Protein−Ligand Binding Affinity by Nonequilibrium Free Energy Methods

Benjamin P. Cossins,[†] Sebastien Foucher,[†] Colin M. Edge,[‡] and Jonathan W. Essex*,[†]

School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K., and GlaxoSmithKline, New Frontiers Science Park, Coldharbour Road, Harlow CM19 5AD, U.K.

Received: April 23, 2008; Revised Manuscript Received: September 5, 2008

Nonequilibrium (NE) free energy methods are embarrassingly parallel and may be very conveniently run on desktop computers using distributed computing software. In recent years there has been a proliferation of NE methods, but these approaches have barely, if at all, been used in the context of calculating protein−ligand binding free energies. In a recent study by these authors, different combinations of NE methods with various test systems were compared and protocols identified which yielded results as accurate as replica exchange thermodynamic integration (RETI). The NE approaches, however, lend themselves to extensive parallelization through the use of distributed computing.[1] Here the best performing of those NE protocols, a replica exchange method using Bennett's acceptance ratio as the free energy estimator (RENE), is applied to two sets of congeneric inhibitors bound to neuraminidase and cyclooxygenase-2. These protein−ligand systems were originally studied with RETI,[2] giving results to which NE and RENE simulations are compared. These NE calculations were carried out on a large, highly distributed group of low-performance desktop computers which are part of a Condor pool.[3] RENE was found to produce results of a predictive quality at least as good as RETI in less than half the wall clock time. However, non-RE NE results were found to be far less predictive. In addition, the RENE method successfully identified a localized region of rapidly changing free energy gradients without the need for prior investigation. These results suggest that the RENE protocol is appropriate for use in the context of predicting protein−ligand binding free energies and that it can offer advantages over conventional, equilibrium approaches.

1. Introduction

In recent years, free energy calculations have become more applicable to rational drug design due to the availability of high-performance computing and better algorithms.[2,4−7] However, these examples are still very system specific and generally limited to groups of ligands which have a common structure (congeneric series of ligands). While approaches are being developed to simulate more diverse ligands,[8−10] the computational cost of these equilibrium free energy calculations remains a handicap. There have been many studies focused on nonequilibrium (NE) free energy methods, most of which are related to the Jarzynski equality (JAR).[11−13] Many of these NE methods have the potential for large scale parallelization of calculations. Through the use of large-scale distributed computing, these methods may be able to offer larger amounts of simulation sampling than established equilibrium methods such as thermodynamic integration (TI) and thus provide more widely applicable calculations with more precise results. In a recent study by the present authors,[1] many highly parallelizable NE methods were compared to equilibrium approaches using simple harmonic oscillator test systems and a more demanding test system of water and methane. This study demonstrated that some of the nonequilibrium free energy methods are capable of producing free energies of comparable quality to the best equilibrium approaches. NE methods offer an additional advantage, however, in that they are ideally suited to working in a distributed computing environment, where the available computational resources are inhomogeneous, not fixed in number, and calculations are not guaranteed to complete. In the present study, Monte Carlo (MC) free energy calculations are used in conjunction with NE methods to predict the binding free energies of sets of inhibitors for neuraminidase and cyclooxygenase-2 (COX2), with comparisons being made to replica exchange thermodynamic integration (RETI) results with the same systems from another study.[2] A simple wall clock time comparison is made between these NE calculations carried out with stolen cycles from many low-performance desktop computers and the RETI calculations of Michel et al. carried out on a smaller number of dedicated high-performance processors.

In sections 2 and 3 the methodolgy used in this study is explained. Section 2 includes explanations of free energy methods: nonequilibrium work methods, thermodynamic integration, replica exchange TI, and replica exchange NE. Also, the "thermodynamic cycle" used in our calculations is explained and a method to measure the "predictive ability" of our methods presented. The results and conclusions are given in sections 4 and 5.

2. Free Energy Methods

2.1. Nonequilibrium Work Methods. Free energy methods are used to calculate free energy differences between two systems, defined by

$$\Delta F = F_B - F_A \tag{1}$$

where F_A is the free energy of system A, F_B is the free energy of system B, and ΔF is the free energy difference. Recently, there has been a renewed interest in free energy methods which employ a continuous switching process from systems A to B.

* Corresponding author. E-mail: J.W.Essex@soton.ac.uk.
† University of Southampton.
‡ GlaxoSmithKline.

10.1021/jp803533w CCC: $40.75
Published on Web 10/31/2008

This switching process provides a value of the work performed in making the switch, defined by

$$W = \sum_{i=1}^{n_{\text{inc}}} U(q_{t_i})_{\lambda_{i+1}} - U(q_{t_i})_{\lambda_i} \quad (2)$$

where W is the work performed, λ is a variable coupled to the switch ($\lambda = 0$ = system A and $\lambda = 1$ = system B), $U(q_{t_i})_{\lambda_i}$ is the potential at the current value of λ (where (q_{t_i}) denotes the set of coordinates at the time of λ increment i), and n_{inc} is the number of λ increments. This new interest is mainly due to the work of Jarzynski and the equality he derived,[11,14] which relates a distribution of these nonequilibrium work values to the equilibrium free energy difference:

$$\langle \exp\{-\beta W\} \rangle = \exp\{-\beta \Delta F\} \quad (3)$$

$$\Delta F = -1/\beta \ln\langle \exp\{-\beta W\} \rangle \quad (4)$$

where $\beta = 1/k_{\text{B}}T$ and $\langle \cdots \rangle$ denotes an ensemble average.

For equality (3) to be true, the switches must have initial configurations taken from the same equilibrium ensemble (an ensemble of switches). Also, all switches in an ensemble must be of the same internal structure (same numbers of λ increments and simulation steps).[11]

An interesting development of this switching methodology is the use of Bennett's acceptance ratio (BAR)[15] as a free energy estimator, with switches in both forward (A → B) and backward (B → A) directions being used to calculate ΔF.[16,17] Here ΔF is estimated: starting with the lower bound estimate of ΔF, found through the average work, the ΔF estimate is then increased slowly and iteratively until ΔF satisfies

$$\sum_{i=1}^{n_{\text{swi}}} (1 + \exp\{\beta(W_i^{\text{F}} - \Delta F)\})^{-1} - \sum_{j=1}^{n_{\text{swi}}} (1 + \exp\{-\beta(W_j^{\text{R}} - \Delta F)\})^{-1} = 0 \quad (5)$$

where n_{swi} is the number of switches in one direction (assuming both directions always have the same number of switches) and WiF and WjR are work values in the forward and backward directions, respectively. BAR produces the ΔF estimate with the lowest variance for a given set of forward and backward work values and thus may be the most efficient nonequilibrium work estimator. However, in some cases, a JAR estimate using work values from one direction has been shown to be more efficient,[18] although in our experience it is difficult to identify a priori which direction is the more accurate[1] and the BAR estimator is generally to be preferred. As an alternative to regularly incrementing λ and using eq 2 to calculate the work performed, it has been suggested that by judiciously choosing either the size of the λ increment, or the particular system configuration to which the increment is applied, smaller work values that are closer to equilibrium may be obtained.[12,13] However, in our experience,[1] no one particular algorithm performs consistently better than any other, and given their computational costs, the simple use of eq 2 at regular intervals is to be preferred.

2.2. Thermodynamic Integration. Thermodynamic Integration (TI) is a well-established rigorous free energy method that is well described in many texts.[19,20] Simulations are run at different values of the coupling parameter λ, and the free energy gradient $(\partial F/\partial\lambda)_\lambda$ is accumulated at each value of λ. ΔF from A to B is then found by integrating over the measured gradients:

$$\Delta F = \int_0^1 \left(\frac{\partial F}{\partial \lambda}\right)_\lambda \mathrm{d}\lambda \quad (6)$$

The free energy gradients can be approximated numerically by the finite difference as in eq 7. TI which uses a finite difference approximation is called finite difference thermodynamic integration (FDTI) and will be used in this study over other forms of TI due to its simplicity and generality.[21−23]

$$\left(\frac{\partial F}{\partial \lambda}\right)_\lambda = \left(\frac{\Delta F}{\Delta \lambda}\right)_\lambda \quad (7)$$

ΔF in eq 7 can be found using the Zwanzig equation[24] and potential values at λ and $\lambda + \Delta\lambda$, i.e.

$$\Delta F = -\frac{1}{\beta} \ln\langle \exp(-\beta(U_{\lambda+\Delta\lambda} - U_\lambda)) \rangle_\lambda \quad (8)$$

The size of the $\Delta\lambda$ evaluation made to find a gradient measurement in FDTI must be small in order that the exact gradient at the correct point is obtained. In this study a $\Delta\lambda$ of 0.001 was found to be sufficiently small.

2.3. Replica Exchange TI. Replica exchange thermodynamic integration (RETI) combines TI with Hamiltonian replica exchange moves between adjacent λ simulations.[25,26] λ moves are made in such a way that each configuration is exchanged with a neighbor. Detailed balance is conserved through the use of a double Metropolis test, one for each configuration introduced to a new simulation. λ swap moves are accepted if

$$\exp\{\beta[U_{\text{B}}(j) - U_{\text{B}}(i) - U_{\text{A}}(j) + U_{\text{A}}(i)]\} \geq \text{rand}(0, 1) \quad (9)$$

is true, where i and j are configurations being exchanged and A and B are the Hamiltonians of the replicas exchanging.

RETI increases sampling especially of the explicit solvent by providing the possibility of simulations making large jumps in phase space. Also, as simulations are able to move freely across λ, configurations which are more favorable to a particular area of λ may migrate there.

2.4. Replica Exchange NE. It has been demonstrated that when using the Jarzynski estimator, it may be optimal to divide the λ coordinate into intervals, for which the work values and associated free energy differences are evaluated independently, as follows:[18,27]

$$\Delta F = \sum_{i=0}^{n_{\text{int}}} -\frac{1}{\beta} \ln\langle \exp\{-\beta W_i\} \rangle \quad (10)$$

where W_i denotes the work values from one of n_{int} intervals in the λ coordinate. Strangely, this possible optimization is often not explored in studies of chemical systems.[28,29] The division of the λ coordinate in NE free energy calculations has also been investigated in the recent study by the present authors.[1] We found that although this approach did not yield improved accuracy for the BAR estimator, improvements using JAR were

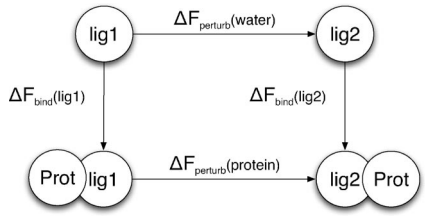

Figure 1. Thermodynamic cycle used in the calculation of the relative binding free energy of protein−ligand complexes.

observed. More importantly, however, this subdivision of the coupling parameter allows optimization of the calculation through two routes. First, multiple equilibrium simulations at each value of λ may be run in parallel to generate the seed configurations for each switch. Second, the fact that multiple equilibrium simulations are being run allows for a replica exchange approach to be used, where configurations adjacent in λ are exchanged. The λ swap moves as part of this equilibrium simulation are again accepted on the basis of eq 9. This so-called replica exchange nonequilibrium (RENE) method was shown to be significantly more precise than other approaches when applied to the simple water−methane system,[1] presumably by generating a more diverse set of equilibrium seed simulations for each nonequilibrium switch. All RENE calculations in this study will divide λ into 10 equally sized partitions (0−0.1, 0.1−0.2...0.9−1.0). These calculations will then be called RENE-BY10, with BY10 referring to the 10 equal divisions of the λ coordinate. Where the total calculation is subdivided, but replica exchange moves are not used to optimize the initial equilibrium simulations, the protocol will be referred to as NE-BY10.

2.5. Thermodynamic Cycle. This study is concerned with calculating relative binding free energies ($\Delta\Delta F_{\text{bind}}$) of inhibitors for enzymes. To find $\Delta\Delta F_{\text{bind}}$s for protein−ligand systems, it is necessary to run two free energy calculations which are part of a thermodynamic cycle as shown in Figure 1. $\Delta\Delta F_{\text{bind}}$s are then calculated through eq 11:

$$\begin{aligned}\Delta\Delta F_{\text{bind}} &= \Delta F_{\text{perturb}}(\text{protein}) - \Delta F_{\text{perturb}}(\text{water}) \\ &= \Delta F_{\text{bind}}(\text{lig2}) - \Delta F_{\text{bind}}(\text{lig1})\end{aligned} \tag{11}$$

2.6. Predictive Ability. The predictive index PI is a quantitative measure of how useful a predicted set of protein−ligand relative binding free energies are in terms of rank order.[30] PI is calculated using the formula

$$\text{PI} = \frac{\sum_{j>i}\sum_{i} w_{ij}C_{ij}}{\sum_{j>i}\sum_{i} w_{ij}} \tag{12}$$

with

$$w_{ij} = |E(j) - E(i)| \tag{13}$$

and

$$C_{ij} = \begin{cases} -1 & \text{if } E(j) - E(i)/P(j) - P(i) < 0 \\ +1 & \text{if } E(j) - E(i)/P(j) - P(i) > 0 \\ 0 & \text{if } P(j) - P(i) = 0 \end{cases} \tag{14}$$

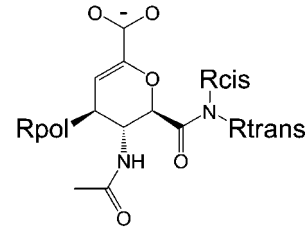

Figure 2. 2D chemical structure of the basic scaffold for the neuraminidase ligands used in this study.

TABLE 1: Experimental Activity of the Sialic Acid Analogues against Neuraminidase[31]

no.	Rtrans	Rcis	Rpol	IC_{50} (μM)
1	Me	H	NH_3^+	190
2	Et	H	NH_3^+	13
3	Me	Me	NH_3^+	2.4
4	Et	Et	NH_3^+	0.003
5	Me	H	$NHC(NH_2)_2^+$	7
6	Me	Me	$NHC(NH_2)_2^+$	0.025
7	Et	Et	$NHC(NH_2)_2^+$	0.001
8	$(CH_2)_2Ph$	Pr	$NHC(NH_2)_2^+$	0.005
9	$(CH_2)_2Ph$	H	NH_3^+	12
10	$(CH_2)_2Ph$	Pr	NH_3^+	0.005

In eqs 12−14, $P(i)$ is the calculated binding affinity of ligand i and $E(i)$ is the experimental binding affinity of the same ligand. Thus, a PI of 1.0 for a set of ligands means that all ligands are in the correct rank order with respect to the experimental order. A PI of −1.0 means that the predicted rank order of the ligand is the opposite of the experimental order, and a PI of 0 means the predicted order is random.

3. Simulation Methods

The model systems used in this study are identical to the explicit solvent simulations of Michel et al. in their recent study.[2] As a result, some important system setup information is not described here.

Figure 2 is a schematic of the basic scaffold for all neuraminidase ligands, which are based on the substrate sialic acid. The scaffold positions Rcis, Rtrans, and Rpol are filled by different substituents for the ligands listed in Table 1.

From the IC_{50} data in Table 1 it seems that it is favorable for binding to fill both the cis and trans pockets. There is little difference between binding affinities of ligands 4, 7, 8, and 10, suggesting there is little profit in groups bulkier than Et lying in the trans and cis pockets. The difference between binding affinities of ligands 4, 7, 8, and 10 are within the experimental errors of approximately half an order of magnitude.[31]

The substitution of an amino group at position Rpol for a guanadino group is also strongly favorable (compare ligands 3 and 6 in Table 1). This is due to the displacement of a bound water by the guanadino group, which would be present with ligands 1−4, 9, and 10.[32]

The series of COX2 ligands evaluated in this study have the same scaffold as celecoxib. This common scaffold is displayed in Figure 3, and the 10 R group substitutions with corresponding experimental binding affinities are listed in Table 2. It is clear that placing polar groups at position R reduces binding affinity (ligands 4, 7, and 8) for COX2, electron-donating groups increase affinity (ligands 5 and 6), and larger groups also decrease affinity (ligand 3). This makes sense considering the nature of the environment into which the R group resides.

The AMBER99[35] force field was used for the protein molecules, and the GAFF[36] force field, with partial charges

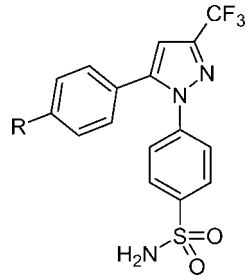

Figure 3. Common structure of the series of celecoxib analogues under study.

TABLE 2: Experimental Activity of the Celecoxib Analogues against COX2[33]

no.	R	IC_{50} (μM)
1	CH_3	0.040
3	CH_2CH_3	0.86
4	CH_2OH	93.3
5	SCH_3	0.009
6	OCH_3	0.008
7	CF_3	8.23
8	OH	>100
9	Cl	0.01
10	F	0.041
11	H	0.032

found using the AM1/BCC[37] method, was used for the inhibitors. Atoms in residues further than 15 Å away from any heavy atom of a representative ligand were discarded to give a protein scoop in an attempt to save computational cost. These scoops were solvated with a 22 Å sphere of TIP4P[34] water molecules centered on the geometric center of the ligand and held in place by a soft half-harmonic potential (1.5 kcal mol^{-1}). All backbone atoms were kept rigid throughout after an initial minimization and side chain bond angles and dihedrals in residues within 10 Å of the ligand were sampled using a modified version of the ProtoMS2.1 simulation program.[38] It is worth noting that within ProtoMS2.1 the Hamiltonians of these simulations are coupled to λ through linear scaling the force field parameters. The MC trial probability ratios used for bound state simulations were protein 9/solute 1/solvent 60, and for unbound state simulations they were solute 1/solvent 60.

The RETI analysis of Michel et al. equilibrated each λ simulation for 10 million MC trials (after any general system equilibration) before any data were collected. The NE protocol of this study equilibrated each λ simulation (here, for NE methods we call these simulations "seed simulations") for 2 million MC trials. This reduction in the size of the λ equilibration for NE protocols was made to save time. The smaller size of these equilibration simulations was found to be sufficient through studying their energy fluctuations for evidence of drift or instability.

Switches were performed using a single topology methodology similar in style to the water to methane study of Cossins et al.[1] The growth of large groups such as benzene rings was made possible by hiding all atoms of the larger group to be grown within the smaller group from which it appears. All bonds in the group to be grown were initially set to 0.2 Å, and atoms were switched to dummy atoms. For example, the perturbation from ligand 2 to 9 grows a benzene ring from a hydrogen. The end state with ligand 2 has a benzene ring consisting of dummy atoms with bonds 0.2 Å long. Even with the benzene ring in its shrunken state, it is unable to hide completely behind the influence of the hydrogen atom from which it grows. However, it is important that as the dummy atoms of the ring are turned

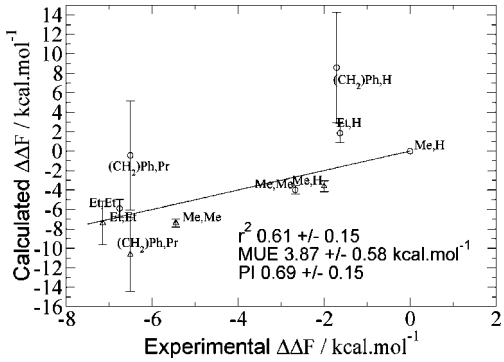

Figure 4. Average of eight repeated sets of calculated $\Delta\Delta F_{bind}$s compared to experimental values, relative to ligand 1 (Me, H, NH_2), for 10 neuraminidase ligands using NE-BY10. Circle data points denote an amino group at position Rpol while triangle data points denote a guanadino group. The solid line represents perfect agreement with experimental values. Errors for all values are standard deviations on the averages.

on they are not close to any other atoms, as this may cause extremely large energies and lead to issues of free energy convergence.

NE protocols for this study were devised with the results and conclusions of the recent study by the present authors in mind.[1] The RENE-BY10 protocol of that study and BAR estimator were deemed most suitable for these and possibly all protein−ligand systems. Also, calculations were carried out with the NE-BY10 protocol to give a comparison with non-RE methods. NE-BY10 calculations used 400 switches with 750 MC trials between each of 1000 equally spaced λ increments with 71 300 MC trials of equilibrium simulation between each NE starting configuration. This gives a total of 316 million MC trials to obtain a $\Delta F_{perturb}$(protein) or $\Delta F_{perturb}$(water) which is similar to the 360 million MC trials used, after equilibration, for the same system in the RETI calculations of Michel et al.[2] RENE-BY10 calculations used 800 switches of 375 MC trials between each of 1000 λ increments with 50 000 MC trials of seed simulation between each starting configuration. This protocol uses 322 million MC trials to obtain a $\Delta G_{perturb}$(protein) or $\Delta G_{perturb}$(water). RENE-BY10 calculations used twice as many switches as NE-BY10 calculations due to a compromise between obtaining sufficient λ swaps and using switches which are as long as possible. NE calculations were carried out with NE-BY10 using 800 switches and found to be slightly less efficient (unpublished results).

All NE simulations reported here were performed on an inhomogeneous distributed computing cluster at the University of Southampton, consisting of ~1500 processors, running under the Condor software.[3] The equilibrium seed simulation were also successfully run on the distributed computing cluster. Each calculation discussed here used 100 condor processors and are thus considered well parallelized.

4. Results

4.1. Neuraminidase. The neuraminidase NE-BY10 calculations described here were repeated eight times using different random number seeds. Figure 4 shows the average of eight repeated NE-BY10 neuraminidase results. Despite the relatively high r^2 and PI scores of the first NE-BY10 analysis, in general, this protocol may not reliably produce results of high quality. The PI scores of the eight repeated calculations range from 0.95 to 0.44.

It is clear that the NE-BY10 protocol used here is unable to produce consistently accurate free energies. Although the first

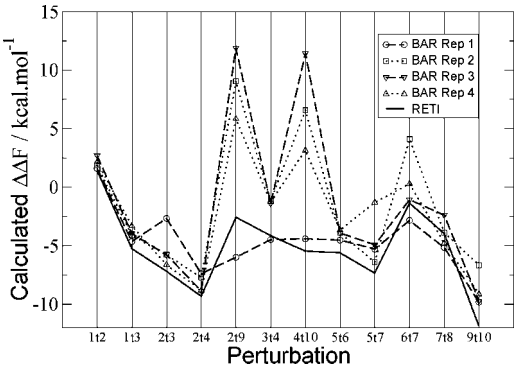

Figure 5. Comparison of $\Delta\Delta F_{\text{bind}}$s for the NE-BY10 repeated neuraminidase calculations. Statistical errors, found through a block averaging method, were below 0.6 kcal mol^{-1} for all $\Delta\Delta F_{\text{bind}}$s with perturbations 2 to 9, 4 to 10, and 7 to 8 generally being higher than the others.

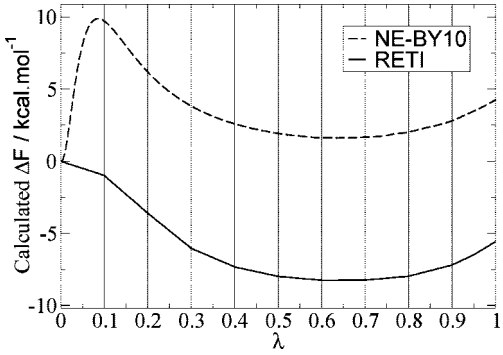

Figure 6. Comparison of RETI and NE-BY10 PMFs of perturbation 2 to 9 in vacuum. Statistical errors, found through a block averaging method, are too small to be visible.

NE-BY10 repetition is of similar quality to the RETI analysis,[2] the subsequent seven NE-BY10 repeats are all far less useful. This poor predictive quality of the eight NE-BY10 repeats is due to a subset of the perturbations carried out in this neuraminidase analysis. Figure 5 compares the individual perturbation $\Delta\Delta F_{\text{bind}}$s for the first four NE-BY10 repeats.

Figure 5 makes it clear that in the three poorer NE-BY10 repeats perturbations 2 to 9 and 4 to 10 differ the most from RETI and are the major source of the poor predictive accuracy in repeats 2−4. The source of the lack of accuracy in estimates of perturbations 2 to 9 and 4 to 10 in repeats 2−4 is probably poor convergence due to random factors. It is worth noting that perturbations 2 to 9 and 4 to 10 both involve the growth of a phenyl ring as discussed above.

Some of the underlying $\Delta F_{\text{perturb}}$(protein)s and $\Delta F_{\text{perturb}}$(water)s of these NE-BY10 analyses (see Supporting Information) display significant differences from their RETI counterparts.[2] In particular, the NE-BY10 $\Delta F_{\text{perturb}}$(protein), $\Delta F_{\text{perturb}}$(water), and ΔF_{vac} values for perturbations 2 to 9, 4 to 10, and 7 to 8 show significant differences from their RETI counterparts, but strangely the differences seem to be much smaller when comparing RETI and NE-BY10 $\Delta\Delta F_{\text{bind}}$s and $\Delta\Delta F_{\text{solv}}$s. It is no coincidence that each of these perturbations showing the largest discrepancy between NE-BY10 and RETI involve the growth of a phenyl into the trans pocket. In investigating this discrepancy, it was convenient to concentrate on vacuum perturbations as simulations are faster and more precise than protein and water perturbations. Figure 6 shows the PMF across the λ coordinate for perturbation 2 to 9 in vacuum with black vertical lines marking the points at which FDTI λ simulations are run. The large disparity in the PMFs of NE-BY10 and RETI

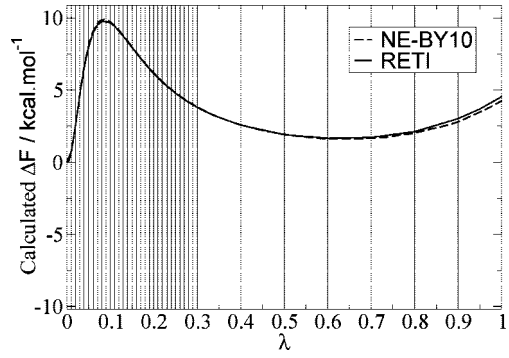

Figure 7. Comparison of RETI with extra λ sampling and NE-BY10 PMFs of perturbation 2 to 9 in vacuum. The RETI analysis has 30 λ values between $\lambda = 0$ and $\lambda = 0.3$. Statistical errors, found through a block averaging method, are too small to be visible.

for perturbation 2 to 9 in vacuum is clear. The NE-BY10 PMF shows a large peak between $\lambda = 0$ and $\lambda = 0.1$ which is missed by RETI.

Apart from this the two PMFs seem similar. From this it can be deduced that the RETI calculation misses this large peak seen with BAR due to the discrete and nondiscrete nature of sampling with RETI and NE-BY10, respectively, along the λ coordinate. Figure 7 shows the same comparison as Figure 6 but with 30 RETI λ sampling windows between $\lambda = 0$ and $\lambda = 0.3$ rather than only 3. The difference between Figures 7 and 6 is very clear and verifies the idea that the disparity seen between RETI and NE-BY10 for perturbations 2 to 9, 4 to 10, and 7 to 8 is because of a large error in the RETI calculation due to its discrete nature.

The large peak in the PMF in Figure 7 is due to very large intramolecular Lennard-Jones (LJ) and Coulombic energies. These large LJ and Coulombic energies are produced by the very close proximity of the atoms of the phenyl ring as they are turned on at the beginning of the perturbation (they are 0.2 Å apart at $\lambda = 0$). This large PMF peak in Figure 7 is therefore present as a result of an inappropriate choice of perturbation scheme. The use of larger bond distances for the phenyl group in its shrunken state removes this artifact although increases the risk of clashes with other molecules as the nonbonded terms are turned on. This suggests that the use of a soft core potential[39,40] for perturbations involving the growth a large group may produce a more favorable path. However, the discovery of this error in an equilibrium approach with an NE method does highlight the importance of sampling across the whole λ coordinate. As this disparity is an internal solute nonbonded effect, it is very well behaved and, as it is present in both protein and water perturbations, is canceled out in the resulting RETI $\Delta\Delta F_{\text{bind}}$ values.[2]

A RENE-BY10 analysis described in the methods section was repeated eight times. Figure 8 shows RENE-BY10 predicted $\Delta\Delta F_{\text{bind}}$s, relative to ligand 1 (Me, H, NH_2), compared to experiment. In general, the eight RENE-BY10 analyses have a similar PI score to the RETI analysis of Michel et al. Also, all of the r^2 values of the RENE-BY10 analyses are higher than that of the RETI analysis. This suggests that this RENE-BY10 protocol is at least as predictive as the RETI protocol.

Figure 9 compares the PI score against MC trials through the first four RENE-BY10 analyses and the RETI analysis of Michel et al.[2] It is fair to say all are quite similar. RETI seems to take slightly longer to converge to a high PI score; this may be due to the fact that in the RETI calculations the λ equilibration simulations did not include RE moves. Figure 10

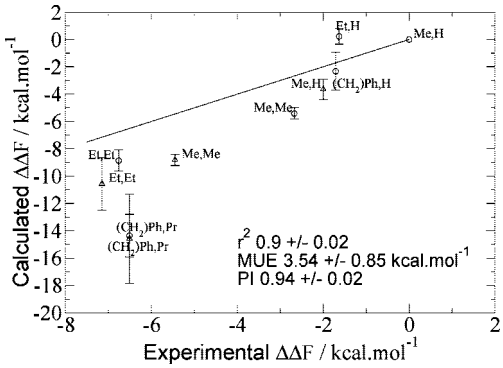

Figure 8. Average of eight repeated sets of calculated $\Delta\Delta F_{bind}$s compared to experimental values, relative to ligand 1 (Me, H, NH_2), for 10 neuraminidase ligands using RENE-BY10. Circle data points denote an amino group at position Rpol while triangle data points denote a guanadino group. The solid line represents perfect agreement with experimental values. Errors for all values are standard deviations on the averages.

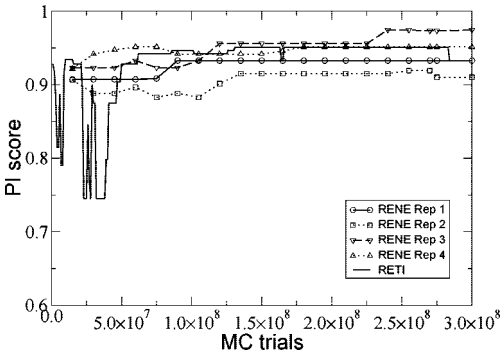

Figure 9. Comparison of PI scores for 4 RENE-BY10 and 1 RETI neuraminidase analyses. The number of MC trials on the *x*-axis represents the number of MC trials from one perturbation leg; i.e., all protein and water simulations are run in parallel.

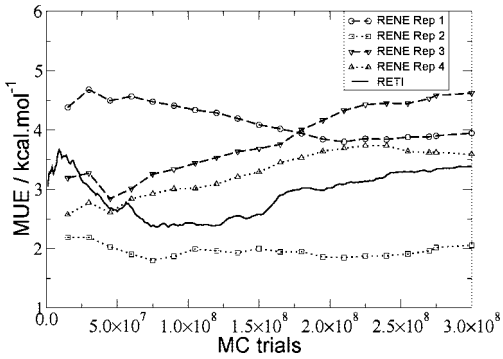

Figure 10. Comparison of the MUE for 4 RENE-BY10 and 1 RETI neuraminidase analyses. The number of MC trials on the *x*-axis represents the number of MC trials from one perturbation leg; i.e., all protein and water simulations are run in parallel.

compares the MUE of the first four neuraminidase RENE-BY10 analyses through the calculations compared to that for the equivalent RETI analysis.[2] RENE-BY10 repeat 2 starts low and does not fluctuate as much as the other analyses. Again, it is interesting how a low MUE does not necessarily translate to a comparatively high PI score, and vice-versa, as seen with repeats 2 and 4.

As these NE analyses discussed above are completed in a relatively low wall clock time (around 10 h on 100 condor nodes for each calculation, including λ equilibration), it is possible to investigate any possible improvement found by increasing the number of MC trials used. A RENE protocol with 3 times the sampling of the simulations above (918 million MC trials for each protein and water ΔF) was devised. This longer protocol has 800 switches of 1125 MC trials between each of 1000 λ increments with 100 000 MC trials of seed simulation between each equilibrium starting configuration (RENE-BY10x3). This more expensive protocol can still be completed in ~20 h using 100 condor nodes for each calculation in parallel. This study yielded an r^2 of 0.91, a MUE of 3.48 kcal mol^{-1}, and a PI of 0.95, suggesting that the extra sampling offers little improvement in the PI.

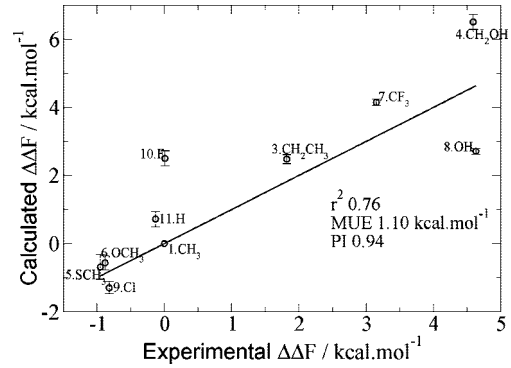

Figure 11. Comparison of RENE-BY10 calculated and experimental $\Delta\Delta F$s relative to ligand 1 for 10 COX2 ligands. Errors were calculated through a block averaging method. The solid line represents perfect agreement with experimental values.

4.2. Cyclooxygenase-2. Figure 11 compares RENE-BY10 calculated and experimental ligand binding results for COX2 ligands relative to ligand 1. The r^2 of 0.76 and PI of 0.94 show this RENE-BY10 analysis to be slightly less predictive than the RETI analysis of Michel et al.[2] (r^2 of 0.85 and PI of 0.96). It is possible that this difference in predictive quality, which can be attributed to the overestimation of $\Delta\Delta F_{bind}$s for ligands 10 (F) and 4 (CH_2OH), is a random effect.

These COX2 results show much closer agreement with experiment compared to the neuraminidase analysis. This is most probably due to the differences in the size of the ligands, the complexity of some of the perturbations, and the different nature of the binding pockets.

Another COX2 analysis was carried out using the RENE-BY10x3 protocol defined above. This more costly RENE-BY10x3 analysis has a MUE of $\Delta\Delta F_{bind}$s relative to ligand 1 of 0.89 kcal mol^{-1}, an r^2 of 0.79, and a PI of 0.94 which is a slight improvement over the RENE-BY10 analysis above.

It is important to study the level of predictability throughout the calculations presented above as this may fluctuate to some extent. Figures 12 and 13 show the fluctuations in PI and MUE through the calculations of the COX2 analyses. Figure 12 shows that the RETI analysis converges to a high PI score more quickly than the RENE-BY10 analyses. Figure 13 shows that the RETI analysis has a lower MUE than the RENE-BY10 analyses at all points of the calculations.

5. Conclusions

This study builds on our previous work by applying our optimum NE free energy protocols to the calculation of protein−ligand binding affinities.[1] To our knowledge, this represents the first such application of these approaches to protein systems. Other studies have investigated slightly different NE methods on less complex systems and found that NE methods are not as efficient as established equilibrium methods.[5,41]

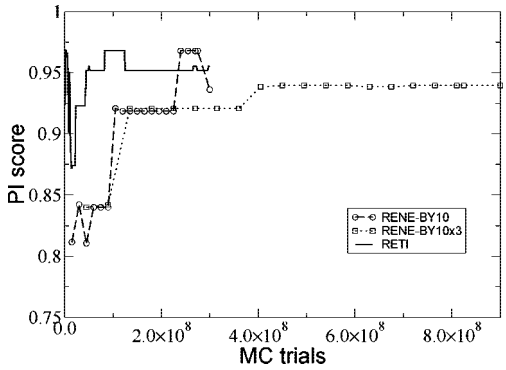

Figure 12. Comparison of PI scores for COX2 analyses. The number of MC trials on the *x*-axis represents the number of MC trials from one perturbation leg; i.e., all protein and water simulations are run in parallel.

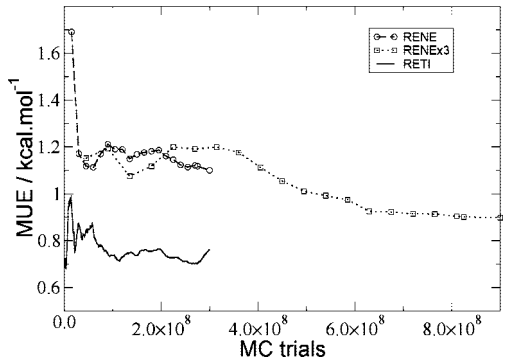

Figure 13. Comparison of the MUE for COX2 analyses. The number of MC trials on the *x*-axis represents the number of MC trials from one perturbation leg; i.e., all protein and water simulations are run in parallel.

However, here we find that our NE methods can perform to a similar standard as established equilibrium methods like RETI.[25]

NE-BY10 was shown to have a very poor predictive performance for the neuraminidase relative binding free energy analysis of this study. The PI scores ranged from 0.95 to 0.44 in Figure 4. The RENE-BY10 protocols, on the other hand, were much more consistent, with PI scores ranging from 0.91 to 0.97 (Figure 8). These results compare favorably with the RETI analysis of Michel et al. Indeed, all RENE-BY10 r^2 scores improved upon the RETI r^2 of 0.81.[2] For clarity, we do not suggest that RENE-BY10 will give more predictive results than RETI for free energy analyses of this kind only that it is able to give equally predictive results using similar amounts of sampling. In addition, the NE calculations identified a large error in the RETI simulations involving the growth of a phenyl ring, highlighting the need for time-consuming exploratory steps when using any free energy method based on TI. TI is unable to calculate accurately the free energy of PMFs with high curvature if insufficient simulations at different values of λ are performed. It is generally thought that for biosystems such as those studied here a PMF of high curvature is relatively unlikely. The large peak in the PMF seen in Figures 6 and 7 arises from the system setup, and fortunately is well behaved, and cancels when constructing the appropriate thermodynamic cycle to determine the relative binding free energies. However, that the NE methods were able to unearth this large discrepancy represents an advantage over conventional, equilibrium approaches and highlights the importance of evaluating the whole perturbation path when using equilibrium methods, even if only in an exploratory capacity.

All the COX2 binding free energy analyses discussed above are very similar and of good predictive quality. Thus, again the RENE-BY10 protocol used in this study has been shown to be the equal of RETI. The RETI analysis of Michel et al. seems to converge to a high PI score and low MUE using fewer MC trials than the RENE-BY10 protocol (Figure 12). RETI converges to a PI score above 0.9 at around 2.46×10^7 MC trials while RENE-BY10 reaches a PI above 0.9 at around 1×10^8 MC trials. It could be argued that this slower convergence of RENE-BY10 is because of the smaller λ equilibration compared to the RETI protocol, although the difference could just as easily be random. This possible advantage should be tempered by the fact that the RENE-BY10 protocol is generally completed much faster than it is possible to complete the RETI protocol.

When considering which method to use for calculations such as these, it is clear that NE methods offer something new and useful. Here the RENE-BY10 protocol has been shown to perform as well as RETI. An advantage of NE methods realized in this study is the possible use of extreme parallelisation. The RENE-BY10 protocol is able to produce a single $\Delta\Delta F_{\text{bind}}$ result in around 10 h on 100 condor nodes while RETI takes around 24 h on 12 2.2 Ghz Opterons. It is clear that NE methods have an advantage in being able to easily utilize large clusters of processors, which may be inhomogeneous and unstable, with there being no guarantee that a submitted switch will necessarily complete. This advantage may become larger as an era of powerful computer processors with many independent processing cores is realized.

Acknowledgment. We thank GlaxoSmithKline, the EPSRC, and the University of Southampton.

Supporting Information Available: Tables of experimental and calculated relative binding energies for 12 neuraminidase perturbations; figures showing comparison of calculated and experimental $\Delta\Delta F$s for 10 neuraminidase COX2 ligands. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Cossins, B. P.; Foucher, S.; Edge, C. M.; Essex, J. W., submitted for publication, 2008.
(2) Michel, J.; Verdonk, M. L.; Essex, J. W. *J. Med. Chem.* **2006**, *49*, 7427–7439.
(3) Litzkow, M.; Livny, M.; Mutka, M. *Proc. 8th Int. Conf. Distributed Comput. Syst.* **1988**, 104–111.
(4) Price, M. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 9455–9466.
(5) Oostenbrink, C.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **2004**, *54*, 237–246.
(6) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; erson, K. S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663–667.
(7) Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domoal, R. A.; Wang, L. G.; Anderson, K. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2006**, *128*, 15372–15373.
(8) Deng, Y. Q.; Roux, B. *J. Chem. Theory. Comput.* **2006**, *2*, 1255–1273.
(9) Oostenbrink, C.; van Gunsteren, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6750–6754.
(10) Michel, J.; Verdonk, M. L.; Essex, J. W. *J. Chem. Theory. Comput.* **2007**, *3*, 1645–1655.
(11) Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018–5035.
(12) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 204104.
(13) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103.
(14) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
(15) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.
(16) Crooks, G. E. *Phys. Rev. E* **1999**, *60*, 2721–2726.
(17) Shirts, M. R.; Blair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601–140604.
(18) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.
(19) Leach, A. R. *Molecular Modelling, Principles and Applications*, 1st ed.; Longman: London, 1996.

(20) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: New York, 1996.
(21) Mezei, M. *J. Chem. Phys.* **1987**, *86*, 7084–7088.
(22) Guimaraes, C. W.; Bicca de Alencastro, R. *J. Med. Chem.* **2002**, *45*, 4995–5004.
(23) Fernandes De Oliveira, C. R.; Guimaraes, C. W.; De Mello, H.; Echevarria, A.; Bicca de Alencastro, R. *Int. J. Quantum Chem.* **2005**, *102*, 42–553.
(24) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
(25) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.
(26) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13711–13718.
(27) Hummer, G. *J. Chem. Phys.* **2001**, *114*, 7330–7337.
(28) Oostenbrink, C.; van Gunsteren, W. F. *Chem. Phys.* **2006**, *323*, 102–108.
(29) Ytreburg, M. F.; Swendsen, R. H.; Zuckerman, D. *Physics* **2006**, *2*, 0602088.
(30) Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 3417–3423.
(31) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. *J. Med. Chem.* **1999**, *42*, 5142–5152.
(32) Taylor, N. R.; Cleasby, O.; Singh, T.; Skarzynski, A. J.; Wonacott, P. W.; Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. *J. Med. Chem.* **1998**, *41*, 798–807.
(33) Penning, T.; Talley, J.; Bertenshaw, S.; Carter, J.; Collins, P.; Docter, S.; Graneto, M.; Lee, L.; Malecha, J.; Miyashiro, J.; Rogers, R.; Rogier, D.; Yu, S.; erson, G.; Burton, E.; Cogburn, J.; Gregory, S.; Koboldt, C.; Perkins, W.; Seibert, K.; Veenhuizen, A.; Zhang, Y.; Isakson, P. *J. Med. Chem.* **1997**, *40*, 1347–1365.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(35) Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. AMBER8, University of California, San Francisco, CA, 2004.
(36) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(37) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.
(38) Woods, C. J.; Michel, J. ProtoMS2.1: in house Monte Carlo software, 2002−2006.
(39) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
(40) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
(41) Oberhofer, H.; Dellago, C.; Geissler, P. L. *J. Phys. Chem. B* **2005**, *109*, 6902–6915.

JP803533W