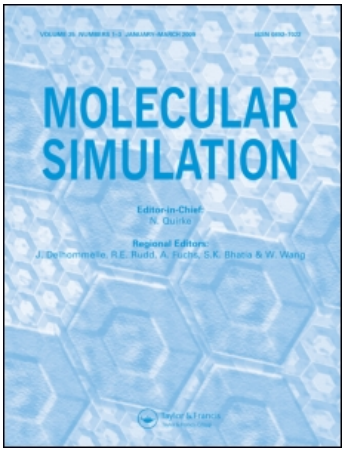

Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Linear interaction energy approximation for binding affinities of nevirapine and HEPT analogues with HIV-1 reverse transcriptase

Antonios Tavlarakis[ab]; Ruhong Zhou[ac]

[a] IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA [b] Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA, USA [c] Department of Chemistry, Columbia University, New York, NY, USA

To cite this Article Tavlarakis, Antonios and Zhou, Ruhong(2009) 'Linear interaction energy approximation for binding affinities of nevirapine and HEPT analogues with HIV-1 reverse transcriptase', Molecular Simulation, 35: 15, 1224 — 1241

To link to this Article: DOI: 10.1080/08927020902929828

URL: http://dx.doi.org/10.1080/08927020902929828

Molecular Simulation
Vol. 35, No. 15, December 2009, 1224–1241

Taylor & Francis
Taylor & Francis Group

Linear interaction energy approximation for binding affinities of nevirapine and HEPT analogues with HIV-1 reverse transcriptase

Antonios Tavlarakis[ab] and Ruhong Zhou[ac]*

[a]*IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA;* [b]*Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA;* [c]*Department of Chemistry, Columbia University, New York, NY 10027, USA*

(*Received 16 January 2009; final version received 24 March 2009*)

The binding affinities and binding mechanism of nevirapine and HEPT analogues with HIV-1 reverse transcriptase (HIV-1RT) have been studied in this paper with a recently proposed linear interaction energy method based on a surface generalised Born continuum solvent model (LIE–SGB). Various LIE–SGB fitting schemes are presented and discussed on this relatively large binding set, consisting of a total of more than 50 ligands. For a training subset of 40 ligands (20 nevirapine and 20 HEPT analogues), the LIE–SGB method gives an RMS error (RMSE) of 0.89 kcal/mol with a correlation coefficient r^2 of 0.74. The leave-one-out cross validation results also show an encouraging RMSE of 1.00 kcal/mol, with a correlation coefficient r^2 of 0.69. The further blind tests on seven mostly high-potent candidates (not originally in the training subset) also show reasonable accuracies. The binding mechanism of this HIV-1RT binding set is found to be mainly driven by van der Waals interactions (i.e. a good geometrical fit is important) and the net loss of ligand cavity energies (i.e. the burial of solvent accessible surface area is favourable). The net electrostatic interactions, however, are found to be anti-binding. A secondary amide indicator is found to be necessary for nevirapine analogues to account for the deficiencies in quantum partial charges, and it is also shown that it represents a critical π-type hydrogen bond between the secondary amide fragment of nevirapine analogues and the phenyl ring of Tyr188A residue, which explains their otherwise surprising binding affinities. More quantum charge fittings, with the protein environment included, gave similar results, indicating that an out-of-plane polarisability might be needed to fully capture this π-type hydrogen bond in classical force fields. Finally, six new ligands are designed for optimal binding based on our predictions for further experimental validations.

Keywords: linear interaction energy; HIV-1 reverse transciptase; binding affinity; non-nucleoside inhibitors

1. Introduction

The design and development of anti-HIV drugs have been of great interest in recent decades due to the wide spread infection of AIDS (according to a 2008 UNAIDS Global Report, there are approximately 33 million individuals in the world living with HIV) [1–7]. One of the key enzymes packaged within the HIV virion capsid is a reverse transcriptase which plays an essential role in the replication of the virus. It is for this reason that HIV-1 reverse transcriptase (HIV-1RT) has emerged as one of the prime targets for the development of drugs for HIV/AIDS therapy. The HIV-1RT protein has both RNA-dependent DNA polymerase and RNaseH activities that are required for the conversion of genomic viral RNA to DNA; this viral DNA is subsequently incorporated into the host cell's genome. Inhibitors of HIV-1RT fall into two main classes, nucleoside inhibitors and non-nucleoside inhibitors [1–3]. Nucleoside inhibitors are compounds that mimic normal nucleoside substrates but lack the 3′-OH group required for DNA chain elongation. Nucleoside inhibitors compete with native nucleosides and effectively stall polymerase activity by becoming incorporated into the growing DNA strand thereby causing premature chain termination [1–3]. The non-nucleoside inhibitors are molecules that bind to a region of HIV-1RT located near the polymerase catalytic site. The binding event alters the conformation of critical residues and thereby inhibits the ability of the enzyme to perform normal RT functions.

Currently, the use of HIV-1RT nucleoside inhibitors, non-nucleoside inhibitors, HIV protease inhibitors, fusion inhibitors, and/or a combination of these is the best method for controlling the HIV infection [6]. In this study, we focus on the non-nucleoside inhibitors of HIV-1RT, particularly the nevirapine and HEPT (1-[(2-hydroxy-ethoxy)methyl]-6-(phenylthio)thymine) analogues. Nevirapine was the first FDA approved non-nucleoside inhibitor, and the HEPT analogue MKC-442 (also known as emivirine) was previously chosen as a drug candidate for clinical trials [2,3]. Unfortunately, MKC-442 was reported to trigger the liver enzyme cytochrome P450, leading to drug interactions between MKC-442 and protease inhibitors [8]. Non-compliance and non-ideal pharmacokinetics are major factors in the rise of drug resistance. Thus, more highly potent and reliable analogues are in great demand for lead optimisation. A total of more than 50 HEPT and nevirapine analogues

*Corresponding author. Email: ruhongz@us.ibm.com

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020902929828
http://www.informaworld.com

are studied in this paper with a computational approach using the linear interaction energy (LIE) method.

The LIE method combines molecular mechanics calculations with experimental data to build a model scoring function for the fast evaluation of ligand–protein binding free energies. LIE type methods were first suggested by Aqvist [9–12] based on approximating the charging integral in the free energy perturbation formula with a mean value approach in which the integral is represented as half the sum of the values at the endpoints, namely the free and bound states of the ligand. Since then, the LIE method has been pursued by many research groups for a number of ligand binding data sets with promising results [13–27]. There is an earlier, but similar approach, named linear response approximation (LRA), proposed by Warshel and co-workers [28]. The difference between LIE and LRA is that the LIE method neglects the electrostatic contribution over trajectories from the ligand's non-polar state (or uncharged state) [29–32]. The method we have recently developed is based on Aqvist's LIE approach, and it couples the LIE method with a continuum solvent model, surface generalised Born (SGB), for relevant physical interaction terms [20]. The LIE–SGB method is found to be significantly more efficient than the previous LIE methods based on explicit solvents with comparable accuracy [20].

From a computational standpoint, the LIE method has a number of highly attractive features. In contrast to free energy perturbations (FEP) where a large number of intermediate windows must be evaluated, the LIE method only requires simulations of the two ending windows [9]: the ligand in pure solvent named the 'free state', and the ligand bound to the solvated receptor called the 'bound state'. The idea is that one views the binding event as a replacement of the aqueous environment of the ligand with a mixed aqueous/protein environment. Also in contrast to FEP where small changes between ligands are usually computationally tractable, the LIE method can handle disparate ligands as long as they share similar binding modes. Furthermore, only interactions between the ligand and either the protein or the aqueous environment enter into the quantities that are accumulated during the simulation; the protein–protein and protein–water interactions are part of the 'reference' Hamiltonian and hence are used to generate conformations in the simulation, but are not used as descriptors in the resulting model for the binding free energy. This eliminates a considerable amount of noise and systematic uncertainties in the calculations, for example arising from different conformations of the protein obtained from co-crystallised structures of different ligands.

As pointed out in the previous paper [20], if the LRA was rigorously valid, the coefficient of the electrostatic term, β, would be 0.5, corresponding to the mean value approximation to the charging integral. In fact, one can recover a value very close to this for less complex systems, such as solvation of small molecules in water [10,13]. However, some of the steps involved in the binding event, such as the removal of water from the protein cavity and subsequent introduction of the ligand, are unlikely to be accurately described by a fully linear model. In practice, therefore, optimisation of fitting parameters can yield the electrostatic coefficient different from the ideal value of 0.5 [10,14,15]. By allowing this empirical element, one is sacrificing generality because the method probably requires the ligands to have similar binding modes, and new parameters must be developed for each receptor. In return, however, one can obtain a reasonable level of accuracy. In this study, our LIE–SGB method has been applied to the binding of nevirapine and HEPT analogues with HIV-1RT, which consists of a total of more than 50 ligands. For a training subset of 40 ligands (20 HEPT analogues and 20 nevirapine analogues), our results show that an RMSE of 0.89 kcal/mol (average unsigned error of 0.71 kcal/mol) with a correlation coefficient r^2 of 0.744 can be achieved for this relatively large binding set. The leave-one-out cross validation also shows an encouraging RMSE of 1.00 kcal/mol (average unsigned error of 0.81 kcal/mol) with a correlation coefficient of 0.69. The additional blind tests on seven mostly high-potent ligands (not originally in the training subset) also show decent accuracies. The binding affinities for this binding set are found to be driven mainly by van der Waals interactions between ligands and the receptor (i.e. a good geometric fit is important), and the net loss of ligand cavity energies (i.e. the burial of solvent accessible surface area is favourable). The net electrostatic interactions, however, are found to be anti-binding for this binding set. A secondary amide indicator is found to be necessary to account for the deficiencies in partial charges which are directly from quantum fitting, and it is shown that it represents a critical π-type hydrogen bond between NH group of nevirapine's secondary amide fragment and the phenyl ring of Tyr188A residue (contributing $\approx$2.25 kcal/mol favourably), which explains their otherwise surprising binding affinities. We have also performed more quantum charge fittings with the protein environment included [33], and similar results are obtained with the leave-one-out cross validation showing an RMSE of 1.02 kcal/mol and a correlation coefficient of 0.67, indicating that an out-of-plane polarisability might be needed to fully catch this π-type hydrogen bond in classical force fields. Finally, six novel ligands are designed for optimal binding based on our predictions for further experimental validations.

2. Method and system

2.1 LIE–SGB method

We have developed an implementation of the LIE method, in the context of the IMPACT program [34,35], using the SGB continuum solvation model [36] and the OPLSAA force field of Jorgensen and co-workers [37]. The details of the LIE–SGB method have been described previously [20], and we only give a brief review of the method here.

The LIE–SGB binding free energy can be described as a sum of three terms in the empirical formula based on electrostatic energy, van der Waals energy, and cavity energy:

$$\begin{aligned}\Delta G = \alpha\Big(\langle U^b_{\mathrm{vdw}}\rangle - \Big\langle U^f_{\mathrm{vdw}}\Big\rangle\Big) \\ + \beta\Big(\langle U^b_{\mathrm{elec}}\rangle - \Big\langle U^f_{\mathrm{elec}}\Big\rangle\Big) \\ + \gamma\big(\langle U^b_{\mathrm{cav}}\rangle - \langle U^f_{\mathrm{cav}}\rangle\big),\end{aligned} \tag{1}$$

where $\langle\cdots\rangle$ means ensemble averages from Monte Carlo or molecular dynamics (MD) simulations. All terms are evaluated only for interactions between the ligand and its environment, with '*f*' denoting the free state and '*b*' denoting the bound state, and α, β and γ being the LIE fitting parameters. It should be noted that earlier LIE methods based on explicit solvent models [9,10] use only two energy terms in the fitting, the van der Waals and electrostatic energies, while later explicit solvent-based LIE models use more terms including non-physical interaction energy terms, such as the number of hydrogen bonds and hydrophobic solvent accessible surface area [15]. In the continuum solvent models, part of the solute–solvent van der Waals interactions are implicitly included in the cavity term, so a third cavity term is necessary to catch that part of interactions. The use of the continuum solvent model provides a much more rapid convergence of the simulations, and a potentially higher accuracy in some cases for the long-range electrostatic forces [20]. For example, there is no need to keep the protein system neutral, as is done in the explicit solvent models to avoid the Born correction due to the finite size of the solvent sphere (typically a radius of 20 Å from the active site) [9,14].

In the generalised Born models, there are two so called reaction field energy and cavity energy terms in the total sum of the solvation free energy [36,38]

$$U_{\mathrm{SGB}} = U_{\mathrm{rxn}} + U_{\mathrm{cav}}, \tag{2}$$

and there is no explicit electrostatic energy and no van der Waals energy between solute and solvent in implicit solvent models. The van der Waals energy is implicitly included in the cavity term, while the reaction field energy (or the charging free energy) will be half the Coulombic energy between solute and solvent if the LIE approximation is exact. Even though this might not be true for complicated systems such as the solvation of ligands in the binding complex (bound state), it is proved to be the case for the solvation of small molecules (free state) by Jorgensen and co-workers [13]. Therefore, we use the following approximation [20]

$$U_{\mathrm{elec}} \approx U_{\mathrm{coul}} + 2U_{\mathrm{rxn}}, \tag{3}$$

as our total electrostatic energy term in LIE–SGB, where U_{coul} stands for the Coulomb interaction between the ligand and protein [20]. This gives a total of four possible LIE interaction components, van der Waals energy and Coulomb energy between the ligand and the protein (they are both zero in the free state), the reaction field energy and cavity energy between the ligand and continuum solvent. The Coulomb energy and reaction field energy are combined as a total electrostatic energy term in LIE–SGB, as shown in Equation (3).

The calculation of the reaction field energy for the ligand in the free state is straightforward [36]. However, the reaction field energy between the ligand and continuum solvent in the bound state is more complicated since SGB generally calculates and reports the total reaction field energy for the entire solute (i.e. for the ligand and protein together), even though the reaction field energy for the protein is not needed in our LIE–SGB equation. The total reaction field energy in SGB is expressed as

$$U_{\mathrm{rxn}} = \sum_i U_{\mathrm{se}}(q_i, \mathbf{r}_i) + \sum_{i<j} U_{\mathrm{pr}}(q_i, q_j, \mathbf{r}_i, \mathbf{r}_j), \tag{4}$$

where the single energy U_{se} is

$$\begin{aligned}U_{\mathrm{se}} = -\frac{1}{8\pi}(1 - 1/\varepsilon)\int_S \frac{q_k^2}{|\mathbf{R} - \mathbf{r}_k|^4} \\ (\mathbf{R} - \mathbf{r}_k)\cdot\mathbf{n}(\mathbf{r})\mathrm{d}^2\mathbf{R},\end{aligned} \tag{5}$$

and the pairwise screened Coulomb energy is

$$U_{\mathrm{pr}} = -(1 - 1/\varepsilon)\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 \mathrm{e}^{-D}}}, \tag{6}$$

with parameter $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$ (α_i and α_j are the Born radius) and parameter $D = r_{ij}^2/(2\alpha_{ij})^2$.

Therefore, for the reaction field energy U_{rxn} in the bound state, we accumulate the single energies U_{se} from ligand atoms only; and count the pairwise screened Coulombic energies U_{pr} as 100% if both atoms are from the ligand, and 50% if one is from the ligand and the other is from the protein, and zero if both are from the protein.

The obtained U_{rxn} energy is then combined with the ligand–receptor Coulomb energy to get the total electrostatic energy for the bound state using Equation (3).

Since the cavity energy in SGB depends solely on the total solvent accessible surface area [36,39],

$$U_{cav} = c_1 \times \text{SASA} + c_2, \tag{7}$$

where c_1 and c_2 are empirical coefficients, with $c_1 = 0.00486$ kcal/mol Å^2 and $c_2 = 1.092$ kcal/mol [36,39], we can easily convert γ to be based on the solvent accessible surface area ($\gamma = c_1 \times \gamma$), which can then be more directly compared with previous explicit solvent models [14,20]. Thus, in Section 3, we use the following expression for the binding free energy instead

$$\Delta G = \alpha\left(\langle U^b_{vdw}\rangle - \left\langle U^f_{vdw}\right\rangle\right) + \beta\left(\langle U^b_{elec}\rangle - \left\langle U^f_{elec}\right\rangle\right) + \gamma(\langle \text{SASA}^b\rangle - \langle \text{SASA}^f\rangle). \tag{8}$$

In other words, we report the γ parameter based on the solvent accessible surface area rather than the cavity energy itself. For more complicated cavity energy terms in other continuum solvent models [38], the cavity energy can depend on each atom's type (i.e. the coefficient c_1 can depend on atom types), and thus the first term in the cavity energy becomes a sum over all atoms, and the γ parameter based on total SASA cannot be defined. For such cases, the original γ based on the cavity energy should be used.

2.2 *Systems and models*

The total binding set consists of more than 50 ligands (more below), with the training subset consisting of 40 ligands (20 HEPT analogues and 20 nevirapine analogues). The molecular structures of the HEPT and nevirapine analogues are shown in Figure 1, with the substitute groups R1, R2 and R3 listed in Tables 1 and 2. In the simplified notations used in Tables 1 and 2, 'Me' stands for the methyl group, and 'Et' for the ethyl group, 'Pr' for the propyl group, 'Bu' for the butyl group, and 'Ph' for the phenyl group. The experimental activities for both the HEPT [40–43] and nevirapine analogues [44] are also listed in Tables 1 and 2.

The initial system setup for the 40 ligand–receptor systems follows a similar approach as previously reported [15,20]. Since the protein HIV-1RT has a large size (about 980 residues in two chains), we only include the necessary nearby residues from the binding pocket in order to save the computational time. Using the initial co-crystal structure of MKC-442 (H11 in this study) binding with HIV-1RT (PDB entry 1rt1), a representative model was constructed by including only those residues within ~20 Å from atom C6 of the HEPT uracil core, which results in a reduced protein size of 123 residues. Hydrogen atoms were added, and clipped residues were then capped with acetyl (ACE) and methylamine (NMA) groups. The capping residues and residues with backbone atoms outside the 15 Å sphere from C6 were constrained with a harmonic potential with a force constant of 100 kcal/mol/Å^2). Other HEPT analogues were then built based on the MKC-442 template (after the minimisation and equilibration of MKC-442 within the complex system). See Figure 2 for one example of the docked protein system, ligand HEPT (H01) in HIV-1RT, which shows both an overview of the system and a closer view to the binding pocket. The nevirapine analogues were treated similarly starting from the coordinates of the X-ray structure of nevirapine bound to HIVRT (PDB entry 1vrt). The residues included in both cases are made identical, and hence the protein atom size is the same in both HEPT analogues and nevirapine analogues in the following. The force field used for the various ligands and the protein is OPLSAA [37], except that the charges for ligands are from quantum HF/6-31G* ChelpG fitting using Gaussian95 [45] because there are no generic charge types for these ligands in OPLSAA yet. These charges, which are directly from quantum electrostatic potential fitting, are not like typical OPLSAA charges which undergo series of fine tunings by running numerous tests against experiments including

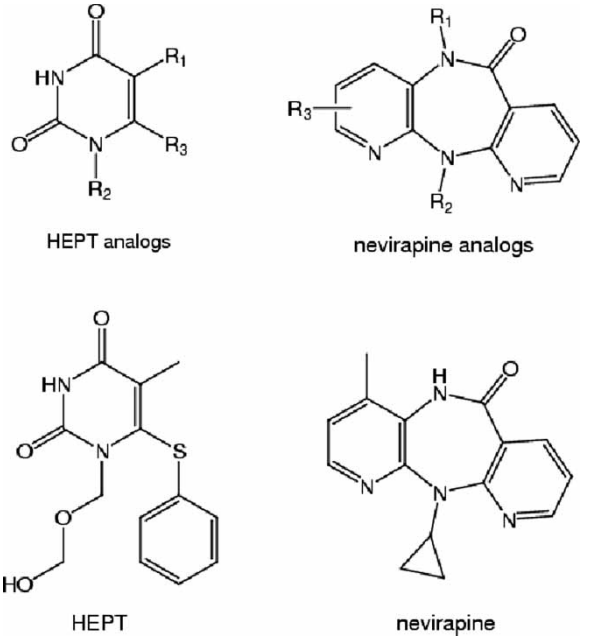

Figure 1. Molecular structures of HEPT and nevirapine analogues binding to HIV-1 reverse transcriptase. See Tables 1 and 2 for details of R1, R2 and R3 groups.

Table 1. Molecular structures of HEPT analogues, with R1, R2, and R3 groups shown in Figure 1.

Number	R1	R2	R3	EC_{50}	ΔG_{bind}
H01	Me	$CH_2OCH_2CH_2OH$	SPh	7.0	−7.32
H02	Me	$CH_2OCH_2CH_2CH_3$	SPh	3.6	−7.73
H03	Me	$CH_2OCH_2CH_3$	SPh	0.33	−9.20
H04	Me	CH_2OCH_3	SPh	2.1	−8.06
H05	Me	CH_2OCH_2Ph	SPh	0.088	−10.01
H06	i-Pr	CH_2OCH_2Ph	SPh	0.0027	−12.16
H07	Me	Et	SPh	2.2	−8.03
H08	Me	Me	SPh	>150	>−5.43
H09	Et	$CH_2OCH_2CH_3$	SPh	0.019	−10.96
H10	i-Pr	$CH_2OCH_2CH_3$	SPh	0.012	−11.24
H11	i-Pr	$CH_2OCH_2CH_3$	CH_2Ph	0.0042	−11.89
H12	c-Pr	$CH_2OCH_2CH_3$	SPh	0.1	−9.93
H13	Me	$CH_2OCH_2CH_2OH$	CH_2Ph	23.0	−6.52
H14	Me	$CH_2OCH_2CH_2OH$	OPh	85.0	−5.78
H15	Me	$CH_2OCH_2CH_2OH$	SPh-3,5 di-Me	0.26	−9.35
H16	Et	$CH_2OCH_2CH_2OH$	SPh-3,5 di-Me	0.013	−11.19
H17	i-Pr	$CH_2OCH_2CH_2OH$	SPh-3,5 di-Me	0.0027	−12.16
H18	Et	CH_2OCH_2Ph	SPh	0.0059	−11.68
H19	Me	H	SPh	>250	>−5.11
H20	Me	Bu	SPh	1.2	−8.40

The experimental activities are taken from the works of Tanaka et al. [40–43], EC_{50} in μM at 310 K, and $\Delta G_{bind} \approx RT\ln(EC_{50})$ in kcal/mol.

liquid simulations to account for polarisations and many-body interactions, etc. Though very successful for many ligands such as HEPT analogues studied before [20], these quantum fitted charges might have some difficulty in treating other cases such as ligands with secondary amides and tertiary amides as we will soon see. Interested readers can refer to the previous paper [15] for more details.

2.3 Simulation details

Both MD and the hybrid Monte Carlo method [46,47] have been used for the underlying LIE conformation sampling in the present work [20]. The difference in the final LIE–SGB binding free energies is minimal from these two sampling techniques. Unless explicitly specified, the ensemble averages reported here are obtained from MD

Table 2. Molecular structures of nevirapine analogues with R1, R2, and R3 groups shown in Figure 1.

Number	R1	R2	R3	IC_{50}	ΔG_{bind}
N01	Me	Et	H	0.125	−9.42
N02	Me	Et	2-Me	0.17	−9.24
N03	Me	Et	2-Cl	0.15	−9.31
N04	Me	Et	3-Me	0.76	−8.35
N05	Me	Et	3-Cl	>1.0	>−8.19
N06	Me	Et	4-Me	1.9	−7.81
N07	H	Et	H	0.44	−8.67
N08	H	Et	4-Me	0.035	−10.17
N09	H	Et	4-Cl	0.095	−9.58
N10	H	c-Pr	4-Me	0.084	−9.65
N11	Me	c-Pr	4-Me	>1.0	>−8.19
N12	Me	Pr	H	0.45	−8.66
N13	Me	t-Bu	H	11.0	−6.77
N14	Me	$COCH_3$	H	15.3	−6.57
N15	H	Et	4-Et	0.11	−9.49
N16	Me	CH_2SCH_3	H	0.85	−8.28
N17	H	c-Pr	4-CH_2OH	3.0	−7.54
N18	H	c-Pr	4-CN	1.25	−8.05
N19	Me	CH_2CH_2F	H	2.9	−7.56
N20	H	c-Pr	H	0.45	−8.66

The experimental activities are taken from works of Hargrave et al. [44], IC_{50} in μM at 298 K, and $\Delta G_{bind} \approx RT\ln(IC_{50})$ in kcal/mol.

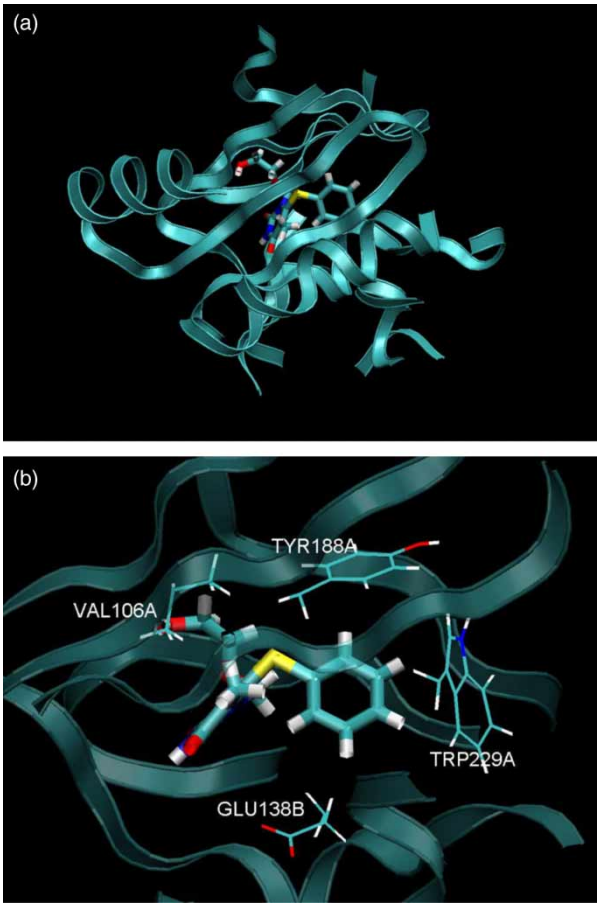

Figure 2. Structure of the docked HEPT (H01) inside the active site of HIV-1RT. (a) An overview of the system (the ligand is shown in heavy sticks). (b) A closer view of the binding pocket, with key residues involved in binding (within 5 Å from the ligand) shown in detail while the rest shown in ribbons.

sampling. All simulations are carried out with the IMPACT package [34]. A conjugate gradient minimisation is performed first, starting from the initial docked structures; and then a 100 ps MD equilibration is followed with temperature smoothly increasing from 0 to 310 K for HEPT analogues (experimental data at 310 K), and from 0 to 298 K for nevirapine analogues (experimental data at 298 K) by Berendsen [48] velocity rescaling and Andersen [49] velocity resampling; finally a 3 ns MD simulation with time step 2.0 fs is run for data collection (longer than our previous simulations [20] for better convergence). No cut-off is used for the long-range electrostatic interactions in the simulations, and a dielectric constant of 1.0 is used for the protein and ligands and 80.0 for the water solvent in the SGB model. The capped residues at clips in HIV-1RT are restrained during MD simulations using a harmonic constraint potential with a force constant of 25 kcal/mol. The details of a few explicit solvent simulations used for comparison are described in the results section.

3. Results and discussion

3.1 Energy components

As explained in Section 2, there are a total of four possible LIE energy components in our continuum solvent-based LIE–SGB method: Coulomb energy, van der Waals energy, reaction field energy and cavity energy (the Coulomb and van der Waals energies in the free state are zero). In order to determine how long the MD simulation is needed for the four LIE components to converge, we have plotted the time averages for each component versus MD time for HEPT analogue H01 in Figure 3(a). The cavity energy and van der Waals energy converge faster than the Coulomb energy and reaction field energy. This is understandable because electrostatic interactions are long-range interactions and thus converge more slowly in general. On the other hand, the van der Waals interactions are typically regarded as short-range forces which die off quickly, and the cavity energy is normally based on the solvent accessible surface area which depends on the overall molecular shape and typically converges very fast as well. In this particular case, the cavity energy converges for H01 in less than 100 ps in both the free and bound states, and the van der Waals energy converges in about 1–2 ns, the reaction field energy converges in less than 2 ns in both the free and bound states, and the long-range Coulomb energy also converges fairly well after 2 ns (convergence here means the average energy component changes no more than 0.25 kcal/mol when doubling the data collection time). We have also examined other ligands and the conclusion is basically the same – a 3 ns MD data collection run yields reasonably converged results. For convergence tests, we have also run longer (e.g. 6 ns) MD simulations, as shown in this test case (Figure 3).

For comparison, we also plotted the time average of Coulomb and van der Waals interactions for the explicit solvent-based LIE model in Figure 3(b). A 20 Å water sphere (SPC water model) around the ligand for the explicit solvation is used [20]. For the ligand H01 (HEPT), this gives a total size of 3604 atoms for the free state and 3298 atoms for the bound state. Similarly, the solvated systems are first minimised with conjugate gradient minimiser and then equilibrated for 100 ps for both the free and bound states. Larger fluctuations in Coulomb and van der Waals energies are observed. For example, in the free state, it takes up to 20 ns before the electrostatic energy fully converges. The bound state shows relatively smaller fluctuations, since the ligand is confined in the cavity of the active site thus the number of conformations accessible is somewhat reduced, but it still takes about 4 ns to converge. It is interesting to note that for the ligand in the free state, the converged reaction field energy from continuum solvent model (which is the

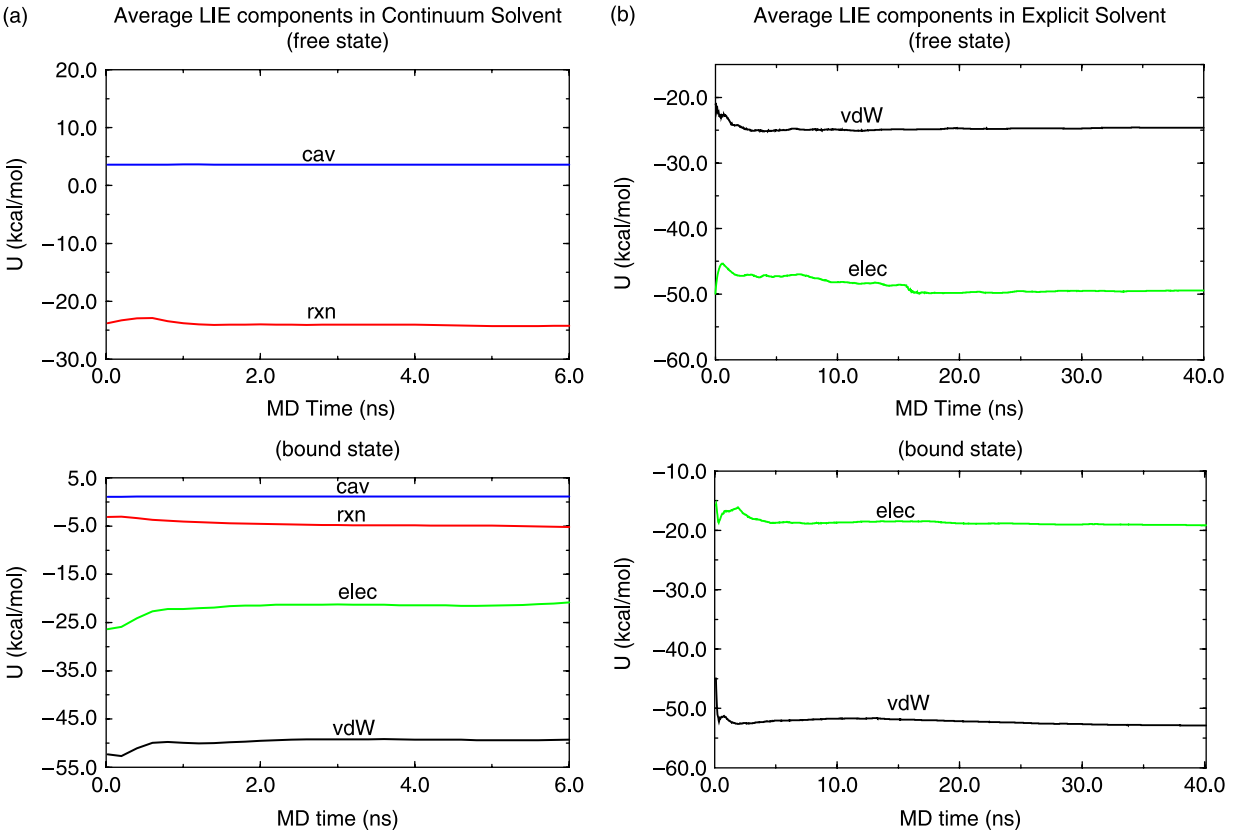

Figure 3. Time averages of the LIE interaction components versus MD time for H01 in both the free state and bound state. (a) Continuum solvent model: time averages up to 6 ns MD data collection are shown. (b) Explicit solvent model: time averages up to 40 ns MD data collection are shown.

charging free energy of solvation) is −23.0 kcal/mol while the Coulomb energy from the explicit solvent model is −49.1 kcal/mol (see Figure 3). This indicates a linear response coefficient β of 0.47 (23.0/49.1) for this ligand's solvation free energy, which is consistent with the previous findings [10,13] that for less complex systems such as the solvation free energy of small ligands, the linear response method works very well and a coefficient β of close to 0.5 can be recovered. For more complicated systems and processes, however, the β coefficient can deviate significantly from 0.5 as we have seen in many cases [10,14,15,20]. Overall, similar to the previous results [20], these data show our continuum solvent-based LIE method is more than one order of magnitude faster than the explicit solvent-based LIE methods for a comparable convergence. It is worth pointing out that proteins in the explicit solvent used in our above calculations and many others' previous LIE studies are only 'partially solvated', since a water sphere of about 20 Å around the ligand is often used in these models [9,13–15]. There are essentially no solvent molecules around the protein beyond that sphere, so the protein is not completely solvated. This might affect the conformations of the receptor near the boundary. Since, the electrostatic interactions are very long-ranged, this kind of treatment for solvent may induce large fluctuations in LIE components and other artefacts, especially if there are net charges. If the entire protein system needs to be solvated and a periodic boundary condition is to be used, then the bound state in explicit-solvent-based LIE will take much longer CPU time, and the speedup will be even greater.

Table 3 lists the time averages of the LIE–SGB energy components for the HIV-1RT binding set. Notice that the Coulomb and van der Waals energies in the free state, U^f_{coul} and U^f_{vdw}, are all zero as explained above. Overall, in going from the free state to the bound state, the magnitude of the reaction field energy and cavity energy decreases as a result of the burial of the ligand in the binding pocket; on the other hand, the magnitude of the Coulomb energy and van der Waals energy increase (from zero) to compensate for the loss of the former two energies. A linear combination of the differences in the above ensemble averaged energy terms from the two states gives the LIE binding affinity.

Table 3. The time averages of LIE interaction components for HEPT and nevirapine analogues binding to HIV-1RT (all energies are in kcal/mol).

Lig	$\langle U^f_{coul}\rangle$	$\langle U^f_{vdw}\rangle$	$\langle U^f_{rxn}\rangle$	$\langle U^f_{cav}\rangle$	$\langle U^b_{coul}\rangle$	$\langle U^b_{vdw}\rangle$	$\langle U^b_{rxn}\rangle$	$\langle U^b_{cav}\rangle$	$I_{2\text{-amide}}$
H01	0.0	0.0	−22.967	3.648	−24.765	−47.227	−3.133	1.098	0
H02	0.0	0.0	−19.116	3.659	−11.019	−51.766	−6.084	1.110	0
H03	0.0	0.0	−20.254	3.601	−13.759	−49.404	−5.668	1.097	0
H04	0.0	0.0	−19.445	3.420	−12.438	−49.149	−5.614	1.103	0
H05	0.0	0.0	−21.310	3.872	−16.977	−58.484	−2.733	1.110	0
H06	0.0	0.0	−20.651	4.161	−16.157	−63.157	−4.619	1.107	0
H07	0.0	0.0	−18.095	3.344	−13.680	−46.266	−4.312	1.095	0
H08	0.0	0.0	−18.494	3.225	−14.918	−44.929	−3.189	1.098	0
H09	0.0	0.0	−21.060	3.613	−15.125	−53.478	−7.016	1.098	0
H10	0.0	0.0	−22.021	3.782	−15.915	−51.081	−6.636	1.095	0
H11	0.0	0.0	−21.171	3.707	−15.380	−52.597	−6.050	1.094	0
H12	0.0	0.0	−21.701	3.780	−14.500	−52.467	−5.275	1.103	0
H13	0.0	0.0	−25.014	3.558	−23.204	−48.898	−4.047	1.096	0
H14	0.0	0.0	−24.748	3.494	−20.121	−47.898	−4.962	1.094	0
H15	0.0	0.0	−24.690	3.954	−21.300	−52.915	−5.661	1.101	0
H16	0.0	0.0	−25.958	4.054	−22.714	−57.412	−6.157	1.096	0
H17	0.0	0.0	−23.442	4.032	−20.313	−59.537	−5.708	1.098	0
H18	0.0	0.0	−20.912	4.044	−15.343	−60.402	−5.777	1.100	0
H19	0.0	0.0	−21.539	3.126	−15.935	−43.689	−4.122	1.093	0
H20	0.0	0.0	−18.657	3.639	−17.906	−45.385	−4.445	1.107	0
N01	0.0	0.0	−14.247	3.313	−8.928	−46.568	−5.623	1.092	0
N02	0.0	0.0	−14.803	3.447	−6.654	−42.568	−5.659	1.125	0
N03	0.0	0.0	−14.303	3.419	−7.181	−48.514	−6.718	1.103	0
N04	0.0	0.0	−14.664	3.457	−9.164	−47.026	−4.705	1.097	0
N05	0.0	0.0	−13.639	3.420	−4.264	−46.914	−6.071	1.110	0
N06	0.0	0.0	−14.455	3.400	−4.256	−42.200	−6.889	1.131	0
N07	0.0	0.0	−17.153	3.198	−9.197	−47.302	−5.441	1.092	1
N08	0.0	0.0	−17.614	3.319	−9.035	−48.214	−6.135	1.094	1
N09	0.0	0.0	−15.441	3.299	−5.705	−47.089	−5.596	1.093	1
N10	0.0	0.0	−17.980	3.389	−5.586	−46.427	−5.877	1.095	1
N11	0.0	0.0	−14.894	3.459	−6.167	−49.771	−4.828	1.101	0
N12	0.0	0.0	−14.335	3.459	−5.052	−45.459	−5.849	1.106	0
N13	0.0	0.0	−15.365	3.399	−8.435	−51.738	−5.317	1.093	0
N14	0.0	0.0	−21.103	3.285	−13.000	−45.088	−7.680	1.093	0
N15	0.0	0.0	−16.869	3.458	−7.008	−46.486	−6.263	1.094	1
N16	0.0	0.0	−15.896	3.474	−6.247	−49.059	−5.228	1.092	0
N17	0.0	0.0	−22.908	3.442	−8.094	−43.816	−7.957	1.114	1
N18	0.0	0.0	−17.090	3.420	−3.888	−48.268	−6.653	1.095	1
N19	0.0	0.0	−14.342	3.325	−9.808	−45.260	−4.216	1.099	0
N20	0.0	0.0	−17.620	3.258	−6.804	−43.055	−5.696	1.093	1

The data are collected from a 3 ns MD simulation. Each ligand shows the LIE components for both free and bound states, with '*f*' denoting the free state and '*b*' denoting the bound state. The last column is the secondary amide indicator, which has a value of 1 if there is an secondary amide group in the ligand and 0 otherwise.

3.2 Three parameters model

Using the three parameter model presented in Section 2, a least square fit based on singular value decomposition is performed to obtain the binding affinities of the 40-ligand training set. The following LIE formula is obtained,

$$\Delta G = 0.0919\left(\langle U^b_{vdw}\rangle - \left\langle U^f_{vdw}\right\rangle\right) + 0.164\left(\langle U^b_{elec}\rangle - \left\langle U^f_{elec}\right\rangle\right) + 0.0134(\langle \mathrm{SASA}^b\rangle - \langle \mathrm{SASA}^f\rangle). \quad (9)$$

The results are shown in Figure 4, which plots the LIE–SGB binding energy predictions versus the experimental data. If a predicted binding energy agrees exactly with the experimental value, a data point (represented by diamonds) will appear precisely on the diagonal line. To help with the visualisation of these data points, a lower and upper bound line are also plotted in the figure (dashed lines), with 1.0 kcal/mol below and above the experimental values. From the figure, the majority of the data points are within or close to these two bound lines, which means that most of them have either less than or about 1.0 kcal/mol error. The numerical binding affinity

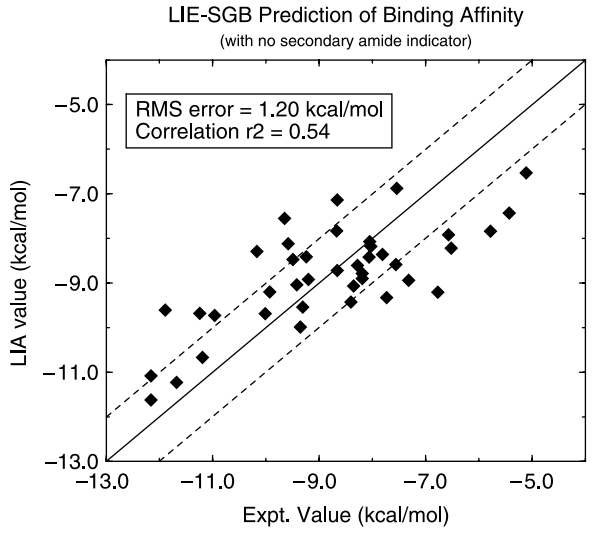

Figure 4. LIE binding energies for the HEPT and nevirapine analogues binding to HIV-1RT using the traditional three energy term model without a secondary amide indicator. The overall RMSE is 1.19 kcal/mol, with correlation $r^2 = 0.54$. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line. The two dashed lines are for lower and upper bounds, with 1.0 kcal/mol below and above the experimental values.

results are also summarised in Table 4. The overall RMS deviation of the LIE–SGB predictions for this 40 ligand binding set is 1.19 kcal/mol and the average unsigned error is 1.01 kcal/mol, which is satisfactory given that a simple model was applied to such a large binding set. However, the correlation coefficient r^2 is only 0.54, which is somewhat low. This may be due to the fact that the binding affinity orders of many nevirapine analogues are not being correctly predicted. Even though the model easily distinguishes the low binders from the high binders, it fails to distinguish many ligands with close binding affinities, particularly those nevirapine analogues with secondary amide groups, N07, N08, N09, N10, N15, N17, N18 and N20, all of which are underestimated. A more careful examination reveals that there exist some nevirapine analogue pairs, such as N08 versus N06, and N10 versus N11, which differ only in R1 group, from 'H' to 'Me', i.e. from secondary amide fragment to tertiary amide fragment. Their experimental binding affinities, however, differ significantly, with N08 2.36 kcal/mol more potent than N06, and N10 1.46 kcal/mol more potent than N11. In other words, those with secondary amide groups (N08, N10) show significantly higher binding affinities than those with tertiary amide groups (N06, N11). However, the LIE–SGB predictions show exactly the opposite, with N08 0.07 kcal/mol less potent than N06 and N10 1.24 kcal/mol less potent than N11.

3.3 Quantum charge deficiencies with secondary amides

It turns out that the HF/6-31G* ChelpG fitted charges have some deficiency in treating secondary amides versus tertiary amides. Both the previous study with explicit solvent model (TIP4P water) [15] and the current work with continuum solvent model found that acceptable correlations could not be obtained for the nevirapine analogues unless an indicator variable was included for secondary amides. Since most ligands in this binding set have no generic OPLSAA charge types parameterised yet, we resorted to quantum electrostatic potential fitted charges for these ligand atoms. As mentioned earlier, these charges are not well parameterised as compared to the rest of the OPLSAA force field, for example, no liquid simulations have been run to fine-tune these charges to account for polarisations and many-body interactions. Clearly, these quantum charges in nevirapine analogues are causing some difficulties in treating secondary amides versus tertiary amides. It seems that using these charges will overestimate the solvation (hydration) free energy differences between the secondary amide and the tertiary amide in the free state, and these charges might also underestimate the many-body interactions (implicitly expressed in charges, since the force field is pairwise potential) shown in a π-type hydrogen bond between the NH group in a secondary amide and the π electrons in a phenyl ring (see more discussions below) [15]. As pointed out in a previous study by Rizzo et al. [50], one way to check this is to calculate the hydration free energy change of a model tertiary amide to secondary amide conversion, DMA→NMA, using these HF/6-31G* ChelpG fitted charges by FEP. The computed hydration free energy change is -2.47 ± 0.24 kcal/mol, while the experimental value is -1.53 kcal/mol [51]. Thus, there is about a 1 kcal/mol overestimation in the difference. By analogy, nevirapine analogues with secondary amides would be expected to be too well hydrated in the free state (too negative reaction field energies in LIE–SGB terms) and, hence, pay an artificially high desolvation penalty upon binding and thus a lower binding affinity. So an indicator similar to the one used in the previous paper [15], $I_{\text{2-amide}}$, was introduced to the LIE–SGB equation, that is for ligands with secondary amide fragments it has a value of 1, and for ligands with no secondary amide fragments it has a value of 0. See Table 3 for the details of which ligands have the indicator value of 1, namely N07, N08, N09, N10, N15, N17, N18 and N20. From the above analysis, it is also easy to see that one should expect the coefficient of the secondary amide indicator term to be negative (i.e. a negative δ coefficient) to correct the artificially high desolvation penalty for secondary amides.

Table 4. The LIE–SGB binding affinity predictions without the secondary amide indicator.

Ligand	Expt.	LIE	Jackknife
H01	−7.32	−8.94	−9.11
H02	−7.73	−9.33	−9.37
H03	−9.20	−8.92	−8.91
H04	−8.06	−8.42	−8.43
H05	−10.01	−9.69	−9.66
H06	−12.16	−11.62	−11.59
H07	−8.03	−8.18	−8.19
H08	−5.43	−7.43	−7.54
H09	−10.96	−9.73	−9.64
H10	−11.24	−9.68	−9.56
H11	−11.89	−9.61	−9.54
H12	−9.93	−9.20	−9.16
H13	−6.52	−8.22	−8.35
H14	−5.78	−7.84	−8.05
H15	−9.35	−9.99	−10.07
H16	−11.19	−10.67	−10.64
H17	−12.16	−11.08	−11.03
H18	−11.68	−11.23	−11.20
H19	−5.11	−6.53	−6.78
H20	−8.40	−9.43	−9.68
N01	−9.42	−9.04	−9.00
N02	−9.24	−8.41	−8.32
N03	−9.31	−9.54	−9.57
N04	−8.35	−9.07	−9.11
N05	−8.19	−8.90	−8.94
N06	−7.81	−8.36	−8.41
N07	−8.67	−7.83	−7.72
N08	−10.17	−8.29	−8.16
N09	−9.58	−8.12	−8.05
N10	−9.65	−7.55	−7.36
N11	−8.19	−8.79	−8.82
N12	−8.66	−8.72	−8.72
N13	−6.77	−9.21	−9.63
N14	−6.57	−7.92	−7.96
N15	−9.49	−8.47	−8.44
N16	−8.28	−8.61	−8.62
N17	−7.54	−6.88	−6.67
N18	−8.05	−8.07	−8.07
N19	−7.56	−8.59	−8.64
N20	−8.66	−7.14	−7.03
RMS error		1.19	1.31
Ave error		1.01	1.09

Both the LIE fitting and leave-one-out cross validation (Jackknife) results for HEPT and nevirapine analogues binding to HIV-1RT are listed (all energies are in kcal/mol). An overall RMS error of 1.19 kcal/mol is obtained for this LIE fitting, and an RMS error of 1.31 kcal/mol for the cross validations. Even though the RMS error in LIE fitting is satisfactory given that a simple model was used for this reasonably large binding set, the correlation coefficient r^2 is somewhat low, only 0.54. See the next table for fitting results with a secondary amide indicator which corrects the deficiencies in HF/6-31G* ChelpG charges.

It should be pointed out that before the charge deficiency and secondary amide indicator were identified, we tried many other descriptors. For example, we tried to add a constant to the three traditional descriptors, van der Waals energy, electrostatic energy and cavity energy. The results show no meaningful improvement at all. The RMSE is found to decrease from 1.19 to 1.18 kcal/mol, and the correlation coefficient r^2 to increase from 0.54 to 0.55. Thus, adding a new term into the fitting equation does not necessarily improve the results if there is no correlation between the descriptor and the binding affinity, i.e. if there is no real physical meaning or no force field parameters improvement behind the descriptor. We also tried some other descriptors, such as separating the reaction field energy and the Coulomb energy to be two separate electrostatic terms, and the fitting results are slightly better, but there were no significant improvements either. Only this secondary amide indicator which identifies the charge deficiency for amides improves the results dramatically. Rizzo et al. [15,52] also tried other non-physical interaction energy descriptors, such as the number of hydrogen bonds and hydrophobic solvent accessible surface area, in their explicit solvent LIE models. Here, we stick with the traditional physical energy terms, the van der Waals energy, electrostatic energy and the cavity energy, as the core terms in our LIE–SGB model (with the addition of the following secondary amide indicator for nevirapine analogues only to handle the quantum charge deficiencies).

It should also be pointed out that we have performed more quantum charge fittings with the protein environment included [33], but no significant improvement was observed (more results below), indicating that there might be a more subtle problem with the point charge approach (i.e. the out-of-plane polarisability of benzene rings cannot

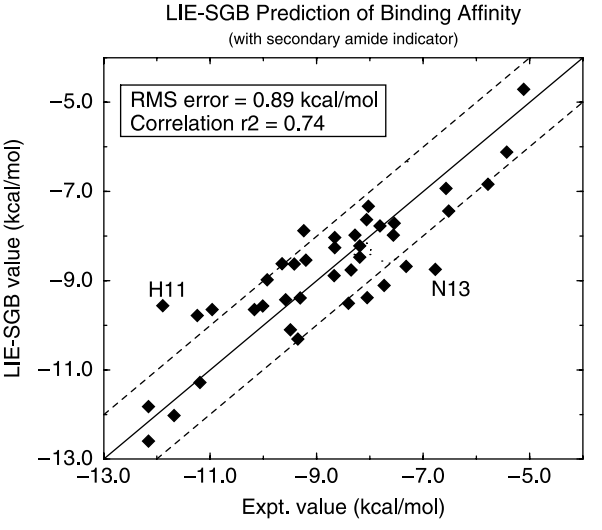

Figure 5. LIE binding energies for the HEPT and nevirapine analogues binding to HIV-1RT using the traditional three energy term model plus a secondary amide indicator, similar to the previous LIE method based on an explicit solvent model [15]. The overall RMSE is 0.89 kcal/mol, with correlation $r^2 = 0.74$. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line. The two dash lines are for lower and upper bounds, with 1.0 kcal/mol below and above the experimental values.

Table 5. The LIE–SGB binding affinity predictions with the secondary amide indicator included.

Ligand	Expt.	LIE-SGB	$\alpha\Delta U_{vdw}$	$\beta\Delta U_{ele}$	$\gamma\Delta$SASA	$\delta I_{2\text{-amide}}$	Jackknife
H01	−7.32	−8.68	−4.93	3.78	−12.74	0.00	−8.82
H02	−7.73	−9.11	−5.41	3.82	−12.74	0.00	−9.16
H03	−9.20	−8.54	−5.16	3.91	−12.51	0.00	−8.52
H04	−8.06	−7.63	−5.13	3.86	−11.58	0.00	−7.60
H05	−10.01	−9.57	−6.11	5.12	−13.80	0.00	−9.47
H06	−12.16	−12.60	−6.60	4.04	−15.26	0.00	−12.74
H07	−8.03	−7.33	−4.83	3.52	−11.24	0.00	−7.29
H08	−5.43	−6.12	−4.69	3.98	−10.63	0.00	−6.22
H09	−10.96	−9.65	−5.58	3.29	−12.57	0.00	−9.54
H10	−11.24	−9.78	−5.33	3.77	−13.43	0.00	−9.65
H11	−11.89	−9.56	−5.49	3.77	−13.06	0.00	−9.46
H12	−9.93	−8.98	−5.48	4.66	−13.38	0.00	−8.89
H13	−6.52	−7.44	−5.11	4.75	−12.30	0.00	−7.54
H14	−5.78	−6.84	−5.00	4.94	−11.99	0.00	−7.01
H15	−9.35	−10.31	−5.53	4.25	−14.26	0.00	−10.48
H16	−11.19	−11.28	−6.00	4.28	−14.78	0.00	−11.29
H17	−12.16	−11.82	−6.22	3.85	−14.66	0.00	−11.76
H18	−11.68	−12.02	−6.31	3.79	−14.71	0.00	−12.08
H19	−5.11	−4.71	−4.56	4.80	−10.16	0.00	−4.55
H20	−8.40	−9.51	−4.74	2.67	−12.65	0.00	−9.80
N01	−9.42	−8.63	−4.86	2.11	−11.10	0.00	−8.51
N02	−9.24	−7.88	−4.45	2.95	−11.60	0.00	−7.65
N03	−9.31	−9.39	−5.07	2.03	−11.57	0.00	−9.41
N04	−8.35	−8.76	−4.91	2.73	−11.79	0.00	−8.79
N05	−8.19	−8.47	−4.90	2.76	−11.54	0.00	−8.49
N06	−7.81	−7.77	−4.41	2.76	−11.34	0.00	−7.76
N07	−8.67	−8.89	−4.94	3.61	−10.53	−2.25	−8.94
N08	−10.17	−9.65	−5.03	3.53	−11.12	−2.25	−9.55
N09	−9.58	−9.43	−4.92	3.55	−11.03	−2.25	−9.40
N10	−9.65	−8.62	−4.85	4.72	−11.46	−2.25	−8.45
N11	−8.19	−8.22	−5.20	3.54	−11.78	0.00	−8.22
N12	−8.66	−8.26	−4.75	3.02	−11.76	0.00	−8.23
N13	−6.77	−8.75	−5.40	2.96	−11.52	0.00	−9.12
N14	−6.57	−6.93	−4.71	3.51	−10.96	0.00	−6.96
N15	−9.49	−10.10	−4.85	3.60	−11.81	−2.25	−10.22
N16	−8.28	−7.98	−5.12	3.83	−11.90	0.00	−7.97
N17	−7.54	−7.71	−4.58	5.53	−11.63	−2.25	−7.78
N18	−8.05	−9.38	−5.04	4.31	−11.62	−2.25	−9.58
N19	−7.56	−7.98	−4.73	2.65	−11.13	0.00	−8.02
N20	−8.66	−8.03	−4.50	4.32	−10.82	−2.25	−7.92
RMS error		0.89					1.00
Ave error		0.71					0.81

Both the LIE fitting and leave-one-out cross validation (Jackknife) results are listed (all energies are in kcal/mol). An overall RMS error of 0.89 kcal/mol is achieved for this LIE fitting with a reasonable correlation coefficient r^2 of 0.74. The Jackknife cross validation test shows an RMS error of 1.00 kcal/mol and a correlation coefficient of 0.69. For the LIE fitting, the detailed contributions from each of the four descriptors are also listed (the individual contributions for the Jackknife tests are omitted for simplicity).

be captured by point-charge models and thus π-type hydrogen bonds might not be fully represented as shown in previous studies as well [53,54]).

3.4 Four parameters model

After including this secondary amide indicator, the binding affinity prediction results have been improved dramatically. Using a similar equation to Rizzo et al. [15] with four descriptors and one constant term, an RMSE of 0.89 kcal/mol has been achieved with a correlation coefficient r^2 of 0.74. The following is the optimised LIE–SGB fitting equation:

$$\begin{aligned}\Delta G = {} & 0.104\left(\langle U^b_{vdw}\rangle - \left\langle U^f_{vdw}\right\rangle\right) \\ & + 0.253\left(\langle U^b_{elec}\rangle - \left\langle U^f_{elec}\right\rangle\right) \\ & + 0.0243(\langle \mathrm{SASA}^b\rangle - \langle \mathrm{SASA}^f\rangle) \\ & - 2.25 I_{2-\mathrm{amide}} + 5.22. \end{aligned} \quad (10)$$

The predicted LIE–SGB binding affinities are shown in Figure 5, and the numerical results are summarised in

Table 5. Table 5 also lists the detailed contributions from each descriptor in the LIE–SGB fitting. As can be seen, significant improvement has been achieved as compared to earlier results shown in Figure 4. The overall RMSE has been reduced from 1.19 to 0.89 kcal/mol, and the average unsigned error has been reduced from 1.01 to 0.71 kcal/mol. More importantly, the correlation coefficient r^2 has been improved from 0.54 to 0.74. The coefficient δ for the secondary amide is found to be -2.25 kcal/mol, which is indeed negative as expected. The magnitude of -2.25 kcal/mol is somewhat larger than that from the FEP calculation using the simple DMA→NMA model, but it is in line with Rizzo et al.'s result, -2.82 kcal/mol, based on the explicit solvent LIE model using four descriptors, the number of hydrogen bonds, hydrophobic solvent accessible surface area, van der Waals energy and secondary amide indicator [15]. Most of the analogues have less than or about 1.0 kcal/mol error as seen from the upper and lower bounds in Figure 5, except the two outliers H11 and N13. H11 and N13 show errors of 2.33 and 1.98 kcal/mol, respectively. We will examine these two in more detail below. Overall, this 0.89 kcal/mol RMSE from experiment is very reasonable for this type of LIE method. Note that this 0.89 kcal/mol RMSE is much better than the previous finding of 1.24 kcal/mol RMSE using an 'analogous' LIE approach based on the explicit solvent model [15], i.e. using the traditional three energy terms plus the secondary amide indicator [15]. The less impressive results from explicit solvent models might be related to the fact that the electrostatic energies in explicit solvent models might not be fully converged due to their limited simulation length (10 million Monte Carlo steps) and/or some boundary effects (only a 20 Å radius of water sphere was used). That is probably why the authors used other non-physical interaction terms as descriptors, such as the number of hydrogen bonds and hydrophobic solvent accessible surface area [15]. Nevertheless, our result of 0.89 kcal/mol RMSE is still slightly better than the explicit solvent results of 0.94 kcal/mol [15] after using non-traditional descriptors [15]. As mentioned above, we think this might indicate one advantage of the continuum solvent-based LIE methods: the energy terms in continuum solvent models converge not only faster but also better in some cases, since there is no size or boundary effects as in the explicit solvent models.

A cross validation is then used to further check how well the model can predict unknown binding energies. It is done by a so called Jackknife or leave-one-out test, that is to leave one ligand out of the training set and then use the parameters obtained from the rest of the binding set (39 ligands in this case) to predict the affinity of the one left out. Table 5 also lists the Jackknife cross-validation results, with each ligand's binding energy being predicted from the rest 39 ligands. The overall RMSE is 1.00 kcal/mol and the average unsigned error is 0.81 kcal/mol, which is very encouraging for this comparatively large binding set with only four descriptors. The correlation coefficient r^2 is 0.69, which also indicates a good prediction ability. There are also two outliers, which are again H11 (2.43 kcal/mol) and N13 (2.35 kcal/mol; more discussions below). We have also performed more quantum charge fittings with the protein environment included [33], and similar results are obtained with the leave-one-out cross validation showing an RMSE of 1.02 kcal/mol and a correlation coefficient of 0.67, indicating an out-of-plane polarisability might be needed to fully capture this π-type hydrogen bond in classical force fields.

3.5 Binding mechanism

It is found that for this binding set the van der Waals interactions between the ligand and the receptor contributes favourably (negative) to the binding affinity, and the net loss of cavity energy also contributes favourably to the binding affinity, but the net electrostatic energy contributes unfavourably (positive) to the binding affinity. In other words, there are stronger electrostatic interactions in the free state than the bound state, meaning that the ligands interact electrostatically more favourably with pure water than with the receptor HIV-1RT. This is also seen in the electrostatic interactions from our explicit solvent simulation on the H01 example, as shown in Figure 3(b), the electrostatic energy from the explicit solvent model in the free state is about -49.5 kcal/mol, but it is only about -19.4 kcal/mol in the bound state. This finding is also consistent with the previous conclusion from the explicit solvent-based LIE model [15], where the loss of hydrogen bonds which mimics the loss of electrostatic interactions is also found to be unfavourable. The ranges of contributions from each energy term, van der Waals energy, electrostatic energy and cavity energy, are 2.04, 2.45 and 5.10 kcal/mol respectively for HEPT analogues, and 0.99, 3.50 and 1.37 kcal/mol, respectively, for nevirapine analogues. The secondary amide indicator only applies to some nevirapine analogues and the contribution is 2.25 kcal/mol. Overall, these results agree quite well with the findings from the explicit solvent LIE model [15]. A closer look at each energy component for each ligand reveals more important facts about this binding set. In the following, we will make a series of comparisons between various ligands, which will help us not only understand the differences between these particular ligands, but also the binding mechanism of this entire HIV-1RT binding set.

Let us first examine some simple cases in the HEPT analogue set. One case is H06 versus H19, where H06 is the highest binder and H19 is the lowest binder; another is

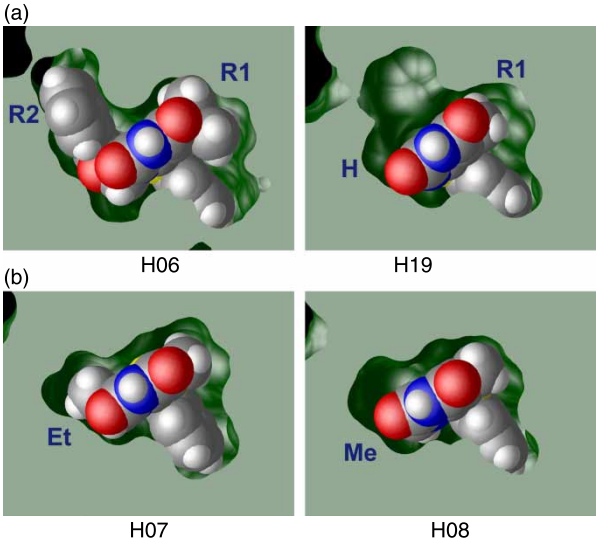

Figure 6. A closer look at the ligands H06 versus H19 and H07 versus H08 in the binding pocket. The figure shows the cross section surfaces of the receptor near the binding pocket when the complex is being cut in half (Connolly surface [55]), the ligand is represented in full atoms with CPK colour scheme. (a) H06 versus H19: it is clear that H06 has larger R1 (i-Pr vs. Me) and R2 (CH_2OCH_2Ph vs. H) groups than H19 and it has much better geometrical fit in the binding pocket than H19; (b) H07 versus H08: it is seen that H07's ethyl group (R2) fits better than the H08's methyl group (R2) in the binding pocket. A better geometrical fit means stronger van der Waals interactions between the ligand and the receptor.

H07 versus H08, where a 2.60 kcal/mol difference is seen in experimental binding affinity but with only a Me→Et difference in structure. The highest binder (most negative) H06 has i-Pr in R1 and CH_2OCH_2Ph in R2, while the lowest binder H19 has Me in R1 and H in R2 (they share the same R3 group SPh). The large size differences in R1 and R2 results in an about 7.0 kcal/mol difference in the binding affinity for H06 and H19. Figure 6(a) shows the ligands H06 and H19 in the binding pocket, with the receptor HIV-1RT represented by the cross section surfaces (Connolly surface [55]) when the complex is being cut in half, and the ligand represented by full atoms in CPK colour scheme. It is clear that H06 has a much better hydrocarbon packing with the surrounding receptor than H19, which yields a higher ligand–protein van der Waals interaction. Also H06 has a much greater loss in solvent accessible surface area upon binding to the pocket than H19, which reduces the cavity penalty for the bound state thereby enhancing the binding. From H19 to H06, the van der Waals energy has contributed about a net of 2.04 kcal/mol binding affinity, the loss of solvent accessible surface area or the cavity energy has contributed about 5.10 kcal/mol, and the electrostatic energy has contributed about 0.76 kcal/mol. Hence, the binding is mainly driven by the van der Waals interactions and the net loss of cavity energies. The electrostatic interactions only contribute slightly in this case. Since the cavity energy in continuum solvent models implicitly includes the van der Waals interactions between the ligand and water, it is understandable that the contribution from the cavity energy in continuum solvent-based LIE method is somewhat larger than that from the pure solvent accessible surface area term in the explicit solvent-based LIE model [15]. These results show that a good geometrical fit is very important for this binding set which typically means a larger van der Waals interaction, and the burial of the solvent accessible surface area is also favourable for the binding. Similarly, Figure 6(b) shows the difference in geometrical packing of H07 and H08 in the binding pocket, with H07 having a larger Et group and H08 a smaller Me group in the R2 position. Clearly H07 has a better geometrical fit than H08. H07 is found to be 1.2 kcal/mol more potent than H08 in LIE–SGB, in qualitative agreement with experiment. The net increase in binding affinity from H08 to H07 is again found to be largely from the van der Waals interactions and the loss of cavity penalty, consistent with the above H06 versus H19 arguments.

Some more complicated cases also show the same behaviour. The experiments reveal that when the R1 group goes from Me group to Et group (H03 to H09, H15 to H16, and H05 to H18) while R2 and R3 groups stay the same, the ligand potency increases by about 1.7 kcal/mol (1.76, 1.84, 1.67 kcal/mol respectively). The LIE–SGB results show the increase of 1.11, 0.97 and 2.45 kcal/mol, respectively, for H03 to H09, H15 to H16, and H05 to H18. Even though the exact numbers are not reproduced, the predicted trends are all correct. Similarly, when further changing R1 group from Et group to i-Pr group in H09 to H10, and H16 to H17, the experimental ligand potency increases by about 0.28 and 0.97 kcal/mol, respectively. This is also confirmed in our LIE–SGB predictions, with 0.13 and 0.54 kcal/mol increase, respectively. Again, even though the potency increases are slightly smaller, LIE–SGB predicts both trends correctly. Thus, both experiment and simulation show that the binding affinities increase when R1 group goes from Me to Et, and Et to i-Pr. Similar to the above two simpler cases, the increase in van der Waals interactions between ligand and protein and the increase in net cavity energy loss due to the burial of solvent accessible surface area have contributed to the increase of the binding affinity, when R1 group changes from smaller Me to larger Et and even larger i-Pr.

Interestingly, H02 is an exception to the above 'pattern rules'. From H03 to H02, and to H05, the R2 group size increases from $CH_2OCH_2CH_3$, to $CH_2OCH_2CH_2CH_3$, and then to CH_2OCH_2Ph. The experimental binding affinity changes from −9.20 to −7.73 kcal/mol and

to − 10.01 kcal/mol, while the LIE–SGB prediction increases monotonically from − 8.54 to − 9.11 kcal/mol and to − 9.57 kcal/mol. From examining the geometrical packing, van der Waals interactions, and loss of surface areas, it seems that H02 should have a higher binding affinity. We do not know exactly why H02 is so different from the others (no steric clashes were observed for H02 during MD simulations, and no steric problem for the even larger H05 either). It is possible that our LIE–SGB model might be unable to capture some features in the H02 binding, but we also think some additional experimental investigations of H02 might be useful (more discussions on this regard later).

Another interesting pair is H01 versus H14. H01 and H14 only differ in group R3, from SPh to OPh. The experimental results show that there is a 1.54 kcal/mol difference, with H01 (SPh) more potent than H14 (OPh). Our LIE–SGB results show a 1.84 kcal/mol difference, which agrees with experiment quite well. For this pair, the ligand sizes are comparable, and the difference is found to be mainly from electrostatic interactions, 1.1 kcal/mol (the ChelpG charges near S and O atoms are quite different), and the burial of solvent accessible surface area, 0.7 kcal/mol (S atom has a larger radius than O atom in SGB continuum solvent model). The difference due to van der Waals interactions is only about 0.07 kcal/mol, which is relatively small. This is different from the previous result using explicit solvent model, [15] where the difference in H01 and H14's binding affinity is found to be 0.5 kcal/mol and is mainly from van der Waals interactions. This discrepancy could be due to the fact that the number of hydrogen bonds used in the explicit solvent-based LIE method [15] might not be able to capture the total electrostatic interactions. Nevertheless, for the majority of the ligands in this binding set, the binding affinity trends and binding mechanism from the current continuum solvent-based LIE model are very similar to the previous study using the explicit solvent-based LIE model. [15]

3.6 π-type H-bonds account for the differences in secondary versus tertiary amides

The somewhat surprising binding affinity trend in nevirapine analogues with secondary amide groups or tertiary amide groups is found to be from a favourable NH-aryl π-type hydrogen bond between the secondary amides and the phenyl ring of residue Y188A (Y188 in chain A). Experimental results show that N08 is 2.36 kcal/mol more potent than N06, and N10 is 1.46 kcal/mol more potent than N11. In both cases, the only difference is that one has a secondary amide fragment (N08 and N10) and the other has a tertiary amide fragment (N06 and N11) near R1 group. Figure 7 shows a representative structure for each

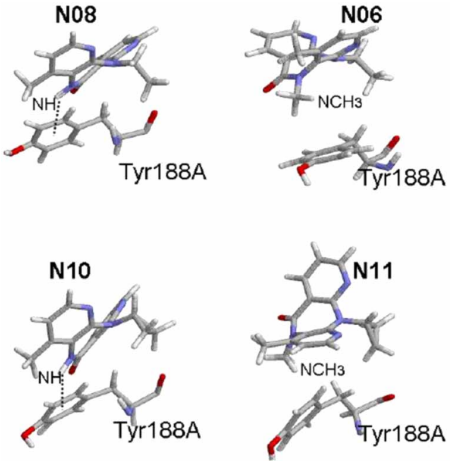

Figure 7. The sketch of representative structures of four nevirapine analogues in the binding pocket with residue Tyr188A of HIV-1RT. It shows the possible π-type hydrogen bonds between the NH group in the secondary amide fragments of N08 and N10 and the phenyl ring of residue Tyr188A in HIV-1RT. However, it is impossible for the tertiary amide groups in N06 and N11 to form π-type hydrogen bonds. (a) N08 versus N06: the structures show a distance of 2.7 Å from H(NH) to the centre of the phenyl ring for N08 and a distance of 4.5 Å from C(NCH_3) to the centre of the phenyl ring for N06; and (b) N10 versus N11. Similarly, the structures show a distance of 2.6 and 5.3 Å, respectively.

ligand and the corresponding Y188A residue nearby from the MD simulation. It is shown that the distance from the H atom in NH group to the centre of the phenyl ring of Y188A is about 2.7 and 2.6 Å for N08 and N10, respectively, which indicates that there might exist a π-type hydrogen bond [56] between the NH group and the phenyl ring; while on the other hand, there is no hydrogen bond possible between NCH3 group and the phenyl ring (the distance from the C atom in NCH3 group to the centre of the phenyl ring is about 4.5 and 5.3 Å, respectively, for N06 and N11). Other configurations from the MD simulations show similar results. In support of this analysis, it is known that the Y188C mutation causes the activity constant IC_{50} of nevirapine (N10) to be 100- to 1000-fold less than the wild type [57]. In principle, this π-type hydrogen bond should be captured in the electrostatic interactions just like the normal hydrogen bonds, [56] but we speculate that these HF/6-31G* ChelpG fitted charges might have underestimated the many-body interactions in this π-type hydrogen bond. This fact, coupled with the above overestimation in the solvation free energy for secondary amides in the free state, might explain the magnitude of the coefficients (− 2.25 kcal/mol) for the secondary amide indicator. It should be noted, that obtaining correct properties such as the relative solvation free energies for amines and amides has been a long-standing problem in the computational community

[50,58–61]. Even for small amines and amides it is non-trivial to get agreement with experiment [50,60].

3.7 Outliers re-examined

There are two outliers as pointed out earlier, with about 2.0 kcal/mol or more errors, H11 (2.33 kcal/mol) and N13 (1.98 kcal/mol). H11, which is MKC-442 (emivirine), is an interesting case. In all the simulations done here, done in our previous study [20], and done with explicit solvent LIE model by Rizzo et al. [15] this MKC-442 always showed a large deviation from the experimental binding affinity, no matter which fitting procedures were used. The LIE predictions are constantly less potent than the experimental affinity by about 2 kcal/mol. As speculated in our previous study [20], we suspected that experimental potency of MKC-442 (0.0042 μM) might be overestimated [43]. Recent measurements by Aly et al. [7] indeed showed a less potent binding for MKC-442, 0.03 ± 0.005 μM versus the earlier 0.0042 μM (i.e. -10.67 kcal/mol vs. -11.89 kcal/mol), which will bring our LIE–SGB prediction within 1 kcal/mol error range (in our current study, we still keep this earlier measurement of MKC-442 in our fitting to be consistent with other ligands measured by the same group;[40–43] if we adopt the new measurement, the overall fitting results will be slightly better).[1] These findings indicate that in some cases, the LIE method is even capable of identifying possible experimental measurement errors.

Analogue N13 is another interesting case. It shows that continuum solvent models might have trouble handling cases where explicit bridging waters might play an important role. Experimentally, N13 is found to have the least favourable affinity in the nevirapine analogue series, only -6.77 kcal/mol, while the LIE–SGB prediction overestimates it by 1.98 to -8.75 kcal/mol, due to the fact that the large t-Bu group in R2 position of N13 has relatively high van der Waals contributions. N13 is the only compound in nevirapine series with a bulkier tertiary butyl group. It is found, from explicit solvent simulations, that the large t-Bu group blocks the possible bridging water hydrogen bonds between nevirapine's pyridine nitrogen and Lys101A residue, which is found in many other ligands such as N01 [15]. This displacement of bridging waters in the bound state contributes a higher penalty in the loss of hydrogen bonds (or the loss of electrostatic interactions here) in explicit solvent model as compared to other ligands with smaller non-tertiary R2 groups. Unfortunately, our continuum solvent model cannot easily catch the bridging water effects, which results in a lower penalty for the N13 ligand and thus a higher binding affinity.

3.8 Design of novel ligands

Finally, we want to design some novel ligands with optimal binding affinity based on the binding mechanism learnt above. Due to the potential issues with the quantum charges for secondary amide groups in nevirapine analogues, we chose the HEPT analogues for this designing effort. Before composing novel compounds, we did more blind tests as further validations using some recently measured high potent compounds that are not included in our earlier training subset [5–7,43,62] Seven new HEPT analogues were chosen, with five, H01v–H05v ('v' stands for validation), having very high affinities (about -11 kcal/mol or higher), and the other two, H06v and H07v, having halide groups (bromine and iodine) [5–7,43] We include these last two ligands because compounds with halide groups are, in general, harder to

Table 6. Newly designed ligands (named 'Hxxd') for optimal binding after further blind test validations (named 'Hxxv').

No.	R1	R2	R3	Derived from	ΔG_{expt}	ΔG_{pred}
H01v	Et	$CH_2OCH_2CH_3$	CH_2Ph-3,5 di-Me		-12.48	-10.94
H02v	i-Pr	$CH_2OCH_2CH_3$	CH_2Ph-3,5 di-Me		-13.08	-11.32
H03v	Et	$CH_2OCH_2CH_2OH$	CH_2Ph-3,5 di-Cl		-11.14	-11.54
H04v	Et	CH_2OCH_2Ph	SPh-3,5 di-Me		-12.05	-12.20
H05v	Et	$CH_2OCH_2CH{=}CH_2$	CH_2Ph-Br		-10.93	-10.01
H06v	Me	$CH_2OCH_2CH_2OH$	SPh-Br		-7.44	-8.43
H07v	Me	$CH_2OCH_2CH_2OH$	SPh-I		-7.09	-8.88
H01d	i-Pr	CH_2OCH_2Ph	SPh-3,5 di-Me	H04v,R1		-12.42
H02d	i-Pr	$CH_2OCH(CH_3)_2$	SPh-3,5 di-Me	H17,R2		-9.67
H03d	i-Pr	CH_2OCH_2Ph	CH_2Ph-3,5 di-Me	H06,R3		-12.34
H04d	i-Pr	$CH_2OCH(CH_3)_2$	CH_2Ph-3,5 di-Me	H02v,R2		-10.92
H05d	i-Pr	CH_2OCH_2Ph	CH_2Ph-3,5 di-Cl	H03v,R1&R2		-11.12
H06d	i-Pr	CH_2OCH_2Ph	SPh-3,5 di-Cl	H06,R3		-12.56

The seven blind test candiates are chosen from recent experiments with mostly high-potent binders that are not originally in the training set (five high-potent binders plus two compounds with halide groups, see text for more details). The R1, R2 and R3 groups are defined in the same way as in Figure 1. We design the new ligands by optimising one or two groups from the exisiting high binders (deriving parent compounds) using the knowledge learned from LIE–SGB analysis. The experimental binding affinities are again calculated from EC_{50} values [6,7,43] ($\Delta G_{bind} \approx RT \ln(EC_{50})$ with unit in kcal/mol).

predict and also because some recent experiments show that adding bromine (bromo derivatives) might help binding [7]. Table 6 lists the details of the compound structures with R1, R2, R3 groups, as well as the experimental binding affinities.

For the five high-potent HEPT analogues (H01v–H05v) with binding affinity about −11 kcal/mol or higher, our LIE–SGB predictions work very well, with all of them predicted to be high-potent compounds (binding affinity all ≤ -10 kcal/mol), as shown in Table 6. The two ligands with halide groups also show fairly good agreement with experiment, even though their binding affinities are not as high as other ligands. The overall RMSE for the seven ligands is only 1.23 kcal/mol. Encouraged by these results, we then designed six more novel compounds in the following way (denoted as H01d–H06d, 'd' stands for design). As mentioned above, the most potent HEPT analogues share some common features in the R1, R2 and R3 groups, for example, the best R1 group from the current data is i-Pr. We then used a combinatorial approach to assemble optimal combinations based on these strong performing groups. We then ran simulations on these ligands and used our model to predict their binding affinity without knowledge of their actual experimental data. The results are summarised in Table 6. As can be seen, three ligands, H01d, H03d and H06d, improved their binding affinity from their deriving parent compounds. One interesting ligand that we would like to point out is H01d. This ligand differs from H04v only in the R1 group, from Et (H04v) to i-Pr (H01d). The binding affinity difference between these two ligands is found to be 0.22 kcal/mol based on our LIE–SGB calculations. This reinforces our observations made earlier for the R1 group, that from Me to Et, and from Et to i-Pr, the binding affinity increases. Making the R3 group from SPh to Sph-3,5 di-Cl also improves the binding affinity by about 0.40 kcal/mol (H06d vs.H06), due to slightly more favourable van der Waals and electrostatic interactions with halide groups. On the other hand, three ligands, H02d, H04d and H05d, showed no improvement in their binding affinity from their deriving parents, with the most noticeable 'failure' in the modification of the R2 group of H17 (H02d). It is also found that the R2 group is most sensitive to the combination of size and functional fragments capable of making H-bonds, as shown by the lower affinity of $-CH_2OCH(CH_3)_2$ as compared to either $-CH_2OCH_2Ph$ (larger size) or $-CH_2OCH_2CH_2OH$ (smaller size but capable of H-bonds). It should be noted that it is non-trivial to design ligands with extremely high potentcy for HIV-1RT, as shown in the recent experiments by Aly et al. [7]. Most of the bromo derivatives the authors designed have lower binding affinity than MKC-442 (H11), and the one tested here H05v is actually their highest binder ($EC_{50} = 0.02 \pm 0.005\,\mu M$, or -10.93 kcal/mol) [7]. This indicates that our results based on LIE–SGB predictions, with about half showing some improvements over their deriving parent compounds, are indeed encouraging. Of course, further experimental validations are needed. Nevertheless, our current approach might help shed some light on the future lead optimisation for HIV-1RT non-nucleoside inhibitors.

4. Conclusion

In this work, we have used a recently developed LIE method based on LIE–SGB, for binding affinity predictions of nevirapine and HEPT analogues binding to HIV-1RT. The method uses three traditional energy terms, van der Waals energy, electrostatic energy and cavity energy, plus a secondary-amide indicator (for nevirapine analogues only) to predict binding affinities. The secondary amide indicator is found to be necessary for nevirapine analogues to account for deficiencies in quantum partial charges. It is later shown that the indicator represents a critical π-type hydrogen bond between the secondary amide fragment of nevirapine analogues and the phenyl ring of Tyr188A residue. The improved model predicts the binding affinities of a 40-ligand training subset with an RMSE of 0.89 kcal/mol (average unsigned error of 0.71 kcal/mol) and a correlation coefficient r^2 of 0.74. The small RMSE and high correlation coefficient indicate a very good agreement with experiment for this comparatively large binding set. The leave-one-out cross validation results also show a 1.00 kcal/mol RMSE with a correlation coefficient of 0.69, which shows a good predictive power. The additional blind tests on seven mostly high-potent candidates, not originally in the training subset, also show reasonable accuracies. More quantum charge fittings with protein environment included are also performed, and similar results are obtained with the leave-one-out cross validation showing an RMSE of 1.02 kcal/mol and a correlation coefficient of 0.67, indicating an out-of-plane polarisability might be needed to fully capture the π-type hydrogen bonds in classical force fields.

The binding affinities of these HEPT and nevirapine analogues are found to be mainly driven by van der Waals interactions between ligands and the HIV-1RT receptor (i.e. a good geometric fit is important), and the net loss of the cavity energy penalty (i.e. the burial of solvent accessible surface area is favourable). The net electrostatic interactions, however, are found to be anti-binding, i.e. the ligands interact electrostatically more favourably with pure water than with the protein HIV-1RT. In addition, for nevirapine analogues, a π-type hydrogen bond between secondary amide fragment of nevirapine analogues and the phenyl ring of Tyr188A of HIV-1RT is found to be critical for their otherwise surprising binding affinities.

These findings are consistent with the previous results from explicit solvent-based LIE methods. Finally, six novel ligands are designed for optimal binding based on our predictions for further experimental validations.

Acknowledgements

We would like to thank William Jorgensen for sending us part of the initial setups of the binding system from their earlier study and for helpful discussions. We also thank Tiziana Jonas-Mordasini for quantum partial charge fittings with the protein environment included. We thank Ajay Royyuru, Robert Rizzo, Rich Friesner and David Silverman for useful discussions. We would also like to thank Frank Suits for help with visualisations.

Note

1. This less potent MKC-442 seems to have been acknowledged by the company conducting the clinical trials as well (press statement from Triangle Pharmaceuticals Inc., http://www.tripharm.com).

References

[1] B.G. Turner and M.F. Summers, *Structural biology of HIV*, J. Mol. Biol. 285 (1999), pp. 1–32.

[2] J. Cohen, *AIDS therapies: the daunting challenge of keeping HIV suppressed*, Science 277 (1997), pp. 32–33.

[3] E.K. Wilson, *AIDS conference highlights hope of drug cocktails, chemokine research*, Chem. Eng. News 74 (1996), pp. 42–46.

[4] R.A. Katz and A.M. Skalka, *The retroviral enzymes*, Annu. Rev. Biochem. 63 (1994), pp. 133–173.

[5] A.L. Hopkins, J. Ren, H. Tanaka, M. Baba, M. Okamato, D.I. Stuart, and D.K. Stammers, *Design of MKC-442 (emivirine) analogues with improved activity against drug-resistant HIV mutants*, J. Med. Chem. 42 (1999), pp. 4500–4505.

[6] M. Wamberg, E.B. Pedersen, N.R. El-Brollosy, and C. Nielsen, *Synthesis of 6-arylvinyl analogues of the HIV drugs SJ-3366 and emivirine*, Bioorg. Med. Chem. 12 (2004), pp. 1141–1149.

[7] Y.L. Aly, E.B. Pedersen, P. La Colla, and R. Loddo, *Synthesis and anti-HIV-1 activity of new MKC-442 analogues with an alkynyl-substituted 6-benzyl group*, Arch. Pharm. (Weinheim) 340 (2007), pp. 225–235.

[8] G.M. Szczech, P. Furman, G.R. Painter, D.W. Barry, K. Borroto-Esoda, T.B. Grizzle, M.R. Blum, J. Sommadossi, R. Endoh, T. Niwa, et al., *Safety assessment,* in vitro *and* in vivo*, and pharmacokinetics of emivirine, a potent and selective nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1*, Antimicrob. Agents Chemother. 44 (2000), pp. 123–130.

[9] J. Aqvist, C. Medina, and J.E. Samuelsson, *A new method for predicting binding affinity in computer-aided drug design*, Protein Eng. 7 (1994), pp. 385–391.

[10] J. Aqvist and T. Hansson, *On the validity of electrostatic linear response in polar solvents*, J. Phys. Chem. 100 (1996), pp. 9512–9521.

[11] J. Aqvist, V.B. Luzhkov, and B.O. Brandsdal, *Ligand binding affinities from MD simulations*, Acc. Chem. Res. 35 (2002), pp. 358–365.

[12] V.B. Luzhkov, J. Nilsson, P. Arhem, and J. Aqvist, *Computational modelling of the open-state kV 1.5 ion channel block by bupivacaine*, Biochim. Biophys. Acta 1652 (2003), pp. 35–51.

[13] H.A. Carlson and W.L. Jorgensen, *An extended linear response method for determining free energies of hydration*, J. Phys. Chem. 99 (1995), pp. 10667–10673.

[14] D.K. Jones-Hertzog and W.L. Jorgensen, *Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method*, J. Med. Chem. 40 (1997), pp. 1539–1549.

[15] R.C. Rizzo, J. Tirado-Rives, and W.L. Jorgensen, *Estimation of binding affinities for HEPT and nevirapine analogues with HIV-1 reverse transcriptase via Monte Carlo simulations*, J. Med. Chem. 44 (2001), pp. 145–154.

[16] A.C. Pierce and W.L. Jorgensen, *Estimation of binding affinities for selective thrombin inhibitors via Monte Carlo simulations*, J. Med. Chem. 44 (2001), pp. 1043–1050.

[17] M.B. Smith Kroeger, M.L. Lamb, J. Tirado-Rives, W.L. Jorgensen, C.L. Michejda, and R.H. Smith, *Monte Carlo calculations on HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor 8-Cl TIBO: contribution of the L100I and Y181C variants to protein stability and biological activity*, Protein Eng. 13 (2000), pp. 413–421.

[18] I.D. Wall, A.R. Leach, D.W. Salt, M.G. Ford, and J.W. Essex, *Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method*, J. Med. Chem. 42 (1999), pp. 5142–5152.

[19] W. Wang, J. Wang, and P.A. Kollman, *What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations?*, Proteins 34 (1999), pp. 395–402.

[20] R. Zhou, R.A. Friesner, A. Ghosh, R.C. Rizzo, W.L. Jorgensen, and R.M. Levy, *New linear interaction method for binding affinity calculations using a continuum solvent model*, J. Phys. Chem. B 105 (2001), pp. 10388–10397.

[21] A.M. Asi, N.A. Rahman, and A.F. Merican, *Application of the linear interaction energy method (LIE) to estimate the binding free energy values of* Escherichia coli *wild-type and mutant arginine repressor c-terminal domain (argrc)-L-arginine and argrc-L-citrulline protein-ligand complexes*, J. Mol. Graph. Model. 22 (2004), pp. 249–262.

[22] U. Bren, V. Martinek, and J. Florian, *Free energy simulations of uncatalyzed DNA replication fidelity: structure and stability of T.G and dTTP.G terminal DNA mismatches flanked by a single dangling nucleotide*, J. Phys. Chem. B 110 (2006), pp. 10557–10567.

[23] Z. Zhou, M. Bates, and J.D. Madura, *Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design*, Proteins 65 (2006), pp. 580–592.

[24] N. Foloppe and R. Hubbard, *Towards predictive ligand design with free-energy based computational methods?* Curr. Med. Chem. 13 (2006), pp. 3583–3608.

[25] M.K. Gilson and H.X. Zhou, *Calculation of protein-ligand binding affinities*, Annu. Rev. Biophys. Biomol. Struct. 36 (2007), pp. 21–42.

[26] N.J. English, *Calculation of binding affinities of HIV-1 RT and beta-secretase inhibitors using the linear interaction energy method with explicit and continuum solvation approaches*, J. Mol. Model. 13 (2007), pp. 1081–1097.

[27] C. Oostenbrink, *Efficient free energy calculations on small molecule host–guest systems – a combined linear interaction energy/one-step perturbation approach*, J. Comput. Chem. 30 (2009), pp. 212–221.

[28] F.S. Lee, Z.T. Chu, M.B. Bolger, and A. Warshel, *Calculations of antibody–antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603*, Prot. Eng. 5 (1992), pp. 215–228.

[29] Y. Sham, Z. Chu, H. Tao, and A. Warshel, *Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease*, Proteins Strut. Funct. Genet. 39 (2000), pp. 393–407.

[30] F.S. Lee, Z.T. Chu, and A. Warshel, *Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs*, J. Comp. Chem. 14 (1993), pp. 161–185.

[31] A. Warshel, H. Tao, M. Fothergill, and Z.T. Chu, *Effective methods for estimation of binding energies in computer-aided drug design*, Israel J. Chem. 34 (1994), pp. 253–256.

[32] I. Muegge, H. Tao, and A. Warshel, *A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding*, Prot. Eng. 10 (1997), pp. 1363–1372.
[33] A. Curioni, T. Mordasini, and W. Andreoni, *Enhancing the accuracy of virtual screening: molecular dynamics with quantum-refined force fields*, J. Comput. Aid. Mol. Des. 18 (2004), pp. 773–784.
[34] D.B. Kitchen, F. Hirata, J.D. Westbrook, R.M. Levy, D. Kofke, and M. Yarmush, *Conserving energy during molecular dynamics simulations of water, proteins and proteins in water*, J. Comp. Chem. 11 (1990), pp. 1169–1180.
[35] F. Figueirido, R. Zhou, R. Levy, and B.J. Berne, *Large scale simulation of macromolecules in solution: combining the periodic fast multipole method with multiple time step integrators*, J. Chem. Phys. 106 (1997), pp. 9835–9849.
[36] A. Ghosh, C.S. Rapp, and R.A. Friesner, *Generalized Born model based on a surface integral formulation*, J. Phys. Chem. B 102 (1998), pp. 10983–10990.
[37] W.L. Jorgensen, D. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom froce field on conformational energetics and properties of organic liquids*, J. Am. Chem. Soc. 118 (1996), pp. 11225–11236.
[38] W. Clark Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson, *Semianalytical treatment of solvation for molecular mechanics and dynamics*, J. Am. Chem. Soc. 112 (1990), pp. 6127–6129.
[39] M. Friedrichs, R. Zhou, S. Edinger, and R.A. Friesner, *Poisson–Boltzman analytical gradients for molecular modeling calculations*, J. Phys. Chem. 103 (1999), pp. 3057–3061.
[40] H. Tanaka, M. Baba, H. Hayakawa, T. Sakamaki, T. Miyasaka, M. Ubasawa, H. Takashima, K. Sekiya, I. Nitta, S. Shigeta, et al., *A new class of HIV-1-specific 6-substituted acyclouridine derivatives: synthesis and anti-HIV-1 activity of 5- or 6-substituted analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (hept)*, J. Med. Chem. 34 (1991), pp. 349–357.
[41] H. Tanaka, H. Takashima, M. Ubasawa, K. Sekiya, I. Nitta, M. Baba, S. Shigeta, R.T. Walker, E. De Clercq, and T. Miyasaka, *Structure-activity relationships of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine analogues: effect of substitutions at the c-6 phenyl ring and at the c-5 position on anti-HIV-1 activity*, J. Med. Chem. 35 (1992), pp. 337–345.
[42] H. Tanaka, H. Takashima, M. Ubasawa, K. Sekiya, I. Nitta, M. Baba, S. Shigeta, R.T. Walker, E. De Clercq, and T. Miyasaka, *Synthesis and antiviral activity of deoxy analogs of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (hept) as potent and selective anti-HIV-1 agents*, J. Med. Chem. 35 (1992), pp. 4713–4719.
[43] H. Tanaka, H. Takashima, M. Ubasawa, K. Sekiya, N. Inouye, M. Baba, S. Shigeta, R.T. Walker, E. De Clercq, and T. Miyasaka, *Synthesis and antiviral activity of 6-benzyl analogs of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (hept) as potent and selective anti-HIV-1 agents*, J. Med. Chem. 38 (1995), pp. 2860–2865.
[44] K.D. Hargrave, J.R. Proudfoot, K.G. Grozinger, E. Cullen, S.R. Kapadia, U.R. Patel, V.U. Fuchs, S.C. Mauldin, et al., *Novel nonnucleoside inhibitors of HIV-1 reverse transcriptase. 1. Tricyclic pyridobenzodiazepinones and dipyridodiazepinones*, J. Med. Chem. 34 (1991), pp. 2231–2241.
[45] M.J. Frisch, et al., *Gaussian 95, Development Version (Revision E.1)*, Gaussian Inc., Pittsburgh, PA, 1996.
[46] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B 195 (1987), pp. 216–220.
[47] R. Zhou and B.J. Berne, *Smart walking: a new method for Boltzmann sampling of protein conformations*, J. Chem. Phys. 107 (1997), pp. 9185–9196.
[48] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684–3690.
[49] H.C. Andersen, *Rattle-A velocity version of the SHAKE algorithm for Molecular-Dynamics calsulations*, J. Comp. Phys. 52 (1983), pp. 24–34.
[50] R.C. Rizzo and W.L. Jorgensen, *OPLS all-atom model for amines: resolution of the amine hydration problem*, J. Am. Chem. Soc. 121 (1999), pp. 4827–4836.
[51] R. Wolfenden, *Interaction of the peptide bond with solvent water: a vapor phase analysis*, Biochemistry 17 (1978), pp. 201–204.
[52] E.M. Duffy and W.L. Jorgensen, *Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water*, J. Am. Chem. Soc. 122 (2000), pp. 2878–2888.
[53] J. Banks, G. Kirminski, R. Zhou, B.J. Berne, and R.A. Friesner, *Parametrizing a polarizable force field from* ab initio *data. I. The fluctuating point charge model*, J. Chem. Phys. 110 (1999), pp. 741–754.
[54] H. Stern, G. Kirminski, J. Banks, R. Zhou, B.J. Berne, and R.A. Friesner, *Fluctuating charge, polarizable dipole, and combined models: parametrization from* ab initio *quantum chemistry*, J. Phys. Chem. B 103 (1999), pp. 4730–4737.
[55] M.L. Connolly, *Analytical molecular surface calculation*, J. Appl. Cryst. 16 (1983), pp. 548–558.
[56] W.L. Jorgensen and D.L. Severance, *Aromatic aromatic interactions: free energy profiles for the benzene dimer in water, chloroform and liquid benzene*, J. Am. Chem. Soc. 112 (1990), pp. 4768–4774.
[57] R.W. Buckheit, et al., *Resistance to 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives is generated by mutations at multiple sites in the HIV-1 reverse-transcriptase*, Virology 210 (1995), pp. 186–193.
[58] P.Y. Morgantini and P.A. Kollman, *Solvation free energies of amides and amines: disagreement between free energy calculations and experiment*, J. Am. Chem. Soc. 117 (1995), pp. 6057–6063.
[59] Y.B. Ding, D.N. Bernardo, K. Kroghjespersen, and R.M. Levy, *Solvation free energies of small amides and amines from molecular dynamics free energy perturbation simulations using pairwise additive and many-body polarizable potentials*, J. Phys. Chem. 99 (1995), pp. 11575–11583.
[60] E.C. Meng, J.W. Caldwell, and P.A. Kollman, *Investigating the anolalous solvation free energies of amines with a polarizable potential*, J. Phys. Chem. 100 (1996), pp. 2367–2371.
[61] B. Marten, K. Kim, C. Cortis, R.A. Friesner, R.B. Murphy, M.N. Ringnalda, D. Sitkoff, and B. Honig, *New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects*, J. Phys. Chem. 100 (1996), pp. 11775–11788.
[62] L. Petersen, C.H. Jessen, E.B. Pedersen, and C. Nielsen, *Synthesis and evaluation of new potential HIV-1 non-nucleoside reverse transcriptase inhibitors. New analogues of MKC-442 containing michael acceptors in the c-6 position*, Org. Biomol. Chem. 1 (2003), pp. 3541–3545.