Ligand Binding Affinities from MD Simulations

JOHAN ÅQVIST,* VICTOR B. LUZHKOV, AND BJØRN O. BRANDSDAL†

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden

Received September 11, 2001

ABSTRACT

Simplified free energy calculations based on force field energy estimates of ligand−receptor interactions and thermal conformational sampling have emerged as a useful tool in structure-based ligand design. Here we give an overview of the linear interaction energy (LIE) method for calculating ligand binding free energies from molecular dynamics simulations. A notable feature is that the binding energetics can be predicted by considering only the intermolecular interactions of the ligand in the associated and dissociated states. The approximations behind this approach are examined, and different parametrizations of the model are discussed. LIE-type methods appear particularly promising for computational "lead optimization". Recent applications to protein−protein interactions and ion channel blocking are also discussed.

Introduction

It may be argued that molecular recognition is what biochemistry is all about on the microscopic level. To characterize the structure and energetics of molecular complexes is therefore the key to understanding many biological functions. It is especially worth emphasizing that energetics often provides the most important and useful link between structure and function of biomolecular systems. To be able to predict the strength of noncovalent binding between molecules and the 3D structures of the corresponding complexes has thus been a longstanding goal in computational chemistry. Aside from the theoretical challenge this problem presents, activities in this field are also spurred by the potential impact on structure-based drug design. Significant progress has been made in computer-aided ligand design during the past decade, and methodologies based on force field calculations, such as molecular mechanics (MM), molecular dynamics (MD), and Monte Carlo (MC) simulations, have been important for many of these developments.[1−3] Another category of methods that has become increasingly useful for rapid ligand screening and docking utilizes empirical and knowledge-based scoring functions for binding affinity estimation.[4,5]

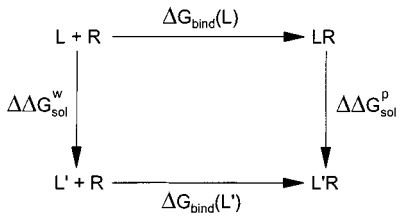

FIGURE 1. Thermodynamic cycle used in FEP calculations of the relative binding free energies of two ligands, L and L′, to a receptor molecule R. The absolute binding free energy of L′ can, in principle, be obtained by considering L as a dummy ligand with no intermolecular interactions.

The free energy perturbation (FEP) approach for calculating relative binding free energies between two ligands (L and L′) and a given receptor (R) employs the thermodynamic cycle of Figure 1.[2,6−8] Here, $\Delta\Delta G^{w}_{sol}$ and $\Delta\Delta G^{p}_{sol}$ denote the differences in ("solvation") free energy between L and L′ in water and when bound to the solvated receptor (protein) site, respectively, and the ΔG_{bind}'s are the corresponding binding energies:

$$\Delta G_{bind}(L') - \Delta G_{bind}(L) = \Delta\Delta G^{p}_{sol} - \Delta\Delta G^{w}_{sol} \qquad (1)$$

With the FEP method one calculates the free energies associated with the two unphysical paths L→L′(aq) and LR→L′R(aq) corresponding to a mutation of L into L′ in the free and bound states, respectively, and MD or MC simulations are used to collect ensemble averages along these paths. The paths are typically discretized into a number of points, each represented by a separate potential energy function constructed as a linear combination of the inital and final state potentials.[2,6−8]

While the FEP approach remains the most important technique for free energy calculations by MD or MC simulations. its drawbacks have been well recognized for some time.[8,9] They mainly pertain to the extensive conformational sampling that is usually required in order to obtain meaningful, convergent results from such simulations. Most of the computer time is also spent on uninteresting configurations that correspond to unphysical "mixtures" of L and L′. This makes the method computationally less attractive for applications where the objective is to estimate binding free energy differences between a large set of ligands, as is often encountered in inhibitor design. The fact that the "perturbations" or transformations involved in FEP cannot be too drastic is also a clear limitation. It restricts the diversity of ligands that can be treated and also the possibility of readily examining the effect of protein mutations on the affinity

Johan Åqvist received his M.Sc. in engineering physics from Uppsala University in 1983 and his Ph.D. in structural biology from the Swedish University of Agricultural Sciences in 1987. Following two years of postdoctoral research at the University of Southern California, he returned to Uppsala University, where he is now Professor of Theoretical Chemistry and a senior researcher of the Swedish Research Council.

Victor Luzhkov received his Ph.D. in biophysics in 1979 at the Moscow Physico-Technical Institute and earned the degree Doctor of Science in 1992 at the Institute of Chemical Physics of the Russian Academy of Sciences in Chernogolovka, where he was a senior scientist until 1999. He is currently a visiting researcher at Uppsala University.

Bjørn Brandsdal was born in 1973 in Harstad, Norway. He has been a graduate student in the Tromsø protein crystallography group of Arne Smalås, where he received his Ph.D. in the fall of 2001.

* Corresponding author. Phone: +46-18-471 4109. Fax: +46-18-536971. E-mail: aqvist@xray.bmc.uu.se.

† Permanent address: Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway.

10.1021/ar010014p CCC: $22.00
Published on Web 02/16/2002

for a given ligand or substrate, since such mutations often fall into the "too large perturbation" category.

Other types of applications of free energy calculations are, however, much better suited for the FEP approach. Typical cases where the method can be pushed to very high accuracy are those where only a few van der Waals parameters and charges are transformed from one set into another and where no large conformational changes are expected to occur. Such well-behaved problems may include, e.g., calculations of relative solvation free energies of simple ions[10] and organic compounds that differ by only a few substituents.[11] Also, the calculation of relative binding free energies of ligands to host molecules can, in favorable cases (small perturbations), become accurate enough for quantitative conclusions to be drawn. Some recent examples from our group in this respect are studies on ligand binding to cyclodextrins[12] and dihydrofolate reductase[13] and of the ion permeation mechanism in potassium channels.[14−16] The FEP method has also proven to be very useful in calculations on enzyme reaction mechanisms.[17,18] What is perhaps still not fully realized is that it is most often the creation and annihilation of atoms (i.e., Lennard-Jones potential terms) that cause problems in FEP rather than changes in atomic charges (electrostatic terms), even though the latter are more long-ranged. These issues have been discussed elsewhere,[8,9,19] and some improvements of the FEP approach have been devised to circumvent the problems.[20]

Several comprehensive reviews of the FEP method have been published during the past decade,[2,6,7] and here we will instead focus on simplified free energy calculation methods, which has become a field of considerable activity.[1−5,9,21−23] Recent reviews by Kollman and coworkers[1,22] and by Simonson et al. in this issue[23] provide a good overview of such methodologies, and since the emphasis here will be mainly on our own results, we also recommend those more general accounts to the interested reader.

The Linear Interaction Energy (LIE) Method

The problems encountered when trying to use FEP calculations for estimating binding affinities of series of diverse ligands prompted us to examine whether it would be possible to extract any useful information on the binding energetics from simulations of only the physically relevant states (free and bound) of the ligand.[9] The idea was to consider the absolute binding free energy of a ligand (l) as the change in free energy when it is transferred from aqueous solution (free state) to its solvated receptor binding site (bound state), that is,

$$\Delta G_{\mathrm{bind}}(\mathrm{l}) = \Delta G^{\mathrm{p}}_{\mathrm{sol}}(\mathrm{l}) - \Delta G^{\mathrm{w}}_{\mathrm{sol}}(\mathrm{l}) \qquad (2)$$

where the superscripts "p" and "w" again denote protein (receptor) and water, respectively. The "solvation" energy of the ligand in a given environment, $\Delta G^{i}_{\mathrm{sol}}(\mathrm{l})$, in turn reflects the process of transferring the molecule from the gas phase to this environment. Such a process can at least formally be considered as consisting of two separate steps: (1) creating the molecular van der Waals cavity in the given environment and (2) turning on the electrostatic interactions between the molecule and its surroundings.[9]

The original version of the LIE method employed the linear response approximation to estimate the electrostatic part of the solvation/binding free energies. The linear response result for this component of the solvation energies, $\Delta G^{i}_{\mathrm{el}}$ (where i = p or w), can be written as[9,24]

$$\Delta G^{\mathrm{i}}_{\mathrm{el}} = \frac{1}{2}\{\langle V^{\mathrm{el}}_{\mathrm{l-s}}\rangle_{\mathrm{on}} + \langle V^{\mathrm{el}}_{\mathrm{l-s}}\rangle_{\mathrm{off}}\} \qquad (3)$$

where the two averages are sampled with the electrostatic interactions between the ligand and the surrounding (l−s) turned on and off, respectively. One of the simplifying features of the LIE method is that the term $\langle V^{\mathrm{l-s}}_{\mathrm{el}}\rangle_{\mathrm{off}}$ is neglected, which has been found to be a good approximation in water[25] (see ref 3 for a further discussion).

Nonpolar contributions to the binding affinity, e.g., hydrophobic effects and van der Waals interactions, may appear to be less straightforward to quantify. We decided to try the simple idea of measuring the nonelectrostatic part of the interaction between the ligand and its surrounding environment in the associated and dissociated states, and then just scale these energies by an empirically derived coefficient. These energies are typically given by a Lennard-Jones potential. The basic idea was *not* that hydrophobic effects somehow reside in these energy terms, but rather based on the following observations. Solvation free energies for typical nonpolar compounds are experimentally found to scale linearly with solute size measures such as accessible surface area.[26] We also found from MD simulations that the average van der Waals (Lennard-Jones) interaction energies scaled approximately linearly with solute size both in polar and nonpolar solvents.[9] Combining these two observations would thus suggest that it might be possible to use average ligand van der Waals energies for estimating nonpolar binding contributions, simply because they are correlated with the same variables as "hydrophobic free energies".

The above considerations thus led us to explore an approximate equation for the binding free energy of the following general type:

$$\Delta G_{\mathrm{bind}} = \alpha\Delta\langle V^{\mathrm{vdw}}_{\mathrm{l-s}}\rangle + \beta\Delta\langle V^{\mathrm{el}}_{\mathrm{l-s}}\rangle + \gamma \qquad (4)$$

where $\langle\ \rangle$ denotes MD or MC averages of the nonbonded van der Waals (vdw) and electrostatic (el) interactions between the ligand and its surrounding environment (l−s), i.e., either the solvated receptor binding site (bound state) or just solvent (free state). The Δ in eq 4 denotes the difference between such averages in the bound and free states. In other words, two simulations are required: one with the ligand free in solution and one with it bound to solvated receptor. The parameters of this equation are the weight coefficients α and β for the nonpolar and polar binding energy contributions, respectively, and possibly an additional constant γ.[9,27] As noted above, the linear response approximation predicts a value of $\beta = {}^1/_2$ for the electrostatic coefficient.

Our initial studies using the LIE model involved calculations on several different receptor−ligand systems, namely, inhibitor complexes with endothiapepsin,[9] HIV-1 protease,[28,29] trypsin,[19] and sugar recognition by a bacterial glucose/galactose binding protein.[30] It came as somewhat of a surprise that the original parametrization of eq 4 with $\alpha = 0.16$, $\beta = {}^1/_2$ and $\gamma = 0$ for a small calibration set of four endothiapepsin complexes[9] also was reasonably predictive for the other systems mentioned above. This model contained only one free parameter (α), since β was set to the linear response value ($^1/_2$) and the constant γ was set to zero. The average unsigned error in the absolute binding free energies for the 18 different complexes considered in these studies was 1.2 kcal/mol,[27] a result that we found rather promising.

However, several questions still remained unanswered, particularly with regard to the general applicability of the simplified binding energy equation (eq 4), the validity of the linear response approximation, possible dependencies on the chemical nature of ligands and protein binding sites, force fields, and computational protocols.

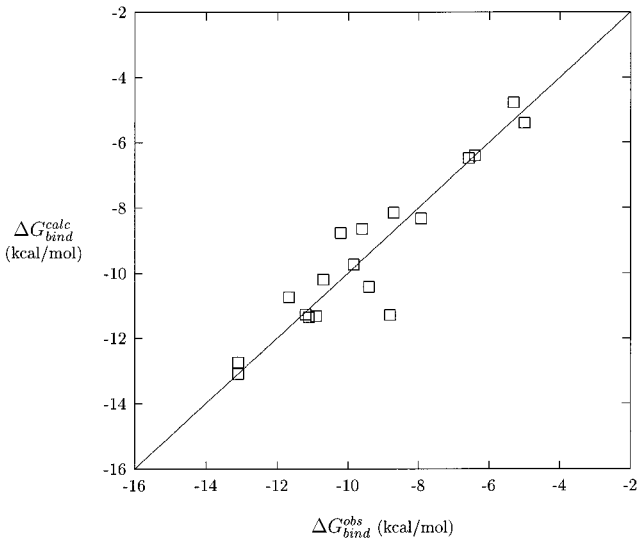

FIGURE 2. Calculated vs observed free energies of binding for the 18 receptor−ligand complexes used to derive the LIE model of ref 27. This model uses the β values 0.5 for ionic ligands and 0.43, 0.37, and 0.33 for neutral ligands with zero, one, and two or more OH groups, respectively.

Relaxing the Linear Response Assumption

While the linear response approximation (eq 3) appeared to be a useful route to more rapid calculations of electrostatic free energies,[9,24,31] it was not clear how well the approximation was actually obeyed at the microscopic level. That is, although it is the basis of many continuum treatments, its validity in real molecular systems was not fully established. To examine this issue in more detail, Åqvist and Hansson[25] used FEP simulations of a number of different solute/solvent systems and separately calculated the left- and right-hand sides of eq 3. It was then found that the linear response approximation holds well for ionic solutes while significant deviations occur in some cases, such as for neutral dipolar solutes containing hydroxyl groups that can interact by specific hydrogen bonding with the solvent. Such deviations from the linear response behavior are reflected by the coefficient β not being exactly $^1/_2$ in eq 4, but typically assuming values between 0.3 and 0.5, depending on the chemical composition of the solute.

The β parameters determined for different solutes in water[25] were subsequently used to derive a refined version of the LIE model where ligands were assigned one out of four possible β values depending on their chemical structure.[27] This model was parametrized for the 18 complexes mentioned above (Figure 2) with respect to the α coefficient and yielded an optimal value of $\alpha = 0.18$, with a mean unsigned error of only 0.6 kcal/mol for the absolute binding free energies.[27] Other notable results from this recalibration of the method were that inclusion of the additional constant γ did not improve the model but yielded $\gamma = 0$, and that attempting to use different β values for the ligand in the bound and free states resulted in very similar values of β in the two cases. The latter point is of particular interest since it strongly suggests that the basic characteristics of the electrostatic response are similar inside a solvated protein and in water.

This revised LIE model[27] has subsequently been employed in studies of dihydrofolate reductase (DHFR)[32−34] and human thrombin inhibitors,[35] as well as a number of complexes with ligand recognition and transport proteins, namely arabinose, lysine, fatty acid, and retinol binding protein.[3] The work on DHFR inhibitor binding involved calculations both on analogues of the classical antifolate methotrexate[32] and on newly designed lipophilic ester soft drugs against the *Pneumocystis carinii* enzyme.[33,34] An essential aspect of these studies was to examine not only the ranking of different inhibitors but also the selectivity of a given inhibitor for different DHFR enzymes. Hence, in ref 32 the effects of point mutations of the human enzyme on methotrexate affinity were addressed, while the calculations on the nonclassical ester inhibitors focused on the selectivity between the human and the *P. carinii* enzyme.[33,34] The pneumonia caused by this fungus is the major cause of death in patients with AIDS, and DHFR is in this case a prime target for pharmaceutical therapy. The soft drug concept, where the inhibitors are designed to undergo a fast metabolism after having exerted their effect at the site of action, holds considerable promise for local treatment (e.g., inhalation) without causing systemic side effects. The LIE calculations on different DHFR complexes showed that the method can, indeed, be useful also for prediction of inhibitor selectivity (Figure 3).

The ability of the LIE method to rank a series of thrombin inhibitors has also recently been examined.[35] That work again demonstrated the capability of this approach in predicting the relative affinities of chemically very different ligands, as well as the possibility of estimating stereoselectivity. However, in the case of thrombin it was found that eq 4 does require a constant term ($\gamma =$

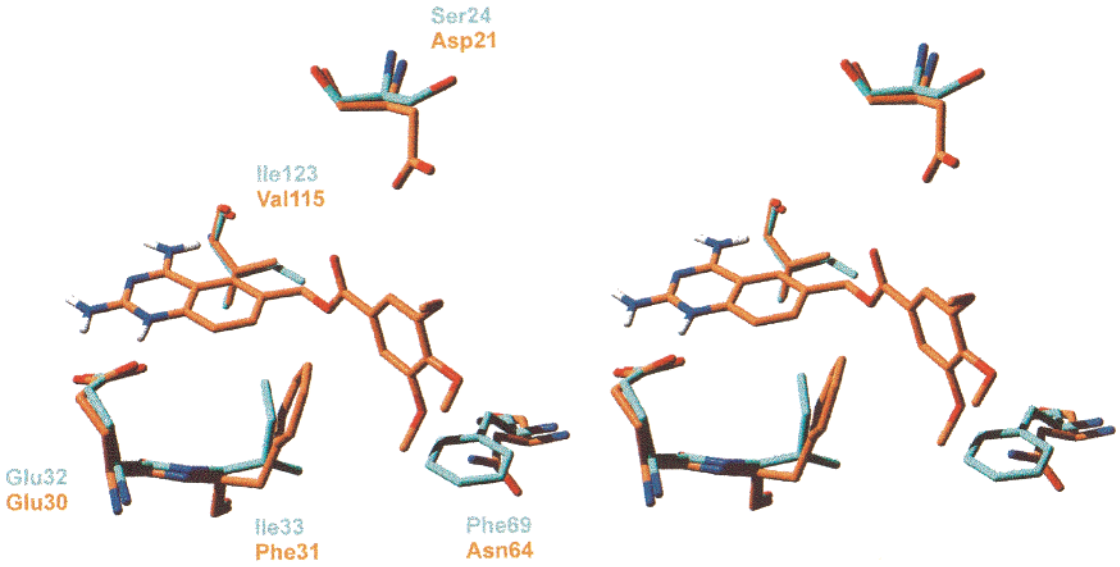

FIGURE 3. Stereoview showing amino acid residues that differ between the active sites of human (blue) and *P. carinii* DHFR (orange). The glutamate interacting with the 2,4-diaminoquinazoline ring of the inhibitor is also shown. The selectivities of three such newly synthesized inhibitors were successfully predicted.[33]

−2.9 kcal/mol) in order to reproduce the absolute binding free energies. The revised LIE model discussed in the previous section with such an additional constant gives a mean unsigned error of 0.6 kcal/mol for the data set of eight thrombin inhibitors. Interestingly, it was also found that a free parametrization of all three coefficients in eq 4 yielded essentially the same values of α and β as before. In our view, this suggests that the possible system dependence of the parametrization of eq 4 might be reducible to different constant terms (γ) for different types of receptor sites.

The above conclusion is also supported by simulations of complexes of the previously mentioned recognition and transport proteins.[3] That is, ligand binding to the polar binding sites of arabinose, lysine, and muscular fatty acid binding protein appears well described by the revised LIE model with no additional constant. On the other hand, the absolute binding free energies of four examined complexes with the entirely hydrophobic cavity of retinol binding protein require a constant term of about −7 kcal/mol in eq 4 in order to reproduce the experimental data.[3] These results, as well as those for the thrombin complexes, seem to indicate that the hydrophobicity of the receptor site may be a source of system dependency that can be alleviated by including a specific constant γ. A similar idea has also been put forward by Wang et al.,[36] who suggested an interesting method based on desolvated nonpolar surface areas in the complex as a means to distinguish between different types of binding sites.

The possibility of introducing a constant term γ in eq 4 was suggested already in the original description of the LIE method,[9] where it was also noted that such a term is in general needed if the approach is to be used for estimating solvation free energies. Jorgensen and co-workers have instead used a third term in eq 4 containing the difference in solvent-accessible surface area (SASA) of the ligand, scaled by an empirical coefficient.[37−39] They have obtained excellent results with this approach, but it does seem to involve an unnecessary computational operation since the SASAs have been found to be strongly correlated with the intermolecular van der Waals energy term.[9,27,37,39] That is, the use of a SASA term appears to be essentially equivalent to the introduction of a constant γ. The work of Jorgensen and co-workers also employs MC simulations where the protein backbone is kept rigid.[38,39] This may offer a substantial improvement in speed, since large-scale protein motions are suppressed, although that might sometimes be a risky assumption.

Protein−Protein Interactions

The formation of protein−protein complexes is expected to obey the same physical principles that underlie protein−ligand interactions, but in the former case a more delicate balance between entropic and enthalpic contributions might be anticipated. The evaluation of absolute binding free energies of protein−protein complexes is thus an extremely difficult task to address with computer simulation approaches. Calculation of such energies with the LIE approach would require that one obtains convergent values of the entire interaction energies, and these quantities can be in the order of several thousand kilocalories per mole. This would require extremely long simulations to get stable averages and result in an inefficient approach with corresponding high uncertainties in the energetics. As far as FEP or "potential of mean force" approaches are concerned, calculation of absolute association energies for large complexes is completely beyond the scope of such methodologies.

Protein−protein interfaces are, however, frequently composed of a cluster of "hot spot" residues at the center of the interface surrounded by energetically less important residues. The interface between the serine proteases and their canonical protein inhibitors is composed of such

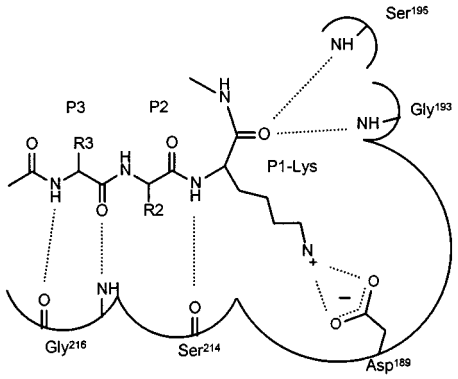

FIGURE 4. Schematic drawing of the P1−S1 interactions in the complex between bovine trypsin and BPTI.

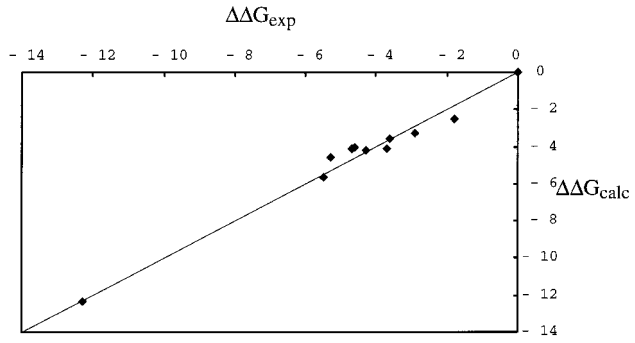

FIGURE 5. Scatter diagram of calculated vs observed binding free energies (in kilocalories per mole) of trypsin−BPTI complexes relative to P1−Gly (excluding the P1−Asp and P1−Glu variants).[43] The mean unsigned error in the calculated binding free energies is 0.38 kcal/mol.

"hot spot" residues (Figure 4), and the P1 residue has been found to be responsible for almost 70% of the interaction free energy in the binding of the bovine pancreatic trypsin inhibitor (BPTI) to trypsin.[40] Instead of trying to predict the absolute binding free energies, one can try to calculate the relative effect of different amino acids at specific positions on the association energy. This is not such a serious drawback, since protein−protein affinities are often more useful to analyze in relative terms, i.e., in terms of point mutations. The crystal structure analysis of 10 P1 variants of BPTI in complex with trypsin showed that the secondary interactions are virtually unchanged, regardless of amino acids at the P1 position.[41] The P1−Gly variant does not have a side chain that enters the S1 site, and the K_a value for this variant thus corresponds to the free energy from secondary interaction sites. In earlier FEP calculations,[42] the P1−Gly variant was successfully used as a reference state when other P1 variants (Ala and Met) were mutated to Gly to obtain the P1−S1 interaction free energy. These calculations also showed that, in order to obtain quantitative results, the water-mediated hydrogen-bonding network at the P1−S1 interface had to be correctly modeled.

Because of the variation in the chemical composition and structure of the natural amino acids, finding a general FEP strategy for point mutations is very difficult. Instead, we tried to apply the LIE approach to investigate the effect of mutations at the P1 position of BPTI on the association energy when bound to trypsin.[43] The idea was to treat the P1 residue as the "ligand" within the LIE framework while the rest of the inhibitor was considered as part of the surroundings. That is, we tried to use the method to obtain the relative affinities of BPTI variants differing only in the P1 position. This strategy actually turns out to be very useful, since the ligand−surroundings interaction energies converge quite rapidly in the MD simulations.

Out of the 20 possible different trypsin−BPTI complexes differing only at the P1 position, 13 were selected such that most of the binding range was covered, including those with the highest and lowest association constants. The LIE method was found to reproduce the experimental association energies in an impressive manner, and a correlation coefficient of 0.99 was obtained (Figure 5), excluding the P1−Asp and P1−Glu that are associated with uncertainties regarding their protonation and possible counterions. It is, however, clear from the calculations that these two P1 variants do not bind in their charged form.[43] A subsequent LIE study of cold-active trypsin from Atlantic salmon revealed that its enhanced binding affinities[44] for positively charged ligands are entirely caused by electrostatic effects.[45] Earlier continuum calculations had also pointed toward the role of different electrostatic surface potentials for cold adaptation of trypsins.[46]

The trypsin−BPTI calculations were carried out with the Amber95 force field,[47] using the LIE model of ref 27, and it is noteworthy that excellent results were obtained without any reparametrization of the method for this force field. Wang et al.[36] have also demonstrated that the present parametrization works well with the Amber95 potential on the trypsin−benzamidine complex.

Combining LIE with Automated Docking: K^+ Channel Blocking by Tetraethylammonium Ion

Binding affinity prediction methods of the LIE type are at present still too slow for use in virtual screening or docking applications. That is, if the affinities of a very large number of (chemically or structurally) different complexes have to be evaluated, it is critical that the "scoring" method is very rapid. For such applications, empirical scoring functions[4,5] are currently the only feasible strategy. However, once a limited set of putative complexes has been obtained, for instance by a docking algorithm, it may be a useful idea to try to refine these by the LIE approach. This type of hierarchical procedure has been adopted by us both in studies of DHFR inhibitors[13,33,34] and, more recently, in a first computational investigation of the binding of blockers to K^+ channels.[48]

The crystallographic structure of the KcsA potassium channel[49] was used together with results from earlier work on the permeation mechanism[14,15] to examine the binding of tetraethylammonium (TEA) to KcsA. Experimental investigations[50] have made it possible to identify two major binding regions for quartenary ammonium ions near the extracellular and intracellular entrances to the K^+ channel pore (Figure 6). An automated docking ap-

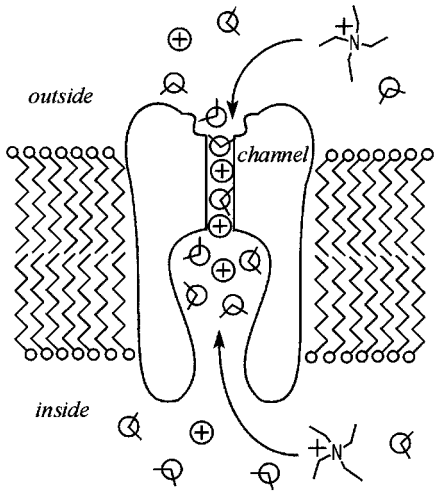

FIGURE 6. Schematic picture of K^+ channel blocking by external and/or internal TEA.

proach (AutoDock 3.0)[51] was first employed to generate a set of docked complexes that were grouped into clusters. This was done both for external and internal binding modes and for different K^+ ion loading states of the four-site selectivity filter[14−16] of the KcsA channel. The highest ranking, low-energy complexes predicted by the AutoDock scoring function for different binding modes, and K^+ loading states were then selected for further refinement by the LIE method, now also allowing flexibility of the protein and including explicit water solvation. No specific reparametrization of eq 4 was attempted in this case, either.

The automated docking yielded several clusters of TEA^+−KcsA complexes both at the external and internal entrances to the selectivity filter. In cases where docking to the bare protein was attempted, with no ions or water molecules occupying the filter, the most stable complexes were predicted to be those with partial insertion of one of the TEA ethyl groups into the selectivity filter (Figure 7). When TEA was docked with the channel filter fully occupied, binding was predicted (by AutoDock) for TEA in the quasi-planar conformation at the inner and outer pore entrances.

A number of these complexes were then further refined by LIE calculations for different loading states of the channel filter. The most stable complexes at the entrances to the selectivity filter were predicted to have dissociation constants in the millimolar region, in agreement with experimental data.[50] The subsequently solved crystal structure of KcsA with tetrabutylantimony (a heavy-atom analogue of tetrabutylammonium, TBA)[52] indicates that TBA binds in the internal cavity on the pore axis with the central nitrogen atom at a distance of ∼4.9 Å from the fourth (innermost) binding site of the selectivity filter. The position of the central antimony atom in the blocker is found to be in exactly the same position as that predicted by the calculations for the nitrogen in TEA. Furthermore, AutoDock/MD simulations of internal TBA binding also reproduce the tetrabutylantimony crystal structure regarding the position and conformation of the blocker.

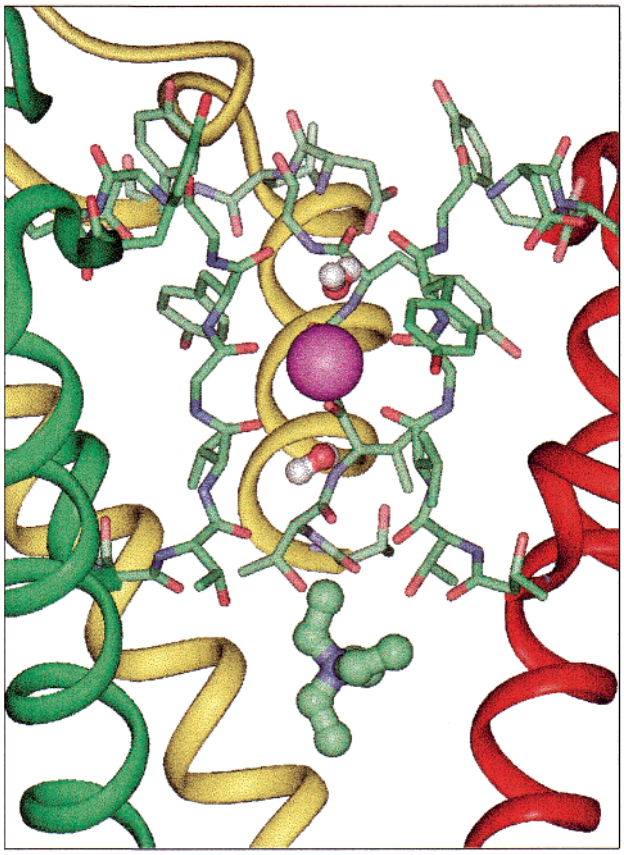

FIGURE 7. Side view of the selectivity filter region for the internal docked structure of TEA after 300 ps of MD simulation.[48] One of the channel monomers is omitted for clarity. A K^+ ion in the filter is also shown (solid sphere).

Experimental data show that the binding of TEA at the extracellular side of K^+ channels to a large extent depends on a ring of aromatic residues near the entrance to the pore. In KcsA, mutation of the corresponding tyrosine residues to valine (Y82V) leads to a reduced TEA blocking effect.[50] The influence of this mutation on the externally docked complexes was also examined by LIE calculations, and a loss of TEA affinity by around a factor of 30 was obtained, again in good agreement with experiments.[50] It was further concluded that the stabilizing effect of aromatic (Tyr of Phe) side chains on TEA at this position is not due to cation−π interactions, but rather to stabilizing hydrophobic as well as dipolar interactions.

Conclusions and Outlook

We have attempted here to give an overview of ligand−receptor binding affinity calculations using the linear interaction energy approach. The method was developed as an alternative to more time-consuming free energy perturbation calculations, in particular for predicting affinities of sets of ligands that are too diverse to fall into the "small perturbation" category required by the FEP method. The LIE approach was originally based on the linear response assumption for electrostatics together with an empirical scaling of nonpolar interaction energies intended to capture nonpolar or hydrophobic binding contributions.

A number of research groups have reported LIE calculations on various systems using different programs, force fields, and computational procedures, and the resulting optimizations of eq 4 can accordingly vary considerably.[27,35−39,53−56] Since it is of fundamental interest to try to understand the physical basis of the parameter values in eq 4, it is important to realize that a significant part of the reported spread in parameter values probably stems from different computational procedures rather than from intrinsic properties of the actual molecular systems. One of the critical technical issues in this type of calculations seems to be the treatment of electrostatic interactions, at least for charged ligands. The implementation of boundary conditions, cutoffs, system neutrality, etc. is thus of considerable importance, as well as ensuring compatibility of the simulations of the bound and free states with respect to electrostatic solvation energies.[3,19]

Regardless of whether one acribes significance to the nonempirical linear response considerations behind the electrostatic part of the calculated binding free energy, or whether one simply regards the parameters of the LIE equation (eq 4) as freely optimizable, the method reveals some rather unexpected features. That is, it came as somewhat of a surprise that (i) binding free energies could be so reasonably predicted by just considering the *intermolecular* interactions of the ligand and (ii) absolute affinities, and not only relative ones, could be reasonably well predicted from MD or MC simulations, something that is not really within the scope of FEP calculations. Furthermore, the idea that (iii) the intermolecular van der Waals energies could be correlated with hydrophobic binding contributions turned out to be quite useful. An explanation for this in terms of atom number densities of the different microenvironments in the bound and free states has been given.[27] We have also reiterated here that the introduction of a surface area term in eq 4 is unnecessary since it is basically equivalent to using the constant γ in the equation. The fact that only intermolecular energies are needed for the binding estimate has sometimes been interpreted in such a way that intramolecular relaxation/strain, entropy, receptor desolvation, etc. are neglected. We have argued elsewhere[3,27] that this is not really the case, but that these effects are rather embedded in the linear response approximation and the hydrophobic binding estimate. As far as electrostatics is concerned, we also find it noteworthy that optimization of eq 4 with different β's for the bound and free states yields very similar values of this coefficient in the two environments. This seems to indicate that the fundamental electrostatic response properties of proteins are not very different from those of aqueous solvent.

While binding in a variety of different receptor−ligand systems appears to be well modeled by our LIE parameters given in ref 27, there seem to be exceptions that are not due to different force fields or simulation setups. A notable case here is the more or less entirely hydrophobic binding of retinoids to RBP,[3] but also the reported calculations on P450cam[53] and avidin[36] seem to point in the same direction. To summarize, these results basically suggest that binding which is dominated by hydrophobic interactions requires a different and significantly higher value of the nonpolar scaling factor α, or the addition of a constant γ. It therefore seems worthwhile to try to understand whether such a scaling factor or constant behaves in a systematic way and could be predicted without using simulations on a training set.

The possibility of extending the applicability of the LIE approach to study protein−protein and protein−peptide interactions has also been illustrated by recent simulations of trypsin−BPTI complexes.[43,45] It is our feeling that treating specific residues as "ligands" in LIE calculations could provide a useful way of quantitatively analyzing the energetics and specificity of more complex protein and peptide recognition processes. Another promising line of development seems to be the combination of LIE calculations with automated docking and empirical scoring methods in a hierarchical approach.[33,34,48,57] This may be an efficient way of bridging the gap between virtual screening and docking with fast, but less accurate, scoring functions and "lead optimization" that requires a more detailed treatment. By selecting hits from the former type of computations to further refinement by fully microscopic LIE calculations (including explicit solvent, etc.), the overall computational effort can be more efficiently distributed.

The authors wish to sincerely thank the co-workers who have contributed to this work and whose names are given in the references list. Most of the reported calculations have been carried out using the program Q (that is freely available from the authors).[58] J.Å. gratefully acknowledges support from the Swedish Research Council. B.O.B. acknowledges financial support from NorFa.

References

(1) Kollman, P. Recent Advances in Structure-Based Ligand Design Using Molecular Dynamics and Monte Carlo Methods. *Pharm. Res.* **1998**, *15*, 368−370.
(2) Lamb, M. L.; Jorgensen, W. L. Computational Approaches to Molecular Recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449−457.
(3) Åqvist, J.; Marelius, J. The Linear Interaction Energy Method for Predicting Ligand Binding Affinities. *Combin. Chem. High Throughput Screen.*, **2001**, *4*, 613−626.
(4) Oprea, T. I.; Marshall, G. R. Receptor-Based Prediction of Binding Affinities. *Perspect. Drug Discov. Des.* **1998**, *9*, 35−61.
(5) Böhm. H.-J.; Stahl, M. Rapid Empirical Scoring Functions in Virtual Screening Applications. *Med. Chem. Res.* **1999**, *9*, 445−462.
(6) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.
(7) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(8) van Gunsteren, W. F.; Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992−1023.
(9) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.
(10) Åqvist, J. Ion−Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(11) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(12) Luzhkov, V. B.; Åqvist, J. Free Energy Perturbation Calculations of Binding and Transition-State Energies: Hydrolysis of Phenyl Esters by β-Cyclodextrin. *Chem. Phys. Lett.* **1999**, *302*, 267−272.

(13) Graffner-Nordberg, M.; Marelius, J.; Ohlsson, S.; Persson, Å.; Swedberg, G.; Andersson, P.; Andersson, S. E.; Åqvist, J.; Hallberg, A. Computational Prediction of Binding Affinities to Dihydrofolate Reductase: Synthesis and Biological Evaluation of Methotrexate Analogues. *J. Med. Chem.* **2000**, *43*, 3852−3861.
(14) Åqvist, J.; Luzhkov, V. Ion Permeation Mechanism of the Potassium Channel. *Nature* **2000**, *404*, 881−884.
(15) Luzhkov, V. B.; Åqvist, J. A Computational Study of Ion Binding and Protonation States in the KcsA Potassium Channel. *Biochim. Biophys. Acta* **2000**, *1481*, 360−370.
(16) Luzhkov, V. B.; Åqvist, J. K^+/Na^+ Selectivity of the KcsA Potassium Channel from Microscopic Free Energy Perturbation Calculations. *Biochim. Biophys. Acta* **2001**, *1548*, 194−202.
(17) Åqvist, J.; Warshel, A. Simulations of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches. *Chem. Rev.* **1993**, *93*, 2523−2544.
(18) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. Elucidating the Nature of Enzyme Catalysis Utilizing a New Twist on an Old Methodology: Quantum Mechanical-Free Energy Calculations on Chemical Reactions in Enzymes and in Aqueous Solution. *Acc. Chem. Res.* **2001**, *34*, 72−79.
(19) 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.
(20) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(21) Guo, Z.; Brooks, C. L., III; Kong, X. Efficient and Flexible Algorithm for Free Energy Calculations Using the λ-Dynamics Approach. *J. Phys. Chem.* **1998**, *102*, 2032−2036.
(22) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein−Ligand, Protein−Protein and Protein-Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.
(23) Simonson, T.; Archontis, G.; Karplus, M. Free Energy Simulations Come of Age: Protein−Ligand Recognition. *Acc. Chem. Res.* **2002**, *35*, 430−437.
(24) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. Calculations of Antibody Antigen Interactions: Microscopic and Semi-Microscopic Evaluation of the Free Energies of Binding of Phosphorylcholine Analogs to MCPC603. *Protein Eng.* **1992**, *5*, 215−228.
(25) Åqvist, J.; Hansson, T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* **1996**, *100*, 9512−9521.
(26) Blokzijl, W.; Engberts, J. B. F. N. Hydrophobic Effects: Opinions and Facts. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1545−1579.
(27) Hansson, T.; Marelius, J.; Åqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(28) Hansson, T.; Åqvist, J. Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137−1144.
(29) Hultén, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Åqvist, J.; Classon, B.; Danielsson, U. H.; Karlén, A.; Kvarnström, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived from Mannitol. Synthesis, Inhibitory Potency and Computational Prediction of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885−897.
(30) 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.
(31) Del Buono, G. S.; Figueiridо, F. E.; Levy, R. M. Intrinsic pK_as of Ionizable Residues in Proteins: An Explicit Solvent Calculation for Lysozyme. *Proteins* **1994**, *20*, 85−97.
(32) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Åqvist, J. Computation of Affinity and Selectivity: Binding of 2,4-Diaminopteridine and 2,4-Diaminoquinazoline Inhibitors to Dihydrofolate Reductases. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 119−131.
(33) Graffner-Nordberg, M.; Kolmodin, K.; Åqvist, J.; Queener, S. F.; Hallberg, A. Design, Synthesis, Computational Prediction and Biological Evaluation of Ester Soft Drugs as Inhibitors of Dihydrofolate Reductase from *Pneumocystis carinii*. *J. Med. Chem.* **2001**, *44*, 2391−2402.
(34) Graffner-Nordberg, M.; Kolmodin, K.; Åqvist, J.; Queener, S. F.; Hallberg, A. Design, Synthesis and Computational Activity Prediction of Ester Soft Drugs as Inhibitors of Dihydrofolate Reductase from *Pneumocystis carinii*. Submitted.
(35) Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Nordén, B.; Åqvist, J. Computational Modelling of Inhibitor Binding to Human Thrombin. *Eur. J. Pharm. Sci.* **2001**, *12*, 441−446.
(36) Wang, W.; Wang, J.; Kollman, P. A. What Determines the van der Waals Coefficient β in the LIE (Linear Interaction Energy) Method to Estimate Binding Free Energies Using Molecular Dynamics Simulations? *Proteins* **1999**, *34*, 395−402.
(37) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.
(38) 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.
(39) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the Binding Affinities of FKBP12 Inhibitors Using a Linear Response Method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.
(40) Krowarsch, D.; Dadlez, M.; Buczek, O.; Krokoszynska, I.; Smalås, A. O.; Otlewski, J. Interscaffolding Additivity: Binding of P1 Variants of Bovine Pancreatic Trypsin Inhibitor to Four Serine Proteases. *J. Mol. Biol.* **1999**, *289*, 175−186.
(41) Helland, R.; Otlewski, J.; Sundheim, O.; Dadlez, M.; Smalås, A. O. The Crystal Structures of the Complexes Between Bovine Beta-Trypsin and Ten P1 Variants of BPTI. *J. Mol. Biol.* **1999**, *287*, 923−942.
(42) Brandsdal, B. O.; Smalås, A. O. Evaluation of Protein−Protein Association Energies by Free Energy Perturbation Calculations. *Protein Eng.* **2000**, *13*, 239−245.
(43) Brandsdal, B. O.; Åqvist, J.; Smalås, A. O. Computational Analysis of Binding of P1-Variants to Trypsin. *Protein Sci.* **2001**, *10*, 1584−1595.
(44) Outzen, H.; Berglund, G. I.; Smalås, A. O.; Willassen, N. P. Temperature and pH Sensitivity of Trypsins from Atlantic Salmon (*Salmo salar*) in Comparison with Bovine and Porcine Trypsin. *Comp. Biochem. Physiol. B* **1996**, *115*, 33−45.
(45) Brandsdal, B. O.; Smalås, A. O.; Åqvist, J. Electrostatic Effects Play a Central Role in Cold Adaptation of Trypsin. *FEBS Lett.* **2001**, *499*, 171−175.
(46) Gorfe, A. A.; Brandsdal, B. O.; Leiros, H. K.; Helland, R.; Smalås, A. O. Electrostatics of Mesophilic and Psychrophilic Trypsin Isoenzymes: Qualitative Evaluation of Electrostatic Differences at the Substrate Binding Site. *Proteins* **2000**, *40*, 207−217.
(47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; 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.
(48) Luzhkov, V. B.; Åqvist, J. Mechanisms of Tetraethylammonium Ion Block in the KcsA Potassium Channel. *FEBS Lett.* **2001**, *495*, 191−196.
(49) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. A. The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity. *Science* **1998**, *280*, 69−77.
(50) Meuser, D.; Splitt, H.; Wagner, R.; Schrempf, H. Exploring the Open Pore of the Potassium Channel from *Streptomyces Lividans*. *FEBS Lett.* **1999**, *462*, 447−452.
(51) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olsen, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(52) Zhou, M.; Morais-Cabral, J. H.; Mann, S.; MacKinnon, R. Potassium Channel Receptor Site for the Inactivation Gate and Quaternary Amine Inhibitors. *Nature* **2001**, *411*, 657−661.
(53) Paulsen, M. D.; Ornstein, R. L. Binding Free Energy Calculations for P450cam-Substrate Complexes. *Protein Eng.* **1996**, *9*, 567−571.
(54) Gorse, A.-D.; Gready, J. E. Molecular Dynamics Simulations of the Docking of Substituted N5-Deazepterins to Dihydrofolate Reductase. *Protein Eng.* **1997**, *10*, 23−30.
(55) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method. *J. Med. Chem.* **1999**, *42*, 5142−5152.
(56) Chen, J. G.; Wang, R. X.; Taussig, M.; Houk, K. N. Quantitative Calculations of Antibody−Antigen Binding: Steroid DB3 Binding Energies by the Linear Interaction Energy Method. *J. Org. Chem.* **2001**, *66*, 3021−3026.
(57) Marelius, J.; Ljungberg, K. B.; Åqvist, J. Sensitivity of an Empirical Affinity Scoring Function to Changes in Receptor−Ligand Complex Conformations. *Eur. J. Pharm. Sci.* **2001**, *14*, 87−95.
(58) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *Q*: An MD Program for Free Energy Calculations and Empirical Valence Bond Simulations in Biomolecular Systems. *J. Mol. Graph. Model.* **1998**, *16*, 213−225.

AR010014P