 6

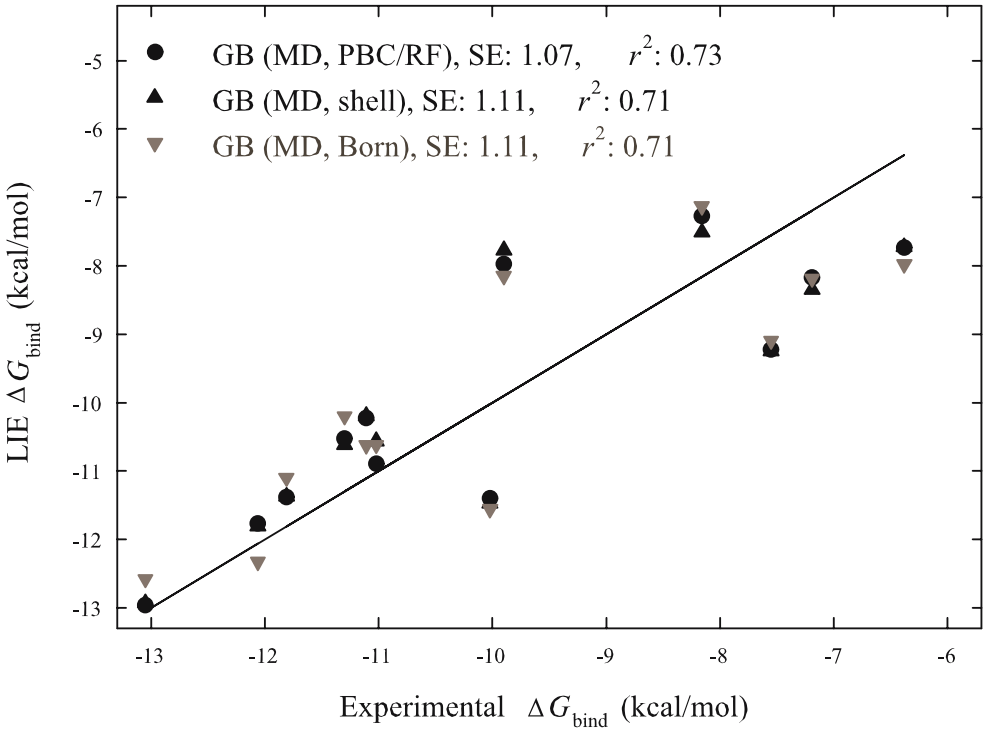


Fig. 8 LIE-predicted binding free energies using GB electrostatics analysis versus experimental values for the BACE set, with MD sampling under PBC, shell and Born solvation. The line is as in Fig. 6

and β coefficients of comparable values, a substantially larger standard deviation of 0.089 for the β coefficients in comparison to 0.003 for BACE, and the slightly greater disparity between the r^2 and SE's for various solvation approaches: the range of r^2 and SE's differ by 15 and 14%, respectively, for HIV-1RT with respective values of 14 and 13 % for BACE. Inspection of Fig. 6 shows that there is a consistent overestimation of binding strength with Born solvation for ligands with experimental binding free energies of greater than −8 kcal mol^{-1}. Although there are clearly differences in energy terms when using explicit and implicit solvation subject to cut-off (*e.g.* the Coulombic

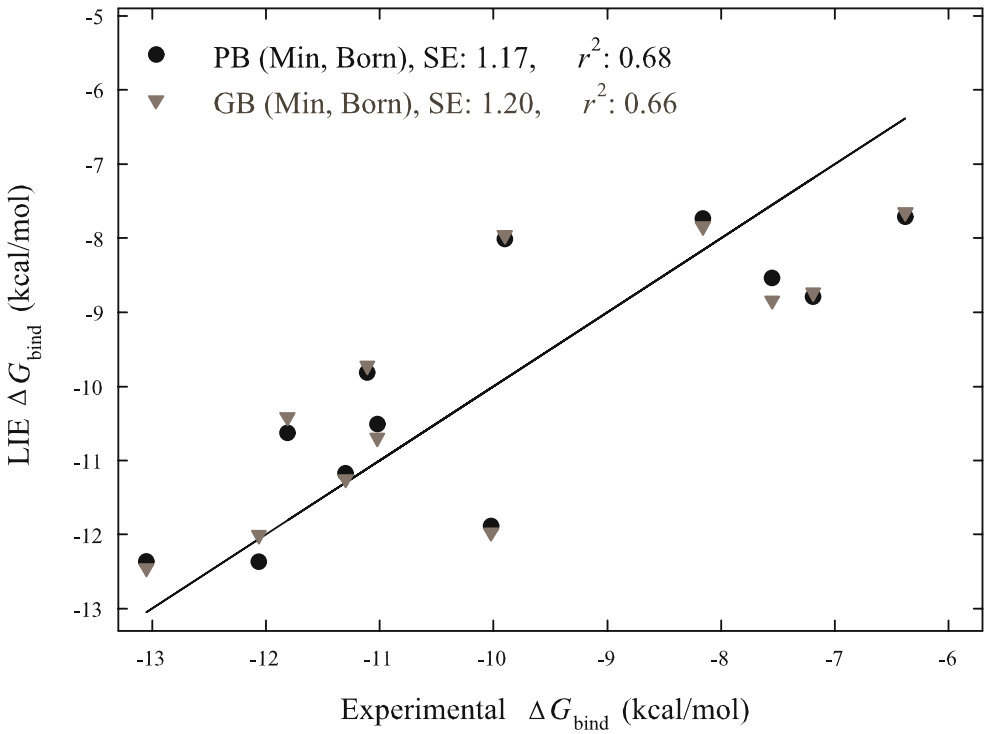


Fig. 9 LIE predictions of binding free energies using PB and GB electrostatics analysis versus experimental values for the BACE set, using energy minimization for sampling. The line is as in Fig. 6

interaction term with explicit water molecules for PBC and shell solvation in the free state, cf. Fig. 3a), these differences are reduced significantly with the use of PB or GB analysis without cut-off, minimizing the difference between results with different solvation methods.

Summary

LIE models have been derived for the binding free energies of congeneric series of HEPT analogues in HIV-1RT and oligopeptide inhibitors in BACE for PBC, shell and Born solvation. Both MD and energy minimization were used for sampling, and PB and GB electrostatics analysis approaches were applied to the sampled configurations without cut-off to compute the LIE van der Waals and electrostatic terms for a two-component fit. Although PB analysis applied to MD sampling under PBC with reaction field electrostatics was found to result in the best fits (r^2: 0.76, SE: 1.11 kcal mol^{-1} for HIV-1RT and r^2: 0.75, SE: 1.03 kcal mol^{-1} for BACE), the other approaches resulted in fits with r^2 and SE's of within a maximum (for the HIV-1RT set) of 15 and 14%, respectively, at a considerably lower computational expense. The cross-validation results were reasonable for the HIV-1RT set (*e.g.* r^2: 0.69, SE: 1.25 kcal mol^{-1} for MD under PBC with PB analysis) and for BACE (*e.g.* r^2: 0.61, SE: 1.30 kcal mol^{-1} for the same sampling).

For some more strongly-binding ligands, the models predicted weaker binding, *e.g.* H11 and H17 and also H03, H06 and H12 without Born solvation MD sampling. Interestingly, the charged ligands OM99-2 and OM00-3 (labeled B01 and B02, respectively) are predicted quite well. There are several possible reasons for errors by the models, not least the assumption that such a simple form of Eq. (18) may be used to approximate the binding process. There may be errors in the experimental estimates of ΔG_{bind}: standard errors of 1 kcal mol^{-1} would be quite common and this limits the accuracy of the LIE fit. The resolution of the crystal structures, in addition to the docked structures of the ligands for which co-crystal structures are not available, would also be important. It would be worthwhile to carry out a systematic study of the effect of different initial (docked) ligand poses on the LIE results. The finding for both of these series that MD with PBC solvation offers no real advantage over shell solvation, nor over Born solvation with MD or energy minimization sampling, suggests that a reasonable strategy for an LIE study on a larger set of compounds for these receptors could proceed with energy minimization and Born solvation with subsequent GB analysis. As the relative timings in Table 3 show, this approach is three orders of magnitude faster than the equivalent MD under PBC with PB analysis.

Acknowledgements The author acknowledges useful discussions with Dr. Jonathan Heal (Prosarix Ltd.).

References

1. Lazaridis T (2002) Curr Org Chem 6:1319–1332
2. Muegge I (2000) Perspect Drug Discovery Des 20:99–114
3. Stahl M, Rarey M (2001) J Med Chem 44:1035–1042
4. Pearlman DA, Charifson PS (2001) J Med Chem 44:3417–3423
5. Beveridge DL, Dicapua FM (1989) Ann Rev Biophys Chem 18:431–492
6. Straatsma TP, McCammon JA (1992) Ann Rev Phys Chem 43:407–435
7. Kollman PA (1993) Chem Rev 93:2395–2417
8. Pearlman DA, Rao BG (1998) Encyl Comput Chem 1036–1061
9. Woods CJ, Essex JW, King MA (2003) J Phys Chem B 107:13703–13710
10. Jayachandran G, Shirts MR, Park S, Pande VS (2006) J Chem Phys 125:084901
11. Åqvist J, Medina C, Samuelsson JE (1994) Protein Eng 7:385–391
12. Hansson T, Åqvist J (1995) Protein Eng 8:1137–1144
13. Åqvist J, Hansson T (1996) J Phys Chem 100:9512–9521
14. Åqvist J (1996) J Comp Chem 17:1587–1597
15. Carlson HA, Jorgensen WL (1995) J Phys Chem 99:10667–10673
16. Jones-Hertzog DK, Jorgensen WL (1997) J Med Chem 40:1539–1549
17. Marelius J, Graffner-Nordberg M, Hansson T, Hallberg A, Åqvist J (1998) J Comput-Aided Mol Des 12:119–131
18. Smith RH Jr, Jorgensen WL, Tirado-Rives J, Lamb ML, Janssen PAJ, Michejda CJ, Smith MBK (1998) J Med Chem 41:5272–5286
19. Hansson T, Marelius J, Åqvist J (1998) J Comput-Aided Mol Des 12:27–35
20. Wall ID, Leach AR, Salt DW, Ford MG, Essex JW (1999) J Med Chem 42:5142–5152
21. Wang W, Wang J, Kollman PA (1999) Proteins: Struct, Funct, Genet 34:395–402
22. Lamb ML, Tirado-Rives J, Jorgensen WL (1999) Biorg Med Chem 7:851–860
23. Sham YY, Chu ZT, Tao H, Warshel A (2000) Proteins: Struct, Funct, Genet 39:393–407
24. Chen J, Wang R, Taussig M, Houk KN (2001) J Org Chem 66:3021–3026
25. Hou TJ, Zhang W, Xu XJ (2001) J Phys Chem B 105:5304–5315
26. Zhou R, Friesner RA, Ghosh A, Rizzo RC, Jorgensen WL, Levy RM (2001) J Phys Chem B 105:10388–10397
27. Åqvist J, Marelius J (2001) Comb Chem High Throughput Screening 4:613–626
28. Tounge BA, Reynolds CH (2003) J Med Chem 46:2074–2082
29. Svab I, Alexandru D, Vitos G, Flonta ML (2004) J Cell Mol Med 8:551–562
30. Schrøder Leiros H-K et al (2004) Protein Sci 13:1056–1070
31. Carlsson J, Andér M, Nervall M, Åqvist J (2006) J Phys Chem B 110:12034–12041
32. Åqvist J (1990) J Phys Chem 94:8021–8024
33. Roux B, Yu H-A, Karplus M (1990) J Phys Chem 94:4683–4688
34. Zhou Z, Madura JD (2004) Proteins: Struct Funct Genet 57:493–503
35. Lee FS, Warshel A (1992) J Chem Phys 97:3100–3107
36. Tanaka H et al (1991) J Med Chem 34:349–357
37. Tanaka H et al (1992) J Med Chem 35:346–350
38. Tanaka H et al (1992) J Med Chem 35:4713–4719

39. Tanaka H et al (1995) J Med Chem 38:2860–2865
40. Hopkins AL, Ren J, Esnouf RM, Willcox BE, Jones EY, Ross C, Miyasaka T, Walker RT, Tanaka H, Stammers DK, Stuart DI (1996) J Med Chem 39:1589–1600
41. Ghosh AK, Bilcer G, Harwood C, Kawahama R, Shin D, Hussain KA, Hong L, Loy JA, Nguyen C, Koelsch G, Ermolieff J, Tang J (2001) J Med Chem 44:2865–2868
42. Hong L, Koelsch G, Lin X, Wu S, Terzyan S, Ghosh AK, Zhang XC, Tang J (2000) Science 290:150–153
43. Hong L, Turner RT, Koelsch G, Shin D, Ghosh AK, Tang J (2002) Biochemistry 41:10963–10967
44. MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1010 Sherbrooke St. W., Suite 910, Montréal, Québec, Canada H3A 2R7
45. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 117:11225–11236
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
47. Onufriev A, Bashford D, Case DA (2000) J Phys Chem B 104:3712–3720
48. Hawkins GD, Cramer CJ, Truhlar DG (1995) Chem Phys Lett 246:122–129
49. van Gunsteren WF, Berendsen HJC (1990) Angew Chem Int Ed Engl 29:992–1023
50. Hart TN, Read RJ (1992) Proteins: Struct Funct Genet 13:206–222
51. Neumann M (1985) J Chem Phys 82:5663–5672
52. Lekner J (1989) Physica A157:826; Lekner J (1991) Physica A157:826–838
53. English NJ, MacElroy JMD (2002) Molec Phys 100:3753–3769
54. English NJ, Sorescu DC, Johnson JK (2006) J Phys Chem Solids 67:1399–1409
55. Allen MP, Tildesley DJ (1987) Computer Simulation of Liquids. Clarendon, Oxford
56. Izaguirre JA, Reich S, Skeel RD (1999) J Chem Phys 110:9853–9864
57. Grant JA, Pickup BT, Nicholls A (2001) J Comput Chem 22:608–640
58. Rizzo RC, Aynechi T, Case DA, Kuntz ID (2006) J Chem Theory Comput 2:128–139