HHS Public Access

Author manuscript

Mol Simul. Author manuscript; available in PMC 2016 July 21.

Published in final edited form as:

Mol Simul. 2016 ; 42(13): 1079–1089. doi:10.1080/08927022.2015.1110583.

Accelerating the weighted histogram analysis method by direct inversion in the iterative subspace

Cheng Zhang[a], **Chun-Liang Lai**[a], and **B. Montgomery Pettitt**[a,*]

[a]Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, Texas 77555-0304, USA

Abstract

The weighted histogram analysis method (WHAM) for free energy calculations is a valuable tool to produce free energy differences with the minimal errors. Given multiple simulations, WHAM obtains from the distribution overlaps the optimal statistical estimator of the density of states, from which the free energy differences can be computed. The WHAM equations are often solved by an iterative procedure. In this work, we use a well-known linear algebra algorithm which allows for more rapid convergence to the solution. We find that the computational complexity of the iterative solution to WHAM and the closely-related multiple Bennett acceptance ratio (MBAR) method can be improved by using the method of direct inversion in the iterative subspace. We give examples from a lattice model, a simple liquid and an aqueous protein solution.

Keywords

free energy; WHAM; MBAR; DIIS

1. Introduction

An important problem in computational physics and chemistry is to obtain the best estimate of a quantity of interest from a given set of data[1]. For free energy calculations, the multiple histogram method[2–5] or its generalisation, the weighted histogram analysis method (WHAM)[1, 6–17], is an effective tool for addressing such a problem. WHAM is statistically efficient in using the data acquired from molecular simulations, and it has become a standard free energy analysis tool, particularly popular for enhanced-sampling simulations, such as umbrella sampling[18, 19], and simulated[20, 21] and parallel[22–26] tempering.

Given several distributions collected at different thermodynamic states, WHAM obtains the optimal estimate of the free energies of the states. This problem arises in the calculation of potentials of mean force and a variety of free energy difference methods. The central idea of WHAM is to find an optimal estimate of the density of states, or the unbiased distribution, which then allows the free energies to be evaluated as weighted integrals. The optimal density of states is computed from a weighted average of the reweighted energy histograms

*Corresponding author. bmpettitt@utmb.edu.

(hence the name of the method) from different distribution realisations or trajectories. The weights, however, depend on the free energies, so that the free energies and the density of states must be determined self-consistently.

WHAM can be reformulated[1, 17] in the limit of zero histogram bin width as an extension of the Bennett acceptance ratio (BAR) method[27]. This form, adopted by the multistate BAR (MBAR) method[1], avoids the histogram dependency[1, 6, 12].

The straightforward implementation of WHAM or MBAR, in which the equations regarding the free energies are solved by direct iteration, can suffer from slow convergence in the later stages[14, 16, 28]. Several remedies have been proposed[1, 14, 16, 28, 29]. For example, one may use the Newton-Raphson method, which involves a Hessian-like matrix, although the approach sometimes can be unstable[1]. Other more advanced techniques include the trust region and Broyden-Fletcher-Goldfarb-Shanno (BFGS) methods[16].

An elegant non-iterative alternative is the statistical-temperature WHAM (ST-WHAM)[28, 29], which determines the density of states through its logarithmic derivative, or the statistical temperature. In this way, the method estimates the density of states non-iteratively with minimal approximation. ST-WHAM can be regarded as a refinement of the more approximate umbrella integration method (UIM)[30, 31]. However, the extension to multidimensional ensembles, e.g., the isothermal-isobaric ensemble, can be numerically challenging[28].

Here, we discuss a numerical improvement of the implementation of WHAM and MBAR using the method of direct inversion in the iterative subspace (DIIS)[32–36]. DIIS shares characteristics with other optimisation techniques which use a limited (non-spanning) basis of vectors which produce the most gain towards the optimum. Although still iterative in nature, this implementation can often improve the rate of convergence significantly in difficult cases.

2. Method

2.1. WHAM

WHAM is a method of estimating the free energies of multiple thermodynamic states with different parameters, such as temperatures, pressures, etc. Below, we shall first review WHAM in the particular case of a temperature scan, since it permits simpler mathematics without much abstraction. Generalisations to umbrella sampling and other ensembles are discussed afterwards.

Consider K temperatures, labelled by $\beta = 1/(k_B T)$, as $\beta_1, \ldots, \beta_K$. Suppose we have performed the respective canonical (NVT) ensemble simulations at the K temperatures, and we wish to estimate the free energies at those temperatures.

In WHAM, we first estimate the density of states, $g(E) = \int \delta(\mathscr{E}(\mathbf{x}) - E)\, d\mathbf{x}$, as the number density of configurations, $\mathbf{x}$, with energy E, from

$$g(E)=\frac{\sum_{k=1}^{K} n_k(E)}{\sum_{k=1}^{K} N_k \exp(-\beta_k E)/Z_k}, \quad (1)$$

where $n_k(E)$ is the unnormalised energy distribution observed from trajectory k, which is usually estimated from the energy histogram as the number of independent trajectory frames whose energies fall in the interval $(E - \Delta E/2, E + \Delta E/2)$ divided by ΔE (we shall omit "independent" below for simplicity); thus, $n_k(E)/N_k$ is the normalised distribution with N_k being the total number of trajectory frames from simulation k; and finally,

$$\exp(-f_k)=Z_k=\int g(E)\ \exp(-\beta_k E)dE, \quad (2)$$

with f_k and Z_k being the dimensionless free energy and partition function, respectively.

To understand Eq. (1), we first observe from the definition the single histogram estimate

$$g(E)=\frac{n_k(E)}{d_k(E)} \text{ for } k=1,\ldots,K, \quad (3)$$

where $d_k(E) \equiv N_k \exp(-\beta_k E + f_k)$. That is, the observed distribution, $n_k(E)/N_k$, should roughly match the exact one, $g(E)\ w_k(E)$, where $w_k(E) \equiv \exp(-\beta_k E + f_k)$ is the normalised weight of the canonical ensemble. The values of Eq. (3) from different k can be combined to improve the precision, and Eq. (1) is the optimal combination[4, 5, 7, 12, 13]. To see this, recall that in an optimal combination, the relative weight is inversely proportional to the variance. Assuming a Poisson distribution so that $\mathrm{var}(n_k \Delta E) \approx \langle n_k \rangle \Delta E$, we have $\mathrm{var}(n_k \Delta E/d_k) \approx \langle n_k \rangle \Delta E/d_k^2 \propto 1/d_k$ (here, $\langle \cdots \rangle$ means an ensemble average). Averaging the values from Eq. (3) using $(1/d_k)^{-1}$ as the relative weight yields Eq. (1).[1] Since for a fixed E, $d_k(E)$ is proportional to $\langle n_k(E) \rangle$, several variants of WHAM may be derived by using $n_k(E)$ in place of $d_k(E)$ as the relative weight for $g(E)$[37] or related quantities[7, 28–31, 38, 39].

From Eqs. (1) and (2), we find that f_i satisfies

$$f_i = -\log \int \frac{\sum_{k=1}^{K} n_k(E)\ \exp(-\beta_i E)}{\sum_{k=1}^{K} N_k\ \exp(-\beta_k E + f_k)} dE\ \equiv -\log\ \mathscr{Z}_i(\mathbf{f}), \quad (4)$$

[1] We can also view Eq. (1) as the result of applying Eq. (3) to the composite ensemble of the K canonical ensembles[13], using the following substitutions: $n_k(E) \to \sum_{k=1}^{K} n_k(E)$, $N_k \to \sum_{k=1}^{K} N_k \equiv N_{\mathrm{tot}}$, and $w_k(E) \to \sum_{k=1}^{K} (N_k/N_{\mathrm{tot}}) w_k(E)$. The last expression is the normalised ensemble weight in the composite ensemble. Note that the composite ensemble is closely related to the expanded ensemble that underlies the simulated tempering method[20, 21]. However, in the latter case, one can show that the N_k in Eq. (1) should be replaced by the average, $\langle N_k \rangle$, and the combination $\langle N_k \rangle / Z_k$ is proportional to the weight $\exp(\eta_k)$ in the acceptance probability of temperature transitions, $A(\beta_k \to \beta_{k'}) = \min\{1, \exp[\eta_{k'} - \eta_k - (\beta_{k'} - \beta_k)E]\}$. This variant is thus more convenient as it depends only on the known weight, η_k, instead of on the unknown f_k.

where $\mathscr{L}_i(\mathbf{f})$ denotes the integral on the right-hand side as a function of $\mathbf{f} = (f_1, \ldots, f_K)$. Once all f_i and $g(E)$ are determined, the free energy at a temperature not simulated, β, can be found from Eq. (2) by substituting β for β_k.

Histogram-free form—The histogram dependency of WHAM [in using $n_k(E)$] can be avoided by noticing from definition that[12]

$$n_k(E)=\sum_{\mathbf{x}}^{(k)}\delta(\mathscr{E}(\mathbf{x})-E), \quad (5)$$

where, $\mathscr{E}(\mathbf{x})$ is the energy function, and $\sum_{\mathbf{x}}^{(k)}$ denotes a sum over trajectory frames of simulation k. Using Eq. (5) in Eq. (4) yields the histogram-free, or the MBAR, form[1, 6, 12, 17]:

$$f_i=-\log\sum_{j=1}^{K}\sum_{\mathbf{x}}^{(j)}\frac{q_i(\mathbf{x})}{\sum_{k=1}^{K}N_k q_k(\mathbf{x})\exp(f_k)}. \quad (6)$$

where $q_i(\mathbf{x}) \equiv \exp[-\beta_i \mathscr{E}(\mathbf{x})]$. The $K = 2$ case is the BAR result[27], and Eq. (6) also holds for a general setting, which permits, e.g., a nonlinear parameter dependence (see Appendix A for derivation). In this sense, MBAR is not only the zero-bin-width limit of WHAM[1, 17], but also a generalisation[1]. As we shall see, the structural similarity of Eqs. (4) and (6) allows our acceleration technique to be applicable to both cases. Since both Eqs. (4) and (6) are invariant under $f_i \to f_i + c$ for all i and an arbitrary c, f_i are determined only up to a constant shift.

Extensions to umbrella sampling—We briefly mention a few extensions. First, for a general Hamiltonian with a linear bias

$$\mathscr{H}(\mathbf{x};\lambda_i)=\mathscr{H}_0(\mathbf{x})+\lambda_i\mathscr{W}(\mathbf{x}),$$

such that $q_i(\mathbf{x}) = \exp[-\mathscr{H}(\mathbf{x}; \lambda_i)]$, we can show, by inserting $1 = \int \delta(\mathscr{W}(\mathbf{x}) - W)\, dW$ into Eq. (6), that

$$f_i=-\log\int\frac{\sum_{k=1}^{K}n_k(W)\,\exp(-\lambda_i W)}{\sum_{k=1}^{K}N_k\,\exp(-\lambda_k W+f_k)}dW,$$

where $n_k(W)\equiv\sum_{\mathbf{x}}^{(k)}\delta(\mathscr{W}(\mathbf{x})-W)$ is understood to be the unnormalised distribution of the bias $\mathscr{W}(\mathbf{x})$. Equation (4) is the special case of $\mathscr{H}_0(\mathbf{x}) = 0$, $\mathscr{W}(\mathbf{x}) = \mathscr{E}(\mathbf{x})$, and $\lambda_i = \beta_i$. Another common example is a system under a quadratic restraint (umbrella)

$\mathscr{E}_i=\mathscr{E}_0+\frac{1}{2}A(\xi-\lambda_i)^2$ for some reaction coordinate $\xi \equiv \xi(\mathbf{x})$. In this case,

$\mathscr{H}_0(\mathbf{x})=\beta(\mathscr{E}_0+\frac{1}{2}A\xi^2)$, and $\mathscr{W}(\mathbf{x}) = -\beta A\xi$. The configuration independent term, $\frac{1}{2}\beta A\lambda_i^2$, can be added back to f_i after the analysis.

Extensions to other ensembles—Further, λ_i and W can be generalised to vectors as $\boldsymbol{\lambda}_i$ and $\mathbf{W}$, respectively. For example, for simulations on multiple isothermal-isobaric (NpT) ensembles with different temperatures and pressures, we set $\boldsymbol{\lambda}_i = (\beta_i, \beta_i p_i)$ and $\mathbf{W} = (E, V)$ with p_i and V being the pressure and volume, respectively. However, if the vector dimension is high, or if the Hamiltonian $\mathscr{H}(\mathbf{x}; \lambda_i)$ depends nonlinearly on λ_i, the histogram-free form, Eq. (6), is more convenient[1]. Besides, the factor $\exp(-\beta_i E)$ can be replaced by a non-Boltzmann (e.g., the multicanonical[40–42], Tsallis[43], microcanonical[44, 45]) weight for multiple non-canonical simulations[28].

Solution by iteration—Numerically, the f_i are most often determined by treating Eq. (4) as an iterative equation,

$$f_i^{(\mathrm{new})} = -\log\, \mathscr{Z}_i\left(\mathbf{f}^{(\mathrm{old})}\right).$$

However, this approach, referred to as direct WHAM below, can take thousands of iterations to finish[14, 16, 28] (cf. Appendix B). In Secs. 2.3 and 2.4, we give a numerical technique to accelerate the solution of Eq. (4) or (6).

2.2. ST-WHAM and UIM

For comparison, we briefly discuss two non-iterative alternatives, ST-WHAM[28, 29] and UIM[30, 31]. By taking the logarithmic derivative of the denominator of Eq. (1), we get,

$$\frac{d}{dE}\log\sum_{k=1}^{K} d_k(E) = -\frac{\sum_{k=1}^{K} d_k(E)\beta_k}{\sum_{k=1}^{K} d_k(E)} = -\frac{\sum_{k=1}^{K} n_k(E)\beta_k}{\sum_{k=1}^{K} n_k(E)},$$

where we have used Eq. (3) to make the final expression independent of $d_k(E)$, hence f_k. So

$$g(E) = \left[\sum_{k=1}^{K} n_k(E)\right] \exp\left[\int^{E} \frac{\sum_{k=1}^{K} n_k(E')\beta_k}{\sum_{k=1}^{K} n_k(E')} dE'\right].$$

This is the ST-WHAM result. In evaluating the integral, we may encounter an empty bin with $\sum_{k=1}^{K} n_k(E')\Delta E = 0$, which leaves the integrand indeterminate. An expedient fix is to let the integrand borrow the value from the nearest non-empty bin [note, however, setting the integrand to zero would cause a larger error in $g(E)$]. ST-WHAM is most convenient in one dimension, and its results usually differ only slightly from those of WHAM[28]. In UIM[30, 31], the distribution $n_k(E)$ is further approximated as a Gaussian. We note that ST-WHAM is difficult when we have more thermodynamic variables in the ensemble of interest. Below we show that WHAM with the DIIS method can handle such a case for the Lennard-Jones (LJ) fluid, readily.

2.3. DIIS

DIIS is a technique useful for solving equations[32–36], and we use it to solve Eq. (4) here. A schematic illustration is shown in Fig. 1. We first represent an approximate solution by a trial vector, $\mathbf{f} = (f_1, \ldots, f_K)$, which is, in our case, the vector of the dimensionless free energies. The target equations can be written as

$$R_i(\mathbf{f})=0 \text{ for } i=1,\ldots,K, \quad (7)$$

which is $-\log \mathscr{Z}_i(\mathbf{f}) - f_i$ in our case. The left-hand side of Eq. (7) also forms a K-dimensional vector, $\mathbf{R} = (R_1, \ldots, R_K)$, which is referred to as the residual vector. The magnitude $\|\mathbf{R}\|$ represents the error, and $\mathbf{R}(\mathbf{f})$ should optimally point in a direction that reduces the error of $\mathbf{f}$.

If Eq. (7) is solved by direct iteration, $\mathbf{f}$ is replaced by $\mathbf{f} + \mathbf{R}(\mathbf{f})$ in each time. This can be a slow process because the residual vector $\mathbf{R}$ does not always have the proper direction and/or magnitude to bring $\mathbf{f}$ close to the true solution $\mathbf{f}^*$. The magnitude of $\mathbf{R}$, however, can be used as a reliable measure of the error of $\mathbf{f}$. Thus, in DIIS, we try to find a vector $\hat{\mathbf{f}}$ with minimal error $\|\mathbf{R}(\hat{\mathbf{f}})\|$, which would be more suitable for direct iteration.

Suppose now we have a basis of M trial vectors $\mathbf{f}_1, \ldots \mathbf{f}_M$ (where M can be much less than K), and the residual vectors are $\mathbf{R}_1, \ldots R_M$ [where $\mathbf{R}_j \equiv \mathbf{R}(\mathbf{f}_j)$ for $j = 1, \ldots, M$], respectively. We wish to construct a vector $\hat{\mathbf{f}}$ with minimal error from a linear combination of the trial vectors.

To do so, we first find the combination of the residual vectors $\hat{\mathbf{R}}=\sum_{i=1}^{M} c_i \mathbf{R}_i$ that minimises the error $\|\hat{\mathbf{R}}\|$ under the constraint

$$\sum_{i=1}^{M} c_i = 1. \quad (8)$$

Mathematically, this means that we solve for the c_i simultaneously from Eq. (8) and

$$\sum_{j=1}^{M} (\mathbf{R}_i \cdot \mathbf{R}_j) c_j = \lambda, \quad (9)$$

for all i, with λ being an unknown Lagrange multiplier that is to be determined from along with c_i. Now the corresponding combination of the trial vectors, $\hat{\mathbf{f}}=\sum_{i=1}^{M} c_i \mathbf{f}_i$, should be close to the desired minimal-error vector. This is because around the true solution $\mathbf{f}^*$, Eq. (7) should be nearly linear; so the residual vector

$$\mathbf{R}(\hat{\mathbf{f}})=\mathbf{R}\left(\sum_{i=1}^{M}c_i\mathbf{f}_i\right)\approx\sum_{i=1}^{M}c_i\mathbf{R}(\mathbf{f}_i)=\hat{\mathbf{R}},$$

has the minimal magnitude. In other words, $\hat{\mathbf{f}}$, among all linear combinations of $\{\mathbf{f}_i\}$, is the closest to the true solution, under the linear approximation. Thus, an iteration based on $\hat{\mathbf{f}}$ should be efficient.

With $M = K + 1$ independent bases, one can show that it is possible to find a combination with zero $\hat{\mathbf{R}}$, which means that $\hat{\mathbf{f}}$ would be the true solution if the equations were linear. A particularly instructive case is that of two vectors ($M = 2$) in one dimension ($K = 1$)[2]. We then recover the secant method[46], as shown in Fig. 1(d). The number of bases, M, however, should not exceed $K + 1$ (or, in our case, K because of the arbitrary shift constant of f_i) to keep Eqs. (9) independent, although this restriction may be relaxed by using certain numerical techniques[46].

We now construct a new trial vector $\mathbf{f}^{(n)}$ as

$$\mathbf{f}^{(n)}=\hat{\mathbf{f}}+\alpha\hat{\mathbf{R}}(\hat{\mathbf{f}}),$$

where the factor α is 1.0 in this study (although a smaller value is recommended for other applications[35, 36]). The new vector $\mathbf{f}^{(n)}$ is used to update the basis as shown next.

2.4. Basis updating

In each iteration of DIIS, the basis is updated by the new trial vector $\mathbf{f}^{(n)}$ from the above step. Initially, the basis contains a single vector. As we add more vectors into the basis in subsequent iterations, some old vectors may be removed to maintain a convenient and efficient maximal size of M.

A simple updating scheme[35] is to treat the basis as a queue: we add $\mathbf{f}^{(n)}$ to the basis, if the latter contains fewer than M vectors, or substitute $\mathbf{f}^{(n)}$ for the earliest vector in the basis. If, however, $\mathbf{f}^{(n)}$ produces an error greater than K_r times the error of $\mathbf{f}_{\min}$, the least erroneous vector in the basis, we rebuild the basis from $\mathbf{f}_{\min}$. Here, the error of a vector $\mathbf{f}$ is defined as $\|\mathbf{R}(\mathbf{f})\|$, and $K_r = 10.0$ is recommended[35].

We used the following modification in this study. First, we find the most erroneous vector, $\mathbf{f}_{\max}$, from the basis. If the new vector, $\mathbf{f}^{(n)}$, produces an error less than $\mathbf{f}_{\max}$, we add $\mathbf{f}^{(n)}$ into the basis or, if the basis is full, substitute $\mathbf{f}^{(n)}$ for $\mathbf{f}_{\max}$. Otherwise, we remove $\mathbf{f}_{\max}$ from the basis, and if this empties the basis, we rebuild the basis from $\mathbf{f}^{(n)}$.

Since the DIIS process is reduced to the direct iteration if $M = 1$, the method is effective only if multiple basis vectors are used.

[2]This is the case for solving the BAR equation, because f_i are determined up to a constant shift, the virtual dimension is 1.

3. Results

We tested DIIS WHAM and MBAR on three systems: Ising model, LJ fluid, and the villin headpiece (a small protein) in aqueous solution (see Secs. 3.2, 3.3, and 3.4, respectively, for details). We tuned the parameters such that direct WHAM and MBAR would take thousands of iterations to finish.

The main results are summarised in Fig. 2, from which one can see that DIIS can speed up WHAM and MBAR dramatically in these cases. The real run time roughly matched the number of iterations, suggesting a negligible overhead for using DIIS. This is unsurprising, for it is often much more expensive to compute the right-hand side of Eq. (4) or (6).

3.1. Set-up

For simplicity, we assumed equal autocorrelation times from different temperatures (and pressures). The approximation should not affect the convergence behaviour of the methods.

In testing WHAM and MBAR, the initial free energies were obtained from the single histogram method:

$$\Delta f_i = \log \langle \exp(\Delta \beta_i E) \rangle_{i+1},$$

where $\Delta A_i \equiv A_{i+1} - A_i$, for any quantity A, and $\langle \ldots \rangle_{i+1}$ denotes an average over trajectory $i + 1$. Then, $f_i = f_1 + \sum_{j=1}^{i-1} \Delta f_j$. Iterations are continued until all $|R_i|$ are reduced below a certain value.

For comparison, we also computed f_i from three approximate formulae. The first is[16, 48]

$$\Delta f_i \approx \overline{\langle E \rangle}_i \Delta \beta_i, \quad (10)$$

where $\bar{A}_i \equiv (A_{i+1} + A_i)/2$. The second is an improvement by the Euler-Maclaurin expansion[49–52]:

$$\Delta f_i \approx \overline{\langle E \rangle}_i \Delta \beta_i + (\langle \delta E^2 \rangle_{i+1} - \langle \delta E^2 \rangle_i) \frac{\Delta \beta_i^2}{12}, \quad (11)$$

where $\langle \delta E^2 \rangle_k \equiv \langle (E - \langle E \rangle_k)^2 \rangle_k$ for $k = i$ and $i + 1$. The third formula is derived from the same expansion but using E instead of β as the independent variable (after integration by parts, $\int E \, d\beta = E\beta - \int \beta \, dE$):

$$\Delta f_i \approx \overline{\langle E \rangle}_i \Delta \beta_i - (\left\langle \delta E^2 \right\rangle_{i+1}^{-1} - \left\langle \delta E^2 \right\rangle_i^{-1}) \frac{(\Delta \langle E \rangle_i)^2}{12}. \quad (12)$$

3.2. Ising model

The first system is a 64 × 64 Ising model. We used parallel tempering[22–26] Monte Carlo (MC) for eighty temperatures: T = 1.5, 1.52, …, 3.08,

To study the accuracy, we generated a large sample with 10^9 single-site MC steps for each temperature. Figure 3 shows that DIIS and direct WHAMs produced identical dimensionless free energies. The differences between the ST-WHAM and WHAM results were subtle, whereas the approximate UIM produced more deviation in the results, especially around the critical region.

To study the rate of convergence, we generated independent samples with 10^7 MC steps at each temperature. Figure 4 shows a faster decay of the error in DIIS WHAM than in direct WHAM.

3.3. LJ fluid

We tested the DIIS method on the 256-particle LJ fluid in the isothermal-isobaric (NpT) ensemble. This is a case for the two-dimensional (p and T) WHAM, which can be difficult for ST-WHAM[28]. The potential interaction between particles was cutoff at half the box size. We simulated the system using parallel tempering MC. We considered the system at N_T × N_p = 6 × 3 conditions, with temperatures T = 1.2, 1.3, …, 1.7, and pressures p = 0.1, 0.15, 0.2. The bin sizes for energy and volume were 1.0 and 2.0, respectively.

As shown in Fig. 2, DIIS WHAM effectively reduced the run time, although the efficiency of DIIS does not always increase with the number of basis set members, M.

3.4. Villin headpiece

We tested the methods on a small protein, the villin headpiece (PDB ID: 1VII), in aqueous solution. This is a well-known test system[55]. The protein was immersed in a dodecahedron box with 1898 TIP3P water molecules and two chloride ions. The force field was AMBER99SB[56, 57] with the side-chain modifications[58]. Molecular dynamics (MD) simulations were performed using GROMACS 5.0[59–65], with a time step of 2 fs. Velocity rescaling[66] was used as the thermostat with the time constant being 0.1 ps. The electronic interaction was handled by the particle mesh Ewald method[67]. The constraints were handled by the LINCS method[68] for hydrogen-related chemical bonds on the protein and by the SETTLE method[69] for water molecules.

We simulated the system at 12 temperatures T = 300 K, 310 K, …, 410 K, each for approximately 200 ns. The energy distributions were properly overlapped, as shown in Fig. 5(e). The energies of individual trajectory frames were saved every 0.1 ps, so that there were about 2 million frames for analysis at each temperature.

As shown in Fig. 2, direct WHAM suffered from slow convergence, while the DIIS methodology again delivered a speed-up of two orders of magnitude, in the number of iterations or in real time. The MBAR case was similar, although MBAR was slower than WHAM as it did not use histogram to aggregate data.

To compare the errors of the methods, we prepared two types of samples of different sizes. A larger sample was randomly selected from roughly $r = 1\%$ of all trajectory frames from every temperature. In a smaller one, r was reduced to 0.01%. The reference values of f_i were computed from all trajectory frames using MBAR, which is the zero-bin-width limit of WHAM[1, 17]. In terms of accuracy [Figs. 5(a) and 5(c)], WHAM, MBAR, BAR, ST-WHAM, Eqs. (11) and (12) were comparably good; UIM was slightly inferior for the larger sample; Eq. (10) was the worst. The accuracy was largely independent of the sample size. In terms of precision [Figs. 5(b) and 5(d)], the differences were small. Generally, WHAM was insensitive to the bin size, whereas ST-WHAM was slightly affected by a small bin size for the smaller samples.

4. Conclusions

In this work, we showed that the DIIS technique can often significantly accelerate WHAM and MBAR to produce free energy difference. The technique achieves rapid convergence by an optimal combination of the approximate solutions obtained during iteration. DIIS does not require computing the Hessian-like matrix, $-\partial \mathbf{R}(\mathbf{f})/\partial \mathbf{f}$, and is numerically stable with minimal run time overhead. Compared to other advanced techniques[1, 16], DIIS is relatively simple and easy to implement. However, methods based on Hessian matrices may further accelerate the solution process for final stages. Other related free energy methods[7, 37–39] may also benefit from this technique.

There are some non-iterative alternatives to WHAM, although they may be less general and/or accurate in some aspect. The use of DIIS makes scanning more than one thermodynamic state variable or a nonlinear variable computationally convenient. This was demonstrated here in the NpT LJ fluid case. Problems with only one thermodynamic variable to scan, such as temperature, are amenable to the non-iterative ST-WHAM, or the more approximate UIM.

Acknowledgments

It is a pleasure to thank Dr. Y. Zhao, J. A. Drake, Dr. M. R. Shirts, Dr. M. W. Deem, Dr. J. Ma, Dr. J. Perkyns and Dr. G. Lynch for helpful discussions.

Funding

The authors acknowledge partial financial support from the National Science Foundation (CHE-1152876), the National Institutes of Health (GM037657), and the Robert A. Welch Foundation (H-0037). Computer time on the Lonestar supercomputer at the Texas Advanced Computing Center at the University of Texas at Austin is gratefully acknowledged.

Appendix A

Probabilistic derivations of Eq. (6)

Here we give some derivations of Eq. (6). First, we show that Eq. (6) is a generalisation of Eq. (4) from the energy space to the configuration space. We follow the probabilistic argument[8–11, 16, 17] for simplicity. We assume that the system is subject to an unknown

underlying configuration-space field, $g(\mathbf{y})$, such that the distribution of state k is $w_k(\mathbf{y}) \equiv g(\mathbf{y})\, q_k(\mathbf{y})/Z_k[g]$, with

$$Z_k[g] = \int g(\mathbf{y}) q_k(\mathbf{y}) d\mathbf{y}. \quad \text{(A1)}$$

We now seek the most probable $g(\mathbf{y})$ from the observed trajectory. Given a certain $g(\mathbf{y})$, the probability of observing the trajectories, $\{\mathbf{x}\}$, is given by

$$p(\{\mathbf{x}\}|g) \propto \prod_{k=1}^{K} \prod_{\mathbf{x}}^{(k)} \frac{g(\mathbf{x}) q_k(\mathbf{x})}{Z_k[g]}.$$

This is also the likelihood of g given the observed trajectory, $\{\mathbf{x}\}$. Thus, to find the most probable $g(\mathbf{y})$, we only need to maximise log $p(\{\mathbf{x}\}|g)$ by taking the functional derivative with respect to $g(\mathbf{y})$ and setting it to zero, which yields

$$g(\mathbf{y}) = \frac{\sum_{j=1}^{K} \sum_{\mathbf{x}}^{(j)} \delta(\mathbf{x} - \mathbf{y})}{\sum_{k=1}^{K} N_k q_k(\mathbf{y})/Z_k[g]}, \quad \text{(A2)}$$

where we have used $\delta \log g(\mathbf{x})/\delta g(\mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})/g(\mathbf{y})$, and $\delta Z_k[g]/\delta g(\mathbf{y}) = q_k(\mathbf{y})$.[3] Using Eq. (A2) in Eq. (A1), and then setting g to 1.0 yields Eq. (6).

We can also show Eq. (6) without introducing the configuration-space field [thus, $g(\mathbf{x}) = 1$ below]. Instead, we now assume that each trajectory frame $\mathbf{x}$ is free to choose the state i according to the Bayes' rule[70, 71]: since the prior probability of state i is N_i/N_{tot}, and the conditional probability of observing $\mathbf{x}$ in state i is $w_i(\mathbf{x}) \equiv q_i(\mathbf{x})/Z_i$, the posterior probability of $\mathbf{x}$ being in state i is given by[11]

$$p(i|\mathbf{x}) = \frac{(N_i/N_{\text{tot}}) w_i(\mathbf{x})}{\sum_{k=1}^{K} (N_k/N_{\text{tot}}) w_k(\mathbf{x})}.$$

Summing this over all trajectory frames $\mathbf{x}$ (no matter the original state j) yields the expected population of state i,

$$\hat{N}_i = \sum_{\mathbf{x}} p(i|\mathbf{x}) = \sum_{\mathbf{x}} \frac{N_i w_i(\mathbf{x})}{\sum_{k=1}^{K} N_k w_k(\mathbf{x})}, \quad \text{(A3)}$$

[3] We can also show Eq. (A2) by considering the composite ensemble of the K states. The number of visits to a phase-space element at $\mathbf{y}$ is given by the product of the total sample size, $N_{\text{tot}} = \sum_{k=1}^{K} N_k$, the distribution of the composite ensemble $w_{\text{tot}} = \sum_{k=1}^{K} (N_k/N_{\text{tot}}) w_k(\mathbf{y})$, and the volume $d\mathbf{y}$. This gives $g(\mathbf{y}) \sum_{k=1}^{K} N_k q_k(\mathbf{y})/Z_k(g) d\mathbf{y}$, and it is expected to match the observed number of visits to the element, $\sum_{j=1}^{K} \sum_{\mathbf{x}}^{(j)} \delta(\mathbf{x} - \mathbf{y}) d\mathbf{y}$, regardless of the state j.

where $\sum_{\mathbf{x}} \equiv \sum_{j=1}^{K} \sum_{\mathbf{x}}^{(j)}$. If we demand $\hat{N_i}$ to be equal to the N_i, Eq. (6) is recovered. Alternatively, by following the argument used in the expectation-maximisation algorithm[11, 71–74], we can view Eq. (A3) as the result of maximising the probability of observing the trajectory $\{\mathbf{x}\}$,

$$p(\{\mathbf{x}\}|N_1,\ldots,N_K) \propto \prod_{\mathbf{x}} \sum_{k=1}^{K} (N_k/N_{\mathrm{tot}}) w_k(\mathbf{x}),$$

with respect to variations of N_i under the constraint $\sum_{i=1}^{K} N_i = N_{\mathrm{tot}}$.

Appendix B

Models for the convergence of direct WHAM

Here, we use simple models to study the convergence of direct WHAM. We shall show that slow convergence is associated to a wide temperature range, especially with a large spacing.

B.1. Linearised WHAM equation

Consider K distributions, $\rho_i(E)$, at different temperatures, β_i (i = 1, …, K), normalised as $\int \rho_i(E)\, dE = 1$. For simplicity, we assume equal sample sizes, N_i. Then, Eq. (4), can be written in the iterative form as

$$f_i^{(\mathrm{new})} = -\log \int \frac{\sum_{k=1}^{K} \rho_k(E)\, \exp(-\beta_i E)}{\sum_{k=1}^{K} \exp(-\beta_k E + f_k^{(\mathrm{old})})} dE. \quad \text{(B1)}$$

Around the true solution, $\mathbf{f}^* = (f_1^*, \ldots, f_K^*)$, the equation can be linearised as

$$\delta f_i^{(\mathrm{new})} = \sum_{j=1}^{K} A_{ij} \delta f_j^{(\mathrm{old})},$$

where $\delta f_i^{(\mathrm{new/old})} \equiv f_i^{(\mathrm{new/old})} - f_i^*$, and

$$A_{ij} = \int \frac{\sum_{k=1}^{K} \rho_k(E) e^{-(\beta_i+\beta_j)E + f_i^* + f_j^*}}{D^2(E;\mathbf{f}^*)} dE. \quad \text{(B2)}$$

with $D(E;\mathbf{f}^*) = \sum_{k=1}^{K} e^{-\beta_k E + f_k}$. In matrix form, we have $\delta\mathbf{f}^{(\mathrm{new})} = \mathbf{A}\, \delta\mathbf{f}^{(\mathrm{old})}$.

The elements of matrix $\mathbf{A}$ are positive, i.e., $A_{ij} > 0$, and normalised, i.e.,

$$\sum_{i=1}^{K} A_{ij}=1. \quad \text{(B3)}$$

The latter can be seen from Eq. (B1) with $f_i^{(\text{new/old})} \to f_i^*$. Besides, **A** is symmetric:

$$A_{ij}=A_{ji}. \quad \text{(B4)}$$

Thus, **A** can be regarded as a transition matrix[4]. Each left eigenvector **c** = (c_1, …, c_K) of **A** is associated to a mode of δ**f** and the eigenvalue λ gives the rate of error reduction during iteration. That is, $\sum_{i=1}^{K} c_i \delta f_i$ decays as λ^n asymptotically with the number of iterations, *n*.

The largest eigenvalue is λ_0 = 1.0, and its eigenvector **c** = (1, …, 1) corresponds to a uniform shift of all δf_i, which is unaffected by the iteration, i.e., $\sum_{i=1}^{K} \delta f_i^{(\text{new})} = \sum_{i=1}^{K} \delta f_i^{(\text{old})}$, as a consequence of Eq. (B3). The next largest eigenvalue, λ_1, determines the rate of convergence of the slowest mode, and a larger value of $1 - \lambda_1$ means faster convergence. Below, we determine λ_1 in a few solvable cases.

B.2. Exact distribution approximation (EDA)

To proceed, we assume that the observed distributions are exact, and thus the solution, $(f_1^*, \ldots, f_K^*)$, is also exact. Then, for any *k*, we have

$$\rho_k(E)=g(E)e^{-\beta_k E+f_k^*}, \quad \text{(B5)}$$

with *g*(*E*) being the density of states. This simplifies Eq. (B2) as

$$A_{ij}=\int \frac{g(E)e^{-(\beta_i+\beta_j)E+f_i^*+f_j^*}}{D(E;\mathbf{f}^*)}dE. \quad \text{(B6)}$$

B.3. Gaussian density of states

Further, we approximate *g*(*E*) as a Gaussian[14]. With a proper choice of the multiplicative constant of *g*(*E*), we have,

$$g(E)=\sqrt{\frac{\alpha}{2\pi}}\exp\left[-\frac{1}{2}\alpha(E-E_c)^2\right]. \quad \text{(B7)}$$

It follows

$$f_i = -\log \int g(E)\ \exp(-\beta_i E) dE = \beta_i E_c - \frac{\beta_i^2}{2a}, \quad \text{(B8)}$$

and

$$\rho_i = \sqrt{\frac{a}{2\pi}} \exp\ \left[-\frac{1}{2} a \left(E - E_c + \frac{\beta_i}{a} \right)^2 \right].$$

Thus, the distribution at any temperature is a Gaussian of the same width $\sigma_E = 1/\sqrt{a}$. The inverse temperature β_i affects the energy distributions only as a linear shift to the distribution centre. It follows that both the origin of β and that of E, represented by E_c, can be set to any values that help calculation, without affecting the value of A_{ij}.

B.4. Two-temperature case

For the two-temperature case, the matrix **A** has only one free variable because of Eqs. (B3) and (B4), and it can be written as

$$\mathbf{A} = \begin{pmatrix} 1 - A_{12} & A_{12} \\ A_{12} & 1 - A_{12} \end{pmatrix}.$$

Thus, the second largest eigenvalue is $\lambda_1 = 1 - 2\ A_{12}$.

Under EDA, we have, from Eqs. (B5) and (B6),

$$A_{12} = \int \frac{\rho_1(E)\rho_2(E)}{\rho_1(E) + \rho_2(E)} dE\ \leq \int \frac{\rho_1(E) + \rho_2(E)}{4} dE = \frac{1}{2},$$

where the equality is achieved only for identical distributions. Geometrically, A_{12} represents the degree of overlap of the distributions, as shown in Fig. B1(a), and it decreases with the separation of the two distributions, from the maximal value, 1/2, achieved at $\beta_1 = \beta_2$.

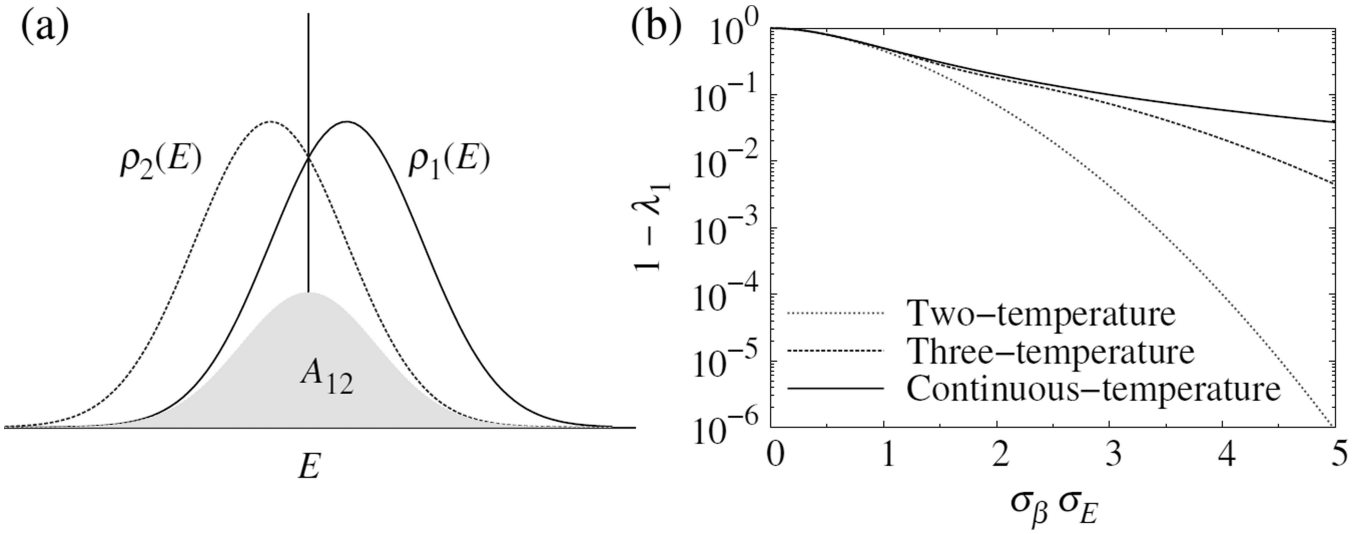

Figure B1. (a) A_{12}, which determines the rate of convergence in the two-temperature case, as a measure of overlap of the two energy distributions. (b) Comparison of the rates of convergence in the two-, three-, and continuous-temperature cases, Eqs. (B9), (B10), and (B12), respectively. Here, σ_β is the standard deviation of the β distribution; σ_E is that of the energy distribution at any temperature. A larger value of $1-\lambda_1$ means faster convergence.

By further assuming a Gaussian density of states, we can, without loss of generality, set $\beta_1 = -\sigma_\beta$, $\beta_2 = \sigma_\beta$, and $E_c = 0$. Then, from Eqs. (B6), (B7) and (B8), we have

$$1-\lambda_1=2A_{12}=\sqrt{\frac{a}{2\pi}}\int\frac{\exp\left[-aE^2/2-\sigma_\beta^2/(2a)\right]}{\cosh(\sigma_\beta E)}dE\;\approx\frac{e^{-\sigma_\beta^2\sigma_E^2/2}}{\sqrt{1+\sigma_\beta^2\sigma_E^2}}\left[1+\frac{\sigma_\beta^4\sigma_E^4}{4(1+\sigma_\beta^2\sigma_E^2)^2}\right],\tag{B9}$$

where $\sigma_E=1/\sqrt{a}$, and we have used $1/\cosh x \approx \exp(-x^2/2)(1 + x^4/12)$. This model shows a rapid decrease in the rate of convergence of direct WHAM with the temperature separation.

B.5. Three-temperature case

Similarly, for three evenly-spaced temperatures, we can, without loss of generality, set $\beta_1 = -\Delta\beta$, $\beta_2 = 0$, and $\beta_3 = \Delta\beta$, with $\Delta\beta=\sqrt{\frac{3}{2}}\sigma_\beta$. Then,[4]

$$1-\lambda_1=A_{12}+2A_{13}\;\approx\mathscr{A}\left(\sqrt{\frac{3}{8}}\sigma_\beta\sigma_E,-\frac{5}{2}\right)+2\mathscr{A}\left(\sqrt{\frac{3}{2}}\sigma_\beta\sigma_E,\frac{1}{2}\right),\tag{B10}$$

[4]For A_{12}, we set $E_c = -\Delta\beta/(2a)$, and $D(E;\mathbf{f}^*)=\exp\left(\frac{1}{2}\Delta\beta E\right)J(E)$, with

where

$$\mathscr{A}(b,t)=\frac{e^{-b^2/2}/(1+\frac{1}{2}e^{tb^2})}{2\sqrt{1+b^2/(1+\frac{1}{2}e^{tb^2})}}\left[1+\frac{(2-\frac{1}{2}e^{tb^2})b^4/8}{(1+b^2+\frac{1}{2}e^{tb^2})^2}\right].$$

B.6. Continuous-temperature case

If there are a large number K of temperatures in a finite range, we can approximate them by a continuous distribution, $w(\beta)$. In this case, the sum over temperatures can be replaced by an integral: $\sum_{k=1}^{K} \rightarrow K\int d\beta w(\beta)$.

The eigenvalue λ_l and eigenvector $c_l(\beta)$ are now determined from the integral equation

$$K\int w(\beta')c_l(\beta')A(\beta,\beta')d\beta'=\lambda_l c_l(\beta), \quad \text{(B11)}$$

where $A(\beta, \beta')$ is A_{ij} with $\beta_i \rightarrow \beta$ and $\beta_j \rightarrow \beta'$.

Equation (B11) can be solved in a special case, in which we assume EDA, Eq. (B6), a Gaussian density of states, Eq. (B7), and a Gaussian β distribution:

$$\omega(\beta)=\frac{1}{\sqrt{2\pi a'}}\exp\left[-\frac{(\beta-\beta_c)^2}{2a'}\right],$$

with width $\sigma_\beta=\sqrt{a'}$. The physical solution is

$$\lambda_l=\left(\frac{a'}{a+a'}\right)^l,$$

$$c_l(\beta)=H_l\left(\frac{\beta-\beta_c}{\sqrt{2a'}}\right),$$

$$\begin{aligned} J(E) &\equiv 2\cosh(\frac{1}{2}\Delta\beta E)+\exp(-\frac{3}{2}\Delta\beta E-\frac{1}{\alpha}\Delta\beta^2) \\ &\approx 2\cosh(\frac{1}{2}\Delta\beta E)+\exp\left(-\frac{u\Delta\beta^2}{4\alpha}\right) \\ &\approx 2\theta\exp\left(\frac{\Delta\beta^2E^2}{8\theta}\right)/\left[1+\frac{(3-\theta)\Delta\beta^4E^4}{24\cdot 2^4\cdot\theta^2}\right], \end{aligned}$$

where $\theta=1+\frac{1}{2}\exp[-u\Delta\beta^2/(4a)]$. The parameter u should be 5/2 to match the small Δβ behaviour, although the overall accuracy can be improved by a value of $u \approx 2.11$.

where $H_l(x)$ is the Hermite polynomial[49–51], generated as $e^{-t^2+2xt}=\sum_{l=0}^{\infty}H_l(x)t^l/l!$. Thus, for the second largest eigenvector, we have

$$1-\lambda_1=\frac{1}{1+\sigma_\beta^2\sigma_E^2}, \quad \text{(B12)}$$

which decreases with increasing temperature range, σ_β, albeit more slowly than the two- and three-temperature values, as shown in Fig. B1(b). Thus, Eq. (B12) only gives a upper bound, and nonzero spacing between temperatures can further slow down the convergence. Besides, sampling error, which is ignored in the above calculation, may also slow down the convergence.

References

1. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. The Journal of Chemical Physics. 2008; 129:124105. [PubMed: 19045004]
2. Ferrenberg AM, Swendsen RH. New monte carlo technique for studying phase transitions. Phys Rev Lett. 1988 Dec.61:2635–2638. [PubMed: 10039183]
3. Ferrenberg AM, Swendsen RH. Optimized monte carlo data analysis. Phys Rev Lett. 1989 Sep. 63:1195–1198. [PubMed: 10040500]
4. Newman, E.; Barkema, G. Monte carlo methods in statistical physics. 1st. Oxford: Clarendon Press; 1999.
5. Frenkel, D.; Smit, B. Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science). 2nd. Academic Press; 2001.
6. Kumar S, Rosenberg JM, Bouzida D, et al. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. Journal of Computational Chemistry. 1992; 13:1011–1021.
7. Roux B. The calculation of the potential of mean force using computer simulations. Computer Physics Communications. 1995; 91:275–282.
8. Bartels C, Karplus M. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. Journal of Computational Chemistry. 1997; 18:1450–1462.
9. Gallicchio E, Andrec M, Felts AK, et al. Temperature weighted histogram analysis method, replica exchange, and transition paths. The Journal of Physical Chemistry B. 2005; 109:6722–6731. [PubMed: 16851756]
10. Habeck M. Bayesian reconstruction of the density of states. Phys Rev Lett. 2007 May.98:200601. [PubMed: 17677679]
11. Habeck M. Bayesian estimation of free energies from equilibrium simulations. Phys Rev Lett. 2012 Sep.109:100601. [PubMed: 23005272]
12. Souaille M, Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Computer Physics Communications. 2001; 135:40–57.
13. Chodera JD, Swope WC, Pitera JW, et al. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. Journal of Chemical Theory and Computation. 2007; 3:26–41. [PubMed: 26627148]
14. Bereau T, Swendsen RH. Optimized convergence for multiple histogram analysis. Journal of Computational Physics. 2009; 228:6119–6129.
15. Hub JS, de Groot BL, van der Spoel D. g wham–a free weighted histogram analysis implementation including robust error and autocorrelation estimates. Journal of Chemical Theory and Computation. 2010; 6:3713–3720.

16. Zhu F, Hummer G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. Journal of Computational Chemistry. 2012; 33:453–465. [PubMed: 22109354]
17. Tan Z, Gallicchio E, Lapelosa M, et al. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. The Journal of Chemical Physics. 2012; 136:144102. [PubMed: 22502496]
18. Torrie GM, Valleau JP. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. Chemical Physics Letters. 1974; 28:578–581.
19. Laio A, Parrinello M. Escaping free-energy minima. Proceedings of the National Academy of Sciences. 2002; 99:12562–12566.
20. Marinari E, Parisi G. Simulated tempering: A new monte carlo scheme. EPL (Europhysics Letters). 1992; 19:451.
21. Lyubartsev AP, Martsinovski AA, Shevkunov SV, et al. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. The Journal of Chemical Physics. 1992; 96:1776–1783.
22. Swendsen RH, Wang JS. Replica monte carlo simulation of spin-glasses. Phys Rev Lett. 1986 Nov. 57:2607–2609. [PubMed: 10033814]
23. Geyer, CJ. Proceedings of the 23rd Symposium on the Interface. New York: American Statistical Association; 1991. p. 156
24. Hukushima K, Nemoto K. Exchange monte carlo method and application to spin glass simulations. Journal of the Physical Society of Japan. 1996; 65:1604–1608.
25. Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. Chemical Physics Letters. 1997; 281:140–150.
26. Earl DJ, Deem MW. Parallel tempering: Theory, applications, and new perspectives. Phys Chem Chem Phys. 2005; 7:3910–3916. [PubMed: 19810318]
27. Bennett CH. Efficient estimation of free energy differences from monte carlo data. Journal of Computational Physics. 1976; 22:245–268.
28. Kim J, Keyes T, Straub JE. Communication: Iteration-free, weighted histogram analysis method in terms of intensive variables. The Journal of Chemical Physics. 2011; 135:061103. [PubMed: 21842919]
29. Fenwick MK. A direct multiple histogram reweighting method for optimal computation of the density of states. The Journal of Chemical Physics. 2008; 129:125106. [PubMed: 19045067]
30. Kästner J, Thiel W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: umbrella integration. The Journal of Chemical Physics. 2005; 123:144104. [PubMed: 16238371]
31. Kästner J. Umbrella integration in two or more reaction coordinates. The Journal of Chemical Physics. 2009; 131:034109. [PubMed: 19624183]
32. Pulay P. Convergence acceleration of iterative sequences. the case of scf iteration. Chemical Physics Letters. 1980; 73:393–398.
33. Pulay P. Improved scf convergence acceleration. Journal of Computational Chemistry. 1982; 3:556–560.
34. Hamilton TP, Pulay P. Direct inversion in the iterative subspace (diis) optimization of openshell, excitedstate, and small multiconfiguration scf wave functions. The Journal of Chemical Physics. 1986; 84:5728–5734.
35. Kovalenko A, Ten-no S, Hirata F. Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace. Journal of Computational Chemistry. 1999; 20:928–936.
36. Howard J, Pettitt B. Integral equations in the study of polar and ionic interaction site fluids. Journal of Statistical Physics. 2011; 145:441–466. [PubMed: 22383857]
37. Shen J, McCammon J. Molecular dynamics simulation of superoxide interacting with superoxide dismutase. Chemical Physics. 1991; 158:191–198.

38. Woolf TB, Roux B. Conformational flexibility of o-phosphorylcholine and o-phosphorylethanolamine: A molecular dynamics study of solvation effects. Journal of the American Chemical Society. 1994; 116:5916–5926.
39. Crouzy S, Woolf T, Roux B. A molecular dynamics study of gating in dioxolane-linked gramicidin a channels. Biophysical Journal. 1994; 67:1370–1386. [PubMed: 7529578]
40. Mezei M. Adaptive umbrella sampling: Self-consistent determination of the non-boltzmann bias. Journal of Computational Physics. 1987; 68:237–248.
41. Berg BA, Neuhaus T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. Phys Rev Lett. 1992 Jan.68:9–12. [PubMed: 10045099]
42. Lee J. New monte carlo algorithm: Entropic sampling. Phys Rev Lett. 1993 Jul.71:211–214. [PubMed: 10054892]
43. Tsallis C. Possible generalization of boltzmann-gibbs statistics. Journal of Statistical Physics. 1988; 52:479–487.
44. Yan Q, de Pablo JJ. Fast calculation of the density of states of a fluid by monte carlo simulations. Phys Rev Lett. 2003 Jan.90:035701. [PubMed: 12570508]
45. Martin-Mayor V. Microcanonical approach to the simulation of first-order phase transitions. Phys Rev Lett. 2007 Mar.98:137207. [PubMed: 17501239]
46. Press, WH.; Teukolsky, SA.; Vetterling, WT., et al. Numerical recipes in c: The art of scientific computing. 2nd. Cambridge: Cambridge University Press; 1992.
47. Efron B. Bootstrap methods: Another look at the jackknife. The Annals of Statistics. 1979; 7:1–26.
48. Park S, Pande VS. Choosing weights for simulated tempering. Phys Rev E. 2007 Jul.76:016703.
49. Arfken, GB.; Weber, HJ. Mathematical methods for physicists. 5th. Academic Press; 2001.
50. Abramowitz, M.; Stegun, I. Applied mathematics series. Dover Publications; 1964. Handbook of mathematical functions: With formulas, graphs, and mathematical tables.
51. Wang, Z.; Guo, D. Special functions. World Scientific; 1989.
52. Whittaker, ET.; Watson, GN. A course of modern analysis. 4th. Cambridge University Press; 1927.
53. Ferdinand AE, Fisher ME. Bounded and inhomogeneous ising models. i. specific-heat anomaly of a finite lattice. Phys Rev. 1969 Sep.185:832–846.
54. Beale PD. Exact distribution of energies in the two-dimensional ising model. Phys Rev Lett. 1996 Jan.76:78–81. [PubMed: 10060438]
55. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science. 1998; 282:740–744. [PubMed: 9784131]
56. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry. 2000; 21:1049–1074.
57. Hornak V, Abel R, Okur A, et al. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics. 2006; 65:712–725.
58. Lindorff-Larsen K, Piana S, Palmo K, et al. Improved side-chain torsion potentials for the amber ff99sb protein force field. Proteins: Structure, Function, and Bioinformatics. 2010; 78:1950–1958.
59. Berendsen H, van der Spoel D, van Drunen R. Gromacs: A message-passing parallel molecular dynamics implementation. Computer Physics Communications. 1995; 91:43–56.
60. Lindahl E, Hess B, van der Spoel D. Gromacs 3.0: a package for molecular simulation and trajectory analysis. Journal of Molecular Modeling. 2001; 7:306–317.
61. Van Der Spoel D, Lindahl E, Hess B, et al. Gromacs: Fast, flexible, and free. Journal of Computational Chemistry. 2005; 26:1701–1718. [PubMed: 16211538]
62. Hess B, Kutzner C, van der Spoel D, et al. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation. 2008; 4:435–447. [PubMed: 26620784]
63. Pronk S, Páll S, Schulz R, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013; 29:845–854. [PubMed: 23407358]
64. Páll, S.; Abraham, M.; Kutzner, C., et al. Tackling exascale software challenges in molecular dynamics simulations with gromacs. In: Markidis, S.; Laure, E., editors. Solving software

challenges for exascale. Vol. 8759 of Lecture Notes in Computer Science. Springer International Publishing; 2015. p. 3-27.

65. Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015; 1–2:19–25.
66. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. The Journal of Chemical Physics. 2007; 126:014101. [PubMed: 17212484]
67. Essmann U, Perera L, Berkowitz ML, et al. A smooth particle mesh ewald method. The Journal of Chemical Physics. 1995; 103:8577–8593.
68. Hess B, Bekker H, Berendsen HJC, et al. Lincs: A linear constraint solver for molecular simulations. Journal of Computational Chemistry. 1997; 18:1463–1472.
69. Miyamoto S, Kollman PA. Settle: An analytical version of the shake and rattle algorithm for rigid water models. Journal of Computational Chemistry. 1992; 13:952–962.
70. Leonard, T.; Hsu, J. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press; 1999. Bayesian methods: An analysis for statisticians and interdisciplinary researchers.
71. Duda, R.; Hart, P.; Stork, D. Pattern classification. 2nd. Wiley: 2000.
72. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, Series B. 1977; 39:1–38.
73. Neal, R.; Hinton, G. Vol. 89 of NATO ASI Series. Springer Netherlands; 1998. A view of the em algorithm that justifies incremental, sparse, and other variants; p. 355-368.
74. Press, WH.; Teukolsky, SA.; Vetterling, WT., et al. Numerical recipes in c: The art of scientific computing. 3rd. Cambridge: Cambridge University Press; 2007.

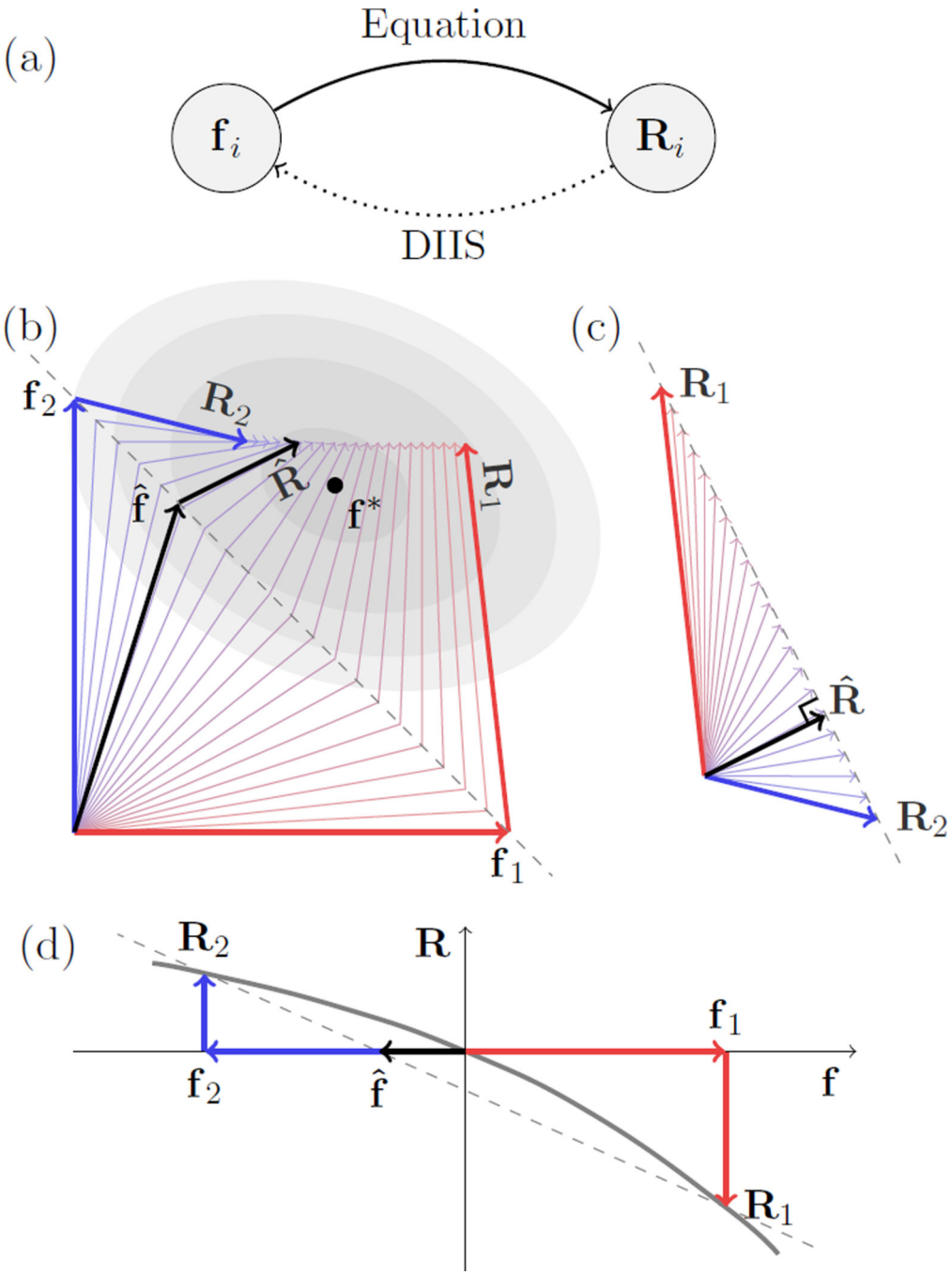

Figure 1.

Schematic illustrations of the method of direct inversion of the iterative subspace (DIIS). (a) As an iterative method, DIIS solves the equation, $\mathbf{R}(\mathbf{f}) = \mathbf{0}$, by a feedback loop. When the equation is applied to a trial solution, represented by vector $\mathbf{f}_i$, we get a residual vector $\mathbf{R}_i \equiv \mathbf{R}(\mathbf{f}_i)$, whose magnitude indicates the error. Ideally, if $\mathbf{f}_i$ were the true solution, $\mathbf{R}_i$ would be zero. In the feedback step, we correct $\mathbf{f}_i$ using $\mathbf{R}_i$. The task of DIIS is to construct from a few previous vectors accumulated during the iteration an optimal trial vector $\hat{\mathbf{f}}$ with hopefully the smallest error $\|\hat{\mathbf{R}}\|$ to help the next round of iteration. (b) and (c) Given a basis of M (two

here) trial vectors, DIIS seeks the combination $\hat{\mathbf{R}}=\sum_{i=1}^{M} c_i \mathbf{R}_i$ that minimises the magnitude $\|\hat{\mathbf{R}}\|$ under the constraint $\sum_{i=1}^{M} c_i=1$ [panel (c)]. The corresponding combination of the trial vectors, $\hat{\mathbf{f}}=\sum_{i=1}^{M} c_i \mathbf{f}_i$, is expected to be close to true solution, **f***. Then, we construct the new trial vector as $\mathbf{f}^{(n)} = \hat{\mathbf{f}} + \hat{\mathbf{R}}$ and use it to update the basis for the next round of iteration. (d) If the vectors are one-dimensional (i.e., scalars), it is possible to find a vanishing $\hat{\mathbf{R}}$, and DIIS is equivalent to the secant method.

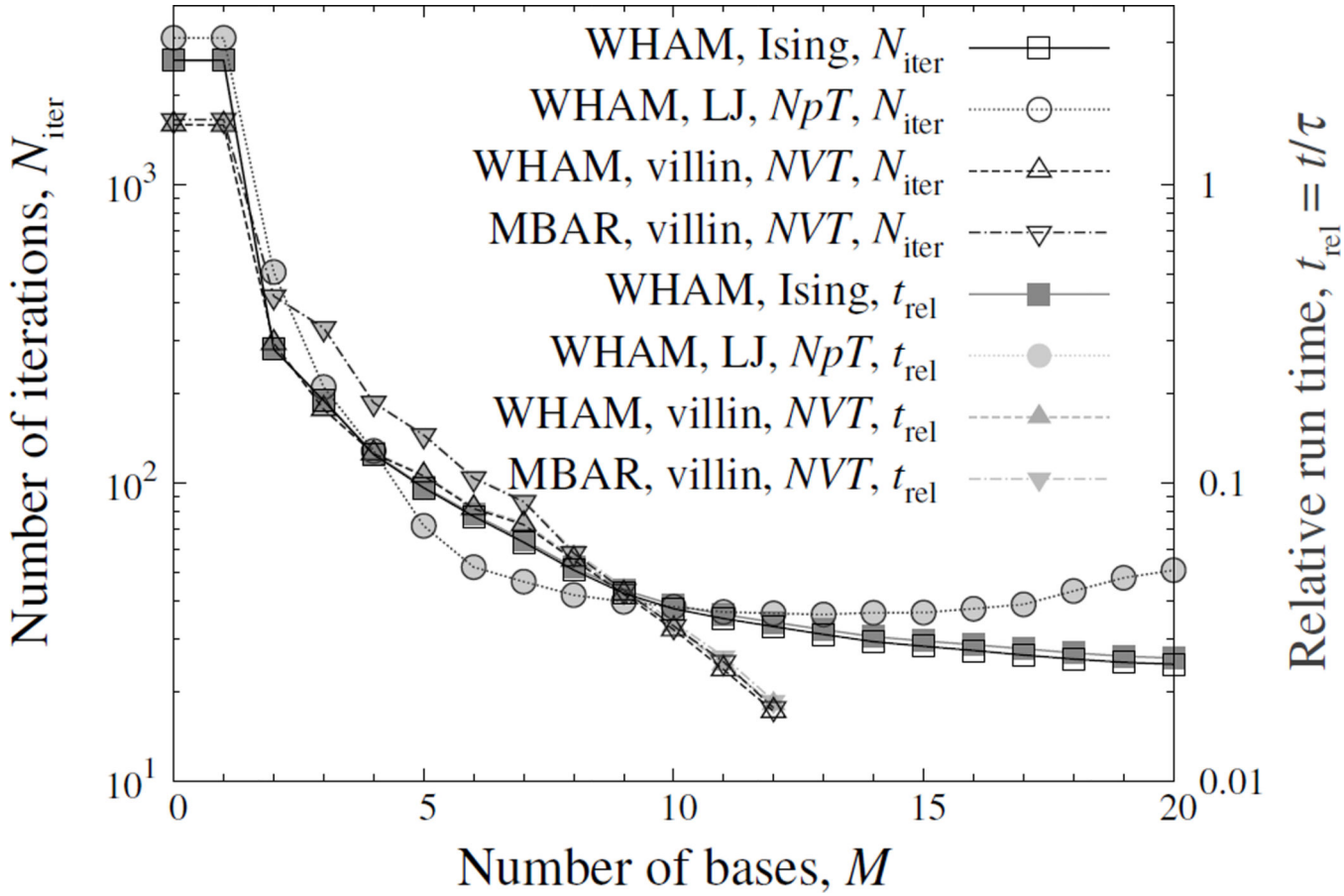

Figure 2.

Number of iterations and run time versus the number of bases, *M*, in DIIS. The four test cases are (1) WHAM on the Ising model, (2) WHAM on the LJ fluid in the *NpT* ensemble, (3) WHAM on the mini-protein villin headpiece in the *NVT* ensemble (with the bin width of energy histograms being 1.0), and (4) MBAR on the same protein system. The $M = 0$ points represent direct WHAM. The run times are inversely scaled by a factor τ for better alignment with the numbers of iterations. The scaling factors τ for cases 1–4 are 12.5, 1.87×10^3, 14.3, and 260 seconds, respectively. The results were averaged over independent samples for the Ising model and LJ fluid. For the villin headpiece, the results were averaged over bootstrap[4, 15, 47] samples for WHAM, or over random subsamples with about 1% of the trajectory frames for MBAR. To mimic the correlation in the trajectories, each data point in the bootstrap sample is either randomly drawn from the trajectory, or the same as the previous one (if any) with probability $\exp(-\Delta t_{MD}/t_{act})$, where Δt_{MD} and t_{act} are the time step of molecular dynamics and the autocorrelation time of the potential energy, respectively. Here, the error tolerance $\max\{|R_i|\}$ is 10^{-8}. The lines are a guide for the eyes.

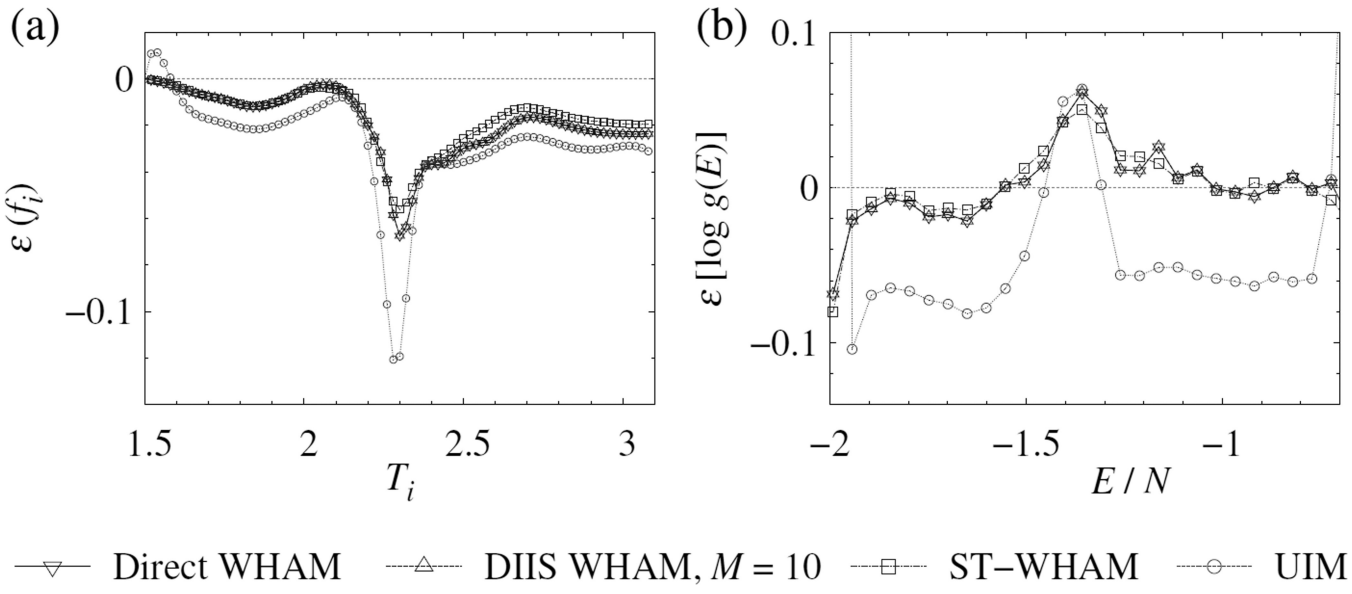

Figure 3.

Errors of (a) the dimensionless free energies, f_i, and (b) the logarithmic density of states, log $g(E)$, for the $N = 64 \times 64$ two-dimensional Ising model (plotted with an energy spacing of 200). Here, $\varepsilon(a) \equiv a - a^{\text{ref}}$, and the reference values for f_i and $g(E)$ were computed using the methods in Refs. [53] and [54], respectively. The bin size of the energy histograms is $\Delta E = 4$. Lines are a guide for the eyes.

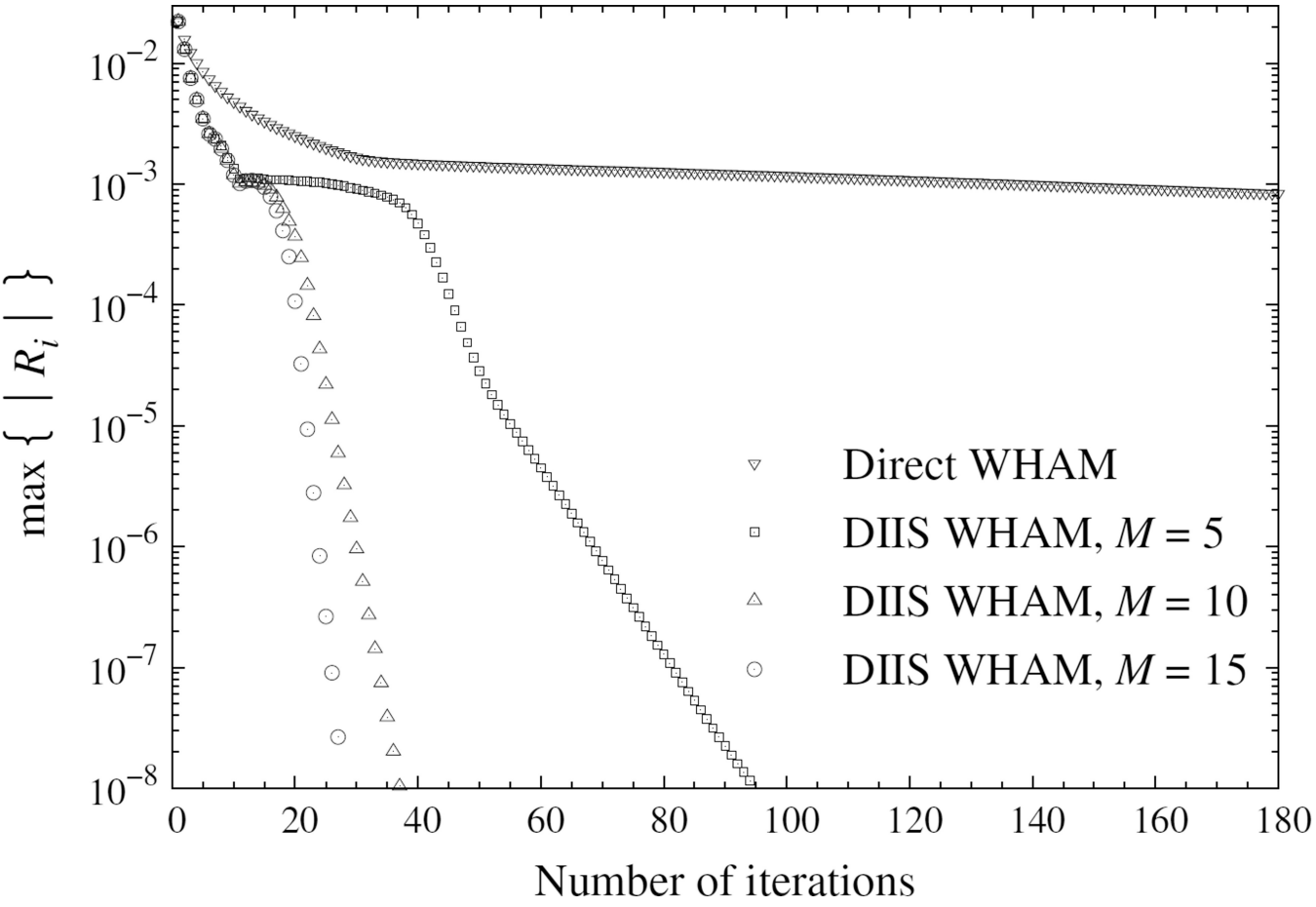

Figure 4.

Convergence error, max{|R_i|}, versus the number of iterations in direct and DIIS WHAMs for the 64 × 64 two-dimensional Ising model. Results were geometrically averaged over independent samples.

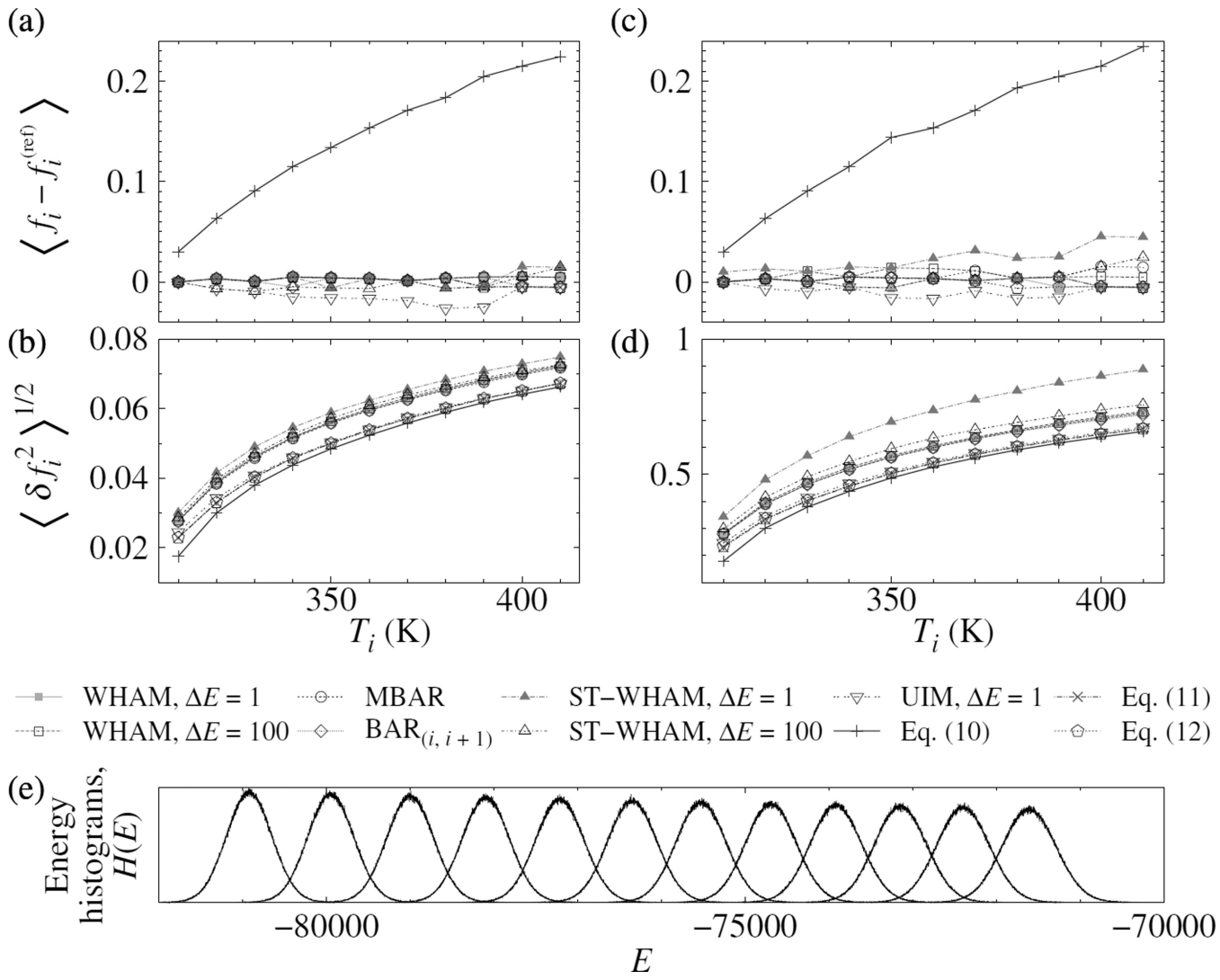

Figure 5.

Accuracy [(a) and (c)] and precision [(b) and (d)] of the dimensionless free energies, f_i, from WHAM, BAR, ST-WHAM, UIM, and approximate formulae for the villin headpiece in solution. Two types of samples were used. A large sample [(a) and (b)] and a small sample [(c) and (d)] contain roughly 1% and 0.01% of all trajectory frames, respectively. The results were averaged over random samples. The MBAR results computed from all trajectory frames were used as the reference. f_i at $T = 300$ K is fixed at zero. The lines are a guide for the eyes. (e) Energy histograms collected with bin width $\Delta E = 1.0$.