

A novel approach to the calculation of the free energy due to molecular flexibility

Mihaly Mezei

*Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University,
New York, NY 10029, USA*

Received 7 January 2000

Dedicated to the memory of Arpad Mezei

The contribution of the molecular flexibility to the solvation excess free energy is expressed in terms of probabilities of reaching hard limits on intramolecular coordinates in a series of calculations successively relaxing those limits. Numerical tests on the harmonic oscillator are also presented and used to make suggestion about computational issues.

KEY WORDS: free energy, molecular flexibility, Monte Carlo computer simulation, thermodynamic integration

1. Introduction

The calculation of the solvation free energy difference between different molecules or different conformational states of a molecule in general requires a series of simulations over a path in the configuration space connecting these two molecules or conformations [1–3]. Such calculations have to deal with a large number of degrees of freedom, typically of the order of 10^4 – 10^5 .

These degrees of freedom can be separated into two classes: intermolecular and intramolecular. For a solute in a continuum solvent model the flexibility is the only source of nontrivial contribution to the free energy, thus any procedure that addresses the free energy of flexibility already addresses the solvation free energy as well. For systems with explicit solvent models, on the other hand, the flexibility provides only a relatively small (but by no means insignificant) contribution to the free energy as the contribution of solute-solvent terms generally dominate in such cases.

While it is possible, and indeed it is customary, to perform the calculation of solvation free energies with explicit models including simultaneously the intermolecular and intramolecular degrees of freedom (i.e., including the molecular flexibility in all the intermediate states required by the calculation), it is clear that the inclusion of flexibility to all the intermediate states requires significant additional effort that can be saved if there

is a way to efficiently calculate the contribution of flexibility at the endpoints only and deal with rigid solutes in-between.

To belabor this point, note that for a large solute in a simple solvent like water, the sampling of the intermolecular degrees of freedom is significantly easier (although by no means easy) due to the interchangeability of the solvent molecules. Sampling of the intramolecular degrees of freedom may appear to be easy as long as the molecule retains a conformation near a local minimum. However, there is no way of telling *a priori* how many other local minima are there that are to be sampled. Thus modeling a flexible system not only faces the problem of having to sample potentially a great number of local minima, but it is relatively easy to obtain a result that appear ‘converged’ when it is actually quite far from it. In addition, it has been argued as well as demonstrated that a particular path, called polynomial path, that creates and annihilates rigid solutes simultaneously in a particular manner lends itself very efficiently to thermodynamic integration [4] – typically using only five quadrature points (independent of the size of the molecules in question) [5–8]. It follows that the development of an efficient procedure for the calculation of the free energy of flexibility is of great interest and potential use.

The definition of the free energy of flexibility requires some care, though. It is not simply the free energy difference between frozen and flexible states of the molecule since that difference is infinite – a consequence of the fact that when the underlying ideal gas system is undergoing the freezing process, the entropy change corresponding to the degrees of freedom just frozen is $kT \ln(V_1/V_0)$, and the volume of the frozen degrees of freedoms is zero. However, the *excess* free energy of the freezing process is finite and so are *differences* between the free energies of freezing around different conformations.

The purpose of this paper is to present a new method that calculates the excess solvation free energy difference between rigid and flexible molecules. This method called the *scaled constraint method*. Once fully developed, combined with the methods based on rigid solutes as referred to above, it is expected to provide a significantly improved procedure (both in terms of efficiency and reliability) for the determination of solvation free energy differences of large solutes, both for system using continuum solvent and explicit solvent.

2. The scaled constraint method

The calculation of the free energy due to flexibility requires a path connecting the rigid and flexible states. Formally this is done by the introduction of the coupling parameter λ that – through its change from 0 to 1 – transforms the system from its initial to its final state. There are two obvious approaches to the problem. The first approach introduces λ by dividing with it all potential parameters that represent the strength of the intramolecular constraints. Thus at $\lambda = 1$ one obtains the flexible system while at $\lambda = 0$ the strengths of all constraints are scaled up to infinity, effectively yielding a system that can not leave its equilibrium conformation (at finite temperatures). The second approach considers the limits on all internal coordinates that define the allowed region

of the conformation space for the particular conformation in question and introduces λ as a multiplier of these limits – hence the name ‘scaled constraint method’. This means that at $\lambda = 0$ all intramolecular degrees of freedom are limited to a single value while at $\lambda = 1$ the flexible system is recovered.

The present study proposes the use of the second approach. It has a conceptual and a technical advantage over the first one: it allows a convenient way to define the allowed region around a conformation and it also provides a way to eliminate the singularity referred to in the introduction. Since the source of the singularity is the underlying ideal gas, it appears plausible that it would be eliminated if one applies the coupling parameter to calculate the excess free energy of flexibility (i.e., the difference between the system under consideration and the underlying ideal gas) and it will be shown below that this is indeed the case.

2.1. The scaled constraint method in one dimension

Let us consider a system with a single internal degree of freedom x limited to a finite interval $[0, X]$ with $x = 0$ being the rigid state and $e(x)$ is the intramolecular Hamiltonian, $e(0) = 0$. First, we will develop the a thermodynamic integration (TI) formalism for the free energy difference ΔA between the rigid and non-rigid systems. The coupling parameter is introduced as a scaling factor to the limit X of the interval where x is allowed to move:

$$A(\lambda) = -kT \ln \int_0^{\lambda X} \exp[-e(x)/kT] dx. \quad (1)$$

Writing ΔA as the integral of its derivative with respect to λ (the standard starting point of the TI methods) and carrying out the requisite derivations (noting that the derivative of an integral with respect to its limit is just the integrand) we obtain

$$\begin{aligned} \Delta A &= \int_0^1 \frac{\partial}{\partial \lambda} \left\{ -kT \ln \int_0^{\lambda X} \exp[-e(x)/kT] dx \right\} d\lambda \\ &= -kT \int_0^1 \frac{(\partial/\partial \lambda) \int_0^{\lambda X} \exp[-e(x)/kT] dx}{\int_0^{\lambda X} \exp[-e(x)/kT] dx} d\lambda \\ &= -kT \int_0^1 \frac{X \exp[-e(\lambda X)/kT]}{\int_0^{\lambda X} \exp[-e(x)/kT] dx} d\lambda = -kT \int_0^1 X \langle \delta(x - \lambda X) \rangle_\lambda d\lambda. \end{aligned} \quad (2)$$

The last expression of equation (2) is just X times the Boltzmann factor at λX for a system where the variable x is limited to the range $[0, \lambda X]$. It can be estimated from a simulation that generates a Boltzmann-distributed ensemble as the probability $P(\lambda X)$ of $x = \lambda X$:

$$P(\lambda X) = \lim_{\Delta x \rightarrow 0} \frac{P(x \in [\lambda X - \Delta x, \lambda X])}{\Delta x}. \quad (3)$$

A quick analysis, however, shows that the resulting TI integral is divergent (as it should be, since it includes the divergent ideal gas contribution): as λ goes to zero, the interval gets smaller and smaller thus the Boltzmann factor will approach a constant which is just $1/\lambda X$, a divergent integrand.

The advantage of this path lies in the fact that it can be extended in such a way that the inclusion of the change in the underlying ideal gas free energy results in a finite and computationally manageable integrand, as will be shown next.

Consider the TI integrand for the difference between two ΔA 's:

$$\Delta A_{01}(\lambda) = kT X \left\{ \frac{\exp[-e_1(\lambda X)/kT]}{\int_0^{\lambda X} \exp[-e_1(x)/kT] dx} - \frac{\exp[-e_0(\lambda X)/kT]}{\int_0^{\lambda X} \exp[-e_0(x)/kT] dx} \right\}. \quad (4)$$

We will show that in the limit of $\lambda \rightarrow 0$, $\Delta A_{01}(\lambda)$ is finite. Consider the leading terms of e_0 and e_1 :

$$e_0(x) = c_0 x^p \quad \text{and} \quad e_1(x) = c_1 x^q \quad (5)$$

and assume $q \geq p \geq 1$. Expanding the exponentials and bringing the two fractions to a common denominator we obtain

$$\begin{aligned} \Delta A_{01}(\lambda) = kT & \left[\left(1 - \frac{c_1 X^q \lambda^q}{kT} + \dots \right) \left(\lambda X - \frac{c_0 X^p \lambda^{p+1}}{kT(p+1)} + \dots \right) \right. \\ & \left. - \left(1 - \frac{c_0 X^p \lambda^p}{kT} + \dots \right) \left(\lambda X - \frac{c_1 X^q \lambda^{q+1}}{kT(q+1)} + \dots \right) \right] \\ & \times \left[\left(\lambda X - \frac{c_0 X^p \lambda^{p+1}}{kT(p+1)} + \dots \right) \left(\lambda X - \frac{c_1 X^q \lambda^{q+1}}{kT(q+1)} + \dots \right) \right]^{-1}. \quad (6) \end{aligned}$$

The leading term in the denominator will be $X^2 \lambda^2$ (independent of the value of the exponents p, q) while the leading term of the numerator will be of the order of λ^{p+1} (since the leading λX terms of each product in the numerator cancel). Therefore, there will be a finite limit of $\Delta A_{01}(\lambda)$ as $\lambda \rightarrow 0$. Furthermore, it is also clear from equation (6) that the limit remain finite when $c_0 = 0$, i.e., when one of the systems is the ideal gas.

2.2. The scaled constraint method in N dimensions

For systems with multiple (n) degrees of freedom, generalization of equation (1) introduces the coupling parameter λ as a common scale factor of the limits of integration:

$$A(\lambda) = -kT \ln \int_{\lambda X'_1}^{\lambda X_1} \dots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(x)/kT] dx d\lambda \quad (7)$$

thus ΔA can be obtained as

$$\Delta A = \int_0^1 \frac{I_\lambda}{\int_{\lambda X'_1}^{\lambda X_1} \dots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(x)/kT] dx} d\lambda, \quad (8)$$

where

$$I_\lambda = \frac{\partial}{\partial \lambda} \int_{\lambda X'_1}^{\lambda X_1} \cdots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(x)/kT] dx. \quad (9)$$

Applying the standard limit procedure to evaluate I_λ we obtain

$$I_\lambda = \lim_{\Delta\lambda \rightarrow 0} \frac{1}{\Delta\lambda} \left(\int_{(\lambda+\Delta\lambda)X'_1}^{(\lambda+\Delta\lambda)X_1} \cdots \int_{(\lambda+\Delta\lambda)X'_n}^{(\lambda+\Delta\lambda)X_n} \exp[-e(x)/kT] dx \right. \\ \left. - \int_{\lambda X'_1}^{\lambda X_1} \cdots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(x)/kT] dx \right). \quad (10)$$

Breaking up each integral to integrals over the intervals $[(\lambda + \Delta\lambda)X', \lambda X']$, $[\lambda X', \lambda X]$ and $[\lambda X, (\lambda + \Delta\lambda)X]$ the first term in the numerator will break up into 3^n terms. The (only) integral without any $\Delta\lambda$ in its upper limit will be canceled by the original second term in the numerator. Assuming e to be continuous, use of the mean value theorem allows the approximation of all of the remaining terms in such a form that each term will have a factor $(\Delta\lambda)^k$, $1 \leq k \leq n$, multiplying finite integrals. Since each such $\Delta\lambda$ factor comes from an integration over one of the intervals $[(\lambda + \Delta\lambda)X', \lambda X']$ or $[\lambda X, (\lambda + \Delta\lambda)X]$, any integral that has more than two variables integrated over this interval will be zero in the limit since the denominator contains $\Delta\lambda$ raised to the first power only. Thus the derivative with respect to λ will be

$$I_\lambda = X'_1 \int_{\lambda X'_2}^{\lambda X_2} \cdots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(X'_1, x_2, \dots, x_n)/kT] dx_2 \dots dx_n \\ + X_1 \int_{\lambda X'_2}^{\lambda X_2} \cdots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(X_1, x_2, \dots, x_n)/kT] dx_2 \dots dx_n \\ + \cdots \\ + X'_n \int_{\lambda X'_2}^{\lambda X_2} \cdots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(x_1, \dots, x_{n-1}, X'_n)/kT] dx_1 \dots dx_{n-1} \\ + X_n \int_{\lambda X'_2}^{\lambda X_2} \cdots \int_{\lambda X'_n}^{\lambda X_n} \exp[-e(x_1, \dots, x_{n-1}, X_n)/kT] dx_1 \dots dx_{n-1} \quad (11)$$

leading to

$$\Delta A = kT \int_0^1 \sum_{i=1}^n X'_i \langle \delta(x - \lambda X'_i) \rangle_\lambda + X_i \langle \delta(x - \lambda X_i) \rangle_\lambda d\lambda. \quad (12)$$

The quantities $\langle \rangle_\lambda$ are again just the Boltzmann probabilities of each variable reaching its lower or upper limits thus we obtain

$$\Delta A = kT \int_0^1 \sum_{i=1}^n X'_i P(\lambda X'_i) + X_i P(\lambda X_i) d\lambda. \quad (13)$$

A meaningful application of hard limits to delineate the region of a conformation requires the use of internal coordinates. Thus, we first have to generalize the method to non-Cartesian systems. Transformation to the internal coordinates $\mathbf{R}$ introduces the Jacobian $J(\mathbf{r})$:

$$A(\lambda) = -kT \ln \int_{\lambda R'_1}^{\lambda R_1} \cdots \int_{\lambda R'_n}^{\lambda R_n} \exp[-e(\mathbf{r})/kT] J(\mathbf{r}) \, d\mathbf{r} \, d\lambda \quad (14)$$

with

$$J(\mathbf{r}) = \det \begin{pmatrix} \frac{\partial x_1}{\partial r_1} & \cdots & \frac{\partial x_1}{\partial r_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial r_1} & \cdots & \frac{\partial x_n}{\partial r_n} \end{pmatrix}. \quad (15)$$

The arguments leading from equation (7) to equation (11) apply equally to equation (14). Furthermore, since $J(\mathbf{r}) \, d\mathbf{r}$ is the volume element associated with $d\mathbf{r}$, the transition to probabilities works equally well:

$$\Delta A = kT \int_0^1 \sum_{i=1}^m R'_i P(\lambda R'_i) + R_i P(\lambda R_i). \quad (16)$$

The replacement of the summation's upper limit n with m indicates the fact that when transforming to internal coordinates 6 degrees of freedom can be integrated out (assuming that no external field is present). Their contributions to ΔA cancels when forming ΔA_{01} thus they were omitted from equation (16) as well. Finally, note that while ΔA of equation (16) is a divergent integral,

$$\Delta A_{01} = kT \int_0^1 \sum_{i=1}^m R'_i [P_1(\lambda R'_i) - P_0(\lambda R'_i)] + R_i [P_1(\lambda R_i) - P_0(\lambda R_i)] \quad (17)$$

is not (according to the argument represented by equations (4)–(6)).

2.3. Numerical considerations

The actual application of equation (16) for the calculation of ΔA_{01} requires a numerical integration where at each quadrature point the integrand at λ has to be estimated from computer simulations. The simulations are to be performed with the hard limits on the degrees of freedoms that are being sampled. The actual limiting values are determined by the value of the integration variable λ (as seen, e.g., in equation (7)). Each simulation will have to estimate the probability density of for each degree of freedom at the limiting value defined by λ .

There are several questions that emerge when equation (16) is to be used for the calculation of ΔA_{01} : (i) the most efficient way of evaluating the integrand; (ii) the choice of $\mathbf{R}$; (iii) the most efficient quadrature.

2.3.1. The evaluation of the integrand

Equation (16) appears ‘benign’ computationally since it requires only the Boltzmann probabilities of each internal degrees of freedom at their limiting values, averaged over the remaining degrees of freedom. The averaging over all other degrees of freedom means that the increase the number of variables will only proportionally increase the computational expenses.

The evaluation of the integrand requires the Boltzmann probability that the variable r is at the limit λR . Such probabilities are usually estimated using statistics over finite intervals (grids). There are conflicting requirements for the size of these grids. First, larger grids (i.e., wider intervals) provide the better statistics. Large gridsize is the source of two possible errors, both related to the nonlinearity of $P(r)$. First, the larger the grid size, the farther one has to extrapolate from the center of the grid to the endpoint of the grid. Second, the nonlinearity of $P(r)$ implies that the average of $P(r)$ over the grid differs from the value of $P(r)$ at the midpoint.

Estimation of $P(R)$ as defined by equation (3) involves gathering statistics of the frequency of sampling of each grid i (representing the interval $[r_i, r_i + \Delta]$, with $r_1 = R$) yielding P_i :

$$P_i = \int_{r_i}^{r_i + \Delta} P(r) dr / \Delta. \quad (18)$$

For Δ small enough the approximation

$$P_i = P(r_i + \Delta/2) \quad (19)$$

is reasonable.

To obtain $P(R)$ an extrapolation procedure is required. This requires that first we fit a function to the data gathered. The following fitting procedures were tested (the formulae used are described in the appendix).

1. Fit a polynomial of order o to P_i .
2. Fit a quadratic to P_1, P_2, P_3 .
3. Fit an exponential of the form $a \exp(-br)$ P_1, P_2 .
4. Fit the integral of a quadratic to P_1, P_2, P_3 .
5. Fit the integral of a polynomial of order o to P_i .

Each fitting was followed by an extrapolation of the fitted function to R . The first two procedures used the approximation of equation (19). The fitting of the exponential turns out to give identical result irrespective of using equation (19).

2.3.2. The choice of hard limits R

For the degrees of freedoms that represent vibrations around a single point governed by a harmonic potential (i.e., bond lengths and bond angles) the choice of the hard limit R is a matter of computational convenience. Any value R larger than the largest

value sampled $R_{\max}$ should provide the correct answer. The only computational consideration is the observation that the more R exceeds $R_{\max}$, the more the integration will sample λ regions that contribute little to the integral and as a result it is likely that more quadrature points will be needed to achieve a given level of accuracy.

The choice of the hard limit for torsions is a more complex issue. Since in general the full range of a torsion angle is accessible energetically, one obvious choice for such R , R' is $\pm 180^\circ$. However, the torsional potentials generally have multiple minima and the barrier between them makes crossing infrequent and thus thorough sampling very difficult. For continuum models, the mixed MC–MD algorithm [9] provides an efficient way for barrier crossing thus for the gradual exploration of the torsional space (as provided by the increase of the λ parameter).

An other aspect of the torsional degree of freedom is that in general different conformations of a molecule are specified by the local minima of certain torsion angles (e.g., conformations labeled *trans*, *gauche*, etc.). Free energy simulations aimed at determining conformational free energy differences usually rely on these torsion angles staying in the region of their local minima or driven from one to an other. The scaled constraint method has the advantage of providing a natural way to control the preservation of the conformation ‘type’ by setting the hard limits at the top of the barriers surrounding the minimum.

2.3.3. The choice of quadrature

Earlier work has shown that Gaussian quadratures work well for thermodynamic integrations. It is proposed that the integral using equation (8) be evaluated with such quadrature. The number of quadrature points that are necessary for a given precision should be established by trial and error. Note that, beside the obvious reduction in computing expenses, there is an additional incentive to reduce the number of quadrature points: the fewer points are used, the larger the initial λ value will be, thus the less severe the problem of calculating small difference between large numbers will be.

2.3.4. Optimizing the runlength at each quadrature point

The magnitude of the TI integrand can vary over two orders of magnitude. This means that to obtain a given level of precision, different relative errors are permissible at the various quadrature points. In a general case the run length at each quadrature point can be based on initial estimates using a certain minimum length. In our case, however, the variation of the integrand at quadrature points r_q is well represented by $p_q \propto r_q^{-1}$. It has been shown [7] that if the error at quadrature point q is proportional to p_q then the optimal length of the run is proportional to $\sqrt{c_q} p_q$ (where c_q is the corresponding quadrature coefficient) giving the optimal number of simulation steps as

$$N_q \propto \sqrt{c_q}/r_q. \quad (20)$$

2.3.5. Choice of the simulation methodology

The introduction of the hard limits into the formalism makes Monte Carlo a natural choice since such limits are very easy to implement into the Metropolis procedure. The

test described below on the harmonic oscillator have all used the standard Metropolis algorithm.

Monte Carlo methods, however, have been found so far to scale rather poorly and despite of serious efforts and significant progress molecular dynamics remained the method of choice for modeling intermolecular degrees of freedom of macromolecules. Thus, unless the calculations are restricted to systems that can be simulated efficiently with Monte Carlo methods, the molecular dynamics method has to be generalized to incorporate hard limits on internal coordinates.

3. Numerical examples

3.1. Description of the system

Tests were run in this study on the harmonic oscillator where

$$e(r) = cr^2 \quad (21)$$

leading to

$$\begin{aligned} \Delta A_{01} &= -kT \left[\ln \int_{-R}^R \exp(-c_1 r^2 / kT) dr - \ln \int_{-R}^R \exp(-c_0 r^2 / kT) dr \right] \\ &= -kT \left[\ln 2 \int_0^R \exp(-c_1 r^2 / kT) dr - \ln 2 \int_0^R \exp(-c_0 r^2 / kT) dr \right] \\ &= -kT \left[\ln \frac{\operatorname{erfu}(R\sqrt{c_1/kT})}{\sqrt{c_1/kT}} - \ln \frac{\operatorname{erfu}(R\sqrt{c_0/kT})}{\sqrt{c_0/kT}} \right], \end{aligned} \quad (22)$$

where erfu is the unnormalized error function:

$$\operatorname{erfu}(r) = \int_0^r \exp(-u^2) du. \quad (23)$$

When the reference system is the ideal gas (i.e., $c_0 = 0$) instead of equation (22) we have

$$\Delta A_{01} = -kT \left[\ln \int_{-R}^R \exp(-c_1 r^2 / kT) dr - \ln(2R) \right]. \quad (24)$$

The TI integral (equation (2)) will become

$$\Delta A_{01} = -2kT \int_0^1 \frac{-R \exp[-c(R\lambda)^2 / kT]}{\operatorname{erfu}(R\sqrt{c/kT}) / \sqrt{c/kT}} d\lambda. \quad (25)$$

Having an analytical expression for both the integrand and the integral will allow us to test the errors in approximating the integrand and the quadrature error separately.

Table 1
Quadrature errors.

Method	R	$N_q = 3$	$N_q = 5$	$N_q = 8$
Trapezoid	2	0.0580	0.0191	0.0073
Simpson	2	0.0395	0.0122	0.0022
Gaussian	2	0.0426	0.0003	0.0000
Trapezoid	3	0.1931	0.0434	0.0163
Simpson	3	0.1566	0.0381	0.0056
Gaussian	3	0.0177	0.0098	0.0003
Trapezoid	4	0.3197	0.0992	0.0295
Simpson	4	0.4441	0.0090	0.0260
Gaussian	4	0.1812	0.0212	0.0021

Legend: R is the limit used in equation (22); N_q is the number of quadrature points used.

3.2. Comparison of quadratures and hard limit choices

Comparisons were made using the trapezoid rule, the Simpson integration and the Gaussian quadrature of orders 3, 5 and 8. These comparisons used the exact expression for the integrand, equation (25).

The calculations were performed for $c_0 = 1$, $c_1 = 5$ and $R = 2, 3$ and 4. The reference values computed using equation (25) are 0.8000, 0.8047 and 0.80047, respectively. Table 1 shows the quadrature errors using 3, 5 and 8 quadrature points. It is clear that the Gaussian quadrature is the best and the trapezoid rule produces the largest quadrature error. It can also be seen that the extension of the limit R to regions that are sampled with negligible frequency or not at all increases significantly the quadrature error with a given number of quadrature points, irrespective of the integration method used.

Table 1 also suggests that even a 3-point Gaussian quadrature could be used if a few percent error in the calculated ΔA_{01} is acceptable.

The calculated free energy seems to be relatively insensitive to the value of the hard limit, as long as the Boltzmann probability at the hard limit is small. In this example, $P(X)$ was 0.022, 0.000 and 0.000 for $R = 2, 3$ and 4, respectively. This shows that setting the hard limit at the 2% probability level results only in about half percent error in the free energy – comparable to the precision of the Monte Carlo procedure described here.

3.3. Comparison of extrapolation procedures

All five fitting procedures described in section 2.3.1 have been tested on the harmonic oscillator. The tests compared the exact values of the integrand at the five quadrature points with the values extrapolated with the different procedures. For these tests, $c_0 = 0$ was chosen to avoid possible compensating errors from the two extrapolations that would have been needed otherwise. In this case, the ratio $R/\sqrt{c_1}$ already characterizes the system so c_0 was kept fixed at 2.0 and only R was varied.

Table 2
Comparison of fitting polynomial orders.

R	c	N_g	o	$\Delta I(1)$	$\Delta I(2)$	$\Delta I(3)$	$\Delta I(4)$	$\Delta I(5)$
2.00	2.00	8	2	0.001	0.058	0.235	0.355	0.175
2.00	2.00	8	3	0.000	0.007	-0.067	-0.387	0.536
2.00	2.00	10	2	0.001	0.052	0.290	0.303	0.132
2.00	2.00	10	3	0.000	0.005	-0.055	-0.309	-0.432
2.00	2.00	10	4	0.001	-0.002	-0.028	0.050	0.249
2.00	2.00	12	2	0.000	0.049	0.270	0.271	0.102
2.00	2.00	12	3	0.000	0.005	-0.049	-0.271	-0.377
2.00	2.00	12	4	0.001	-0.001	-0.022	0.036	0.185
2.00	2.00	12	5	-0.011	-0.001	0.009	0.055	0.025
2.00	2.00	20	2	0.000	0.044	0.235	0.212	0.043
2.00	2.00	20	3	0.000	0.004	-0.041	-0.216	-0.293
2.00	2.00	20	4	-0.001	-0.001	-0.014	0.025	0.123
2.00	2.00	20	5	-0.010	-0.001	0.006	0.035	0.023
2.00	2.00	20	6	2.427	0.051	-0.571	1.004	3.686
2.00	2.00	20	7	0.040	0.002	0.000	0.034	0.119
2.00	2.00	20	8	0.009	-0.002	0.003	0.004	0.003
2.00	2.00	20	9	-0.046	-0.013	0.027	0.060	0.227
1.50	2.00	8	2	0.000	0.019	0.144	0.284	0.289
1.50	2.00	8	3	0.000	0.003	-0.002	-0.107	0.239
1.50	2.00	10	2	0.000	0.018	0.130	0.252	0.250
1.50	2.00	10	3	0.000	0.002	-0.002	-0.086	-0.191
1.50	2.00	10	4	0.001	0.000	-0.010	-0.022	0.009
1.50	2.00	12	2	0.000	0.017	0.122	0.232	0.226
1.50	2.00	12	3	0.000	0.002	0.002	0.076	-0.168
1.50	2.00	12	4	0.001	0.000	-0.007	-0.018	0.006
1.50	2.00	12	5	-0.008	-0.006	0.002	0.013	0.037
1.50	2.00	20	2	0.000	0.015	0.108	0.199	0.184
1.50	2.00	20	3	0.000	0.002	-0.002	-0.063	-0.135
1.50	2.00	20	4	0.000	0.000	-0.005	-0.011	0.005
1.50	2.00	20	5	0.006	0.000	0.000	0.010	0.022
1.50	2.00	20	6	-3.126	0.130	-0.119	-0.247	0.585
1.50	2.00	20	7	0.043	-0.013	0.002	-0.007	-0.006
1.50	2.00	20	8	0.027	-0.005	0.001	-0.001	-0.001
1.50	2.00	20	9	0.448	-0.052	0.007	-0.026	-0.018

Legend: R is the limit used in equation (22); c is the potential coefficient of equation (21); N_g is the number of grid points used in collecting the sampling frequencies; o is the order of the fitting polynomial; $\Delta I(q)$ is the error of the integrand at the quadrature point q .

The first comparison was aimed at testing the effect of the order o on the polynomial approximations. Table 2 shows the results of the tests using the polynomial integral fits (similar results were obtained using the pointwise fit, except that the anomalously large errors seen with $o = 6$ were not seen there). For the last two quadrature points the fits show large variations with the order and increasing o quite often brings in larger errors. The fits for the first three quadrature points, on the other hand, show the expected

Table 3
Comparison of fitting procedures.

R	c	N_g	q	I_x	$\Delta I(p^2)$	$\Delta I(p_{av}^2)$	$\Delta I(e)$	$\Delta I(p)$	$\Delta I(p_{av}^2)$
2.00	2.00	8	1	-21.068	0.000	0.003	-0.020	0.003	0.000
2.00	2.00	10	1	-21.068	0.000	0.002	-0.012	0.000	0.001
2.00	2.00	12	1	-21.068	0.000	0.001	-0.008	0.000	0.001
2.00	2.00	20	1	-21.068	0.000	0.000	-0.002	0.000	-0.001
2.00	2.00	8	2	-3.237	0.037	0.038	-0.079	0.010	0.007
2.00	2.00	10	2	-3.237	0.018	0.019	-0.047	-0.002	-0.002
2.00	2.00	12	2	-3.237	0.010	0.011	-0.032	-0.001	-0.001
2.00	2.00	20	2	-3.237	0.002	0.002	-0.010	-0.001	-0.001
2.00	2.00	8	3	-0.453	0.120	0.101	-0.064	-0.115	-0.067
2.00	2.00	10	3	-0.453	0.035	0.025	-0.037	-0.048	-0.028
2.00	2.00	12	3	-0.453	0.012	0.005	-0.024	-0.048	-0.028
2.00	2.00	20	3	-0.453	0.000	-0.003	-0.007	-0.018	-0.014
2.00	2.00	8	4	-0.028	-0.120	-0.094	-0.013	-0.503	-0.387
2.00	2.00	10	4	-0.028	-0.097	-0.093	-0.007	0.074	0.050
2.00	2.00	12	4	-0.028	-0.058	-0.059	-0.004	0.052	0.036
2.00	2.00	20	4	-0.028	-0.010	-0.011	-0.001	0.140	0.123
2.00	2.00	8	5	-0.002	-0.316	-0.268	-0.002	-0.660	-0.536
2.00	2.00	10	5	-0.002	-0.144	-0.140	-0.001	0.347	0.249
2.00	2.00	12	5	-0.002	-0.065	-0.066	-0.001	0.248	0.185
2.00	2.00	20	5	-0.002	-0.007	-0.007	-0.000	0.140	0.123

Legend: R is the limit used in equation (22); c is the potential coefficient of equation (21); N_g is the number of grid points used in collecting the sampling frequencies; q is quadrature point index; I_x is the exact integrand at the quadrature point q ; $\Delta I(p^2)$, $\Delta I(p_{av}^2)$, $\Delta I(e)$, $\Delta I(p)$, $\Delta I(p_{av})$ are the errors of the integrand at the quadrature point q using the quadratic fit, integral of quadratic fit, exponential fit, polynomial fit, and integral of polynomial fit, respectively.

improvements with increased orders until o reaches 4 or 5. From this it is concluded that the o value to be used should not exceed this range – we chose $o = 4$ as the highest order recommended and that value was used in the subsequent tests.

Table 3 shows the comparison of the recommended polynomial fits with the other fitting procedures. In principle, a fitting involving integrals should be better than its counterpart using the approximation of equation (2) since it involves one less approximation. The data in Table 3 does indeed conform to this expectation, thus the quadratic and polynomial fittings using the approximation of equation (19) can be safely eliminated from further considerations.

The relative performance of the quadratic, exponential and polynomial fitting procedures depended on the magnitude of the integrand. For small magnitude of the integrand (larger λ values, $q = 4$ or 5) the exponential fitting consistently outperformed the polynomial-based procedures. For the larger magnitudes (smaller λ values, $q = 1$ or 2), the polynomial fittings outperformed both the quadratic and the exponential fits. In the middle range ($q = 3$) the exponential and quadratic fits are competitive with each other.

Table 4
Comparison of runlength strategies.

Strategy	N_g	$\Delta(p^2)$	$SD(p^2)$	$\Delta(e)$	$SD(e)$	$\Delta(\text{cmb})$	$SD(\text{cmb})$
even	8	-0.011	0.042	-0.044	0.042	-0.035	0.035
opt	8	-0.016	0.021	-0.057	0.017	-0.029	0.028
even	10	0.012	0.066	-0.008	0.057	0.021	0.060
opt	10	-0.039	0.018	-0.045	0.015	-0.042	0.020
even	12	0.023	0.037	0.006	0.016	0.013	0.016
opt	12	-0.010	0.025	-0.015	0.020	-0.009	0.025
even	20	-0.026	0.050	-0.018	0.029	-0.007	0.031
opt	20	-0.001	0.039	-0.013	0.028	-0.010	0.023

Legend: N_g is the number of gridpoints used; $\Delta(p^2)$ and $SD(p^2)$ are the error and standard deviation of the integrated quadratic fit; $\Delta(e)$ and $SD(e)$ are the error and standard deviation of the exponential fit; $\Delta(\text{cmb})$ and $SD(\text{cmb})$ are the error and standard deviation of the combined polynomial and exponential fit; the exact value is 1.1606.

3.4. Comparison of runlength strategies

To quantify the savings by using different runlengths at the various quadrature points, calculations were run using constant runlengths and the optimal choice according to equation (20). Each choice was scaled so that the total number of MC steps were always the same. Ten runs using a total of $5 \cdot 10^5$ MC steps were performed for using the two runlength strategies and the root mean square deviations were calculated for each method. Table 4 gives the result of these comparison for $R = 2$, $c_0 = 0$ and $c_1 = 2$. It can be seen that the use of the optimal MC steps has resulted in a decrease of the statistical error by about 50% – without any additional computational investment.

3.5. Selection of the gridsize

Both the calculation using the exact grid averages and the Monte Carlo calculations produced more accurate results as the number of grids was increased from 8 to 12. Further increase to 20 brought generally little if any improvement. As a result, it is suggested that 12 gridpoints would be the best choice.

4. Conclusions

A new formalism has been proposed for calculating the free energy contribution of molecular flexibility, based on gradually releasing constraints on internal coordinates. The method is also well suited for the calculation of the contribution of selected degrees of freedom to the solvation excess free energy and allows for precise control over the range of configurations that are allowed for a given conformation.

Numerical test on a harmonic oscillator showed that runs using $O(10^6)$ Monte Carlo steps can achieve about 2% accuracy using a 5-point Gaussian quadrature, collecting the data on a 12-point grid and applying an extrapolation procedure using either a polynomial or an exponential fit, depending on the coupling parameter's value.

As it is true with any new technique, significant further developments are necessary for its efficient application to larger systems. Incorporation of molecular dynamics with the hard limits on internal coordinates as required by this new method will be of prime importance.

Acknowledgements

Stimulating discussions with Dr. F. Guarnieri are greatly appreciated. Dr. B. Jayaram is thanked for useful comments on the manuscript.

Appendix A. Fitting formulae

The formulae below (except for the exponential fit) have been derived assuming a grid size of unity and extrapolating to zero – this can always be achieved by the appropriate choice of units.

A.1. Quadratic fit

$$P(r) = ar^2 + br + c$$

can be least-square fitted to the points $(0.5, y_1)$, $(1.5, y_2)$; $(2.5, y_3)$ with the expressions

$$a = \frac{2(y_1 - y_2) - (y_1 - y_3)}{2}, \quad b = y_2 - y_1 - 2a, \quad c = y_1 - \frac{b}{2} - \frac{a}{4}.$$

A.2. Quadratic integral fit

$$\int_{R+k\Delta}^{R+(k+1)\Delta} P(r) = (ar^2 + br + c)dr \quad (k = 1, 2, 3)$$

can be least-square fitted to the points y_1, y_2, y_3 , assumed to be the integrals of $P(r)$ over the intervals $[0, 1]$, $[1, 2]$, and $[1, 3]$, respectively, with the expressions

$$a = \frac{y_1 + y_3 - 2.0y_2}{2.0}, \quad b = 3.0y_2 - 2.0y_1 - y_3$$

and

$$c = \frac{11.0y_1 - 7.0y_2 + 2.0y_3}{6.0}.$$

A.3. Exponential fit

$$P(r) = a \exp(-br)$$

can be least-square fitted to the points $(R - 0.5\Delta, y_1)$, $(R - 1.5\Delta, y_2)$ with the expressions

$$b = \frac{1}{\Delta} \log\left(\frac{y_2}{y_1}\right), \quad a = \frac{y_2 b \Delta}{\exp(-bR)(e^2 - e)}$$

with $e = \exp(b\Delta)$. Interestingly, the same formulae are obtained if y_1 and y_2 are assumed to be the integral of $P(r)$ over the two corresponding grid points.

A.4. Polynomial fit

The polynomial of order n

$$P(r) = \sum_{i=0}^{n_g} a_i r^{i-1}$$

can be fitted to the n_g data points $(i - 0.5, y_i)$ ($i = 1, \dots, n_g$) if the coefficients a_i are the solution of a system of linear equations with coefficient matrix

$$A_{p,q} = \sum_{i=1}^{n_g} (i - 0.5)^p (i - 0.5)^q, \quad 0 \leq p, q \leq n_g$$

and right-hand side

$$b_p = \sum_{i=1}^{n_g} y_i (i - 0.5)^p, \quad 0 \leq p \leq n_g.$$

A.5. Polynomial integral fit

Analogously to the quadratic integral fit, the data points y_i can be considered the integrals of the polynomial $P(r)$ over the intervals $[i - 1, i]$. In this case the optimal coefficients a_i are the solution of a system of linear equations with coefficient matrix

$$A_{p,q} = \sum_{i=1}^{n_g} \frac{[i^{p+1} - (i - 1)^{p+1}][i^{q+1} - (i - 1)^{q+1}]}{(p + 1)(q + 1)}, \quad 0 \leq p, q \leq n_g$$

and right-hand side

$$b_p = \sum_{i=1}^{n_g} \frac{y_i [i^{p+1} - (i - 1)^{p+1}]}{p + 1}, \quad 0 \leq p \leq n_g.$$

References

- [1] M. Mezei and D.L. Beveridge, Ann. Acad. Sci. New York 482 (1986) 1.
- [2] D.L. Beveridge and F.M. DiCapua, Annu. Rev. Biophys. Chem. 18 (1989) 431.

- [3] T.P. Straatsma and J.A. McCammon, *Annu. Rev. Phys. Chem.* 43 (1992) 407.
- [4] J.G. Kirkwood, in: *Theory of Liquids*, ed. B.J. Alder (Gordon and Breach, New York, 1968).
- [5] M. Mezei, *J. Comp. Chem.* 13 (1992) 651.
- [6] H. Resat and M. Mezei, *J. Chem. Phys.* 99 (1993) 6052.
- [7] M. Mezei, *Mol. Simul.* 10 (1993) 225.
- [8] P.V. Maye and M. Mezei, *J. Molecular Struct. (Theochem)* 362 (1996) 317.
- [9] F. Guarnieri, *J. Math. Chem.* 18 (1995) 25.