Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method

Ian D. Wall,[†] Andrew R. Leach,[‡] David W. Salt,[§] Martyn G. Ford,[§] and Jonathan W. Essex*[,†]

Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K., Glaxo Research and Development, Glaxo Medicines Research Centre, Chemistry Building, Gunnels Wood Road, Stevenage, Herts SG1 2NY, U.K., Centre for Molecular Design, University of Portsmouth, 1/2 Halpern House, Portsmouth, Hants PO1 2DY, U.K.

Received March 8, 1999

The linear interaction energy (LIE) method has been applied to the calculation of the binding free energies of 15 inhibitors of the enzyme neuraminidase. This is a particularly challenging system for this methodology since the protein conformation and the number of tightly bound water molecules in the active site are known to change for different inhibitors. It is not clear that the basic LIE method will calculate the contributions to the binding free energies arising from these effects correctly. Application of the basic LIE equation yielded an rms error with respect to experiment of 1.51 kcal mol^{-1} for the free energies of binding. To determine whether it is appropriate to include extra terms in the LIE equation, a detailed statistical analysis was undertaken. Multiple linear regression (MLR) is often used to determine the significance of terms in a fitting equation; this method is inappropriate for the current system owing to the highly correlated nature of the descriptor variables. Use of MLR in other applications of the LIE equation is therefore not recommended without a correlation analysis being performed first. Here factor analysis was used to determine the number of useful dimensions contained within the data and, hence, the maximum number of variables to be considered when specifying a model or equation. Biased fitting methods using orthogonalized components were then used to generate the most predictive model. The final model gave a q^2 of 0.74 and contained van der Waals and electrostatic energy terms. This result was obtained without recourse to prior knowledge and was based solely on the information content of the data.

Introduction

The calculation of free energies of binding can play a critical role in rational drug design. Such calculations have traditionally been carried out using the free energy perturbation (FEP) or thermodynamic integration (TI) methods (for reviews, see refs 1−3). These procedures are theoretically exact and, if implemented appropriately, can yield very precise estimates of free energies of binding. However, to obtain good convergence and hence high precision, the extensive simulation of nonphysical intermediate states is often required. Consequently, the methods are highly computationally intensive, making them impractical for application to many industrial pharmaceutical problems. Clearly a primary objective of current research has to be to obtain precision and accuracy similar to that of these methods in a more computationally efficient fashion.

A host of methods have been proposed to increase the efficiency of such calculations either by reducing the amount of data collection required through enhanced sampling or by increasing the amount of information which may be calculated from each simulation.[4−9]

One of the most attractive methods is the linear interaction energy (LIE) method[10] recently proposed by Åqvist and co-workers. This is a semiempirical technique for the calculation of free energy changes based on the simulation of only two states. In the case of the calculation of binding free energies, these two states are (i) the solvated ligand and (ii) the ligand bound to the solvated protein. The theoretical foundation for the method is linear response theory,[10,11] on the basis of which Åqvist and co-workers proposed an equation of the form

$$\Delta G = 0.5\langle \Delta U_{elec} \rangle + \alpha \langle U_{vdw} \rangle \qquad (1)$$

for these calculations. ΔU_{elec} and ΔU_{vdw} are the differences in the averaged inhibitor-environment electrostatic and van der Waals energies, respectively, between the two simulations, and the angled brackets denote ensemble averages. This original work fitted the equation to the experimental free energies of binding for a series of four endothiapepsin inhibitors, generating a value of 0.161 for α.

The proposed method has since been successfully applied to a variety of different systems[10,12−16] including HIV-1 protease,[12,17] trypsin,[18] and thrombin.[15] While Åqvist has reported good transferability of the parameters to a limited number of other systems, other workers[15,16] have not been so successful. Moreover, the data sets to which the method has been applied have mostly been quite small. Consequently, the issue of the transferability of LIE parameters has not been fully addressed.

Recently, however, Wang et al.[19] investigated the transferability of the van der Waals coefficient in more detail. Simulations of trypsin-benzamidine[18] and camphor-P450cam[16] systems had previously been carried

[†] University of Southampton.
[‡] Glaxo Medicines Research Centre.
[§] University of Portsmouth.

10.1021/jm990105g CCC: $18.00 © 1999 American Chemical Society
Published on Web 12/01/1999

out using the GROMOS and CVFF force fields, respectively. These simulations were repeated using the Cornell et al. force field[20] with a similar simulation setup resulting in similar values of α. Hence it was concluded that the value of α was not strongly dependent on the force field. Similar simulations on five further systems, giving a total of seven ligands interacting with five proteins, showed that one fixed α could not give results in agreement with experiment in all cases. Since the work looks at a range of proteins and ligands all simulated under the same protocol, it provides the most reliable data to suggest the value of α is dependent on the local environment. This work was carried out with the value of β fixed to 0.5, and hence a similar analysis is not available for this parameter. It has, however, been noted by Hansson et al. that the value of β becomes smaller than the theoretical 0.5 as the number of −OH groups increases.[13]

Several questions concerning the methodology remain unanswered. It is unclear how intramolecular energy contributions are accounted for when only intermolecular energy contributions are evaluated, whether free energy changes associated with expulsion of bound water from the active site can be reproduced, and whether the entropy change associated with binding is fully explained by the parametrization. Experience of calculating free energies of hydration of simple molecules also suggests that the statistical basis of the LIE method requires careful assessment.[21] In an effort to address these issues, a study of inhibitors of the protein neuraminidase is reported here using LIE. This system provides an excellent test case for which several of the aforementioned points of interest are pertinent. For example, one of the protein residues is known to undergo a conformational change upon the binding of certain inhibitors, and the number of tightly bound water molecules is known to vary depending on the inhibitor. There is also a wide range of binding constants despite the structural differences between inhibitors being small. Hence FEP calculations on these systems are practicable.

Neuraminidase is an influenza enzyme which facilitates the release of virons from infected cells by cleaving sialic acid residues from the carbohydrate side chains. It has a tetrameric structure consisting of four polypeptide chains each containing around 470 amino acid residues depending on the strain of the virus. These chains consist of globular headgroups on the end of a long stalk. The structure of the stalk has not yet been established, but this should not be of practical significance since it is so far from the active site. Structures of the headgroup are known for several strains of neuraminidase, and the work presented here is based on a monomer of the N2 strain of influenza A.

Methods

The original LIE work proposed eq 1 for the calculation of binding free energies, with an α value of 0.161. Subsequent application of the equation in that form by the same research group to HIV-1 protease and trypsin inhibitor systems gave rms deviations from experiment of between 1.0 and 2.2 kcal mol^{-1}.[12,17,18] Other workers have also applied the method,[15,16] but, transfer of the α parameter to their systems did not yield accurate predictions. However, good agreement between simulation and experiment was achieved upon reparametrization of the equation. When Paulsen and Ornstein applied the method to the calculation of the binding affinity of a series of P450cam substrates,[16] an optimal fit was achieved with an α value of 1.043. When Jones-Hertzog and Jorgensen applied the method to a series of thrombin-inhibitor complexes,[15] the application of eq 1 gave an optimal α value of 0.734. However, even having optimized the parameter to the system of interest, the rms deviation from experiment was 4.40 kcal mol^{-1}; the equation in this form did not provide a satisfactory fit to the data. In light of this, modified forms of the equation were applied. First, the electrostatic coefficient β was included in the parametrization:

$$\Delta G = \beta\langle\Delta U_{elec}\rangle + \alpha\langle\Delta U_{vdw}\rangle \quad (2)$$

Optimization of this equation generated values of $\alpha = 0.476$ and $\beta = 0.165$ with a much improved rms deviation of 1.34 kcal mol^{-1}. As an extension of previous work which had been carried out using LIE to calculate free energies of hydration,[22] the equation was extended to include a term to account for the change in solvent accessible surface area (SASA), whence:

$$\Delta G = \beta\langle\Delta U_{elec}\rangle + \alpha\langle\Delta U_{vdw}\rangle + \gamma\langle\Delta SASA\rangle \quad (3)$$

Again, optimizing α, β, and γ yielded an improvement in the fit to the experimental data; values of $\alpha = 0.236$, $\beta = 0.146$, and $\gamma = 0.010$ gave an rms deviation of 1.15 kcal mol^{-1}. This trend is perhaps unsurprising since increasing the number of explanatory variables in the equation will invariably improve the goodness of fit. Indeed if the number of explanatory variables equals the number of observations a perfect fit is achieved by definition, although the model will not have been generalized. Consequently, such a model would have a low power of prediction.

In addition to binding studies, the LIE method has also been applied to the predictions of free energies of solvation and hydration[21−23] with good success. The addition of the SASA term to the LIE equation was first proposed in these papers, its inclusion being rationalized as accounting for the cost of cavitation[22] and, hence, allowing for the possibility of positive free energies of hydration. Most recently another study of free energies of hydration suggested that using all the terms in eq 3 may in fact be overfitting the data.[21] The solvent accessible surface area and van der Waals terms were highly correlated, and other simpler equations were shown to perform almost as well. In particular, fitting to the equation

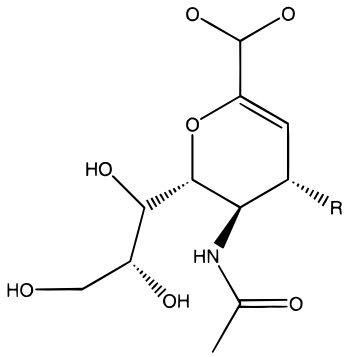
$$\Delta G = \alpha + \beta\langle\Delta U_{elec}\rangle \quad (4)$$

gave an rms deviation from experiment of just 0.74 kcal mol^{-1} for a set of 22 small organic molecules.

In this light, a detailed statistical investigation of the method is called for. Moreover, the questions previously posed regarding the inclusion of additional factors in the equation need addressing such that the optimum fitting equation is generated without making any a priori assumptions regarding the validity of any particular variable. Recently, a basic statistical analysis has been reported,[13] although it is not clear that the statistical methods chosen were the most appropriate for the data sets investigated. A set of 15 diverse neuraminidase inhibitors was studied to investigate these issues.

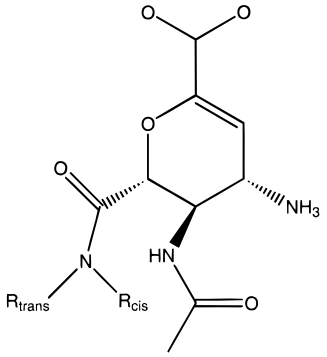
Computational Details

Simulation systems were prepared for a set of 15 neuraminidase inhibitors. Five were based on DANA (2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid) (Figure 1), and the remainder were based on the amide derivatives of this molecule (Figures 2 and 3). Experimental data were available for the DANA structures in the form of K_is,[24] whereas the data for the amide derivatives were in the form of IC_{50}s[25] and were



Inhibitor	R	$K_i/10^{-6}$	ΔG/kcal mol^{-1}
11	$-NH_3^+$	0.04	−10.5
12	$-NH_3^+$ (*R*)	0.1	−9.94
13	$-N(CH_3)_3^+$	10	−7.10
14	−OH	4	−7.66
15	$-NH{=}C(NH_2)_2^+$	0.001	−12.78

Figure 1. Experimental data for DANA type structures.



Inhibitor	R_{trans}	R_{cis}	IC_{50}/nM	$K_i/10^{-9}$	ΔG/kcal mol^{-1}
1	−Me	−H	190 000	24 000	−6.53
2	−Et	−H	13 000	1690	−8.16
3	−Me	−Me	2400	310	−9.20
4	−Et	−Et	3	0.39	−13.29
9	$-(CH_2)_2Ph$	−H	12 000	1570	−8.242
10	$-(CH_2)_2Ph$	−Pr	5	0.65	−13.04

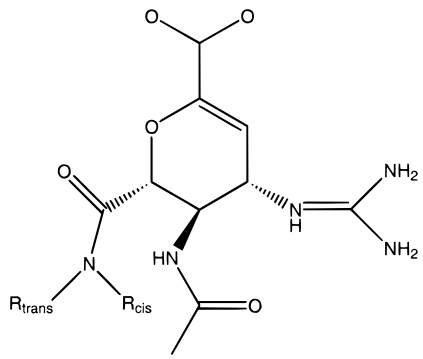
Figure 2. Experimental data for amino type structures.

converted to K_is using the Cheng−Prusoff equation:[26]

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}} \quad (5)$$

K_i is the Michaelis constant and [S] is the substrate concentration. These values were obtained from the authors of the experimental paper and are 15×10^{-6} and 100 μM, respectively.[27] Discussions with the experimentalists also assured us that the application of the Cheng−Prusoff equation to this data set was valid.[27] Experimental errors on the K_is and IC_{50}s were quoted as approximately half an order of magnitude. This corresponds to about a 1 kcal mol^{-1} error in the experimental free energy of binding.

There is excellent diversity in the data, both in terms of the range of inhibition constants and in the range of



Inhibitor	R_{trans}	R_{cis}	IC_{50}/nM	$K_i/10^{-9}$	ΔG/kcal mol^{-1}
5	−Me	−H	7000	910	−8.53
6	−Me	−Me	25	3.3	−11.98
7	−Et	−Et	1	0.13	−13.68
8	$-(CH_2)_2Ph$	−Pr	5	0.65	−13.04

Figure 3. Experimental data for guanadino type structures.

structures of the inhibitors. Of particular interest is the fact that the number of tightly bound waters varies between the hydroxyl and guanadino forms (Figure 4). The figure shows that when the DANA molecule is bound (pdb structure 1nsd), a water molecule is clearly stabilized between the hydroxyl group of the inhibitor and residues GLU[227] and TRP[178]. However, when this hydroxyl substituent is replaced with a guanadino,[28] the water is expelled from the active site. A difference in the number of tightly bound waters has also been reported between the amino and guanadino forms.[25] Moreover, there is a conformational change known to take place in residue GLU[276] between the binding of the glycerol type inhibitors and the amide type inhibitors. Hydrogen bonds are known to exist between GLU[276] and the glycerol chains of the DANA inhibitors (pdb structure 1ivf). However, when this glycerol group is replaced by an amide,[28] GLU[276] has been observed to rotate[25] and form a salt bridge to ARG[224] (see Figure 5).

The protein structure used in the simulations was that of the N2 strain of neuraminidase with sialic acid bound (Brookhaven protein data bank entry 2bat). The crystal structure was examined for the correct orientation of the histidine residues. This was done by considering whether the hydrogen bonding structure would clearly be enhanced by rotating any of the histidines through 180°. The analysis revealed no significant evidence that any residues should be rotated. The pH of the experimental studies was 6.5 so the histidine residues were assumed to be protonated unless there was evidence that a hydrogen bond could be accepted from another residue. No such evidence was found. Since the crystallographic structure does not contain any hydrogen atoms, they were added in the minimum energy conformation of the appropriate torsional potential. However, in the case of hydrogen bond donating residues, this is not necessarily the most appropriate conformation. Hence, all threonine, serine, and tyrosine residues were examined in turn to see whether the hydrogen bonding pattern indicated that an alternative conformation would be more appropriate. For example if the hydrogen had been assigned in the trans configuration, but a clear hydrogen bond was available in the

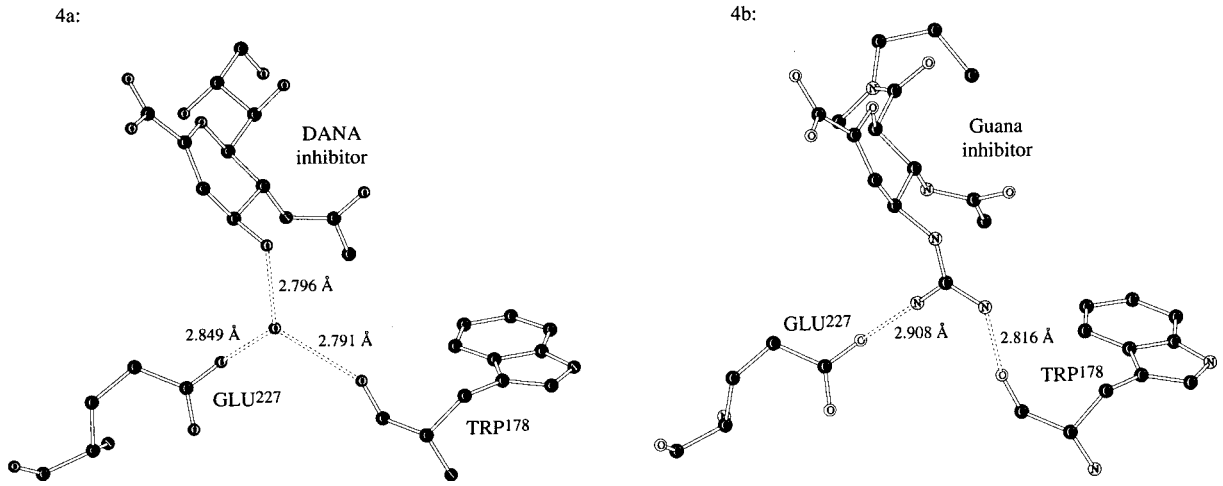


Figure 4. X-ray crystallographic structures of DANA (**4a**, pdb structure 1nsd) and DANA with OH replaced with guanadino (**4b**, unpublished structure obtained from GlaxoWellcome), showing expulsion of water from active site by the guanadino group.

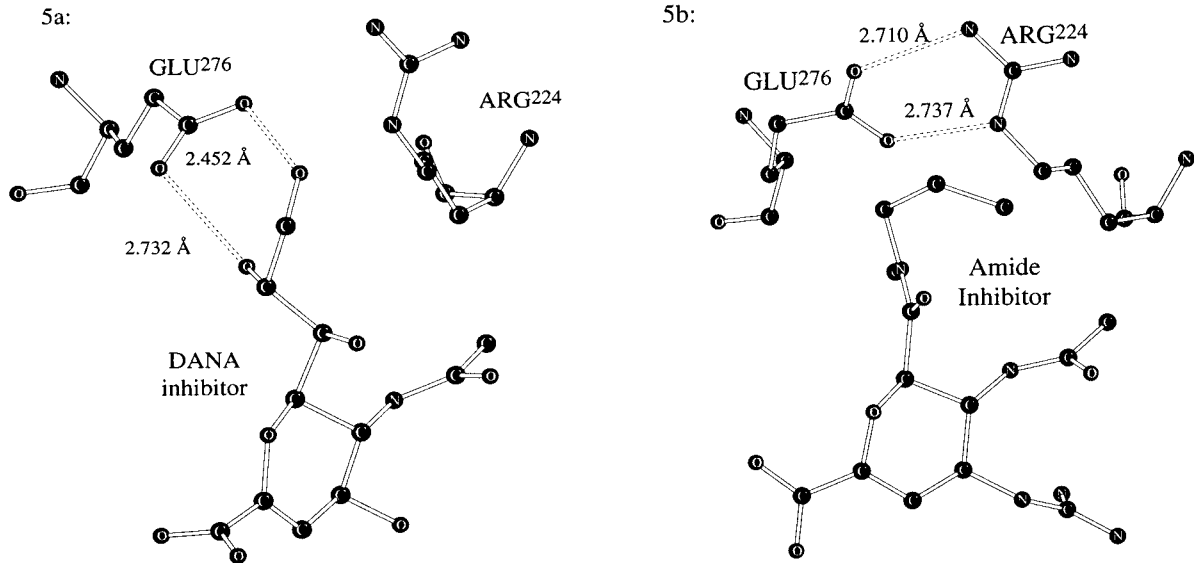


Figure 5. X-ray crystallographic structures of DANA (**5a**, pdb structure 1ivf) and one of its amide derivatives (**5b**, unpublished structure obtained from GlaxoWellcome, showing the two possible conformations of GLU^{276}.

gauche conformation, then a change was made to this conformation. As with the histidines, if there was any doubt about the validity of such a rearrangement, the atom was left in its original position.

The only structure available with an amide type inhibitor (structure **4g** ref 25) was for the N9 strain of neuraminidase.[28] However, the active site structure is known to be conserved between the various strains of neuraminidase,[29] and the binding orientation of DANA inhibitors is the same in N2 and N9 (structures 1ivf and 1nnb). Consequently, the docking of the inhibitors into the N2 system was carried out on the basis of the N9 structure. All the inhibitors were assumed to bind in the same orientation. The structure so obtained was then used as a template for the binding of the other inhibitors, with docking being carried out by superimposing the ring atoms. For the phenyl substituted amides, the docking conformation of the aromatic rings was not clear. Consequently, the amide substituents were placed in the all trans conformation, and an extra 1 M (million) configurations of Monte Carlo equilibration were carried out prior to the standard simulation. During this period only, inhibitor and water moves were carried out using preferential sampling[30] to enhance motion in the vicinity of the active site, and the protein was constrained to its crystallographic coordinates. Approximately 65 crystallographic water molecules were included in the calculations depending on the exact inhibitor structure. All the crystallographic waters in structure 2bat were included. Other neuraminidase structures from the pdb were superimposed, and any waters absent from 2bat were included provided there was no steric clash. As larger inhibitors were docked, any of these waters causing a steric clash were also removed. For all amide type inhibitors GLU^{276} was rotated to form a salt bridge with ARG^{224} as observed in crystallographic studies.[25] The crystallographic conformation was not altered for the DANA based inhibitors. The conformation of ARG^{224} also had to be altered to enable this salt bridge formation; however, crystallographic evidence (1nnb, 1nnc, 1ivf) suggests that this new conformation was also more appropriate for the DANA type structures. It has also been reported that there is little change in the conformation of ARG^{224} as the conformation of GLU^{276} changes,[25] so all systems

were set up with ARG[224] in the conformation shown in the crystal structure with the amide inhibitor bound.[28]

The N2 strain of neuraminidase contains 388 residues. Adequate solvation of the full protein using periodic boundary conditions would have required approximately 7000 water molecules. Simulating a system of this size would have been impracticable. Thus a spherical approximation was used, including only those residues within a specified distance of the active site. In such a system the outer residues must be kept frozen since they are not being restrained by interactions with their neighbors. The Monte Carlo method was therefore chosen since freezing these outer residues is trivial. Monte Carlo simulations were carried out with the MCPRO package.[31] For the simulations of the inhibitors bound to the protein, residues within 20 Å of the active site were included in the simulations, and those within 15 Å of the central atom of the inhibitor were sampled. The system was hydrated by a ball of TIP4P water[32] of radius 20 Å centered on the central atom of the inhibitor. Any water molecules whose oxygen was within 2.5 Å of another non-hydrogen atom were discarded, resulting in the inclusion of approximately 375 water molecules. This simulation protocol has been successfully applied to various protein systems.[15,33] The free ligand simulations were carried out by simply placing the inhibitor at the center of a 20 Å sphere of water containing approximately 1110 water molecules. In both simulations the water molecules were restrained to prevent evaporation using a half harmonic-restraint 20 Å from the central atom with a force constant of 1.5 kcal mol^{-1} $Å^{-2}$. The use of a spherical hydration system rather than conventional periodic boundary conditions has been demonstrated to affect calculated free energies of hydration in simple systems.[34] To reduce the influence of this approximation on the free energies of binding, an identical system size of 20 Å was therefore adopted for the free and bound simulations.

It has been noted that it is important to have the same net charge for the bound and free states to avoid having to make ill-defined Born-type corrections.[10] This was achieved in these systems by neutralizing the charge on the protein. The charge on the 20 Å sphere of protein was initially +5, so to neutralize this the five nearest negatively charged residues outside the 20 Å were also included in the simulation system. All inhibitor−solvent and inhibitor−protein interactions were evaluated. Solvent−solvent interactions were truncated at 20 Å, protein−solvent and intramolecular nonbonded interactions were truncated at 12 Å. A residue based cutoff was used—that is to say that if any atom of a residue is within the cutoff radius all interactions with that residue are evaluated. Similarly, solvent interactions are based on the position of the oxygen atom such that if this is within the cutoff radius then interactions with the whole water molecule are included.

Protein side chains were sampled by varying the bond angles and torsions. The inhibitor was partially flexible with ring side chains again having bond angles and torsions sampled, but the ring was fixed. Crystallographic evidence suggests that there is little variation in the ring structure (pdb structures 1nsd, 1ivf, 1nnb) and hence that this is a reasonable approximation. Attempted Monte Carlo moves for the protein involved randomly selecting a residue and making random variations to all flexible angles and dihedrals therein. Similarly for the inhibitor, the same variations are made with additional whole body translations and rotations being carried out in the bound simulations. In the free simulations total body translations were not allowed. Finally, water molecules were represented by the TIP4P water[32] model; since this is a rigid water model, only total translations and rotations were attempted. Protein residue moves were attempted every five configurations and inhibitor moves every 99 configurations, with the remainder of the attempted moves involving water molecules.

The simulation of the protein environment consisted of 15 M configurations of equilibration with only the water sampled, followed by 3 M configurations of equilibration with angles and dihedrals sampled as outlined above, and finally four batches of 5 M data collection steps. Similarly for the free ligand, 10 M configurations of water equilibration, 3 M steps of fully flexible equilibration, and five batches of 5 M data collection steps were carried out. In addition to the two basic types of simulation required to implement LIE, it was decided to carry out an extra simulation of the free protein. This enabled the calculation of changes in protein energies associated with the binding process in addition to inhibitor energy changes. This simulation followed the same protocol as those for the bound inhibitors. The simulations were carried out at a constant temperature of 37 °C in agreement with the binding experiments. Solvent accessible surface area calculations were carried out after the simulations on coordinates saved every 100 000 configurations by applying the Richmond algorithm[35] based on its implementation in TINKER.[36]

The OPLS united atom force field[37] was used for the protein. Inhibitor charges, van der Waals parameters, and angle bending parameters were assigned by analogy to the OPLS force field, and the nonbonded parameters are given in the Supporting Information. 1,4-Distributed electrostatic and van der Waals interactions were scaled by a factor of a $^1/_2$. However, it was not possible to assign all the inhibitor torsional parameters by analogy to OPLS. Those dihedral parameters not obtainable by analogy with standard OPLS values had to be derived independently. Details of the derivation procedure and the resulting parameters are given in the Supporting Information.

Results and Discussion

Energies. The data collection phase of the simulations was split into four successive batches of 5 M configurations for the protein-bound calculations and five batches of 5 M configurations for the aqueous calculations. The averages of the following terms are calculated for each batch—van der Waals, electrostatic, intramolecular energies, and the solvent accessible surface area. The overall means and standard errors are then determined from these batch averages. The average difference between the bound and free simulations for each of these terms is shown in Table 1. ΔU_{elec} and ΔU_{vdw} are the changes in electrostatic and van der Waals energies, respectively; ΔU_{intra_prot} and ΔU_{intra_inhib} are the changes in the intramolecular energy of the

Table 1. Energy Components Calculated from Monte Carlo Simulation[a]

inhibitor	ΔU_{vdw}	ΔU_{elec}	$\Delta SASA_{inhib}$	$\Delta SASA_{prot}$	ΔU_{intra_inhib}	ΔU_{intra_prot}
1	−8.186	−59.440	−432.941	−170.494	9.879	−44.447
2	−5.166	−66.369	−469.462	−192.273	11.529	−77.308
3	−9.701	−61.202	−453.177	−197.578	6.963	−82.159
4	−11.027	−80.192	−491.695	−242.094	19.165	−76.996
5	−19.671	−29.331	−485.457	−160.203	6.830	−26.480
6	−20.132	−32.002	−502.203	−199.258	10.169	−41.960
7	−19.924	−50.641	−546.761	−176.860	18.040	−31.531
8	−21.539	−42.249	−653.697	−263.619	16.476	14.537
9	−13.604	−31.429	−505.174	−234.388	−9.734	−26.190
10	−11.875	−61.397	−570.817	−230.165	10.033	−46.771
11	−10.943	−71.154	−453.739	−212.135	7.699	−115.124
12	−18.572	−49.297	−456.739	−195.791	12.774	−41.316
13	−6.217	−48.183	−515.188	−206.846	18.179	−78.924
14	−8.945	−69.303	−450.498	−212.408	8.148	−84.122
15	−18.539	−36.616	−506.006	−222.799	4.936	−64.057

[a] All energies in kcal mol^{-1}.

protein and inhibitor, respectively; $\Delta SASA_{prot}$ and $\Delta SASA_{inhib}$ are the corresponding changes in solvent accessible surface area. The nonbonded contribution to the change in the intramolecular energy, ΔU_{intra_prot}, of the protein was particularly noisy owing to cutoff effects. In an effort to prevent this from dominating the statistics, an extra energy variable was calculated consisting of only the change in angle and dihedral energies of the protein—excluding the noisy nonbonded term.

The simulation data (Table 1) exhibit an unusual feature. It is observed that ΔU_{intra_inhib} (the change in intramolecular energy of the inhibitor) is negative for inhibitor **9**. Careful analysis of the simulation trajectory and the use of simulated annealing have helped to identify two possible reasons for this. Gas phase simulated annealing calculations identified several minima of roughly the same energy. Thus it is unclear whether the starting geometry of the inhibitor is in fact the appropriate energy minimum when either solvated or bound to the protein. In the aqueous simulation the inhibitor is only sampling states close to its starting geometry, whereas in the bound state, interactions with the surrounding protein serve to drive conformational change, possibly to a lower energy state. It is also possible that the strong hydrogen bonding network formed with the surrounding waters serves to increase the barrier height between the two states in the aqueous simulation. Furthermore, in solution the torsion about the bond joining the amide to the ring is pulled away from its equilibrium position into a relatively high energy state.

One of the most interesting features of the neuraminidase protein is the ability of the GLU^{276} residue to adopt two very different conformations depending on the inhibitor. Crystallographic evidence indicates that the residue hydrogen-bonds to the inhibitor when a glycerol side chain is present at the C6 position, but when this is replaced by an amide type substituent, GLU^{276} flips to form a salt bridge with ARG^{224}.[25] Those structures which started in the salt bridging form showed no sign of moving away from that conformation during the simulation. However, two of the simulations involving the glycerol type inhibitors, **12** and **13**, did move away from their their starting configurations into the salt bridging form. There was no evidence to suggest that they were returning to the starting configuration although it is not clear whether such a change would be observed in a longer simulation. Previous simulation work on neuraminidase has also observed considerable flexibility in this residue,[24] suggesting that the energy balance between the two conformations is subtle and therefore unlikely to affect binding.

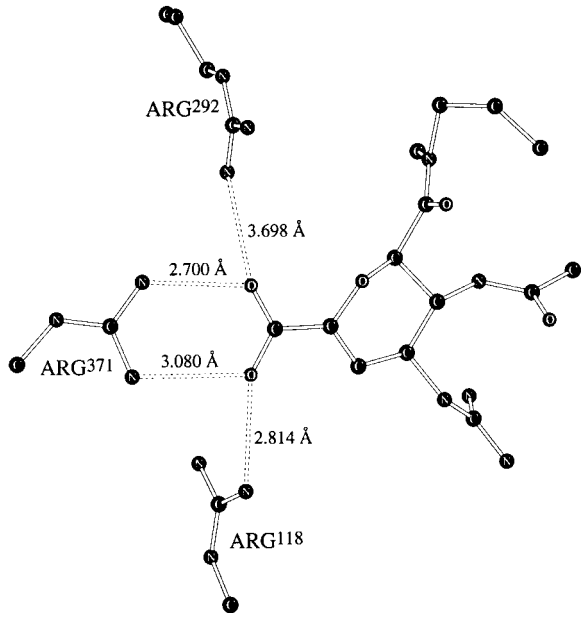


Figure 6. X-ray crystallographic structure obtained from GlaxoWellcome showing the carboxylate part of inhibitors held strongly in position by four hydrogen bonds.

The binding mode of the inhibitors is dominated by a complicated network of hydrogen bonds which hold the carboxylate fragment strongly in position (Figure 6). Consequently, very little motion is observed in this part of the active site during the simulations. Both the experimental and simulation results (see later) show the guanadino forms of the inhibitors to be, on average, stronger binders than their amino analogues, although this effect becomes less apparent as the size of the inhibitors increases to the point where both propyl, phenethyl inhibitors have the same binding constants.

A detailed analysis was undertaken to try to rationalize these observations. First, a hydrogen bonding analysis was carried out whereby the number of hydrogen bonds involving each inhibitor were calculated for both the bound and free inhibitors during the simulations. The change in number of hydrogen bonds on binding

Table 2. Predicted ΔG Using Equations Proposed by Other Workers

inhibitor	experiment	Åqvist $\alpha = 0.161$, $\beta = 0.5$, $\gamma = 0.00$ (ref 10)	Paulsen $\alpha = 1.043$, $\beta = 0.500$, $\gamma = 0.00$ (ref 16)	Jones-Hertzog $\alpha = 0.476$, $\beta = 0.165$, $\gamma = 0.00$ (ref 15)	Jones-Hertzog $\alpha = 0.236$, $\beta = 0.146$, $\gamma = 0.010$ (ref 15)
1	−6.5	−31.0	−38.3	−13.7	−14.9
2	−8.2	−34.0	−38.6	−13.4	−15.6
3	−9.2	−32.2	−40.7	−14.7	−15.8
4	−13.3	−41.9	−51.6	−18.5	−19.2
5	−8.5	−17.8	−35.2	−14.2	−13.8
6	−12.0	−19.2	−37.0	−14.9	−14.5
7	−13.7	−28.5	−46.1	−17.8	−17.6
8	−13.0	−24.6	−43.6	−17.2	−17.8
9	−8.2	−17.9	−29.9	−11.7	−12.9
10	−13.0	−32.6	−43.1	−15.8	−17.5
11	−10.5	−37.3	−47.0	−17.0	−17.5
12	−9.9	−27.6	−44.0	−17.0	−16.2
13	−7.1	−25.1	−30.6	−10.9	−13.7
14	−7.7	−36.1	−44.0	−15.7	−16.7
15	−12.8	−21.3	−37.6	−14.9	−14.8
rms		19.7	30.6	5.2	6.0

[a] All energies in kcal mol^{-1}.

was then plotted against the free energy of binding to determine correlations; there was no evidence of a correlation between the two. The simulation structures were also examined visually, and a justification was found for the preferential binding of guanadino substituents over amino. This is at least in part due to the expulsion of crystallographic water from the active site by the guanadino group. The gain in entropy associated with the expulsion of this water is likely to have a favorable effect on the free energy of binding.

It is also noted that for both the amino and guanadino forms there is a clear trend in the binding affinities: diethyl > dimethyl > methyl, with the propyl, phenethyl fitting between the diethyl and dimethyl. Thus increasing the size of these substituents increases the binding affinity up to a point, but it seems that the bulky phenethyl inhibitors may be bigger than the optimum size for the active site. Examination of the simulation structures has not identified any specific interactions responsible for these trends. Hence it seems that the stronger binding of the molecules with larger amide substituents is likely to be due to hydrophobic effects. The addition of increasingly hydrophobic groups makes the solvation of the inhibitors increasingly unfavorable thereby increasing their affinity to bind to the protein. Indeed this may in part explain the low coefficient for the electrostatic term in the LIE equation derived in the next section.

LIE Calculations. Having calculated the energies the first aim was to investigate whether the linear response parameters derived by other workers were transferable to the neuraminidase system. To assess this, previously proposed equations were applied to the current data set. The equations used, the corresponding calculated free energies, and the rms deviations with respect to experiment are shown in Table 2. It is clear that the coefficients derived by Åqvist et al.[10] and Paulsen and Ornstein[16] are not predictive when applied to our system. This is in some ways not surprising since these workers used different force fields and electrostatic calculation schemes. However, the work of Jones-Hertzog and Jorgensen on thrombin inhibitors[15] used the OPLS force field and a similar simulation protocol, and although the agreement between predicted and experimental values is indeed much closer than in the other cases, an rms deviation of 5.2 kcal mol^{-1} (in the best case) shows that the parameters are still not transferable.

Table 3. Coefficients and rms Errors (in kcal mol^{-1}) Derived by Least Squares Fitting to Equations Proposed by Other Workers

	α	β	γ	rms
$\beta\Delta U_{elec} + \alpha\Delta U_{vdw} + \gamma\Delta SASA_{inhib}$	0.338	0.066	0.004	1.47
$\beta\Delta U_{elec} + \alpha\Delta U_{vdw}$	0.418	0.087	0	1.51
$\beta\Delta U_{elec} + \alpha$	−10.319	−0.001	0	2.44

To derive the coefficients which give the best fit to the neuraminidase data, a series of possible equations were examined based on previous work (see Table 3). Also given are the rms values of the calculated free energies of binding with respect to experiment. The coefficients were calculated by least-squares fitting. Fitting to the equation

$$\Delta G = \beta\Delta U_{elec} + \alpha\Delta U_{vdw} + \gamma\Delta SASA \qquad (6)$$

gave an rms error of 1.47 kcal mol^{-1}, consistent with the type of agreement other workers have observed. It is noteworthy that the calculated coefficients vary considerably between equations and are therefore unreliable. This strongly suggests that regression based on ordinary least-squares fitting is inappropriate for this data set (see later).

Statistical Considerations. The possibility of including extra energy terms in the LIE expression was investigated using the following general equation

$$\Delta G = \beta\Delta U_{elec} + \alpha\Delta U_{vdw} + \gamma\Delta SASA_{inhib} + \delta\Delta SASA_{prot} + \epsilon\Delta U_{intra\_inhib} + \zeta\Delta U_{intra\_prot} + \eta \qquad (7)$$

where η is a constant. However, with only 15 compounds in the data set, eq 7 is an overdefined model, and the contributions of each descriptor must be carefully assessed. Such analysis is often carried out using the multiple linear regression (MLR) method. One important assumption that underpins the MLR method is that the predictor variables are independent. Therefore a correlation analysis was carried out (Table 4). For the current data set with 13 degrees of freedom, correlation coefficients of greater than 0.51 or less than −0.51 are significant at the 5% level. The presence of several significant correlations indicates that MLR is an inap-

Table 4. Correlation Analysis of Energy Components

	ΔG	ΔU_{vdw}	ΔU_{elec}	$\Delta SASA_{inhib}$	$\Delta SASA_{prot}$	ΔU_{intra_inhib}
ΔU_{vdw}	0.600					
ΔU_{elec}	−0.009	−0.663				
$\Delta SASA_{inhib}$	0.612	0.492	−0.321			
$\Delta SASA_{prot}$	0.450	0.063	0.120	0.577		
ΔU_{intra_inhib}	−0.337	−0.006	−0.385	−0.267	−0.026	
ΔU_{intra_prot}	−0.250	−0.689	0.655	−0.655	−0.115	−0.029

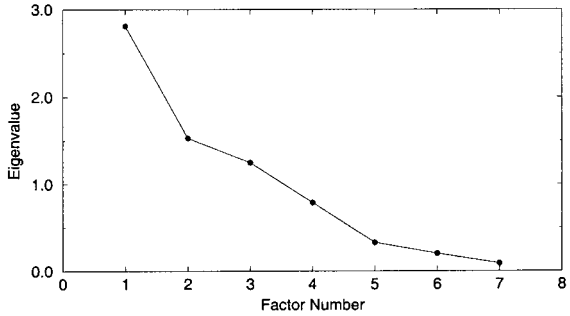


Figure 7. Scree plot from unsupervised factor analysis.

propriate procedure for the analysis of these data. Consequently, a biased regression method, continuum regression (CR), was used to fit to eq 7.

To maximize the predictive ability of any model it is necessary to reduce the dimensionality of the data by identifying the most important variables from which to construct orthogonalized components. As a general rule, parsimonious models are likely to be more predictive since the objects upon which the model has been specified are likely to span the low-dimensional space more efficiently. Before carrying out a fitting procedure based on orthogonalized components, eq 7 needs to be reduced to include only those energy terms that contain useful and unique information.

Selection of Descriptor Variables for Model Specification. The approach to this problem was to carry out a factor analysis.[38] This procedure spans the multidimensional property space of the data using a series of orthogonal factors constructed to simplify the correlations between the original variables and the factors. A variable consisting of the change in bond angle and dihedral energies of the protein was also included in the factor analysis. An unsupervised factor analysis was initially carried out on the simulation variables only, and a scree plot, Figure 7, was used to determine the dimensionality of the descriptor variables (the *x*-block). In general, factors with an eigenvalue of less than 1 may be considered unimportant. For the current data, Table 1, three or four factors may be required to represent the independent data. As a rule of thumb, approximately 3^P data points are required to make predictions with a linear model specified using P orthogonal variables with objects spread evenly through the variable space. Since there are 15 points in the data set, there is enough information to make predictions on two variables and some information about a third. The specified equation should therefore contain two, at most three, components.

A second, supervised factor analysis, based on the experimental free energy (the response variable) as well as the simulation data, was used to identify the variable sets that make uncorrelated contributions to the response variable ΔG. Those factors that are highly correlated with ΔG were identified, and the variables that make the largest contributions to those factors were then selected as important variables for regression analysis. The results revealed that ΔU_{vdw} and $\Delta SASA_{prot}$ were the most closely associated to ΔG. However, it has been established that up to three descriptor variables can be considered for prediction of ΔG. To determine the third, an ordinary least squares (OLS) analysis was carried out using these two variables with leave one out (LOO) cross-validation. This yielded a cross-validated regression coefficient (q^2) of 0.33. The remaining variables were then added in turn, and the procedure was repeated. A substantially improved q^2 of 0.56 was achieved when ΔU_{elec} was added. Hence ΔU_{vdw}, ΔU_{elec}, and $\Delta SASA_{prot}$ were used in the final stage of model specification.

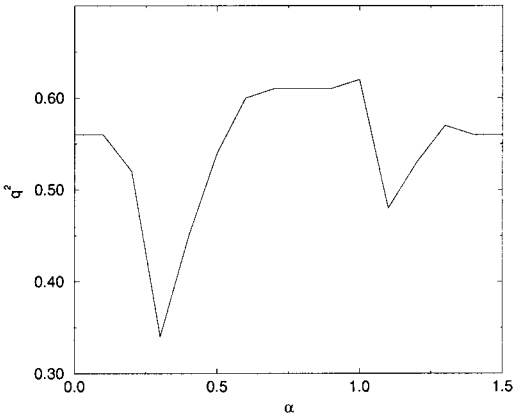


Figure 8. Plot of LOO cross-validated correlation coefficient (q^2) vs CR parameter α for model constructed from ΔU_{vdw}, ΔU_{elec}, and $\Delta SASA_{prot}$.

Model Specification Using Continuum Regression. A new generalized procedure, continuum regression (CR),[39−41] encompasses ordinary least squares (OLS) and a continuum of orthogonalized regression procedures including partial least squares (PLS) and principle components regression (PCR). The Portsmouth formulation of the CR[40] implemented using the PARAGON drug design software[42] uses a parameter, α, to determine the component construction criteria—α takes values between 0 and 1.5 where $\alpha = 0$ corresponds to OLS, $\alpha = 0.5$ is PLS, and $\alpha = 1$ corresponds to PCR. The CR procedure was used to vary the value of α between 0 and 1.5, with the leave one out (LOO) cross-validated regression coefficient (q^2) calculated as a function of α (Figure 8). For most values of α, two components were found to be significant. The most predictive three-variable model was identified on the basis of the q^2. The highest value of $q^2 = 0.62$ was obtained for $\alpha = 1.0$, corresponding to PCR.

The final step in the model specification process was to repeat the CR procedure on each subset of two of the three variables chosen above to check for more predic-

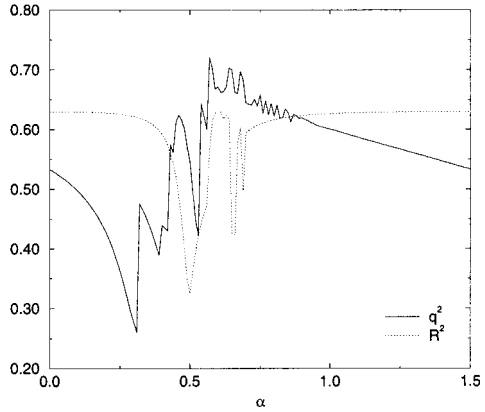


Figure 9. Plot of LOO cross-validated correlation coefficient (q^2) and correlation coefficient (R^2) vs CR parameter α for model constructed from ΔU_{vdw} and ΔU_{elec}.

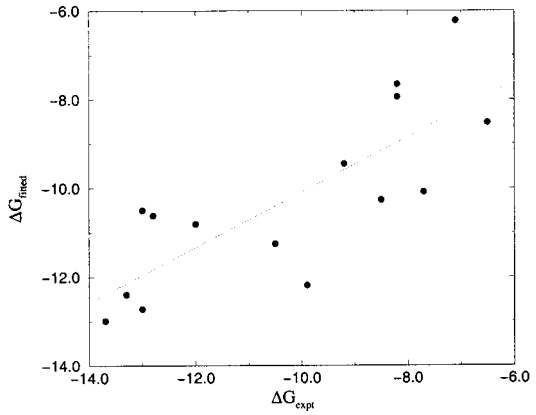


Figure 10. Plot of fitted vs experimental free energies of binding for the most predictive model.

tive models. Application to ΔU_{vdw} and $\Delta SASA_{prot}$, and to ΔU_{elec} and $\Delta SASA_{prot}$, produced no improvement. For many values of α no significant model was found, and where one was identified the best, q^2 was 0.34. However, when the two-variable model comprising ΔU_{vdw} and ΔU_{elec} was specified, the plot of q^2 against α shown in Figure 9 was obtained; the corresponding values of R^2 are also plotted. This showed an improved q^2 in excess of 0.7 at $\alpha = 0.58$. It was also noted that within this α domain, q^2 was sensitive to changes in α. CR was therefore repeated with α incremented in 0.001 steps for the range $0.57 < \alpha < 0.63$. This yielded a further improvement of $q^2 = 0.737$ at $\alpha = 0.584$ for a one-component model. Fitting this model to all 15 points yielded an rms deviation from experiment of 1.61 kcal mol^{-1} and a correlation coefficient (R^2) of 0.622. It is encouraging to note that the CR model ($\alpha = 0.584$) has an R^2 equal to that obtained using OLS. The increased value of q^2 (0.74) confirms that $\alpha = 0.584$ yields a more generalized model than that obtained by OLS ($\alpha = 0$; $q^2 = 0.533$) and should therefore be more predictive. This degree of prediction results from the construction of a low dimensional model comprising a single CR component that summaries all of the relevant information in the descriptor variables required to explain the variation in ΔG.

The physical meaning of the orthogonalized regression model is difficult to interpret. The results were therefore transformed back to the original data space resulting in the model described by eq 8.

$$\Delta G = 2.603 + 0.472\Delta U_{vdw} + 0.122\Delta U_{elec}$$
$$\pm 0.116 \qquad \pm 0.063 \qquad (8)$$

Standard errors have been estimated using bootstrapping based on 1000 runs. A plot of the calculated free energies against experimental values is shown (Figure 10). The plot indicates that the derived equation is unbiased and predicts the order of the binding constants more or less correctly. The rms deviation from experiment is typical of the type of agreement other workers have observed.

Inhibitor 14 was the only charged inhibitor, and it was therefore considered that its inclusion may be biasing the statistics. The statistical procedure was therefore repeated on the data set containing only the neutral inhibitors. The removal of inhibitor 14 had only a marginal influence on the coefficients in the final equation and the cross-validated statistics. The value of q^2 dropped to 0.594, and the values of the constant term and van der Waals and electrostatic coefficients became 1.756, 0.457, and 0.114, within the errors on the coefficients for the full model. Thus the inclusion of inhibitor **14** is not having a detrimental effect on the fitting, as would be expected in light of the precautions taken to avoid having to make ill-defined Born-type corrections.

There is no evidence for this system and force field that there is a factor of 0.5 relating ΔU_{elec} to ΔG. It should, however, be noted that the values of α and β derived in this work are similar to those obtained by Jones-Hertzog and Jorgensen,[15] who also used a very similar simulation protocol to that adopted here. The use of approximations such as spherical systems and nonbonded cutoffs can affect both the accuracy and precision of the energies obtained from a simulation. The current work has used a charge neutralization protocol and has made the sphere size and cutoff lengths as large as possible in an effort to minimize such effects. However, it is noted that these approximations will result in the simulation protocol having a direct effect on the form of the LIE equation and its coefficients. Hence a true test of coefficient transferability can only be achieved by simulating a selection of systems using the same protocol.[19]

Conclusions

A series of inhibitors of the protein neuraminidase have been studied by Monte Carlo simulations and application of the LIE method for the calculation of free energies of binding. Three classes of inhibitors have been studied, those based on DANA, and the amino and guanadino forms of the amide derivatives of that molecule. Studies have shown that the stronger binding of guanadino molecules with respect to their amino analogues is likely to be the consequence of the expulsion of tightly bound water from the active site. Binding affinities are also observed to increase as the size of the amide substituents increases. Since no specific interactions can be identified as responsible for this trend, it seems likely to be due to hydrophobic effects. The simulation results have also been consistent with previous observations that the residue GLU^{276} adopts a different conformation for the amide type inhibitors than for DANA and is conformationally mobile.

A difficult test case for the LIE method has been studied, using a diverse set of 15 inhibitors. Although equations proposed by other workers have not proved to be predictive when applied to the current system, a careful statistical analysis has enabled the deduction of an equation which produces fitted and cross-validated R^2 values that indicate the method could be useful for rational drug design purposes. The evidence also suggests that under the OPLS force field and the simulation protocol used here the linear response result that the change in free energy and the change in electrostatic energy are related by a factor of a half is not transferable to protein systems. It is also noted that great care should be taken over the addition of terms to the LIE equation without evidence of the significance of their inclusion. Consequently it seems that the best way of determining which variables are appropriate for inclusion is by statistical fitting.

It also seems clear that the MLR method typically applied to analyze such data is not valid due to the intrinsic cross-correlations in the data. Such difficulties should be overcome by using a statistical procedure based on a set of orthogonal variables (components) spanning the data space. A fitting procedure based on the orthogonalized variables can be carried out and the optimum model mapped back to the original data space as required. The statistical analysis has also revealed that three or four variables are likely to be required to explain optimally the data. However, to obtain good estimates on three variables a larger data set is required containing at least 27 data points.

The results of the statistical analysis have indicated that despite the complexity of the system, terms accounting for the change in intramolecular energy and solvent accessible surface area do not make significant contributions to the change in free energy on binding. Moreover, in agreement with previous work, ΔU_{elec} and ΔU_{vdw} have been found to be the most important variables for making predictions of ΔG giving a q^2 of 0.74. This result was obtained without recourse to prior knowledge and is therefore based solely on the data. However, it should be noted that the use of appropriately constructed components, chosen using the continuum regression method, has provided a dramatic improvement in the predictive ability of the final equation over one determined by simple least-squares fitting.

It seems the LIE method can produce reasonable predictions of binding free energies. However, it also seems that there is no unique formula which can be universally applied to such calculations. Consequently, a typical application would probably involve the use of a training set of molecules to deduce the appropriate parameters followed by the application of this formula to the molecules of interest. Unfortunately, this means that the method may not be as efficient as was initially hoped, but as sampling methodology and computer speeds increase it will become increasingly practicable.

Acknowledgment. I.D.W. acknowledges the generous support of GlaxoWellcome and the BBSRC in the form of a studentship. J.W.E. is a Royal Society University Research Fellow. We also thank the BBSRC for support of the development of CR, the EPSRC for a grant of time on the Columbus facility, and the University of Southampton for computational resources.

Supporting Information Available: Description of the development of nonstandard force field parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kollman, P. Free-energy calculations − applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(2) Reynolds, C. A.; King, P. M.; Richards, W. G. Free-energy calculations in molecular biophysics. *Mol. Phys.* **1992**, *76*, 251−275.
(3) Beveridge, D. L.; DiCapua, F. M. Free-energy via molecular simulation − applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(4) Liu, H. Y.; Mark, A. E.; van Gunsteren, W. F. Estimating the relative free energy of different molecular states with respect to a single reference state. *J. Phys. Chem.* **1996**, *100*, 9485−9494.
(5) Smith, P. E.; van Gunsteren, W. F. Predictions of free-energy differences from a single simulation of the initial-state. *J. Chem. Phys.* **1994**, *100*, 577−585.
(6) Beutler, T. C.; van Gunsteren, W. F. Molecular-dynamics free-energy calculation in 4 dimensions. *J. Chem. Phys.* **1994**, *101*, 1417−1422.
(7) Radmer, R. J.; Kollman, P. A. The application of three approximate free energy calculations methods to structure based ligand design: trypsin and its complex with inhibitors. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 215−227.
(8) Levy, R. M.; Belhadj, M.; Kitchen, D. B. Gaussian fluctuation formula for electrostatic free-energy changes in solution. *J. Chem. Phys.* **1991**, *95*, 3627−3633.
(9) Pitera, J.; Kollman, P. Designing an optimum guest for a host using multimolecule free energy calculations: predicting the best ligand for rebek's tennis ball. *J. Am. Chem. Soc.* **1998**, *120*, 7557−7567.
(10) Åqvist, J.; Medina, C.; Samuelsson, J. E. New method for predicting binding-affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.
(11) Åqvist, J.; Hansson, T. On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.* **1996**, *100*, 9512−9521.
(12) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137−1144.
(13) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(14) Åqvist, J.; Mowbray, S. L. Sugar recognition by a glucose/galactose receptor − evaluation of binding energetics from molecular-dynamics simulations. *J. Biol. Chem.* **1995**, *270*, 9978−9981.
(15) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.* **1997**, *40*, 1539−1549.
(16) Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* **1996**, *9*, 567−571.
(17) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Åqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 protease inhibitors derived from mannitol: synthesis, inhibitory potencies, and computational predictions of binding affinities. *J. Med. Chem.* **1997**, *40*, 885−897.
(18) Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, *17*, 1587−1597.
(19) Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: a molecular dynamics- based interaction energy study. *Proteins: Struct. Funct. Genet.* **1999**, *34*, 69−81.
(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force-field for the simulation of proteins, nucleic- acids, and organic-molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(21) Henchman, R. H.; Essex, J. W. Free energies of hydration using restrained electrostatic potential derived (REPD) charges via free energy perturbations and linear response. *J. Comput. Chem.* **1999**, *20*, 499−510.
(22) Carlson, H. A.; Jorgensen, W. L. An extended linear-response method for determining free-energies of hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.

(23) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem*. **1997**, *10*, 563−576.
(24) Taylor, N. R.; von Itzstein, M. 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.* **1996**, *10*, 233−246.
(25) Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Cobley, K. N.; Taylor, H.; Whittington, A. R.; Scicinski, J.; Bethell, R. C.; Taylor, N.; Skarzynski, T.; Cleasby, A.; Singh, O.; Wonacott, A.; Varghese, J.; Colman, P. Novel inhibitors of influenza sialidases related to GG167 − structure−activity, crystallographic and molecular dynamic studies with 4H-pyran-2-carboxylic acid 6-carboxamides. *Bioog. Med. Chem. Lett.* **1996**, *6*, 2931−2936.
(26) Cheng, Y.; Prusoff, W. H. The relationship between inhibition constant K_i and the concentraton of inhibitor which causes 50% inhibition I_{50} of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(27) Bethell, R. Personal communication.
(28) Unpublished stucture obtained from GlaxoWellcome.
(29) von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational design of potent sialidase-based inhibitors of influenza-virus replication. *Nature* **1993**, *363*, 418−423.
(30) Jorgensen, W. L. Theoretical studies of medium effects on conformational equilibria. *J. Phys. Chem*. **1983**, *87*, 5304−5314.
(31) Jorgensen, W. L. MCPRO 1.4, Yale University, New Haven, CT, 1996.
(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(33) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. Monte Carlo simulations for proteins: binding affinities for trypsin-benzamidine complexes via free-energy perturbations. *J. Phys. Chem. B* **1997**, *101*, 9663−9669.
(34) Essex, J. W.; Jorgensen, W. L. An empirical boundary potential for water droplet simulations. *J. Comput. Chem*. **1995**, *16*, 951−972.
(35) Richmond, T. J. Solvent accessible surface-area and excluded volume in proteins − analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **1984**, *178*, 63−89.
(36) Ponder, J. W. TINKER, Version 3.6; Washington University, St. Louis, MO, 1998.
(37) Jorgensen, W. L.; Tirado-Rives, J. The OPLS potential functions for proteins − energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(38) Livingstone, D. *Data Analysis for Chemists*; Oxford University Press: Oxford, 1995.
(39) Stone, M.; Brooks, R. J. Continuum regression − cross-validated sequentially constructed prediction embracing ordinary least-squares, partial least-squares and principal components regression. *J. R. Stat. Soc. Ser. B*— *Methodol.* **1990**, *52*, 237−269.
(40) Malpass, J.; Salt, D.; Ford, M.; Wynn, E.; Livingstone, D. Continuum regression: A new algorithm for the prediction of biological activity. In *Methods and Principles in Medicinal Chemistry*, *3*, *Advanced computer assisted techniques in drug discovery*; VCH Publishers: Weinheim, 1995.
(41) Malpass, J.; Salt, D.; Ford, M. Continuum regression: optimised prediction of biological activity. *Pestic. Sci.* **1996**, *46*, 282−284.
(42) Ford, M.; Crichton, R.; Salt, D.; Livingstone, D. PARAGON Drug Design Software, V1.09.01; University of Portsmouth, UK, 1999.

JM990105G