FREE ENERGY DETERMINATION:

1. NORMAL MODE APPROXIMATION
2. UMBRELLA SAMPLING
3. THERMODYNAMIC INTEGRATION
4. PERTURBATION METHOD

$$A = -kT\left(\ln \frac{Se^{-\beta E}}{N!\,\Lambda^{3N}}\right) \qquad \beta = 1/kT$$

$$A^* - A = -kT \cdot \ln\left(Se^{-\beta E^*} / Se^{-\beta E}\right)$$

$$= -kT \cdot \ln\left(Se^{-\beta(E^*-E)} e^{-\beta E} / Se^{-\beta E}\right)$$

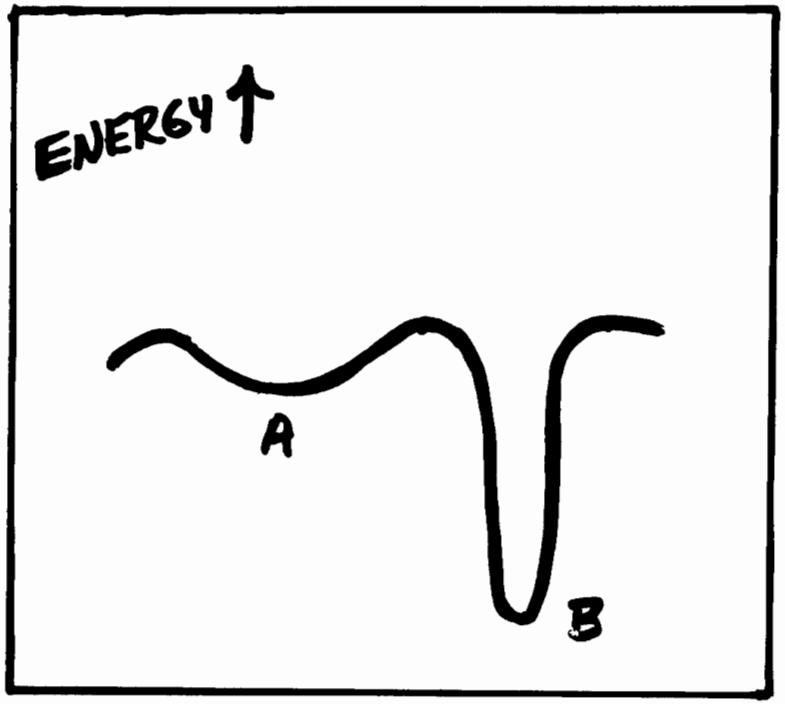
$$= -kT \ln \left\langle e^{-\beta(E^*-E)} \right\rangle_E$$

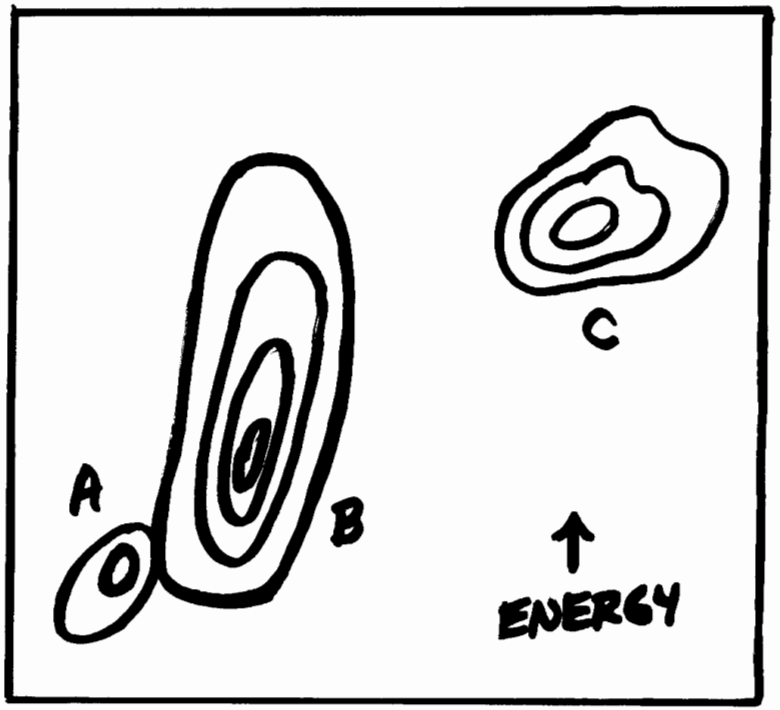
$$E_\lambda = \lambda E_M + (1-\lambda) E_L$$

FOR DISTANT STATES M AND L

— THERMODYNAMIC CYCLE METHOD

CONFORMATIONAL FREE ENERGY



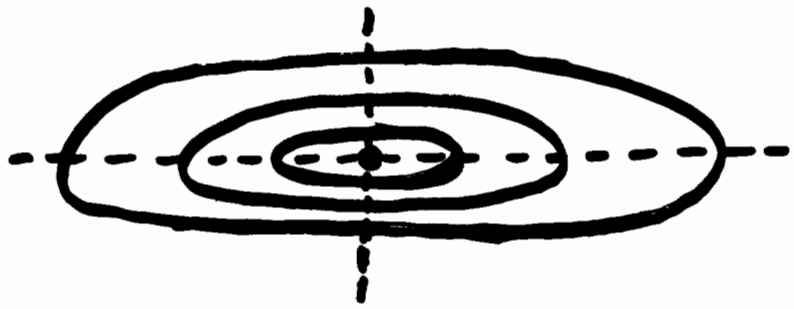


$$Z_{CONF} = \sum_{\tau_1, \tau_2} \exp\left[-E(\tau_1, \tau_2)/kT\right]$$

$$\Delta G^\circ = -kT \ln\left(Z_{CONF\,A}/Z_{CONF\,B}\right)$$

VIBRATIONAL ANALYSIS:

NORMAL MODES



HOW DO WE GET THEM?

DIAGONALIZE MASS WEIGHTED SECOND DERIVATIVES OF THE POTENTIAL ENERGY

USES?

1. COMPARE W/ EXPT. FREQUENCIES
2. ESTIMATE CONFORMATIONAL ENTROPY
3. EXAMINE LOW FREQUENCY MOTIONS OF THE STRUCTURE

LEVITT et al
JMB 181 423 (1985)

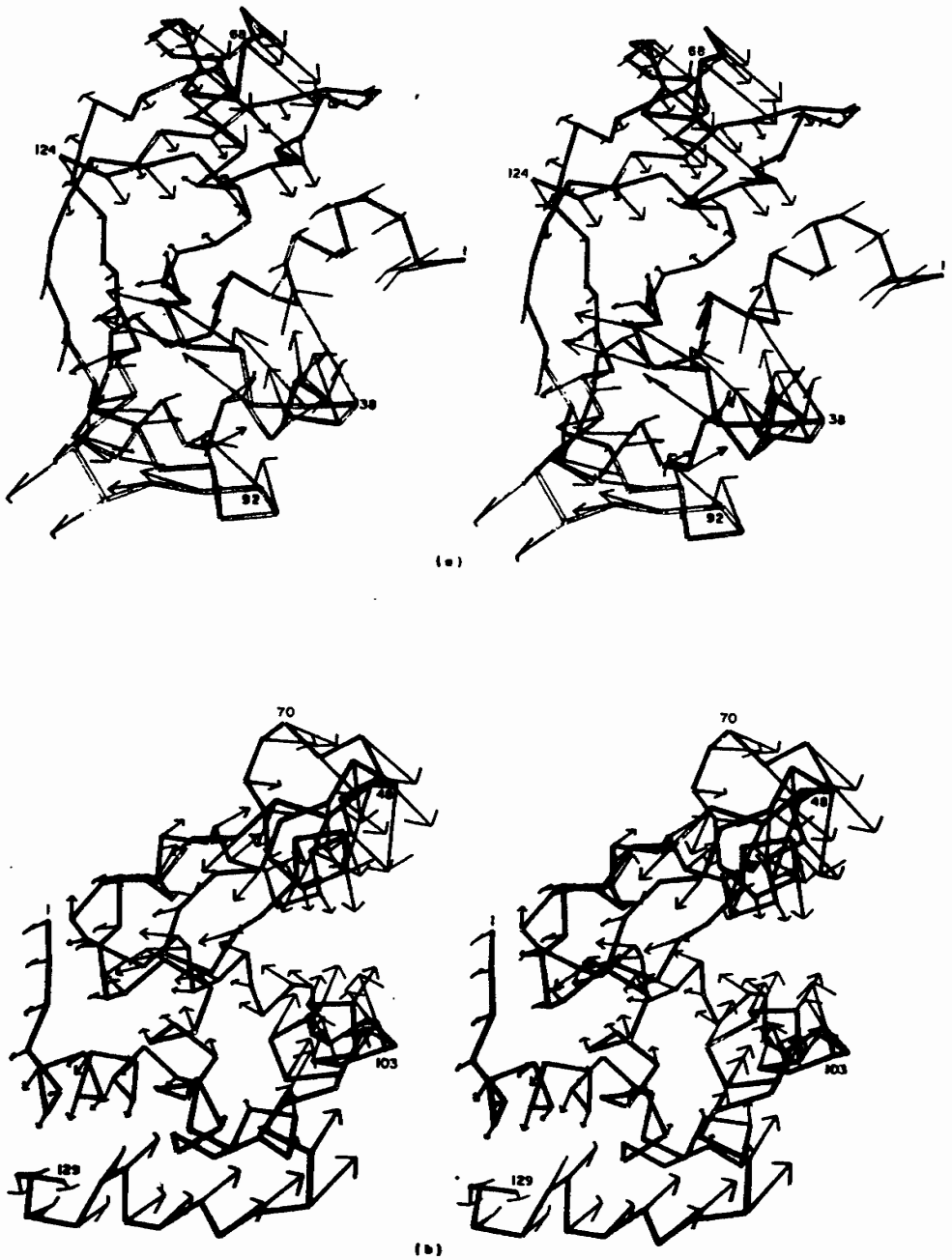


Figure 9. Showing the domain motion due to the lowest frequency modes in ribonuclease and lysozyme. The arrows on the α-carbon backbone are drawn as described in the legend to Fig. 8. (a) Ribonuclease, $\nu_1 = 2{\cdot}43\ \mathrm{cm}^{-1}$, period = 13·7 ps. (b) Lysozyme, $\nu_1 = 2{\cdot}98\ \mathrm{cm}^{-1}$, period = 11·2 ps.

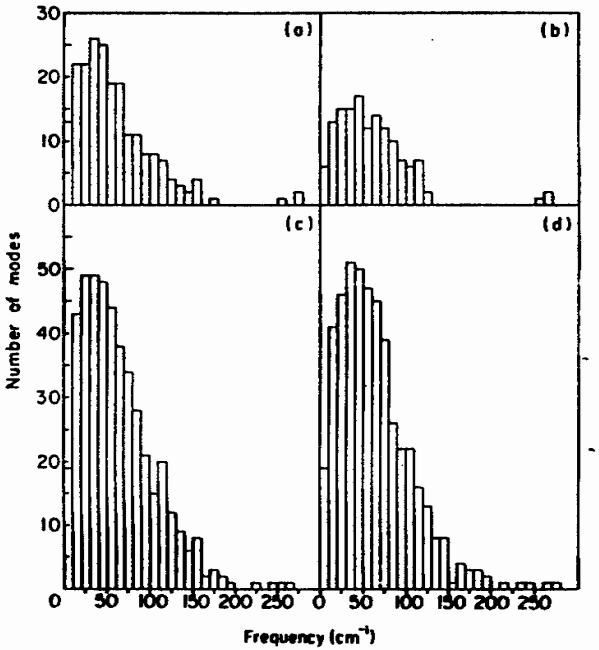


Figure 1. Showing the vibrational spectra calculated here with single-bond torsion angle variables. The number of modes with frequencies in a $10\ \mathrm{cm}^{-1}$ interval is plotted against frequency to give a density of states distribution. If each mode had the same intensity and a half line-width of $5\ \mathrm{cm}^{-1}$ the intensity envelope would be like the distribution shown. The spectra shown are for: (a) BPTI, (b) crambin, (c) ribonuclease and (d) lysozyme.

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:

$$\mathcal{V}'(\mathbf{r}^N) = \mathcal{V}(\mathbf{r}^N) + W(\mathbf{r}^N) \tag{9.27}$$

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

$$W(\mathbf{r}^N) = k_W(\mathbf{r}^N - \mathbf{r}_0^N)^2 \tag{9.28}$$

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 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:

$$\langle A \rangle = \frac{\langle A(\mathbf{r}^N) \exp[+W(\mathbf{r}^N)/k_B T] \rangle_W}{\langle \exp[+W(\mathbf{r}^N)/k_B T] \rangle_W} \tag{9.29}$$

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 $\mathcal{V}'(\mathbf{r}^N)$.

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.

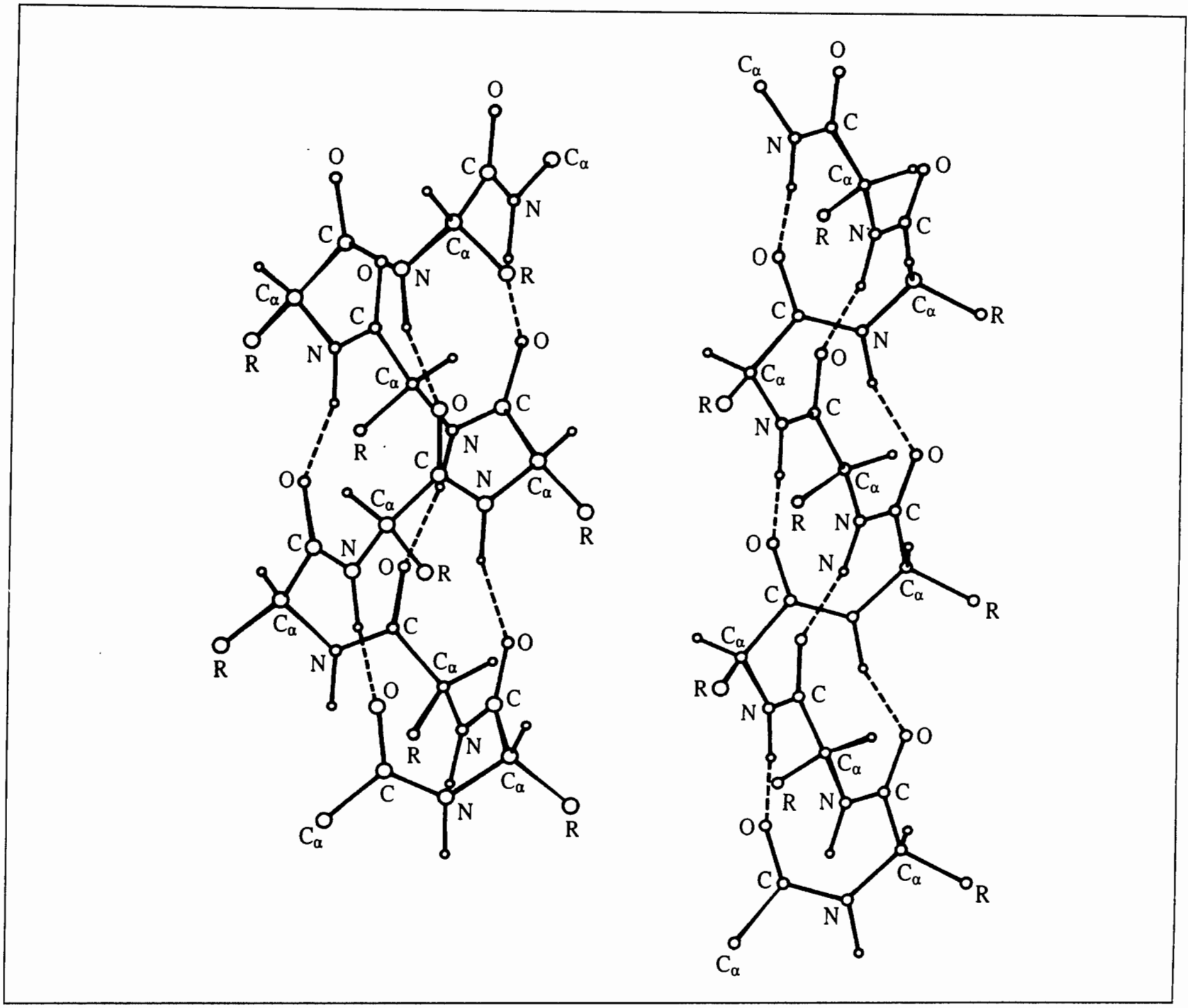
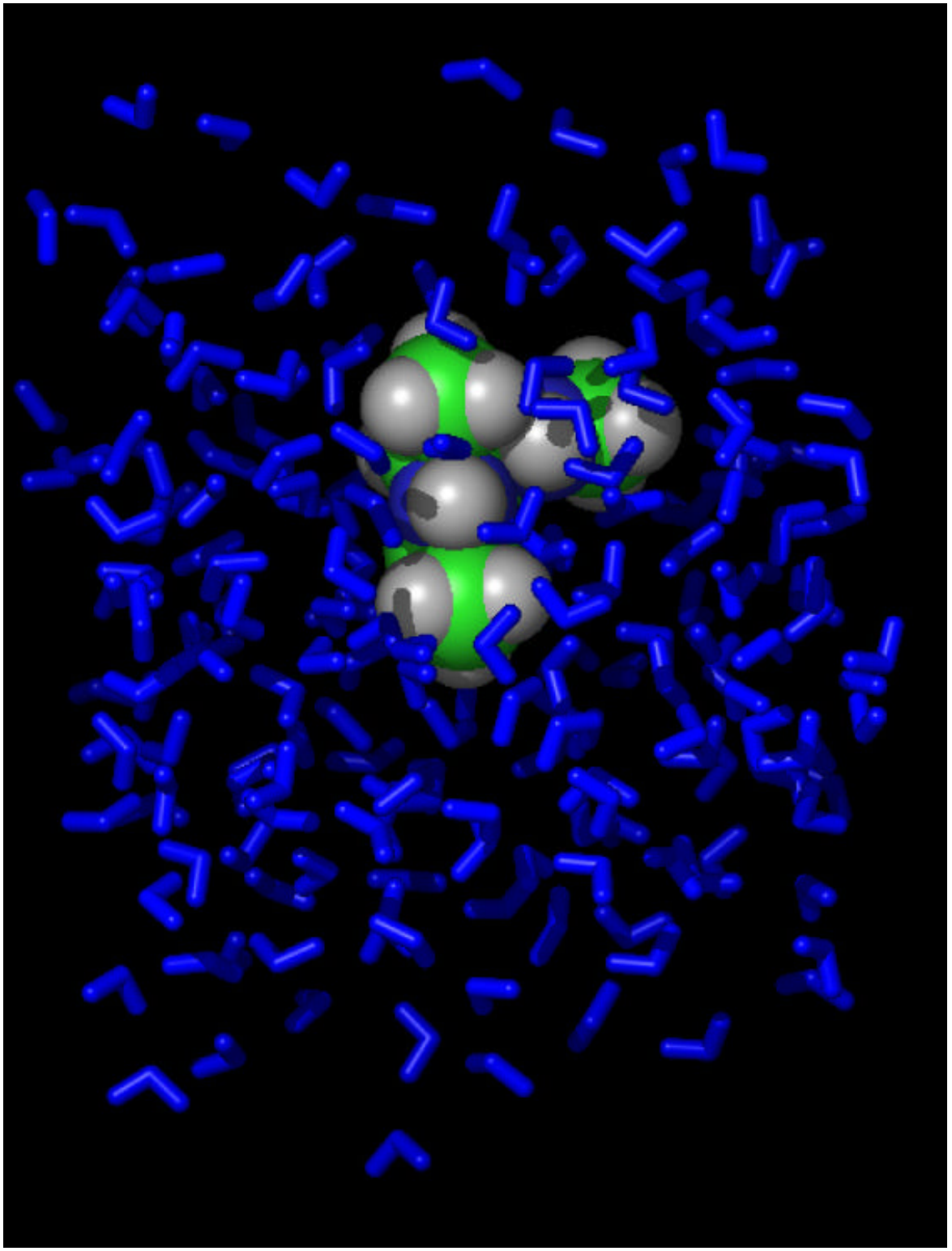


Fig. 11.15: The α-helix (left) and the 3_{10}-helix.



Periodic Box Simulation (alanine dipeptide and 206 water molecules)

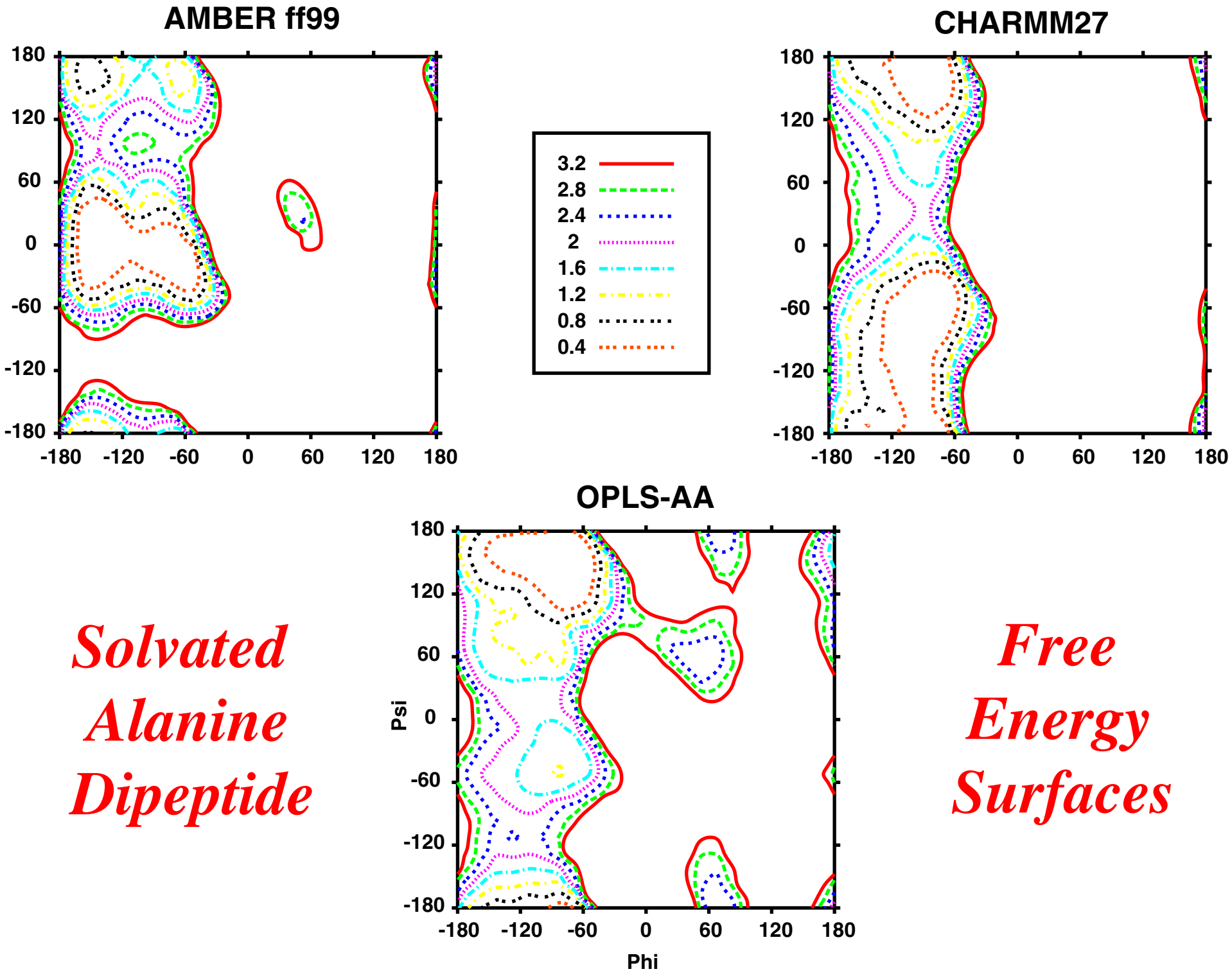
Stochastic Dynamics (576 trajectories of 200 picoseconds each)

Free Energies *via* Umbrella Sampling and 2D-WHAM:

$$r(f,y) = \frac{\sum_{i=1}^{Nw} n_i \, r_{wi}(f,y)}{\sum_{i=1}^{Nw} n_i \, e^{-w_i(f,y)-F_i)/k_bT}}$$

$$F_i = -k_b T \ln\left[\sum_f \sum_y e^{-w_i(f,y)/k_bT} \, r(f,y)\right]$$

$$\Delta G(f,y) = -k_b T \ln r(f,y)$$


AMBER ff99
CHARMM27
OPLS-AA
3.2
2.8
2.4
2
1.6
1.2
0.8
0.4
180
120
60
0
-60
-120
-180
-180 -120 -60 0 60 120 180
Psi
Phi
Solvated Alanine Dipeptide
Free Energy Surfaces

Conformational Populations

	Alpha	*Pass*	*Beta*	*Other*
Amber *ff94*	68	5	26	1
Amber *ff99*	77	10	13	1
CHARMM27	46	2	52	0
OPLS-AA	13	9	75	3
OPLS-AA/L	23	8	65	4
SCCDFTB (Amber)	27	16	48	9
SCCDFTB (CHARMM)	33	14	48	4
SCCDFTB (CEDAR)	27	12	61	0
AMOEBA (Polar Water)	29	16	54	1
AMOEBA (Fixed Water)	32	13	54	1

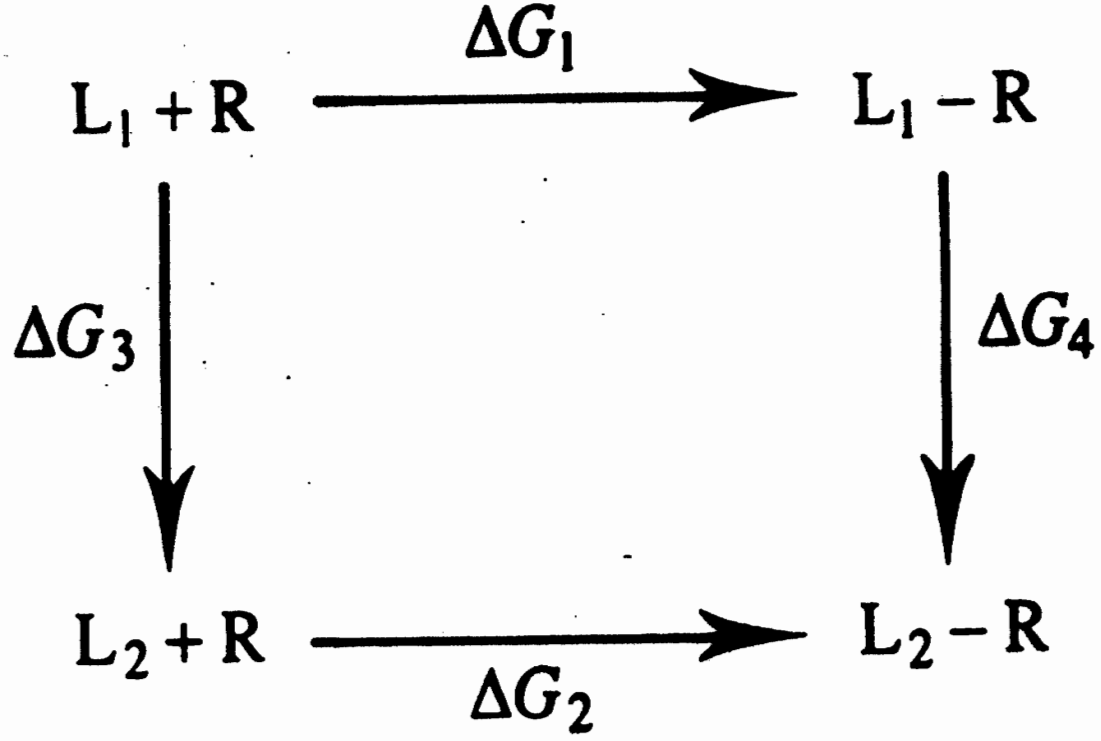


Fig. 9.5 Thermodynamic cycle for binding ligands

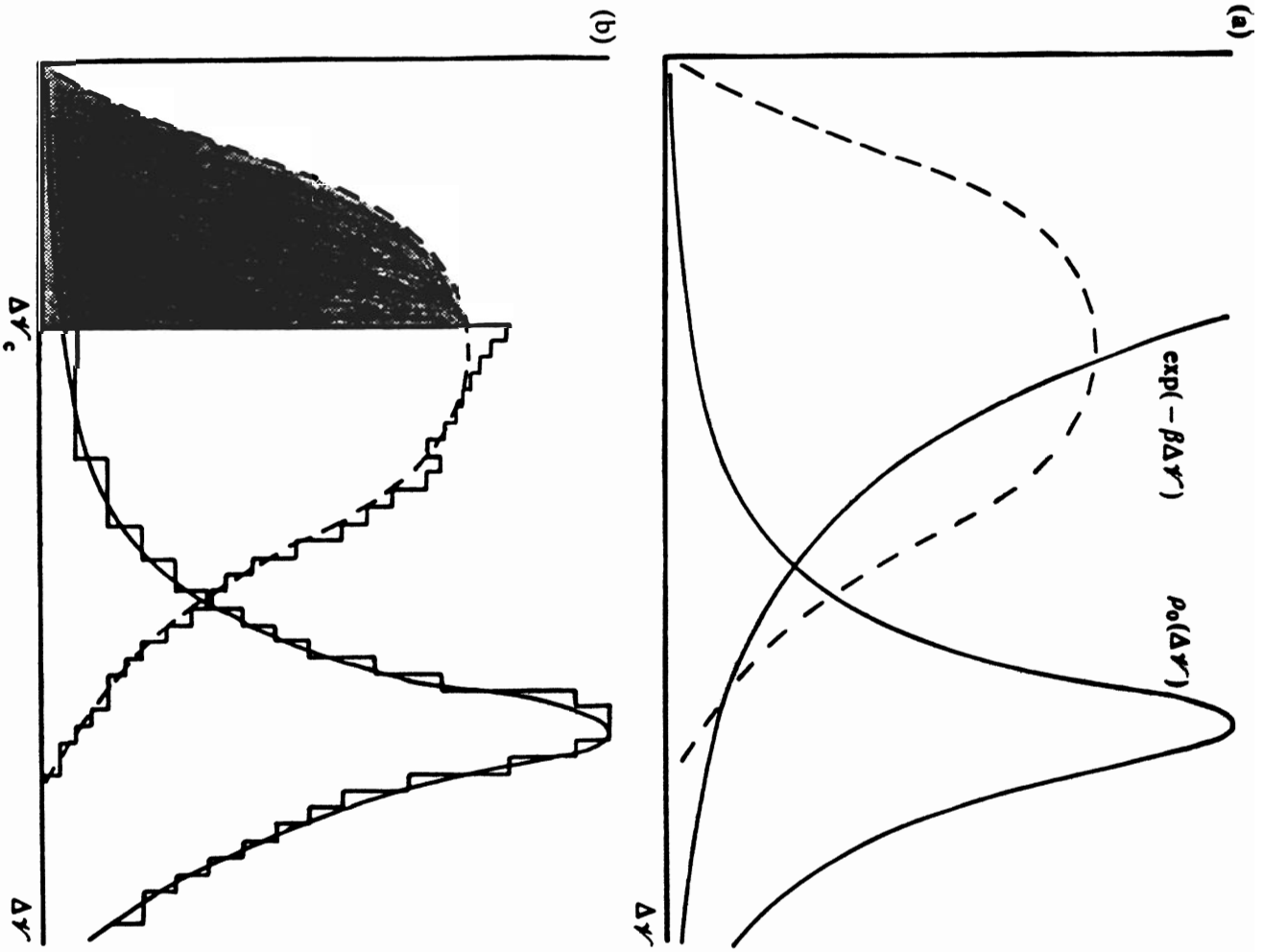


Fig. 7.1 The problem in estimating free energy differences. (a) The functions $\rho_0(\Delta\mathscr{V})$ and $\exp(-\beta\Delta\mathscr{V})$ are shown as solid lines. The product of these two functions, the integrand in eqn (7.3), is shown as a dashed line.
(b) Th y in which these functions are estimated in a finite-length simul t.

Let us begin with a less taxing problem than that of estimating eqn (7.1), specifically the calculation of a free energy difference. Consider two fluids characterized by potentials $\mathscr{V}(\mathbf{r})$ and $\mathscr{V}_0(\mathbf{r})$. If the free energy of the reference fluid, A_0, is known, then the free energy of the fluid of interest, A, can be determined from

$$A - A_0 = -k_B T \ln (Q/Q_0) = -k_B T \ln (\langle \exp -\beta\Delta\mathscr{V} \rangle_0) \qquad (7.2)$$

where $\Delta\mathscr{V}(\mathbf{r}) = \mathscr{V}(\mathbf{r}) - \mathscr{V}_0(\mathbf{r})$ and the ensemble average $\langle \ldots \rangle_0$ is taken in the reference system $\mathscr{V}_0$. Unless the two fluids are very similar, and $\beta\Delta\mathscr{V}$ is small for all the important configurations in this ensemble, the average in eqn (7.2) is difficult to calculate accurately. The reason for this becomes clear if we rewrite the configurational density function $\rho_0(\mathbf{r})$ as a function, $\rho_0(\Delta\mathscr{V})$, of the energy difference. Then

$$Q/Q_0 = \int_{-\infty}^{\infty} \mathrm{d}(\Delta\mathscr{V}) \exp(-\beta\Delta\mathscr{V}) \rho_0(\Delta\mathscr{V}) . \qquad (7.3)$$

1. Thermodynamic Integration

$$\Delta A = \int_0^1 \langle \Delta E \rangle_\lambda \, d\lambda = \sum_i \langle \Delta E (\lambda_i \rightarrow \lambda_{i+1}) \rangle_{\lambda_i}$$

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} \, d\lambda = \int_0^1 \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

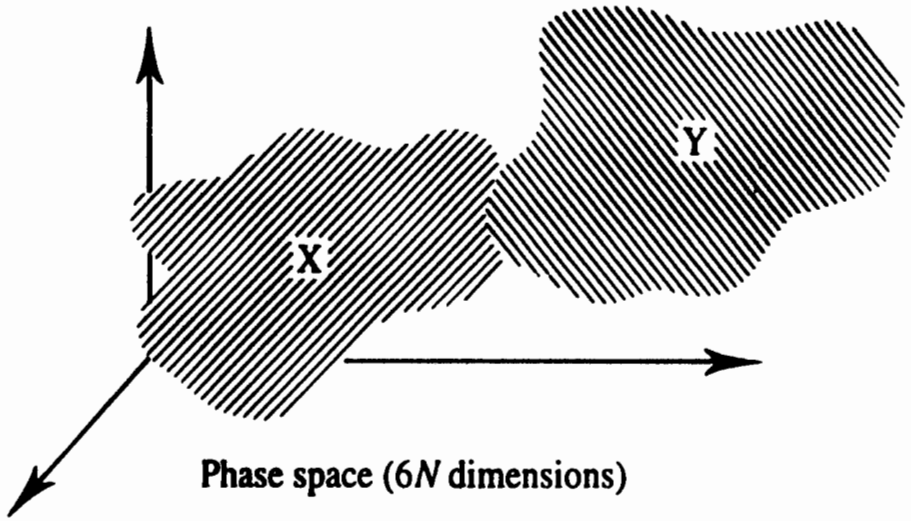
2. Perturbation Method

$$\Delta A = - kT \sum_i \ln \langle e^{- \beta \Delta E [i - (i + 1)]} \rangle_i$$

$$\Delta A = - kT \ln \langle e^{- \beta \Delta E} \rangle_0$$

$$A_1 - A_0 = - kT \ln \left(\frac{Z_1}{Z_0} \right)$$

$$A(\lambda) = -kT \ln Z(\lambda)$$



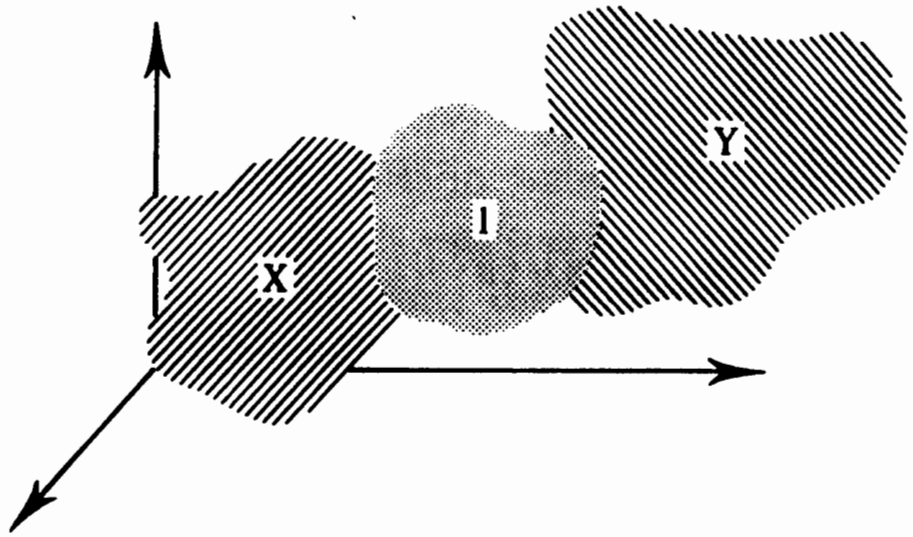
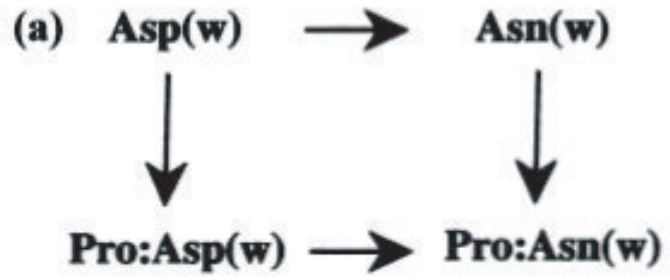
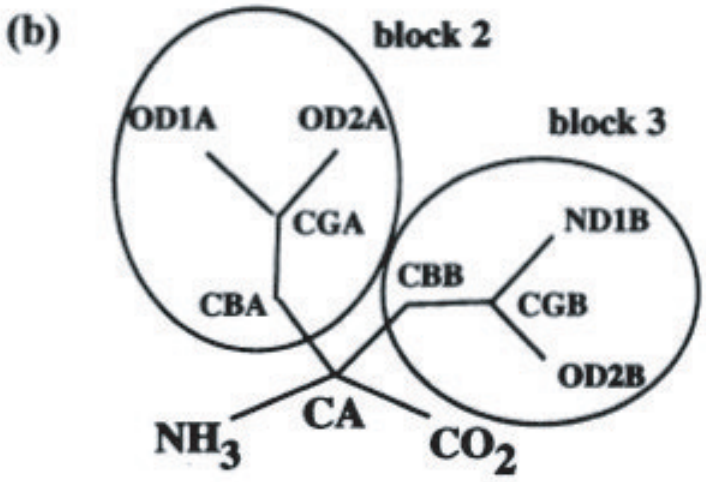


Fig. 9.1 An intermediate state (labelled 1) can improve the degree of overlap in phase space and lead to improved sampling.





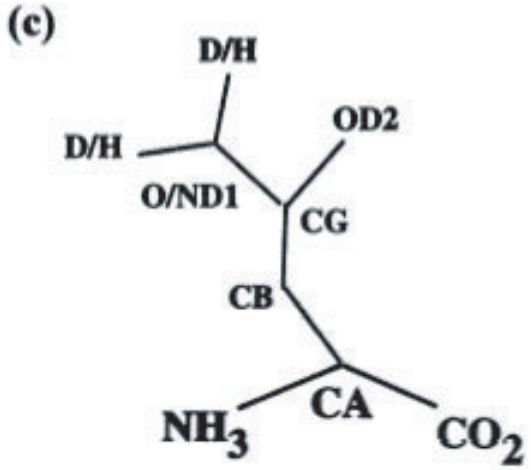


Figure 3 Mutation of a ligand Asp into Asn in solution and bound to a protein. (a) Thermodynamic cycle. (b) Dual topology description: a hybrid ligand with two side chains. "Blocks" are used to define the hybrid energy function [Eq. (14)]. Only the ligand is shown; the environment is either solvent or the solvated protein. (c) Single-topology description.

Bonds:

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

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

Angles:

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

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

Dihedrals:

$$\mathscr{V}_\omega(\lambda) = \lambda\mathscr{V}_\omega(\mathrm{Y}) + (1 - \lambda)\mathscr{V}_\omega(\mathrm{X})$$

Electrostatics:

$$q_\mathrm{i}(\lambda) = \lambda q_\mathrm{i}(\mathrm{Y}) + (1 - \lambda)q_\mathrm{i}(\mathrm{X})$$

van der Waals:

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

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

$$v_{ij}^{LJ}(\lambda) = 4(1-\lambda)\epsilon_X\left(\frac{\sigma_X^{12}}{[\alpha_{LJ}\lambda^2\sigma_X^6 + r_{ij}^6]^2} - \frac{\sigma_X^6}{[\alpha_{LJ}\lambda^2\sigma_X^6 + r_{ij}^6]}\right)$$
$$+ 4\lambda\epsilon_Y\left(\frac{\sigma_Y^{12}}{[\alpha_{LJ}(1-\lambda)^2\sigma_Y^6 + r_{ij}^6]^2} - \frac{\sigma_Y^6}{[\alpha_{LJ}(1-\lambda)^2\sigma_Y^6 + r_{ij}^6]}\right) \qquad (11.31)$$

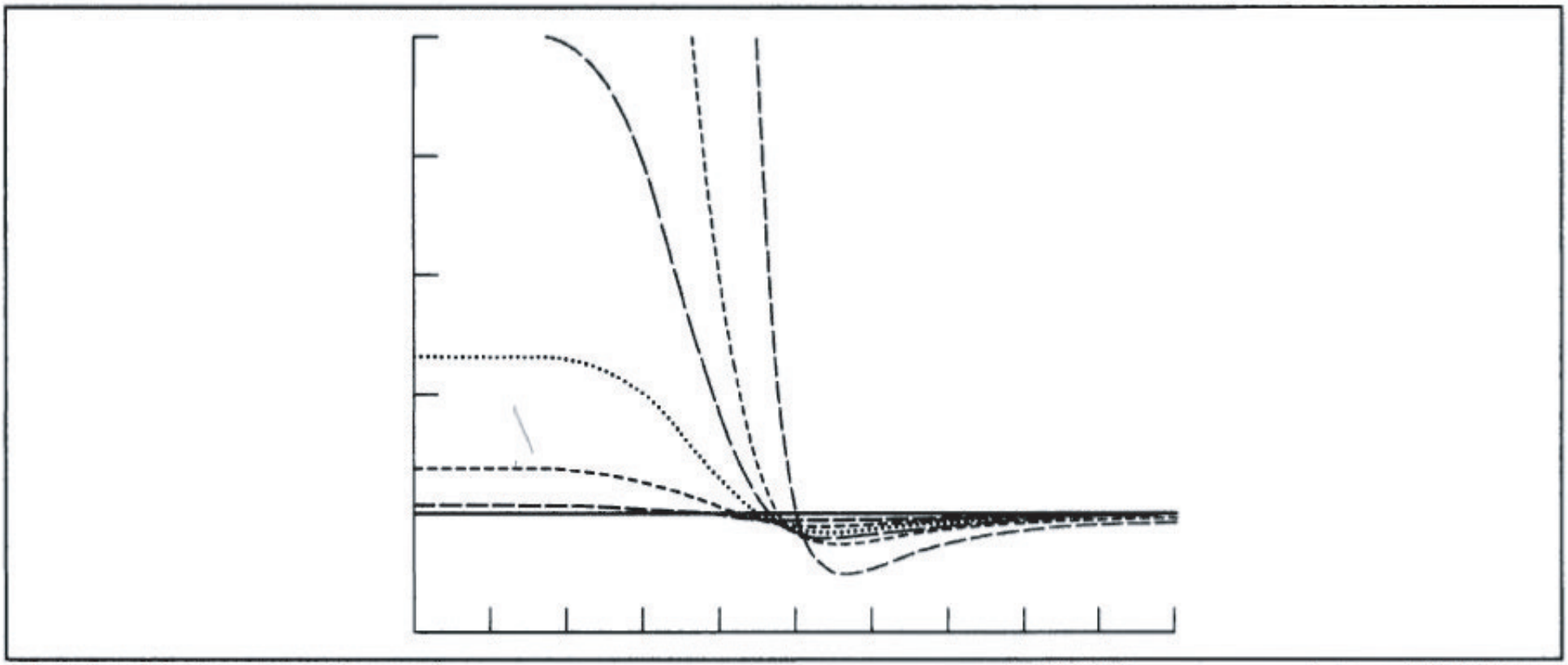


Fig. 11.14: Comparison of scaled and unscaled Lennard-Jones potentials (Equation (11.31)) for the case where a particle disappears at $\lambda = 0$. As λ decreases the curves get progressively closer to the x axis.

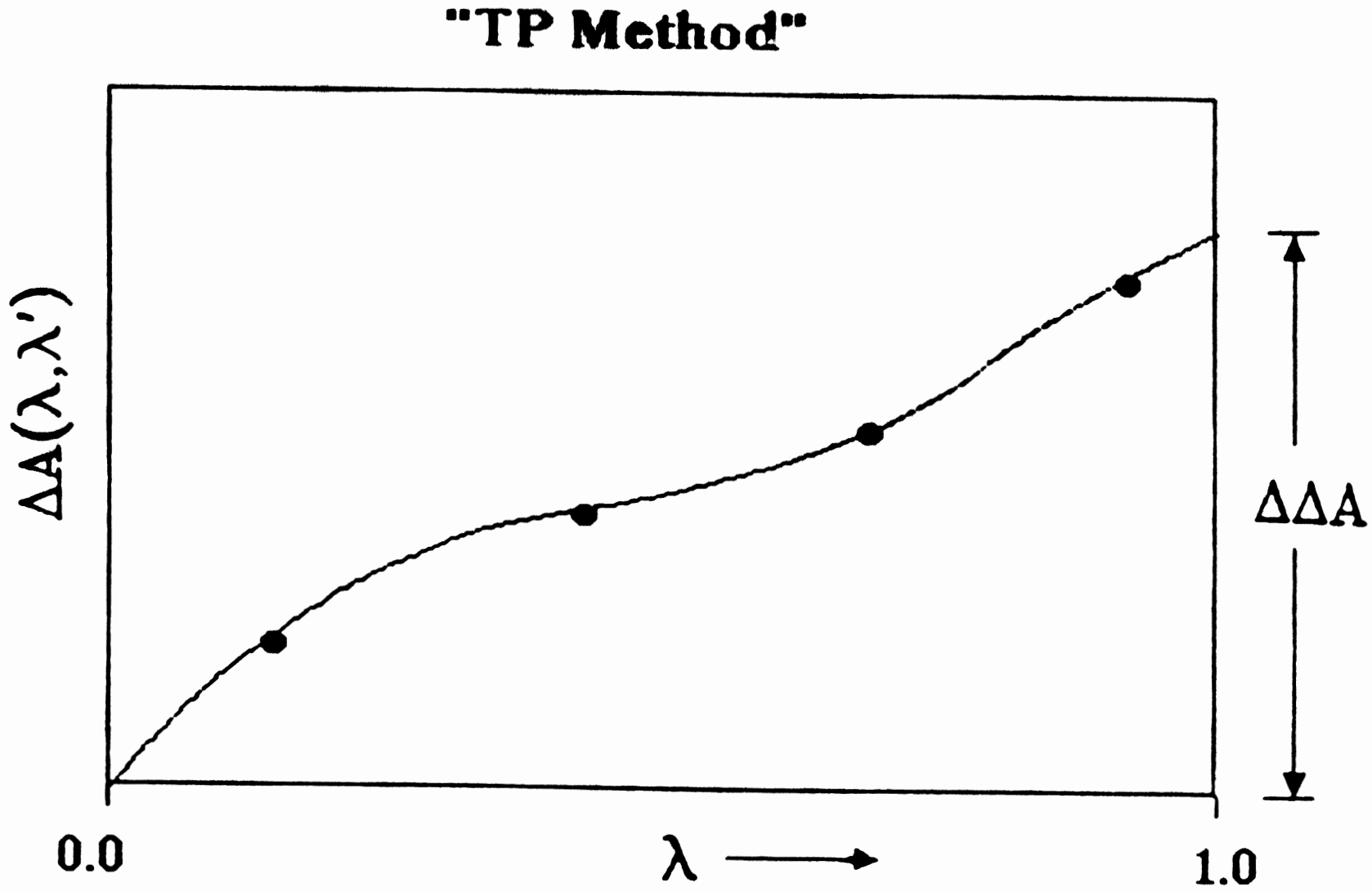


Fig. 1. A schematic representation of results from a free energy simulation using the thermodynamic perturbation (TP) protocol.

Appendix 9.1 Calculating free energy differences using thermodynamic integration

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

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} d\lambda \tag{9.81}$$

Now

$$A(\lambda) = -k_B T \ln Q(\lambda) \tag{9.82}$$

Thus

$$\Delta A = -k_B T \int_0^1 \left[\frac{\partial \ln Q(\lambda)}{\partial \lambda}\right] d\lambda = \int_0^1 \frac{-k_B T}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda} d\lambda \tag{9.83}$$

From the definition of Q (section 5.1.1):

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int\int d\mathbf{p}^N d\mathbf{r}^N \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \tag{9.84}$$

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

$$\frac{\partial Q(\lambda)}{\partial \lambda} = \frac{1}{N!}\frac{1}{h^{3N}} \int\int d\mathbf{p}^N d\mathbf{r}^N \frac{\partial}{\partial \lambda} \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \tag{9.85}$$

Applying the chain rule:

$$\frac{\partial Q(\lambda)}{\partial \lambda} = -\frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{k_B T} \int\int d\mathbf{p}^N d\mathbf{r}^N \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \tag{9.86}$$

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

$$\begin{aligned}\frac{\partial A(\lambda)}{\partial \lambda} &= \frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{Q(\lambda)} \int\int d\mathbf{p}^N d\mathbf{r}^N \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \exp\left[-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right] \\ &= \int\int d\mathbf{p}^N d\mathbf{r}^N \frac{\partial H(\mathbf{p}^N, \mathbf{r}^N)}{\partial \lambda} \left\{\frac{\exp\left[-\dfrac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}\right]}{Q(\lambda)}\right\} \\ &= \left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N, \lambda}{\partial \lambda} \right\rangle_\lambda \end{aligned} \tag{9.87}$$

Thus

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(p^N, \mathbf{r}^N, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{9.88}$$

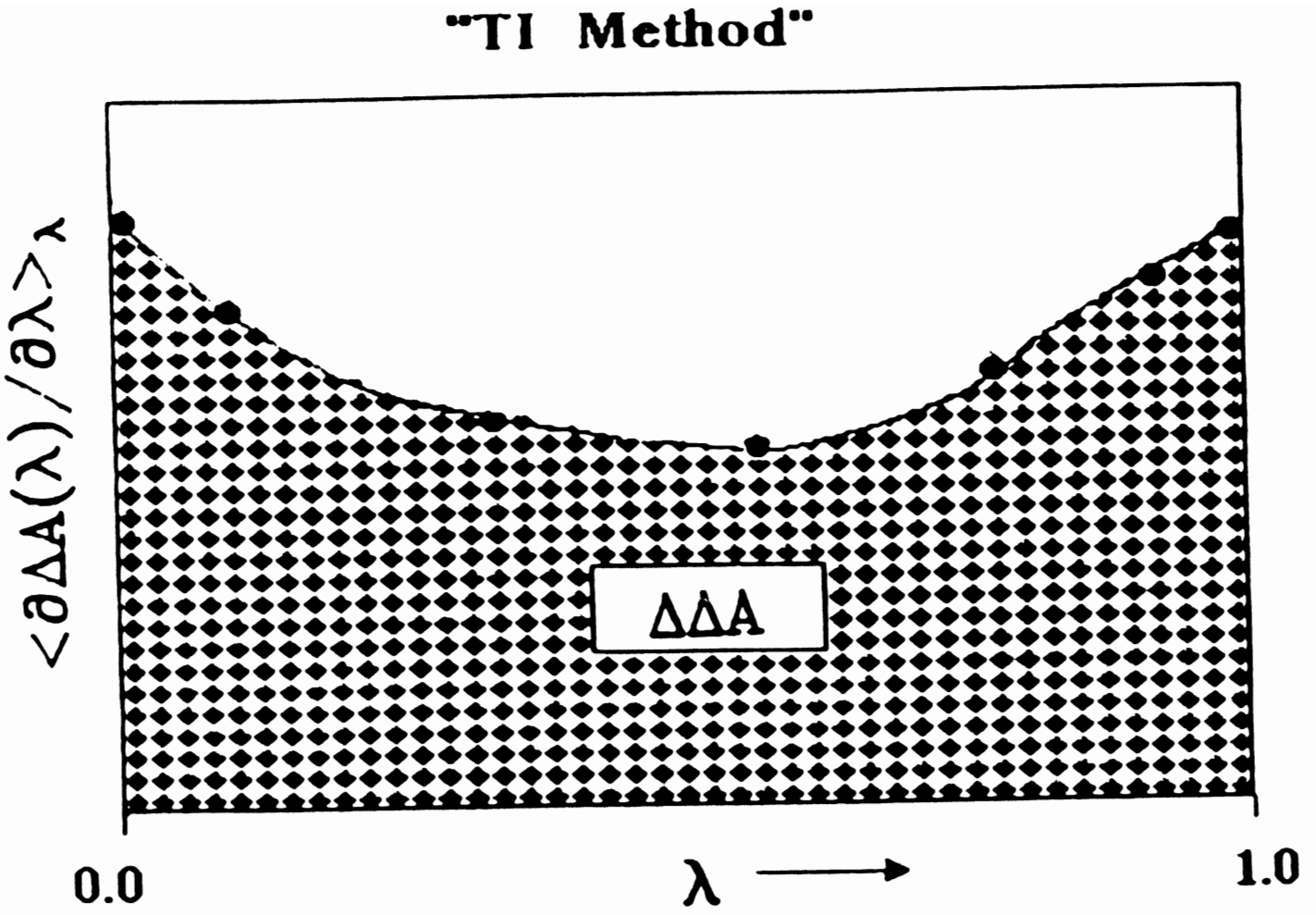


Fig. 2. A schematic representation of results from a free energy simulation using the thermodynamic integration (TI) protocol.

Recent Applications of Free Energy Calculations:

(1) Uses in Force Field Parameterization

(2) Validity of Free Energy Decompositions

(3) Massive Sampling to Reduce Statistical Error

(4) Computing Absolute Free Energies

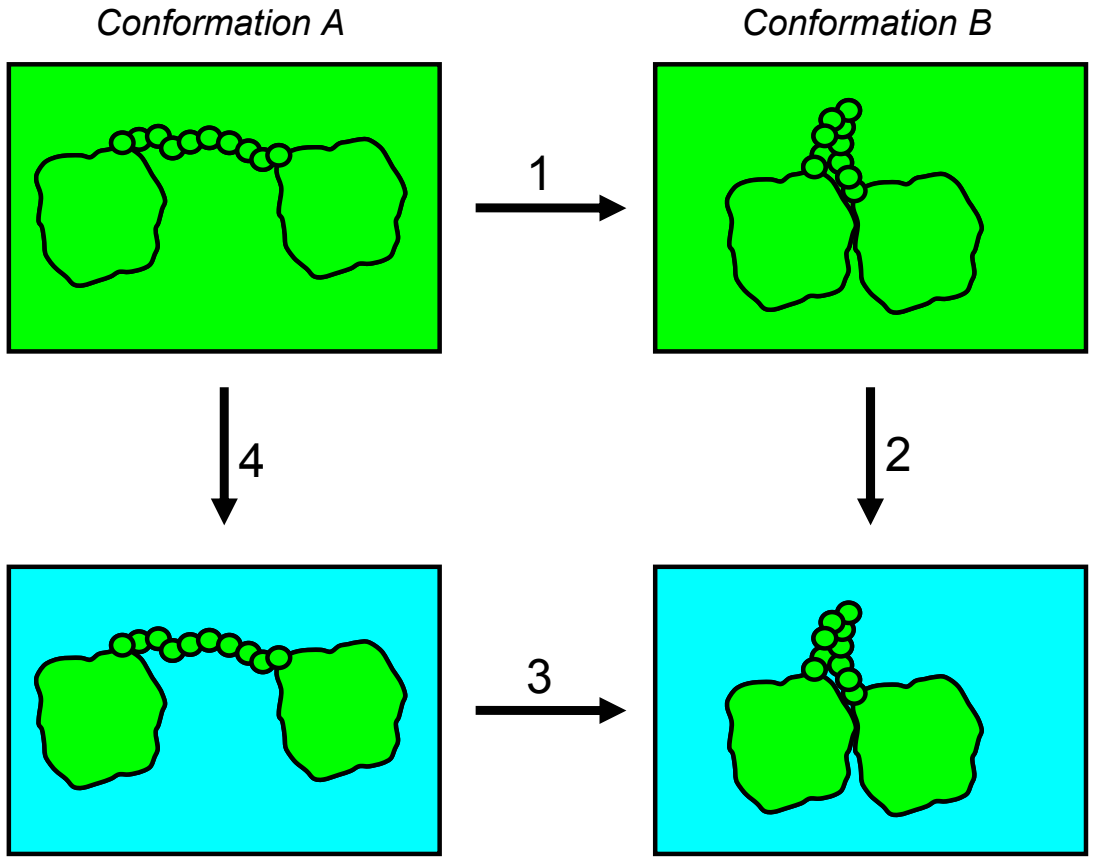
(5) Jarzynski's Non-Equilibrium Work Method

(6) Linear Interaction Energy Method

Free energy cycle

- Thermodynamic energies are “state variables”
- Remember Hess’ law? *Products minus reactants*
- Free energy cycles allow:
  - separate complicated processes into tractable parts
  - multiple ways to calculate energy
  - validation

Example free energy cycle



$$\Delta G_1 + \Delta G_2 - \Delta G_3 - \Delta G_4 = 0$$

$$\Delta G_3 = \Delta G_1 + \Delta G_2 - \Delta G_4$$

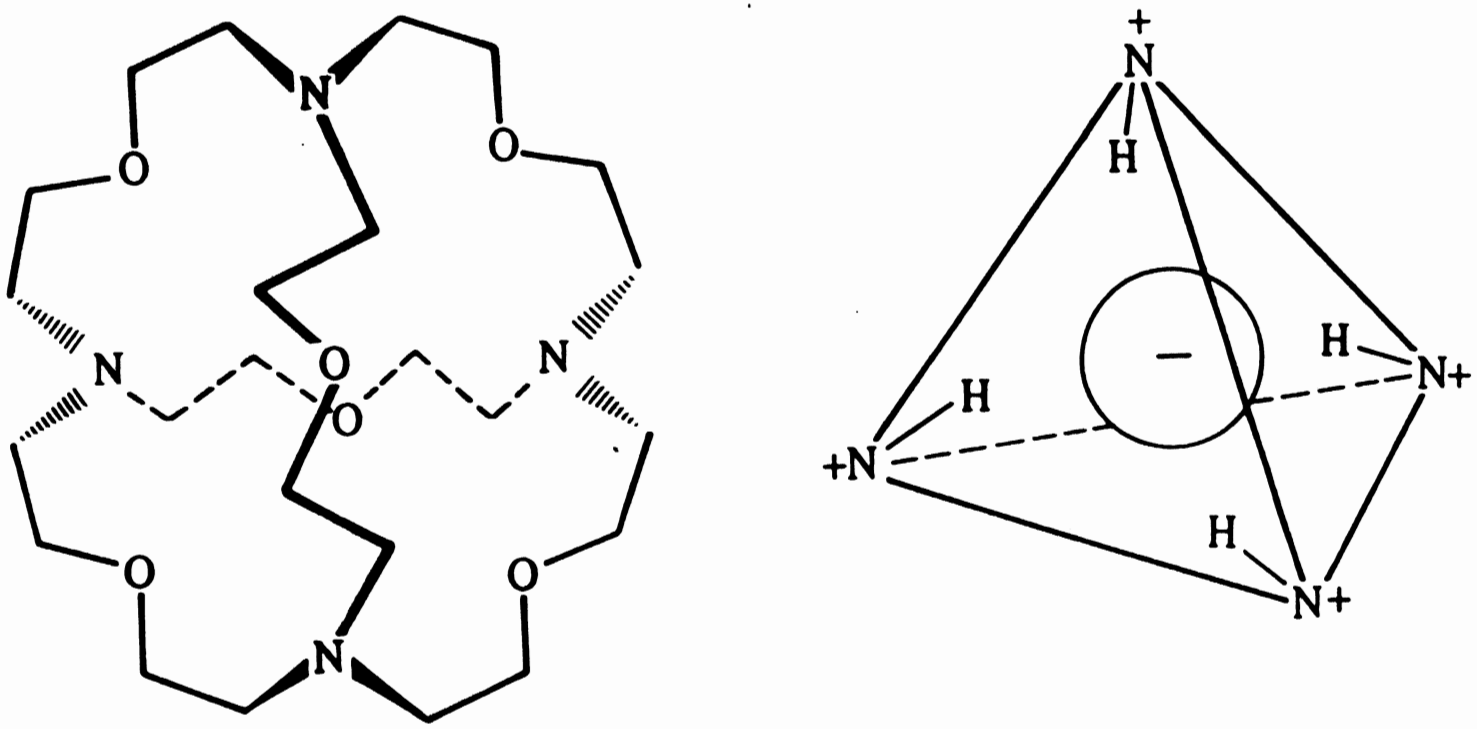


Fig. 9.6 The SC24/halide system. Figure adapted from Lybrand T P, J A McCammon and G Wipff 1986. Theoretical Calculation of Relative Binding Affinity in Host–Guest Systems. *Proceedings of the National Academy of Sciences USA* **83**:833–835.

Table I

Amino Acid Side Chain Potentials					
Solute[a]	Reference[a]	ELE[b]	VDW[c]	ΔΔG[d]	Experimental[e]
arginine (+)	arginine	–	–	-38.28 ± 0.71	–
arginine	alanine	-12.51 ± 0.28	2.75/-1.03 -1.89 ± 1.21	-14.40 ± 1.49	-12.98
cysteine	alanine	-2.02 ± 0.14	1.18/-0.85 -1.02 ± 0.23	-3.04 ± 0.37	-3.18
histidine (+)	histidine	–	–	-35.49 ± 0.38	–
histidine	alanine	-11.02 ± 0.51	1.80/-2.80 -2.30 ± 0.71	-13.32 ± 0.20	-12.30
leucine	alanine	0.30 ± 0.05	-0.24/0.46 0.35 ± 0.16	0.65 ± 0.20	–
lysine (+)	lysine	–	–	-58.85 ± 0.60	–
lysine	alanine	-7.03 ± 0.28	-0.40/0.15 0.28 ± 0.18	-6.75 ± 0.10	-6.34
phenyalanine	alanine	-2.35 ± 0.11	0.75/-0.27 -0.51 ± 0.34	-2.87 ± 0.45	-2.70
serine	alanine	-6.16 ± 0.21	0.98/-0.77 -0.88 ± 0.15	-7.04 ± 0.35	-7.00
tryptophan	alanine	-4.10 ± 0.06	2.90/-3.47 -3.19 ± 0.40	-7.29 ± 0.35	–
methane	nothing	0.03 ± 0.02	-1.47/2.80 2.14 ±0.94	2.12 ± 0.90	1.94
3-methylindole	methane	-3.11 ± 0.21	3.11/-3.58 -3.58 ± 0.73	-6.68± 0.93	-7.82
isobutane	methane	-0.11 ± 0.02	-0.95/1.41 1.18 ± 0.33	1.08 ± 0.35	0.38
phenol	benzene	-3.80 ± 0.15	0.79/-0.73 -0.76 ± 0.04	-4.57 ± 0.19	-4.14
acetamide	methane	-8.93 ± 0.16	1.10/-1.63 -1.37 ± 0.37	-10.30 ± 0.21	-11.77
acetate	acetamide	–	–	-71.86 ± 0.88	–
acetic acid	methane	-6.80 ± 0.30	2.87/-1.72 -2.30 ± 0.81	-9.03 ± 0.45	-8.72

a. Free energies are given for reference state → solute in kcal/mol.

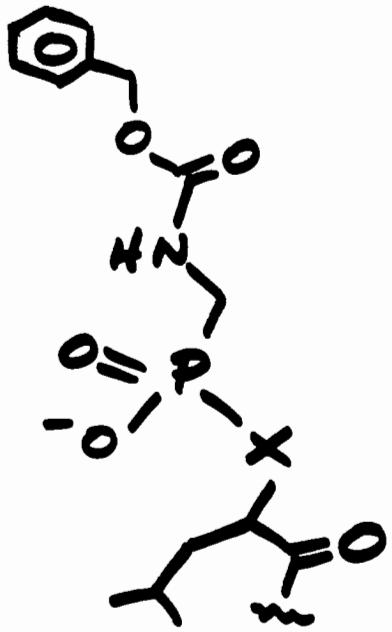
b. Electrostatic free energy using thermodynamic windowing.

BASH et al SCIENCE 236 564 (1987)

(BASH et al SCIENCE 235 574 '87)

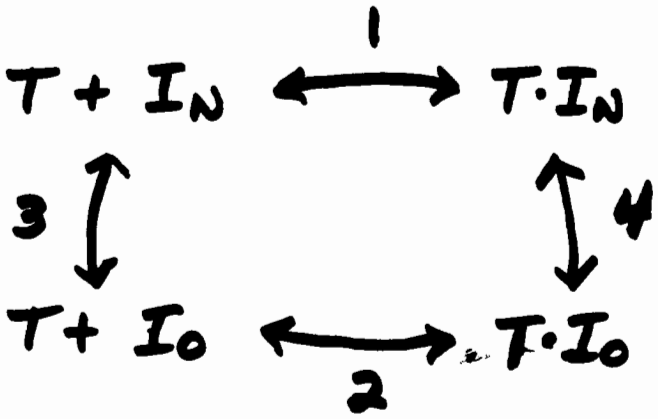
THERMODYNAMIC PERTURBATION METHOD

THERMOLYSIN +



$X = O$ (I_O)

NH_2 (I_N)



EXPT 4.1 KCAL

CALC 4.21 ± .54

$$\Delta G = G(\lambda + d\lambda) - G(\lambda)$$

$$= -RT \left\langle e^{-H(\lambda+d\lambda)+H(\lambda)/RT} \right\rangle$$

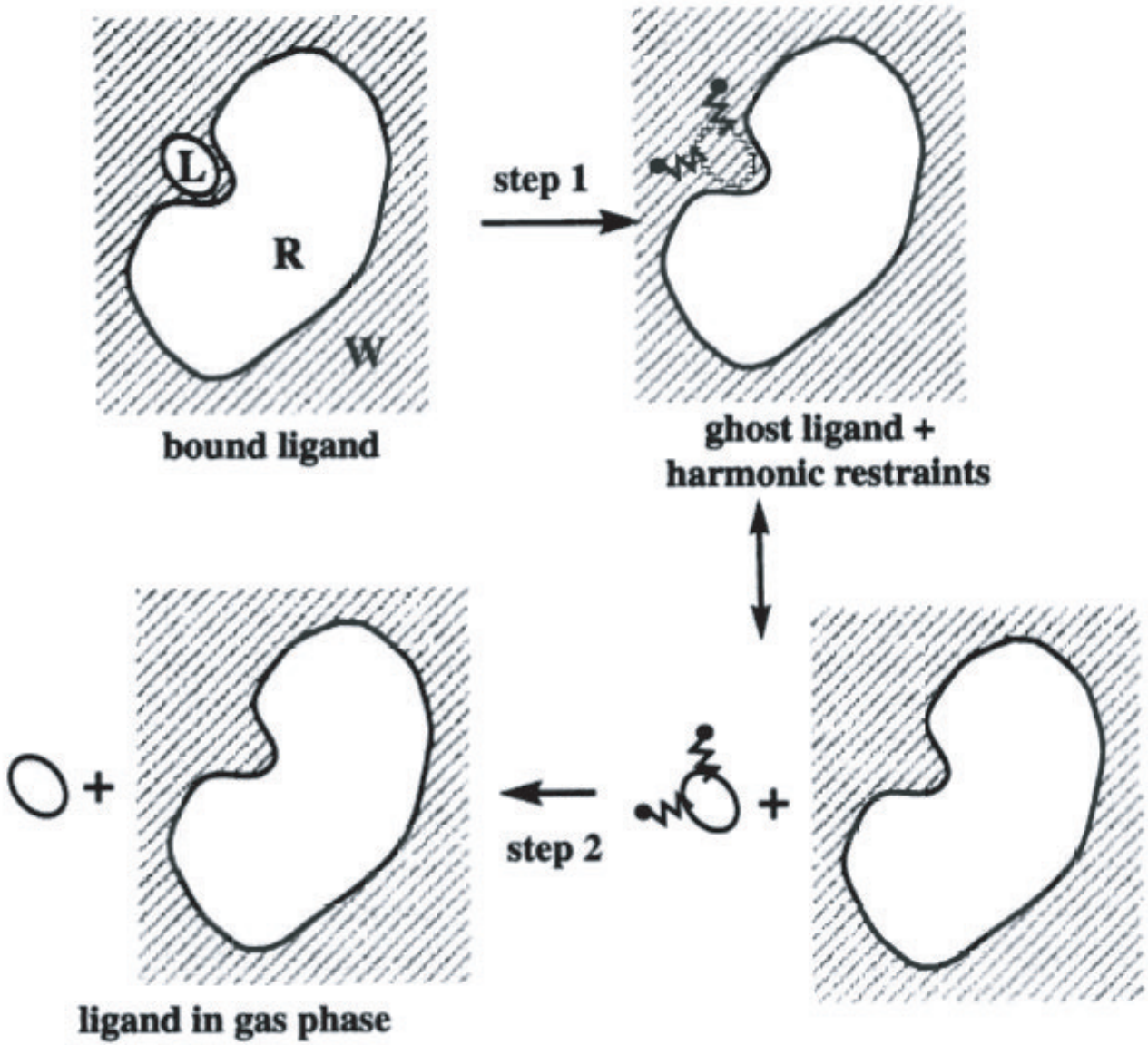
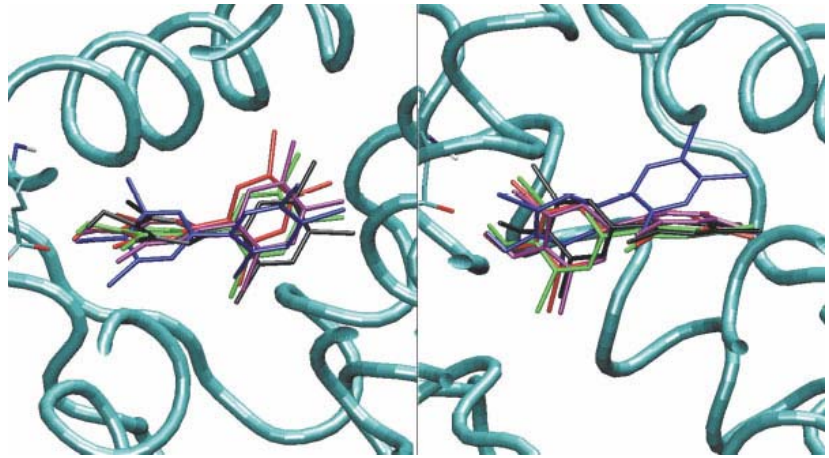
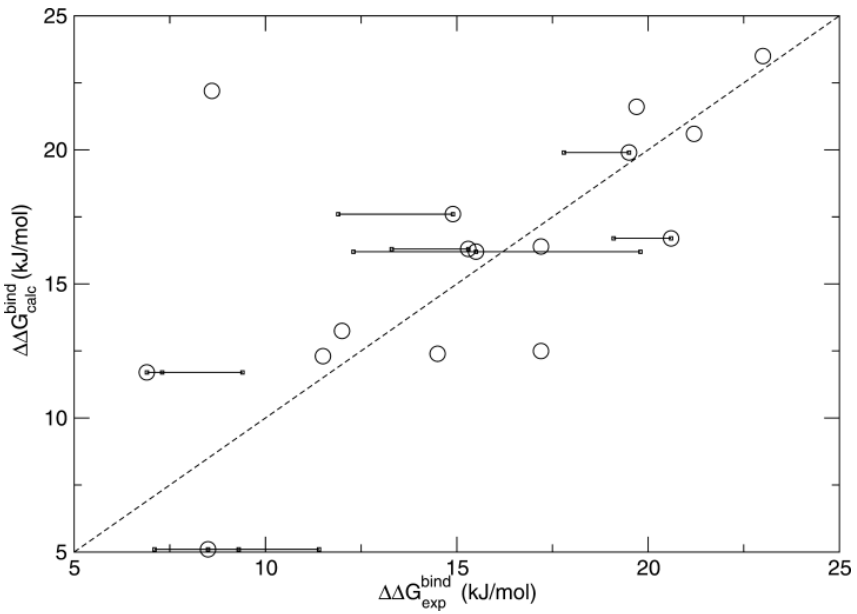


Figure 4 Thermodynamic pathway for the calculation of a standard binding free energy. In the first step, the interactions of the ligand (L) with its environment (receptor R + solvent W) are gradually turned off; at the same time, one or more harmonic restraints are turned on, restricting the translation and possibly the rotation of the ligand. The restrained, ghost ligand (at right in figure) is strictly equivalent to the restrained ligand in the gas phase. The second step removes the harmonic restraints; the corresponding free energy has a simple analytical form (see text).

Free energy perturbation example

- Oostenbrink C, van Gunsteren WF. *Proteins* **54** (2) 234-246, 2004.
- Poly-chlorinated biphenyl binding to estrogen receptor
- “Fast” FEP on 17 compounds
- Good agreement with experiment
- Insight into structural and dynamic aspects of ligand binding





GROMOS Liquid Densities

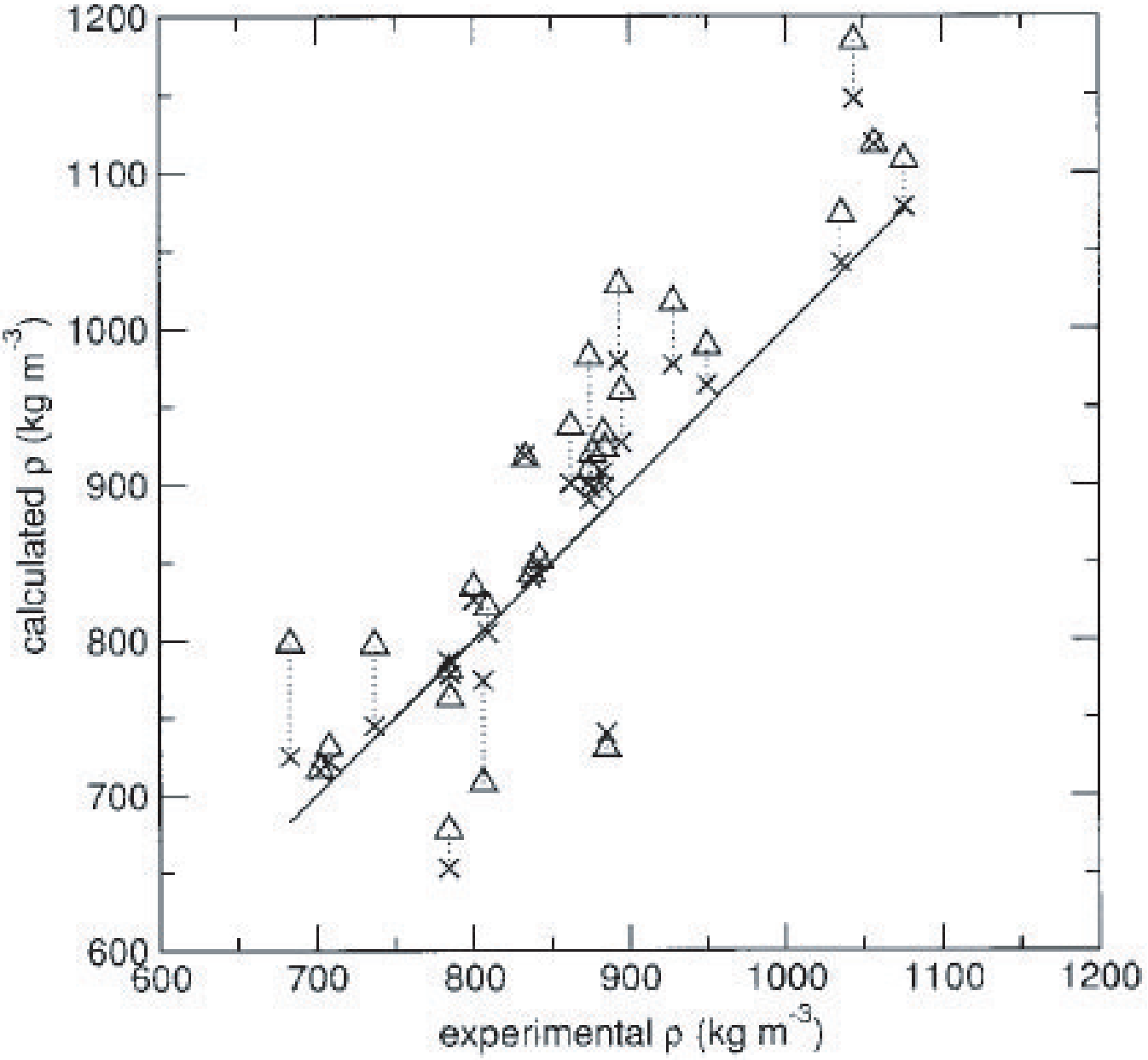


Figure 1. Density for pure liquids. Comparison of experimental densities to calculated values obtained using parameter sets 45A3 (triangles) and 53A5 (crosses) for the 28 compounds listed in Table 16. Diagonal line corresponds to perfect agreement with experiment. Dotted lines are drawn to facilitate comparison between values for the same compound obtained with different parameter sets.

GROMOS Heat of Vaporization

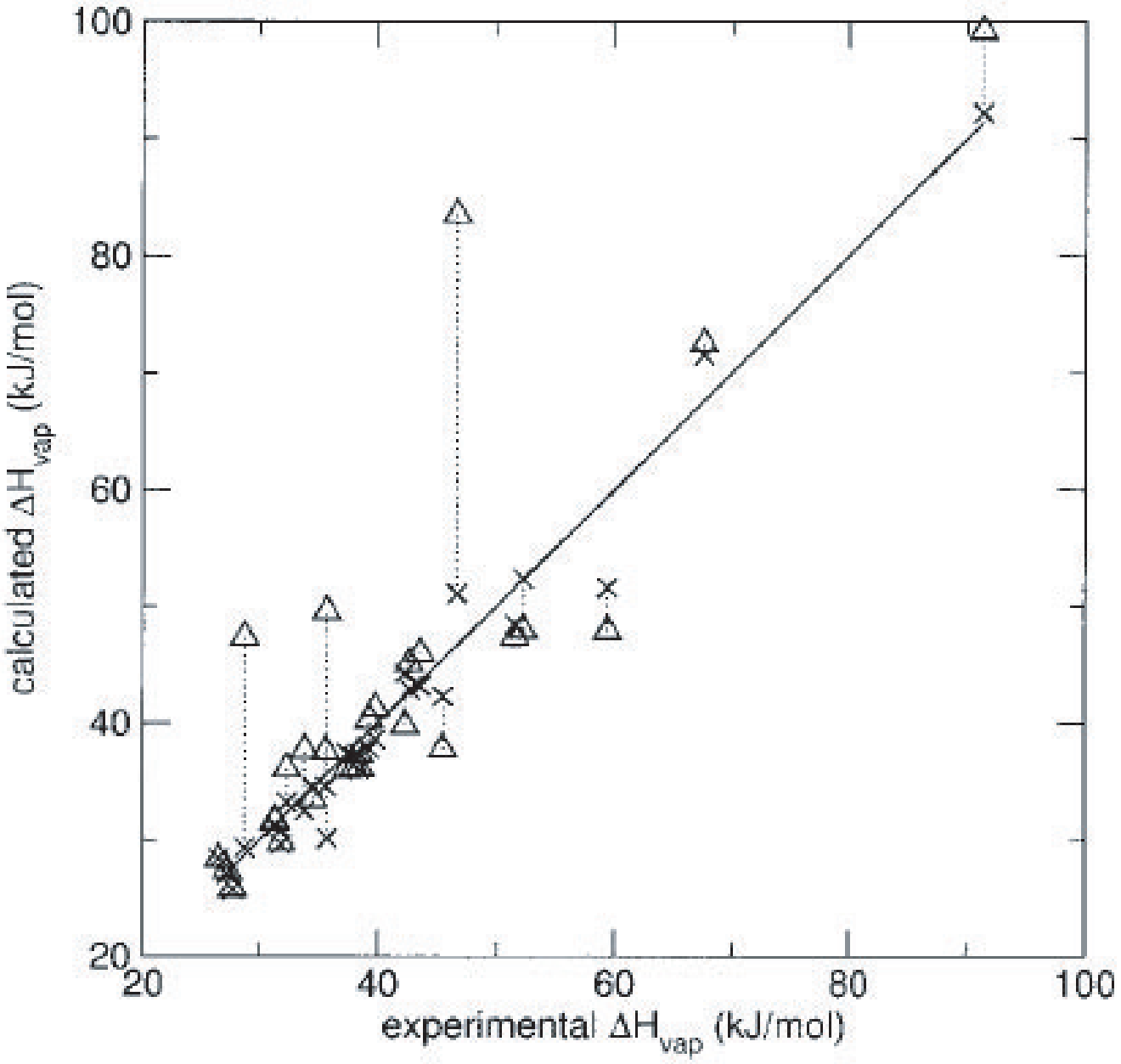


Figure 2. Heat of vaporization for pure liquids. Comparison of experimental heats of vaporization to calculated values obtained using parameter sets 45A3 (triangles) and 53A5 (crosses) for the 28 compounds listed in Table 16. Diagonal line corresponds to perfect agreement with experiment. Dotted lines are drawn to facilitate comparison between values for the same compound obtained with different parameter sets.

GROMOS ΔG_{solv} in Cyclohexane

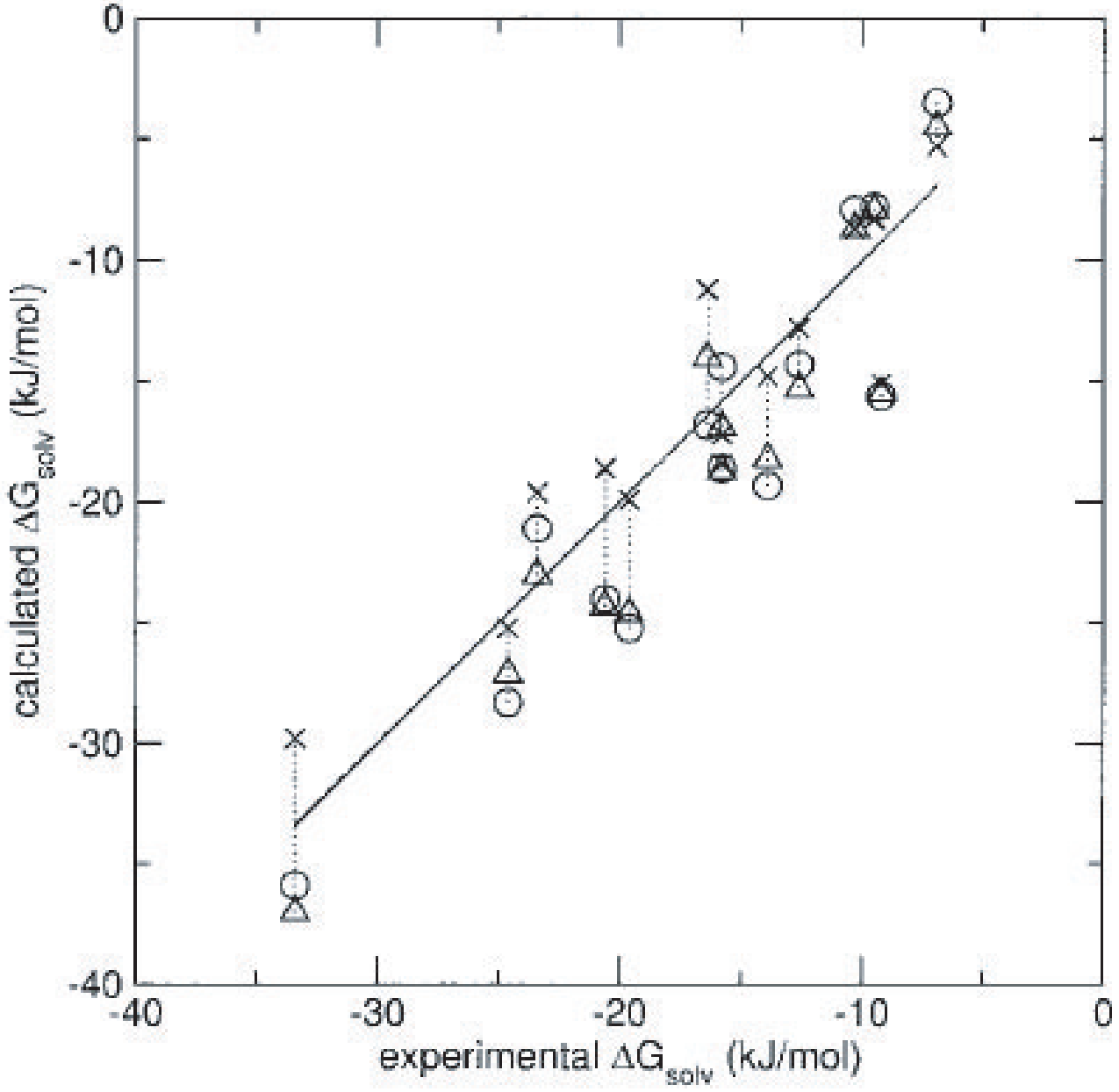


Figure 3. Free enthalpies of solvation in cyclohexane. Comparison of experimental free enthalpies of solvation to calculated values obtained using parameter sets 43A2 (circles), 45A3 (triangles), and 53A5 (crosses) for the 14 compounds listed in Table 17. Diagonal line corresponds to perfect agreement with experiment. Dotted lines are drawn to facilitate comparison between values for the same compound obtained with different parameter sets.

GROMOS ΔG_{solv} in Water

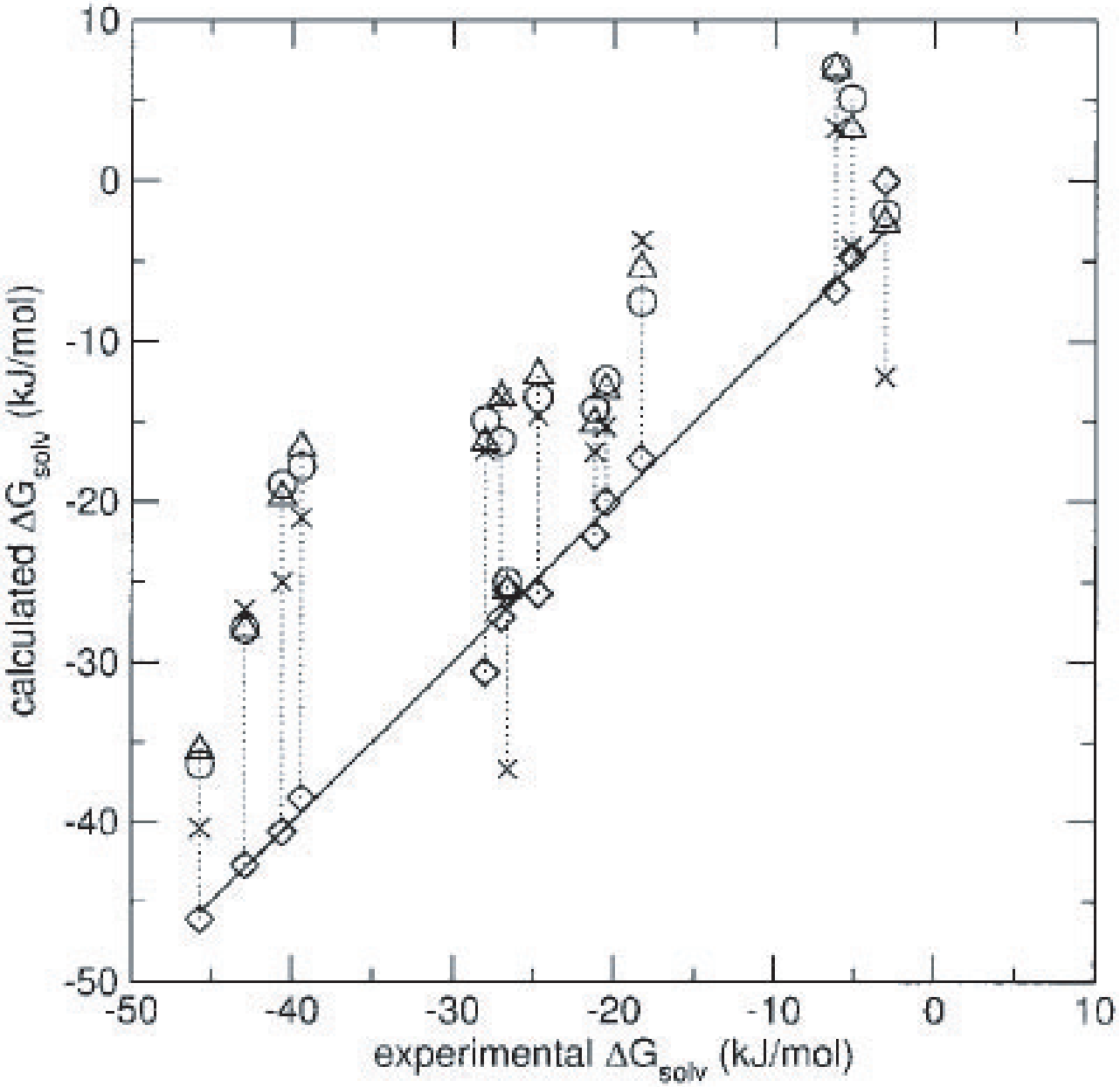


Figure 4. Free enthalpies of solvation in water. Comparison of experimental free enthalpies of solvation to calculated values obtained using parameter sets 43A2 (circles), 45A3 (triangles), 53A5 (crosses), and 53A6 (diamonds) for the 14 compounds listed in Table 18. Diagonal line corresponds to perfect agreement with experiment. Dotted lines are drawn to facilitate comparison between values for the same compound obtained with different parameter sets.

The problem of protein stability can be described in terms of the following thermodynamic cycle.

$$\begin{array}{ccc} & \xrightarrow{\Delta G_{N_F - N_U}} & \\ N_F & \longrightarrow & N_U \\ \Delta G_{N_F - M_F} \downarrow & & \downarrow \Delta G_{N_U - M_U} \\ M_F & \longrightarrow & M_U \\ & \xrightarrow[\Delta G_{M_F - M_U}]{} & \end{array} \qquad (1)$$

In these terms the relative stability of the native, N, as compared to the mutant, M, protein is given by

$$\Delta\Delta G = \Delta G_{N_F - N_U} - \Delta G_{M_F - M_U} = \Delta G_{N_F - M_F} - \Delta G_{N_U - M_U} \qquad (2)$$

where $\Delta G_{N_F - N_U}$ and $\Delta G_{M_F - M_U}$ represent the free energy change going from the folded (F) to the unfolded (U) state for the native and mutant proteins respectively.

Table I. Details of the size of the system and a summary of simulation parameters

	Folded protein	Unfolded peptide
Total number of atoms	2102	2346
Number of water molecules	435	771
Reference temperature (K)	300	300
Boundary conditions		
Boundary type	extended wall	extended wall
Radius of outer sphere (nm)	1.8	1.8
Radius of inner sphere (nm)	1.2	1.2
Cut-offs		
Short range non-bonded (nm)	0.8	0.8
Long range electrostatic (nm)	1.1	1.1
Integration time step		
Bond stretching forces (fs)	0.3	0.3
Other bonded and short range non-bonded forces (fs)	1.5	1.5
Long range electrostatic force (fs)	15	15
Equilibration period (ps)	10	10
λ change over period (ps)	220 (A), 240 (B)	240
Dielectric constant	1	1

Table II. Contributions to the total free energy change separated on the basis of terms in the force field

	$\Delta G_{N_F - M_F}$ (A) (kJ/mol)	$\Delta G_{N_F - M_F}$ (B) (kJ/mol)	$\Delta G_{N_U - M_U}$ (kJ/mol)	$\Delta G_{N_F - M_F} - \Delta G_{N_U - M_U}$ (A) (kJ/mol)	$\Delta G_{N_F - M_F} - \Delta G_{N_U - M_U}$ (B) (kJ/mol)
Lennard–Jones					
Protein	4.4	4.2	3.1	1.3	1.1
Internal	− 3.6	− 2.5	− 0.7	− 2.9	−1.8
Solvent	− 12.1	− 10.6	− 10.3	− 1.8	−0.3
Total	− 11.3	− 8.9	− 7.9	− 3.4	−1.0
Short range electrostatic					
Protein	13.4	− 11.7	− 6.6	20.0	−5.1
Internal	147.9	153.9	154.3	− 6.4	−0.4
Solvent	− 22.0	− 10.1	− 12.0	− 10.0	1.9
Total	139.3	132.1	135.7	3.6	−3.6
Long range electrostatic					
Protein	− 7.3	1.5	0.1	− 7.4	1.4
Internal	0.0	0.0	0.0	0.0	0.0
Solvent	− 0.2	− 1.2	0.8	− 1.0	−2.0
Total	− 7.5	0.3	0.9	− 8.4	−0.6
Bond angles	− 22.1	− 21.3	− 21.9	− 0.2	0.6
Improper dihedral	− 10.6	− 11.3	− 10.1	− 0.5	−1.2
Proper dihedral	− 5.9	− 5.9	− 5.6	− 0.3	−0.3
Total	81.9	85.0	91.1	− 9.2	−6.1

$\Delta G_{N_F - M_F}$ (A) and (B) correspond to the mutation of the folded protein over 220 and 240 ps respectively using slightly different protocols. $\Delta G_{N_U - M_U}$ corresponds to the free energy change in the peptide model of the unfolded state. The final two columns correspond to the net contribution for simulations (A) and (B). Internal contributions arise from interactions within the perturbed residue itself.

Table III. Contributions to the total free energy separated on the basis of amino acid residues for simulations $N_F - M_F$ (B) and $N_U - M_U$

Residue	$\Delta G_{N_F - M_F}$ (B) Lennard–Jones (kJ/mol)	Electrostatic (kJ/mol)	Total (kJ/mol)	$\Delta G_{N_U - M_U}$ Lennard–Jones (kJ/mol)	Electrostatic (kJ/mol)	Total (kJ/mol)	$\Delta\Delta G$ Net (kJ/mol)
Tyr6		−0.3	− 0.3				− 0.3
His64		−1.0	− 1.0				− 1.0
His67		0.3	0.3				0.3
Asn155		−0.5	− 0.5				− 0.5
Gly178		−0.3	− 0.3				− 0.3
Ala179		−0.3	− 0.3				− 0.3
Asp181		1.9	1.9				1.9
Ser188		0.7	0.7				0.7
Phe189	1.6	0.6	2.2				2.2
Ala200		−0.4	− 0.4				− 0.4
Pro201		0.4	0.4				0.4
Gly202	0.4	−0.1	0.3				0.3
Val203	0.3	0.3	0.6				0.6
Ser204	−1.9	−8.9	−10.8				−10.8
Ile205	0.2	− 0.5	− 0.3				− 0.3
Ala216	0.5	1.1	1.7				1.7
Tyr217	1.8	−2.5	− 0.7	0.2	2.3	2.5	− 3.2
Gly219	0.4	0.7	1.1	1.5	0.3	1.8	− 0.7

Only residues for which the net contribution is >0.2 kJ/mol are shown.

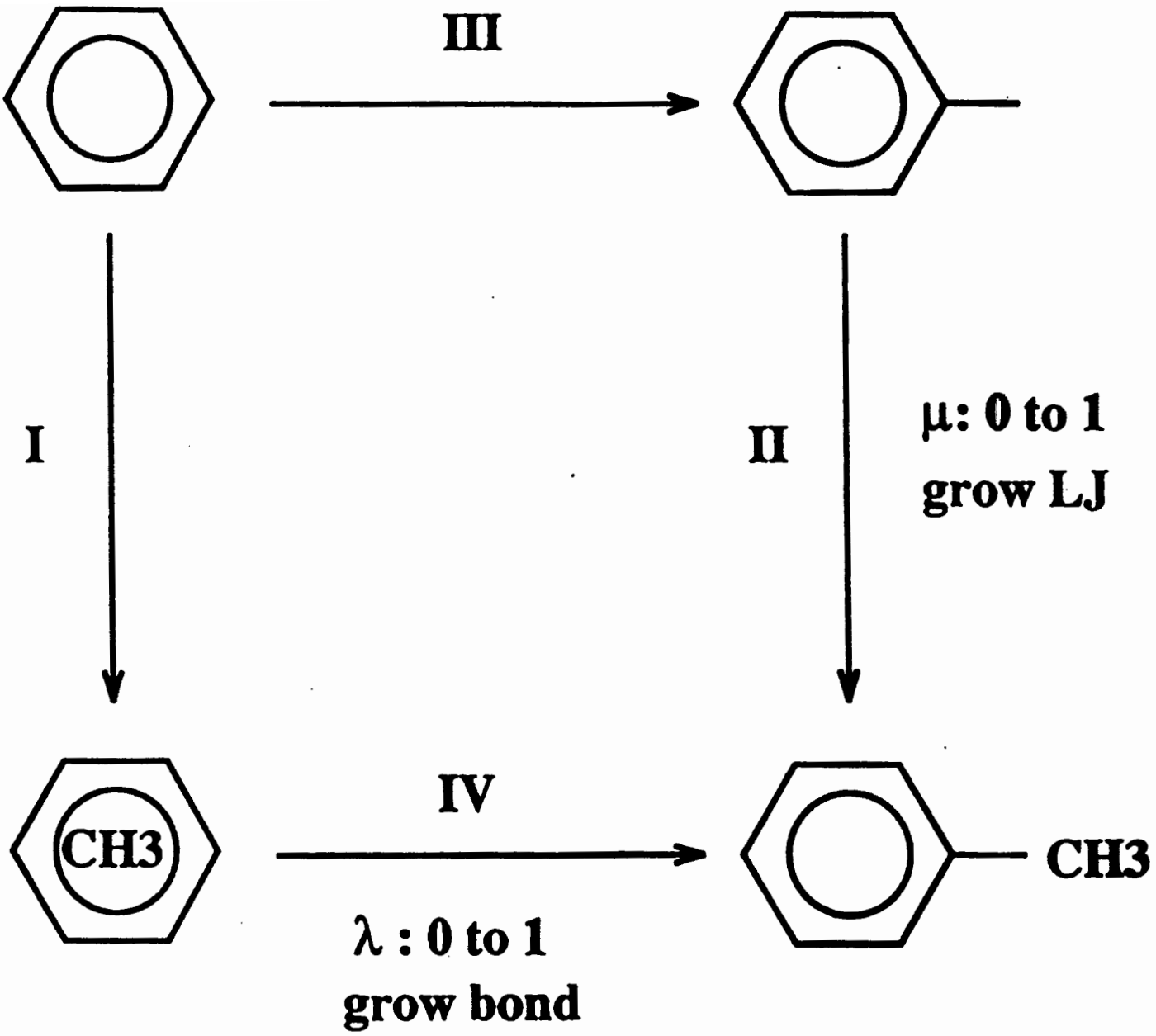


Figure 1. A thermodynamic cycle illustrating 2 possible pathways for the mutation of benzene to toluene in which either the Lennard-Jones (LJ) interaction of the CH3 group ($\mu = 0$ to 1) or the C-CH3 bond ($\lambda = 0$ to 1) is grown first.

JOURNAL OF CHEMICAL PHYSICS VOLUME 119, NUMBER 11 15 SEPTEMBER 2003

Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins

Michael R. Shirts
Department of Chemistry, Stanford University, Stanford, California 94305-5080

Jed W. Pitera and William C. Swope
IBM Almaden Research Center, San Jose, California 95120-6099

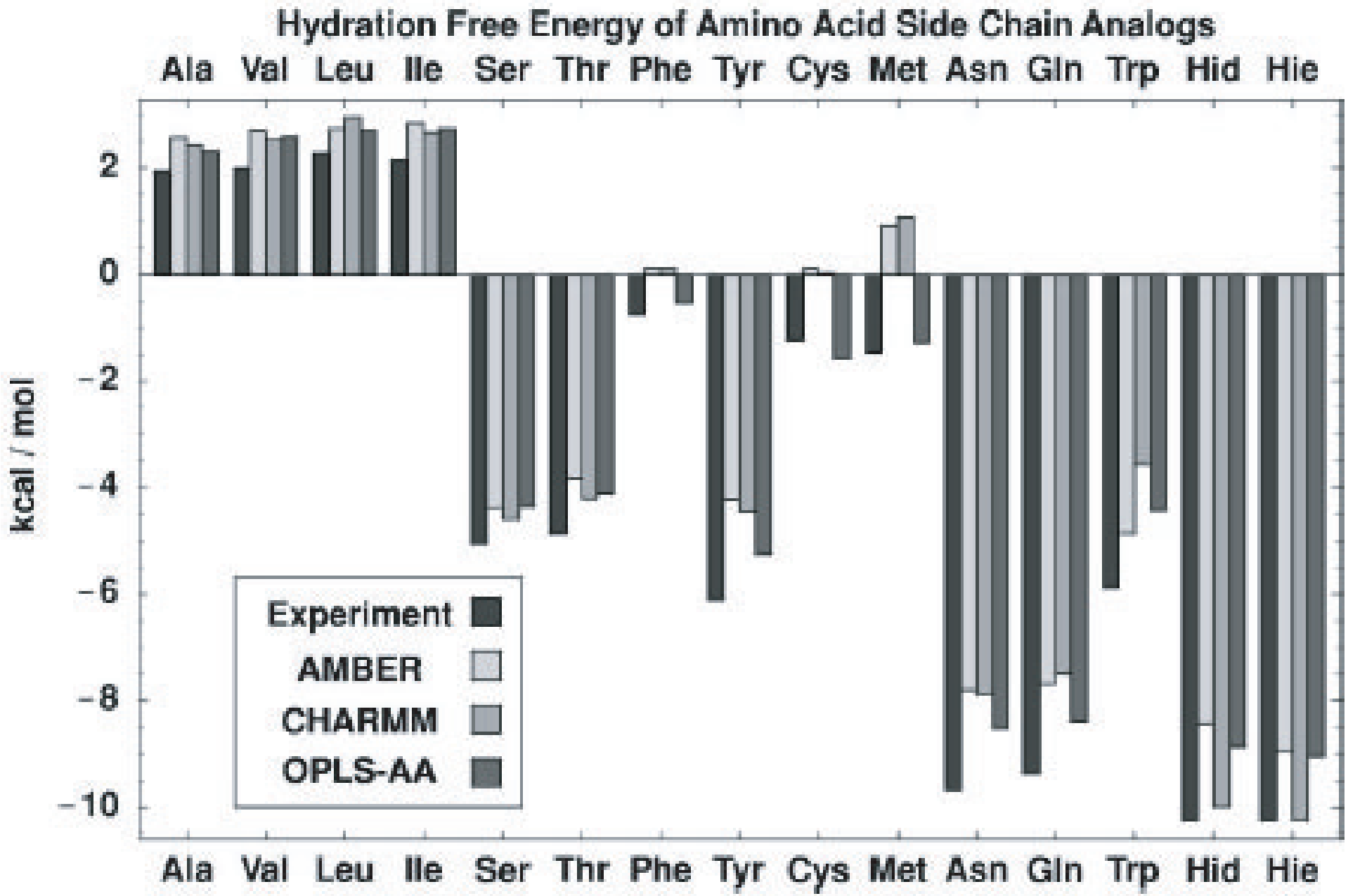
Vijay S. Pande
Department of Chemistry, Stanford University, Stanford, California 94305-5080

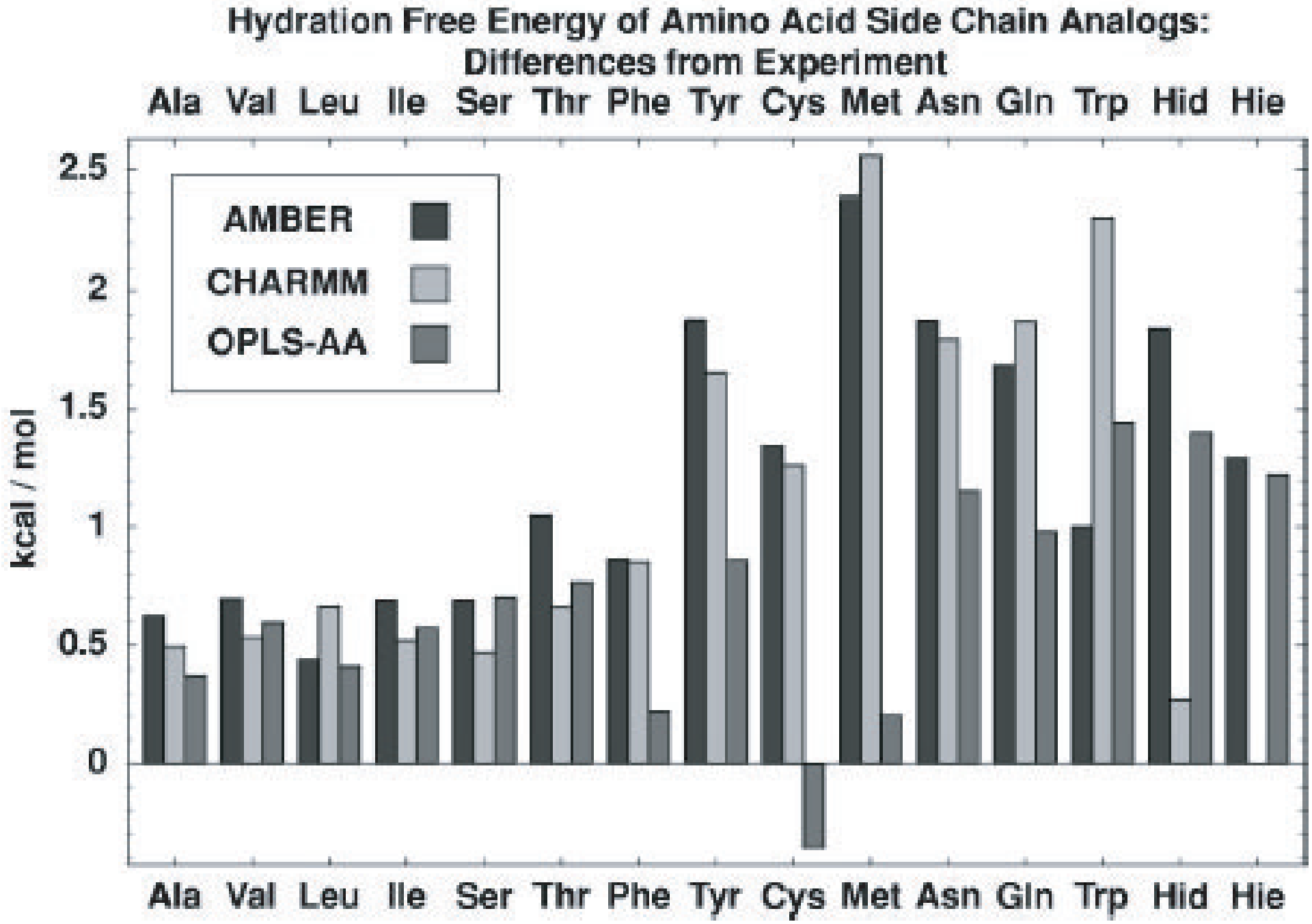
TABLE I. Correspondence between amino acids and the amino acid side chain analogs used in this study.

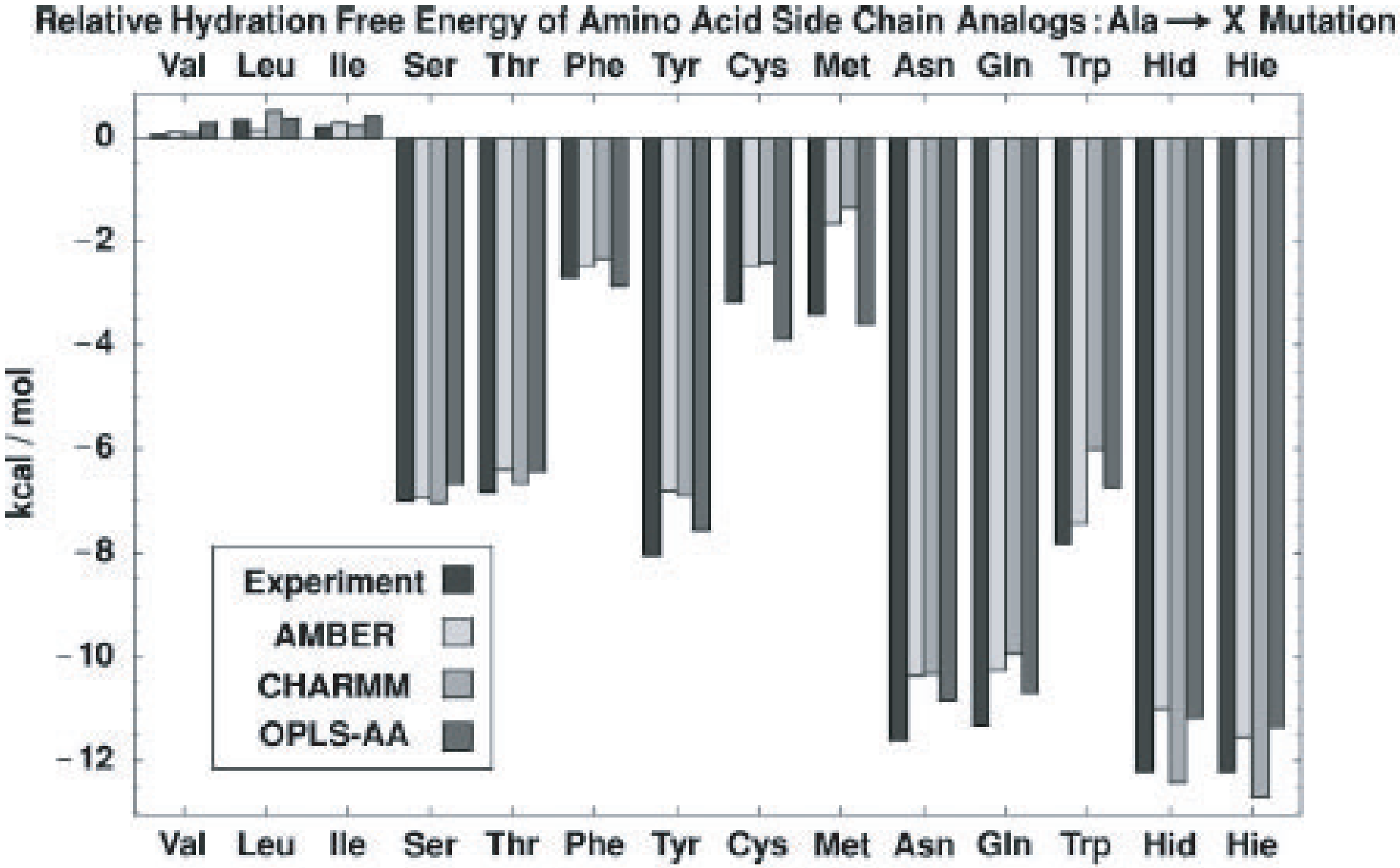
Ala	methane	Val	propane	Leu	iso-butane
Ile	*n*-butane	Ser	methanol	Thr	ethanol
Phe	toluene	Tyr	*p*-cresol	Cys	methanethiol
Met	methyl ethyl sulfide	Asn	acetamide	Gln	propionamide
Trp	3-methylindole	Hid	4-methylimidazole	Hie	4-methylimidazole

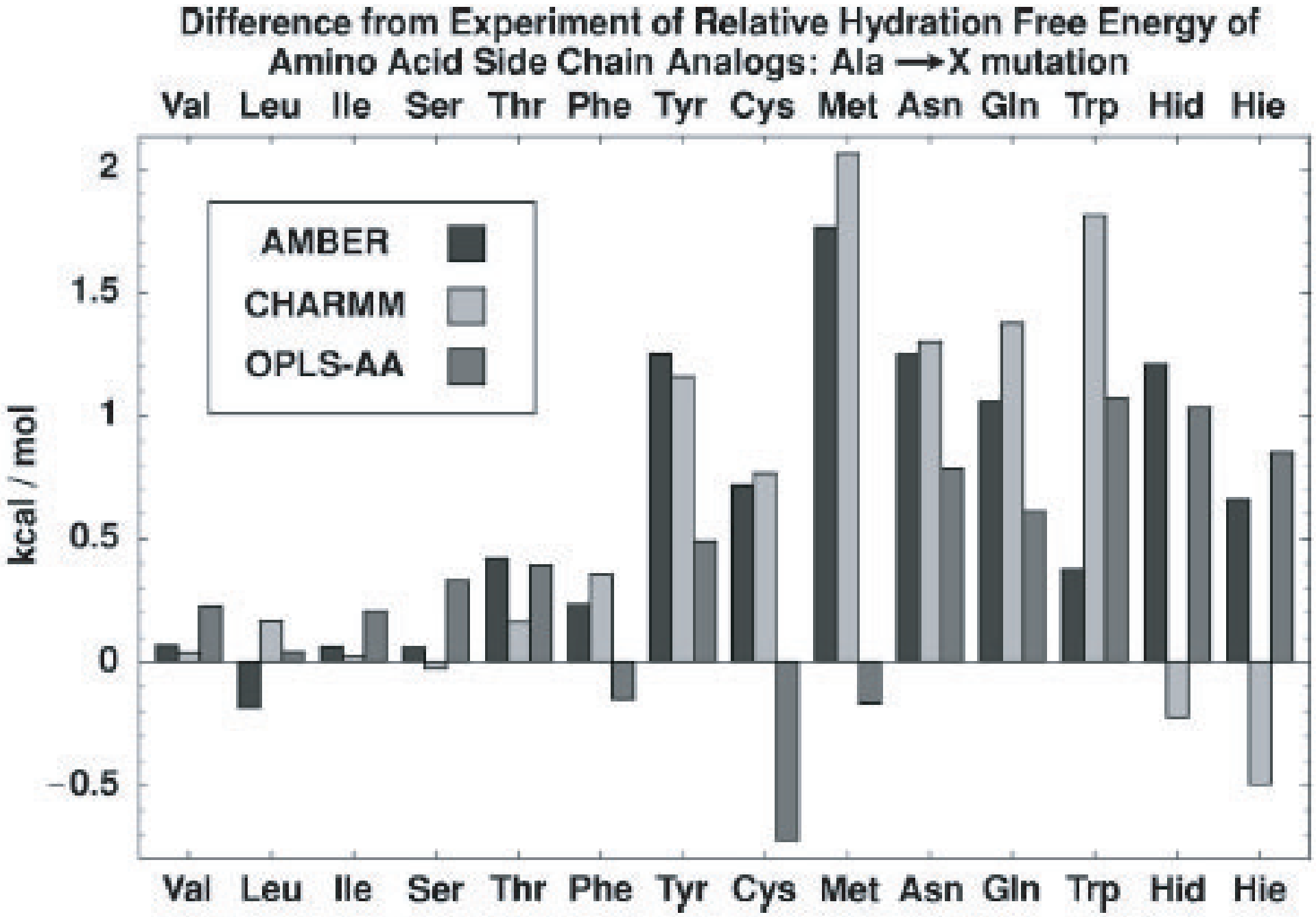
TABLE II. Free energies of hydration for 15 neutral amino acid side chain analogs, for the AMBER($ff94$), CHARMM22, and OPLS-AA force fields, in kcal/mol. Each reported value is the average of five trials. LRC is the long range van der Waals correction as computed by Eq. (9). Uncertainties are the averaged uncertainty of the five runs, and represent one standard deviation from the mean ($\pm 1\sigma$). Experimental values are from Wolfenden *et al.* (Ref. 18).

Side chain analog	Coulombic	van der Waals	Total	LRC	Total (w/LRC)	Expt.
			AMBER			
Ala	$-0.00\pm<0.001$	2.69 ± 0.02	2.68 ± 0.02	-0.12	2.57 ± 0.02	1.94
Val	-0.02 ± 0.001	3.02 ± 0.03	3.00 ± 0.03	-0.31	2.69 ± 0.03	1.99
Leu	-0.05 ± 0.001	3.19 ± 0.03	3.13 ± 0.03	-0.41	2.72 ± 0.03	2.28
Ile	-0.00 ± 0.001	3.27 ± 0.03	3.27 ± 0.03	-0.42	2.84 ± 0.03	2.15
Ser	-6.13 ± 0.01	1.93 ± 0.02	-4.20 ± 0.02	-0.17	-4.37 ± 0.02	-5.06
Thr	-5.77 ± 0.01	2.21 ± 0.03	-3.56 ± 0.03	-0.27	-3.83 ± 0.03	-4.88
Phe	-2.35 ± 0.01	3.03 ± 0.04	0.68 ± 0.04	-0.58	0.10 ± 0.04	-0.76
Tyr	-6.19 ± 0.02	2.65 ± 0.04	-3.54 ± 0.04	-0.69	-4.23 ± 0.04	-6.11
Cys	-2.04 ± 0.01	2.37 ± 0.02	0.33 ± 0.02	-0.22	0.11 ± 0.02	-1.24
Met	-1.61 ± 0.01	2.96 ± 0.03	1.35 ± 0.03	-0.44	0.91 ± 0.03	-1.48
Asn	-9.27 ± 0.02	1.78 ± 0.03	-7.49 ± 0.03	-0.31	-7.80 ± 0.03	-9.68
Gln	-9.35 ± 0.02	2.08 ± 0.03	-7.27 ± 0.04	-0.42	-7.69 ± 0.04	-9.38
Trp	-6.39 ± 0.02	2.35 ± 0.04	-4.05 ± 0.05	-0.83	-4.88 ± 0.05	-5.88
Hid	-9.51 ± 0.02	1.56 ± 0.03	-7.95 ± 0.04	-0.48	-8.43 ± 0.04	-10.27
Hie	-10.07 ± 0.02	1.57 ± 0.03	-8.50 ± 0.04	-0.48	-8.98 ± 0.04	-10.27
	Average absolute error w/o LRC: 1.64				rms error with w/o LRC: 1.76	
	Average absolute error with LRC: 1.22				rms error with LRC: 1.35	


Hydration Free Energy of Amino Acid Side Chain Analogs
Ala Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie
2
0
−2
−4
−6
−8
−10
kcal / mol
Experiment
AMBER
CHARMM
OPLS-AA
Ala Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie


Hydration Free Energy of Amino Acid Side Chain Analogs:
Differences from Experiment
Ala Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie
AMBER
CHARMM
OPLS-AA
kcal / mol
2.5
2
1.5
1
0.5
0
Ala Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie


Relative Hydration Free Energy of Amino Acid Side Chain Analogs : Ala → X Mutation
Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie
0
−2
−4
−6
−8
−10
−12
kcal / mol
Experiment
AMBER
CHARMM
OPLS-AA
Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie


Difference from Experiment of Relative Hydration Free Energy of
Amino Acid Side Chain Analogs: Ala → X mutation
Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie
AMBER
CHARMM
OPLS-AA
kcal / mol
2
1.5
1
0.5
0
−0.5
Val Leu Ile Ser Thr Phe Tyr Cys Met Asn Gln Trp Hid Hie

Jarzynski's Method

Jarzynski's equality is a relation between *equilibrium* free energy differences and work done through *nonequilibrium* processes. Consider a process that changes a parameter λ of a system from λ_0 at time zero to λ_t at time t. The second law of thermodynamics states that the average work done on the system cannot be smaller than the difference between the free energies corresponding to the initial and the final values of λ:

$$\Delta F = F(\lambda_t) - F(\lambda_0) \leqslant \langle W \rangle, \tag{1}$$

where the equality holds only if the process is quasi-static (see, e.g., Ref. 17). According to this inequality, a nonequilibrium process provides only an upper limit for the free energy difference. However, Jarzynski discovered an equality that holds regardless of the speed of the process:

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle. \tag{2}$$

This equality has been tested against computer simulations and experiments.

The major difficulty is that the average of exponential work appearing in Jarzynski's equality is dominated by the trajectories corresponding to small work values that arise only rarely. An accurate estimate of free energy hence requires suitable sampling of such rare trajectories. Therefore, although Jarzynski's equality holds for processes of any speed, practical applications are currently limited to slow processes for which the fluctuation of work is comparable to the temperature.

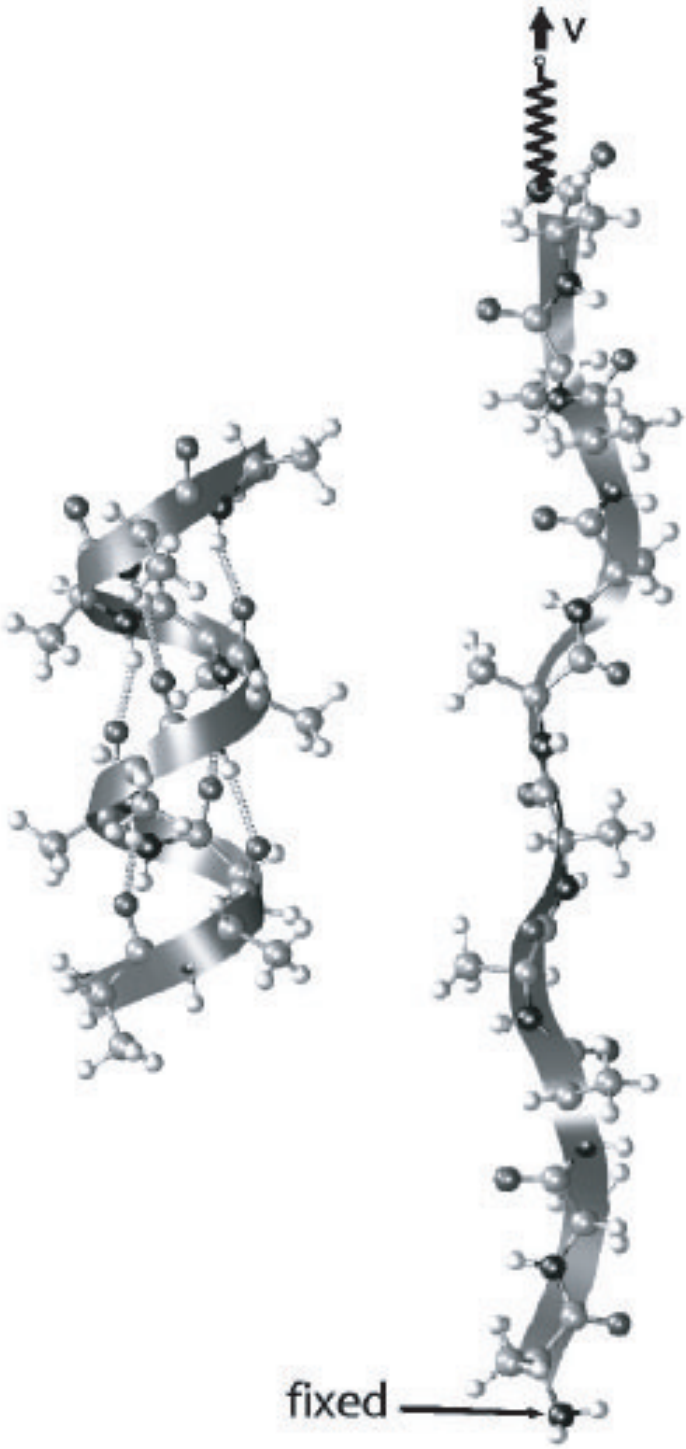


FIG. 1. Unfolding of helical deca-alanine. Left, a folded configuration (α-helix). The six hydrogen bonds that stabilize the helix are shown. Right, an extended configuration (coil). The backbone of the peptide is represented as a ribbon. The N atom of the first residue was fixed during the simulations. The moving guiding potential used in the pulling simulations is represented by a spring which is connected to the C-terminus and pulled with a constant velocity v. Figure made with VMD (Ref. 39).

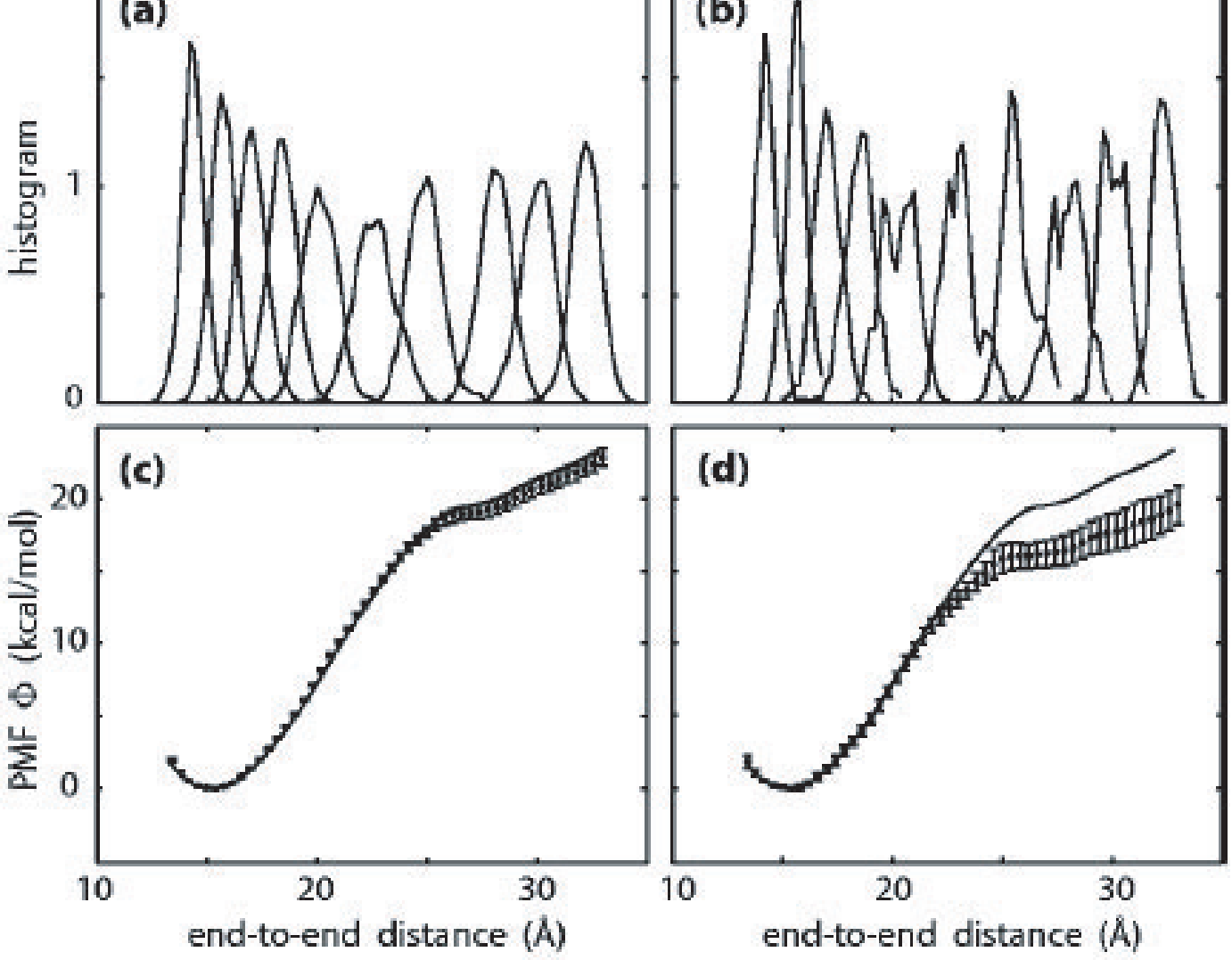


FIG. 8. PMF calculated from umbrella sampling simulations. (a) and (c): 2 ns simulation for each histogram; ten histograms for each block; ten blocks in total. (b) and (d): 0.2 ns simulation for each histogram; ten histograms for each block; ten blocks in total. (a) and (b) show histograms in one block out of the ten blocks. In (c) and (d), the error bars indicate the standard deviation over the blocks, and the exact PMF is plotted as a solid line. The minimum at $\xi = 15.2$ Å was chosen as a reference point for calculating block averages.

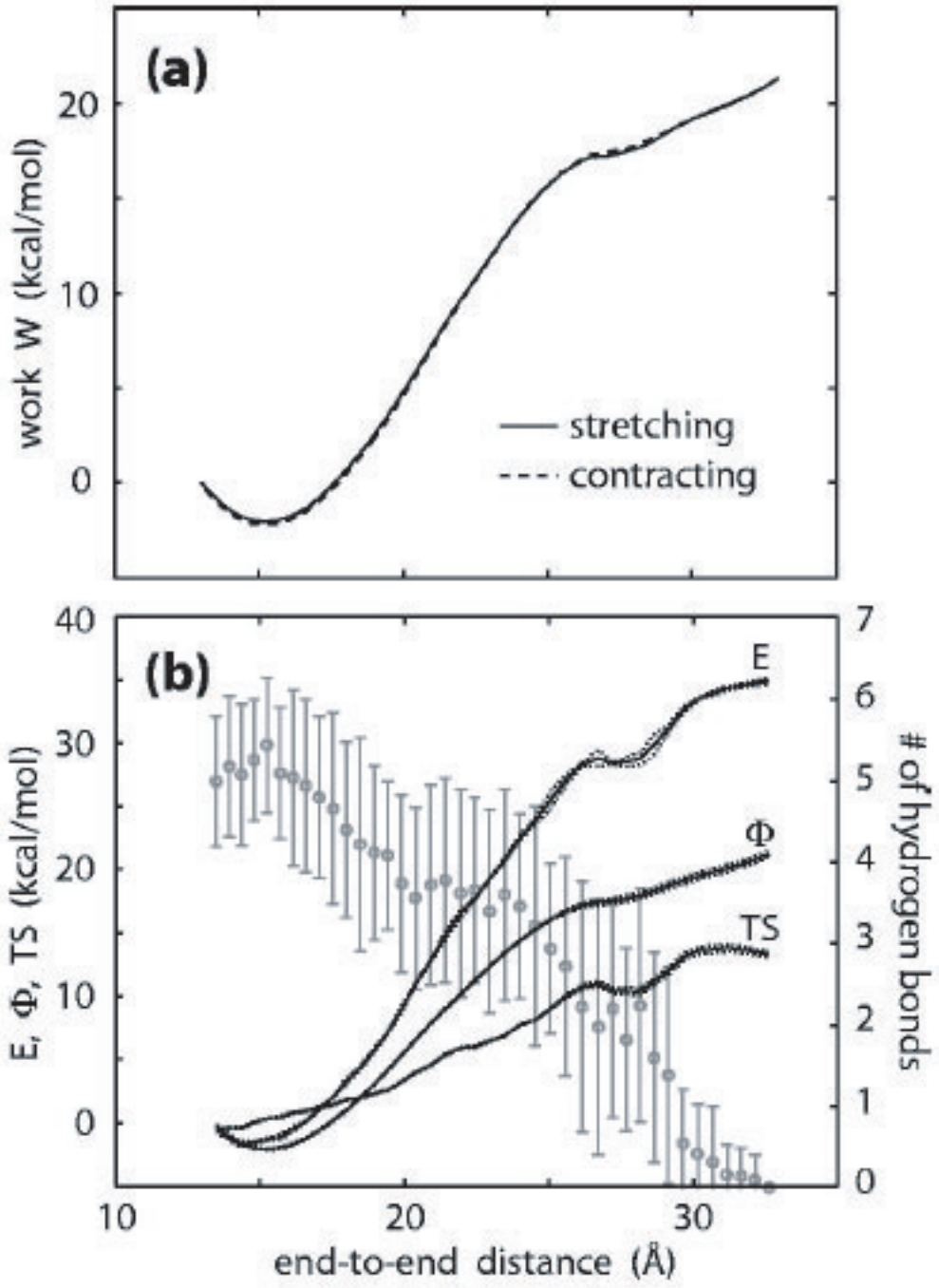


FIG. 3. Reversible pulling ($|v|=0.1$ Å/ns). (a) Work done by forward pulling (stretching) and backward pulling (contracting). For the forward pulling, the position of the constraint center λ is varied from 13 to 33 Å; for the backward pulling, from 33 to 13 Å. For the sake of comparison, the backward-pulling work curve has been shifted vertically so that it coincides with the forward-pulling work curve at $\lambda=33$ Å. (b) Energy E, PMF Φ, and entropy S calculated from four forward pullings. The error bars are shown as dotted lines. Also shown is the number of hydrogen bonds (averaged over time windows) plotted against the end-to-end distance (circles with error bars). A minimum heteroatomic distance of 3.5 Å (between N and O) and a minimum bond angle of 140° (N–H···O) were used for defining a hydrogen bond.

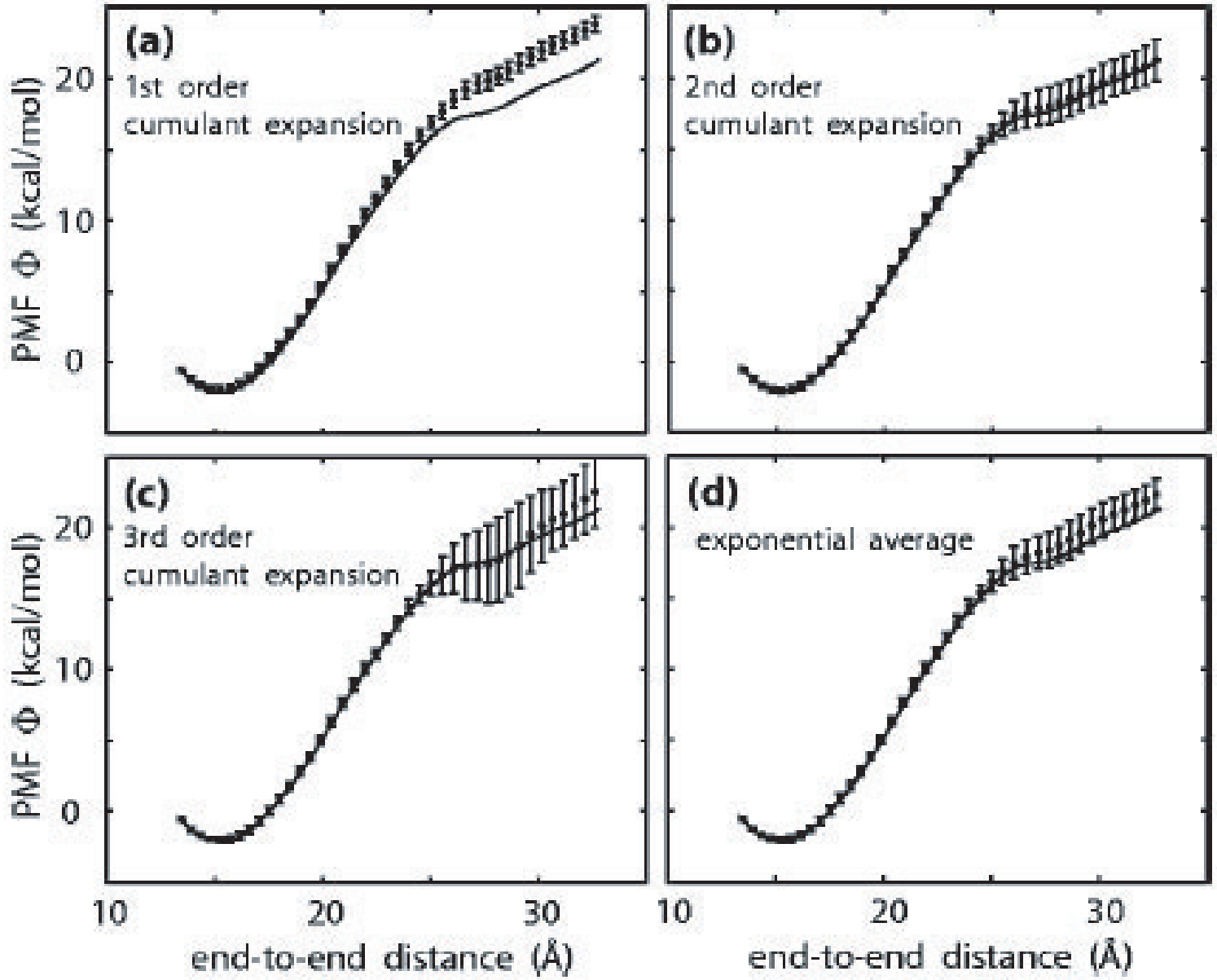


FIG. 4. PMF calculated from irreversible pulling ($v = 10$ Å/ns) through the block average of ten blocks of ten trajectories. The error bars indicate the standard deviation over the blocks. The exact PMF calculated from the reversible pulling is plotted as a solid line in each panel.

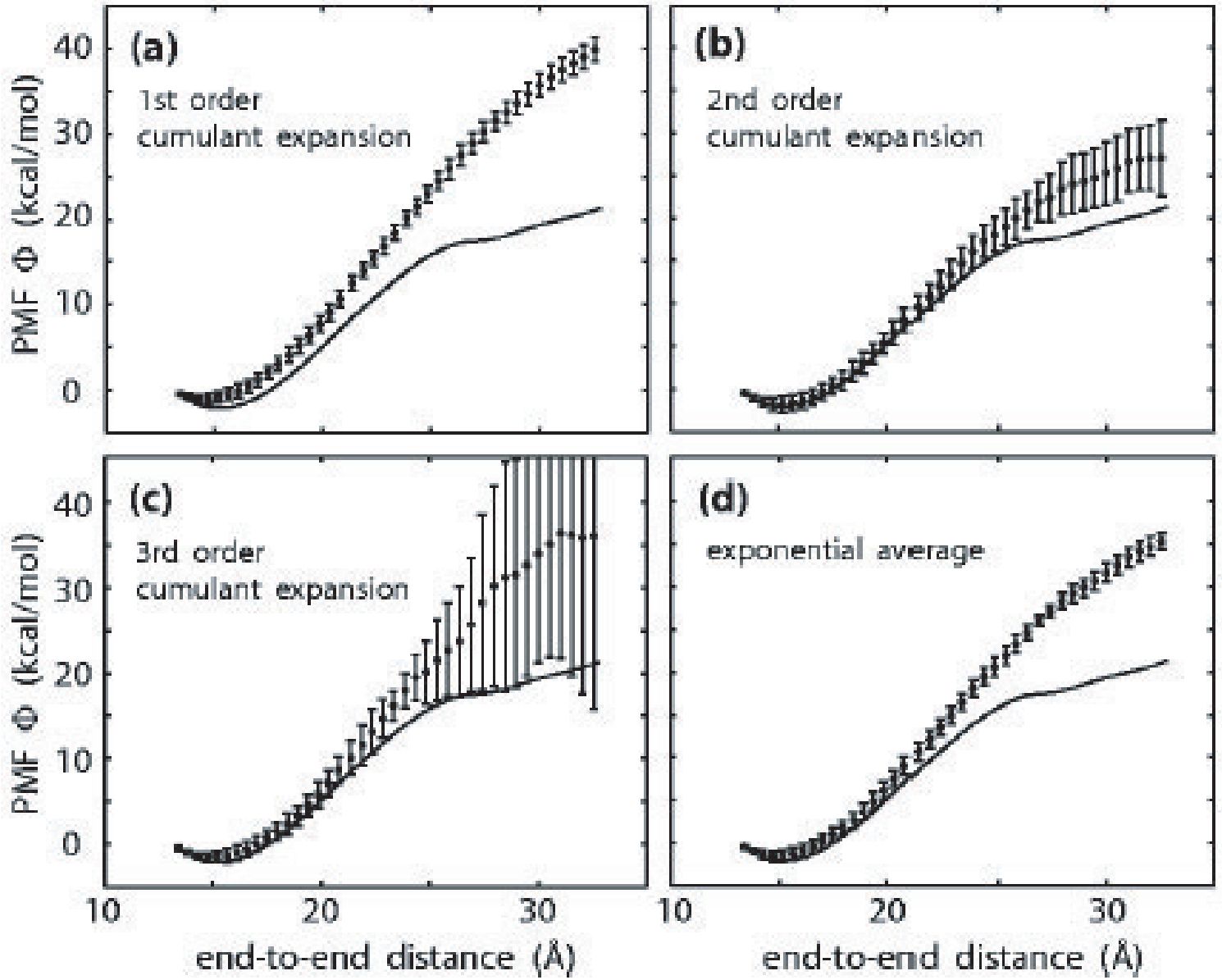


FIG. 5. PMF calculated from irreversible pulling ($v = 100$ Å/ns) through the block average of ten blocks of ten trajectories. The error bars indicate the standard deviation over the blocks. The exact PMF calculated from the reversible pulling is plotted as a solid line in each panel.

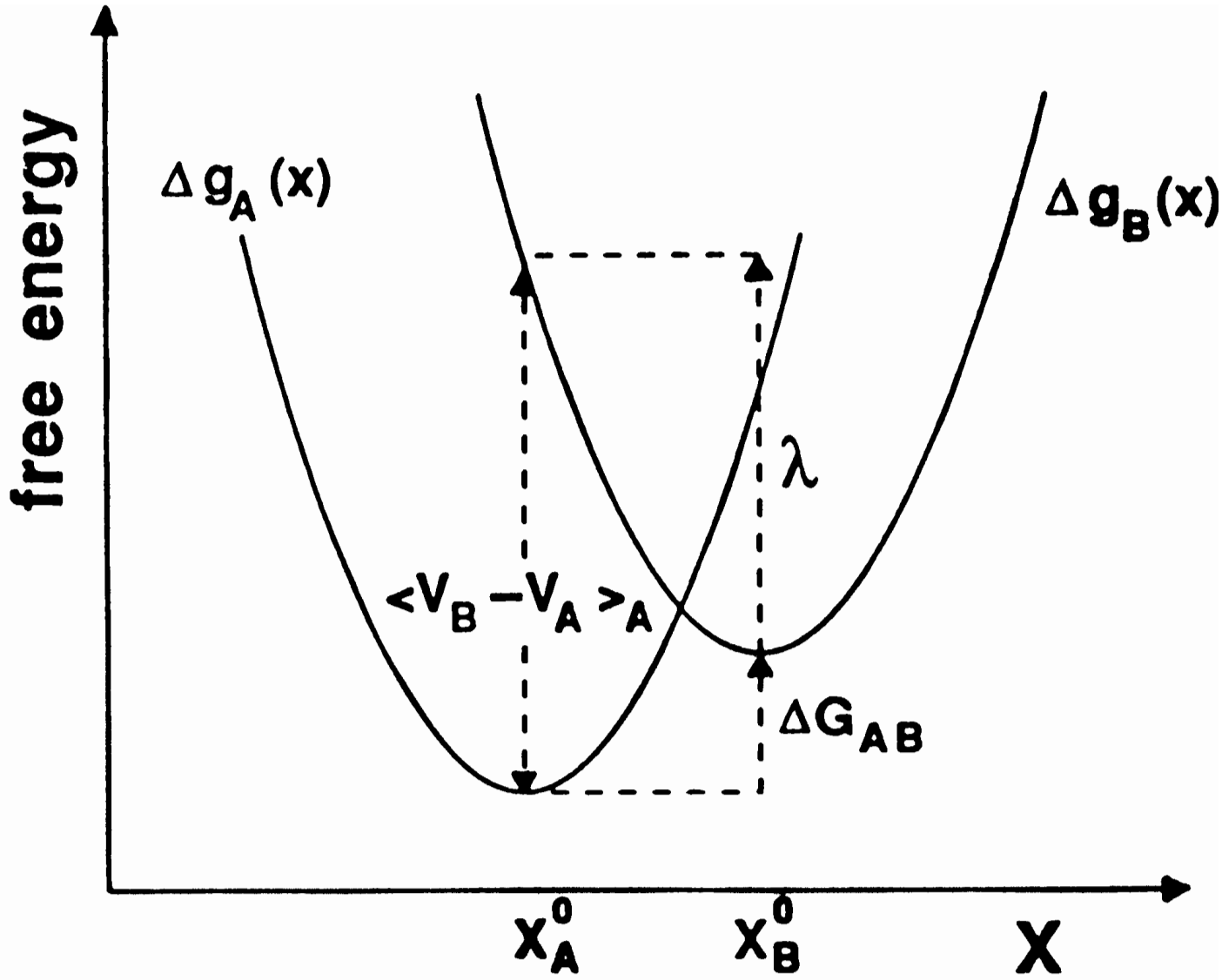


Fig. 5. Schematic free energy curves for a system obeying the linear response approximation.

$$\Delta G = -\beta^{-1} \ln \langle e^{-\beta(V_B - V_A)} \rangle_A$$

Expansion of the exponent and logarithm then yields

$$\begin{aligned}
\Delta G &= -\beta^{-1} \ln \langle 1 - \beta V_B - V_A) + \left(\frac{\beta^2}{2}\right) \\
&\qquad (V_B - V_A)^2 - \ldots \rangle_A \\
&= -\beta^{-1} \ln \{1 - \beta \langle (V_B - V_A) \rangle_A + \left(\frac{\beta^2}{2}\right) \\
&\qquad \langle (V_B - V_A)^2 \rangle_A - \ldots \} \\
&= -\beta^{-1} \{-\beta \langle V_B - V_A \rangle_A + \left(\frac{\beta^2}{2}\right) \\
&\qquad \langle (V_B - V_A)^2 \rangle_A \\
&\quad - \tfrac{1}{2}(-\beta \langle V_B - V_A \rangle_A + \left(\frac{\beta^2}{2}\right) \\
&\qquad \langle (V_B - V_A)^2 \rangle_A)^2 + \ldots \}
\end{aligned}$$

which can be rearranged as

$$\begin{aligned}
\Delta G = \langle \Delta V \rangle_A &- \left(\frac{\beta}{2}\right) \langle (\Delta V - \langle \Delta V \rangle_A)^2 \rangle_A + \\
&\left(\frac{\beta^2}{6}\right) \langle (\Delta V - \langle \Delta V \rangle_A)^3 \rangle_A \\
&- \left(\frac{\beta^3}{24}\right) \{ \langle (\Delta V - \langle \Delta V \rangle_A)^4 \rangle_A - \\
&3 \langle (\Delta V - \langle \Delta V \rangle_A)^2 \rangle_A^2 \} + \ldots
\end{aligned}$$

where $\Delta V = V_B - V_A$. In the same way we obtain by averaging on the potential surface V_B

$$\Delta G = \langle \Delta V \rangle_B + \left(\frac{\beta}{2}\right) \langle (\Delta V - \langle \Delta V \rangle_B)^2 \rangle_B + \ldots$$

Adding the two equations then gives us

$$\begin{aligned}
\Delta G = \tfrac{1}{2} &\{ \langle \Delta V \rangle_A + \langle \Delta V \rangle_B \} \\
&- \left(\frac{\beta}{4}\right) \{ \langle (\Delta V - \langle \Delta V \rangle_A)^2 \rangle_A - \\
&\qquad \langle (\Delta V - \langle \Delta V \rangle_B)^2 \rangle_B \} + \ldots
\end{aligned}$$

The free energy change for a mutation is then given by one-half the change in the electrostatic energy plus the change in the van der Waals energy scaled by an empirical parameter α (eq 1).

$$\Delta G = {}^1/_2\Delta\langle U_{elec}\rangle + \alpha\Delta\langle U_{vdw}\rangle \qquad (1)$$

In their case, the environments for the ligands were isolated in water and bound to the protein in water, and α was evaluated as 0.162 to give the best fit to the experimental binding data. As pursued here, the same procedure can be applied to the fundamental case of computing free energies of hydration. In principle, all that is needed is a simulation for the solute of interest in water to obtain the two energy components.

The factor of ${}^1/_2$ for the Coulombic contribution in eq 1 can be derived in several ways; it is consistent with the Born model of the free energy of hydration for ions (eq 2), which reflects the linear response approximation that polar solutions give a quadratic potential of mean force for a changing electric field.[14-19] In eq 2, Q is the charge on the ion, σ is the solute

$$\Delta G_{hyd} = \frac{Q^2}{2\sigma}\left(\frac{1}{\epsilon} - 1\right) \qquad (2)$$

cavity radius (ion radius), and ϵ is the dielectric constant of the solvent.

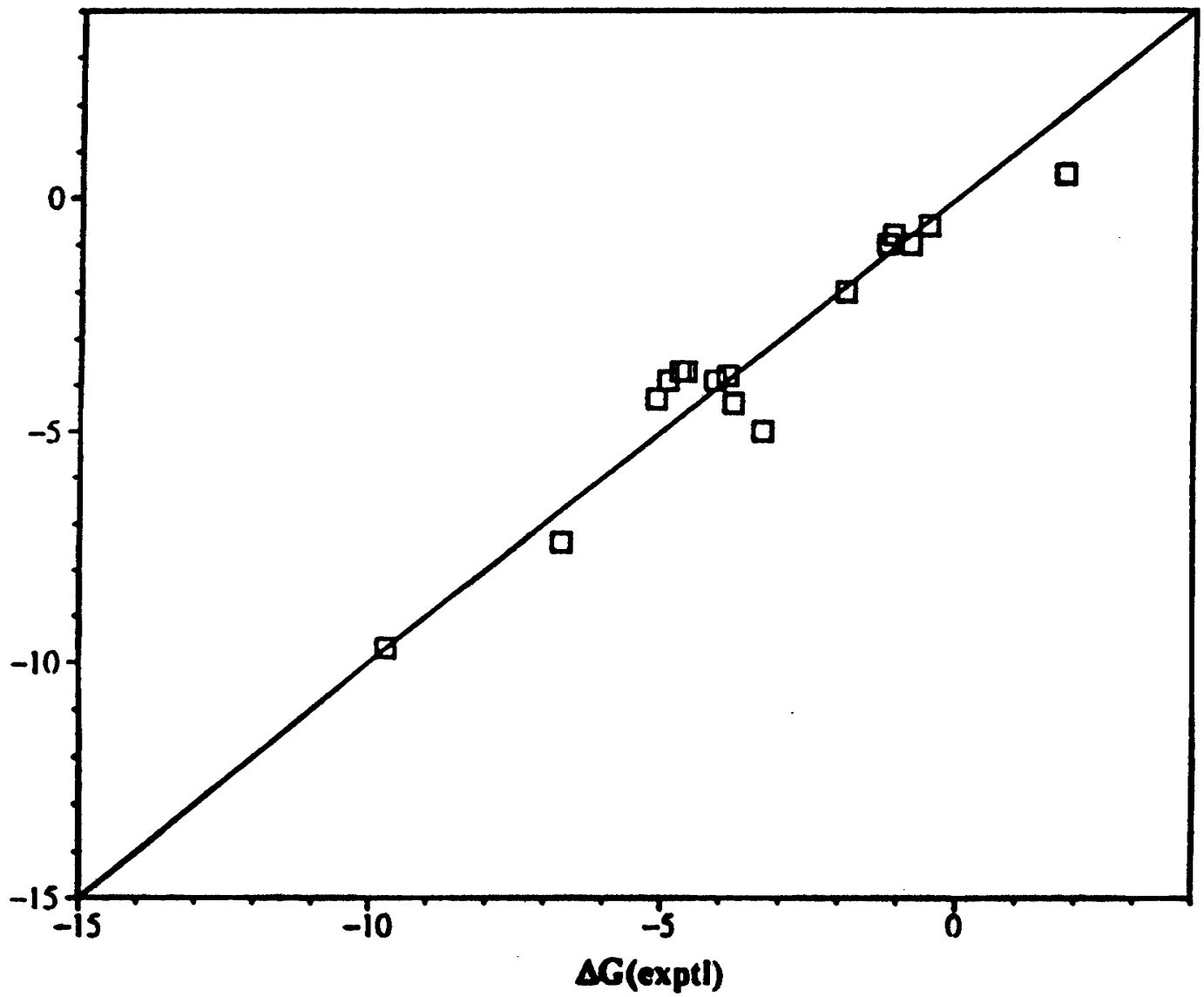


Fig. 5. Correlation between experimental free energies of hydration (kcal/mol) and computed values from MC calculations with the linear response method (Eq. 9).

TABLE 3

FREE ENERGIES OF HYDRATION (kcal/mol) FOR DIVERSE ORGANIC MOLECULES AT 25 °C

Solute	OPLS/UA[a]	EPS/OPLS[b]	EPS/LR[c]	Exptl[d]
Ethane	2.3±0.4	3.3±0.5	0.5	1.8
Methylchloride		0.1±0.5	–0.6	–0.5
Benzene	–0.9±0.4	–0.4±0.4	–1.0	–0.8
Chlorobenzene		–0.5±0.4	–0.8	–1.1
Methanethiol	–1.5±0.4	0.0±0.5	–1.0	–1.2
Dimethylether	–0.7±0.4	–1.4±0.5	–2.0	–1.9
Methylacetate	–0.5±0.6	–5.3±0.5	–5.0	–3.3
Acetone	–1.6±0.6	–3.6±0.5	–4.4	–3.8
Acetonitrile	–2.8±0.5	–4.7±0.5	–3.8	–3.9
Nitrobenzene		–3.4±0.5	–3.9	–4.1
Methylamine	–3.3±0.4	–4.3±0.5	–3.7	–4.6
Pyridine		–4.9±0.4	–3.7	–4.7
Aniline	–6.2±0.4	–5.1±0.4	–3.9	–4.9
Methanol	–4.6±0.3	–4.6±0.4	–4.3	–5.1
Phenol	–6.1±0.4			–6.6
Acetic acid	–5.5±0.5	–8.5±0.4	–7.4	–6.7
Acetamide	–9.5±0.4	–13.4±0.4	–9.7	–9.7
Average error	1.0	1.0	0.6	

[a] Refs. 11, 12, 16, 24–27, 30 and 31. All results are for transfer of the solute from ideal gas to aqueous solution with 1 M standard states.

[b] Ref. 28.

[c] Ref. 50.

[d] Refs. 47–50.

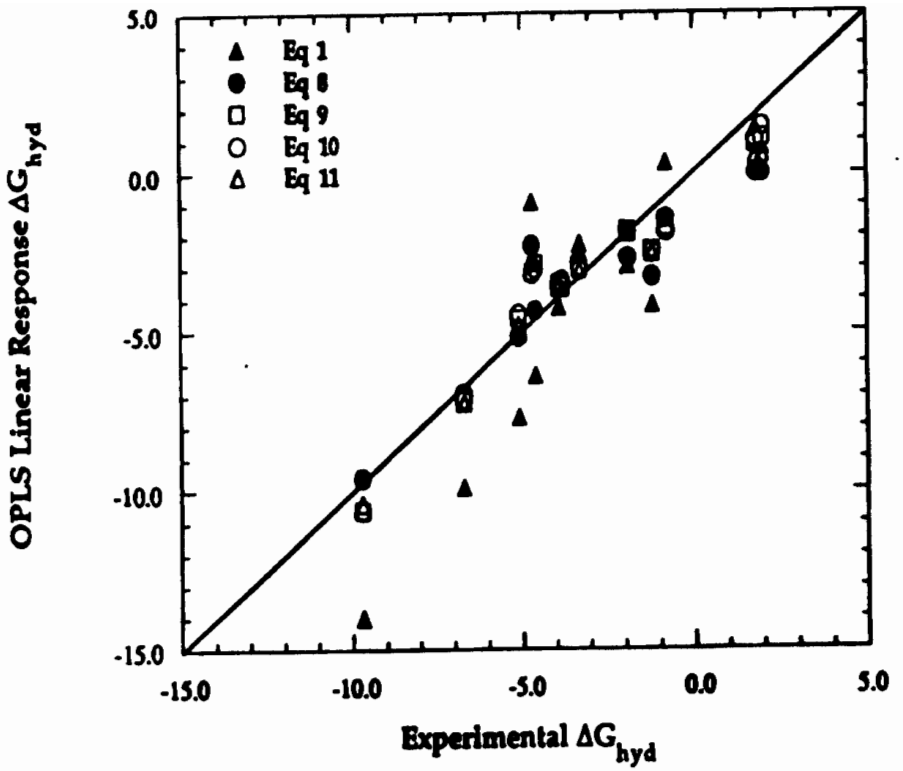


Figure 3. Correlations between the results from the five linear response equations and experimental free energies of hydration (kcal/mol) with the OPLS charges.

TABLE 5: α, β, and γ Parameters Determined by fitting the OPLS-based Results to the Experimental ΔG_{hyd} and $\Delta\Delta G_{hyd}$[a]

linear response method	α	β	γ (kcal/(mol Å²)) or (kcal/(mol Å³))	rms to expt (kcal/mol)
ΔG_{hyd}				
1/2 Coulombic and van der Waals	−0.275	0.5[b]		2.27
	−0.28 ± 0.04			
Coulombic and van der Waals	0.001	0.318		1.21
	0.00 ± 0.03	*0.318* ± 0.007		
Coulombic, van der Waals, and molecular surface area	0.634	0.432	0.059	0.93
	0.6 ± 0.1	*0.43* ± 0.02	*0.06* ± 0.01	
Coulombic, van der Waals, and solvent-accessible surface area	0.489	0.421	0.020	0.88
	0.49 ± 0.08	*0.42* ± 0.02	*0.020* ± 0.003	
Coulombic, van der Waals, and molecular volume	0.683	0.415	0.077	1.03
	0.6 ± 0.2	*0.42* ± 0.03	*0.08* ± 0.02	

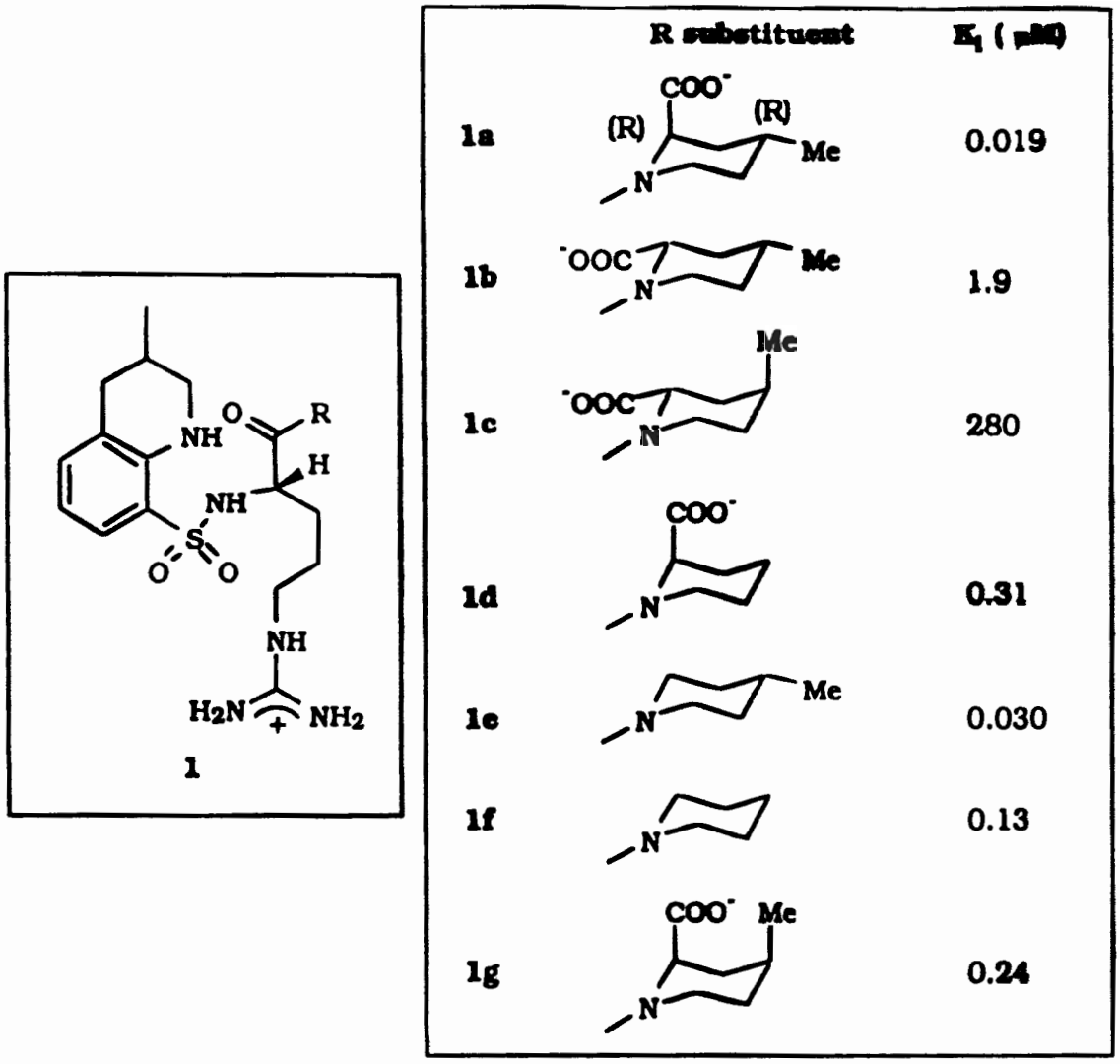


Figure 1. Structures and experimental inhibition constants (ref 7h) for sulfonamide inhibitors of human thrombin.

$$\Delta G_b = \beta\langle\Delta U_{elec}\rangle + \alpha\langle\Delta U_{vdw}\rangle + \gamma\langle\Delta SASA\rangle$$

Table 6. Comparison of Experimental and Calculated ΔG_b (kcal/mol) after Refitting the Scaling Factors[a]

		ΔG_b				
compd	exptl	$\beta = 0.5$, $\alpha = 0.734$, $\gamma = 0.0$ (eq 1)	$\beta = 0.165$, $\alpha = 0.476$, $\gamma = 0.0$ (eq 1)	$\beta = 0.075$, $\alpha = -0.462$, $\gamma = 0.038$ (eq 2)	$\beta = 0.146$, ($\alpha = 0.236$), $\gamma = 0.010$ (eq 2)	$\beta = \alpha = 0.131$, $\gamma = 0.014$ (eq 2)
1a	−11.0	−8.2	−8.6	−9.9	−9.0	−9.1
1b	−8.1	−6.1	−8.6	−7.2	−8.3	−8.1
1c	−5.0	0.9	−6.3	−5.9	−6.3	−6.3
1d	−9.2	−14.0	−10.1	−10.2	−10.3	−10.1
1e	−10.7	−17.7	−11.6	−11.0	−11.6	−11.3
1f	−9.8	−12.8	−10.0	−9.5	−10.0	−9.8
1g	−9.4	−6.6	−7.6	−9.7	−8.2	−8.4
rms		4.40	1.34	0.77	1.15	1.02

[a] Using results from the longest Monte Carlo runs in each case.

$$\Delta G_{\text{calcd}} = -0.94\langle \Delta HB_{\text{total}} \rangle + 0.30\langle EXX\text{-}LJ \rangle + 0.0085 \langle \Delta PHOB_{\text{area}} \rangle - 2.8(\text{secondary amide}) + 4.6 \quad (5)$$

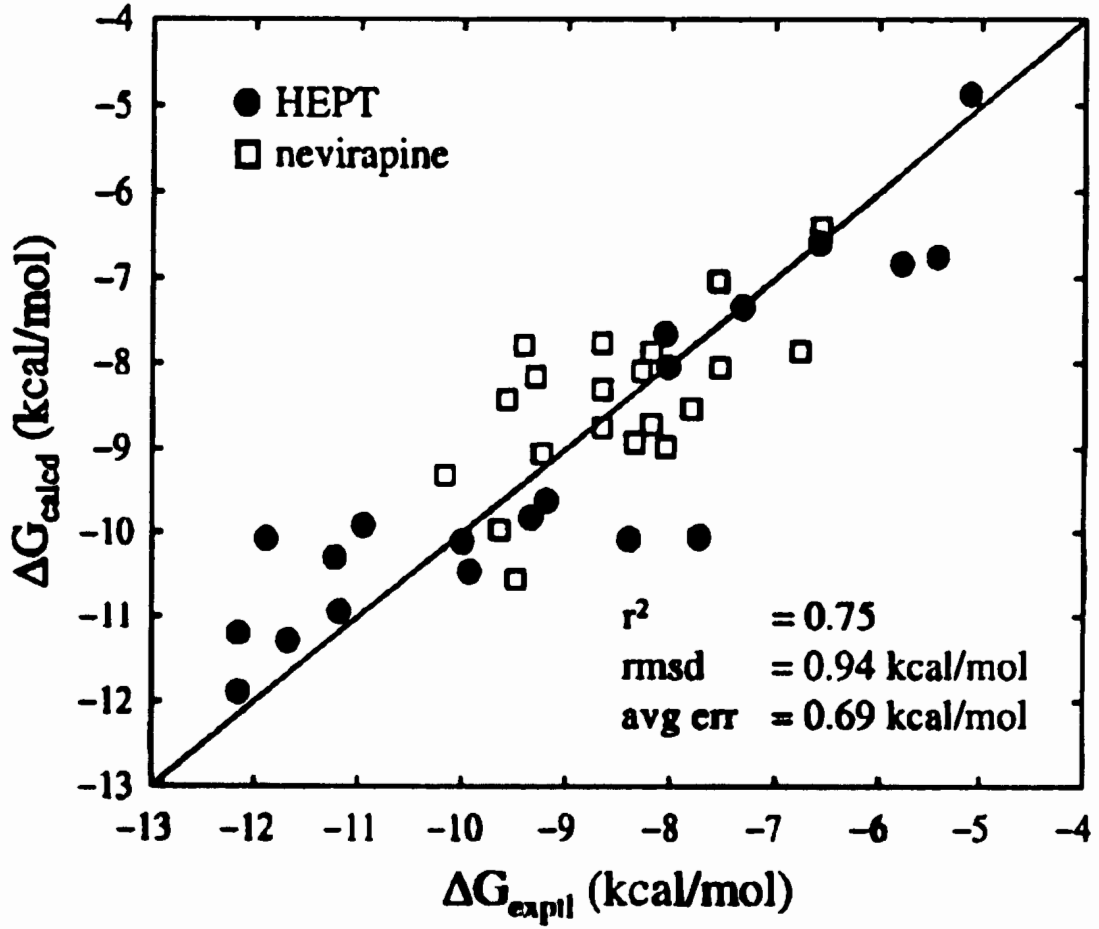


Figure 7. Predicted binding affinities (ΔG_{calcd}) using eq 5 vs experimental activities (ΔG_{exptl}) for 20 HEPT (●) and 20 nevirapine (□) analogues with HIVRT.