

This article was downloaded by:

On: 14 January 2011

Access details: *Access Details: Free Access*

Publisher *Taylor & Francis*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Molecular Simulation

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713644482>

Hysteresis and Statistical Errors in Free Energy Perturbation L to D Amino Acid Conversion

Olle Edholm^a; Indira Ghosh^b

^a Department of Theoretical Physics, Royal Institute of Technology, Stockholm, Sweden ^b Astra Research Center India, Bangalore, India

To cite this Article Edholm, Olle and Ghosh, Indira(1993) 'Hysteresis and Statistical Errors in Free Energy Perturbation L to D Amino Acid Conversion', *Molecular Simulation*, 10: 2, 241 — 253

To link to this Article: DOI: 10.1080/08927029308022167

URL: <http://dx.doi.org/10.1080/08927029308022167>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

HYSTERESIS AND STATISTICAL ERRORS IN FREE ENERGY PERTURBATION L TO D AMINO ACID CONVERSION

OLLE EDHOLM

*Department of Theoretical Physics, Royal Institute of Technology,
S-100 44 Stockholm 70, Sweden*

INDIRA GHOSH

*Astra Research Center India, 18th Cross Road, Malleswaram,
Bangalore, 560 003 India*

(Received December 1992, accepted December 1992)

A theory based on a Langevin equation along the reaction coordinate is developed to explain and calculate systematic and statistical errors in free energy perturbation simulations. The errors are calculated exactly when both the perturbation potential and the mean potential from the surrounding degrees of freedom are harmonic in the reaction coordinate. The effect of the mean potential is small as long as the force constant is small compared to the force constant of the perturbation potential. This indicates that the results obtained with zero mean force may still be valid as long as the second derivative of the mean potential is small compared to that of the perturbation potential. The theory is applied to conversion between L and D amino acids by changing the position of the minimum of the harmonic improper dihedral potential between ± 35.264 degrees. For phenylalanine bound in the active site of a protein (thermolysin) we find from 20 psec. simulations statistical errors and hysteresis that both are about 2.5 kJ/mol in agreement with what is obtained from the theoretical predictions. The statistical errors are proportional to the square root of the coupling to the heat bath and inversely proportional to the square root of integration time while the (positive) hysteresis due to that the reaction coordinate lags behind is linear in the same quantities. This shows that the systematic errors will dominate in short simulations while the statistical ones will dominate for long simulations. The treatment is based on that the systematic influence of the surroundings can be represented by a mean force upon the reaction coordinate. If the relaxation processes of the environment are slow this may not be true. Then additional errors have to be considered.

KEY WORDS: Free energy perturbation, errors, amino acid, hysteresis

INTRODUCTION

The calculation of free energy differences by integration over a parameter in the potential energy is an old idea (1). It is, however, first in the last ten years that it has become widely used (2). For a recent review see (3).

The method is based upon a continuous conversion between two states described by different potential energy functions. This is done by introducing a potential energy function $E(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \lambda)$ that depends not only of all the atomic coordinates but also of a parameter λ that gives the two different potential energy functions for the limiting values zero and one. Then, the free energy as a function of the parameter λ is

$$F(\lambda) = -k_B T \ln \int \dots \int \exp(-E(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \lambda)/k_B T) d^{3N}r + \text{constant terms} \quad (1)$$

with $k_B T$ being the Boltzmann constant times absolute temperature. The integration is over all degrees of freedom of the system. Derivativation with respect to λ , gives

$$\frac{\partial F}{\partial \lambda} = \left\langle \frac{\partial E}{\partial \lambda} \right\rangle \quad (2)$$

with $\langle \dots \rangle$ meaning a canonical ensemble average. Integration of this gives finally:

$$F(\lambda) - F(0) = \int_0^\lambda \left\langle \frac{\partial E(\lambda')}{\partial \lambda'} \right\rangle d\lambda'. \quad (3)$$

The sampling can be done in different ways. The simplest way (4), which also will be used here is to change λ slowly linearly in time during the run and use all data. Alternatively one may change λ in steps, equilibrate and then sample. This is then done over a number of discrete values of λ and is called a windowing method (5), (6). Equation (3) looks misleadingly simple and one should realize that it involves an averaging over all the degrees of freedom in the system. It is not as easy as it may seem to get accurate and reliable results. There are many traps and fallacies some of which have been discussed in (7) and (8). Here we will try to create a framework in which the errors can be analyzed when there is a well defined reaction coordinate for the integration. The error estimation has been discussed in somewhat similar terms in (9) and (10).

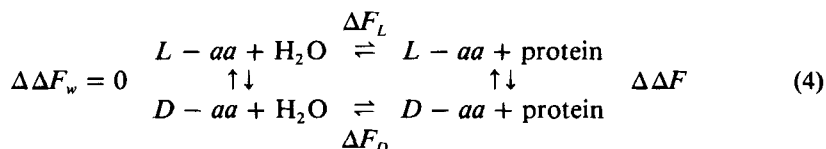
We will here consider a situation where a potential energy minimum is moving during the free energy integration. This means that there exists a well defined simple reaction coordinate. One example is *L* to *D* amino acid conversion. The stereoisomers differ by having opposite signs on the improper dihedral angle at the C_α atom. One may then define an improper dihedral potential that is harmonic and has a minimum that may be shifted continuously between the two states. Thus the free energy difference between the *L* and *D* amino acid bound to a protein may be calculated. Other variables will change during this conversion, but anyhow the improper dihedral will serve as a suitable reaction coordinate. This situation is simple since the reaction coordinate is one-dimensional and the perturbing force is linear in this reaction coordinate. The question of difference in binding between *L* and *D* amino acids is in itself of considerable importance.

However, there are also other situations to which a similar scheme could be applied with simple modifications. One is structural conformation changes in proteins, which may be characterized by the change of a number of dihedral angles (or virtual dihedrals (i.e. dihedrals along an imaginary chain connecting the C_α -atoms)). Then it may be suitable to introduce harmonic potentials in these angles with minima that are slowly shifted between the positions of the angles in the two structures. In this case one may have to slowly switch this potential on in the beginning and then slowly switch it off at the end to get the contributions to the free energy due to the difference in the importance of the harmonic restraining in the two cases.

Both these situation differ from for instance that when the charge of some atom is changed during the free energy integration. Such a change will cause changes in the structure of surrounding protein and solvent and it is not easy to define a simple reaction coordinate.

We will here take the *L* to *D* amino acid conversion as an example of a transition driven by a harmonic potential along a well defined reaction coordinate. The aim is to get a theoretical estimate of statistical and systematic errors, to be able to correct for the latter if necessary and to be able to make a rational choice between different ways of doing the simulations.

The reaction scheme considered here is the following



We want to calculate the difference in free energy of binding between the *L* and *D* amino acids, $\Delta F_D - \Delta F_L$. Since the sum of the free energy changes in the cyclic reaction scheme must be zero and $\Delta\Delta F_w$ is zero we obtain this as:

$$\Delta F_D - \Delta F_L = \Delta\Delta F \quad (5)$$

We will for simplicity drop one Δ and denote this difference ΔF . The stereoisomeric state is characterized by an improper dihedral angle that defines the direction of the $C_\alpha - C_\beta$ bond with respect to the backbone. The hydrogen atom of the C_α is as for all the other aliphatic carbons included into a united atom in the simulations here. The conversion can therefore be done by replacing the ordinary potential for the improper dihedral angle, χ with the λ -dependent perturbing potential:

$$U_p(\chi, \lambda) = \frac{C_\chi}{2} (\chi - (1 - 2\lambda)\chi_0)^2 \quad (6)$$

where, C_χ is a force constant ($340 \text{ kJ mol}^{-1} \text{ rad}^{-2}$) and $\chi_0 = 0.61547$ radians (35.264 degrees) is the position of the minimum of the potential for the *L*-amino acid ($\lambda = 0$). When λ is varied from 0 to 1 the position of the minimum changes successively from $+35.264$, passes through zero to becomes -35.264 degrees (for $\lambda = 1$ corresponding to the *D*-isomer). The free energy is then

$$\Delta F(\lambda) = F(\lambda) - F(0) = \int_0^\lambda \left\langle \frac{\partial U_p}{\partial \lambda'} \right\rangle d\lambda' \quad (7)$$

and the final difference becomes

$$\Delta F = F(1) - F(0) = \int_0^1 \left\langle \frac{\partial U_p}{\partial \lambda} \right\rangle d\lambda \quad (8)$$

with $F(1)$ referring to the *D*-amino acid and $F(0)$ to the *L* one. The integrand is

$$\left\langle \frac{\partial U_p}{\partial \lambda} \right\rangle = 2C_\chi \chi_0 (\langle \chi \rangle - (1 - 2\lambda)\chi_0) \quad (9)$$

THE MODEL

We will here describe the time development along the one-dimensional reaction coordinate (λ , the improper dihedral in this case) with a Langevin equation:

$$I \frac{d^2\chi}{dt^2} = -I\beta \frac{d\chi}{dt} - C_\chi(\chi - \chi_0(1 - 2\lambda)) + A(t) + F_{\text{mean}}(\chi) \quad (10)$$

Here I is the moment of inertia, β the damping, $A(t)$ the random force and $F_{\text{mean}}(\chi)$ the mean field force originating from the other degrees of freedom. $A(t)$ has zero mean, is uncorrelated to itself at different times and has its size give by a fluctuation-dissipation theorem:

$$\langle A(t) \rangle = 0 \quad (11)$$

$$\langle A(t)A(t') \rangle = 2I\beta\Theta\delta(t - t'), \quad (12)$$

with Θ being the Boltzmann constant times absolute temperature. If one did know the mean force, there would be no need for an integration procedure, since the free energy difference is

$$F(\chi) - F(\chi_0) = - \int_{\chi_0}^{\chi} F_{\text{mean}}(\chi') d\chi' + U_p(\chi) - U_p(\chi_0). \quad (13)$$

It is, however not easy to estimate this mean force, and we may see the integration procedure as one way to do this. We will show in the next sections that for a couple of simple forms for the mean force we may solve the Langevin equation and thus calculate the free energy that would be obtained from a free energy perturbation integration. This free energy will contain systematic and statistical errors compared to the correct one obtained from equation (13).

One should observe that this means that we make two assumptions. The most serious one is that we describe the systematic influence from the surroundings by a mean force. This is true in equilibrium, which means that the integration has to be performed so slow that the surroundings (water and protein) is given enough time to relax as a response to the change in the reaction coordinate. This may often be difficult to achieve and monitor. In cases where the mean force varies slowly along the reaction coordinate and the changes in the surroundings are local and small, the errors calculated here will be the dominating ones. In other cases, the treatment here gives a lower limit to the systematic and statistical errors.

Further, to assume that the random influence can be described by Brownian motion theory as done above is of course an approximation, but considering the other uncertainties it is not necessary to introduce a more sophisticated stochastic description. In fact we will show in a practical example that this approximation is quite good.

To simplify the treatment a bit we will specialize to a λ that is varied linearly in time. This means that

$$\lambda = t/T \quad (14)$$

with T being the total integration time.

SYSTEMATIC ERRORS

We will first consider the situation with mean force $F_{\text{mean}}(\chi)$ equal zero. Then the free energy is zero. We can solve the Langevin equation above exactly and from that calculate hysteresis and statistical errors. We write the solution as:

$$\chi(t) = \chi_{\text{mean}}(t) + \xi(t) \quad (15)$$

Here $\xi(t)$ is the stochastic process that is obtained as a solution to the Langevin equation for Brownian motion in a harmonic oscillator with its minimum at zero. $\chi_{\text{mean}}(t)$, just extracts the contribution due to the shift of the minimum position of the harmonic potential with time. Since $\xi(t)$ has average zero (11), it will not contribute to the average free energy (but give rise to statistical fluctuations as shown in the next section). The mean solution is linear in time and is easily found to be

$$\chi_{\text{mean}}(t) = \chi_0 + \frac{2I\beta\chi_0}{C_x T} - \frac{2\chi_0 t}{T}. \quad (16)$$

The average free energy change in an integration then becomes

$$\langle \Delta F \rangle = \int_0^1 \left\langle \frac{\partial U_p}{\partial \lambda} \right\rangle d\lambda = \frac{4I\beta\chi_0^2}{T}. \quad (17)$$

The simplest way of reversing the direction of integration is to change the sign of χ_0 . The average free energy change will still be positive since it is quadratic in χ_0 . Thus, we have hysteresis. Even if the mean force from the surroundings is zero we will always get a positive free energy change irrespective of the integration direction. This is due to that the reaction coordinate is lagging behind the position of the potential minimum. This time-lag can be reduced by increasing the integration time. The hysteresis is inversely proportional to the integration time and proportional to the damping β . It is the same in both directions so that an average over forward and backward runs gives no systematic errors.

In a realistic situation, there is always a mean force (otherwise the free energy difference is zero). The simplest case to consider is that of a harmonic mean potential

$$U_{\text{mean}}(\chi) = \frac{1}{2} D(\chi - \chi_1)^2 \quad (18)$$

D may be positive or negative. This gives rise to a free energy difference between the L and D amino acids as calculated from eq. (13)

$$F(-\chi_0) - F(\chi_0) = 2D\chi_1\chi_0. \quad (19)$$

This is nonzero unless the minimum or maximum of the harmonic mean potential occurs at the improper dihedral angle value zero ($\chi_1 = 0$). The Langevin equation can be treated in exactly the same way as with zero mean force. There will just be different constants. This gives after some straightforward algebra the average difference

$$\langle \Delta F \rangle = 2D\chi_1\chi_0 \left(1 - \frac{D}{C_x + D} \right) + \frac{4I\beta\chi_0^2}{T} \left(\frac{C_x}{C_x + D} \right)^2. \quad (20)$$

As long as the force constant D of the mean potential is much smaller than the force constant of the perturbing potential, which is the typical situation, the result is the same as without a mean force. If the force constant D is larger one gets increased hysteresis if it is negative (barrier) and decreased hysteresis if it is positive (well). The hysteresis will now no longer be symmetric around the correct average since

there is a term that is linear in χ_0 . That means that, a correct average no longer is obtained by averaging forward and backward runs.

For a more general shape of the mean force as a function of the reaction coordinate, we cannot easily estimate the systematic errors. It seems, however, reasonable that one result shown above could be generalized to non harmonic potentials. We suggest that the hysteresis is still symmetric and given by equation (17) as long, as the derivative of the mean force is much smaller than the force constant of the perturbing harmonic potential (for all λ -values).

NON-SYSTEMATIC ERRORS

The non-systematic errors are related to big fluctuations in the derivative of the perturbation potential with respect to λ . With λ varied linearly in time, we have

$$\Delta F = \frac{1}{T} \int_0^T \frac{\partial U_p}{\partial \lambda} dt. \quad (21)$$

The integrand has typically fluctuations around 30–40 kJ/mol while the final value of the integral is about 5% of that. The possibility to estimate small free energy changes is thus strongly related to the statistical properties of these fluctuations. Most of them are in fact very short lived, which makes the effort feasible.

With the mean force zero we get using equations (15) and (16)

$$\Delta F = \frac{4I\beta\chi_0^2}{T} + \frac{2C_\chi\chi_0}{T} \int_0^T \xi(t) dt. \quad (22)$$

The second term is an integral over the stochastic process $\xi(t)$ which has the average zero. Thus the average of the integral will be zero. Since the first term is constant, we will get the fluctuations in ΔF as

$$\langle \Delta F^2 \rangle - \langle \Delta F \rangle^2 = \left(\frac{2C_\chi\chi_0}{T} \right)^2 \left\langle \left(\int_0^T \xi(t) dt \right)^2 \right\rangle \quad (23)$$

which also may be written

$$\langle \Delta F^2 \rangle - \langle \Delta F \rangle^2 = \left(\frac{2C_\chi\chi_0}{T} \right)^2 \int_0^T \int_0^T dt du \langle \xi(t) \xi(u) \rangle = \quad (24)$$

$$= 2 \left(\frac{2C_\chi\chi_0}{T} \right)^2 \int_0^T (T-t) \langle \xi(s) \xi(s+t) \rangle_s dt. \quad (25)$$

This formula contains the autocorrelation function of Brownian motion in a harmonic potential. This autocorrelation function decays to zero in much shorter times than T for all sensible systems and times T . We may therefore put $T-t = T$ inside the integral and do get:

$$\sqrt{\langle \Delta F^2 \rangle - \langle \Delta F \rangle^2} = 2C_\chi\chi_0 \sqrt{\frac{2}{T} \int_0^T \langle \xi(s) \xi(s+t) \rangle_s dt}. \quad (26)$$

For practical purposes, the upper limit in the integral can be put equal to infinity and the autocorrelation function could be characterized by a correlation time τ_c . This gives:

$$\sqrt{\langle \Delta F^2 \rangle} - \langle \Delta F \rangle^2 = 2C_x \chi_0 \sqrt{\frac{2\tau_c}{T} \langle \xi^2 \rangle}. \quad (27)$$

Brownian motion in a harmonic potential is characterized by two parameters, the frequency of the oscillator, ω and the damping, β , reflecting influence of the surroundings. The time autocorrelation function is (11)

$$\langle \xi(s) \xi(s+t) \rangle_s = \langle \xi^2 \rangle \exp(-\beta t/2) \left(\cos \beta_1 t/2 + \frac{\beta}{\beta_1} \sin \beta_1 t/2 \right) \quad (28)$$

with

$$\beta_1 = \sqrt{4\omega^2 - \beta^2}. \quad (29)$$

The correlation time can be calculated analytically in terms of the parameters β and ω

$$\tau_c = \beta/\omega^2. \quad (30)$$

Further, we may use that the average energy in the improper dihedral degree of freedom is $\Theta/2$ (Boltzmanns constant times absolute temperature) in equilibrium giving

$$\langle \xi^2 \rangle = \Theta/C_x. \quad (31)$$

We may also eliminate C_x by the relation

$$C_x = I\omega^2. \quad (32)$$

This gives the final formula for the rms fluctuation in the free energy

$$\sqrt{\langle \Delta F^2 \rangle} - \langle \Delta F \rangle^2 = 2\chi_0 \sqrt{\frac{2\beta I \Theta}{T}}. \quad (33)$$

Note that both the statistical errors and the systematic ones are functions of β/T . The statistical ones are proportional to the square root of this quantity while the systematic ones are proportional to the quantity itself (equation 17). This means that these two different types of errors scale in a different way with the length of the simulation. For short runs the hysteresis may dominate over the statistical errors, while the situation is opposite for long runs.

AN ALTERNATIVE REACTION COORDINATE

The most straightforward way of doing the conversion is to use the improper dihedral as reaction coordinate as done above. It has, however, the disadvantage that an energy barrier is introduced since the three valence angles surrounding the center atom for the improper dihedral are forced out of there minima during the integration. The deviation is maximum for the planar configuration at $\lambda = 0.5$ corresponding to about 10 degrees and an energy of about 20 kJ/mol. In practice, this causes the improper dihedral to lag behind its minimum value during the first half of the integration and then increase beyond the minimum during the second half of the integration. This nonlinear variation of the improper dihedral and the

free energy with λ is in principle not a problem. It affects, however, the efficiency of the sampling and may increase systematic errors. If this barrier is approximated with a harmonic potential, this will have a force constant of about $-100 \text{ kJ/mol}^{-1} \text{ rad}^{-2}$. This is not negligible compared to the force constant of the perturbing potential (about $340 \text{ kJ/mol}^{-1} \text{ rad}^{-2}$). According to equation (20) this results in a doubling of the hysteresis. Since the position of the maximum is at zero, this should not have any other systematic effect as long as other contributions to the mean force are small.

It might be a good idea to let the minimum of the valence angles follow the improper dihedral minimum. With the same minimum value of for the potentials of all three valence angles one obtains with the help of a little geometry the following formula for the valence angle minimum

$$\gamma_0 = 2 \arctan(\sqrt{3} \cos(\chi_0(1 - 2\lambda))). \quad (34)$$

If this relation is fulfilled both the valence angles and the improper dihedral can take the values of minimum(zero) energy at the same time. This means that the integration can be performed without passing a free energy maximum at $\lambda = 0.5$. The integrand becomes

$$\left\langle \frac{\partial U_p}{\partial \lambda} \right\rangle = 2C_\chi \chi_0 (\langle \chi \rangle - (1 - 2\lambda)\chi_0) + \sum_{i=1}^3 C_\gamma (\langle \gamma_i \rangle - \gamma_0(\lambda)) \frac{d\gamma_0}{d\lambda}. \quad (35)$$

This could be generalized to situations where the three valence angles have potentials with minima at different values. We have performed stochastic dynamics integration conversions of a single phenylalanine molecule using this approach and verified that a reduction of the hysteresis is obtained.

AN EXAMPLE

We have studied the binding of L- and D-phenylalanine in the active site of thermolysin and calculated free energy differences from this. The detailed results and methods will be published elsewhere (12). Here we just give the average free energy differences obtained together with standard deviations and the hysteresis in table 1. There are results with two different force fields, the standard GROMOS one and one with 30% higher fractional charges. The figures in table 1 are the result of the averaging over 20 psec. integrations (4 in each direction). The result using the standard GROMOS force field, -2 kJ/mol , agrees within the standard deviation divided by the square root of the number of runs with the experimental value, -2.8 kJ/mol (13). These simulations were done with a perturbation potential involving only the improper dihedral. Therefore the analysis here will be based upon

Table 1 Free energies differences of binding for L- and D-phenylalanine into the active site of thermolysin in kJ/mol obtained from 8 conversion runs each using two different force fields.

average L to D	+0.9	-4.0
average D to L	+4.9	+8.1
total average (as L to D)	-2.0	-6.0
hysteresis	+2.9	+2.0
fluctuations	± 3.5	± 1.6

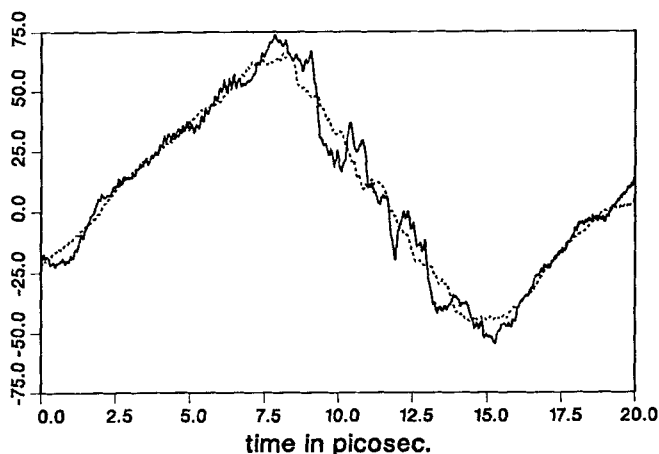


Figure 1 The smoothed part part of $\partial U_p/\partial\lambda$ in a typical conversion in kJ/mol The full curve is a running average over 1 psec. and the dashed curve over 2.5 psec.

equation (9) not (35). The aim is to show that the results for fluctuations and hysteresis are consistent with the theory (equation (33) and (20)).

To do this we will first analyze the statistical properties of $\partial U_p/\partial\lambda = 2C_\chi\chi_0(\chi - \chi_0(1 - 2t/T))$ as this quantity is obtained from the simulation. We therefore try to separate it into a slowly varying part $F_{\text{mean}}(\lambda(t))$ and a strongly fluctuating one, $\eta(t)$ which has mean zero and hopefully the properties of Brownian motion in a harmonic oscillator

$$\frac{\partial U_p}{\partial\lambda} = F_{\text{mean}}(\lambda(t)) + \eta(t). \quad (36)$$

To obtain the mean force, a running average over time intervals Δt

$$F_{\text{mean}}(t) = \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} \frac{\partial U_p}{\partial\lambda} dt \quad (37)$$

is calculated. $\eta(t)$ is then defined as the remaining part from equation (36). This was performed on one of the 20 psec. conversion runs. The time interval Δt is obviously a parameter here, but the result is not very sensitive to the precise value chosen in an interval between 0.5 and 2.5 psec. In the figure 1 the $F_{\text{mean}}(t)$ obtained using two time intervals, 1 and 2.5 psec., are shown. In the further analysis a 1 psec. interval corresponding to 0.05 in λ has been used. The remaining part $\eta(t)$ is shown in figure 2. Obviously some information about the shape of $F_{\text{mean}}(\lambda(t))$ is lost if the interval is chosen too big. If its taken to small the function becomes noisy.

The idea is now that $F_{\text{mean}}(\lambda(t))$ is a deterministic function that will be the same in different runs while $\eta(t)$ is stochastic process with average value zero. Alternatively one may consider $\partial U_p/\partial\lambda$ as a stochastic process that has an average value that is a function of λ , $F_{\text{mean}}(\lambda)$. We have then just estimated and subtracted this average value to obtain a process $\eta(t)$ with average zero. This can be done in a unique way if the autocorrelation time of the stochastic process is much shorter

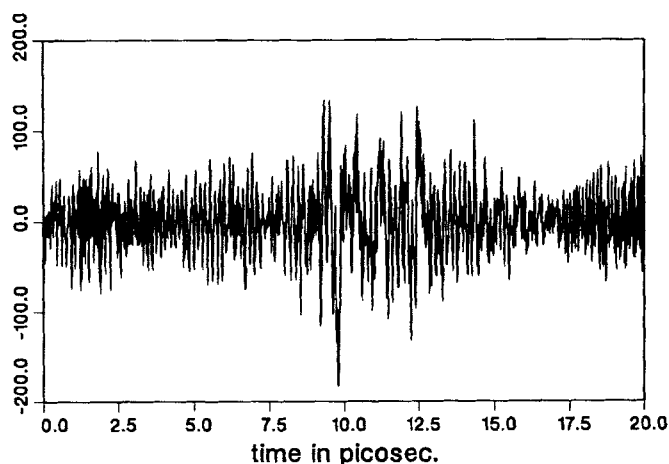


Figure 2 The remaining part of $\partial U_p/\partial \lambda$ after subtracting running 1 psec. averages in kJ/mol

than the times over which there is any appreciable variation of $F_{\text{mean}}(\lambda(t))$. If the sampling time T , is long enough this necessary separation of time scales is obtained.

The rms fluctuations calculated from figure 2 are 36 kJ/mol. This agrees perfectly with the theoretical value

$$\sqrt{\langle \eta^2 \rangle} = 2C_x \chi_0 \sqrt{\langle \chi^2 \rangle - \langle \chi \rangle^2} = 2\chi_0 \sqrt{C_x \Theta} \quad (38)$$

which also is 36 kJ/mol at 300 K with $\chi_0 = 0.615$ rad and $C_x = 340$ kJ mol⁻¹ rad⁻². The normalized time autocorrelation function of $\eta(t)$ is shown in figure 3. It is clear that much more data are necessary to get a proper estimate of it for times larger than 0.5 psec. If it is integrated over increasing times, the integral reaches a value slightly less than 0.03 after 0.1 and 0.2 psec. After that the integral drops and fluctuates between positive and negative values. We therefore believe that the time autocorrelation function has dropped to zero after 0.2 psec. and what is seen for longer times is noise due to poor statistics. A nonlinear least square fit of an exponentially damped sinusoidal oscillation like equation (28), gives the values $\beta = 47$ psec⁻¹ and $\beta_1 = 81$ psec⁻¹ using all data points up to 2 psec. This curve is also shown in figure 3. The result did not change when the part of the function used for the fit was reduced down to 0.3 psec. These values for β and β_1 give (equation 29) $\omega = 33$ psec⁻¹.

The value obtained for ω is comparable to what is obtained from the force constant, masses and distances involved (equation (32)). We assume then that it is the lighter carboxy- and amino groups that move and not the phenyl ring. The moment of inertia depends then also on the orientation of the oxygen (and hydrogens). For the motion of the lighter NH_2 -group ω becomes about 30 psec⁻¹, for the COO -group about 15 psec⁻¹. The motion is slightly underdamped but very close to critical damping since $\beta/2\omega \approx 1$. The correlation time (equation (30)) is very short, somewhere in the range $\tau_c = 0.02$ – 0.06 psec.

Using 47 psec⁻¹ for β and $6 \cdot 10^{-46}$ k gm² for the moment of inertia (the NH_2 group) we obtain for a 20 psec. run the statistical error ± 2.5 kJ/mol from equation (33) and $+2.6$ kJ/mol from the second term in equation (20). This agrees with the data from the simulations in table 1.

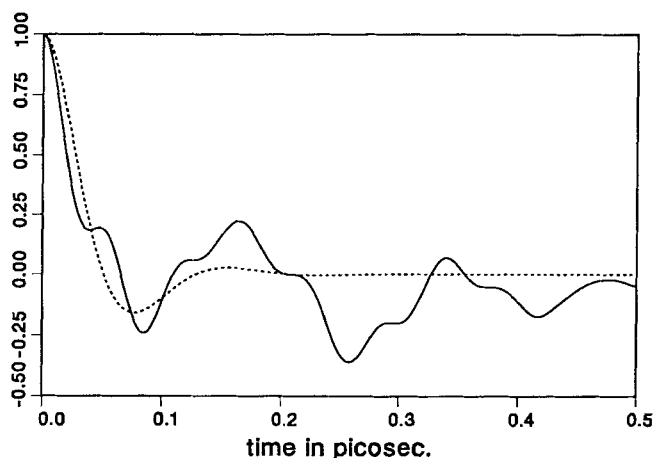


Figure 3 Time autocorrelation function of the function in figure 2 together with the best nonlinear least square fit to an exponentially damped sinusoidal oscillation according to equation (28).

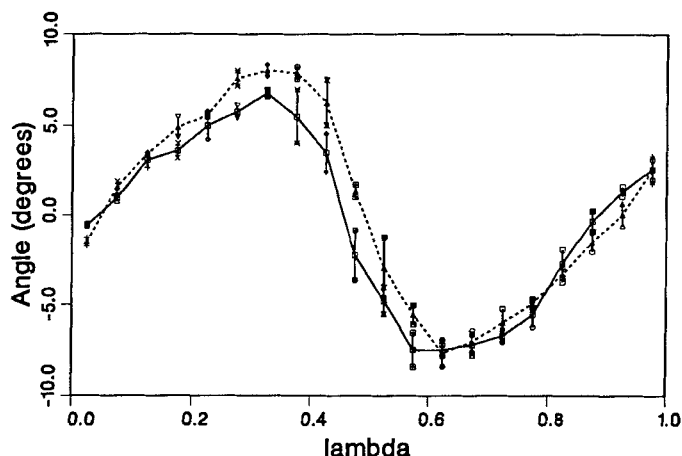


Figure 4 Average deviation of the improper dihedral from the value at the minimum of the λ -dependent potential as a function of λ . The full curve is an average over three L to D conversions and the dashed one an average over three D to L conversions. The error bars indicate the standard deviation between the three runs divided by $\sqrt{3}$.

In the free energy integration, χ will approximately follow the changing position of the minimum. However, the mean potential representing the influence from the other degrees of freedom makes different values of χ more or less favourable. Therefore, it will as seen in figure 4 not follow the position of the minimum perfectly. The main reason for this is that there is a barrier due to the valence angles that makes it lag behind during the first half of the simulation and then run ahead in the second half. If there is no hysteresis, the average value of the improper

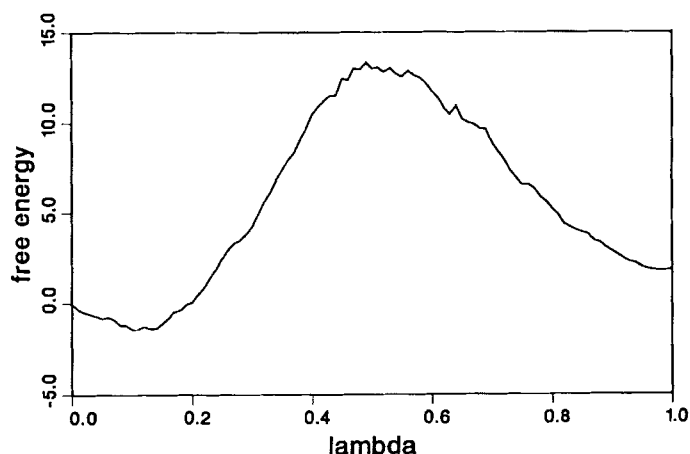


Figure 5 Free energy as a function of λ for an L to D conversion in kJ/mol.

dihedral as a function of λ should, however, be the same irrespective of the direction of the integration. From figure 4 it is seen that there is a small but significant difference between the runs in the L to D and the D to L direction. This is what causes the hysteresis seen in table 1. In figure 5 the free energy as a function of λ is shown as calculated from a simulation. There is a barrier of about 15 kJ/mol.

We also tried the windowing technique by making separate simulations at λ values 0.1 units apart. Each of these simulations were 1 psec. and were preceded by 0.5 psec. equilibration. The change of λ was done in a 0.5 psec. continuous integration. The effect of that was as could have been anticipated a reduced hysteresis, since the system now gets more time to equilibrate at each λ . Therefore, the time-lag is less. On the other hand, only half the time was used for sampling. This means larger statistical errors. Further, a systematic error due to the finite sampling of the configurational space will occur (14). Hence, it is only justified to use windowing techniques when one has reason to believe that non-symmetric systematic errors due to time-lag are more important than the statistical ones and the symmetric hysteresis. This situation occurs mainly when the simulations are short. For long simulations the statistical errors will dominate also in a continuous integration.

CONCLUSIONS

In a free energy perturbation where the integration is performed along a reaction coordinate and driven by a moving harmonic perturbation potential, the systematic and statistical errors can be estimated under the following conditions: The influence from the other degrees of freedom is described by a mean force that is weak compared to the perturbation force and a random force that is modeled by Brownian motion theory. There is then a positive hysteresis due to that the reaction coordinate will lag behind the minimum of the perturbation potential. The hysteresis is symmetric and inversely proportional to the simulation time. The non-systematic error is inversely proportional to the square root of the simulation time. This

means that the hysteresis will dominate for short runs while the statistical errors will dominate for long runs. The hysteresis is less of a problem if most of it is symmetric. Then an averaging over runs in both directions will considerably reduce the error. If the mean force varies slowly compared to the perturbation force this is the case.

For a conversion between the two stereoisomers of phenylalanine bound to a protein (in the active site of thermolysin), both systematic and non-systematic errors are about 2.5 kJ/mol in a 20 psec. integration. This is comparable to the difference in free energy of binding. The errors are proportional to respectively proportional to the square root of the moment of inertia of the moving group (the NH_2 -group in this case). This means that the corresponding error will be considerably larger in applications where entire loops in a protein are moved by restraining dihedrals with harmonic potentials.

The estimate here gives only a lower limit, on the errors. If the integration along the reaction coordinate is accompanied (by slow relaxation processes of the environment, the mean force description will not be valid until these have occurred. Then it may be necessary to use windowing techniques with very long equilibrations between the samplings. Further, there are of course systematic errors due to that the force fields used are inaccurate or inappropriate.

References

- [1] R.W.J. Zwanzig, "High-temperature equation of state by perturbation method. I nonpolar gases", *J. Chem. Phys.*, **22**, 1420 (1954).
- [2] J.P.M. Postma, H.J.C. Berendsen and J.R. Haak, "Thermodynamics of cavity formation in water", *Faraday Symp. Chem. Soc.*, **17**, 55 (1982).
- [3] W.F. van Gunsteren and P.K. Weiner, Computer simulation in biomolecular systems: Theoretical and experimental applications, Escom Science, the Netherlands, (1989).
- [4] T.P. Straatsma, H.J.C. Berendsen and J.P.M. Postma, "Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water", *J. Chem. Phys.*, **85**, 6720 (1986).
- [5] D.A. Pearlman and P.A. Kollman, "A new method for carrying out free energy perturbation calculations: Dynamically modified windows", *J. Chem. Phys.*, **90**, 2460 (1989).
- [6] D.A. Pearlman and P.A. Kollman, "The lag between Hamiltonians and the system configuration in free energy perturbation calculations", *J. Chem. Phys.*, **91**, 7831 (1989).
- [7] W.F. van Gunsteren and A.E. Mark, "On the interpretation of biochemical data by molecular dynamics computer simulation", *Eur. J. Biochem.*, **204**, 947 (1992).
- [8] M.J. Mitchell and J.A. McCammon, "Free energy difference calculations by thermodynamic integration: Difficulties in obtaining a precise value", *J. Comp. Chem.*, **12**, 271 (1991).
- [9] T.P. Straatsma, H.J.C. Berendsen and A.J. Stam, "Estimation of statistical errors in molecular dynamics simulations", *Mol. Phys.*, **57**, 89 (1986).
- [10] R.H. Wood, "Estimation of errors in free energy calculations due to the lag between the Hamiltonian and the system configuration", *J. Phys. Chem.*, **95**, 4838 (1991).
- [11] S. Chandrasekhar, "Stochastic problems in physics and astronomy", *Rev. Modern Phys.*, **15**, 1 (1943).
- [12] I. Ghosh and O. Edholm, "Molecular dynamics study of the binding of phenylalanine stereoisomers to thermolysin", to be published, 1992.
- [13] J. Feder, L.R. Brougham and B.A. Wildi, "Inhibition of thermolysin by dipeptides", *Biochemistry*, **13**, 1186 (1974).
- [14] R.H. Wood, W.C.F. Mühlbauer and P.T. Thompson, "Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: Sample-size hysteresis", *J. Phys. Chem.*, **95**, 6670 (1991).