# Harmonic Fourier beads method for studying rare events on rugged energy surfaces

Ilja V. Khavrutskii,[a),b)] Karunesh Arora, and Charles L. Brooks III[a),c)]
*Department of Molecular Biology, The Scripps Research Institute, TPC6, 10550 North Torrey Pines Road, La Jolla, California 92037*

We present a robust, distributable method for computing minimum free energy paths of large molecular systems with rugged energy landscapes. The method, which we call harmonic Fourier beads (HFB), exploits the Fourier representation of a path in an appropriate coordinate space and proceeds iteratively by evolving a discrete set of harmonically restrained path points—beads—to generate positions for the next path. The HFB method does not require explicit knowledge of the free energy to locate the path. To compute the free energy profile along the final path we employ an umbrella sampling method in two generalized dimensions. The proposed HFB method is anticipated to aid the study of rare events in biomolecular systems. Its utility is demonstrated with an application to conformational isomerization of the alanine dipeptide in gas phase. © *2006 American Institute of Physics.* [DOI: 10.1063/1.2363379]

## I. INTRODUCTION

Elucidating rare events, specifically, chemical and conformational reorganizations of molecules that require activation energies in excess of the thermally available $k_B T$ value is essential for understanding the mechanics of complex biomolecular systems. Typically, reactant and product states, pertinent to the event in question, are known experimentally, but the underlying transitions remain unresolved. These large-scale and long-time conformational transitions are difficult to study experimentally and require state-of-the-art methods to be employed in computational pursuit.[1,2]

The most prominent computational methods that attempt to delineate transition paths between given reactant and product states are the nudged elastic band,[3] line integral,[4–8] and action based methods[9–12] as well as conjugate peak refinement,[13] transition path sampling,[14,15] and finite temperature string[16–18] methods. Except for transition path sampling and finite temperature string methods, the majority of these approaches relies on gradient-based optimization techniques that require smooth potential energy surfaces to delineate adiabatic transition paths. However, the dynamics of the molecular machines is governed by the free energy that, unlike the potential energy, is not readily available from computer simulations. Therefore, mapping out the transition paths on the free energy surface constitutes a challenge.

Another challenging task is to compute the free energy profile along a given transition path. One common way is to perform constrained dynamics that can then be used to compute the mean forces along the path using either vectorial constraints in the reaction coordinate space or hyperplanes orthogonal to the path.[19–22] Subsequently, the free energy profile can be computed from the corresponding mean forces

along the path. However, practical application of these approaches requires performing complex numerical procedures, particularly in the former case, and is not guaranteed to give reliable free energy estimates.[22–25] Another alternative is to use an umbrella sampling procedure along the reaction coordinate to obtain the free energy profile.[26] However, for biomolecular systems with multidimensional reaction coordinates, umbrella sampling requires substantial sampling in each coordinate dimension and this limits the scope of the method to low dimensional reaction coordinates. In either case, both constrained dynamics and umbrella sampling procedures require a predefined reaction coordinate that accurately describes the transformation in question. However, in large complex molecular systems the reaction coordinate is often difficult to obtain, which further limits the scope of the referred procedures.

In the present paper we provide a simple, robust approach to tackle the study of rare events in complex molecular systems, namely, computing minimum free energy paths and the corresponding free energy profiles. In the core of our approach is a novel method that we call the harmonic Fourier beads (HFB) method.

In what follows, we first describe our HFB method as it pertains to computing minimum free energy paths, and then demonstrate how it can be used to compute the corresponding free energy profiles. Finally, to demonstrate the utility of the HFB method we apply it to study conformational rearrangement of the simplest proteinlike prototype, namely, the alanine dipeptide in gas phase. Future applications of the method to investigate dynamics of larger biological molecules is forthcoming.

## II. METHODOLOGY

The HFB method we present here has some similarities with the recently proposed finite temperature string (FTS)

---
[a)]Authors to whom correspondence should be addressed.
[b)]Electronic mail: ikhavru@scripps.edu
[c)]Electronic mail: brooks@scripps.edu

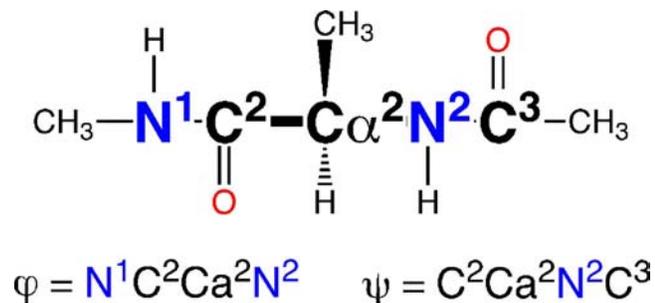$$\varphi = N^1 C^2 C\alpha^2 N^2 \qquad \psi = C^2 C\alpha^2 N^2 C^3$$

FIG. 1. The chemical structure of alanine dipeptide. The essential atoms comprising the RCS are shown in large bold font. These atoms are used in the definition of the $\varphi$ and $\psi$ dihedral angles.

method.[16,27] To provide a basis for comparison with the FTS method, we first describe in detail our method and then discuss its advantages over the FTS method.

## A. The Fourier representation of the path

The Fourier representation of the path—Fourier path—is a key ingredient of the HFB method. The Fourier path describes a curve in a multidimensional coordinate space that is a function of a single progress variable $\alpha \in [0;1]$ and connects a reactant $R(\alpha=0)$ and a product $R(\alpha=1)$, each comprising $N$ atoms. The functional form of the Fourier path includes a linear part in $\alpha$ that corresponds to a trivial line interpolation between the reactant and the product, and the nonlinear part—the Fourier series—that describes the nontrivial features of the path:[9]

$$q_i(\alpha) = q_i(0) + (q_i(1) - q_i(0))\alpha + \sum_{n=1}^{P} b_n^i \sin(n\pi\alpha). \quad (1)$$

In Eq. (1) $q_i$ is the $i$th component of the configuration vector $R=(q_1,\dots,q_{3N})$, also called a bead, describing the positions of all $N$ atoms in Cartesian space, $\{b_n^i\}_{n=1,P}^{i=\overline{1,3N}}$ are the amplitudes of the corresponding Fourier basis functions, and $P$ is the series truncation index that controls the path fidelity. The Fourier path is therefore fully defined by the two end points and a set of the Fourier amplitudes.

## B. Generating a Fourier path

To generate a Fourier path we compute the Fourier amplitudes by performing Fourier transform of a discrete set of total $K$ initial beads $\{R_k^{\text{init}}\}_{k=\overline{1,K}}$ that progresses from the reactant to the product. The coordinates of the beads are prepared for the Fourier transform according to the following procedure.[8]

First, a subset of atoms (say, $S<N$) that is representative of the reaction coordinate space (RCS) of interest is chosen. The remaining atoms ($N-S$) comprise the spectator coordinate set (SCS) that is orthogonal to the RCS. For example, to describe the conformational rearrangements of the 22 atom ($N=22$) alanine dipeptide (see Fig. 1), we choose the five backbone atoms ($S=5$) depicted in bold in Fig. 1 that define the $\varphi$ and $\psi$ dihedral angles for the RCS, leaving 17 atoms ($N-S=17$) in the SCS. The coordinates of the individual beads are then translated into the center of mass of the RCS,

and then rotated to superpose the RCS of all the consecutive beads onto the RCS of the reactant. The resulting sequence of aligned structures $\{R(\alpha_k)=(q_1(\alpha_k),\dots,q_{3N}(\alpha_k)) = W^{k,1}R_k^{\text{init}}\}_{k=\overline{1,K}}$ is then used to compute the Fourier path amplitudes according to Eq. (2) on an even-spaced grid with $\alpha_k=(k-1)/(K-1)$ using a simple trapezoid rule.[28] Here $W^{k,1}$ is the $3\times3$ matrix describing the alignment transformation of the Cartesian coordinates of bead $k$ onto bead 1 (the reactant),

$$b_n^i = 2\int_0^1 [q_i(\alpha) - q_i(0) - (q_i(1) - q_i(0))\alpha]\sin(n\pi\alpha)d\alpha. \quad (2)$$

Note that only the first $P \leq K$ amplitudes and only for the atoms in the RCS need to be determined.

Once the Fourier amplitudes are computed, the beads are redistributed along the path curve according to a chosen metric. In the present paper we use arc length as the metric. The redistribution, therefore, ensures that the arcs between the adjacent beads in the RCS are of equal length. This is done by computing the length of the path curve as a function of $\alpha(L(\alpha))$ using an analytical expression for the differential line segment and then finding the values of $\alpha$ that split the curve in RCS into segments of equal length by simple bracketing and bisection.[28] The solution vector, $\{\alpha_k^{\text{eql}}\}_{k=\overline{1,K}}$, along with the set of Fourier amplitudes are then used in Eq. (1) to generate the corresponding Cartesian coordinates of the reference beads $\{R(\alpha_k^{\text{eql}})\}_{k=\overline{1,K}}$ for the subsequent evolution of the path (the "eql" superscript indicates that the values correspond to the points on the Fourier path that split the path curve in RCS into segments of equal length, or, simply, redistributed points).

## C. Evolving the Fourier path

### 1. Minimum free energy path

The HFB method evolves the Fourier path toward the minimum free energy path by running independent, short molecular dynamics (MD) simulations for each bead $k$ with a corresponding absolute positional harmonic restraint,

$$V(R;\alpha_k^{\text{eql}}) = \frac{f}{M_S}\sum_{j=1}^{S}\left(m_j\sum_{i=3j-2}^{3j}(q_i - q_i(\alpha_k^{\text{eql}}))^2\right). \quad (3)$$

For each bead $k$ the restraint is applied with respect to the corresponding reference bead $R(\alpha_k^{\text{eql}})$ and only to the RCS atoms. In Eq. (3), $f$ is the force constant and $m_j$ and $M_S$ are the mass of the $j$th atom and the total mass of the atoms in the RCS. The fact that each bead is evolved completely independent of all the other beads makes the HFB method efficiently distributable. Each independent simulation generates a MD trajectory that is used to compute the dynamics-averaged coordinates of the corresponding bead. The resulting dynamics-averaged coordinates represent the evolved beads and are used to generate a successive Fourier path. Normally, one has to wait for all the $K$ MD simulations corresponding to the $K$ beads to be completed before generating the next Fourier path.

### 2. Minimum potential energy path

The Fourier path can also be used to compute the minimum potential energy path. In this case, optimization techniques are used to evolve the beads instead of MD simulations, while keeping the same harmonic restraint (3). The final optimized structures, instead of the dynamics-averaged structures, serve as the evolved beads to generate the next Fourier path.

### 3. Convergence of the Fourier path evolution

Regardless of the evolution technique, the coordinates of the evolved beads will exhibit an offset from the corresponding reference beads used in the harmonic restraint (3) in a direction orthogonal to the Fourier path until a corresponding energy valley is reached. The direction and, in particular, the magnitude of the offset vector depend on the value of the force constant used in the harmonic restraint (3).

The convergence of the Fourier path evolution is monitored using the root-mean-square deviation (RMSD) of the pairwise RMSDs computed in the RCS between the beads in the newly evolved and the preceding Fourier paths. Other ways of monitoring convergence of the path will be discussed elsewhere.[29] The final Fourier path, corresponding to either minimum free energy or minimum potential energy path, potentially describes a rather complex transformation in a multidimensional RCS while remaining an analytical function of a single progress variable $\alpha$.

### D. Computing the free energy profile along a Fourier path

Once the minimum free energy path is obtained, the free energy profile underlying the transition between the reactant and the product along the corresponding Fourier path may be computed by performing standard umbrella sampling simulations with two simultaneous best-fit RMSD restraints (4) along the path. Similar to the Fourier path evolution, collecting statistics with umbrella sampling requires running a MD simulation with the two restraints (4) for each bead,

$$V(R;\alpha_k^{\text{eql}};\alpha_{\text{end}}^{\text{eql}}) = \frac{f}{M_S}\left( \sqrt{\sum_{j=1}^{S} m_j \sum_{i=3j-2}^{3j} (W^{k,\text{end}}q_i - q_i(\alpha_{\text{end}}^{\text{eql}}))^2} - \text{RMSD}(\alpha_k^{\text{eql}};\alpha_{\text{end}}^{\text{eql}})\sqrt{M_S} \right)^2. \tag{4}$$

Because such a MD simulation for a particular bead is completely independent of all the other beads, the umbrella sampling procedure is also efficiently distributable. The two restraints (4) are applied to each bead in the RCS using the corresponding reference bead from the optimized Fourier path to compute the offset values (5) with respect to the reactant and the product reference bead, respectively:

$$\text{RMSD}(\alpha_k^{\text{eql}};\alpha_{\text{end}}^{\text{eql}})$$
$$= \sqrt{\frac{\sum_{j=1}^{S} m_j \sum_{i=3j-2}^{3j} (W_{\text{eql}}^{k,\text{end}}q_i(\alpha_k^{\text{eql}}) - q_i(\alpha_{\text{end}}^{\text{eql}}))^2}{M_S}}. \tag{5}$$

In Eqs. (4) and (5), "end" designates either the reactant or the product state; $W^{k,\text{end}}$ is the $3 \times 3$ matrix describing the RMSD best-fit transformation of the Cartesian coordinates of the current dynamic bead $k$ onto the reference bead end and has to be recomputed at every dynamics step; $W_{\text{eql}}^{k,\text{end}}$ is the corresponding transformation for the reference bead $k$ onto the reference bead end and, hence, has to be computed only once (recall that eql refers to the structures corresponding to the reference beads that are equidistant from each other in the RCS). Due to the nature of the applied restraints, it is imperative to initiate the simulations in the vicinity of the corresponding reference bead to ensure a smooth transition and significant overlap between adjacent beads. In cases where the density of points used during the path optimization does not provide the necessary overlap in umbrella sampling, more beads can be trivially inserted using Eq. (1).

The umbrella sampling procedure with the specified restraints (4) for computing the free energy profile reduces the multidimensional RCS to only two generalized RMSD coordinates,[8]

$$\text{RMSD}^{\text{fit}}(R;\alpha_{\text{end}}^{\text{eql}})$$
$$= \sqrt{\frac{\sum_{j=1}^{S} m_j \sum_{i=3j-2}^{3j} (W^{k,\text{end}}q_i - q_i(\alpha_{\text{end}}^{\text{eql}}))^2}{M_S}}. \tag{6}$$

Furthermore, this procedure only requires sampling along the Fourier path projection on the two-dimensional (2D) RMSD space, and is, therefore, effectively one dimensional. Once enough statistics are gathered, the free energy strip along the Fourier path projection is reconstructed using the 2D weighted histogram analysis method (WHAM).[30,31] If only the relative free energies between the reactant and product are of interest, the path along which the free energy profile is computed does not have to be the minimum free energy path. Thus, the procedure for computing the free energy profiles can be applied to any path to yield the corresponding free energy profile.

For complex paths that encompass multiple intermediates and transition states the reduced 2D representation may yield a path curve projection that will cross itself, possibly multiple times. In such cases, the path should be subdivided into consecutive segments between the free energy minima such that the crossings are avoided. The profile of the averaged potential energies over the corresponding evolution tra-

jectories can be used to pinpoint approximate positions of suitable free energy minima in this case. An alternative solution for computing the free energy profiles along the self-crossing paths will be discussed elsewhere.[29]

### E. Practical considerations for the use of the harmonic Fourier beads method

#### 1. Initializing the Fourier path evolution

To initialize the Fourier path evolution toward the minimum free energy path, different strategies can be employed. See, for example, recent works that detail generating initial paths.[8,32–34] Alternatively, one can initiate path evolution using an activated HFB evolution approach. In the activated HFB evolution the reactant and the product reference beads are initially connected with a straight line, comprising a total of $K$ structures; however, only the end point beads and a few beads immediately adjacent to them that lack steric clashes are activated for evolution. The activated beads farthest from each end inward are bridged with a new line segment during reparametrization to maintain the total number of $K$ beads in the path. The structures at both ends of the line segment are considered for activation based on their potential energy. As more structures become activated, this line segment shrinks and eventually disappears. Although the beads on the line segment are not used in the evolution, they ensure that the distances between the beads remain consistent with the complete path. If the arc length between the adjacent beads becomes too large, more beads can be inserted at any redistribution step. The activated HFB evolution achieves the same goal as the growing string method.[17]

#### 2. The choice of the force constant and step size

Some consideration for the choice of force constant and the step size along the offset vector is required to optimize the performance of the HFB method. If the chosen force constant is too small, the beads will slide down to the nearest local minima and the Fourier path for the next step will comprise local minima interconnected with straight lines, with no information about the transition states involved. The choice of a force constant that is too stiff will greatly impede evolution to the minimum free energy valley. Although employing relatively stiff force constants permits fairly small structural changes in the RCS, the resulting offset directions can be used to generate a new set of structures by stepping a certain length along each offset vector. We are currently investigating different options that will allow enhancing the HFB method performance in the future.[29] In this work we use the evolved, mean bead positions directly as an initial guess to compute the Fourier amplitudes for a successive Fourier path. The evolution procedure is iterated until converged, yielding the minimum free energy path.[27,29]

#### 3. Choosing the best starting structure

For some choices of the RCS/SCS partition, averaging and redistribution procedures might scramble positions of the SCS atoms in the newly generated reference beads. Although scrambling of the SCS atoms does not affect the restraint, it makes the reference bead a poor starting structure for the MD simulations in the next evolution step. In such cases, the best option for starting the next evolution step is to use the last configuration or, even better, the restart file from the previous MD trajectory for each bead.

#### 4. Avoiding kinks along the path

In the flat regions of the free energy surface the path curve may develop kinks, especially if the series truncation index $P$ is close to the number of structures in the path $K$. This problem can be resolved by pairwise averaging of the adjacent reference structures and then reparametrizing the path curve or, alternatively, by simply decreasing the index $P$.

### F. Comparison with the FTS method

We are now in a position to compare our HFB method with the FTS method. In the FTS method a path between a reactant and a product is represented by a parametric curve in a multidimensional coordinate space using local piecewise polynomial basis functions—splines.[17] In contrast, our HFB method employs global sinusoidal Fourier basis functions.[9,35] The FTS method interpolates a discrete set of structures between the reactant and the product to generate the path curve that is then reparametrized according to a chosen metric. The HFB method also involves these essential steps, however, the underlying procedures become particularly trivial in the Fourier basis, as they involve a conventional Fourier transform.

Another advantage of the Fourier representation is that it provides a systematic way of controlling the fidelity of the path by controlling the number of the basis functions in the series. Such a simple control does not increase the complexity of the computer code and enhances the stability of the HFB method during the evolution stage.

The FTS method evolves its path curve by running constrained dynamics in the hyperplanes normal to the curve at discrete points along it. However, using such constrained dynamics often requires additional constraints to further localize sampling in the planes and makes the method fairly unstable. To combat the instability, the FTS method necessitates a curve-smoothing step. In contrast, the HFB method employs harmonic absolute and/or best-fit RMSD restraints that avoid the problems associated with the hyperplanes and thus provides greater robustness. Furthermore, using the best-fit RMSD restraints, unlike hyperplane constraints, allows one to exclude unproductive configurations associated with rotations and translations of molecules if necessary.[22]

Both the FTS and HFB methods accept various pathways or even trajectories from a variety of methods as an initial guess for the path. For example, they can be used to anneal pathways generated at high temperatures, such as protein unfolding pathways. Yet if no other alternatives for the initial path exist, the HFB method can grow the path curve into an unknown region of configuration space, similar to the growing string method.[17]

The HFB method has all the features required for studying rare events in complex systems with rough energy land-

scapes. It does not require explicit knowledge of the free energy or that the potential energy surface be smooth; it can generate an initial path and does not need an *a priori* defined reaction coordinate. Furthermore, the HFB method is simple, flexible, robust, and naturally distributable. These properties make the HFB method particularly attractive for locating minimum free energy paths in biomolecular systems. Complemented with the proposed method for computing the free energy surface strip in a reduced 2D space of generalized RMSD coordinates, our approach provides quantitative free energy information essential for computing reaction rates from transition state theory.

Interestingly, after giving the name harmonic Fourier beads to our method we discovered that a similar name, bead-Fourier approximation, has been used in the context of quantum mechanical path integral simulations.[36,37] Despite the similarity of the name and that of the classical isomorph of the referred method, our method is significantly different in that we do not approximate the path as a dynamic trajectory as required by the path integral approach. As such we do not have to couple the evolution of the beads and the curve together that would make the method impractical for the proposed applications.

In the following paragraphs we illustrate the application of the HFB/umbrella sampling approach by studying a conformational isomerization of the alanine dipeptide in gas phase.

## III. COMPUTATIONAL DETAILS

The HFB method was implemented into the c33a2 version of the CHARMM program.[38] All simulations reported in this paper were performed with the CHARMM22 all-atom force field.[39] For visualization purposes, the isomerization pathways of the alanine dipeptide were projected onto the $\varphi/\psi$ adiabatic potential energy surface computed as described previously.[8] Here we only considered the transition of the alanine dipeptide from the $C_{7eq}$ to the $C_{7ax}$ conformation in the gas phase. Both electrostatic and van der Waals (vdW) interactions employ 16 Å cutoffs that are truncated with switching functions at 18 Å. MD was run with the leap frog integrator using a 2 fs time step and the velocities were reassigned from the Boltzmann distribution centered at $T = 298$ K. All bonds involving hydrogen atoms were constrained using SHAKE (Refs. [40–42]) with tolerance of $1.0 \times 10^{-8}$ Å.

### A. Generating an initial Fourier path and its evolution toward the minimum free energy path

To initialize the Fourier path between the $C_{7eq}$ and $C_{7ax}$ conformations of the alanine dipeptide, the intermediate beads were generated by a linear interpolation in Cartesian coordinate space. The number of beads $K$ and the Fourier series truncation index $P$ were set to 32, unless noted otherwise. To evolve the Fourier path toward the minimum free energy path, the beads were initially optimized subject to the harmonic restraint (3) on atoms in the RCS and SCS with a mass normalized force constant ($f/M_s$) of 25.0 and 10 kcal/(g Å$^2$), respectively. The optimization employed

1000 steps of steepest descent (SD) and subsequently 1000 steps of adaptive basis Newton-Raphson (ABNR) optimizations. During the minimization process the restraint on the atoms in the SCS was relaxed from 10.0 to 0.0 kcal/(g Å$^2$) in 1.0 kcal/(g Å$^2$) decrements. The Fourier path evolution toward the minimum free energy path was run with MD simulations following the minimization. During the dynamics simulations a harmonic restraint identical to that used in minimization of the RCS atoms was employed with respect to the appropriate reference structure. Each bead was equilibrated for 200 ps, while heating the system to the target temperature $T = 298.0$ K. A production run of 8 ns length followed the equilibration.

To evolve the Fourier path toward the minimum potential energy path rather than the minimum free energy path only minimization of the beads was performed as described above, except that the mass normalized force constant on the RCS atoms was set to 50 kcal/(g Å$^2$). In this case the minimized structures were used as evolved beads to generate the subsequent Fourier path.

The evolution process in the search for either the minimum free energy path or the minimum potential energy path is iterated until the root-mean-square deviation of the pairwise RMSDs in the RCS between the reference structures in the newly evolved and preceding paths reaches the convergence threshold of $5.0 \times 10^{-4}$ Å.

### B. Computing the free energy profile along the Fourier path

To compute the free energy profile along the minimum free energy path we initially minimized the SCS of the final reference beads from the optimized Fourier path as described above, and then equilibrated them for 20 ps using the restraint (3) with the corresponding reference bead. We then reequilibrated each bead for another 10 ps using two restraints (4) simultaneously, one with respect to the reactant and the other with respect to the product. The offset values (5) for each bead were computed as the RMSD of the reference bead in each window with respect to the reactant and the product beads using the Fourier path that corresponds to the minimum free energy path. These latter restraints were employed during the production run of 20 ns in each window. The mass normalized force constants for the restraints were 25 kcal/(g Å$^2$), identical to that used during minimum free energy path optimization. The coordinates for the subsequent potential of mean force calculations were saved every 200 fs. The WHAM (Refs. [30] and [31]) was used to obtain the two-dimensional free energy profile along the final Fourier path.

## IV. RESULTS AND DISCUSSIONS

To demonstrate the utility of our method we investigated the conformational isomerization of the 22-atom alanine dipeptide (see Fig. 1) in gas phase. Specifically, we located both the minimum potential energy path and minimum free energy path between the following conformations: the $C_{7eq}$ at $[-81.4; 70.5]$ ($E = 0.0$ kcal/mol) and the $C_{7ax}$ at $[69.7; -67.6]$ ($E = 2.1$ kcal/mol). The path between the $C_{7eq}$ and
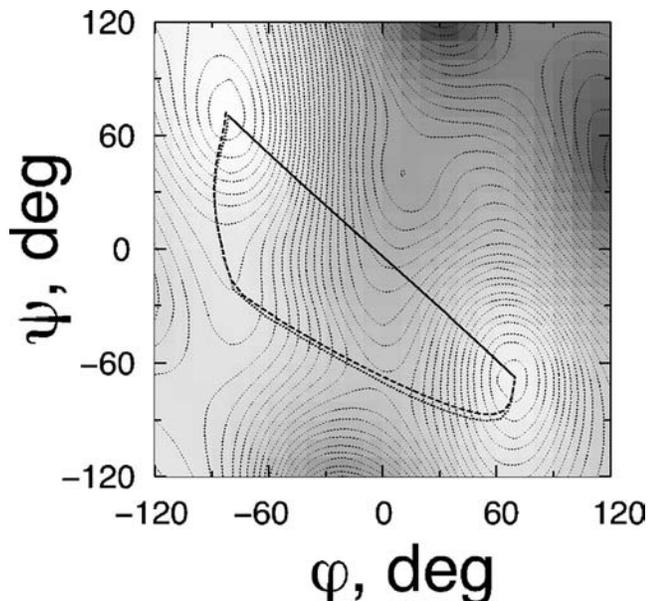
FIG. 2. The projections of the minimum potential and free energy paths onto the $\varphi/\psi$ adiabatic map of the alanine dipeptide. The initial path is shown as a solid line, whereas the minimum potential and minimum free energy paths are shown with short- and long-dashed lines, respectively. See the description in the text for the path optimization details.

$C_{7ax}$ used 32 beads. The RCS included only five atoms ($S = 5$) that define the backbone dihedral angles $\varphi$ and $\psi$.

Figure 2 shows the final minimum potential energy path and minimum free energy path optimized with the HFB method, starting from a linear interpolation in Cartesian coordinate space. The minimum free energy path converges within 60 iterations using MD simulations at $T = 298.0$ K and the mass normalized force constant in the restraint (3) of 25.0 kcal/(g Å$^2$). The final free energy minima of the $C_{7eq}$ and $C_{7ax}$ conformations are at $[-82.6; 72.5]$ and at $[69.7; -69.1]$, respectively. The minimum potential energy path converges in 80 Fourier path evolution steps using optimization that corresponds to $T = 0.0$ K and the force constant of 50.0 kcal/(g Å$^2$). Stiffer force constants $[75.0-125.0$ kcal/(g Å$^2$)] appeared to produce kinks, but the string evolved to the same minimum energy valley (results not shown). Doubling the number of beads in the path from 32 to 64 and increasing the force constant to 125.0 kcal/(g Å$^2$) yielded a path practically identical to that with 32 beads and a force constant of 50.0 kcal/(g Å$^2$) (results not shown). The minimum potential energy path and minimum free energy path differ only slightly as can be seen from Fig. 2. Neither path is perpendicular to the adiabatic constant energy contours. Interestingly, the minimum potential energy path between the same reactant and product optimized in the "essential" internal coordinate space of $\varphi$ and $\psi$ (Ref. 41) is similar to the minimum potential energy path optimized in the Cartesian RCS in the present paper.

Figure 3 depicts the free energy strip along the minimum free energy path. As seen from Fig. 3 the path proceeds through the free energy valley as expected. The relative free energy of the two minima is 2.5 kcal/mol and the free energy of the activation is 8.4–8.6 kcal/mol. These values compare well with the corresponding potential energy profile values of 2.1 and 8.4 kcal/mol, respectively.
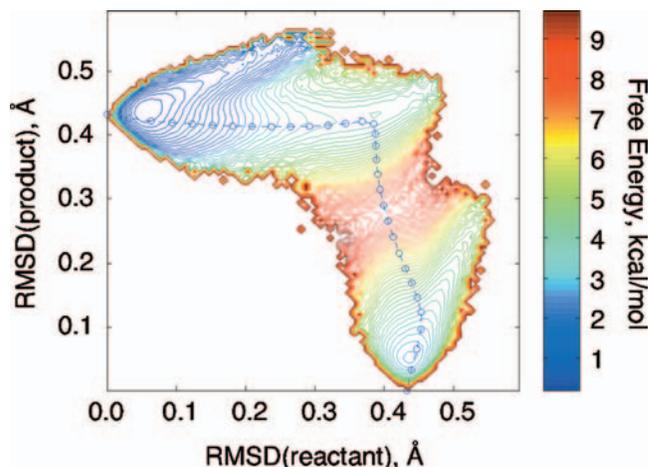


FIG. 3. (Color) The free energy strip along the minimum free energy path. The line with the circles depicts the optimized minimum free energy path in the generalized RMSD coordinates. Each circle provides the position of the actual bead used during umbrella sampling simulations. The path corresponds to $T = 298$ K.

## V. CONCLUDING REMARKS

In conclusion, we have presented a novel approach to computing minimum free energy paths and corresponding free energy profiles. Our approach features the HFB method for the minimum free energy path optimization and employs well-established umbrella sampling in a generalized 2D space along the optimized path to compute the corresponding free energy profile. The HFB method includes all the features required for studying rare events in complex systems with rough energy landscapes. It does not require an explicit knowledge of the free energy and that the potential energy surface be smooth; it can itself generate an initial path and does not need an *a priori* defined reaction coordinate. Complemented with the procedure for computing the free energy surfaces in the reduced 2D space of generalized RMSD coordinates, the HFB method provides quantitative free energy information essential for computing reaction rates with the transition state theory. Furthermore, the HFB method is simple, flexible, robust, and naturally distributable. These properties make the HFB method particularly attractive for studying biomolecular systems.

The utility of the method has been demonstrated on the case of a simple proteinlike system—the alanine dipeptide. Further applications and enhancements of the method are currently in progress in our laboratory and will be reported elsewhere.[29] We believe that the HFB method will expand our capabilities in studying rare events in large complex systems with rugged energy landscapes, and in particular, biomolecular systems.

## ACKNOWLEDGMENTS

[1] M. Karplus and J. Kuryian, Proc. Natl. Acad. Sci. U.S.A. **102**, 6679 (2005).

[2] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer-Verlag, New York, 2002).

[3] G. Henkelman and H. Jonsson, J. Chem. Phys. **113**, 9978 (2000).

[4] R. Elber and M. Karplus, Chem. Phys. Lett. **139**, 375 (1987).

[5] S. Huo and J. E. Straub, J. Chem. Phys. **107**, 5000 (1997).

[6] S. Huo and J. E. Straub, Proteins **36**, 249 (1999).

[7] G. Domotor, L. Stacho, and M. I. Ban, J. Mol. Struct.: THEOCHEM **501–502**, 509 (2000).

[8] I. V. Khavrutskii, R. H. Byrd, and C. L. Brooks III, J. Chem. Phys. **124**, 194903 (2006).

[9] A. E. Cho, J. D. Doll, and D. L. Freeman, Chem. Phys. Lett. **229**, 218 (1994).

[10] R. Elber, Curr. Opin. Struct. Biol. **15**, 151 (2005).

[11] K. Arora and T. Schlick, Chem. Phys. Lett. **378**, 1 (2003).

[12] R. Elber, A. Cárdenas, A. Ghosh, and H. Stern, Adv. Chem. Phys. **126**, 93 (2003).

[13] S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).

[14] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).

[15] R. Radhakrishnan and T. Schlick, Proc. Natl. Acad. Sci. U.S.A. **101**, 5970 (2004).

[16] E. Weinan, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).

[17] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, J. Chem. Phys. **120**, 7877 (2004).

[18] W. Ren, Commun. Math. Sci. **1**, 377 (2003).

[19] R. Elber, J. Chem. Phys. **93**, 4312 (1990).

[20] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).

[21] G. H. Johannesson and H. Jonsson, J. Chem. Phys. **115**, 9644 (2001).

[22] K. N. Kudin and R. Car, J. Chem. Phys. **122**, 114108 (2005).

[23] N. Go and H. A. Scheraga, Macromolecules **9**, 535 (1976).

[24] M. Fixman, J. Chem. Phys. **69**, 1527 (1978).

[25] M. Fixman, Proc. Natl. Acad. Sci. U.S.A. **71**, 3050 (1974).

[26] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).

[27] E. Weinan, W. Ren, and E. Vanden-Eijnden, J. Phys. Chem. B **109**, 6688 (2005).

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge University Press, Port Chester, NY, 2001).

[29] I. V. Khavrutskii and C. L. Brooks III (in press).

[30] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).

[31] E. M. Boczko and C. L. Brooks III, J. Phys. Chem. **97**, 4509 (1993).

[32] O. Miyashita, P. G. Wolynes, and J. N. Onuchic, J. Phys. Chem. B **109**, 1959 (2005).

[33] F. Tama, O. Miyashita, and C. L. Brooks III, J. Struct. Biol. **147**, 315 (2004).

[34] O. Miyashita, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **100**, 12570 (2003).

[35] R. Olender and R. Elber, J. Chem. Phys. **105**, 9299 (1996).

[36] P. N. Vorontsov-Velyaminov, R. I. Gorbunov, and S. D. Ivanov, Comput. Phys. Commun. **121–122**, 64 (1999).

[37] S. D. Ivanov, A. P. Lyubartsev, and A. Laaksonen, Phys. Rev. E **67**, 066710 (2003).

[38] A. D. MacKerell, Jr., B. R. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, in *The Encyclopedia of Computational Chemistry*, edited by P. v. R. Schleyer, P. R. Schreiner, N. L. Allinger, T. Clark, J. Gasteiger, P. Kollman, I. Henry, and F. Schaefer (Wiley, Chichester, 1998).

[39] A. D. MacKerell, Jr., D. Bashford, M. Bellott *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[40] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[41] T. Lazaridis, D. J. Tobias, C. L. Brooks III, and M. E. Paulaitis, J. Chem. Phys. **95**, 7612 (1991).

[42] D. J. Tobias and C. L. Brooks III, J. Chem. Phys. **89**, 5115 (1988).