CHAPTER ELEVEN

Four Challenges in Molecular Modelling: Free Energies, Solvation, Reactions and Solid-state Defects

In this chapter we shall consider four important problems in molecular modelling. First, we discuss the problem of calculating free energies. We then consider continuum solvent models, which enable the effects of the solvent to be incorporated into a calculation without requiring the solvent molecules to be represented explicitly. Third, we shall consider the simulation of chemical reactions, including the important technique of *ab initio* molecular dynamics. Finally, we consider how to study the nature of defects in solid-state materials.

11.1 Free Energy Calculations

11.1.1 The Difficulty of Calculating Free Energies by Computer

The free energy is often considered to be the most important quantity in thermodynamics. The free energy is usually expressed as the Helmholtz function, A, or the Gibbs function, G. The Helmholtz free energy is appropriate to a system with constant number of particles, temperature and volume (constant NVT), whereas the Gibbs free energy is appropriate to constant number of particles, temperature and pressure (constant NPT). Most experiments are conducted under conditions of constant temperature and pressure, where the Gibbs function is the appropriate free energy quantity.

Unfortunately, the free energy is a difficult quantity to obtain for systems such as liquids or flexible macromolecules that have many minimum energy configurations separated by low-energy barriers. Associated quantities such as the entropy and the chemical potential are also difficult to calculate. As we showed in Section 6.3, the free energy cannot be accurately determined from a 'standard' molecular dynamics or Monte Carlo simulation, because such simulations do not adequately sample from those regions of phase space that make important contributions to the free energy. Specifically, we showed that the Helmholtz free energy is given by:

$$A = k_B T \ln \left(\iint d\mathbf{p}^N \, d\mathbf{r}^N \exp \left(\frac{+\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T} \right) \rho(\mathbf{p}^N, \mathbf{r}^N) \right) \tag{11.1}$$

The term $\exp[+\mathscr{H}(\mathbf{p}^N,\mathbf{r}^N)/k_BT]$ makes important contributions to the integral. However, a simulation using either Monte Carlo or molecular dynamics sampling seeks out the *lower-energy* regions of phase space. Such simulations will never adequately sample the important high-energy regions and so to calculate the free energy using a conventional simulation will lead to poorly converged and inaccurate values. The grand canonical and particle-insertion methods do provide a route to the free energy, but they are not applicable to many of the systems of interest, which contain complex molecules at high densities.

11.2 The Calculation of Free Energy Differences

Let us consider a closely related but slightly different problem: the calculation of the free energy difference of two states. As an example, we will consider the problem of calculating the free energy difference between ethanol (CH_3CH_2OH) and ethane thiol (CH_3CH_2SH) in water. As we shall see, this is a problem that can be tackled using methods that use Monte Carlo or molecular dynamics sampling. Three methods have been proposed for calculating free energy differences: thermodynamic perturbation, thermodynamic integration and slow growth. We shall consider each of these in turn.

11.2.1 Thermodynamic Perturbation

Consider two well-defined states X and Y. For example, X could be a system comprising a molecule of ethanol in a periodic box of water and Y could be ethane thiol in water. X contains N particles interacting according to the Hamiltonian $\mathscr{H}_X$ Y contains N particles interacting according to $\mathscr{H}_Y$ The free energy difference (ΔA) between the two states is as follows:

$$\Delta A = A_Y - A_X = -k_BT \ln \frac{Q_Y}{Q_X} \tag{11.2}$$

$$\Delta A = -k_BT \left\{ \frac{\int\int d\mathbf{p}^N d\mathbf{r}^N \exp[-\mathscr{H}_Y(\mathbf{p}^N,\mathbf{r}^N)/k_BT]}{\int\int d\mathbf{p}^N d\mathbf{r}^N \exp[-\mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT]} \right\} \tag{11.3}$$

Substituting 1 in the form $\exp[+\mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT]\exp[-\mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT]$ into the numerator gives:

$$\Delta A = -k_BT \left\{ \frac{\int\int d\mathbf{r}^N d\mathbf{p}^N \exp\left(-\frac{\mathscr{H}_Y(\mathbf{r}^N,\mathbf{p}^N)}{k_BT}\right) \exp\left(+\frac{\mathscr{H}_X(\mathbf{r}^N,\mathbf{p}^N)}{k_BT}\right) \exp\left(-\frac{\mathscr{H}_X(\mathbf{r}^N,\mathbf{p}^N)}{k_BT}\right)}{\int\int d\mathbf{r}^N d\mathbf{p}^N \exp\left(\frac{-\mathscr{H}_X(\mathbf{r}^N,\mathbf{p}^N)}{k_BT}\right)} \right\} \tag{11.4}$$

Equation (11.4) can be written in terms of an ensemble average, as follows:

$$\Delta A = -k_BT \left\{ \frac{\int\int d\mathbf{p}^N d\mathbf{r}^N \exp[-\mathscr{H}_Y(\mathbf{p}^N,\mathbf{r}^N)/k_BT] \exp[+\mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT] \exp[-\mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT]}{\int\int d\mathbf{p}^N d\mathbf{r}^N \exp[-\mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT]} \right\}$$

$$= -k_BT \langle \exp -[\mathscr{H}_Y(\mathbf{p}^N,\mathbf{r}^N) - \mathscr{H}_X(\mathbf{p}^N,\mathbf{r}^N)/k_BT] \rangle_0 \tag{11.5}$$

The subscript 0 indicates averaging over the ensemble of configurations representative of the initial state X. If the averaging is over the ensemble corresponding to the final state Y (indicated by the subscript 1) then we are effectively simulating the reverse process, from which ΔA can be determined by:

$$\Delta A = k_B \ln\langle \exp[-(\mathscr{H}_X - \mathscr{H}_Y)/k_B T]\rangle_1 \tag{11.6}$$

This approach to the calculation of free energy differences, Equation (11.6), is generally attributed to Zwanzig [Zwanzig 1954]. To perform a thermodynamic perturbation calculation we must first define $\mathscr{H}_Y$ and $\mathscr{H}_X$ and then run a simulation at the state X, forming the ensemble average of $\exp[-(\mathscr{H}_Y - \mathscr{H}_X)/k_B T]$ as we proceed. Analogously, we could run a simulation at the state Y and obtain the ensemble average of $\exp[-(\mathscr{H}_X - \mathscr{H}_Y)/k_B T]$. Thus, if X corresponds to ethanol and Y to ethane thiol, the free energy difference could be obtained from a simulation of ethanol in a periodic box of water as follows. For each configuration we calculate the value of the energy for every instantaneous conformation of ethanol in which the oxygen atom is temporarily assigned the potential energy parameters of sulphur. Alternatively, we could simulate ethane thiol and for each configuration calculate the energy of the system in which the sulphur is 'mutated' into oxygen.

If X and Y do not overlap in phase space then the value of the free energy difference calculated using Equation (11.6) will not be very accurate, because we will not adequately sample the phase space of Y when simulating X. This problem arises when the energy difference between the two states is much larger than $k_B T$: $|\mathscr{H}_Y - \mathscr{H}_X| \gg k_B T$. How then can we obtain accurate estimates of the free energy difference under such circumstances? Consider what happens if we introduce a state that is intermediate between X and Y, with a Hamiltonian $\mathscr{H}_1$ and a free energy $A(1)$:

$$\begin{aligned}
\Delta A &= A(Y) - A(X) \\
&= (A(Y) - A(1)) + (A(1) - A(X)) \\
&= -k_B T \ln\left[\frac{Q(Y)}{Q(1)} \cdot \frac{Q(1)}{Q(X)}\right] \\
&= -k_B T \ln\langle \exp[-(\mathscr{H}_Y - \mathscr{H}_1)/k_B T]\rangle - k_B T \ln\langle \exp[-(\mathscr{H}_1 - \mathscr{H}_X)/k_B T]\rangle
\end{aligned} \tag{11.7}$$

If we define region 1 so that it overlaps with X and Y we may improve the sampling and obtain a more reliable value. This is shown in Figure 11.1.

An obvious extension is to use several different intermediate states in progressing from $\mathscr{H}_X$ to $\mathscr{H}_Y$:

$$\begin{aligned}
\Delta A &= A(Y) - A(X) \\
&= (A(Y) - A(N)) + (A(N) - A(N-1)) + \cdots \\
&\quad + (A(2) - A(1)) + (A(1) - A(X)) \\
&= -k_B T \ln\left[\frac{Q(Y)}{Q(N)} \cdot \frac{Q(N)}{Q(N-1)} \cdot \frac{Q(N-1)}{Q(N-2)} \cdots \frac{Q(2)}{Q(1)} \frac{Q(1)}{Q(X)}\right]
\end{aligned} \tag{11.8}$$

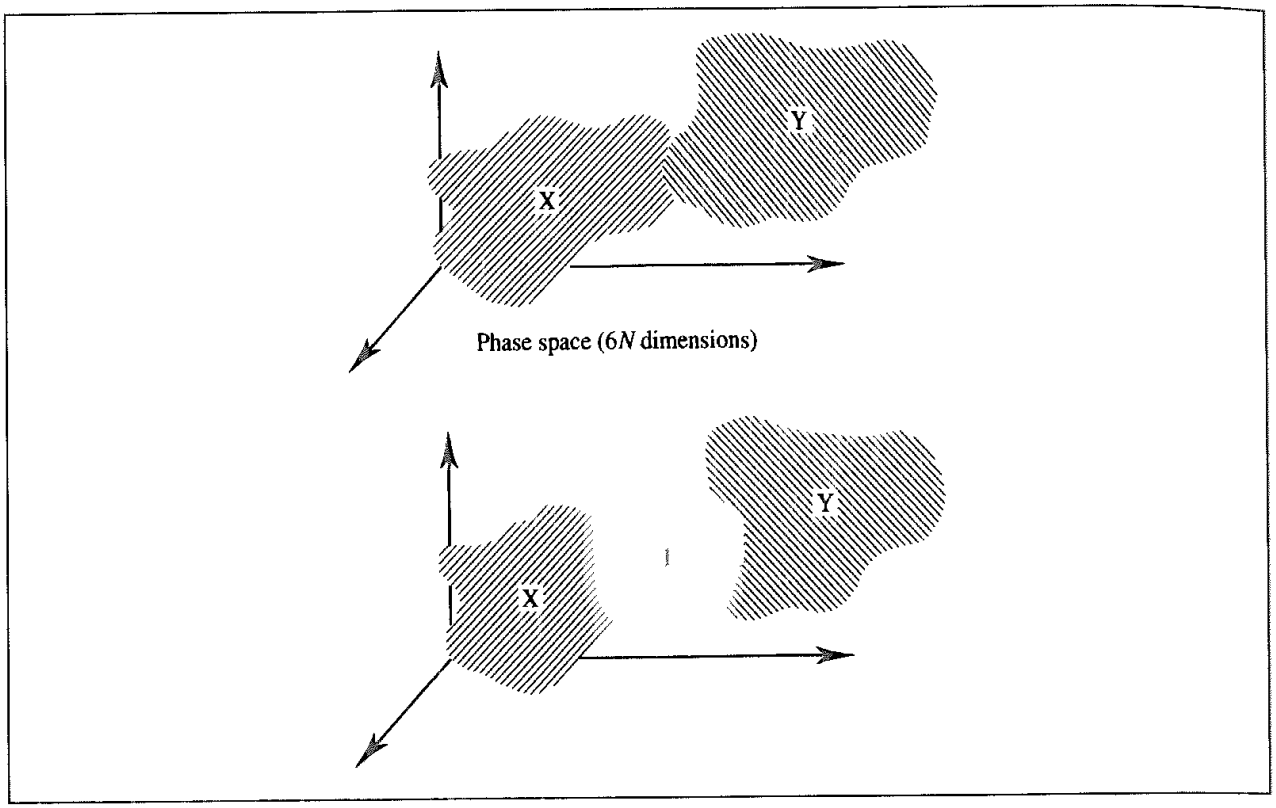

Fig 11.1 An intermediate state (labelled 1) can improve the degree of overlap in phase space and lead to improved sampling

The important point to notice is that the intermediate terms cancel out and so we are free to choose as many intermediate states as are necessary to get good overlaps and thus reliable values of the free energy differences.

11.2.2 Implementation of Free Energy Perturbation

Suppose we are using an empirical energy function such as the following to describe the inter- and intramolecular interactions in our ethanol/ethane thiol system:

$$\mathcal{V}(\mathbf{r}^N) = \sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \tag{11.9}$$

The force-field model for ethanol contains C–O and O–H bond-stretching contributions; in ethane thiol these are replaced by C–S and S–H parameters. Similarly, in ethanol there will be angle-bending terms due to C–O–H, C–C–O and H–C–O angles; in ethane thiol these will be C–S–H, C–C–S and H–C–S. The torsional contribution will be modified appropriately, as will the van der Waals and electrostatic interactions (both those within the

solute and between the solute and solvent). The partial atomic charges for all of the atoms in ethanol may all be different from those for ethane thiol.

The relationship between the initial, final and intermediate states is usefully described in terms of a *coupling parameter*, λ. As λ is changed from 0 to 1, the Hamiltonian varies from $\mathscr{H}_X$ to $\mathscr{H}_Y$. Each of the terms in the force field for an intermediate state λ can be written as a linear combination of the values for X and Y:

1. Bonds:

$$k_l(\lambda) = \lambda k_l(\mathrm{Y}) + (1-\lambda)k_l(\mathrm{X}) \tag{11.10}$$

$$l_0(\lambda) = \lambda l_0(\mathrm{Y}) + (1-\lambda)l_0(\mathrm{X}) \tag{11.11}$$

2. Angles:

$$k_\theta(\lambda) = \lambda k_\theta(\mathrm{Y}) + (1-\lambda)k_\theta(\mathrm{X}) \tag{11.12}$$

$$\theta_0(\lambda) = \lambda \theta_0(\mathrm{Y}) + (1-\lambda)\theta_0(\mathrm{X}) \tag{11.13}$$

3. Dihedrals:

$$v_\omega(\lambda) = \lambda v_\omega(\mathrm{Y}) + (1-\lambda)v_\omega(\mathrm{X}) \tag{11.14}$$

4 Electrostatics:

$$q_i(\lambda) = \lambda q_i(\mathrm{Y}) + (1-\lambda)q_i(\mathrm{X}) \tag{11.15}$$

5. van der Waals:

$$\varepsilon(\lambda) = \lambda \varepsilon(\mathrm{Y}) + (1-\lambda)\varepsilon(\mathrm{X}) \tag{11.16}$$

$$\sigma(\lambda) = \lambda \sigma(\mathrm{Y}) + (1-\lambda)\sigma(\mathrm{X}) \tag{11.17}$$

For each value of λ (λ_i) a simulation is performed (using either Monte Carlo or molecular dynamics as appropriate) with the appropriate force field parameters. First, the system is equilibrated using the force field parameters appropriate to λ_i. A production phase is then performed during which the free energy difference $\Delta A(\lambda_i \rightarrow \lambda_{i+1})$ is accumulated as $-k_B T \ln\langle \exp(-\Delta\mathscr{H}_i/k_B T)\rangle$, where $\Delta\mathscr{H}_i = \mathscr{H}_{i+1} - \mathscr{H}_i$. The total free energy change for $\lambda = 0$ to $\lambda = 1$ is then the sum of the free energy changes for the various values of λ_i, as shown in Figure 11.2.

The approach that we have described so far is known as *forward sampling*, because the free energy is determined for $\lambda_i \rightarrow \lambda_{i+1}$. In *backward sampling*, the free energy difference between λ_i and λ_{i-1} is determined. The coupling parameter λ still increases from 0 to 1; it is just that the free energies are accumulated in a different manner. In *double-wide sampling*, the free energy differences for both $\lambda_i \rightarrow \lambda_{i+1}$ and $\lambda_i \rightarrow \lambda_{i-1}$ are obtained from a simulation as

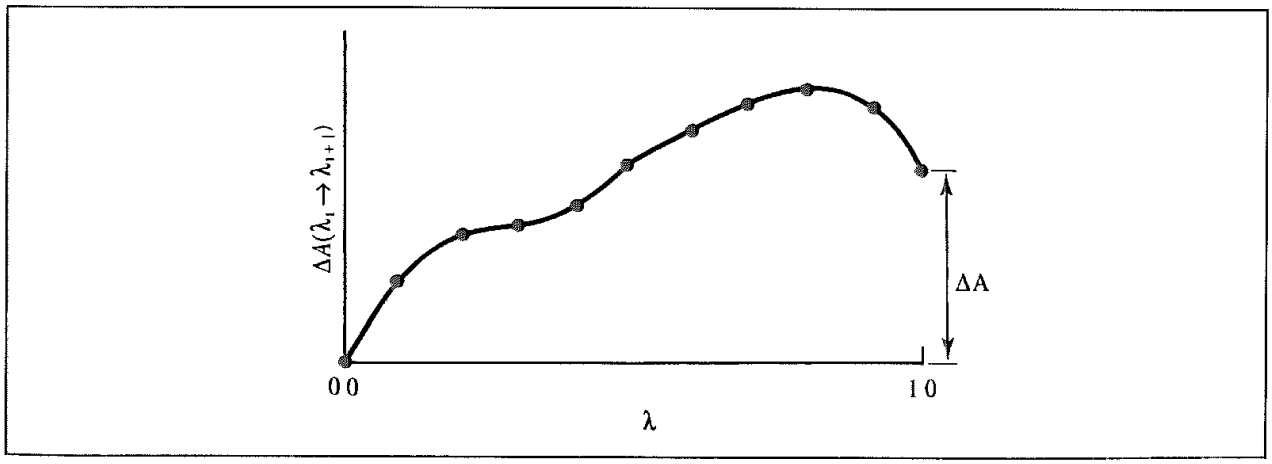

Fig. 11 2: Calculation of the free energy difference using perturbation

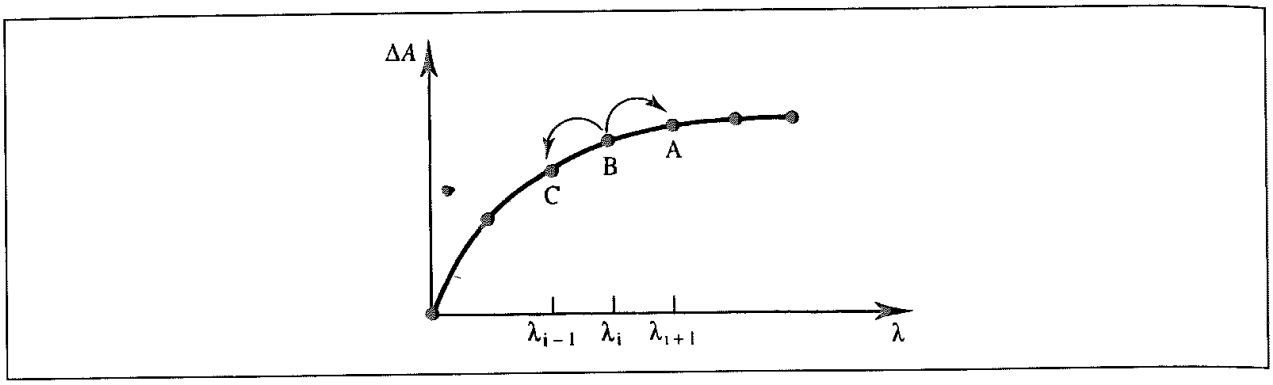

Fig 11 3 Double wide sampling enables two free energies to be accumulated from a single simulation.

illustrated in Figure 11.3. Consider point B in Figure 11.3, which corresponds to a coupling parameter λ_i. A simulation performed using λ_i can be used to obtain values for both the free energy difference $\Delta A(\lambda_i \rightarrow \lambda_{i+1})$ and the free energy difference $\Delta A(\lambda_i \rightarrow \lambda_{i-1})$. This is clearly a more efficient way to obtain the desired free energy as twice as many free energy differences can be obtained from a single simulation.

11.2.3 Thermodynamic Integration

An alternative way to calculate the free energy difference uses thermodynamic integration. The formula for the free energy difference is derived in Appendix 11.1 and is:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{11 18}$$

To calculate a free energy difference using thermodynamic integration we thus need to determine the integral in Equation (11.18). In practice, this is achieved by performing a series of simulations corresponding to discrete values of λ between 0 and 1. For each value of λ, the average of

$$\left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \right\rangle_\lambda \tag{11 19}$$

is determined. These partial derivatives are calculated analytically in some programs but in others a finite difference approximation is used ($\partial \mathcal{H}/\partial \lambda \approx \Delta \mathcal{H}/\Delta \lambda$). The total free energy difference ΔA is then equal to the area under the graph of

$$\left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \right\rangle_\lambda \tag{11.20}$$

versus λ, as illustrated in Figure 11.4.

11.2.4 The 'Slow Growth' Method

A third approach for the calculation of free energy differences from computer simulation is the slow growth method. Here, the Hamiltonian changes by a very small, constant amount at

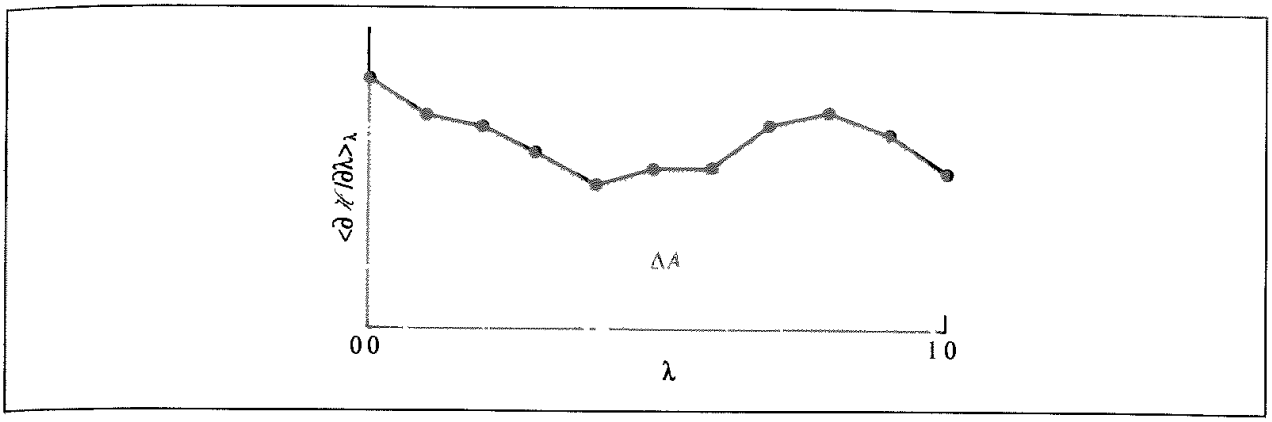

Fig 11 4: Calculation of free energy differences by thermodynamic integration

each step of the calculation. This means that at each stage the Hamiltonian $\mathcal{H}(\lambda_{i+1})$ is very nearly equal to $\mathcal{H}(\lambda_i)$. The free energy difference is given by:

$$\Delta A = \sum_{i=1,\lambda=0}^{i=N_{\text{step}};\lambda=1} (\mathcal{H}_{i+1} - \mathcal{H}_i) \tag{11.21}$$

This expression is derived in Appendix 11.2.

In principle, all three methods for calculating the free energy difference should give the same result, as the free energy is a state function and so independent of path. However, there may be practical reasons for choosing one method or another, as we shall discuss in Section 11.6. The other point to note at this stage is that our formulation of the free energy has been in terms of the partition function Q and the Hamiltonian $\mathcal{H}(\mathbf{p}^N,\mathbf{r}^N)$, which have contributions from both kinetic and potential energies. When the kinetic energy contributions are integrated out they cancel and so the various equations can be written in terms of the difference between the potential function, $\mathcal{V}(\mathbf{r}^N)$, rather than the Hamiltonian, $\mathcal{H}(\mathbf{p}^N,\mathbf{r}^N)$. Q is then replaced by the configurational integral, Z, and the free energy values that are obtained are excess free energies, relative to an ideal gas.

Our discussion so far has considered the calculation of Helmholtz free energies, which are obtained by performing simulations at constant *NVT*. For proper comparison with experimental values we usually require the Gibbs free energy, G. Gibbs free energies are obtained from a simulation at constant *NPT*.

11.3 Applications of Methods for Calculating Free Energy Differences

11.3.1 Thermodynamic Cycles

An early application of the free energy perturbation method was the determination of the free energy required to create a cavity in a solvent. Postma, Berendsen and Haak determined the free energy to create a cavity ($\lambda = 1$) in pure water ($\lambda = 0$) using isothermal-isobaric

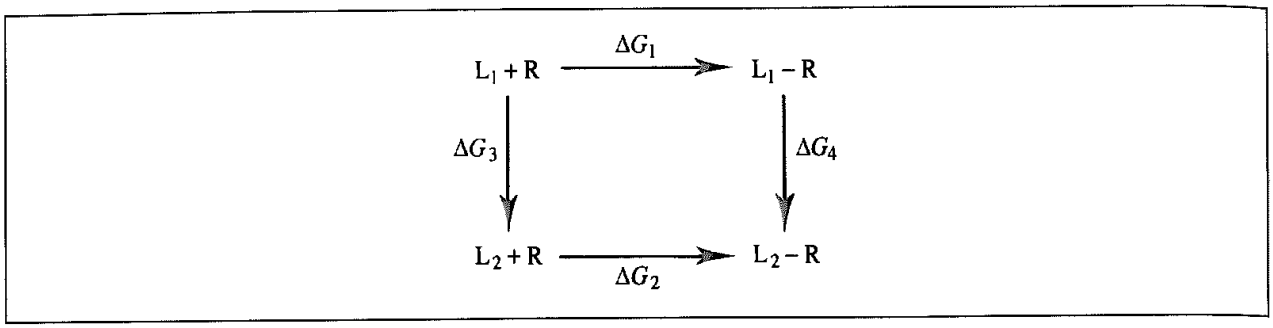

Fig. 11.5 Thermodynamic cycle for binding ligands L_1 and L_2 to receptor R

molecular dynamics simulations [Postma *et al.* 1982]. Five different cavity sizes were considered and the results showed that, as expected, the free energy of cavity formation increased with the size of the cavity, the results being in good agreement with analytical theories. For small cavities (< 1 Å radius) the results were inaccurate due to poor sampling. The calculations provided not only the free energy of cavity formation for the different cavity sizes but also structural and dynamic properties of the water molecules around the cavity. For example, the water structure varied with the cavity size. A cavity of radius 1.78 Å had the most pronounced shell structure, with a high first-neighbour peak and a significant second-neighbour peak in the cavity–water pair distribution function.

Many processes of interest to the molecular modeller involve an equilibrium between molecules that interact via non-covalent forces, the free energy being related to the equilibrium constant by $\Delta G = -RT \ln K$. Let us consider the binding of two different ligands (L_1 and L_2) to a receptor molecule (R). L_1 and L_2 could be putative inhibitors of an enzyme R or two 'guests' for a host R. The thermodynamic cycle for the two binding processes is shown in Figure 11.5. The relative binding affinity of L_1 and L_2 equals $\Delta G_2 - \Delta G_1$ and is commonly written $\Delta\Delta G$. In principle, it would be possible to calculate values of ΔG_1 and ΔG_2 by simulating the actual association process. To do this we would bring the ligand and the receptor together from an initial large separation to gradually form the intermolecular complex. However, in most cases this would involve such a major reorganisation of the receptor, the ligand and the solvent that it would be difficult to ensure adequate sampling of phase space.

The free energy is a state function, and so its value round a thermodynamic cycle must be zero. Thus $\Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$ (Figure 11.5). ΔG_3 corresponds to the free energy difference of the two ligands in solution; ΔG_4 is the free energy difference of the two intermolecular complexes. The changes ΔG_3 and ΔG_4 do not correspond to any transformation that can be performed in the laboratory, but they are quite feasible in the computer. The free energy difference only depends upon the endpoints, and so we are at liberty to change the Hamiltonians in any way we wish. The free energy differences obtained from such non-physical pathways are likely to be much more reliable than the 'physically plausible' processes as they should involve much less reorganisation of the system This is particularly so if the two ligands L_1 and L_2 have similar structures. To calculate the relative free energy of binding of the two ligands we would therefore 'mutate' L_1 into L_2 in solution and L_1 to L_2 within the receptor. This is the *thermodynamic cycle perturbation approach* to calculating relative free energies.

11.3.2 Applications of the Thermodynamic Cycle Perturbation Method

One of the first applications of the thermodynamic cycle perturbation approach to the calculation of relative binding constants was the study by Lybrand, McCammon and Wipff of the synthetic macrocycle SC24, which, when protonated, can bind halide ions (Figure 11 6) [Lybrand *et al.* 1986]. SC24 binds Cl^- 4 30 kcal/mol more strongly than Br^-. Two simulations were performed to determine a theoretical value for this relative free energy using the free energy perturbation method with molecular dynamics. First, Cl^- was mutated to Br in aqueous solution, giving a free energy difference of 3.35 kcal/mol. The same mutation was then performed within the macrocycle, in a periodic box of water. The value obtained for this step was 7.50 kcal/mol, giving an overall relative free energy of binding of 4.15 kcal/mol. The experimental value was approximately 4.3 kcal/mol. Thus, although the free energy to desolvate Cl^- is unfavourable compared with Br^-, this is more than compensated for by favourable interactions between Cl^- and the host; Br^- is slightly too large to fit comfortably in the relatively inflexible SC24 molecule.

One of the most attractive applications of the free energy techniques is for predicting the relative free energies of binding of inhibitors of biological macromolecules such as proteins or DNA. If we know the binding constant of an inhibitor then we can, in principle at least, calculate the binding constant of a related inhibitor. The free energy cycle used to perform this calculation is analogous to that shown in Figure 11.5: we perform two separate free energy calculations: ligand L_1 is mutated to ligand L_2 in solution and within the binding site An early calculation of this type was performed by Bash and colleagues, who studied two inhubitors of thermolysin (an enzyme which cleaves the amide bonds in peptides and proteins) [Bash *et al.* 1987]. The two inhibitors investigated had the general formula carbobenzoxy-Gly^P(X)-L-Leu-L-Leu [Barlett and Marlowe 1987] (Figure 11.7). The experimentally determined binding constants (K_i) of the X ≡ NH and X ≡ O inhibitors were 9.1 nM and 9000 nM, i.e. the former binds 1000 times more strongly. This difference in binding constants is equivalent to 4.1 kcal/mol. X-ray crystallographic analysis showed that the two inhibitors bind in almost identical positions. The calculated free energy difference was determined to

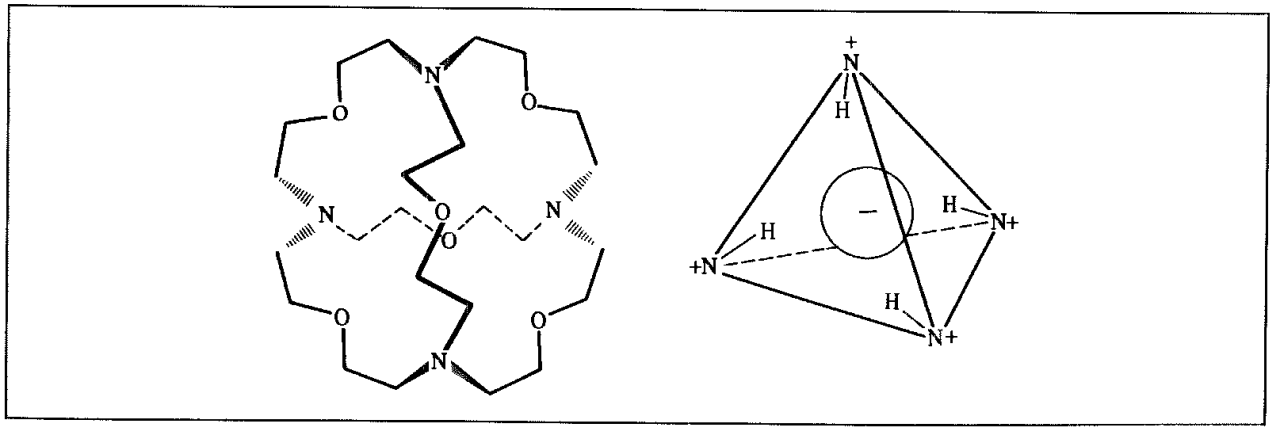

Fig 11 6: The SC24/halide system. (Figure adapted from Lybrand T P, J A McCammon and G Wipff 1986 Theoretical Calculation of Relative Binding Affinity in Host–Guest Systems. Proceedings of the National Academy of Sciences USA *83 833–835)*

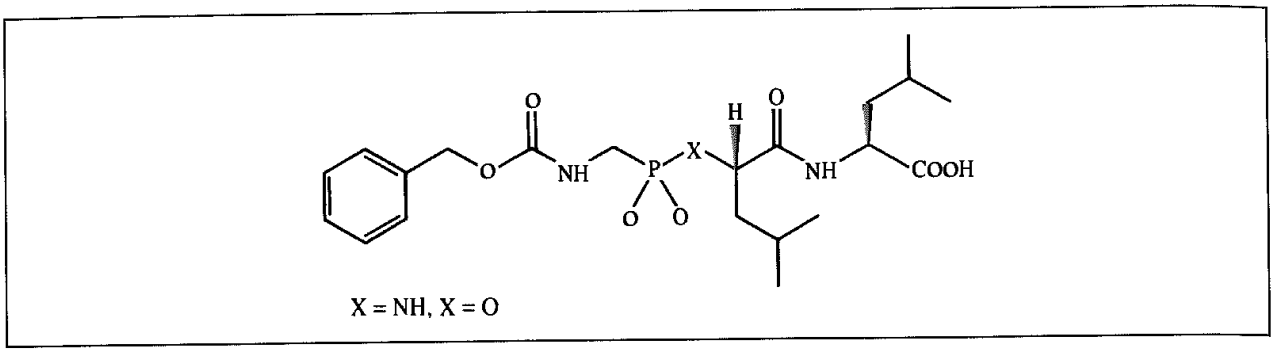

Fig. 11.7. Thermolysin inhibitors [Bartlett and Marlowe 1987]

be 4.2 ± 0.5 kcal/mol, in good agreement with the experimental result. In the active site of the enzyme the X group of the inhibitor interacts with the backbone carbonyl oxygen of one of the amino acids (Ala 113). The ester oxygen interacts unfavourably with this carbonyl, but the amide can form a hydrogen bond. The relative free energy of binding of the amide inhibitor to the protein was calculated to be 7.6 kcal/mol lower than the ester, but this was counteracted by the difference in the free energies of solvation, which was calculated to be 3.4 kcal/mol. The amide inhibitor thus incurs a greater desolvation penalty than the ester.

This study obviously gave very satisfactory agreement with the experimental data. However, a subsequent calculation by Merz and Kollman showed that the results were very sensitive to the charge model used for the inhibitor [Merz and Kollman 1989]. The charges for the inhibitor were obtained by electrostatic potential fitting in each case, though with different basis sets. This second calculation gave a free energy difference of 5.9 kcal/mol. Other studies have also shown that calculated free energies can be very sensitive to the charge model used; we will discuss some of the problems with performing free energy calculations in Section 11.6.

As one final example of the application of free energy calculations we will examine the determination of relative partition coefficients. The partition coefficient (P) is the equilibrium constant for the transfer of a solute between two solvents. The logarithm of the partition coefficient ($\log P$) for transfer between water and a variety of solvents (primarily 1-octanol) is widely used to derive structure–activity relationships (see Section 12.9), in which the biological activity of a molecule is correlated with its physicochemical properties. The thermodynamic cycle for the partition of two solutes, A and B, between two solvents is shown in Figure 11.8. If it were possible to calculate the free energy of transfer from one solvent to another (i.e. ΔG_1 or ΔG_2 in Figure 11.8) then this would give the partition coefficient directly. However, such a simulation would require an inordinate amount of time and probably be very inaccurate. A relative partition coefficient can be determined by mutating one solute into the other in the two separate solvents.

Calculations of relative partition coefficients have been reported using the free energy perturbation method with the molecular dynamics and Monte Carlo simulation methods. For example, Essex, Reynolds and Richards calculated the difference in partition coefficients of methanol and ethanol partitioned between water and carbon tetrachloride with molecular dynamics sampling [Essex *et al.* 1989]. The results agreed remarkably well with experiment

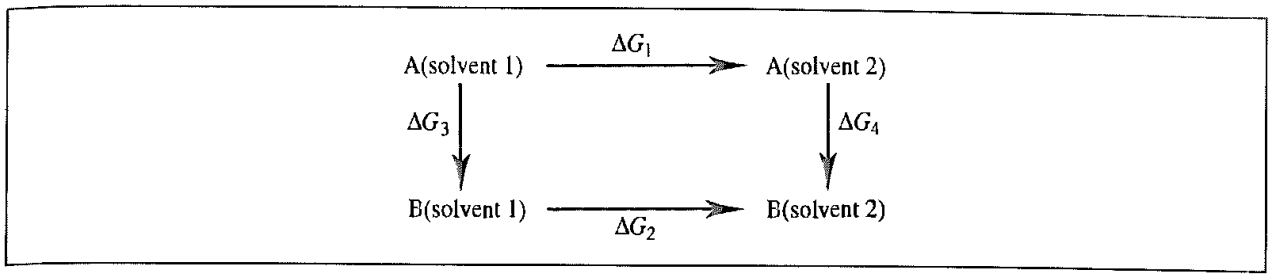

Fig 11.8: Thermodynamic cycle for calculating relative partition coefficients

(within 0.06 kcal/mol). Jorgensen, Briggs and Contreras used Monte Carlo methods to calculate the relative partition coefficients for eight pairs of solutes (including methanol/methylamine, acetic acid/acetamide and pyrazine/pyridine) between water and chloroform [Jorgensen *et al.* 1990]. For these eight systems good qualitative agreement with experimental data was obtained. However, the results involving acetic acid gave too broad a spread of values. This was traced to the relative free energies of hydration, which varied over too wide a range and indicated some areas for improvement in the force field model.

11.3.3 The Calculation of Absolute Free Energies

In some cases, it is possible to devise thermodynamic cycles which enable the absolute free energy of a change to be determined using free energy perturbation methods [Jorgensen *et al.* 1988]. Figure 11.9 shows a thermodynamic cycle for the association of L and R to give a complex LR in both the gas phase and in solution. ΔG_{ass} is the free energy of association in solution and is given by:

$$\Delta G_{ass} = \Delta G_{gas}(L + R \rightarrow LR) + \Delta G_{sol}(LR) - \Delta G_{sol}(L) - \Delta G_{sol}(R) \qquad (11.22)$$

$\Delta G_{sol}(X)$ is the solvation free energy of the species X (the free energy of transfer from the gas phase to solvent). The solvation free energy can be written in terms of perturbations where the species disappear to nothing in the gas phase and in solution, $\Delta G_{sol}(X) = \Delta G_{gas}(X \rightarrow 0) - \Delta G_{sol}(X \rightarrow 0)$. The free energy of association, ΔG_{ass}, can then be written:

$$\begin{aligned}\Delta G_{ass} = {} & \Delta G_{gas}(L + R \rightarrow LR) - \Delta G_{gas}(L \rightarrow 0) + \Delta G_{sol}(L \rightarrow 0) \\ & - \Delta G_{gas}(R \rightarrow 0) + \Delta G_{sol}(R \rightarrow 0) + \Delta G_{gas}(LR \rightarrow 0) \\ & - \Delta G_{sol}(LR \rightarrow 0) \end{aligned} \qquad (11.23)$$

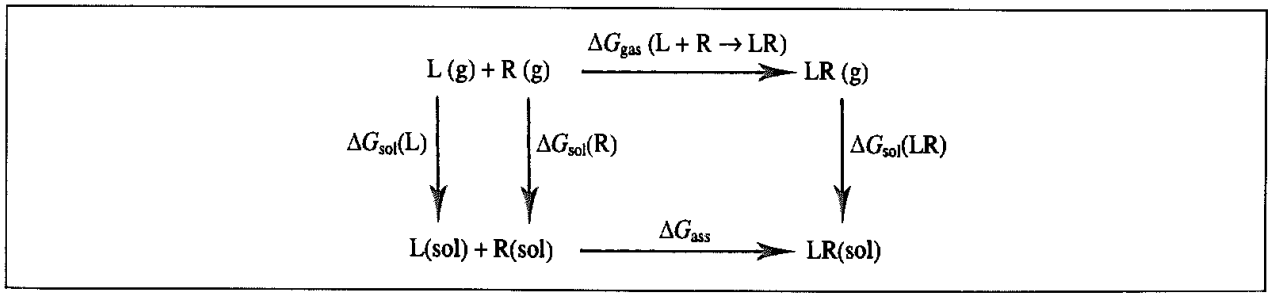

Fig 11.9 Thermodynamic cycle used to calculate absolute free energies [Jorgensen et al *1988]*

The gas-phase terms cancel and $\Delta G_{sol}(LR \rightarrow 0)$ can be written as the sum of two separate calculations:

$$\Delta G_{sol}(LR \rightarrow 0) = \Delta G_{sol}(LR \rightarrow R) + \Delta G_{sol}(R \rightarrow 0) \quad (11\,24)$$

Thus, the overall free energy change can be written:

$$\Delta G_{ass} = \Delta G_{sol}(L \rightarrow 0) - \Delta G_{sol}(LR \rightarrow R) \quad (11.25)$$

We thus need perform only two simulations, L to nothing in water and L to nothing in the LR complex. The first application of this approach was to the association of two methane molecules in water, where both species (L and R) are identical. In general, L should be chosen as the smaller component.

11.4 The Calculation of Enthalpy and Entropy Differences

Free energy changes can now be routinely calculated with errors of less than 1 kcal/mol in favourable cases. How does this compare with the error with which the enthalpy or entropy difference can be determined? One way to determine the enthalpy change would be to perform two separate simulations, one of the initial system and one of the final system. For example, the difference in the enthalpy of solvation of ethanol and ethane thiol in water could be determined by simulating the two species separately and then taking the difference in the total enthalpies of the two systems. These total energies are invariably large numbers, with relatively large errors. The error in the calculated enthalpy difference would be comparable in magnitude to the error in the energy of each system. By contrast, the free energy is determined solely in terms of interactions involving the solute. This means that the free energy can be calculated much more accurately. More efficient ways to calculate the enthalpy and entropy change have been proposed for use with both free energy perturbation and thermodynamic integration schemes [Fleischman and Brooks 1987; Yu and Karplus 1988]. The uncertainties in the enthalpies and entropies so calculated are better than would be obtained by subtracting the differences in total energies, but they are still about one order of magnitude larger than the corresponding free energies.

11.5 Partitioning the Free Energy

The overall free energy can be partitioned into individual contributions if the thermodynamic integration method is used [Boresch *et al.* 1994; Boresch and Karplus 1995]. The starting point is the thermodynamic integration formula for the free energy:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (11.26)$$

The Hamiltonian can be written as a sum of contributions from bond stretching, angle bending, and so on:

$$\left\langle \frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \right\rangle_\lambda = \left\langle \frac{\partial \mathcal{H}_{bonds}(\lambda)}{\partial \lambda} + \frac{\partial \mathcal{H}_{angles}(\lambda)}{\partial \lambda} + \cdots \right\rangle_\lambda \quad (11.27)$$

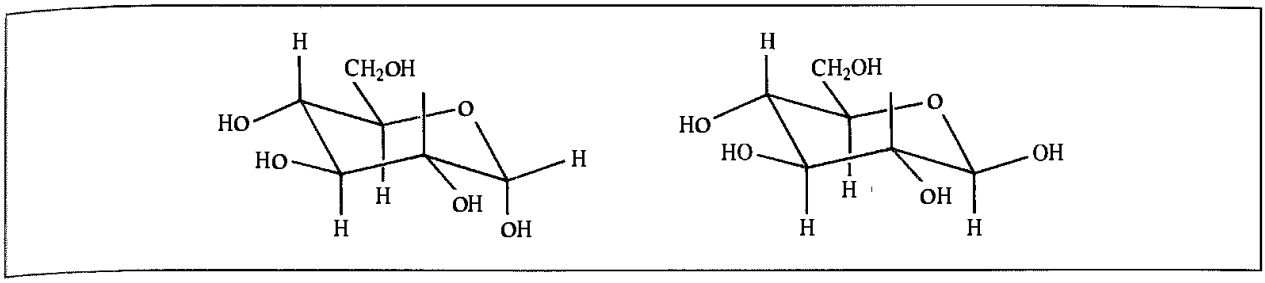

Fig 11 10 The α and β anomers of D-glucose

So the free energy is given by:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}_{\text{bonds}}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}_{\text{angles}}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \cdots$$
$$= \Delta A_{\text{bonds}} + \Delta A_{\text{angles}} + \cdot \qquad (11\,28)$$

We should remember that only the sum of the contributions is truly meaningful, as the individual contributions are not state functions. This has led some to criticise any use of such partitioning schemes [Smith and van Gunsteren 1994b], though they may be useful to indicate which interactions contribute the most to the overall free energy, and may also suggest the source of most of the error in the calculation. It is not possible to perform such a partitioning using thermodynamic perturbation.

An example of this partitioning scheme is the study by Ha and colleagues of the anomeric equilibrium between the α and β anomers of D-glucose [Ha *et al.* 1991]. D-glucose can exist in two tautomeric forms: α-D-glucose, in which the C_1 hydroxyl group is axial; and β-D-glucose, in which it is equatorial (Figure 11.10). In the gas phase, the axial α isomer is more stable than the equatorial β isomer, due to the anomeric effect which is considered to arise from unfavourable dipole–dipole interactions and delocalisation of the lone pair on the ring oxygen into an anti-bonding σ^* orbital. However, the β-D (equatorial) anomer is more stable than the α-D (axial) anomer by 0.3 kcal/mol in aqueous solution. The free energy difference between the two isomers in water was calculated by Ha *et al.* using both free energy perturbation and thermodynamic integration to be -0.3 ± 0.4 kcal/mol for $\beta \rightarrow \alpha$. A partitioning of the free energy showed that this small difference arose from the cancellation of two large terms: the α isomer was predicted to be 3.6 kcal/mol more favourable than the β isomer in the gas phase, due mainly to electrostatic effects. However, the β isomer was favoured over the α isomer in aqueous solution, again due to electrostatic effects, such as the enhanced hydrogen-bonding capability of the β isomer with the solvent. The small free energy difference, the difficulties of obtaining a reliable force field model and the large number of accessible conformations makes this equilibrium particularly difficult to tackle. One way in which the sampling problem has been tackled is by the use of a method called locally enhanced sampling (LES), which uses multiple copies of those parts of the system that can exist in more than one conformation. In the case of glucose these are the hydroxyl hydrogens and the hydroxymethyl group. In LES, each of the copies does not interact with the other copies of the same group and each atom 'sees' the mean force from all the copies. The first application of the technique was the study of the diffusion of

carbon monoxide within the protein myoglobin [Elber and Karplus 1990], but there are many other potential applications. LES reduces the barriers to conformational transitions, which leads to more rapid transitions between the conformational minima [Simmerling and Elber 1995]. However, the free energies calculated for the LES energy surface do need to be corrected to give the corresponding result for the single-copy system. In the case of glucose, it was found that the α isomer was favoured by 0.5–1.0 kcal/mol in the gas phase but that the β isomer was favoured in solution by 0.2 kcal/mol [Simmerling *et al.* 1998]. The gas-phase result was suggested to be a consequence of the tendency of the O–C–O–C linkage to adopt a gauche conformation, whereas the solution result was due to solvation effects. Many quantum mechanical studies have also been performed on this system, some of which have also included solvation effects (using the methods to be discussed in Section 11.10.2). For example, Barrows and co-workers performed high-level *ab initio* calculations on 11 varied low-energy conformations on glucose using large basis sets and including the effects of electron correlation [Barrows *et al.* 1998]. From this data, they calculated a gas-phase equilibrium constant of 0.4 kcal/mol in favour of the α isomer (using a Boltzmann-weighted average), whereas the equivalent value in solution was 0.6 kcal/mol in favour of the β isomer.

Another common practice is to partition the free energy into contributions from van der Waals and electrostatic interactions. This can be achieved rather easily by first perturbing the electrostatic and then the van der Waals parameters One system that has been studied in this way is the biotin/streptavidin complex. This protein–ligand complex is of particular interest due to the extremely strong association constant (−18 3 kcal/mol). The chemical structure of biotin is shown in Figure 11 11. The separate electrostatic and van der Waals free-energy calculations suggested that the largest contribution to the very negative free energy of binding was due to the non-polar van der Waals contribution rather than the electrostatic component [Miyamoto and Kollman 1993a, b]. Despite the presence of many hydrogen bonds between the ligand and the protein in the complex it was suggested that, whilst there was a large and favourable electrostatic interaction between biotin and streptavidin, this was almost cancelled by the free energy of interaction of biotin with water. By contrast, the van der Waals interaction gave a much greater contribution in the protein–ligand complex than for the ligand in water, so leading to its dominance. Indeed, the ligand is almost completely buried within the protein cavity, as can be seen in the structure of the intermolecular complex shown in Figure 11.12 (colour plate section).

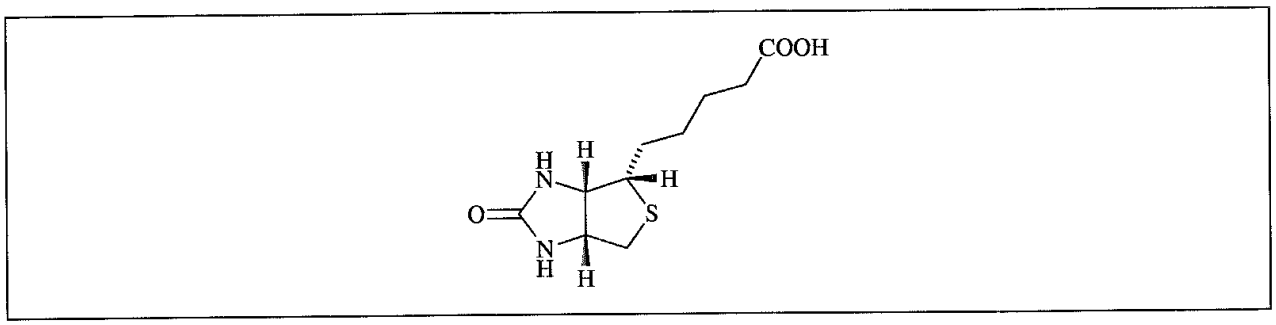

Fig 11 11· Biotin.

11.6 Potential Pitfalls with Free Energy Calculations

There are two major sources of error associated with the calculation of free energies from computer simulations. Errors may arise from inaccuracies in the Hamiltonian, be it the potential model chosen or its implementation (the treatment of long-range forces, etc.) The second source of error arises from an insufficient sampling of phase space.

Unfortunately, there is no set recipe that guarantees adequate coverage of phase space and thus reliable free energy values [Mitchell and McCammon 1991]. The errors associated with inadequate sampling may be identified by running the simulation for longer periods of time (molecular dynamics) or for more iterations (Monte Carlo); the perturbation can be performed in both forward and reverse directions; a different scheme could be used to determine the free energy difference (e.g. thermodynamic perturbation and thermodynamic integration). At the very least, the simulation should be run in both directions; the difference in the calculated free energy values (often referred to as the *hysteresis*) gives a lower-bound estimate of the error in the calculation

One possible pitfall to be aware of when estimating errors is that an excessively short simulation may give an almost zero difference between the forward and reverse directions. If the time of the simulation is much longer than the relaxation time of the system then the change can be performed reversibly. If the simulation time is of the same order of magnitude as the relaxation time then one would expect a significant degree of hysteresis However, if the simulation is much shorter than the relaxation time then approximately zero hysteresis may result, due to the inability of the system to adjust to the changes. In such a situation, the free energies for both forward and reverse directions may be approximately the same, but quite likely incorrect.

11.6.1 Implementation Aspects

The allure of methods for calculating free energies and their associated thermodynamic values such as equilibrium constants has resulted in considerable interest in free energy calculations. A number of decisions must be made about the way that the calculation is performed. One obvious choice concerns the simulation method. In principle, either Monte Carlo or molecular dynamics can be used; in practice, molecular dynamics is almost always used for systems where there is a significant degree of conformational flexibility, whereas Monte Carlo can give very good results for small molecules which are either rigid or have limited conformational freedom.

One must choose from the thermodynamic perturbation, thermodynamic integration and slow growth methods. Each of these methods has been extensively used, but the slow growth method is not now recommended. This method suffers from a phenomenon known as 'Hamiltonian lag'; the system never has time to properly equilibrate for a given value of the coupling parameter, because the potential function changes at every step. An additional advantage of the integration and perturbation approaches is that, should one decide at the end of a simulation that more sampling needs to be done for particular values of λ, or that more λ values are required over a particular range, then this can

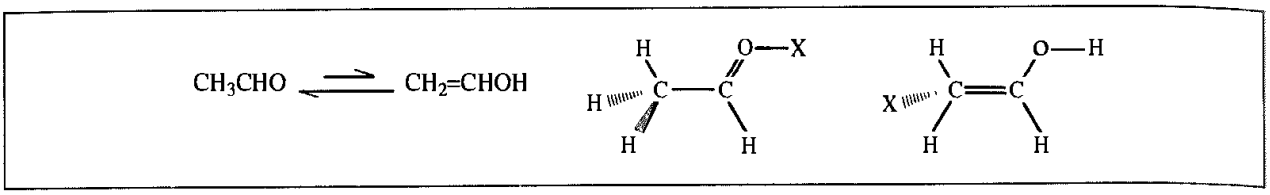

Fig. 11 13 To calculate the free energy difference between the aldehyde and enol forms of acetaldehyde, the single topology method uses dummy atoms (X)

easily be done without losing information from other parts of the calculation With slow growth, one would have to redo the simulation from scratch.

Prior to the calculation, the increment $\delta\lambda$ in the coupling parameter must be specified Traditionally, $\delta\lambda$ is set to a constant value before the simulation commences. It is important that there is enough overlap between successive states λ_i and λ_{i+1} so that reliable values can be obtained. An alternative approach is to use small changes in λ when the free energy is changing quickly and a larger change in λ when the free energy is changing more slowly. This is the basis of a method called *dynamically modified windows*, in which the slope of the free energy versus λ curve is used to determine the value of $\delta\lambda$ to use in the next iteration [Pearlman and Kollman 1989].

As the free energy is a thermodynamic state function the free energy difference between the initial and final states should be independent of the path along which the change is made, so long as it is reversible. It may be possible to proceed from the initial to the final state along more than one pathway. A change that involves high energy barriers will require much smaller increments to be made in the coupling parameter λ to ensure reversibility than a pathway that proceeds via a lower barrier.

Many free energy calculations involve changes in the molecular topologies of the species concerned, there are often different numbers of atoms in the initial and final states, and the atoms may be bonded in different ways. For example, suppose we wish to determine the free energy difference between acetaldehyde and its enol (Figure 11.13), in which a hydrogen atom migrates from the methyl carbon atom to the carbonyl oxygen. The system can be represented in the calculation using either a 'single' topology or a 'dual' topology. In the single-topology method, the molecular topology at all stages is the union of the initial and final states, using dummy atoms where necessary. A dummy atom does not interact with the other atoms in the system. Thus the hydrogen atom bonded to the oxygen in the enol form would be represented as a dummy atom when the simulation reached the endpoint corresponding to the aldehyde as shown in Figure 11.13.

The alternative to the single-topology representation is the dual-topology method. Here, both the molecular topologies are maintained during the entire simulation, such that both species 'exist' (in a topological sense) but do not interact with each other. The Hamiltonian that describes the interaction between these groups and the environment can be described in a number of ways, the simplest of which is the linear relationship:

$$\mathscr{H}(\lambda) = \lambda\mathscr{H}_Y + (1-\lambda)\mathscr{H}_X \tag{11.29}$$

Many free energy calculations involve the creation or annihilation of atoms. A potential problem with such simulations is that a singularity may occur in the function for which

an ensemble average is to be formed. One way to try to deal with this is to scale the initial Hamiltonian by a factor λ^n (rather than just λ) and the final Hamiltonian by $(1-\lambda)^n$ (rather than $(1-\lambda)$). It can be shown that for Monte Carlo simulations the singularity problem can be dealt with, provided n is at least 4 [Buetler *et al.* 1994]. However, a molecular dynamics simulation requires not only the energies to be calculated but also the first and the second derivatives. If λ^n scaling is used then either a steadily decreasing time step must be used as λ approaches zero, or these regions of the simulation must be omitted altogether and their contributions estimated by extrapolation. An alternative approach is to replace the traditional Lennard-Jones interaction with a soft-core potential of the following form [Buetler *et al.* 1994; Liu *et al.* 1996]:

$$v_{ij}^{\mathrm{LJ}} = 4\varepsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{[\alpha_{\mathrm{LJ}}\sigma_{ij}^6 + r_{ij}^6]^2} - \frac{\sigma_{ij}^6}{(\alpha_{\mathrm{LJ}}\sigma_{ij}^6 + r_{ij}^6)}\right) \tag{11.30}$$

where ε_{ij} and σ_{ij} have their usual Lennard-Jones meanings. The parameter α_{LJ} determines the 'softness' of the interaction, which has the effect of making the interaction approach a finite value as the interatomic distance r_{ij} goes to zero. With a suitable choice of the parameter α_{LJ} it is possible to ensure that the position of the minimum in the soft-core potential coincides with that of the unscaled energy curve (Figure 11.14) When used to perturb the system from X at $\lambda = 0$ to Y at $\lambda = 1$ then the soft-core Lennard-Jones interaction between the particle i that is being perturbed and some other particle j at a distance r_{ij} varies as:

$$\begin{aligned} v_{ij}^{\mathrm{LJ}}(\lambda) = {} & 4(1-\lambda)\varepsilon_{\mathrm{X}}\left(\frac{\sigma_{\mathrm{X}}^{12}}{[\alpha_{\mathrm{LJ}}\lambda^2\sigma_{\mathrm{X}}^6 + r_{ij}^6]^2} - \frac{\sigma_{\mathrm{X}}^6}{[\alpha_{\mathrm{LJ}}\lambda^2\sigma_{\mathrm{X}}^6 + r_{ij}^6]}\right) \\ & + 4\lambda\varepsilon_{\mathrm{Y}}\left(\frac{\sigma_{\mathrm{Y}}^{12}}{[\alpha_{\mathrm{LJ}}(1-\lambda)^2\sigma_{\mathrm{Y}}^6 + r_{ij}^6]^2} - \frac{\sigma_{\mathrm{Y}}^6}{[\alpha_{\mathrm{LJ}}(1-\lambda)^2\sigma_{\mathrm{Y}}^6 + r_{ij}^6]}\right) \end{aligned} \tag{11.31}$$

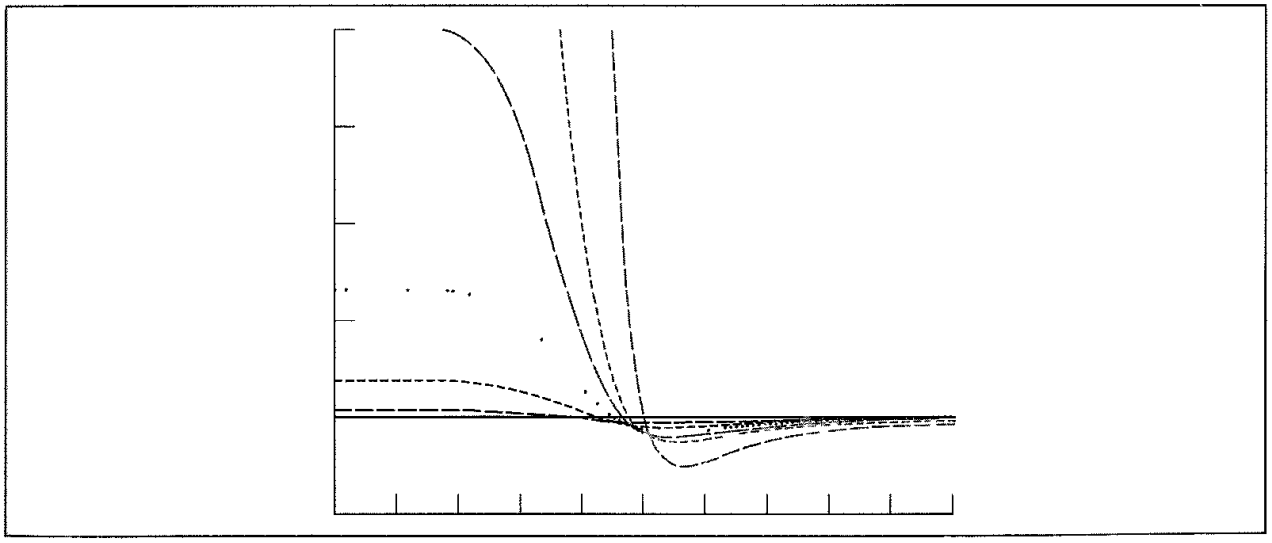

Fig 11 14. Comparison of scaled and unscaled Lennard-Jones potentials (Equation (11 31)) for the case where a particle disappears at $\lambda = 0$ As λ decreases the curves get progressively closer to the x axis

A similar soft-core expression can also be derived for the electrostatic interactions. In the situation where the particle disappears then only one of the terms remains (i.e. the second term if the particle disappears at $\lambda = 0$). The effect of using this type of soft-core potential can be seen in simulations of protein–ligand systems, where the ligand 'disappears' and is replaced by one or more solvent molecules. In a normal perturbation calculation, decreasing the effective radius of the ligand atoms will give rise to a collapse of the protein cavity and disruption to the surrounding protein structure. By contrast, placing the soft-core interaction sites at locations where the atoms are to be created or deleted maintains the protein cavity, because the solvent molecules are able to actually pass through the ligand as it is annihilated. This feature of soft-core potentials can also be useful for other kinds of free energy calculation where it is desired to try to simultaneously derive the relative free energies of binding of several ligands to a receptor and also in certain types of simulated annealing structure refinement.

11.7 Potentials of Mean Force

The free energy changes that we have considered so far correspond to chemical 'mutations'. We may also be interested to know how the free energy changes as a function of some inter- or intramolecular coordinate, such as the distance between two atoms, or the torsion angle of a bond within a molecule. The free energy surface along the chosen coordinate is known as a *potential of mean force* (PMF). When the system is in a solvent, the potential of mean force incorporates solvent effects as well as the intrinsic interaction between the two particles. Potentials of mean force were introduced in our discussion of Langevin dynamics (Section 7.8), where we noted that the ratio of *trans* to *gauche* conformers of 1,2-dichloroethane was significantly different in the liquid than in an isolated molecule. Unlike the mutations so common in free energy perturbation calculations, which are often along non-physical pathways, the potential of mean force is calculated for a physically achievable process. Consequently, the point of highest energy on the free energy profile that is obtained from a PMF calculation corresponds to the transition state for the process, from which it is possible to derive kinetic quantities such as rate constants.

Various methods have been proposed for calculating potentials of mean force. The simplest type of PMF is the free energy change as the separation (r) between two particles is changed. We might anticipate that we could calculate the potential of mean force from the radial distribution function using the following expression for the Helmholtz free energy:

$$A(r) = -k_B T \ln g(r) + \text{constant} \tag{11.32}$$

The constant is often chosen so that the most probable distribution corresponds to a free energy of zero.

Unfortunately, the potential of mean force may vary by several multiples of $k_B T$ over the relevant range of the parameter r. The logarithmic relationship between the potential of mean force and the radial distribution function means that a relatively small change in the free energy (i.e. a small multiple of $k_B T$) may correspond to $g(r)$ changing by an order of magnitude from its most likely value. Unfortunately, standard Monte Carlo or molecular

dynamics simulation methods do not adequately sample regions where the radial distribution function differs drastically from the most likely value, leading to inaccurate values for the potential of mean force. The traditional way to avoid this problem uses a technique called *umbrella sampling*.

11.7.1 Umbrella Sampling

Umbrella sampling attempts to overcome the sampling problem by modifying the potential function so that the unfavourable states are sampled sufficiently. The method can be used with both Monte Carlo and molecular dynamics simulations. The modification of the potential function can be written as a perturbation:

$$\mathcal{V}'(\mathbf{r}^N) = \mathcal{V}(\mathbf{r}^N) + W(\mathbf{r}^N) \tag{11.33}$$

where $W(\mathbf{r}^N)$ is a weighting function, which often takes a quadratic form:

$$W(\mathbf{r}^N) = k_W(\mathbf{r}^N - \mathbf{r}_0^N)^2 \tag{11.34}$$

For configurations that are far from the equilibrium state $\mathbf{r}_0^N$ the weighting function will be large and so a simulation using the modified energy function $\mathcal{V}'(\mathbf{r}^N)$ will be biased along some relevant 'reaction coordinate' away from the configuration $\mathbf{r}_0^N$. The resulting distribution will, of course, be non-Boltzmann. The corresponding Boltzmann averages can be extracted from the non-Boltzmann distribution using a method introduced by Torrie and Valleau [Torrie and Valleau 1977]. The result is:

$$\langle A \rangle = \frac{\langle A(\mathbf{r}^N) \exp[+W(\mathbf{r}^N)/k_B T] \rangle_W}{\langle \exp[+W(\mathbf{r}^N)/k_B T] \rangle_W} \tag{11.35}$$

The subscript W indicates that the average is based on the probability $P_W(\mathbf{r}^N)$, which in turn is determined by the modified energy function $\mathcal{V}'(\mathbf{r}^N)$. For example, to obtain the potential of mean force via the radial distribution function (Equation (11.32)) the distribution function with the forcing potential would be determined and then corrected to give the 'true' radial distribution function, from which the free energy can be calculated as a function of the separation It is usual to perform an umbrella sampling calculation in a series of stages, each of which is characterised by a particular value of the coordinate and an appropriate value of the forcing potential $W(\mathbf{r}^N)$. However, if the forcing potential is too large, the denominator in Equation (11.35) is dominated by contributions from only a few configurations with especially large values of $\exp[W(\mathbf{r}^N)]$ and the averages take too long to converge.

To illustrate the use of umbrella sampling, let us consider how the technique has been used to determine the potential of mean force for rotation of the central C–C bond of butane in aqueous solution. The barrier between the *trans* and *gauche* conformations of butane is approximately 3.5 kcal/mol, which is sufficiently high to give sampling problems in simulations. For example, in the molecular dynamics simulation of Ryckaert and Bellemans the mean time between *gauche–trans* transitions was about 10 ps [Ryckaert and Bellemans 1978]. Jorgensen, Gao and Ravimohan used umbrella sampling with Monte Carlo simulations to calculate the potential of mean force as the central bond in butane is rotated in a periodic box of water molecules, to determine the effect of the solvent on the relative

populations of the different conformations [Jorgensen *et al.* 1985]. The results predicted a shift in the expected populations of *trans* and *gauche* isomers from 68% *trans* in the gas phase to 54% in aqueous solution, a change of 14%. In addition, the barrier height was reduced in solution. Jorgensen and colleagues performed many calculations on similar systems using umbrella sampling and Monte Carlo simulations; he recommended that to reduce the barriers to a value between 1 kcal/mol and 3 kcal/mol was appropriate. In some cases, it is possible to use a barrier height of zero, though the barriers cannot be reduced too severely as this makes the forcing potential too large.

It is also possible to calculate potentials of mean force using the free energy perturbation method with a molecular dynamics or Monte Carlo simulation. As usual, the calculation is broken into a series of steps that are characterised by a coupling parameter λ. With molecular dynamics, holonomic constraint methods are used to fix the desired coordinates without affecting the dynamic motion of the system. This is the essence of the extension of the SHAKE procedure by Tobias and Brooks to cope with general coordinate changes [Tobias and Brooks 1988] (see Section 7.5). In a Monte Carlo simulation the required coordinates are simply fixed at the desired value(s). This contrasts with umbrella sampling, in which the coordinate(s) of interest would be able to vary over their range of values throughout the simulation, subjected to a potential that has been modified using the forcing function. At each step of the perturbation calculation, the difference in the energy between the configuration and the configuration that corresponds to $\lambda + \delta\lambda$ is determined and the free energy accumulated in the appropriate way.

To compare the perturbation and umbrella sampling methods for calculating potentials of mean force, Jorgensen and Buckner repeated the PMF calculation for butane in water using the perturbation method [Jorgensen and Buckner 1987]. The *gauche* population was calculated to increase by 12.3% using this method, in accordance with the previous umbrella sampling calculations. Jorgensen put forward several arguments in favour of the perturbation approach. A major concern with umbrella sampling is that a proper sampling of the phase space may not be achieved. In some cases, the presence of bottlenecks in phase space may be identified if separate simulations starting from different configurations give different results, but even this approach is not fail-safe as all simulations may encounter the same problem. Indeed, Jorgensen suggested that just such a bottleneck may have occurred in a previous simulation of pentane in water using umbrella sampling (which involved 5 million Monte Carlo steps). The only real problem with the perturbation method is the need to choose an appropriate value of $\delta\lambda$ so that there is adequate overlap between the configuration corresponding to λ and that corresponding to $\lambda + \delta\lambda$. Jorgensen and Buckner varied the central torsion angle in their simulation of butane using 15° increments.

11.7.2 Calculating the Potential of Mean Force for Flexible Molecules

To calculate a potential of mean force using free energy perturbation (or indeed umbrella sampling) it is necessary to determine the pathway for the transition of interest. This is trivial for simple problems such as the separation of two particles or the rotation of butane but can be quite complicated for more detailed changes such as conformational interconversions.

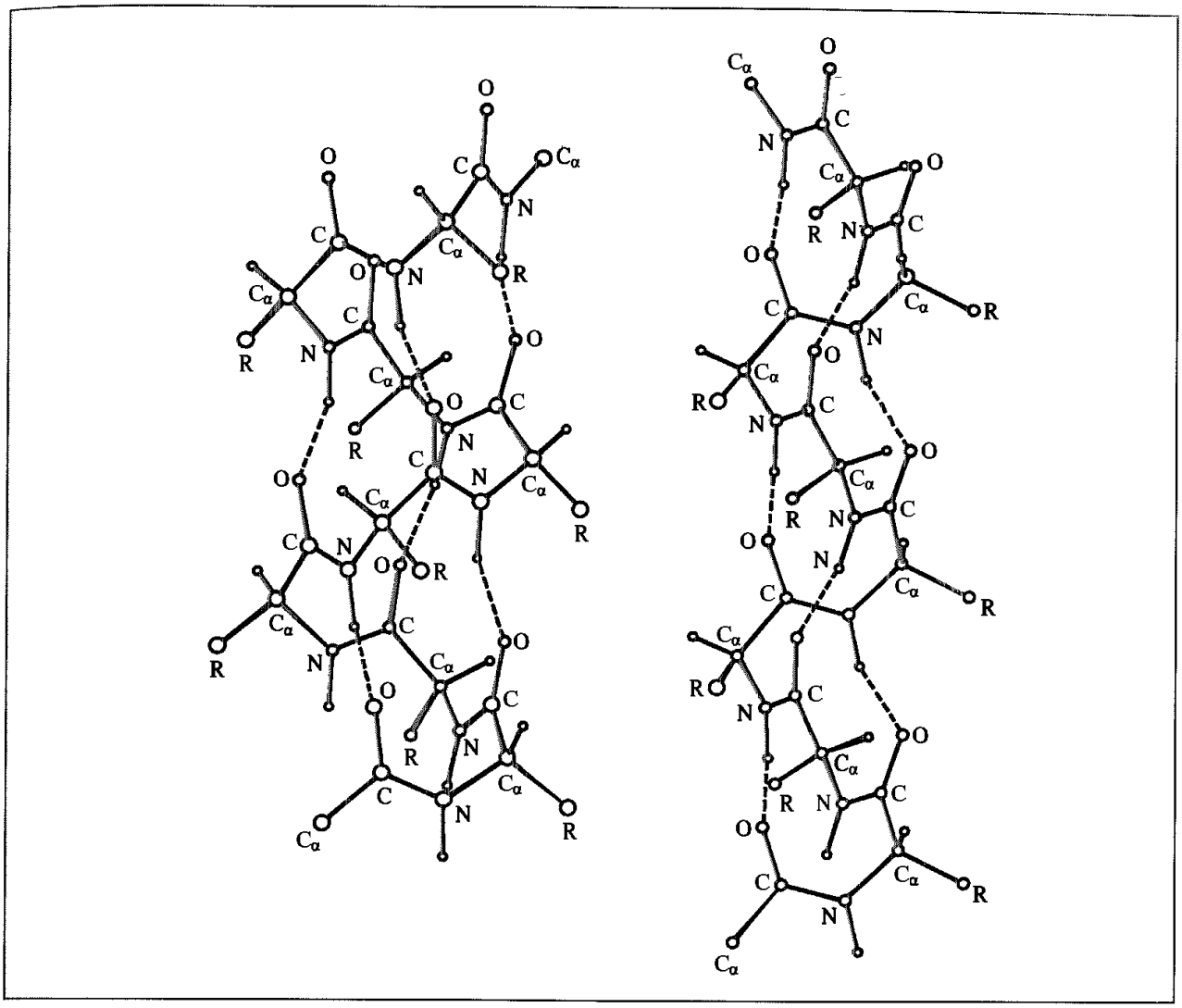

Fig 11 15. The α-helix (left) and the 3_{10}-helix

The reaction path methods discussed in Section 5.9.3 may be helpful in determining these pathways.

To illustrate the calculation of potentials of mean force for flexible systems we will consider helical conformations of polypeptide chains. We have already met the α-helix, which is commonly observed in protein structures (see Section 10.2). In this conformation, hydrogen bonds are formed between residues i and $i + 4$. Polypeptide chains can also form a different type of helix, called a 3_{10}-helix. Here, the hydrogen bonds are formed between residues i and $i + 3$. These two helices are compared in Figure 11.15. The backbone conformations of such helices do not differ significantly: the α-helix has backbone torsion angles ($\phi = -60°$, $\psi = -50°$) and the 3_{10}-helix has ($\phi = -50°$, $\psi = -28°$). The 3_{10}-helix is found to a small extent in protein structures, usually at the ends of α-helices. However, the 3_{10}-helix is much more common in peptides formed from α,α-dialkyl amino acids, which have two alkyl substituents at the α-carbon atom. The prototypical member of this class of amino acids is α-methylalanine (MeA; see Figure 11.16) Peptides containing this amino acid can

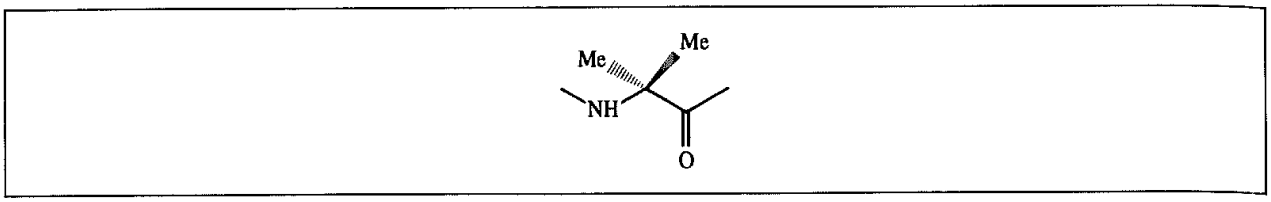

Fig. 11.16. Methylalanine

form both α-helices and 3_{10}-helices with the actual conformation present being rather sensitive to the conditions; such peptides are 3_{10}-helical in $CDCl_3$ and α-helical in $(CD_3)_2SO$.

To calculate the potential of mean force for interconverting the α-helix to the 3_{10}-helix requires an appropriate reaction coordinate to be determined. Here we describe the calculations of three groups who all used different approaches. Smythe, Huston and Marshall studied a decamer of α-methylalanine, $Ch_3CO\text{-}MeA_{10}\text{-}NMe$, using umbrella sampling [Smythe *et al.* 1993, 1995]. They used the self-penalty walk method described in Section 5.9.3 to determine the transition pathway and observed that the reaction coordinate correlated well with a smooth change in the end-to-end distance from the 3_{10}-helix (19 Å) to the α-helix (13 Å). Their umbrella sampling calculations were performed using molecular dynamics, with the end-to-end distance being subjected to a restraining potential. Simulations were performed in various solvents: in water, the free energy change for the α-helix $\rightarrow$ 3_{10}-helix transition was calculated to be 7.6 kcal/mol, with the value in dichloromethane being 5.8 kcal/mol and *in vacuo* 3.2 kcal/mol Although a distinct energy barrier was found for the vacuum calculations, no transition barrier was found for either solution calculation.

Zhang and Hermans studied a 10-residue alanine peptide as well as a 10-residue α-methylalanine peptide, *in vacuo* and in water [Zhang and Hermans 1994]. In their calculations, the transition from one conformation to the other was performed using a restraining potential that forced the structure to exchange one set of hydrogen bonds to the set of hydrogen bonds appropriate to the other structure. This additional potential function could be used to drive the molecule back and forth between the two conformations by varying a coupling parameter, λ, between 0 and 1. Free energy profiles were determined for the α-helix to 3_{10}-helix transition using molecular dynamics and the slow growth method. The results showed that the alanine peptide had a clear preference for the α-helix both *in vacuo* and in water but that the free energy change for the MeA peptide was approximately zero in water and that the 3_{10}-helix was preferred *in vacuo*. It was proposed by Zhang and Hermans that the discrepancy between these results for the α-methylalanine peptide and those obtained by Smythe, Huston and Marshall was probably due to the different force field models employed; Smythe *et al.* used a united atom model, whereas Zhang and Hermans used an all-atom model.

Tirado-Reeves, Maxwell and Jorgensen used yet another approach for calculating the potential of mean force, this time for an undecaalanine peptide in water [Tirado-Reeves *et al.* 1993]. The free energy profile was calculated using the perturbation method and Monte Carlo simulations by gradually varying the ψ backbone torsion angles, keeping the ϕ torsion angles fixed at $-60°$. The free energy difference between the two conformations

was calculated to be 10.6 kcal/mol in favour of the α-helix, with a small activation barrier of 2.8 kcal/mol for the 3_{10}- to α-helical transition. *In vacuo*, a larger free energy difference was predicted (13.6 kcal/mol).

These three studies have been described at some length, in part to illustrate the different approaches available for calculating thermodynamic properties of complex systems but also to emphasise the fact that different methods can given quite different (and sometimes contradictory) results. Such comparative studies serve to highlight the fact that it is necessary to examine critically the methods and models used in a calculation. All three studies were in part prompted by experimental electron spin resonance results that suggested that a 16-residue alanine-based peptide adopted a 3_{10}-helical conformation in water [Miick *et al.* 1992]. These results were contradicted by all the simulations, and indeed prompted Smythe and Marshall to undertake similar experiments on their conformationally constrained peptides, experiments which showed that these peptides were α-helical, in agreement with the calculations.

11.8 Approximate/'Rapid' Free Energy Methods

Free energy calculations are notoriously time-consuming to perform. Whilst one might have anticipated that ever faster computers would have made significant inroads on this problem, in some respects the opposite has happened, as researchers are now able to more fully quantify the need for sufficient sampling of phase space and to attain better convergence. In addition, of course, there is a natural desire to investigate ever larger systems. A practical illustration of the dilemma facing the proponents of free energy methods as a predictive tool, at least in an industrial environment, is that, if the calculation takes longer to perform than a candidate molecule can be synthesised and tested, then there is little practical benefit from attempting the calculation. There has thus been continued interest in the development of alternative methods which, whilst still being based upon 'exact' statistical mechanics, are intended to provide answers with less computational effort than a full-blown free energy calculation. These methods tend to approach the problem from one of two perspectives. Some, such as the λ-dynamics method, enable a single simulation to provide information on a number of molecules. Others, such as the linear response method, aim to limit the amount of simulation that needs to be performed.

In a traditional free energy calculation a coupling parameter, λ, provides the link between the initial and final systems. In most free energy calculations λ varies uniformly from 0 to 1 (or from 1 to 0), one exception being the dynamically modified windows technique discussed in Section 11.6.1. By contrast, the λ-dynamics technique considers lambda to be another 'particle' in the simulation, with its own fictitious mass. As such, λ-dynamics is similar in some respects to those charge calculation schemes where the charges can vary as a dynamic variable (see Section 4.9.6). Specifically, lambda corresponds to the reaction coordinate along which the potential would be modified in an umbrella sampling calculation. The advantage of making this association is that the biasing potentials that are used in the umbrella sampling method can be used in the λ-dynamics technique to provide enhanced sampling in relevant regions of configurational space Indeed, it is possible to

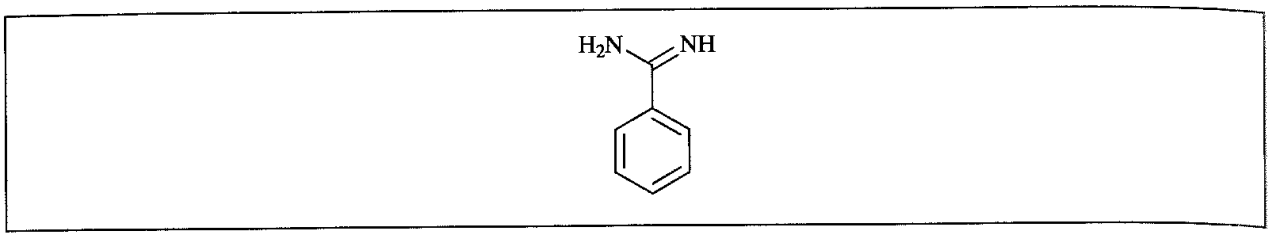

Fig 11.17 Benzamidine

use a set of coupling variables, λ_i, $i = 1, \ldots, n$. In the case where we have just one molecule being changed into another, then these different λ_i represent changes in different components of the interaction potential (for example, the Lennard-Jones and Coulombic interactions). If λ_1 and λ_2 refer to the Coulombic and the van der Waals interactions, respectively, then the potential function can be written:

$$\mathcal{V}(\mathbf{r}^N, \lambda_1, \lambda_2) = (1 - \lambda_1)\mathcal{V}_A^{\text{coul}} + \lambda_1 \mathcal{V}_B^{\text{coul}} + (1 - \lambda_2)\mathcal{V}_A^{\text{L-J}} + \lambda_2 \mathcal{V}_B^{\text{L-J}} + \mathcal{V}_{\text{env}}(\mathbf{r}^N) \quad (11\,36)$$

Here, the term $\mathcal{V}_{\text{env}}(\mathbf{r}^N)$ corresponds to all interactions involving those parts of the system that are not changing (i.e. the solvent and the unchanging part of the solute) The lambda variables move under the influence of a specific term which serves to limit their absolute extent (i.e. between 0 and 1) and which can be used to restrict their value to particular ranges during the simulation in order to provide enhanced sampling at particular points.

The basic λ-dynamics scheme can be used to perform a 'regular' type of free energy calculation in which one solute is perturbed into another such as the perturbation of methanol to ethane or to methane thiol [Kong and Brooks 1996]. However, it can also be used to investigate a number of perturbations simultaneously. As such, it provides a route to assess several free energies from a single simulation. One published example concerns the binding of benzamidine derivatives to the enzyme trypsin [Guo and Brooks 1998; Guo *et al.* 1998] Benzamidine is shown in Figure 11.17; this molecule binds relatively strongly to the enzyme because the positively charged amidine group interacts with a negatively charged aspartate residue in the protein. However, substitution at the para position can affect the strength of binding, with *p*-amino benzamidine binding slightly more strongly, *p*-methyl slightly more weakly and *p*-chloro more weakly still than the parent molecule. When λ-dynamics is applied to this problem, each of the L ligands ($L = 4$ in this case) is represented by a different value of λ_i. Initially, all values of λ_i are set to $1/L$ and their velocities set to zero. This means that each molecule is set on an equal footing at the beginning of the calculation. The system then evolves under the influence of the following hybrid potential:

$$\mathcal{V}(\mathbf{r}^N, \lambda_i) = \sum_{i=1}^{L} \lambda_i^2 (\mathcal{V}_i(\mathbf{r}^{\text{int}}) - F_i) + \mathcal{V}_{\text{env}}(\mathbf{r}^N) \quad (11\,37)$$

As in Equation (11.36), $\mathcal{V}_{\text{env}}(\mathbf{r}^N)$ corresponds to those interactions concerning all atoms not directly involved in the perturbations, whereas $\mathcal{V}_i(\mathbf{r}^{\text{int}})$ concerns those atoms associated with the group being perturbed in ligand i (for which the associated lambda parameter is λ_i). F_i is a reference free energy and can serve two purposes If F_i equals the solvation/desolvation free energy of the relevant ligand then the free energy value obtained from

the calculation corresponds to the free energy change for the full cycle. F_i can also be used as a biasing potential to control the sampling in particular regions of phase space. Finally, there is a constraint on the values of λ_i:

$$\sum_{i=1}^{L} \lambda_i^2 = 1 \tag{11.38}$$

As the simulation proceeds, the values of λ_i fluctuate, subject to the constraint in Equation (11.38). The free energy difference between two molecules i and j can be determined by identifying the probability that each molecule occupies the state $\lambda_i = 1$ or $\lambda_j = 1$, respectively Thus:

$$\Delta\Delta A_{ij} = -\frac{1}{k_B T} \ln \left[\frac{P(\lambda_i = 1, \lambda_{m \neq i} = 0)}{P(\lambda_j = 1, \lambda_{n \neq i} = 0)} \right] \tag{11 39}$$

These relative probabilities can be easily determined by simply counting the number of times during the simulation that the relevant value of lambda reaches unity. In the case of the para-substituted benzamidines it was possible after only a relatively short simulation (110 ps) to observe that the *p*-chloro and *p*-methyl derivatives were significantly weaker than the *p*-amino and the parent compound (Figure 11.18). In this particular case, all four

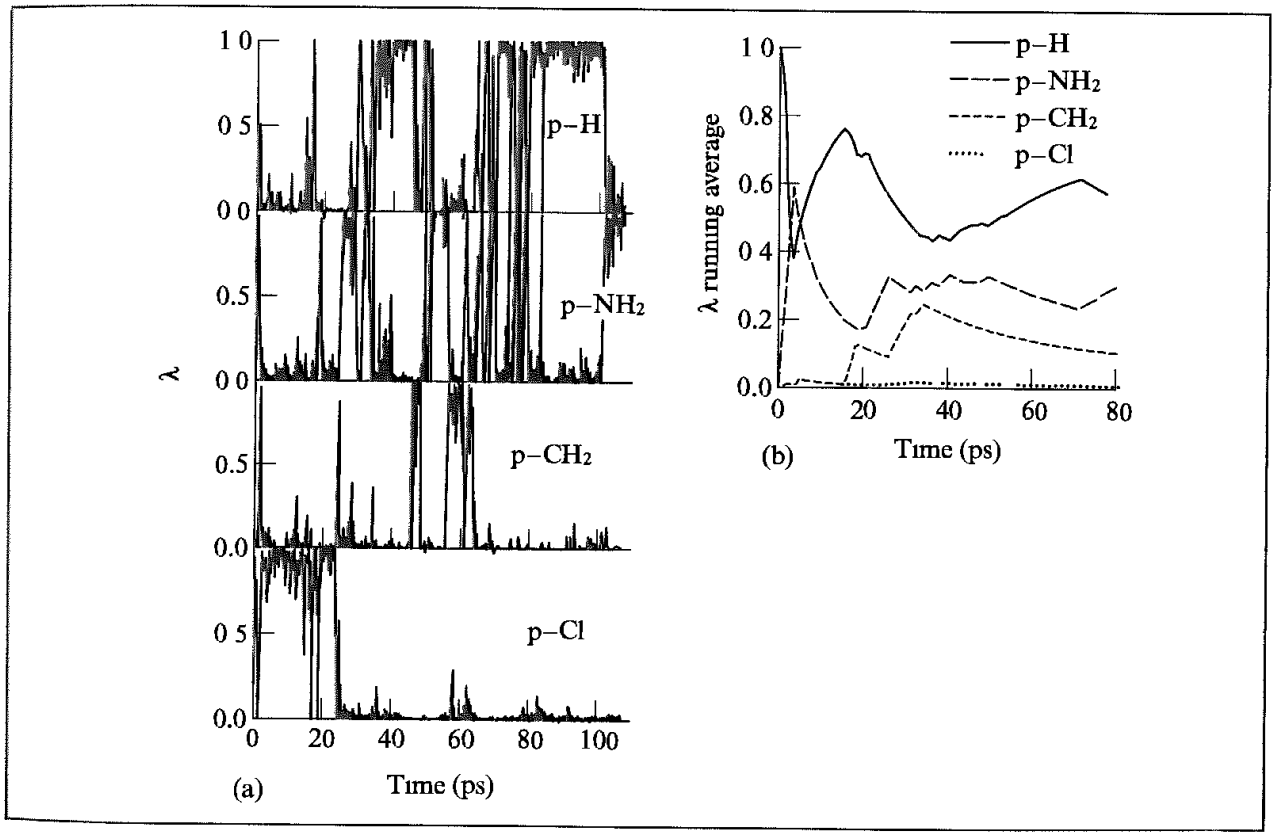

Fig 11 18 λ-dynamics simulation of benzamidine derivatives binding to trypsin. (a) The larger the value of λ the stronger the interaction with the protein at that instant (b) Running average of each value of λ over the course of the simulation (Figure redrawn from Guo Z and C L Brooks III 1998. Rapid Screening of Binding Affinities Application of the λ-Dynamics Method to a Trypsin-Inhibitor System Journal of the American Chemical Society *120* 1920–1921)

inhibitors have rather similar binding affinities (within 1 kcal/mol), thus requiring a relatively long simulation to separate them. In general, a compound with a binding affinity more than 3 kcal/mol worse than the most favourable molecule should be screened out within a few tens of picoseconds, though longer simulation times would be required to provide a correct rank ordering.

Conceptually similar to the lambda-dynamics approach is the so-called 'chemical-Monte Carlo/molecular dynamics' method [Pitera and Kollman 1998; Eriksson *et al.* 1999], which also considers many molecules simultaneously. In this approach, molecular dynamics is used to sample the coordinate space with Monte Carlo moves sampling the various chemical states. To avoid possible problems associated with hybrid states the chemical sampling is restricted to jumps between the relevant end-states. At the end of the simulation the relative free energies of the various chemical states is given by the ratio of the populations. Both host-guest and protein-ligand systems have been successfully investigated with the method, which, like the other methods discussed in this section, is designed to rapidly identify which candidates look most promising for further investigation.

The linear response (LR) method was originally devised by Åqvist and co-workers [Åqvist *et al* 1994] for estimating the binding affinities of ligands binding to proteins. Also known as the linear interaction energy (LIE) approach, it is a semi-empirical method for estimating absolute binding free energies and requires just two simulations, one of the solvated ligand–protein system and one of the ligand alone in solution. In both cases, the interaction between the ligand and its environment is broken down into the electrostatic and van der Waals contributions. The free energy of binding is then given by the following expression.

$$\Delta G = \beta(\langle \mathscr{V}^{\text{el}}_{\text{ligand-protein}} \rangle - \langle \mathscr{V}^{\text{el}}_{\text{ligand-solvent}} \rangle) + \alpha(\langle \mathscr{V}^{\text{vdw}}_{\text{ligand-protein}} \rangle - \langle \mathscr{V}^{\text{vdw}}_{\text{ligand-solvent}} \rangle) \quad (11.40)$$

As usual, the angle brackets $\langle\ \rangle$ indicate ensemble averages. α and β are two parameters. To determine ΔG one thus needs to perform just two simulations, one of the ligand in the solvent and the other of the ligand bound to the protein. The interactions that are accumulated consist solely of the electrostatic and van der Waals interactions between the ligand and its environment. First we consider an expansion of the Zwanzig expression for the free energy difference between two states X and Y (Equation (11.6)). The result obtained (see Appendix 11.3) is:

$$\Delta A = \frac{1}{2}[\langle \Delta\mathscr{H} \rangle_0 + \langle \Delta\mathscr{H} \rangle_1] - \frac{1}{4k_BT}[\langle (\Delta\mathscr{H} - \langle \Delta\mathscr{H} \rangle_0)^2 \rangle_0 - \langle (\Delta\mathscr{H} - \langle \Delta\mathscr{H} \rangle_1)^2 \rangle_1] + \cdots$$

$$\text{where} \quad \Delta\mathscr{H} = \mathscr{H}_Y - \mathscr{H}_X \quad (11.41)$$

For the electrostatic component, the free energy varies in a harmonic fashion with respect to deviations from equilibrium with a constant force constant (Figure 11.19). This is a standard result from dielectric theory and means that the mean square fluctuations of the energy on the two surfaces (the second terms in Equation (11.41)) will cancel, leaving just the first term. This leads to a value of $\frac{1}{2}$ for the electrostatic component (i.e. $\beta = 0\,5$). A simple test of this theory is to calculate the electrostatic contribution to solvation free energies. Here, state X corresponds to the situation where all of the solvent–solvent and intramolecular solute interactions are present but the interaction between the solute and solvent is only described

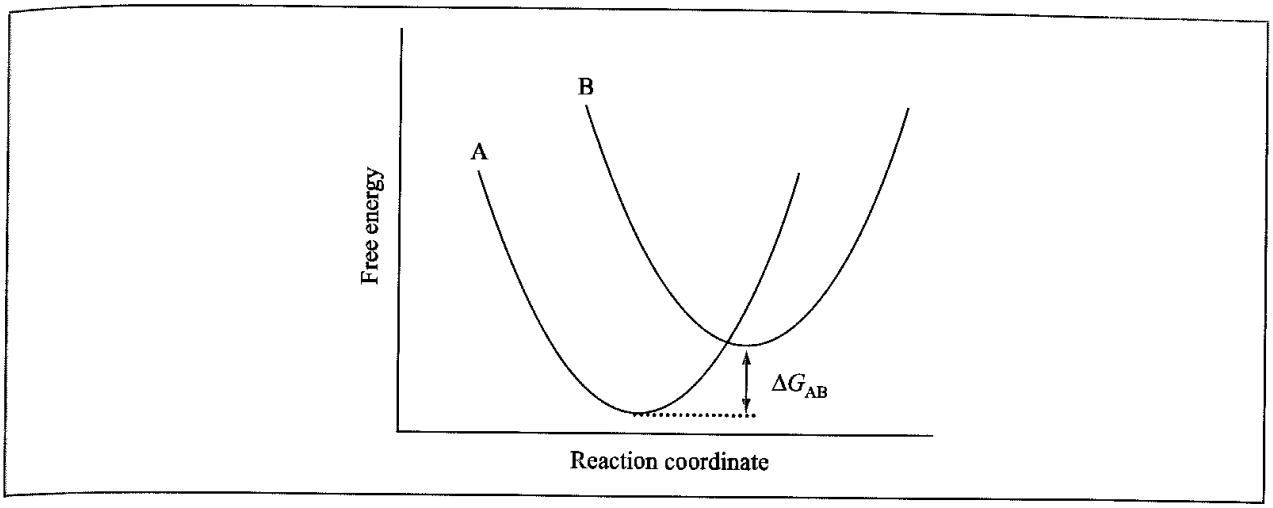

Fig 11 19· Representation of the harmonic variation of the electrostatic component of the free energy according to the linear response approximation

by a Lennard-Jones potential; the solute–solvent electrostatic interactions are missing. In state Y all interactions are included. The only difference between X and Y is thus the presence of the solute–solvent electrostatic terms, and so $\Delta\mathscr{H}$ $(= \mathscr{H}_Y - \mathscr{H}_X)$ in Equation (11.41) is equal to $\mathscr{H}^{el}$(ligand–solvent). Thus:

$$\Delta A^{el}_{sol} = \tfrac{1}{2}\langle \mathscr{H}^{el}_{ligand-solvent} \rangle \tag{11.42}$$

The validity of this result has been confirmed by for example comparing the free energy perturbation result for charging Na^+ and Ca^{2+} ions in water with the ensemble average value of $\mathscr{V}^{el}$(ion–solvent), giving factors of 0.49 and 0.52. Moreover, one can apply the same arguments to the case of a ligand in a protein environment, leading to the following expression for the electrostatic contribution to the free energy of binding (i.e. the first term in Equation (11.40)):

$$\Delta A^{el}_{binding} = \tfrac{1}{2}(\langle \mathscr{H}^{el}_{ligand-protein} \rangle - \langle \mathscr{H}^{el}_{ligand-solvent} \rangle) \tag{11.43}$$

For the van der Waals component no such analytical theory exists. Åqvist and co-workers assumed that a similar linear treatment would work for these interactions but with a different empirical factor, to be determined from calibration experiments. There was some indirect evidence that this approach would be reasonable. For example, the experimental free energies of solvation for various hydrocarbons (e.g. *n*-alkanes) depend in an approximately linear fashion on the length of the carbon chain. In addition, the mean van der Waals solute–solvent energies from molecular dynamics simulations did show a linear variation with chain length (the slope of the line varying according to the solvent).

What remains is to determine a value of the parameter α. In the original publication this was done using a series of ligands which bind to endothiapepsin, an enzyme for which various crystal structures are known. Some of these ligands are shown in Figure 11.20; as can be seen, they are quite substantial. Molecular dynamics simulations of four ligands were performed within the enzyme binding site and in water and accumulating the required average interaction energies. Assuming the factor $\frac{1}{2}$ for the electrostatic contribution and comparing with the experimental binding affinities gave $\alpha = 0.161$. When a fifth ligand was evaluated,

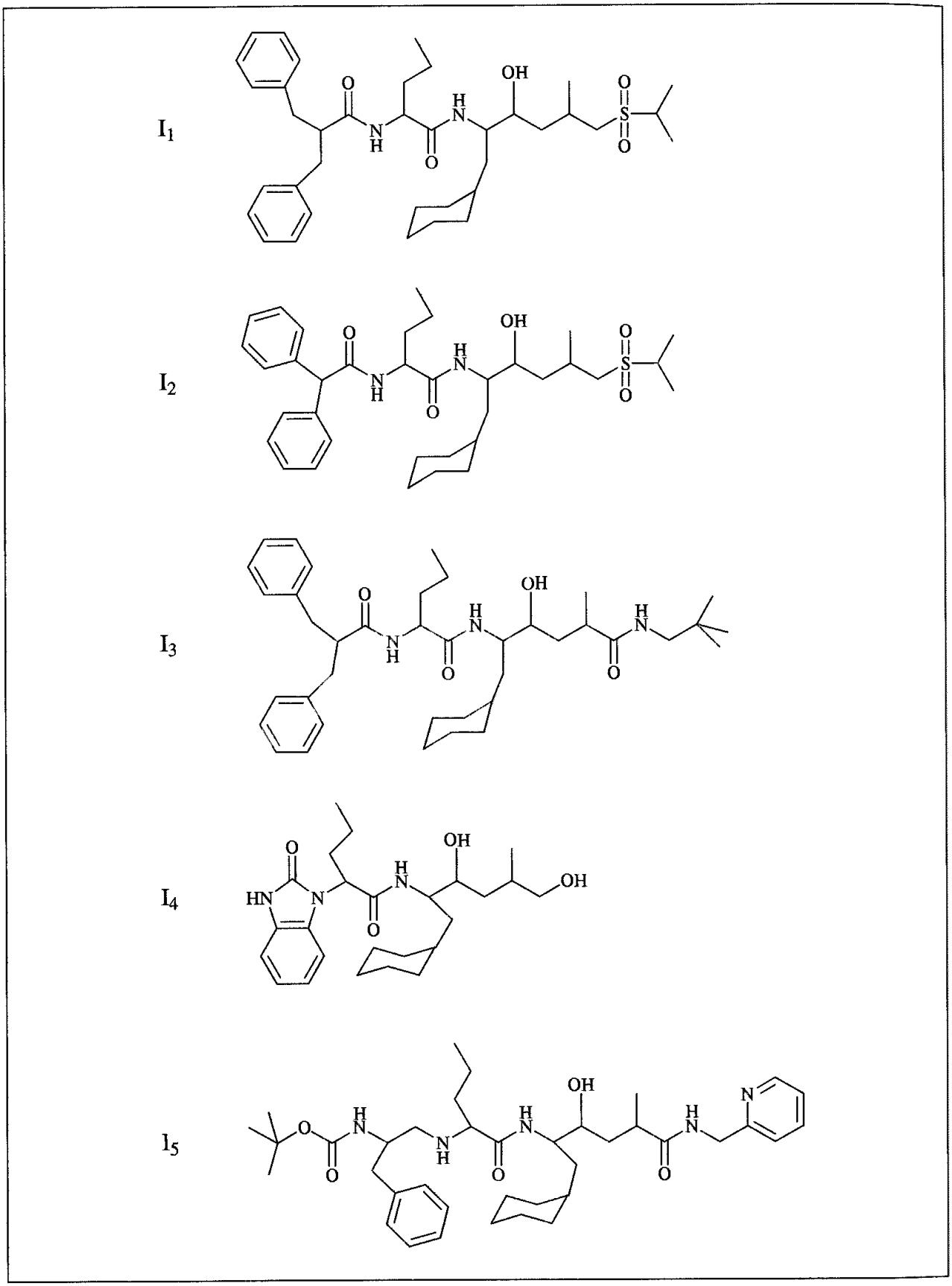

Fig 11 20: Structures of the endothiapepsin ligands used to calibrate the LIE approach.

not present in the calibration set, rather remarkably the predicted free energy of binding was within 0.2 kcal/mol of the experimental result.

Further studies for different ligands and different enzymes appeared to support the approach, and also the constants α and β [Hansson and Åqvist 1995, Hansson *et al* 1998] However, other groups found that different values of the van der Waals parameter, α, were required. One possible reason for this discrepancy could be due to the different protocols (for example, different force fields), but some groups found that different parameters were required for different systems, even when the same protocols were employed. One possible explanation for this is that α depends on the nature of the binding site This is not an unreasonable conclusion, given the different distributions of polar and non-polar groups in different binding sites. Wang and co-workers investigated this variation in more detail and showed that there appeared to be a correlation between the value of α and the 'weighted non-polar desolvation ratio', which is a measure of the hydrophobicity of the binding site [Wang *et al.* 1999]. It was found that, whilst it is generally more accurate to calibrate α for each system if experimental binding data for similar ligands is available, choosing a value based on the weighted non-polar desolvation ratio could give better results for dissimilar compounds.

Other groups have applied the linear response method to problems other than protein-ligand binding. A good problem for any new free energy approach is to predict the free energies of hydration of small organic molecules. Accurate hydration data are available for a wide variety of systems, and the calculations can usually be run relatively quickly. One immediate problem with the two-parameter linear response method is that, as α and β are both positive, it is not possible for any solute to have a positive hydration free energy (both the electrostatic and van der Waals interactions between solutes and water give negative solute–solvent energies) To deal with this problem, Carlson and Jorgensen introduced an additional term which was related to the penalty for forming a solute cavity [Carlsen and Jorgensen 1995]. This third term was proportional to the solvent-accessible surface area:

$$\Delta G_{\text{hyd}} = \beta\langle \mathscr{V}^{\text{el}} \rangle + \alpha\langle \mathscr{V}^{\text{vdw}} \rangle + \gamma SASA \tag{11.44}$$

In their work on hydration, Carlson and Jorgensen attempted to fit the three coefficients α, β and γ, obtaining the best fit for $\alpha = 0.4$, $\beta = 0.45$ and $\gamma = 0.03$ kcal/(mol Å^2). In a subsequent study of the binding of a series of sulphonamide inhibitors to the enzyme thrombin, however, these parameter values were found to be ineffective and that new values were required to give an acceptable fit to the experimental data, with β now being much reduced in value (0.146) [Jones-Hertzog and Jorgensen 1997]. As more variables are considered for inclusion in an LIE-like relationship, it is important that a statistically correct strategy is employed to ensure that the 'optimal' equation is derived (the one with the most predictive power) The techniques to derive such equations are discussed in Section 12.12 on quantitative structure–activity relationships; one study where they were successfully employed considered the binding of a series of inhibitors to the enzyme neuraminidase [Wall *et al.* 1999].

Other attempts to predict free energies from a single simulation have explored the relationship between the coupling parameter, λ, and the free energy. Specifically, the free energy is

expressed as a Taylor series expansion in terms of λ around the point $\lambda = 0$. This expansion is [Smith and van Gunsteren 1994a; Liu *et al.* 1996] (showing just the first two terms explicitly):

$$A(\lambda) = A(\lambda) - A(0) = A'_{\lambda=0}\lambda + \frac{1}{2!}A''_{\lambda=0}\lambda^2 + \frac{1}{3!}A'''_{\lambda=0}\lambda^3 + \cdots$$

$$= \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda \lambda + \frac{1}{k_B T}\left\langle \left(\frac{\partial \mathcal{H}}{\partial \lambda} - \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_0 \right)^2 \right\rangle_0 + \cdots \qquad (11.45)$$

Truncating this series after the first derivative and integrating provides the basis for the thermodynamic integration approach. Moreover, if the Taylor series expansion is continued until it converges then Equation (11.45) is equivalent to the thermodynamic perturbation formula, so providing a link between the two approaches. In practice, it is always necessary to truncate the series; the problem then is whether it is appropriate to assume that the discarded higher-order terms are zero. A good way to test this approach is to consider model systems where the free energy change is known to be zero One such system involves a simple diatomic molecule in a box of water. Each atom of the diatomic molecule is assigned a charge, equal in magnitude (0.25) but of opposite signs. The state to which this system is 'perturbed' corresponds to simply switching the charges (i.e. the start and final states are equivalent) For this system, a standard free energy perturbation calculation can give an answer very close to zero The series expansion was not able to reproduce this result well, even from a 1 ns simulation However, if one considered an alternative problem involving a change from $\lambda = 0$ to $\lambda = 0\,5$ (which corresponds to decharging the system) the series expansion did give a very good result. Nevertheless, all these methods were found to fail for calculations involving the creation or deletion of atoms (a problem we discussed above) In the same paper, Liu and colleagues suggested an interesting method that could not only overcome this problem but also enable many free energy values for a series of related ligands to be obtained from a single simulation At those positions where atoms are created or deleted, soft-core interaction sites are used of the form in Equation (11 30). A single long simulation of this (non-physical) reference state is performed. The soft-core potential has a functional form such that solvent molecules can sometimes penetrate 'within' the usual van der Waals radius. This extends the configurational space accessible to the system. Estimates of free energy differences can be obtained by running through the trajectory, substituting the soft-core sites for the appropriate 'real' atoms and calculating the energy for incorporation into the free-energy perturbation formula. In a 'proof-of-concept' illustration, the free energies of hydration for a series of small molecules were calculated from a single simulation consisting of a soft-core cavity in water [Schäfer *et al.* 1999]. These calculations suggested that the efficiency gains over conventional free energy calculations could reach 2–3 orders of magnitude but that the method did require some further development for certain types of system.

Appendix 11.1 Calculating Free Energy Differences Using Thermodynamic Integration

If the free energy, A, is a continuous function of λ then we can write:

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda}\, d\lambda \tag{11.101}$$

Now

$$A(\lambda) = -k_B T \ln Q(\lambda) \tag{11.102}$$

Thus

$$\Delta A = -k_B T \int_0^1 \left[\frac{\partial \ln Q(\lambda)}{\partial \lambda}\right] d\lambda = \int_0^1 \frac{-k_B T}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda}\, d\lambda \tag{11.103}$$

From the definition of Q (Section 6.1.1):

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \iint d\mathbf{p}^N\, d\mathbf{r}^N \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \tag{11.104}$$

we can write the following for $\partial Q(\lambda)/\partial\lambda$:

$$\frac{\partial Q(\lambda)}{\partial \lambda} = \frac{1}{N!}\frac{1}{h^{3N}} \iint d\mathbf{p}^N\, d\mathbf{r}^N \frac{\partial}{\partial \lambda} \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \tag{11.105}$$

Applying the chain rule:

$$\frac{\partial Q(\lambda)}{\partial \lambda} = -\frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{k_B T} \iint d\mathbf{p}^N\, d\mathbf{r}^N \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \tag{11.106}$$

Substituting back into the expression for $\partial A/\partial\lambda$ gives:

$$\begin{aligned}\frac{\partial A(\lambda)}{\partial \lambda} &= \frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{Q(\lambda)} \iint d\mathbf{p}^N\, d\mathbf{r}^N \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \\ &= \iint d\mathbf{p}^N\, d\mathbf{r}^N \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \left\{\frac{\exp[-\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)/k_B T]}{Q(\lambda)}\right\} = \left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N, \lambda)}{\partial \lambda} \right\rangle_\lambda\end{aligned} \tag{11.107}$$

Thus

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{11.108}$$

Appendix 11.2 Using the Slow Growth Method for Calculating Free Energy Differences

The slow growth expression can be derived from the thermodynamic perturbation expression (Equation (11.7)) if it is written as a Taylor series:

$$\Delta A = -k_B T \sum_{i=0}^{N_{step}-1} \ln\langle \exp(-[\mathcal{H}(\lambda_{i+1}) - \mathcal{H}(\lambda_i)]/k_B T)\rangle_{NVT} \tag{11.109}$$

$$\Delta A \approx -k_B T \sum_{i=0}^{N_{step}-1} \ln\langle 1 - [\mathcal{H}(\lambda_{i+1}) - \mathcal{H}(\lambda_i)]/k_B T + \cdots\rangle_{NVT} \tag{11.110}$$

$$\Delta A \approx -k_B T \sum_{i=0}^{N_{step}-1} \ln\left\{1 - \frac{1}{k_B T}\langle[\mathcal{H}(\lambda_{i+1}) - \mathcal{H}(\lambda_i)]\rangle_{NVT} + \cdot\right\} \tag{11.111}$$

$$\Delta A \approx \sum_{i=0}^{N_{step}-1} \langle[\mathcal{H}(\lambda_{i+1}) - \mathcal{H}(\lambda_i)]\rangle_{NVT} \tag{11.112}$$

Appendix 11.3 Expansion of Zwanzig Expression for the Free Energy Difference for the Linear Response Method

The starting point is the standard expression for the free energy difference, Equation (11.6):

$$\Delta A = -k_B T \ln\langle \exp[-(\mathcal{H}_Y - \mathcal{H}_X)/k_B T]\rangle_0 \tag{11.113}$$

We expand the exponential:

$$\Delta A = -k_B T \ln\left\langle 1 - \frac{(\mathcal{H}_Y - \mathcal{H}_X)}{k_B T} + \frac{(\mathcal{H}_Y - \mathcal{H}_X)^2}{2(k_B T)^2} - \cdots \right\rangle_0$$

$$= -k_B T \ln\left[1 - \frac{\langle \mathcal{H}_Y - \mathcal{H}_X\rangle_0}{k_B T} + \frac{\langle(\mathcal{H}_Y - \mathcal{H}_X)^2\rangle_0}{2(k_B T)^2} - \cdots\right] \tag{11.114}$$

Using the series expansion of $\ln(1+x)$ gives:

$$\Delta A = -k_B T\left\{ -\frac{\langle \mathcal{H}_Y - \mathcal{H}_X\rangle_0}{k_B T} + \frac{\langle(\mathcal{H}_Y - \mathcal{H}_X)^2\rangle_0}{2(k_B T)^2} \right.$$

$$\left. -\frac{1}{2}\left[\left(\frac{\langle \mathcal{H}_Y - \mathcal{H}_X\rangle_0}{k_B T}\right)^2 - \frac{\langle \mathcal{H}_Y - \mathcal{H}_X\rangle_0\langle(\mathcal{H}_Y - \mathcal{H}_X)^2\rangle_0}{2(k_B T)^3} + \left(\frac{\langle(\mathcal{H}_Y - \mathcal{H}_X)^2\rangle_0}{2(k_B T)^2}\right)^2\right]\right\} \tag{11.115}$$

This can be rearranged to:

$$\Delta A = \langle \mathscr{H}_Y - \mathscr{H}_X \rangle_0 - \frac{1}{2k_B T} \langle [(\mathscr{H}_Y - \mathscr{H}_X) - \langle \mathscr{H}_Y - \mathscr{H}_X \rangle_0]^2 \rangle_0 + \cdots \quad (11.116)$$

A similar procedure applied to the result from averaging at Y gives:

$$\Delta A = \langle \mathscr{H}_Y - \mathscr{H}_X \rangle_1 + \frac{1}{2k_B T} \langle [(\mathscr{H}_Y - \mathscr{H}_X) - \langle \mathscr{H}_Y - \mathscr{H}_X \rangle_1]^2 \rangle_1 + \cdots \quad (11.117)$$

When Equations (11.116) and (11.117) are added together and we substitute $\Delta\mathscr{H}$ for $\mathscr{H}_Y - \mathscr{H}_X$ then we obtain:

$$\Delta A = \tfrac{1}{2}[\langle \Delta\mathscr{H} \rangle_0 + \langle \Delta\mathscr{H} \rangle_1] - \frac{1}{4k_B T}[\langle (\Delta\mathscr{H} - \langle \Delta\mathscr{H} \rangle_0)^2 \rangle_0 - \langle (\Delta\mathscr{H} - \langle \Delta\mathscr{H} \rangle_1)^2 \rangle_1] + \cdots \quad (11.118)$$

Further Reading

Allan N L and W C Mackrodt 1997. High-T_c Superconductors in Computer Modelling In Catlow C R A (Editor) *Inorganic Crystallography*, pp. 241–268.

Amara P and M J Field 1998. Combined Quantum Mechanical and Molecular Mechanical Potentials. In Schleyer, P v R, N L Allinger , T Clark, J Gasteiger, P A Kollman H F Schaefer III and P R Schreiner (Editors). *The Encyclopedia of Computational Chemistry*. Chichester, John Wiley & Sons.

Beveridge D L and F M DiCapua 1989. Free Energy via Molecular Simulation: A Primer. In van Gunsteren W F and P K Weiner (Editors) *Computer Simulation of Biomolecular Systems* Leiden, ESCOM, pp. 1–26

Catlow C R A 1994. An Introduction to Disorder in Solids. In NATO ASI Series C 418 (*Defects and Disorder in Crystalline and Amorphous Solids*), pp. 1–23

Catlow C R A 1994. Molecular Dynamics Studies of Defects in Solids In NATO ASI Series C 418 (*Defects and Disorder in Crystalline and Amorphous Solids*), pp. 357–373.

Catlow C R A, R G Bell and J D Gale 1994. Computer Modelling as a Technique in Materials Chemistry. *Journal of Materials Chemistry* 4:781–792

Catlow C R A and W C Mackrodt 1982. Theory of Simulation Methods for Lattice and Defect Energy Calculations in Crystals. In *Lecture Notes in Physics* **166** (Comput. Simul. Solids), pp 3–20.

Chadwick A V and J Corish 1997. Defects and Matter Transport in Solid Materials In NATO ASI Series C 498 (*New Trends in Materials Chemistry*), pp. 285–318.

Cramer C J and Truhlar D G 1995. Continuum Solvation Models· Classical and Quantum Mechanical Implementations In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 6. New York, VCH Publishers, pp. 1–72

Gale J 1999. *General Utility Lattice Program Manual*, Imperial College, London.

Gao J 1995. Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 7 New York, VCH Publishers, pp. 119–185.

Gillan M J 1989. *Ab Initio* Calculation of the Energy and Structure of Solids. *Journal of the Chemical Society Faraday Transactions 2* **85**:521–536.

Gillan M J 1997. The Virtual Matter Laboratory. *Contemporary Physics* **38**:115–130

Harding J H 1997 Defects, Surfaces and Interfaces In Catlow C R A (Editor) *Inorganic Crystallography*, pp. 185–199.

Jorgensen W L 1983. Theoretical Studies of Medium Effects on Conformational Equilibria *Journal of Physical Chemistry* **87**:5304–5314

King P M 1993. Free Energy via Molecular Simulation: A Primer. In van Gunsteren W F, P K Weiner and A J Wilkinson (Editors) *Computer Simulation of Biomolecular Systems* Volume 2 Leiden, ESCOM, pp 267–314.

Kollman P A 1993. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena *Chemical Reviews* **93**:2395–2417

Lybrand T P 1990, Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 1. New York, VCH Publishers, pp 295–320.

Mark A E and van Gunsteren W F 1995 Free Energy Calculations in Drug Design: A Practical Guide In Dean P M, G Jolles and C G Newton (Editors) *New Perspectives in Drug Design* London, Academic Press, pp 185–200.

Mezei M and D L Beveridge 1986. Free Energy Simulations In Beveridge D L and W L Jorgensen (Editors) *Computer Simulation of Chemical and Biomolecular Systems. Annals of the New York Academy of Sciences* **482**:1–23.

Sandre E and A Pasturel 1997. An Introduction to *Ab-Initio* Molecular Dynamics Schemes *Molecular Simulation* **20**:63–77.

Straatsma T P 1996 Free Energy by Molecular Simulation. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 9 New York, VCH Publishers, pp 81–127.

van Gunsteren W F 1989. Methods for Calculation of Free Energies and Binding Constants. Successes and Problems. In van Gunsteren and P K Weiner (Editors) *Computer Simulation of Biomolecular Systems*. Leiden, ESCOM, pp. 27–59.

References

Allan N L and W C Mackrodt 1994. Oxygen Interstitial Defects in High-T_c Oxides. *Molecular Simulation* **12**:89–100.

Åqvist J, C Medina and J-E Samuelsson 1994 A New Method for Predicting Binding Affinity in Computer-aided Drug Design *Protein Engineering* **7**:385–391.

Åqvist J and A Warshel 1993. Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches. *Chemical Reviews* **93**:2523–2544

Åqvist J, M Fothergill and A Warshel 1993. Computer Simulation of the CO_2/HCO_3^- Interconversion Step in Human Carbonic Anhydrase I. *Journal of the American Chemical Society* **115**:631–635

Barrows S E, J W Storer, C J Cramer, A D French and D G Truhlar 1998 Factors Controlling Relative Stability of Anomers and Hydroxymethyl Conformers of Glucopyranose. *Journal of Computational Chemistry* **19**:1111–1129.

Bartlett P A and C K Marlowe 1987. Evaluation of Intrinsic Binding Energy from a Hydrogen-bonding Group in an Enzyme Inhibitor. *Science* **235** 569–571

Bash P A, U C Singh, F K Brown, R Langridge and P A Kollman 1987. Calculation of the Relative Change in Binding Free-Energy of a Protein-Inhibitor Complex. *Science* **235** 574–576.

Bernardi A, A M Capelli, A Comotti, C Gannari, J M Goodman and I Paterson 1990. Transition-State Modeling of the Aldol Reaction of Boron Enolates: A Force Field Approach. *Journal of Organic Chemistry* **55**:3576–3581

Boero M, M Parrinello and K Terakura 1999 Ziegler–Natta Heterogeneous Catalysis by First Principles Computer Experiments *Surface Science* **438**:1–8.

Boresch S, G Archontis and M Karplus 1994 Free Energy Simulations: The Meaning of the Individual Contributions from a Component Analysis. *Proteins: Structure, Function and Genetics* **20**.25–33

Boresch S and M Karplus 1995. The Meaning of Component Analysis· Decomposition of the Free Energy in Terms of Specific Interactions. *Journal of Molecular Biology* **254**:801–807.

Born M 1920 Volumen and Hydratationswärme der Ionen *Zeitschrift fur Physik* 1:45–48

Buetler T C, A E Mark, R C van Schaik, P R Gerber and W F van Gunsteren 1994. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations *Chemical Physics Letters* **222**:529–539

Burger M T, A Armstrong, F Guarnieri, D Q McDonald and W C Still 1994. Free Energy Calculations in Molecular Design: Predictions by Theory and Reality by Experiment with Enantioselective Podand Ionophores. *Journal of the American Chemical Society* **116**:3593–3594

Car R and M Parrinello 1985. Unified Approach for Molecular Dynamics and Density Functional Theory *Physical Review Letters* **55**:2471–2474.

Carlson H A and W L Jorgensen 1995 An Extended Linear Response Method for Determining Free Energies of Hydration *Journal of Physical Chemistry* **99**.10667–10673

Chambers C C, G D Hawkins, C J Cramer and D G Truhlar 1996. Model for Aqueous Solvation Based on Class IC Atomic Charges and First Solvation Shell Effects *Journal of Physical Chemistry* **100** 16385–16398

Chandrasekhar J and W L Jorgensen 1985. Energy Profile for a Nonconcerted S_N2 Reaction in Solution. *Journal of the American Chemical Society* **107**:2974–2975.

Chandrasekhar J, S F Smith and W L Jorgsensen 1985. Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution *Journal of the American Chemical Society* **107**·154–163.

Claverie P, J P Daudey, J Langlet, B Pullman, D Piazzola and M J Huron 1978 Studies of Solvent Effects I. Discrete, Continuum and Discrete-Continuum Models and Their Comparison for Some Simple Cases. NH_4^+, CH_3OH and substituted NH_4^+. *Journal of Physical Chemistry* **82**:405–418.

Constanciel R and R Contreras 1984. Self-Consistent Field Theory of Solvent Effects Representation by Continuum Models - Introduction of Desolvation Contribution. *Theoretica Chimica Acta* **65**.1–11.

Cramer C J and D G Truhlar 1992. AM1-SM2 and PM3-SM3 Parametrized SCF Solvation Models for Free Energies in Aqueous Solution *Journal of Computer-Aided Molecular Design* **6**:629–666

Dapprich S, I Komiromi, K S Byun, K Morokuma and M J Frisch 1999. A New ONIOM Implementation in Gaussian '98 Part I The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *THEOCHEM* **461–462**:1–21.

de Wijs G A, G Kresse, L Vočadlo, D Dobson, D Alfè, M J Gillan and G D Price 1998. The Viscosity of Liquid Iron at the Physical Conditions of the Earth's Core *Nature* **392**:805–807

Eisenberg D and A D McLachlan 1986 Solvation Energy in Protein Folding and Binding. *Nature* **319**·199–203.

Elber R and M Karplus 1990. Enhanced Sampling in Molecular Dynamics: Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion through Myoglobin *Journal of the American Chemical Society* **112**.9161–9175.

Eriksson M A L, J Pitera and P A Kollman 1999. Prediction of the Binding Free Energies of New TIBO-like HIV-1 Reverse Transcriptase Inhibitors Using a Combination of PROFEC, PB/SA, CMC/MD, and Free Energy Calculations. *Journal of Medicinal Chemistry* **42**·868–881

Essex J W, C A Reynolds and W G Richards 1989. Relative Partition Coefficients from Partition Functions A Theoretical Approach to Drug Transport *Journal of the Chemical Society Chemical Communications* 1152–1154

Field M J, P A Bash and M Karplus 1990. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *Journal of Computational Chemistry* **11**:700–733

Fleischman S H and C L Brooks III 1987 Thermodynamics of Aqueous Solvation - Solution Properties of Alcohols and Alkanes. *Journal of Chemical Physics* **87**.3029–3037.

Floris F and J Tomasi 1989 Evaluation of the Dispersion Contribution to the Solvation Energy - A Simple Computational Model in the Continuum Approximation *Journal of Computational Chemistry* **10**.616–627

Freitag S, J Le Trong, P S Stayton and R E Stenkamp 1997. Structural Studies of the Streptavidin Binding Loop *Protein Science* **6** 1157.

Gilson M K and B Honig 1988. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies and Conformational Analysis *Proteins Structure, Function and Genetics* 4:7–18.

Gonzalez C and H B Schlegel 1988. An Improved Algorithm for Reaction Path Following. *Journal of Chemical Physics* **90**:2154–2161.

Grimes R W, C R A Catlow and A M Stoneham 1989 Quantum-mechanical Cluster Calculations and the Mott–Littleton Methodology *Journal of the Chemical Society, Faraday Transactions* **85** 485–495.

Guo Z and C L Brooks III 1998. Rapid Screening of Binding Affinities: Application of the λ-Dynamics Method to a Trypsin–Inhibitor System *Journal of the American Chemical Society* **120**.1920–1921.

Guo Z, C L Brooks III and X Kong 1998 Efficient and Flexible Algorithm for Free Energy Calculations using the λ-Dynamics Approach. *Journal of Physical Chemistry* **B102**:2032–2036

Ha S, J Gao, B Tidor, J W Brady and M Karplus 1991. Solvent Effect on the Anomeric Equilibrium in D-Glucose· A Free Energy Simulation Analysis *Journal of the American Chemical Society* **113**·1553–1557.

Hansson T and J Åqvist 1995 Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations *Protein Engineering* **8**.1137–1144

Hansson T, J Marelius and J Åqvist 1998. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *Journal of Computer-Aided Molecular Design* **12**:27–35

Hasel W, T F Hendrickson and W C Still 1988 A Rapid Approximation to the Solvent Accessible Surface Areas of Atoms *Tetrahedron Computer Methodology* 1:103–116

Honig B and A Nicholls 1995. Classical Electrostatics in Biology and Chemistry. *Science* **268**·1144–1149.

Jones-Hertzog D K and W L Jorgensen 1997. Binding Affinities for Sulphonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method. *Journal of Medicinal Chemistry* **40**·1539–1549

Jorgensen W L, J M Briggs and M L Contreras 1990. Relative Partition Coefficients for Organic Solutes from Fluid Simulations *Journal of Physical Chemistry* **94**.1683–1986.

Jorgensen W L and J K Buckner 1987. Use of Statistical Perturbation Theory for Computing Solvent Effects on Molecular Conformation. Butane in Water *Journal of Physical Chemistry* **91**:6083–6085.

Jorgensen W L, J K Buckner, S Boudon and J Tirado-Reeves 1988. Efficient Computation of Absolute Free Energies of Binding by Computer Simulations - Applications to the Methane Dimer in Water. *Journal of Chemical Physics* **89**·3742–3746

Jorgensen W L, J Gao and C Ravimohan 1985. Monte Carlo Simulations of Alkanes in Water· Hydration Numbers and the Hydrophobic Effect. *Journal of Physical Chemistry* **89**:3470–3473

Kirkwood J G 1934 Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *Journal of Chemical Physics* **2**:351–361.

Klamt A 1995. Conductor-like Screening Model for Real Solvent: A New Approach to the Quantitative Calculation of Solvation Phenomena. *Journal of Physical Chemistry* **99** 2224–2235

Klamt A, V Jonas, T. Bürger and J C W Lohrenz 1998 Refinements and Parametrisation of COSMO-RS *Journal of Physical Chemistry* **102**:5074–5085.

Klamt A and G Schüürmann 1993. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient *Journal of the Chemical Society, Perkin Transations* **2**:799–805

Klapper I, R Hagstrom, R Fine, K Sharp and B Honig 1986 Focusing of Electric Fields in the Active Site of CuZn Superoxide Dismutase: Effects of Ionic Strength and Amino-Acid Substitution *Proteins Structure, Function and Genetics* **1**:47–59

Kong X and C L Brooks III 1996 λ-Dynamics: A New Approach to Free Energy Calculations. *Journal of Chemical Physics* **105**.2414–2423.

Laasonen, M Sprik and M Parrinello 1993. 'Ab Initio' Liquid Water. *Journal of Chemical Physics* **99**.9080–9089.

Liu H, A E Mark and W F van Gunsteren 1996 Estimating the Relative Free Energy of Different Molecular States with Respect to a Single Reference State. *Journal of Physical Chemistry* **100**·9485–9494

Lybrand T P, J A McCammon and G Wipff 1986 Theoretical Calculation of Relative Binding Affinity in Host–Guest Systems. *Proceedings of the National Academy of Sciences USA* **83**:833–835.

Mackrodt W C 1982 Defect Calculations for Ionic Materials *Lecture Notes in Physics* **166** (Computer Simulation of Solids).175–194

Marquart M, J Walter, J Deisenhofer, W Bode and R Huber 1983. The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and its Complexes with Inhibitors. *Acta Crystallographica* **B39**:480–490.

Maseras F and K Morokuma 1995 IMOMM· A New Integrated *Ab Initio* + Molecular Mechanics Geometry Optimisation Scheme of Equilibrium Structures and Transition States. *Journal of Computational Chemistry* **16**.1170–1179.

McRee D E, S M Redford, E D Getzoff, J R Lepock, R A Hallewell and J A Tainer 1990. Changes in Crystallographic Structure and Thermostability of a Cu, Zn Superoxide Dismutase Mutant Resulting from the Removal of Buried Cysteine. *Journal of Biological Chemistry* **265**·14234–14241

Merz K M Jr and P A Kollman 1989 Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor. *Journal of the American Chemical Society* **111**·5649–5658

Miertus S, E Scrocco and J Tomasi 1981 Electrostatic Interaction of a Solute with a Continuum – A Direct Utilization of *Ab Initio* Molecular Potentials for the Provision of Solvent Effects. *Chemical Physics* **55**.117–129.

Miick S M, G V Martinez, W R Fiori, A P Todd and G L Millhauser 1992. Short Alanine-based Peptides May Form 3(10)-Helices and not Alpha-helices in Aqueous Solution *Nature* **359**:653–655

Mitchell M J and J A McCammon 1991 Free Energy Difference Calculations by Thermodynamic Integration· Difficulties in Obtaining a Precise Value. *Journal of Computational Chemistry* **12**.271–275.

Miyamoto S and P A Kollman 1993a Absolute and Relative Binding Free Energy Calculations of the Interaction of Biotin and its Analogues with Streptavidin Using Molecular Dynamics/Free Energy Perturbation Approaches. *Proteins: Structure, Function and Genetics* **16**·226–245.

Miyamoto S and P A Kollman 1993b What Determines the Strength of Noncovalent Association of Ligands to Proteins in Aqueous Solution? *Proceedings of the National Academy of Sciences USA* **90**·8402–8406

Mott N F and M J Littleton 1938. Conduction in Polar Crystals. I Electrolytic Conduction in Solid Salts *Transactions of the Faraday Society* **34**: 485–499.

Onsager L 1936 Electric Moments of Molecules in Liquids *Journal of the American Chemical Society* **58**.1486–1493

Paschual-Ahuir J L, E Silla, J Tomasi and R Bonaccorsi 1987. Electrostatic Interaction of a Solute with a Continuum. Improved Description of the Cavity and of the Surface Cavity Bound Charge Distribution. *Journal of Computational Chemistry* **8** 778–787

Payne M C, M P Teter, D C Allan, R A Arias and D J Joannopoulos 1992 Iterative Minimisation Techniques for *Ab Initio* Total-Energy Calculations. Molecular Dynamics and Conjugate Gradients. *Reviews of Modern Physics* **64**.1045–1097

Pearlman D A and P A Kollman 1989. A New Method for Carrying Out Free-Energy Perturbation Calculations – Dynamically Modified Windows *Journal Of Chemical Physics* **90**:2460–2470.

Pierotti R 1965. Aqueous Solutions of Nonpolar Gases *Journal of Physical Chemistry* **69**:281–288

Pisani C 1999 Software for the Quantum-mechanical Simulation of the Properties of Crystalline Materials: State of the Art and Prospects. *THEOCHEM* **463**:125–137

Pitera J and P Kollman 1998 Designing an Optimum Guest for a Host Using Multimolecule Free Energy Calculations: Predicting the Best Ligand for Rebek's 'Tennis Ball'. *Journal of the American Chemical Society* **120**:7557–7567.

Postma J P M, H J C Berendsen and J R Haak 1982 Thermodynamics of Cavity Formation in Water *Faraday Symposium of the Chemical Society* 17:55–67.

Qiu D, P S Shenkin, F P Hollinger and W C Still 1997. The GB/SA Continuum Model for Solvation A Fast Analytical Method for the Calculation of Approximate Born Radii. *Journal of Physical Chemistry* 101 3005–3014.

Rashin A A 1990. Hydration Phenomena, Classical Electrostatics, and the Boundary Element Method. *Journal of Physical Chemistry* 94·1725–1733.

Rashin A A and B Honig 1985 Reevaluation of the Born Model of Ion Hydration. *Journal of Physical Chemistry* **89** 5588–5593.

Rashin A A and K Namboodiri 1987 A Simple Method for the Calculation of Hydration Enthalpies of Polar Molecules with Arbitrary Shapes. *Journal of Physical Chemistry* **91**.6003–6012.

Remler D K and P A Madden 1990 Molecular Dynamics without Effective Potentials via the Car-Parrinello Approach. *Molecular Physics* **70**.921–966

Reuter N, A Dejaegere, B Maigret and M Karplus 2000. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches *Journal of Physical Chemistry* **A104**:1720–1733.

Rinaldi D, M F Ruiz-Lopez and J L Rivail 1983. *Ab Initio* SCF Calculations on Electrostatically Solvated Molecules Using a Deformable Three Axes Ellipsoidal Cavity *Journal of Chemical Physics* **78**:834–838.

Röthlisberger and M Parrinello 1997. *Ab Initio* Molecular Dynamics Simulation of Liquid Hydrogen Fluoride *Journal of Chemical Physics* **106** 4658–4664.

Ryckaert J-P and A Bellemans 1978 *Molecular Dynamics of Liquid Alkanes, Faraday Discussions* **20**:95–106.

Saitta A M, P D Sooper, E Wasserman and M L Klein 1999 Influence of a Knot on the Strength of a Polymer Strand. *Nature* **399**:46–48

Schäfer H, W F van Gunsteren and A E Mark 1999 Estimating Relative Free Energies from a Single Ensemble Hydration Free Energies. *Journal of Computational Chemistry* **20**:1604–1617

Silvestrelli P L and M Parrinello 1999. Structural, Electronic and Bonding Properties of Liquid Water from First Principles *Journal of Chemical Physics* **111**:3572–3580.

Simmerling C and R Elber 1995. Computer Determination of Peptide Conformations in Water: Different Roads to Structure *Proceedings of the National Academy of Sciences USA* **92**:3190–3193

Simmerling C, T Fox and P A Kollman 1998. Use of Locally Enhanced Sampling in Free Energy Calculations: Testing and Application to the $\alpha \rightarrow \beta$ Anomerisation of Glucose. *Journal of the American Chemical Society* **120**·5771–5782.

Singh U C and P A Kollman 1986. A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $CH_3Cl + Cl^-$ Exchange Reaction and Gas Phase Protonation of Polyethers. *Journal of Computational Chemistry* 7·718–730

Sitkoff D, K A Sharp and B Honig 1994 Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *Journal of Physical Chemistry* **98** 1978–1988

Smith P E and B M Pettitt 1994. Modeling Solvent in Biomolecular Systems *Journal of Physical Chemistry* **98**.9700–9711.

Smith P E and W F van Gunsteren 1994a Predictions of Free Energy Differences from a Single Simulation of the Initial State. *Journal of Chemical Physics* **100**:577–585.

Smith P E and W F van Gunsteren 1994b. When Are Free Energy Components Meaningful? *Journal of Physical Chemistry* **98**:13735–13740

Smythe M L, S E Huston and G R Marshall 1993 Free Energy Profile of a 3_{10} to α-Helical Transition of an Oligopeptide in Various Solvents. *Journal of the American Chemical Society* **115**:11594–11595.

Smythe M L, S E Huston and G R Marshall 1995. The Molten Helix: Effects of Solvation on the α- to 3_{10}-Helical Transition. *Journal of the American Chemical Society* **117** 5445–5452.

Sprik M, J Hutter and M Parrinello 1996. *Ab Initio* Molecular Dynamics Simulation of Liquid Water. Comparison of Three Gradient-corrected Density Functionals *Journal of Chemical Physics* **105**:1142–1152.

Stich I, A De Vita, M C Payne, M J Gilland and L J Clarke 1994. Surface Dissociation from First Principles' Dynamics and Chemistry. *Physical Review* **B49**.8076–8085.

Still W C, A Tempczyrk, R C Hawley and T Hendrickson 1990. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society* **112**:6127–6129

Svensson M, S Humbel, R D J Froese, T Matsubara, S Sieber and K Morokuma 1996. ONIOM. A Multilayered Integrated MO+MM Method for Geometry Optimisations and Single Point Energy Predictions. A Test for Diels–Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *Journal of Physical Chemistry* **100**:19357–19363.

Tapia O and O Goscinski 1975 Self-Consistent Reaction Field Theory of Solvent Effects. *Molecular Physics* **29**:1653–1661.

Taylor M B, G D Barrera, N L Allan, T H K Barron and W C Mackrodt 1997 Free Energy of Formation of Defects in Polar Solids *Faraday Discussions* **106**:377–387

Tirado-Reeves J, D S Maxwell and W L Jorgensen 1993. Molecular Dynamics and Monte Carlo Simulations Favor the α-Helical Form for Alanine-Based Peptides in Water. *Journal of the American Chemical Society* **115**:11590–11593.

Tobias D J and C L Brooks III 1988. Molecular Dynamics with Internal Coordinate Constraints. *Journal of Chemical Physics* **89**:5115–5126.

Torrie G M and J P Valleau 1977 Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation Umbrella Sampling. *Journal of Computational Physics* **23** 187–199

Tuckerman M, K Laasonen, M Sprik and M Parrinello 1995a. *Ab Initio* Molecular Dynamics Simulation of the Solvation and Transport of Hydronium and Hydroxyl Ions in Water. *Journal of Chemical Physics* **103**:150–161

Tuckerman M, K Laasonen, M Sprik and M Parrinello 1995b *Ab Initio* Molecular Dynamics Simulation of the Solvation and Transport of H_3O^+ and OH^- Ions in Water. *Journal of Physical Chemistry* **99**:5749–5752

Wall I D, A R Leach, D W Salt, M G Ford and J W Essex 1999. Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method. *Journal of Medicinal Chemistry* **42**:5142–5152

Wang W, J Wang and P A Kollman 1999. 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: Structure, Function and Genetics* **34**:395–402.

Warshel A 1991. *Computer Modelling of Chemical Reactions in Enzymes and Solutions* New York, John Wiley & Sons

Warshel A and M Levitt 1976. Theoretical Studies of Enzymic Reactions Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *Journal of Molecular Biology* **103**:227–249.

Warwicker J and H C Watson 1982 Calculation of the Electric Potential in the Active-Site Cleft Due to Alpha-Helix Dipoles. *Journal of Molecular Biology* **157**:671–679

Wodak S J and J Janin 1980 Analytical Approximation to the Solvent Accessible Surface Area of Proteins *Proceedings of the National Academy of Sciences USA* **77**:1736–1740.

Wong M W, K B Wiberg and M J Frisch 1992 Solvent Effects 3 Tautomeric Equilibria of Formamide and 2-Pyridone in the Gas Phase and Solution. An *Ab Initio* SCRF Study. *Journal of the American Chemical Society* **114**:1645–1652.

Yu H-A and M Karplus 1988 A Thermodynamic Analysis of Solvation *Journal of Chemical Physics* **89** 2366–2379

Zhang L and J Hermans 1994. 3_{10}-Helix versus α-Helix: A Molecular Dynamics Study of Conformational Preferences of Aib and Alanine. *Journal of the American Chemical Society* **116**:11915–11921.

Zhang X and C R A Catlow 1992 Molecular Dynamics Study of Oxygen Diffusion in $YBa_2Cu_3O_{6.19}$. *Physical Review* **B46**:457–462

Zwanzig R W 1954. High-temperature Equation of State by a Perturbation Method. I. Nonpolar Gases *Journal of Chemical Physics* **22**:1420–1426