

Biophysical Letter

Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions

Fabrizio Marinelli^{1,*} and José D. Faraldo-Gómez^{1,*}¹Theoretical Molecular Biophysics Section, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland

ABSTRACT We introduce an enhanced-sampling method for molecular dynamics (MD) simulations referred to as ensemble-biased metadynamics (EBMetaD). The method biases a conventional MD simulation to sample a molecular ensemble that is consistent with one or more probability distributions known a priori, e.g., experimental intramolecular distance distributions obtained by double electron-electron resonance or other spectroscopic techniques. To this end, EBMetaD adds an adaptive biasing potential throughout the simulation that discourages sampling of configurations inconsistent with the target probability distributions. The bias introduced is the minimum necessary to fulfill the target distributions, i.e., EBMetaD satisfies the maximum-entropy principle. Unlike other methods, EBMetaD does not require multiple simulation replicas or the introduction of Lagrange multipliers, and is therefore computationally efficient and straightforward in practice. We demonstrate the performance and accuracy of the method for a model system as well as for spin-labeled T4 lysozyme in explicit water, and show how EBMetaD reproduces three double electron-electron resonance distance distributions concurrently within a few tens of nanoseconds of simulation time. EBMetaD is integrated in the open-source PLUMED plug-in (www.plumed-code.org), and can be therefore readily used with multiple MD engines.

Received for publication 1 April 2015 and in final form 20 May 2015.

*Correspondence: fabrizio.marinelli@nih.gov or jose.faraldo@nih.gov

Biophysical techniques probing the structural dynamics of biomolecules typically yield signals that arise from an ensemble of molecular conformations, and thus it is often not straightforward to interpret the experimental data unambiguously. For example, double electron-electron resonance (DEER) spectroscopy is increasingly used to measure distances between spin-labeled cysteine residues (1), and to assess conformational mechanisms in proteins. DEER spectra, however, actually translate into distance probability distributions, which are often multimodal and interdependent, and might reflect a variety of protein conformations and rotameric states of the labels.

Molecular dynamics (MD) simulations are arguably the best computational approach to address this problem. The concept is to employ an MD simulation to construct an ensemble of molecular configurations \mathbf{X} that is consistent with the measured probability distribution of an observable $\xi = \xi^f(\mathbf{X})$, while simultaneously representing the molecular system more realistically (solvent, temperature, etc.) than in standard structural-refinement methods. In practice, this approach entails a modification of the simulation energy function, $U(\mathbf{X})$, so that the resulting probability distribution, $\rho(\mathbf{X})$, fulfills the experimental data with the minimum possible bias i.e., the so-called maximum-entropy principle (2,3). If the experimental data for observable ξ is binned into a histogram, a possible modification of $U(\mathbf{X})$ is a linear perturbation, leading to (4)

$$\rho(\mathbf{X}) = \frac{\exp\{-\beta U(\mathbf{X}) + \sum_i \lambda_i h_i[\xi^f(\mathbf{X})]\}}{\int d\mathbf{X}' \exp\{-\beta U(\mathbf{X}') + \sum_i \lambda_i h_i[\xi^f(\mathbf{X}')]\}}, \quad (1)$$

where the i index denotes each of the bins in the measured histogram of ξ , and $h_i[\xi^f(\mathbf{X})] = 1$ if the value of $\xi^f(\mathbf{X})$ is in bin i , while $h_i[\xi^f(\mathbf{X})] = 0$ otherwise (and $\beta = 1/k_B T$, where k_B is the Boltzmann constant and T is the temperature). The λ_i parameters, which must be determined in each case, ensure that the time averages of $h_i[\xi^f(\mathbf{X})]$ are equal to the experimental probability values for each of the bins i . Practical applications of the maximum-entropy formulation in Eq. 1 have so far relied on computationally intensive approaches such as averaging over multiple system replicas simulated concurrently (4–7) or iterative optimization algorithms to determine the values of λ_i (2,3).

Here, we present an alternative, single-replica approach inspired by the metadynamics method (8,9), which is also consistent with the maximum-entropy principle. We refer to this method as ensemble-biased metadynamics (EBMetaD). Let us define $\rho_{\text{exp}}(\xi)$ as the target experimental probability distribution of observable ξ and $F(\xi)$ as the free energy,

Editor: Bert de Groot.

© 2015 by the Biophysical Society

<http://dx.doi.org/10.1016/j.bpj.2015.05.024>

$$F(\xi) = -\frac{1}{\beta} \ln \left[\int d\mathbf{X} \exp\{-\beta U(\mathbf{X})\} \delta(\xi - \xi^f(\mathbf{X})) \right] + C, \quad (2)$$

where C is a constant. In the limit of infinitesimally narrow bins, Eq. 1 becomes (Appendix S1 in the Supporting Material):

$$\rho(\mathbf{X}) = \frac{\exp\{-\beta U(\mathbf{X}) + \ln \rho_{\text{exp}}[\xi^f(\mathbf{X})] + \beta F[\xi^f(\mathbf{X})]\}}{\int d\mathbf{X}' \exp\{-\beta U(\mathbf{X}') + \ln \rho_{\text{exp}}[\xi^f(\mathbf{X}')] + \beta F[\xi^f(\mathbf{X}')]\}}. \quad (3)$$

In EBMetaD, a biasing potential is added to the energy function so that the simulation samples $\rho(\mathbf{X})$ in Eq. 3. Like in standard metadynamics, this biasing potential, denoted by $V(\xi^f(\mathbf{X}), t)$, is constructed throughout the simulation as a cumulative sum of Gaussians, added one at a time at a frequency of $1/\tau$, each centered on the value of ξ at that time. In EBMetaD, however, these Gaussian functions are weighted by the target probability distribution, that is

$$V(\xi, t) = \sum_{t'=\tau, 2\tau, \dots}^t \frac{w \exp\{-[\xi - \xi^f(\mathbf{X}_{t'})]^2/2\sigma^2\}}{\exp\{S_\rho\} \rho_{\text{exp}}[\xi^f(\mathbf{X}_{t'})]}, \quad (4)$$

where $\mathbf{X}_{t'}$ denotes the atomic coordinates at time t' and σ is the Gaussians width, which sets the resolution of $F(\xi)$ and $\rho_{\text{exp}}(\xi)$. The quantity $S_\rho = -\int d\xi \rho_{\text{exp}}(\xi) \ln[\rho_{\text{exp}}(\xi)]$ is the differential entropy of $\rho_{\text{exp}}(\xi)$, i.e., $\exp\{S_\rho\}$ is the effective volume in ξ spanned by $\rho_{\text{exp}}(\xi)$, and serves a normalization factor to ensure that the mean height of the Gaussians added in the range of $\rho_{\text{exp}}(\xi)$ is equal to w . As in standard metadynamics, EBMetaD simulations remain close to equilibrium if w , σ , and τ in Eq. 4 are selected adequately (Appendix S2 in the Supporting Material), and a stationary condition is reached at a certain time t_e after which the biasing potential fluctuates around an average profile that converges asymptotically (10). Specifically, the change of $V(\xi^f(\mathbf{X}), t)$ from this point forward is (11,12)

$$\dot{V}(\xi, t > t_e) \approx \frac{C}{\rho_{\text{exp}}(\xi)} \frac{\exp\{-\beta[F(\xi) + V(\xi, t)]\}}{\int d\xi' \exp\{-\beta[F(\xi') + V(\xi', t)]\}} \approx C, \quad (5)$$

where C is a constant. Provided that the region in which $\rho_{\text{exp}}(\xi) > 0$ is energetically allowed by $U(\mathbf{X})$, the implication of Eq. 5 is that the average biasing potential converges to

$$\bar{V}(\xi, t > t_e) \approx -\frac{1}{\beta} \ln \rho_{\text{exp}}(\xi) - F(\xi). \quad (6)$$

That is, when $t > t_e$, the EBMetaD simulation samples the space of ξ as in the target distribution $\rho_{\text{exp}}(\xi)$.

It is straightforward to generalize this approach to the case of multiple observables ξ_i and probability distributions

thereof, $\rho_{\text{exp}}[\xi_i^f(\mathbf{X})]$, employing a multidimensional biasing potential analogous to that in Eq. 4:

$$V(\xi_1, \xi_2, \dots, t) = \sum_{t'=\tau, 2\tau, \dots}^t \sum_i \frac{w_i \exp\{-[\xi_i - \xi_i^f(\mathbf{X}_{t'})]^2/2\sigma_i^2\}}{\exp\{S_{\rho_i}\} \rho_{\text{exp}}[\xi_i^f(\mathbf{X}_{t'})]}. \quad (7)$$

Owing to the scaling factors S_{ρ_i} , several distributions can be simultaneously targeted even if they have very different effective volumes. The observables ξ_i , however, ought not be a function of each other (2–4).

To test the validity of the EBMetaD method, we first considered the two-dimensional model potential $U(\xi, \xi')$ shown in Fig. 1 A; the corresponding one-dimensional probability distribution $\rho_o(\xi)$, calculated analytically, is shown in Fig. 1 B (gray). We aim to sample instead a hypothetical experimental distribution $\rho_{\text{exp}}(\xi)$, also shown in Fig. 1 B (black). We thus carry out an overdamped Langevin dynamics simulation on the $U(\xi, \xi')$ potential, using EBMetaD to slowly construct the biasing potential $V(\xi, t)$ defined in Eq. 4. As Fig. 1 B shows, the calculated histogram $\rho_{\text{sim}}(\xi)$ evolves gradually until it converges to the target probability distribution. Thereafter, the simulation reaches a stationary condition, and neither $\rho_{\text{sim}}(\xi)$ nor the average bias potential change significantly (Fig. 1 B, inset). To assess whether the ensemble sampled at convergence corresponds to that defined in Eq. 3, i.e., whether EBMetaD indeed fulfills the maximum-entropy principle, we directly compare the calculated simulation histogram $\rho_{\text{sim}}(\xi, \xi')$ with the modified two-dimensional potential, $U(\xi, \xi') - \ln \rho_{\text{exp}}(\xi) - F(\xi)$, calculated analytically ($k_B T = 1$). As Fig. 1 C shows, these distributions match perfectly; that is, the bias introduced so as to reproduce $\rho_{\text{exp}}(\xi)$ does not alter $\rho(\xi')$ for any ξ -value. An extension of this test in which two hypothetical distributions $\rho_{\text{exp}}(\xi)$ and $\rho'_{\text{exp}}(\xi')$ are concurrently targeted further confirms that EBMetaD fulfills the maximum-entropy condition (see Appendix S3 and Fig. S2 in the Supporting Material).

To test EBMetaD in a realistic application, we next considered T4 lysozyme in explicit water (Fig. 2 A, and Weaver and Matthews (13)). Following Roux and Islam (4), three methanethio-sulfonate spin-labels were attached at positions E62C, T109C, and A134C. Experimental distance distributions for each pair of nitroxide groups were obtained via electron spin resonance (ESR)/DEER spectroscopy; data were kindly provided by R. A. Stein and H. S. McHaourab (Vanderbilt University Medical Center, Nashville, TN).

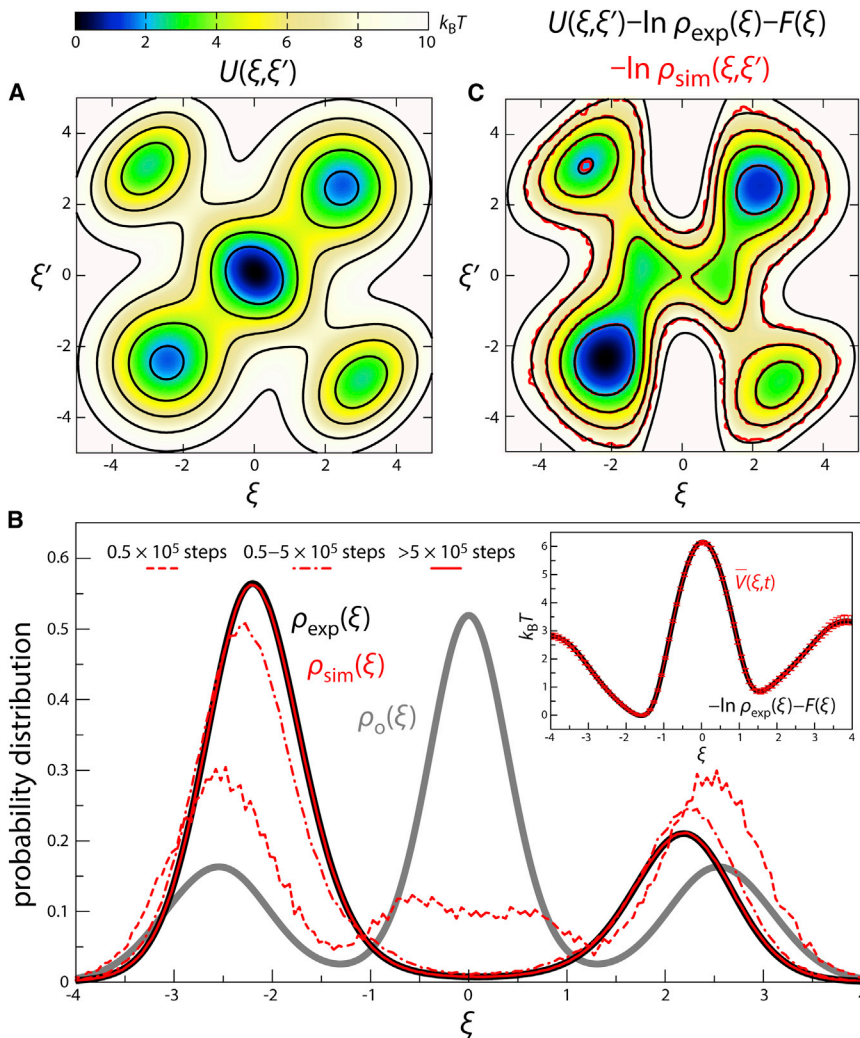


FIGURE 1 (A) Model two-dimensional potential used to test EBMetaD, via an overdamped Langevin dynamics simulation. (B) Histogram of ξ as a function of the number of simulation steps (red lines), compared with the probability distribution associated with the model potential (gray), and with the target distribution (black). (Inset) Average biasing potential, versus $-\ln \rho_{\text{exp}}(\xi) - F(\xi)$ (Eq. 6), with $t_e = 5 \times 10^5$ steps. $F(\xi)$ was calculated analytically, as $F(\xi) = -\ln \int d\xi' \exp\{-U(\xi, \xi')\} + C$. (C) Histogram of ξ and ξ' from EBMetaD (red isolines), overlaid on the ensemble-corrected potential calculated analytically (black isolines) (Eq. 3). Diffusion coefficients in ξ, ξ' were set to 10, the integration time step was 10^{-5} , and $k_B T = 1$. Gaussians of height $10^{-4} k_B T$ and width 0.1 were added every 10^3 steps. Equivalent results were obtained for a wide range of alternative values (Fig. S1).

A single trajectory of ~ 200 ns was then calculated with EBMetaD, using a three-dimensional biasing potential identical to that defined in Eq. 7, i.e., the three experimental distributions are targeted concurrently. For comparison, an

unbiased ~ 270 -ns trajectory was also calculated using a standard MD. As shown in Fig. 2 B, the distance histograms derived from the unbiased trajectory fail to reproduce those obtained experimentally. By contrast, the histograms

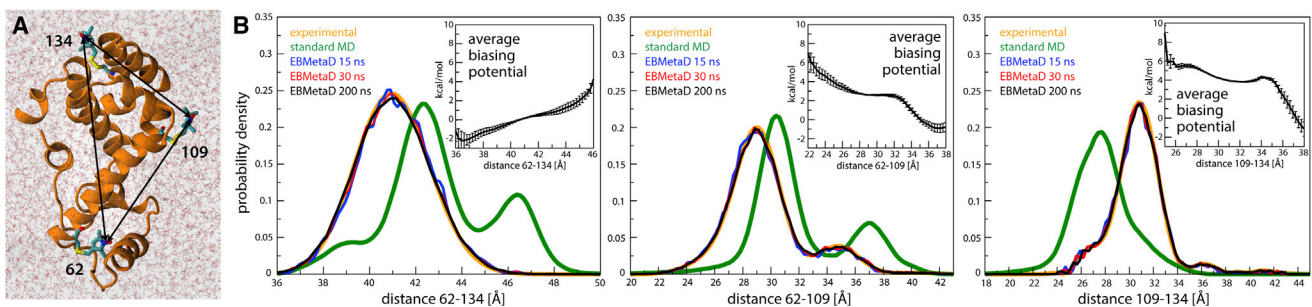


FIGURE 2 (A) Spin-labeled T4 lysozyme simulated in explicit water (PDB:2LZM (13)). The protein is enclosed in a truncated-octahedron periodic box containing 11,895 TIP3P water molecules and 10 Cl^- counterions that neutralize the total charge of the system. The distances between the spin-label nitroxide groups measured by ESR/DEER are indicated (solid arrows). (B) Comparison of the experimental and calculated probability distributions for each of the spin-label pairs, from either unbiased MD simulations or EBMetaD; the latter are given for different simulation times. (Insets) EBMetaD biasing potential, averaged over the simulated trajectory ($t_e = 5$ ns). Error bars are standard errors over three simulation fragments (see Fig. S4 for further details).

derived from the EBMetaD simulation converge to the ESR/DEER data within a few tens of nanoseconds, and preserve that agreement thereafter. To further assess the performance of the method, we compared the time-averaged biasing potential applied to each of the spin-spin distances in three fragments of the EBMetaD trajectory (excluding only the first 5 ns). As shown in Fig. 2 B (insets), the shape of the biasing potentials is largely constant in time, with fluctuations significantly larger than $k_B T$ only in the distal, low-probability regions, thus confirming that EBMetaD reaches an approximately stationary condition (Eqs. 5 and 6). Consistent with the maximum-entropy principle, the ensemble correction introduced by EBMetaD primarily entails a population shift in the rotameric states of the spin labels (Fig. S3 A), with no significant changes in the protein backbone (Fig. S3 B); the root-mean-square deviation of the C α -trace, relative to the starting x-ray structure, is within 2 Å in both the unbiased and EBMetaD trajectories.

In summary, we have introduced an enhanced-sampling MD simulation method to generate molecular ensembles that reproduce probability distributions for one or more independent observables. This method, referred to as ensemble-biased metadynamics, adaptively provides an ensemble correction consistent with the maximum entropy principle (2–6), without mean field approximations (4), multiple simulation replicas (4,5,7), or the iterative optimization of Lagrangian parameters (2,3). Owing to the computational efficiency and practical simplicity of the method, we posit that EBMetaD can be extremely useful in a wide range of applications, such as structure refinement, mechanistic studies based on spectroscopic data, or purely computational simulation studies. EBMetaD is integrated within the PLUMED 1.3 plug-in (14), and can be thus readily used with multiple simulation engines.

SUPPORTING MATERIAL

Three appendices, Supporting Materials and Methods, and four figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(15\)00536-6](http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00536-6).

AUTHOR CONTRIBUTIONS

F.M. and J.D.F.G. designed research and wrote the article. F.M. performed research, contributed analytical tools, and analyzed the data.

ACKNOWLEDGMENTS

This work was funded by the Division of Intramural Research of the National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD.

REFERENCES

- Jeschke, G. 2012. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* 63:419–446.
- Pitera, J. W., and J. D. Chodera. 2012. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* 8:3445–3451.
- White, A. D., and G. A. Voth. 2014. Efficient and minimal method to bias molecular simulations with experimental data. *J. Chem. Theory Comput.* 10:3023–3030.
- Roux, B., and S. M. Islam. 2013. Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron-electron resonance spectroscopy. *J. Phys. Chem. B.* 117:4733–4739.
- Cavalli, A., C. Camilloni, and M. Vendruscolo. 2013. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* 138:094112.
- Roux, B., and J. Weare. 2013. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* 138:084107.
- Islam, S. M., R. A. Stein, ..., B. Roux. 2013. Structural refinement from restrained-ensemble simulations based on EPR/DEER data: application to T4 lysozyme. *J. Phys. Chem. B.* 117:4740–4754.
- Laio, A., and F. L. Gervasio. 2008. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* 71:126601.
- Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA.* 99:12562–12566.
- Bussi, G., A. Laio, and M. Parrinello. 2006. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* 96:090601.
- Barducci, A., G. Bussi, and M. Parrinello. 2008. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100:020603.
- Dama, J. F., M. Parrinello, and G. A. Voth. 2014. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* 112:240602.
- Weaver, L. H., and B. W. Matthews. 1987. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* 193:189–199.
- Bonomi, M., D. Branduardi, ..., M. Parrinello. 2009. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* 180:1961–1972.