Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2005) 147–156
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

# Scoring binding affinity of multiple ligands using implicit solvent and a single molecular dynamics trajectory: Application to *Influenza* neuraminidase

Pascal Bonnet [1], Richard A. Bryce *

*School of Pharmacy and Pharmaceutical Sciences, University of Manchester, Manchester M13 9PL, UK*

Accepted 13 June 2005
Available online 10 August 2005

**Abstract**

We explore a perturbative approach to calculation of binding free energy of multiple ligands, based on a single molecular dynamics simulation of a reference ligand–receptor complex and analysis via a hybrid force field/continuum model potential. The methodology is applied to prediction of relative binding free energies of 10 *Influenza* neuraminidase inhibitors, using Poisson–Boltzmann and generalised Born models of implicit solvent. These single-step MM-PB/SA and MM-GB/SA approaches predict the experimentally most potent ligand as first- or second-ranked according to total binding free energy. Ranking of inhibitors displays only moderate sensitivity to the choice of reference trajectory and ligand partial charge scheme. When ranked according to total electrostatic binding free energy, correlation with experiment improves ($r^2$ of 0.72); this may be related to underestimated first solvation shell effects by the implicit water models. Therefore, to increase the generality of this single-step approach as part of a potential computational compound optimisation strategy, further development of the treatment of short-range solvent interactions is warranted.


*Keywords:* Computer-aided ligand design; MM-PB/SA; MM-GB/SA; Binding free energies; Molecular dynamics

## 1. Introduction

A predictive model of molecular association is recognised as a key element in rational drug design and development strategies. However, a tractable first-principles approach to this problem has remained elusive. Although application of statistical perturbation theory to calculation of relative ligand–receptor binding affinity has achieved in favourable cases predictions to within one kilocalorie of experiment [1,2], the approach necessitates evaluation of closely related ligand congeners connected by numerous interpolating simulations. Consequently, despite its rigour, this computationally intensive approach does not readily synchronise with the tight design/synthesis cycle demanded in a lead discovery and optimisation environment.

As a result, considerable effort has focused on more expedient routes to first-principles evaluation of binding free energies [3,4]. Based upon analysis of just two molecular simulations, of the ligand receptor-bound and free in aqueous solution, the linear interaction energy (LIE) approach [4] connects binding free energy to the ensemble-averaged electrostatic and van der Waals ligand receptor interaction energies. Another expedient approach to theoretical calculation of binding affinity, in its simplest form, evaluates binding free energy from a single molecular dynamics (MD) simulation of the ligand–protein complex in solution [3,5]. An ensemble of configurations generated by an explicit solvent simulation of the complex are subsequently analysed with a hybrid potential, combining molecular mechanics and a dielectric continuum solvent

* Corresponding author. Tel.: +44 161 275 8345; fax: +44 161 275 2481.
*E-mail address:* R.A.Bryce@manchester.ac.uk (R.A. Bryce).
[1] Present address: Johnson & Johnson, Pharmaceutical Research & Development, Division of Janssen-Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium.

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.06.003

approach, typically the Poisson–Boltzmann (PB) or generalized Born (GB) models [6]. These MM-PB/SA and MM-GB/SA methods have proved effective tools in analysis of nucleic acid structure [6], protein folding [7], RNA–protein interactions [8], carbohydrate-lectin [9] and other ligand–protein interactions [10,11].

These studies involve analysis of individual trajectories generated for each distinct molecular complex investigated. However, it is possible to introduce further efficiency into the method, calculating the energetic effect of sub-structural substitutions into the MD trajectory of a single reference system. For example, a MM-PB/SA ''alanine-scanning'' approach was successfully adopted to explore oncoprotein Mdm2-tumour suppressor protein p53 interactions [12] where key amino acid side-chains were replaced with methyl groups for snapshots from the trajectory of the protein–protein complex. For each perturbed trajectory, the binding free energy of the amino acid substitution was evaluated via the hybrid molecular mechanics/continuum model potential, and good agreement with mutagenesis experiments was obtained. Conversely, this approach can be applied to perturbation of the ligand structure, and has been applied to analysis of ligand interactions with (strept)avidin [3,5], HIV protease [13] and matrix metalloproteinase-1 (MMP-1) [14]. These studies indicate that the perturbative MM-PB/SA approach has potential to rank the binding affinity of a series of ligands from a single molecular dynamics trajectory.

As part of a detailed molecular dynamics study [15] of *Influenza* neuraminidase, a viral glycoprotein, we have applied a variant of this single-step perturbative MM-PB/SA approach to predict binding free energies for four inhibitors of this enzyme, based on a single reference ligand–protein MD simulation. The approach we employed differed from earlier work [3,5,13,14] in that a rotameric search for the substituted fragment was introduced for each MD snapshot, coupled to energy minimization around the ligand that includes solvent effects; conformers were then selected according to the total MM-PB/SA or MM-GB/SA energy. The 4 inhibitors considered in the study, based on the molecular framework of 2-deoxy-2,3-didehydro-*N*-acetyl-neuraminic acid (DANA), vary in substituent at C4 (Fig. 1).

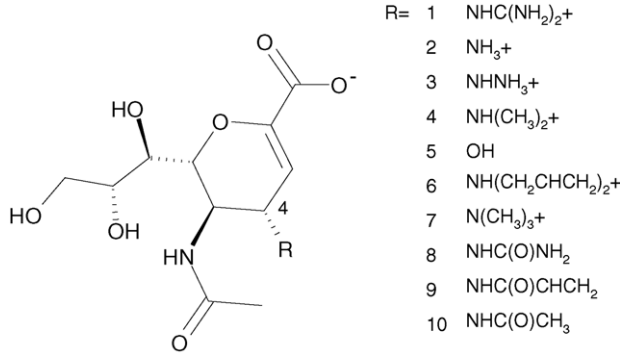


Fig. 1. Influenza neuraminidase inhibitors.

The common framework and binding mode exhibited by this series of ligands, in addition to the availability of a body of good quality structural and energetic data from experiment, make neuraminidase an appealing test case for the single-step methodology. Interestingly, the larger substituents at C4 displace bound water molecules in an anionic pocket (Fig. 2). An anticipated advantage of a hybrid MM/continuum model potential is obviation of the need to explicitly consider displaced structural water in the substitution of smaller functional groups with larger functional groups.

The aim of this study, therefore, is to assess a single-step perturbative approach for ranking multiple inhibitors as a tool for ligand design. We apply single-step MM-PB/SA and MM-GB/SA approaches to order a set of 10 DANA-based *Influenza* neuraminidase inhibitors, **1**–**10** (Fig. 1), based on a single MD simulation. To examine dependence of the approach on choice of reference trajectory, we compare energetic analyses based on 2 ns MD simulations of three different inhibitor–protein complexes. In addition to solvent model and reference trajectory, we also explore the effect of ligand electrostatic model, using both RESP-derived [16] and Gasteiger–Marsili [17] partial atomic charge sets.

## 2. Computational methods

### 2.1. Molecular dynamics simulations

Atomic coordinates of subtype N9 (A/TERN/AUSTRALIA/G70C/75) *Influenza* neuraminidase inhibitor complexes with inhibitors **5** and **1** (Fig. 1) were taken from the Protein Data Bank crystal structures (respective PDB codes 1F8B [18] and 1NNC [19], both of resolution 1.8 Å). Non-electrostatic parameters were modelled from the parm99 force field [20] for use with the AMBER molecular dynamics package [21]. Inhibitor charges were derived via multi-conformational RESP [16] fitting using the HF/6-31G* electrostatic potential, calculated using the Gaussian 98 package [22]. These RESP-derived charges were employed for molecular dynamics of neuraminidase complexes of **1** and **5**. All waters of hydration within 3 Å of the active site were preserved. The enzyme–inhibitor complexes were immersed in a box of TIP3P water model [23] (~5900 atoms). Periodic boundary conditions were imposed in conjunction with the particle-mesh Ewald (PME) method [24] for long-range electrostatic interactions. A cutoff of 8 Å was used for Lennard–Jones interactions and SHAKE [25] to constrain covalent bonds involving hydrogen. Molecular dynamics simulations were carried out at 300 K with a time step of 2 fs. NVT/NPT equilibration of 110 ps was followed by 2 ns NVT production trajectories. Details of the MD simulation of the neuraminidase/inhibitor **2** complex are given elsewhere [15].

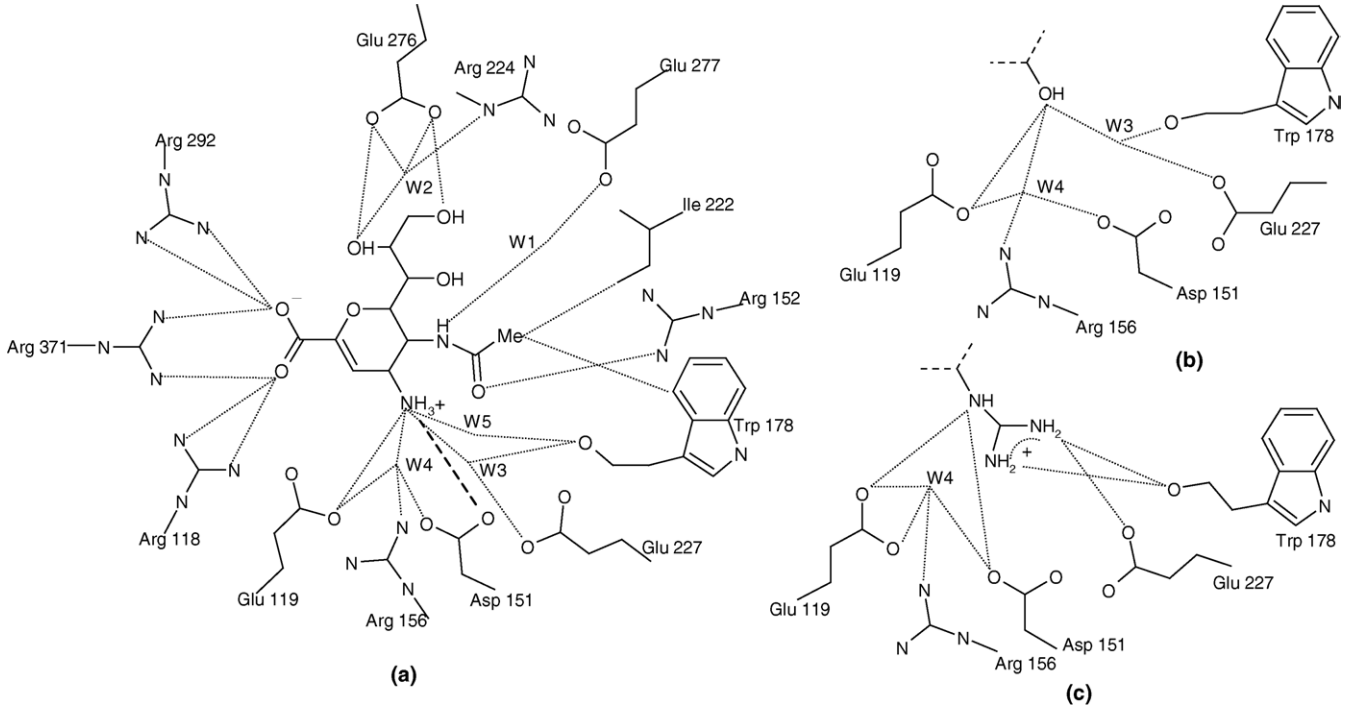


Fig. 2. (a) Direct and water-mediated crystallographic interactions between neuraminidase and **2**, (b) crystallographic interactions at 4-position of **5** with neuraminidase, (c) crystallographic interactions at 4-position of **1** with neuraminidase.

### 2.2. Binding free energy calculations

Energetic post-processing of the trajectories was performed via molecular mechanics and a continuum solvent model, calculated using the *mm_pbsa* module of AMBER6 [21]. This method calculates a gas-phase contribution to binding using an all-atom force field and incorporates the influence of solvent via the PB [26] or GB solvent models [27,28], leading to the MM-PB/SA and MM-GB/SA approaches, respectively. A free energy difference between bound and free protein receptor and ligand is averaged over the trajectories, using the equation:

$$\Delta G_{\text{bind}} = (\Delta H_{\text{bind}} - T\Delta S_{\text{bind}})_{\text{gas}} + \Delta\Delta G_{\text{solv}} \quad (1)$$

where $\Delta H_{\text{bind}}$ and $\Delta S_{\text{bind}}$ is the enthalpy and entropy of receptor–ligand binding, respectively, without solvent contributions and $T$ is temperature. $\Delta H_{\text{bind}}$ is composed of ensemble-averaged electrostatic and van der Waals interaction energies, $\Delta H_{\text{elec}}$ and $\Delta H_{\text{vdw}}$. Vibrational solute entropies were calculated from summation over normal modes using AMBER6 [21]; for computational tractability, residues further than 6 Å from the active site were excluded; the resulting subsystem was minimized using a distance-dependent dielectric constant of 4r and entropic contributions calculated [3,11,29], using the average of three snapshots. $\Delta\Delta G_{\text{solv}}$ is the ensemble-averaged difference in solvation free energy between receptor and ligand, bound and free. The solvation energies of the receptor and ligand are calculated using geometries taken from the structure of the complex. The electrostatic contribution to solvation, $\Delta\Delta G_{\text{pol}}$, is calculated either by: (i) solving the linear Poisson–Boltzmann equation using the DelPhi program [26] with radii from the Cornell et al. force field [30], a 0.5 Å grid extending 20% beyond the solute, a solute dielectric of unity and a solvent dielectric of 80 or (ii) generalized Born calculations, employing a parametrization consistent for use with AMBER [27]. Non-polar contributions to solvation ($\Delta\Delta G_{\text{np}}$) were determined from solvent-accessible surface areas [6], computed using MSMS [31]. Analysis was performed using 20 equispaced snapshots drawn from the last nanosecond of the trajectories.

### 2.3. Perturbative binding free energy calculations

To enable perturbation of inhibitor chemical groups within the thermodynamic analysis approach, an algorithm was implemented within the *mm_pbsa* module of AMBER6 [21] to substitute a new fragment functional group on to an existing ligand molecular framework via rigid body superposition [32]. For each of 20 trajectory snapshots sampled, a search of possible fragment rotamers was performed; conformational sampling of amino acid sidechains was represented by MD configurations of the protein. We initially explored energetic post-processing without energy minimization of perturbed ligand or surrounding protein. Although perturbations that reduce substituent size were feasible, such as for ligands **1** → **2** and **1** → **5**, the converse was sterically forbidden. Therefore, a more general approach required relaxation of the perturbed ligand and

its environment: a range of minimization protocols was explored, using a distance-dependent dielectric (DDD) [33] of 4r to model the effect of solvent on structure. The most suitable minimization procedure involved full minimization of the ligand and amino acid residues that fall within an inclusive 8 Å radius around it. Schemes involving ligand-only relaxation, or optimisation of the most immediate environment around the 4-substituent did not permit sufficient structural relaxation in response to ligand perturbation. Thus, the overall conformational search protocol consists of a rotameric search, with minimization under a DDD solvent model and subsequent energetic evaluation of each resultant structure using the full MM-PB/SA or MM-GB/SA potential, for selection of the lowest energy conformer.

Two alternative inhibitor charge sets were employed for energetic analysis of **1**–**10**: firstly, via the RESP-based approach, as described for the molecular dynamics simulations above. Here, to obviate small differences in partial charges on atoms of the molecular framework common to the inhibitor set, charges for those atoms on zwitterionic inhibitors were constrained to the set obtained from ligand **1**, and for anionic inhibitors, from ligand **5**. Secondly, the partial orbital electronegativity equalization scheme of Gasteiger and Marsili was employed [17]. This topology-dependent method automatically assigns equivalent charges to atoms of the common molecular framework and thus constitutes an interesting alternative to the RESP-based charges for the perturbative methodology. For each of the 10 ligands, vibrational solute entropies were calculated from summation over normal modes based on the average of 10 snapshots of the given ligand complex using the ligand **2**/neuraminidase reference trajectory.

## 3. Results

### *3.1. Molecular dynamics simulations of neuraminidase complexes and energy analysis*

We first describe MD simulations of neuraminidase in complex with ligands **1** and **5**, which form the basis for subsequent application of the one-step MM-PB(GB)/SA methodology to predict binding affinity of the set of 10 inhibitors. The receptor is *Influenza* neuraminidase, a viral surface protein that cleaves terminal sialic acid residues from cell-surface glycoconjugates. Inhibition of neuraminidase leads to delayed spread of the virus, increasing the effect of the host immune response. Structure-based drug design has led to the development of zanamivir, based on the natural neuraminidase inhibitor **5** (Fig. 1) [19]. Substitution of the 4-hydroxyl of **5** with an amino group (**2** in Fig. 1) leads to improved binding affinity ($K_i$), from 4 μM to 40 nM [34]. Subsequent replacement of the 4-amino group with a guanidino functionality gives the tight binding 1 nM zanamivir inhibitor (**1** in Fig. 1). Crystal structures highlight

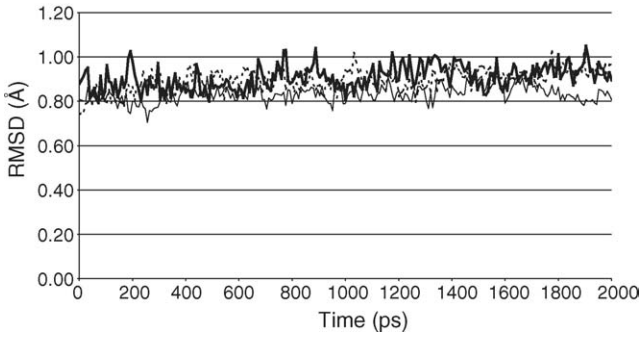


Fig. 3. Protein RMS deviation (Å) from crystal structure during the simulations of the neuraminidase/**5** complex (dotted line), neuraminidase/**2** complex (solid line) and neuraminidase/**1** complex (bold line).

common binding features for ligands **1**, **2** and **5** (Fig. 2): the ligand is tethered to the protein at the carboxylate via a triad of arginines (Arg118, 292 and 371), at the acetamido moiety via particularly Arg152, through hydrophobic interactions with Trp178 and Ile222, and at the glycerol chain by Glu276. Several water-mediated protein–ligand interactions also exist (Fig. 2a). The 4-OH on **5** is adjacent to a pocket containing acidic residues Glu119, Asp151 and Glu227, but makes a polar contact with only Glu119 (Fig. 2b). For ligands **1** and **2**, Asp151 rotates, to make direct and water-mediated interactions with $NH_3^+$ of **2** (Fig. 2a) and directly with the terminal N atoms of the guanidino group of **1** (Fig. 2c). These N atoms of the guanidino group (Fig. 2c) replace the location of two resolved waters of hydration observed in the ligand **2**/neuraminidase structure (W3 and W5 in Fig. 2a).

Over the 2 ns molecular dynamics trajectories for protein complexes of **1** and **5**, a conformation close to the crystal structure is preserved with an average backbone RMSD of 0.8 Å and 0.9 Å, respectively. The RMSD time series for ligands **1** and **5** are comparable with that from our previous study of **2** (Fig. 3) [15]; crystallographic ligand–protein polar contacts are generally well-populated, including at the key C4 substituent position on the sugar ring (see Supplementary information).

Based upon these three MD ensembles for complexes of **1**, **2** and **5**, we proceed to calculation of binding affinity for each of the three inhibitors to neuraminidase according to MM-PB/SA and MM-GB/SA methods (Table 1). Employing

Table 1
Total free energies of binding, $\Delta G_{bind}$, for neuraminidase inhibitors **1**, **2** and **5**, based on MM-PB/SA or MM-GB/SA analysis of individual ligand–protein trajectories (kcal/mol); ligand atoms use RESP-based charges

| Method | Ligand **1** | Ligand **2** | Ligand **5** |
|---|---|---|---|
| MM-PB/SA | −12.1 (0.0) | −5.2 (6.9) | 8.3 (20.4) |
| MM-GB/SA | −37.8 (0.0) | −37.3 (0.5) | −31.8 (6.0) |
| MM-PB/SA[a] | −24.9 (0.0) | −15.1 (9.8) | 0.4 (25.3) |
| MM-GB/SA[a] | −50.7 (0.0) | −47.2 (3.5) | −39.7 (11.0) |
| Experiment [35] | −12.4 (0.0) | −10.2 (2.2) | −7.4 (5.0) |

Standard errors range from 1.0 to 1.5 kcal/mol. Energies relative to ligand **1** in parentheses.
[a] Excludes $-T\Delta S_{bind,gas}$ contribution.

the individual ligand–protein trajectories, the three inhibitors are correctly ranked using the PB solvent model: the total binding free energies, $\Delta G_{bind}$, for **1**, **2** and **5** are −12.1, −5.2 and 8.3 kcal/mol, respectively (Table 1). Employing instead the GB solvent model, the three inhibitors are also correctly ranked, although the predicted affinity of ligands **1** and **2** differ by only 0.5 kcal/mol (Table 1). For both PB- and GB-based approaches, the binding affinity of ligand **5** is understabilised relative to the zwitterionic inhibitors, as observed in other studies [15,18,35]. In this work, ligand **5** is 10.7 and 2.7 kcal/mol underpredicted by total $\Delta G_{bind}$ relative to **2**, using PB and GB solvent models, respectively. We note that this underprediction is also found in the absence of the problematic [9,12] solute $-T\Delta S_{bind,gas}$ contribution, although the three ligands still rank in accord with experiment (Table 1). Insight into the magnitude of protein–ligand interactions can be obtained from analysis of binding enthalpy ($\Delta H_{bind,gas}$, Eq. (1)) according to ligand group contributions (Fig. 4): as expected, the ring carboxylate group is significant for binding of all three ligands, forming salt bridges with the arginine triad. Interestingly, the unfavourable interaction of the 4-hydroxyl substituent of **5** is also apparent, contrasting with strongly favourable enthalpic interactions for the cationic groups at the C4 position of **1** ($NHC(NH_2)_2^+$) and **2** ($NH_3^+$).

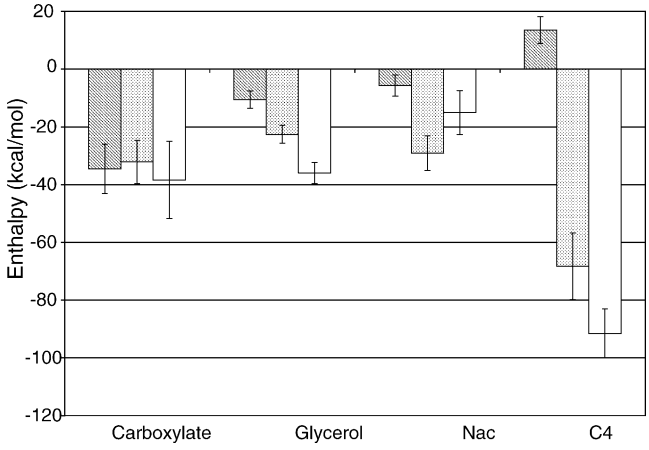


Fig. 4. Contributions of ligand groups (carboxylate, glycerol, *N*-acetyl and C4 substituent groups) to gas-phase binding enthalpy (kcal/mol) for ligands **1** (white), **2** (stippled) and **5** (hatched).

### 3.2. Single-step perturbative calculations of binding affinity

We now apply perturbative binding free energy analysis to calculate relative binding energies for ligands **1**–**10**. We note that the $-T\Delta S_{bind,gas}$ contribution shows considerable variation amongst the 10 ligands (Table 2), most probably representing the largest source of uncertainty in calculation of $\Delta G_{bind}$; accordingly, inclusion of this term in calculated $\Delta G_{bind}$ reduced correlation with experimental affinities. Therefore this contribution is excluded from further analysis, and we subsequently focus on relative free energies of binding. We first consider $\Delta G_{bind}$ calculated using a MM-PB/SA potential, the inhibitor **2**/neuraminidase trajectory as the reference ensemble, and RESP-based or Gasteiger charge sets on the 10 inhibitors; we denote these two variants as Models 1 and 2, respectively (Table 2). Ligand **1** is correctly identified in both cases as the tightest binding inhibitor, with a calculated $\Delta G_{bind}$ of −41.9 kcal/mol for Model 1 and −53.9 kcal/mol for Model 2. For both models, a general ranking of inhibitors with charged C4 substituents (**1**–**4**, **6**, **7**) above those with neutral C4 groups (**5**, **8**–**10**) is obtained. Here, $\Delta G_{bind}$ is qualitatively useful in identifying the most potent inhibitor, with an energy difference to the second-ranked ligand of 3.1 and 7.2 kcal/mol for Models 1 and 2, respectively. However, for overall ranking of the inhibitor set, the square of the coefficient of determination, $r^2$, for these calculations is low (0.38 and 0.42, respectively).

Interestingly, when solely the total electrostatic contribution to binding is considered, $\Delta G_{elec}$ ($=\Delta H_{elec,gas} + \Delta\Delta G_{pol}$), correlation with experiment improves to a $r^2$ of 0.53 and 0.67 for RESP-based or Gasteiger ligand charge sets (Models 1E and 2E, respectively, Table 2). The improved agreement appears to arise from preferential stabilisation of ligands with smaller C4 substituents, such as

Table 2
Relative binding free energies for ligands **1**–**10** from perturbative MM-PB/SA calculations compared to experiment (kcal/mol)

| Ligand | Model 1 | Model 2 | Model 1E | Model 2E | $\Delta G_{solv}$ | $-T\Delta S_{bind,gas}$ | $\Delta G_{expt}$ [34] |
|---|---|---|---|---|---|---|---|
| **1**[a] | 0.0 (1) | 0.0 (1) | 0.0 (3) | 0.0 (2) | 195.5 | 22.3 | 0.0 |
| **2** | 6.8 (5) | 15.8 (6) | −3.9 (1) | 1.7 (3) | 192.7 | 12.8 | 2.2 |
| **3** | 7.4 (6) | 13.7 (5) | −1.9 (2) | −0.2 (1) | 195.5 | 18.1 | 2.8 |
| **4** | 4.7 (4) | 7.2 (2) | 2.0 (4) | 1.8 (4) | 175.8 | 11.6 | 4.8 |
| **5** | 39.1 (10) | 25.4 (10) | 38.8 (9) | 19.2 (7) | 97.6 | 11.2 | 5.0 |
| **6** | 3.1 (2) | 7.6 (3) | 5.5 (6) | 6.7 (5) | 171.8 | 19.6 | 5.4 |
| **7** | 3.7 (3) | 12.5 (4) | 5.3 (5) | 11.5 (6) | 165.1 | 11.3 | 5.5 |
| **8** | 27.6 (7) | 20.1 (7) | 32.4 (7) | 22.0 (8) | 117.2 | 12.0 | 6.6 |
| **9** | 29.3 (8) | 22.4 (9) | 38.5 (8) | 23.2 (9) | 104.7 | 17.3 | 6.7 |
| **10** | 30.1 (9) | 21.8 (8) | 39.8 (10) | 23.4 (10) | 99.3 | 14.1 | 7.3 |

See text for description of models. Rank in parentheses. Standard errors range from 1 to 2 kcal/mol. Computed binding free energies omit solute entropy contribution, $-T\Delta S_{bind,gas}$ (kcal/mol). Absolute solvation energies (kcal/mol) for Model 1 are also shown ($\Delta G_{solv}$).

[a] Absolute energies (kcal/mol) of ligand **1**: −41.9 (Model 1), −53.9 (Model 2), −33.8 (Model 1E), −24.6 (Model 2E).

Table 3
Relative binding free energies for ligands **1–10** from perturbative MM-GB/SA calculations compared to experiment (kcal/mol)

| Ligand | Model 3 | $\Delta G_{solv}$ [a] | Model 4 | $\Delta G_{solv}$ [b] | Model 5 | $\Delta G_{solv}$ [c] | Model 5E | $\Delta G_{expt}$ [34] |
|---|---|---|---|---|---|---|---|---|
| **1** [d] | 0.0 (2) | 173.5 | 0.0 (2) | 178.7 | 0.0 (1) | 170.8 | 0.0 (2) | 0.0 |
| **2** | 3.2 (6) | 167.6 | 5.0 (6) | 168.9 | 3.9 (6) | 165.8 | 0.9 (3) | 2.2 |
| **3** | 3.0 (5) | 162.8 | 3.9 (5) | 169.8 | 3.2 (4) | 159.0 | −0.3 (1) | 2.8 |
| **4** | 2.5 (4) | 151.0 | 3.5 (4) | 160.9 | 3.3 (5) | 151.4 | 7.0 (4) | 4.8 |
| **5** | 6.5 (7) | 50.6 | 6.9 (7) | 66.1 | 7.4 (8) | 76.1 | 9.0 (5) | 5.0 |
| **6** | 1.2 (3) | 142.9 | 0.3 (3) | 157.9 | 2.4 (3) | 140.2 | 11.3 (6) | 5.4 |
| **7** | 15.9 (10) | 156.8 | 17.7 (10) | 160.9 | 14.4 (10) | 148.1 | 21.5 (10) | 5.5 |
| **8** | −0.4 (1) | 52.2 | −0.5 (1) | 76.6 | 1.6 (2) | 85.1 | 12.4 (7) | 6.6 |
| **9** | 8.4 (8) | 54.1 | 9.6 (8) | 73.5 | 6.4 (7) | 85.1 | 15.7 (8) | 6.7 |
| **10** | 11.0 (9) | 52.0 | 10.4 (9) | 75.3 | 8.6 (9) | 90.3 | 17.0 (9) | 7.3 |

See text for description of models. Rank in parentheses. Standard errors range from 1 to 2 kcal/mol. Computed binding free energies omit solute entropy contribution $-T\Delta S_{bind,gas}$ (kcal/mol). Absolute solvation energies (kcal/mol) also shown ($\Delta G_{solv}$).

[a] Model 3.
[b] Model 4.
[c] Model 5.
[d] Absolute energies (kcal/mol) of ligand **1**: −63.8 (Model 3), −64.4 (Model 4), −58.6 (Model 5), −50.5 (Model 5E).

**2** ($NH_3^+$), **3** ($NHNH_3^+$) and **5** (OH). However, based on $\Delta G_{elec}$, ligand **1** is no longer predicted as the most potent inhibitor. It also appears that the Gasteiger approach to estimating ligand electrostatics affords a more balanced estimate of zwitterionic and anionic inhibitor binding affinities, with a narrower range of computed affinities than via RESP-based charges (Model 1). This may be due in part to the uniformity of the Gasteiger charge-partitioning scheme, which allocates identical charges on common atoms across the full range of DANA-based inhibitors. Thus, for example, the amide nitrogen of all ligands are −0.27e via the Gasteiger approach, whereas for the RESP-based approach, anionic inhibitors, such as **5**, have a charge of −0.70e, and zwitterionic inhibitors, −0.54e. However, the Gasteiger method here does not significantly improve ranking, and typically provides a lower estimate of charge magnitude than via the RESP-based approach. Indeed, these charges are more gas-phase-like, correlating with core electron binding energies, and thus are not strictly consistent with the Cornell et al. protein force field, which incorporates average condensed phase polarization effects.

We turn now to consider energetic analysis of inhibitors **1–10**, employing a single-step MM-GB/SA approach (Table 3). Using the ligand **2**/neuraminidase simulation as reference trajectory and RESP-based ligand charges (denoted Model 3, Table 3), ranking of the inhibitors according to $\Delta G_{bind}$ changes relative to PB, notably with overstabilisation of **8** to the extent that **1** is no longer top-ranked. We also note a smaller range in energies here: for example, the five top-ranked ligands are found within 3.0 kcal/mol using GB solvent (Model 3, Table 3), as opposed to within 6.8 kcal/mol via the PB model (Model 1, Table 2). The ranking from $\Delta G_{bind}$ does not appear to be strongly sensitive to choice of reference ligand/protein trajectory: identical ranking and similar absolute energies are found using the two zwitterionic ligand/protein trajectories for **2** and **1** (Models 3 and 4, respectively, Table 3). For the reference trajectory involving **5** (Model 5), the ordering is similar (ranks 1/2, 4/5 and 7/8 interchange) but the stability of inhibitor **8** is still overestimated (Table 3). Considering the electrostatic contribution to binding for this trajectory, denoted Model 5E (Table 3, Fig. 5), an improved correlation with experiment is obtained ($r^2 = 0.72$). It is possible that the consequent improved ranking of **8** arises from an overestimated van der Waals binding component, due to inaccuracies in conformer selection in the protein active site via use of the GB solvent model [36]; however, we suggest below an alternative possibility.

## 4. Discussion

The single-step perturbative approach consistently predicts the experimentally most potent ligand, **1**, as first- or second-ranked according to $\Delta G_{bind}$. We find that the ranking of inhibitors does not appear strongly dependent on reference trajectory (Table 3), initiated from different ligand–neuraminidase crystal structures. This is encouraging, given the difference in size and charge of the

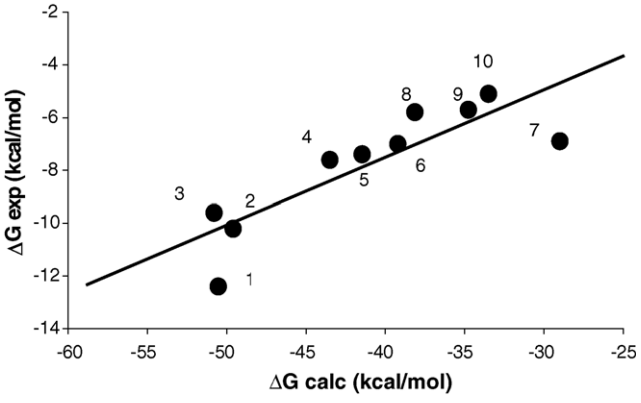


Fig. 5. Plot of experimental free energy of binding against calculated electrostatic free energy of binding for neuraminidase ligands (**1–10**, shown), using Model 5E ($r^2 = 0.72$).

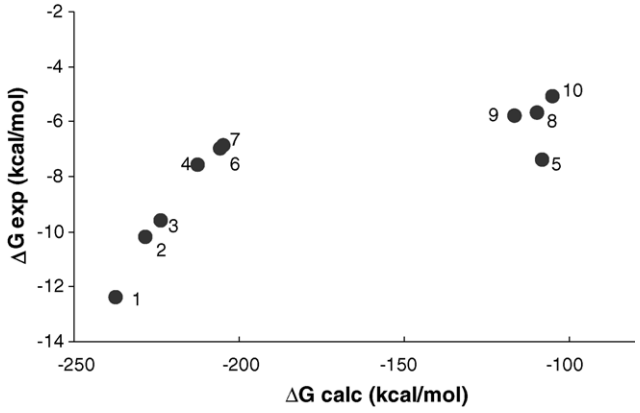


Fig. 6. Plot of experimental free energy of binding against calculated total gas-phase binding enthalpy for neuraminidase ligands (**1–10**, shown), using Model 3G.

substituents at position 4 of the reference ligand. With the exception of ligand **8**, the ranking of ligands does not appear strongly dependent on choice of PB or GB solvation model: although the absolute solvation components differ between the two methods, the relative values of $\Delta G_{solv}$ are more comparable (Tables 2 and 3). Appropriately parametrized GB models can provide good estimates of relative solvation energies [37]. Interestingly, excluding the GB or PB solvation components altogether, to provide an in vacuo binding enthalpy, generally yields a reasonable ranking of inhibitors relative to experiment (see, for example, Model 3G in Table 2S, and Fig. 6). However, this calculation also clearly highlights the large difference in computed gas-phase binding enthalpy for the zwitterionic and anionic inhibitors. Inclusion of solvation contributions leads to relative binding energies in closer agreement with experiment.

Nevertheless, despite the inclusion of solvent effects, the computed $\Delta G_{bind}$ of anionic ligands is still significantly underestimated with respect to zwitterionic inhibitors, both from the single simulation and independent simulations calculations. This effect has also been observed in previous studies using various ligand charge sets and the PB solvent model [15,18,35]. A study of MMP inhibitors [14] found that the PB solvent model underpredicted the solvation free energies of unbound zwitterionic ligands relative to anionic ligands. Inclusion of two explicit waters in the unbound zwitterionic inhibitor trajectories placed the computed affinities in closer proximity to the range of binding values for the anionic inhibitors. In this connection, a recent MM-PB/SA study of four neuraminidase ligands, sialic acid, DANA (**5**), Relenza (**1**) and Tamiflu, predicted a binding free energy for **1** over **5** in good agreement with experiment [38]. These calculations employed 1 ns of molecular dynamics using a model neuraminidase system, comprising protein, ligand and a 30 Å explicit solvent sphere. Snapshots were sampled from dynamics of a 17 Å sphere of motion around the protein active site. Significantly, energetic analysis included explicit crystallographic waters. The resulting total binding free energy for ligand **5** relative to ligand **1** was predicted by the individual trajectory MM-PB/SA approach to be 6.5 kcal/mol [38], close to an experimental value of ~5.0 kcal/mol and considerably less than the computed value of 20.4 kcal/mol obtained in this study from separate trajectories (Table 1).

To more fully understand the energetic basis for this difference in estimate of $\Delta\Delta G_{bind}$ of 13.9 kcal/mol, we compare the absolute $\Delta G_{bind}$ for ligands **1** and **5** from this work and the study of Masukawa et al. [38]: excluding internal energy and solute entropy contributions to binding, the respective computed $\Delta G_{bind}$ are −34.9 and −24.0 kcal/mol from Masukawa et al. [38]. From our analysis, calculated $\Delta G_{bind}$ for ligand **1** is −24.9 kcal/mol (Table 1). This is in closer agreement with the earlier work than for ligand **5**, where $\Delta G_{bind}$ is predicted to be 24.4 kcal/mol less favourable than in the preceding work (Table 1) [38]. In this work, for ligands **1** and **5**, we calculate the respective van der Waals contribution to $\Delta G_{bind}$ as −30.1 and −26.7 kcal/mol, in reasonable agreement with estimates from the earlier study of −30.0 and −23.2 kcal/mol, respectively [35]. As non-polar solvation terms differ by less than 0.5 kcal/mol between the two studies, the difference in calculated $\Delta\Delta G_{bind}$ for the two studies (13.9 kcal/mol) arises principally from the difference in overall electrostatic contributions to protein–ligand interactions (17.9 kcal/mol). Examination of absolute values of the net electrostatic binding energy indicates that this is due to lower predicted stability of ligand **5** in this work relative to that of Masukawa et al. Given that both systems use the same solute radii and use HF/6-31G* RESP-based solute charge set, the most plausible reason for the 17.9 kcal/mol difference in relative electrostatic energy for ligands **1** and **5** lies in use of explicit crystallographic waters associated with the protein in the MM-PB/SA analysis of Masukawa et al. [38]. In particular, this study concluded that explicit water was necessary to mediate the interactions of the weaker binding ligands, such as **5**. Inclusion of explicit water appears to be necessary for quantitative agreement of the MM-PB/SA approach with neuraminidase inhibition data. A similar observation of the need for first solvation shell effects in energetic analysis was made for MM-PB/SA studies of ligand binding to streptavidin [5] and to an RNA aptamer [39].

Consequently, when applied to prediction of relative binding free energies of 10 *Influenza* neuraminidase inhibitors, the single-trajectory perturbative approach provides only qualitative agreement using $\Delta G_{bind}$. Agreement with experiment improves when purely electrostatic contributions to binding are considered; omission of the van der Waals binding contribution penalises inhibitors with larger C4 substitutents, typically the stronger inhibitors, and may be compensating for neglect of explicit active site solvent molecules by stabilisation of the smaller inhibitor complexes. Although quantitative solvent effects may in some cases require a more complete account of short-range

hydrogen bonding effects than afforded by implicit solvent models [40,41], our perturbative MM-PB/SA approach is designed to evaluate multiple ligand structures of differing substituent size. The inclusion of explicit water molecules in this process raises the same complex issues as for computational docking methods.

The best correlation observed here ($r^2$ of 0.72) is comparable in magnitude to a LIE study [4] of fifteen neuraminidase inhibitors [42] where a cross-validation regression coefficient $q^2$ of 0.74 was found. A study of the binding of six inhibitors to two proteins, gelatinase-A and stromelysin-1, using individual trajectory MM-GB/SA analysis, also found a reasonable correlation with experiment ($r^2$ of 0.71) [43]. The latter work employed molecular dynamics in implicit GB solvent to generate the trajectories for post-processing, suggesting again that explicit solvent effects need not always be included to obtain a level of agreement with experiment. Interestingly, correlation decreased slightly through inclusion of an estimate of protein–ligand binding entropy. A robust route to calculation of this quantity is still an open question.

An advantage of the single-step perturbative MM-PB(GB)/SA approach over conventional free energy pertubation techniques is the ability to study large ligand variation in structure. Although the LIE approach also allows this, the method here is computationally more efficient, requiring one reference trajectory. This single reference trajectory is based upon a representative ligand in complex to a protein. We note a single-step perturbative approach to calculation of binding free energies which exhibited reasonable accuracy in ranking inhibitors of the estrogen alpha receptor [44]; this method is based on two non-physical reference trajectories, where a softened non-bonded interaction potential is applied on ligand sites of substitution. The implicit assumption made by both this one-step approach [44] and the work here is that the ligand derivatives adopt a homologous binding mode to the reference ligand, despite some variation in chemical structure. Although further work needs to be done in determining the degree of diversity allowed via this perturbative approach, we have found this to be an acceptable assumption for neuraminidase ligands considered here: using our protocol of a local conformational search/minimization of the perturbed functional group, the predicted structures give reasonable agreement with crystallographic information [15]; however, at the site of perturbation, ligand–protein hydrogen bond distances generally tend to be shorter than observed from the X-ray or MD average values, due to cooling of the 300 K structure via energy minimization [15].

The single-step MM-PB(GB)/SA algorithm is suited for focussed exploration of derivative structures related to hit molecules obtained from preceding virtual screening of large compound databases. In the past, considerable effort has been required to derive reliable electrostatic and non-electrostatic force field parameters for organic small molecule ligands featuring in MM-PB(GB)/SA approaches. We note recent efforts in the direction of more automated parameter assignment, with the advent of a general AMBER force field (GAFF) for organic molecules [45]; and the AM1-BCC (bond charge correction) method [46], designed to produce atom-centred partial charges consistent with more computationally demanding HF/6-31G* RESP calculations. Innovations such as these will facilitate application of single-step MM-PB(GB)/SA approaches to considerably greater numbers of derivative compounds.

## 5. Conclusions

We have explored single-trajectory MM-PB/SA and MM-GB/SA approaches to ranking the binding affinity of 10 *Influenza* neuraminidase inhibitors. The single-step perturbative approach, using either implicit solvent model, predicts the experimentally most potent ligand as first- or second-ranked according to total binding free energy. Ranking of inhibitors is not strongly dependent on choice of reference trajectory or ligand partial charge scheme. When ranked according to net electrostatic binding free energy, reasonable agreement with experiment is obtained ($r^2$ of 0.72).

In its current form, this perturbative approach constitutes a qualitative computational tool for analysis and prediction of neuraminidase–inhibitor interactions. We may conclude that, in the shorter term, the methodology requires validation on protein–ligand systems where the underlying MM-PB(GB)/SA potential is able to furnish an accurate estimate of binding free energy in the absence of critical, short-range solvent effects. In the longer term, further elaboration of the methodology is warranted, to encompass an ability to describe important explicit solute-solvent interactions in the receptor binding pocket, as for example, through direct solvent insertion methods [47] as part of the functional group perturbation process. Other issues such as calculation of solute entropic effects also require further work. In this way, the generality of the approach can be optimised, for use in lead discovery and development strategies in structure-based design.

## Acknowledgment

The authors acknowledge support from Wellcome Trust Grant 063492.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2005.06.003.

## References


[1] V. Helms, R.C. Wade, Computational alchemy to calculate absolute protein–ligand binding free energy, J. Am. Chem. Soc. 120 (1998) 2710–2713.

[2] P.A. Bash, L.L. Ho, A.D. MacKerell, D. Levine, P. Hallstrom, Progress toward chemical accuracy in the computer simulation of condensed phase reactions, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 3698–3703.

[3] B. Kuhn, P.A. Kollman, Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models, J. Med. Chem. 43 (2000) 3786–3791.

[4] J. Aqvist, C. Medina, J.E. Samuelsson, New method for predicting binding affinity in computer-aided drug design, Protein Eng. 7 (1994) 385–391.

[5] B. Kuhn, P.A. Kollman, A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning, J. Am. Chem. Soc. 122 (2000) 3909–3916.

[6] J. Srinivasan, T.E. Cheatham, P. Cieplak, P.A. Kollman, D.A. Case, Continuum solvent studies of the stability of DNA. RNA and phosphoramidate-DNA helices, J. Am. Chem. Soc. 120 (1998) 9401–9409.

[7] M.R. Lee, Y. Duan, P.A. Kollman, Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece, Proteins Struct. Funct. Genet. 39 (2000) 309–316.

[8] C.M. Reyes, P.A. Kollman, Structure and thermodynamics of RNA–protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change, J. Mol. Biol. 297 (2000) 1145–1158.

[9] R.A. Bryce, I.H. Hillier, J.N. Naismith, Carbohydrate–protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A, Biophys. J. 81 (2001) 1373–1388.

[10] T. Hou, S. Guo, X. Xu, Predictions of binding of a diverse set of ligands to gelatinase-A by a combination of molecular dynamics and continuum solvent models, J. Phys. Chem. 106 (2002) 5527–5535.

[11] S. Hou, J. Wang, J. Cioslowski, P.A. Kollman, I.D. Kuntz, Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design, J. Med. Chem. 45 (2002) 1412–1419.

[12] I. Massova, P.A. Kollman, Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies, J. Am. Chem. Soc. 121 (1999) 8133–8143.

[13] P. Kalra, T.V. Reddy, B. Jayaram, Free energy component analysis for drug design: a case study of HIV-1 protease-inhibitor binding, J. Med. Chem. 44 (2002) 4325–4338.

[14] O.A.T. Donini, P.A. Kollman, Calculation and prediction of binding free energies for matrix metalloproteinases, J. Med. Chem. 43 (2000) 4180–4188.

[15] P. Bonnet, R.A. Bryce, Molecular dynamics and free energy analysis of neuraminidase–ligand interactions, Protein Sci. 13 (2004) 946–957.

[16] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model, J. Phys. Chem. 97 (1993) 10269–10280.

[17] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3288.

[18] B.J. Smith, P.M. Colman, M. von Itzstein, B. Daylec, J.N. Varghese, Analysis of inhibitor binding in influenza virus neuraminidase, Protein Sci. 10 (2001) 689–696.

[19] J.N. Varghese, V.C. Epa, P.M. Colman, 3-D structure of the complex of 4-guanidino-neu5ac2en and influenza virus neuraminidase, Protein Sci. 4 (2001) 1081–1087.

[20] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (2000) 1049–1074.

[21] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, AMBER6, University of California, San Fransisco, 1996 .

[22] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzekski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, B. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I.I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Gaussian Inc., Pittsburg, PA, 1998.

[23] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[24] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.

[25] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, J. Comput. Phys. 23 (1977) 327–341.

[26] M.K. Gilson, K.A. Sharp, B.H. Honig, Calculating the electrostatic potential of molecules in solution—method and error assessment, J. Comput. Chem. 9 (1987) 327–335.

[27] B. Jayaram, D. Sprous, D.L. Beveridge, Solvation free energy of biomolecules: parameters for a modified generalized Born model consistent with the AMBER force field, J. Phys. Chem. B 102 (2002) 9571–9576.

[28] W.C. Still, A. Tempczyk, R.C. Hawley, T.J. Hendricksen, Semianalytical treatment of solvation for molecular mechanics and dynamics, J. Am. Chem. Soc. 112 (1990) 6129.

[29] J. Wang, P. Morin, W. Wang, P.A. Kollman, Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA, J. Am. Chem. Soc. 123 (2001) 5221–5230.

[30] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.

[31] M.F. Sanner, A.J. Olson, J.C. Spehner, Reduced surface: an efficient way to compute molecular surfaces, Biopolymers 38 (1996) 305–320.

[32] S.K. Kearsley, On the orthogonal transformation used for structural comparisons, Acta Cryst. A45 (1989) 208–210.

[33] B.R. Gelin, M. Karplus, Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor, Proc. Natl. Acad. Sci. U.S.A. 72 (1975) 2002–2006.

[34] C.T. Holzer, M. von Itzstein, B. Jin, M.S. Pegg, W.P. Stewart, W.-Y. Wu, Inhibition of sialidases from viral, bacterial and mammalian sources by analogs of 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid modified at the C-4 position, Glycoconjugate J. 10 (1993) 40–44.

[35] N.R. Taylor, M. von Itzstein, A structural and energetics analysis of the binding of a series of *N*-acetylneuraminic-acid-based inhibitors to influenza virus sialidase, J. Comput. Aided Mol. Des. 10 (1996) 233–246.

[36] J. Srinivasan, M.W. Trevathan, P. Beroza, D.A. Case, Application of a pairwise generalized Born model to proteins and nucleic

acids: inclusion of salt effects, Theor. Chem. Acc. 101 (1999) 426–434.
[37] S.R. Edinger, C. Cortis, P.S. Shenkin, R.A. Friesner, Solvation free energies of peptides: comparison of approximate continuum solvation models with accurate solution of the Poisson–Boltzmann equation, J. Phys. Chem. B 101 (1997) 1190–1197.
[38] K.M. Masukawa, P.A. Kollman, I.D. Kuntz, Investigation of neuraminidase substrate recognition using molecular dynamics and free energy calculations, J. Med. Chem. 46 (2004) 5628–5637.
[39] H. Gouda, I.D. Kuntz, D.A. Case, P.A. Kollman, Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PB/SA and thermodynamics integration methods, Biopolymers 68 (2003) 16–34.
[40] S.A. Maw, R.A. Bryce, R.J. Hall, A.J. Masters, I.H. Hillier, Integral equation and ab initio study of the effect of solvation on anomeric equilibria in aqueous solution: application to 4,6-dimethyl-2-methoxytetrahydropyran, J. Phys. Chem. B 102 (1998) 4089–4095.
[41] B. Marten, K. Kim, C. Cortis, R.A. Friesner, R.B. Murphy, M.N. Ringnalda, D. Sitkoff, 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) 11775–11788.
[42] I.D. Wall, A.R. Leach, D.W. Salt, M.G. Ford, J.W. Essex, Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method, J. Med. Chem. 42 (1999) 5142–5152.
[43] R.C. Rizzo, S. Toba, I.D. Kuntz, A molecular basis for the selectivity of thiadiazole urea inhibitors with stromelysin-1 and gelatinase-A from generalized Born molecular dynamics simulations, J. Med. Chem. 47 (2004) 3065–3074.
[44] C. Oostenbrink, W.F. van Gunsteren, Free energies of binding of polychlorindated biphenyls to the estrogen receptor from a single simulation, Proteins Struct. Funct. Bioinf. 54 (2004) 237–246.
[45] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general Amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[46] M. Jakalian, B.L. Bush, D.B. Jack, C.I. Bayly, Fast efficient generation of high-quality atomic charges. AM1-BCC model I: method, J. Comput. Chem. 21 (2004) 132–146.
[47] M. Rarey, B. Kramer, T. Lengauer, The particle concept: placing discrete water molecules during protein–ligand docking predictions, Proteins Struct. Funct. Genet. 34 (1999) 17–28.