Acta Biochimica Polonica

Vol. 42 No. 4/1995

525–536

QUARTERLY

Text of lecture presented at the 5th International Symposium on ***Molecular Aspects of Chemotherapy***

Rapid non-empirical approaches for estimating relative binding free energies*

Alan E. Mark**, Yingwu Xu , Haiyan Liu and Wilfred F. van Gunsteren

Laboratorium für Physikalische Chemie, Eidgenössische Technische Hochschule, ETH-Zentrum, CH 8092 Switzerland

Key words: free energy, molecular dynamics, binding constants, drug design

Rapid non-empirical methods for estimating binding free energies are reviewed. A novel approach based on the application of the free energy perturbation formula to a biased ensemble is presented. Preliminary results demonstrating the applicability of this approach in protein systems are shown and the potential of this method in structure-based drug design is discussed.

From a physical perspective drug design is primarily concerned with estimating the difference in the free energy of a compound in different environments. This is because thermodynamically each of the binding, transport and processing events, which together give rise to a specific biological activity, are directly dependent on the associated change in free energy. Historically drug design studies have relied primarily on empirical estimates of free energy differences [1, 2]. The recent availability of structural data on protein targets, however, has led to the growing popularity of free energy estimates based on a variety of energy-based scoring functions [3, 4]. Unfortunately, both approaches have met only with limited success. This is not surprising. Free energy is a statistical property and reflects a thermal average over microscopic configurations. Free energy cannot be reliably deduced from a single structure nor simply approximated as a sum of independent terms [5, 6].

To reliably calculate free energy differences one must turn to statistical mechanical approaches. A number of such methods have been developed to determine free energy differences based on molecular simulation techniques [7]. The theory underling these calculations is straightforward. The change in free energy is determined either from the relative probability of finding the system in a given state or from the work required to go from an initial to a final state *via* a reversible path. Free energy calculations have been used to predict relative binding affinities in good agreement with experiment. Nevertheless, substantial time and computational cost are associated with obtaining a single number. This greatly inhibits the wider application of this methodology in structure based drug design. For example, free energy calculations are not practical for the large-scale screening of compounds. For such applications rapid non-empirical methods that can be used to reliability estimate differences in binding free energies are highly sought after.

In this manuscript recent progress toward the development of general non-empirical methods to estimate free energy differences suitable for application in structure-based drug design is described. A brief introduction into free en-

*Financial support from the Huber-Kudlich Foundation (grant 2-89-100-91) is gratefully acknowledged.

**To whom correspondence should be addressed.

ergy calculations is first given. Alternative rapid non-empirical and semi-empirical approaches to estimate free energy differences are then reviewed. Finally, preliminary results from the application to the binding of small hydrophobic ligands to T4-lysozyme of a recently proposed method to estimate the difference in binding free energy for multiple modifications of a lead compound based on a single simulation of a non-physical initial state are presented.

THEORY

Using statistical mechanics it can be readily shown that the difference in free energy, ΔF, between two states of a system, A and B, described by the Hamiltonian $H(\lambda)$, where $H(0) = H_A$ and $H(1) = H_B$, can be given by [8]

$$\Delta F = F(1) - F(0) = \int_0^1 F'(\lambda)d\lambda = \int_0^1 < \partial H/\partial\lambda >_\lambda d\lambda \quad (1)$$

where $<...>_\lambda$ denotes an average over the ensemble at the corresponding λ value. The coupling parameter λ effectively describes the path taken from the initial to the final state. Formula (1) is commonly referred to as the thermodynamic integration formula. Using (1) the difference in free energy between A and B can be determined either by slowly changing the coupling parameter λ during the course of a simulation and approximating the integral as a continuous sum or separate simulations can be performed at discrete intermediate λ values and the integral determined numerically. An alternative to thermodynamic integration is to compute the difference in free energy between two states of a system directly using the so called free energy perturbation formula [9]

$$\Delta F = \sum_{\lambda=0}^{\lambda=1-\delta\lambda} -k_B T \ln <\exp\{-[H(\lambda+\delta\lambda) - H(\lambda)]/k_B T\}>_\lambda \quad (2).$$

Using either approach a series of simulations are normally performed over a range of intermediate λ values to ensure convergence. This imposes a pathway between the initial and final states determined by the dependence of the potential energy function on λ and means that a separate series of calculations must be performed for each mutation. For a detailed introduction to the theory of free energy calculations the reader is referred to [7, 10, 11] for reviews.

FREE ENERGY EXTRAPOLATION METHODS

In principle, the difference in free energy between a reference state and any other state of a system could be determined if the equilibrium fluctuations of the reference state were completely known. Essentially the free energy of an alternative state can be extrapolated from the behaviour of the system in the reference state. This principle can be readily understood from the following example. Consider a lead compound bound to a receptor as illustrated schematically in Fig. 1. In the complex the average geometry of the binding pocket (configuration (A)), is optimized to fit the lead compound. In this configuration the binding pocket will not accommodate a derivative which includes an extra functional group, -R. Thermal motion of the protein, however, will lead to fluctuations in the geometry of the binding site which may transiently create a cavity optimal for the larger derivative (configuration (B)). If we could determine the correctly weighted probability of sampling this alternate geometry, for example, from a simulation of the protein motions, we would know the difference in binding free energy between the two compounds.

For structure-based drug design, estimating binding free energies from such an extrapolation approach would have many advantages. First, only a single initial reference state need be considered which in many cases would be experimentally known. Second, the fluctuations in the initial state are independent of the changes to be considered. Thus, potential modifications of the reference compound do not have to be pre-defined. Third, the same set of configurations is used to estimate the change in free energy for any possible derivative making such an approach potentially very efficient. The question is, however, whether the change in free energy can be accurately estimated for physically relevant mutations.

There have been several recent studies whose aim has been to predict the change in free energy associated with multiple perturbed states from a single simulation. Gerber *et al.* [12] proposed computing the first-order partial deriva-

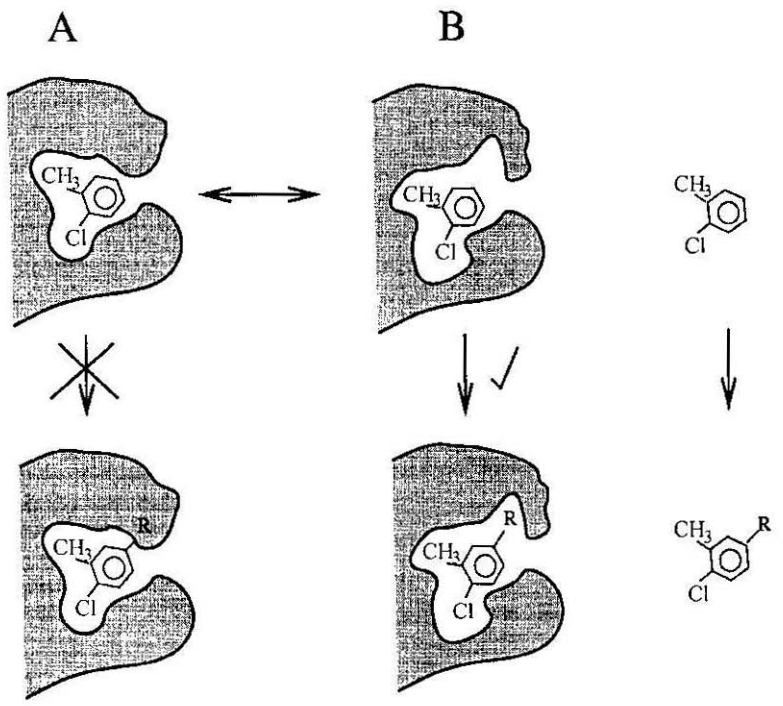

Fig. 1. A diagrammatic representation of the binding of a lead compound to a protein receptor. Configuration A shows the equilibrium conformation of the binding pocket which is unable to accommodate the larger derivative. Configuration B shows an alternative conformation of the binding pocket sampled by thermal motion that can accommodate the larger derivative.

tives of the change in free energy with respect to modification of each atomic site independently. Then, using a linear combination of these derivatives, they attempted to approximate the total change in free energy associated with a series of alternative states. The assumptions in this approach were that the response of the system to a given mutation is small and that the individual contributions to the change in free energy are independent. The relative binding constants of different trimethoprim analogues with dihydrofolate reductase were estimated using this approach. The correlation with experimental data was poor, indicating the approximation was too crude. Later Smith *et al.* [13] refined the approach by expanding the free energy as a function of a coupling parameter λ into a Taylor series around $\lambda = 0$ (reference state) as follows,

$$\Delta F(\lambda) = F(\lambda) - F(0) = F' \mid_{\lambda=0} \lambda + \frac{1}{2!} F'' \mid_{\lambda=0} \lambda^2 + \frac{1}{3!} F''' \mid_{\lambda=0} \lambda^3 + \ldots, \quad (3)$$

where the values of the higher-order derivatives F'', F'''... at $\lambda = 0$ were computed as averages over the ensemble of the reference state [13]. Using a 1 ns simulation, they showed that the change in free energy associated with substantial charge rearrangements (± 0.25 *e*) of a model diatomic dipolar molecule in water could be predicted truncating the series beyond the second or third-order terms. Higher-order derivatives, however, converged slowly. A similar approach, described by Levy *et al.* [14], is equivalent to including only the first and the second-order terms in the Taylor series.

A special case of this approach has been proposed by King and Barford [15] based on linear response theory. They showed that reliable estimates of free energy differences associated with charge rearrangements could be obtained from a single simulation of an intermediate state composed of a linear combination of the two end states by effectively considering only the first two terms of the Taylor series (3). The advantage of using a precisely defined mixed state was that larger modifications of the charge distribution could be handled. However, separate simulations were required for each modification.

Aqvist *et al.* [16] also proposed estimating relative binding free energies based on linear response theory. Rather than considering a mixed state, separate simulations of the two end states were performed. Linear response theory was then used to derive an expression for the electrostatic contribution to the change in free energy. As with the approach of King & Barford [15], this method could be shown to yield good results for substantial charge rear-

rangements. To extend their approach to include mutations involving changes in topology and the creation or deletion of atoms Aqvist *et al.* [16], proposed treating modifications of the van der Waals interactions by an expression analogous to that used for the electrostatic interactions. Although the theoretical basis of such a treatment is dubious, Aqvist *et al.* [16] did show that, by inclusion of an empirically derived scaling factor for the van der Waals contribution, the model could be fitted to reproduce the binding free energies for a series of endothiapepsin inhibitors. The advantage of the approach taken by Aqvist *et al.* [16], as with that of King & Barford [15], is that large modifications can be treated. The disadvantage is that independent simulations must be performed for each mutation. Both methods are at best semi-empirical. A specific functional form for the reponse of the system to the perturbation is assumed (linear response). In cases where this approximation holds, the change in electrostatic free energy can often be reliably estimated using a simple continuum method, e.g. a finite difference solution to the Poisson-Boltzmann equation. To estimate binding free energies, however, the calculated electrostatic contribution must be combined with an empirically derived estimate of the other contributions to the change in free energy. The application of continuum approaches to the estimation of hydration free energies and binding energies are described in detail by Gilson and Honig [17] and in later reviews [18, 19].

If the series expansion (3) used by Smith *et al.* [13] converges, the approach is formally equivalent to the application of the perturbation formula (2). This was noted by Liu *et al.* [20] who demonstrated that, for the water dipole system considered by Smith *et al.* [13], comparable results could be obtained more simply using a perturbation approach. Liu *et al.* [20] went on to show that the perturbation formula could also be used for modest changes in van der Waals parameters, but failed for mutations involving the creation or deletion of atoms.

To obtain a meaningful estimate of the change in free energy associated with a given perturbation applying the perturbation formula, the ensemble generated for the reference state must overlap with that for the alternative state. When a computer-based molecular simulation technique such as Monte Carlo or molecular dynamics is used to generate the relevant ensemble [7, 10, 11], the region of configuration space that is sampled corresponds only to low energy configurations of the state that is simulated. A well-known problem in free energy calculations is that low energy regions of the configurational space in the reference state do not correspond to low energy regions in the end state when atoms are created or deleted. For this reason a given mutation is normally broken into a number of intermediate steps in order to obtain a reliable estimate of the relative free energy. The essential difficulty is that of sampling. A general mechanism to reduce sampling problems in molecular simulations is to incorporate an appropriate biasing or umbrella potential energy term into the Hamiltonian of a given physical reference state. Liu *et al.* [20] demonstrated that a so-called soft-core interaction function could serve as an appropriate biasing term. At positions where atoms were to be created or deleted, soft interaction sites were introduced which interacted with the surroundings *via* a modified Lennard-Jones 6–12 interaction which approached a finite value at short interatomic distances [21]. The idea of the introduction of such interaction sites at positions where atoms are to be created or deleted is to extend the parts of configuration space that can be sampled in the reference state such that it encompasses the parts of configuration space accessible to the system in the relevant final state. The change in free energy associated with mutating the system from this non-physical reference state to a given physical alternative state can then calculated by applying the perturbation formula. The difference in free energy between alternative physical states can be determined by constructing thermodynamic cycles. The difference in free energy for any closed cycle is zero. Liu *et al.* [20] tested this approach by comparing the results from thermodynamic integration calculations to predictions based on a single 300 ps simulation of a non-physical reference state for a series of para-substituted phenols in water illustrated in Fig. 2. The correlation between the target values (thermodynamic integration) and the values based on the extrapolation from the biased ensemble is shown in Fig. 3. For the range of mutations investigated by Liu *et al.* [20] the approach was clearly very successful.

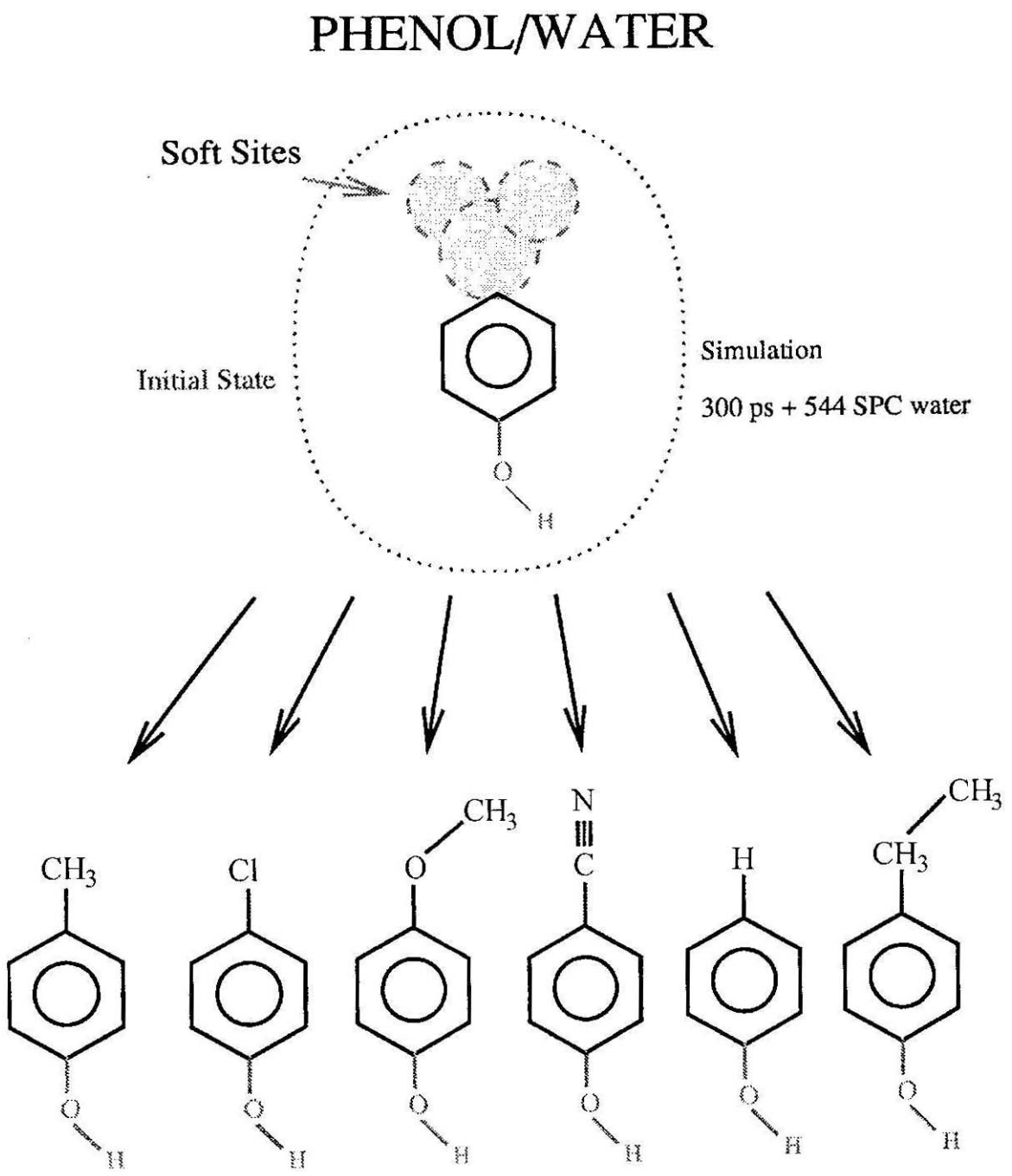

Fig. 2. Illustration of the mutations investigated by Liu et al. [20] to test extrapolation based on a biased ensemble.

ESTIMATING RELATIVE BINDING AFFINITIES USING THE BIASED ENSEMBLE APPROACH: T4-LYSOZYME A CASE STUDY

To test the effectiveness of the approach of Liu *et al.* [20] for estimating differences in binding affinities, the binding of a series of small aromatic ligands to a hydrophobic cavity in T4-lysozyme (L99A) has been investigated. This test case was chosen because, in addition to detailed thermodynamic data for a series of related ligands, high resolution x-ray crystal structures are available for a number of the corresponding complexes [22, 23]. In this preliminary study the relative binding affinities of five ligands, *viz*: benzene, toluene, *o*-xylene, *p*-xylene and *m*-xylene, shown in Fig. 4, were investigated.

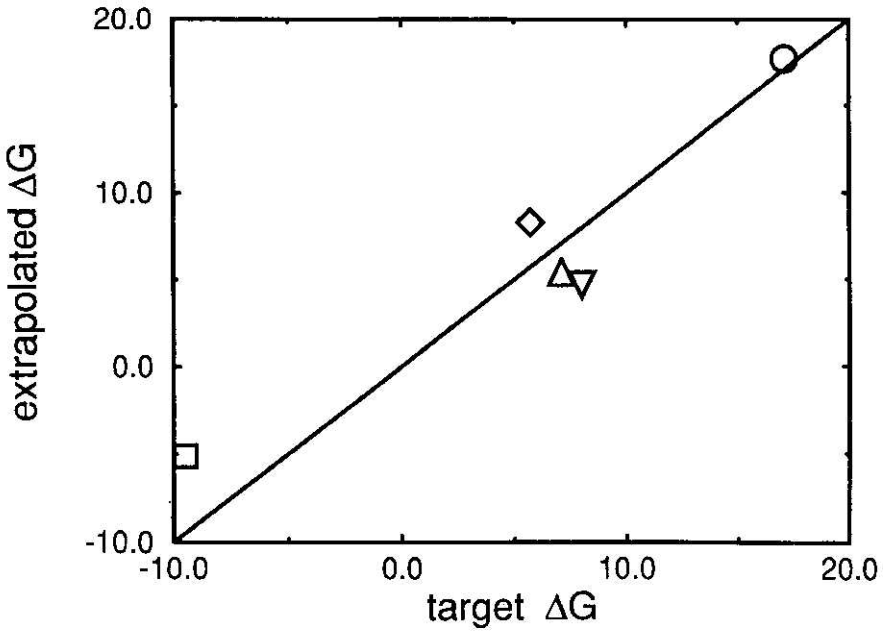

Fig. 3. A plot of the change in free energy calculated using thermodynamic integration (target) versus the change in free energy based on extrapolation from a biased ensemble for the mutations shown in Fig. 2. The change in free energy relative to *p*-methyl phenol is shown for p-Cl-phenol (○), p-CN-phenol (□), p-OCH_3-phenol (▽), p-H-phenol (△), and p-CH_3-CH_3-phenol (◇). Values in kJ/mole are taken from Liu *et al.* [20] for non-physical reference state S_a.

COMPUTATIONAL DETAILS

To estimate the relative binding affinity of two compounds it is necessary to determine the difference in free energy between the two compounds in water and also between the two compounds bound to the receptor. Two systems were, therefore, considered. In the first the reference molecule was placed at the centre of a rectangular periodic box containing 1177 SPC [24] water molecules. The second system contained the same reference molecule bound to T4-lysozyme in a rectangular periodic box containing 4547 SPC water molecules. The reference molecule consisted of a benzene ring backbone to which six soft-interaction sites had been added, one attached to each carbon as illustrated in Fig. 4. The bonded and non-bonded parameters of the soft-interaction sites were set to those of a methyl group in toluene. The scaling parameter α which determines the softness of the interaction function was set to 0.8 (see Beutler *et al.* [21]). The starting configuration for the simulation of the complex was taken from the crystal structure of T4-lysozyme (L99A) complexed with *o*-xylene, entry 188l of the Brookhaven Protein Data Bank. After an initial equilibration period of 80 ps, the *o*-xylene was then slowly converted into the non-physical reference state by slowly growing in the soft-interaction function over a period of 50 ps. After equilibration both systems were simulated for a total of 500 ps at a temperature of 300 K and a pressure of 1 atm. The temperature and pressure were maintained by weak coupling to an external bath using a relaxation time of 0.1 ps and 0.5 ps, respectively [25]. A twin-range cutoff for non-bonded interactions was used. Short-range (< 0.9 nm) Lennard-Jones and Coulombic interactions were updated at every time step (2 fs). Longer range Coulombic interactions (< 1.2 nm) were updated every 10 time steps at the same time as the pairlist generation. Bond lengths were constrained to reference values using the SHAKE procedure with a geometric tolerance of 10^{-4} [26]. All simulations were performed using a modified version of the GROMOS87 [27] simulation package in conjunction with a modified version of the force field as described by Smith *et al.* [28] Configurations from the trajectory were stored every 10 fs. The difference in free energy between the reference state and each of the five ligands shown in Fig. 4 was determined using the stored configurations. After reading each time frame, the soft-interaction sites were replaced with either a hydrogen atom or methyl group as appropriate for a given ligand. All possible permutations of the soft sites were considered. The interactions of each of the perturbed atoms within the appropriate cutoff range were then recalculated and the difference in free energy estimated by applying the perturbation formula (2) with $\delta\lambda = 1$. As only the interactions between the perturbed atoms and the rest of the system need be considered, this

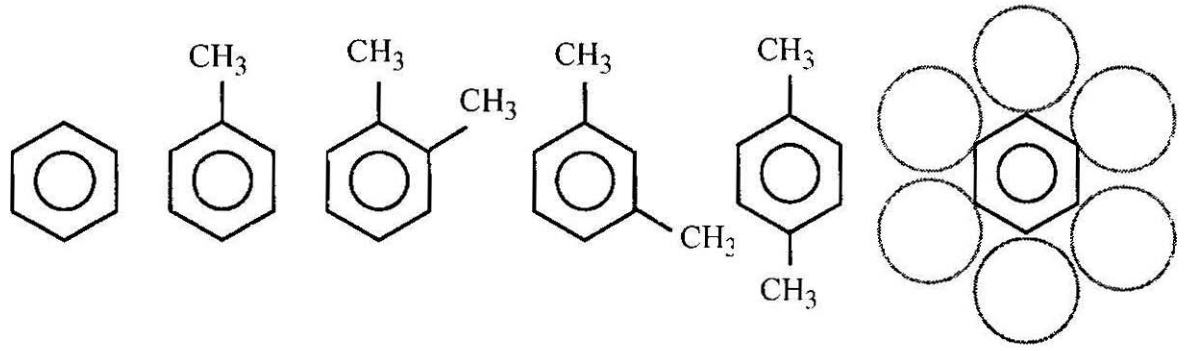

Fig. 4. Structural formula of the five aromatic ligands investigated in this study together with the structural formula of the reference compound (hydrogen atoms are not shown). The shaded lines represent the positions of soft interaction sites on the reference compound.

procedure is rapid and much more efficient than calculating the difference in free energy during the simulation.

RESULTS AND DISCUSSION

The difference in free energy for the mutation from the non-physical reference state to each of the five ligands shown in Fig. 4 in water are listed in Table 1. Also listed in Table 1 are the calculated differences in free energy relative to benzene and the experimental hydration free energies again relative to benzene. It should be noted that, because interactions within the reference molecule were excluded during the simulations, there is no vacuum correction in this case and the values in columns 3 and 4 can be compared directly.

The reliability of any free energy calculation is dependent on two factors. First, whether the calculations have converged. That is, whether the calculated free energy difference corresponds to the value given by the underling model or force field. Second, whether the force field reproduces the experimental free energy differences. Both factors are essential in practical applications. From Table 1 it is evident that, though the first condition may be met, the second is not. The results are in good agreement with experiment for toluene. For the series *m*-xylene, *p*-xylene and *o*-xylene the force field predicts a systematic increase in hydration free energy which is not observed experimentally. This is, however, clearly a problem of the force field and not of the extrapolation itself and must be accounted for when comparing the results in the protein to experiment. The dif-

Table 1

Relative free energy of hydration for the series of aromatic compounds (AC) shown in Fig. 4.

Values were derived from a single 500 ps simulation of a reference compound (RC) in water. The reference compound consisted of a benzene backbone with six soft-interaction sites, one attached to each carbon. The difference in free energy was estimated using the perturbation formula (2) with $\delta\lambda = 1$. Values are in kJ/mole. $\Delta G_{A\to B} \equiv G_{B\to B}$.

Compound	$\Delta G_{RC\to AC}$	$\Delta\Delta G_{benzene\to AC}$	$\Delta\Delta G_{benzene\to AC}$	$\Delta\Delta G_{benzene\to AC}$
	calc.	calc.	exper.[a]	exper. – calc.
Benzene	50.4	0.0	0.0	0.0
Toluene	52.3	1.9	2.1	0.2
m-Xylene	53.2	2.8	0.1	–2.7
p-Xylene	55.5	5.1	0.2	–4.9
o-Xylene	59.7	9.3	–0.2	–9.5

[a]Experimental values taken from Cabani *et al.* [29].

ference between the calculated and experimental free energy of hydration relative to benzene for each ligand is given in column 5 of Table 1.

Table 2 gives the difference in free energy between the reference state and each of the ligands bound to the protein (column 2) together with the relative difference in free energy compared to that of benzene (column 3). The difference between the relative free energy bound to the protein and the relative free energy in solution (column 4) is given in column 5. This number can be compared to the experimental binding free energies relative to the binding of benzene (column 6). The last column in Table 2 contains the calculated free energy of binding relative to that of benzene corrected for the difference between the calculated and the observed free energy of hydration. It is clear that the errors due to the force field in the calculated free energy of hydration for *m*-xylene, *p*-xylene and *o*-xylene dominate the relative binding free energy. Correcting for the difference in the calculated and experimental hydration free energies, the values for the relative binding free energies lie within 2.5 kJ/mole or 1 kT of the experimental values. The

Table 2

Relative free energy of binding for the series of aromatic compounds (AC) shown in Fig. 4 to T4-lysozyme. Values were derived from a single 500 ps simulation of a reference compound (RC) bound to T4-lysozyme. The difference in free energy was estimated using the perturbation formula (2) with $\delta\lambda = 1$. Values are in kJ/mole. $\Delta G_{A\to B} \equiv G_{B\to B}$.

Compound	$\Delta G_{RC\to AC}$ protein	$\Delta\Delta G_{benzene\to AC}$ protein	$\Delta\Delta G_{benzene\to AC}$ water	$\Delta\Delta G_{benzene\to AC}$ protein-water	$\Delta\Delta G_{benzene\to AC}$ protein-water	$\Delta\Delta G_{benzene\to AC}$ protein-water (corr.)
	calc.	calc.	calc.	calc.	exper.[a]	calc.
Benzene	76.7	0.0	0.0	0.0	0.0	0.0
Toluene	76.4	–0.3	1.9	–2.1	–1.4	–2.3
m-Xylene	76.2	–0.5	2.8	–3.3	1.8	–0.6
p-Xylene	79.2	2.5	5.1	–2.6	2.2	2.3
o-Xylene	80.7	4.0	9.3	–5.3	2.5	4.2

[a]Experimental values from Morton *et al.* [23, 24].

range in the experimental values for these initial compounds is, nevertheless, small and further derivatives need to be tested to demonstrate the utility of the approach. However, because the calculated relative binding free energy results from the difference between two large numbers, it is in fact very encouraging that the relative errors are so small.

Extrapolation based on a biased ensemble depends critically on whether configurations appropriate to the different end states are in fact sampled during the simulation of the nonphysical initial state. Fig. 5 shows an overlay of the initial starting configuration of T4-ly-

Fig. 5. A superposition of the initial starting configuration of T4-lysozyme (thick solid lines) and snapshots from the trajectory of the reference molecule (unfilled lines) at approximately 50 ps intervals from 0 ps to 450 ps. Also shown is the crystal orientation of o-xylene (filled lines).

sozyme and snapshots from the trajectory of the position of the reference molecule at approximately 50 ps intervals from 0 ps to 450 ps. As can be seen from Fig. 5, a wide variety of alternative binding configurations are in fact sampled by the reference compound during the simulation. The reference molecule remains close to the starting position for approximately the first 150 ps. It then moves away but returns to close to the initial location several times over the next 300 ps. Toward the end of the simulation it is drifting away again.

To estimate the difference in free energy for a particular ligand, alternative configurations are weighted appropriately by applying equation (2) with $\delta\lambda = 1$. The configurations with the highest weight correspond to low energy configurations for the specific ligand. This is illustrated in Fig. 6 for toluene. Figure 6 shows an overlay of the four lowest energy configurations for toluene extracted from the reference state simulation superimposed on the crystal structure of *p*-xylene. Although the reference compound samples a wide variety of alternate configurations, the low-energy configurations for toluene cluster in two degenerate well-defined orientations. This would indicate that there exist two possible binding orientations for toluene, one co-planar and the other perpendicular to the orientation of *p*-xylene, but in the same pocket.

CONCLUSIONS

Previously, it has been shown that results comparable to those from thermodynamic integration calculations using intermediate states can be obtained for mutations involving moderate charge rearrangements and the creation or deletion of atoms in water by applying the perturbation formula to an appropriately biased ensemble. The current study extends the scope of this work by applying the method to the estimation of relative binding affinities of a range of aromatic ligands bound to a hydrophobic cavity in T4-lysozyme. Despite the very preliminary nature of the results presented, it is evident that extrapolation can be used to predict changes in binding affinity for multiple physically relevant mutations from a single simulation in protein systems. We demonstrate that an appropriately chosen reference state will sample a variety of binding geometries during a molecular dynamics simulation and that, depending on the force field, this can be used to predict the most probable binding geometry for a given compound.

Extrapolation methods have the advantage that a large number of potential modifications can be investigated with a single calculation [12, 13]. In addition, because the extrapolation is based on an unperturbed ensemble, the nature of the perturbation does not have to be predefined and predictions for specific mutations can be efficiently obtained by reanalysis of existing trajectories [20]. Extrapolation based on the application of the perturbation formula to an appropriately biased ensemble holds considerable promise for use as a rapid, non-em-

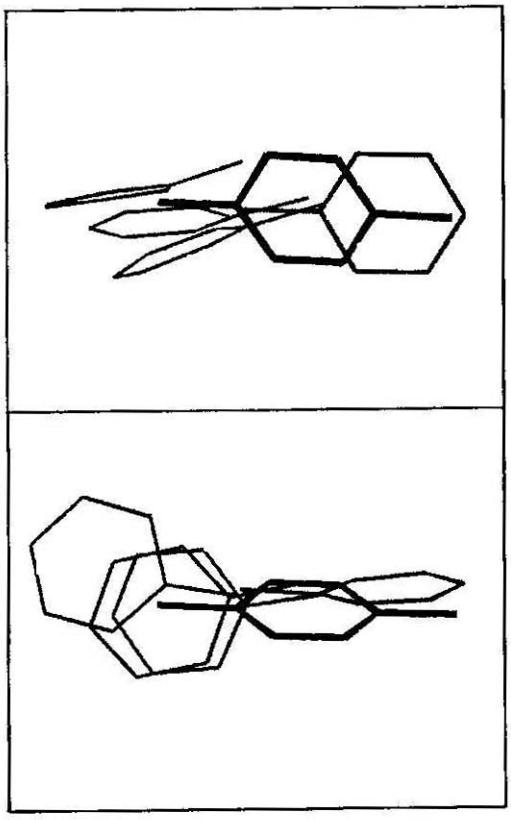

Fig. 6. An overlay of representatives of the four most probable (lowest energy) binding orientations of toluene complexed to T4-lysozyme.
The configurations were extracted from a 500 ps simulation of the T4-lysozyme — reference molecule complex and correspond to configurations at 234 ps, 448 ps, 461 ps, and 476 ps. Before overlaying the structures a best fit rotation was performed using all C_α atoms of the C-terminal sub-domain of T4-lysozyme. The relative orientation of *p*-xylene (thick lines) is also shown for comparison. The two views correspond to a rotation in the plane of 90 degrees.

pirical means of estimating relative binding affinities for a wide range of possible chemical modifications of a lead compound. As such, the approach could be used to guide experimental planning in drug design, a task not practical using normal free energy calculations.

REFERENCES

1. Kubinyi, H. (1993) QASR: *Hansch analysis and related approaches*. VCH, Weinheim, Germany.
2. Martin, Y.C. (1978) *Quantitative drug design: A critical introduction*. Marcel Dekker, New York.
3. Kuntz, I.D., Meng, E.C. & Shoichet, B.K. (1994) Structure-based molecular design. *Acc. Chem. Res.* **27**, 117–123.
4. Colman, P.M. (1994) Structure-based drug design. *Curr. Opin. Struct. Biol.* **4**, 868–874.
5. van Gunsteren, W.F., King, P.M. & Mark, A.E. (1994) Fundamentals of drug design from a biophysical viewpoint. *Quart. Rev. Biophys.* **27**, 435–481.
6. Mark, A.E. & van Gunsteren, W.F. (1994) Decomposition of the free energy of a system in terms specific interactions: Implications for theoretical and experimental studies. *J. Mol. Biol.* **240**, 167–176.
7. Beveridge, D.L. & Di Capua, F.M. (1989) Free energy *via* molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492.
8. Kirkwood, J.G. (1935) Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300–313.
9. Zwanzig, R.W. (1954) High-temperature equation of state by a perturbation method: I. Non--polar gases. *J. Chem. Phys.* **22**, 1420–1426.
10. Jorgensen, W.L. (1991) Computational insights on intermolecular interactions and binding in solution. *Chemtracts: Org. Chem.* **4**, 91–119.
11. van Gunsteren, W.F., Beutler, T.C., Fraternali, F., King, P.M., Mark, A.E. & Smith, P.E. (1993) Computation of free energy in practice: Choice of approximations and accuracy limiting factors; in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications* (van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., eds.) vol. 2, pp. 315–348, ESCOM Science Publi- shers, Leiden, The Netherlands.
12. Gerber, P.R., Mark, A.E. & van Gunsteren, W.F. (1993) An approximate but efficient method to calculate free energy trends by computer simulation: Application to dihydrofolate reductase-inhibitor complexes. *J. Computer-Aided Mole- cular Design* **7**, 305–323.
13. Smith, P.E. & van Gunsteren, W.F. (1994) Predictions of free energy differences from a single simulation of the initial state. *J. Chem. Phys.* **100**, 577–585.
14. Levy, R.M., Belhadj, M. & Kitchen, D.B. (1991) Gaussian fluctuation formula for electrostatic free-energy changes. *J. Chem. Phys.* **95**, 3627–-3633.
15. King, G. & Barford, R.A. (1993) Calculation of electrostatic free energy differences with a time-saving approximate method. *J. Phys. Chem.* **97**, 8798–8802.
16. Aqvist, J., Medina, C. & Samuelsson, J.-E. (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385–391.
17. Gilson, M.K. & Honig, B. (1988) Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins* **4**, 7–18.
18. Sharp, K.A. & Honig, B. (1990) Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301–332.
19. Sharp, K.A. (1994) Electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **4**, 234–-239.
20. Liu, H., Mark, A.E. & van Gunsteren, W.F. (1995) On estimating the relative free energy of different molecular states with respect to a single reference state. *J. Phys. Chem.* (in press).
21. Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R. & van Gunsteren, W.F. (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulation. *Chem. Phys. Lett.* **222**, 529–539.
22. Morton, A., Baase, W.A. & Matthews, B.W. (1995) Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* (in press).
23. Morton, A. & Matthews, B.W. (1995) Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry* (in press).
24. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. & Hermans, J. (1981) Interaction models for water in relation to protein hydration; in *Intermolecular Forces* (Pullman, B., ed.) pp. 331–342, Reidel, Dordrecht, The Netherlands.

25. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. & Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690.

26. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. (1977) Numerical integration of the Cartesian equations of motion with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341.

27. van Gunsteren, W.F. & Berendsen, H.J.C. (1987) Groningen Molecular Simulation Library Manual, Biomos, University of Groningen, The Netherlands.

28. Smith, L.J., Mark, A.E., Dobson, C.M. & van Gunsteren, W.F. (1995) Comparison of MD simulations and NMR experiments for hen lysozyme: Analysis of local fluctuations, cooperative motions, and global changes. *Biochemistry* **34,** 10918–10918.

29. Cabani, S., Gianni, P., Mollica, V. & Lepori, L. (1981) Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Soln. Chem.* **10,** 563–595.