Flooding in GROMACS: Accelerated Barrier Crossings in Molecular Dynamics

OLIVER F. LANGE, LARS V. SCHÄFER, HELMUT GRUBMÜLLER
Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Received 18 January 2006; Revised 28 February 2006; Accepted 11 March 2006
DOI 10.1002/jcc.20473
Published online 9 August 2006 in Wiley InterScience (www.interscience.wiley.com).

Abstract: The major bottleneck of today's atomistic molecular dynamics (MD) simulations is that because of the enormous computational effort involved, only processes at nanoseconds to microseconds time scales or faster can be studied directly. Unfortunately, apart from a few exceptions, relevant processes, such as chemical reactions or many large scale conformational transitions in proteins, occur at slower time scales and therefore are currently far out of reach for conventional MD. The flooding technique addresses this problem by inclusion of a flooding potential into the force field. This flooding potential locally destabilizes the educt state and thereby significantly accelerates the escape from the initial energy well without affecting the reaction pathway. Here, we summarize the theory and method for the computational chemistry community and detail the implementation within the official version 3.3 of the freely available MD program package GROMACS. Two examples shall demonstrate the application of flooding to accelerate conformational transitions and chemical reactions. The second example was carried out within a QM/MM framework.

 J Comput Chem 27: 1693–1702, 2006

Key words: molecular dynamics; chemical flooding; conformational flooding; enhanced sampling; transition path prediction; QM/MM; umbrella sampling

Introduction

Atomistic molecular dynamics (MD) simulations have been successfully used in recent years to study complex systems, such as proteins or DNA.[1–4] In these simulations, atomic motions are described classically. Accordingly, the propagation of a molecular system in time as it samples conformational space is obtained by integrating Newton's equations of motion. The huge computational effort involved currently limits simulation times to nanoseconds. However, most biomolecular processes occur at slower time scales. Often, this is because free energy barriers much higher than the thermal energy have to be overcome. Examples for such slow processes that involve a barrier crossing from an initial (educt) to a final (product) free energy well (Fig. 1, top) are collective conformational transitions or activated chemical reactions. These rare events[5] are often not directly accessible to conventional MD, impeding the elucidation of transition pathways. Many techniques to address this problem have been devised (for an overview, see refs. 6 and 7), the majority of which rely on *a priori* knowledge of both, the educt and the product state. Therefore, if only the educt state is known, these methods cannot be applied, and the prediction of product states becomes a real challenge. This scenario is addressed by the flooding technique,[5,8] which aims at predicting both, the unknown product state as well as the transition pathway.

Here we report on the recent implementation of this method within the latest version (v3.3) of the fast and freely available MD program GROMACS,[9] which can be downloaded from the official GROMACS website (http://www.gromacs.org). The basic idea of flooding is to destabilize the educt conformation to accelerate the escape without bias towards any of the many possible product states. As illustrated in Figure 1, this is achieved by inclusion of an additional flooding potential V_{fl}, which locally rises the educt free energy without affecting transition or product states. Thereby, the free energy barrier for the transition is effectively lowered, and the transition becomes accessible to simulations. A notable strong point of the flooding technique is that no prior knowledge of the putative transition is required. In particular, no reaction coordinate, transition, intermediate, or product states have to be assumed or guessed, and thus the method allows for the unbiased prediction of transition pathways and product states.

A number of related techniques to accelerate transitions by destabilizing potentials have been proposed, such as local elevation,[10] hyperdynamics[11,12] or boosted dynamics,[13] and metadynamics.[14]

Correspondence to: H. Grubmüller; e-mail: hgrubmu@gwdg.de

Contract/grant sponsor: Volkswagen Foundation; contract/grant numbers: A2:I/78839, A2:I/80436

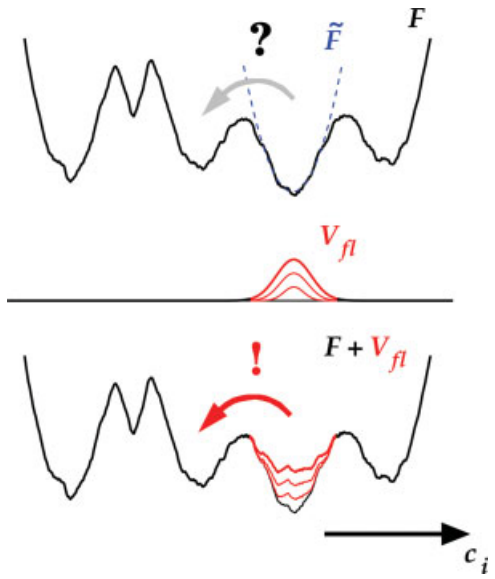

Figure 1. Principle of flooding. The free energy F along a collective coordinate, c, is approximated quasi-harmonically in the local educt minimum to yield $\tilde{F}$. From this, a Gaussian shaped flooding potential V_{fl} is constructed which destabilizes the initial well and accelerates the transition across the barrier.

These, however, have not been implemented in GROMACS and are therefore not considered here.

In this paper, we briefly summarize the theory[5] and detail the implementation. Subsequently, user interface and file formats are described. Finally, two examples will serve to illustrate the application of the method and the interpretation of the results. The first example is the acceleration of the trans-gauche conformational transition of *n*-butane. The second example is a chemical reaction, the opening of a three-membered ring in aqueous solution. The latter example also demonstrates that flooding can be used within the framework of a hybrid quantum mechanical/molecular mechanical (QM/MM) approach.[15] Here, the benefit of using the flooding method is particularly pronounced, since chemical reactions are often connected with tremendous energy barriers. Furthermore, the mandatory application of vastly time consuming QM methods shortens the accessible time scales drastically. Here, particularly large acceleration can be gained using the flooding technique. Finally, most of the conventional methods mentioned earlier cannot be applied to reactions in condensed phase such as done in the second example.

Theory

We first summarize the conceptual framework, as detailed in refs. 5 and 8, to clarify notation (Fig. 1). Flooding involves two steps: first, the free energy landscape of the system is approximated quasi-harmonically.[5] Second, a multivariate flooding potential V_{fl} is constructed from this approximation, which serves to raise the bottom of the educt energy well without affecting regions of higher energy and, in particular, the barriers surrounding the energy well, which determine the transition pathway.

For the quasi-harmonic approximation of the free energy landscape, m linear collective coordinates

$$c_j = \sum_{j}^{3N} a_{ij} x_i \quad j = (1, \ldots, m)$$

are chosen. The x_i denote the $3N$ cartesian coordinates of the molecule and the $3N \times m$ coefficients a_{ij} form an orthogonal matrix $\mathbf{A}$, which defines the collective coordinates $\mathbf{c} = (c_1, \ldots, c_m)^T$. Here, we consider two widely used methods to obtain such collective coordinates, principal component analysis (PCA)[16–18] and normal mode analysis (NMA).[19–21] However, the theory is also valid for other suitably chosen coordinates, e.g., coordinates obtained from a full correlation analysis (FCA) [O.L., Ph.D. thesis, Cuvillier Verlag, Göttingen 2006].

The quasi-harmonic approximation in the collective coordinates and centered at $\mathbf{c}^{(0)} = \langle \mathbf{c} \rangle$,

$$\tilde{F}(c_1, \ldots, c_m) = \frac{1}{2} k_B T \sum_{j=1}^{m} \lambda_j \left(c_j - c_j^{(0)}\right)^2,$$

is chosen such that thermal motions within $\tilde{F}$ have the same amplitudes as the original atomic motion within the original well of the atomistic energy landscape. Accordingly, the curvatures λ_j are determined by the fluctuational amplitudes

$$\lambda_j^{-1} = \alpha^2 \left\langle \left(c_j - c_j^{(0)}\right)^2 \right\rangle,$$

that can, e.g., be obtained from MD. Here, the parameter α was included to allow global scaling of the quasi-harmonic free energy approximation, e.g., to account for insufficient sampling (see General Remark Section).

A Gaussian-shaped flooding potential

$$V_{\mathrm{fl}}(c_1, \ldots c_m) = E_{\mathrm{fl}} \exp\left[-\frac{k_{\mathbf{B}} T}{2 E_{\mathrm{fl}}} \sum_{j=1}^{m} \lambda_j \left(c_j - c_j^{(0)}\right)^2 \right], \qquad (1)$$

is constructed such that its principal axes are parallel to those of $\tilde{F}$, and its extensions (standard deviations) along these axes are proportional to those of the thermal fluctuations in $\tilde{F}$. The flooding strength E_{fl} controls both, the width as well as the height of the flooding potential.

If the harmonic approximation is sufficiently good, e.g., for small molecules *in vacuo*, or under certain conditions also for solvated proteins,[22] the curvatures λ_j of $\tilde{F}$ can also be obtained from the vibrational frequencies ω_j using the equipartition theorem, $\lambda_j \approx \omega_j^2 / k_{\mathrm{B}} T \alpha^2$.[23] This fact can be used to replace MD sampling by NMA, which is computationally more efficient.

Two strategies can be used to determine the flooding strength. The first, rather straightforward, strategy employs a *constant flooding* strength $E_{\mathrm{fl}} \equiv E_{\mathrm{fl}}^{(0)}$ throughout the simulation.[5] In many cases, however, the relation between $E_{\mathrm{fl}}^{(0)}$ and the expected acceleration of the transition is unclear. To this end, the second strategy, *adaptive flooding*,[8] couples the flooding strength E_{fl} to the flooding energy V_{fl}

with the aim to reach and maintain a specified target destabilization free energy $\Delta F_0 = -k_B T \ln\langle e^{-V_{fl}/k_B T}\rangle \approx \langle V_{fl}\rangle$,[5] where the angle brackets denote an appropriate time-average. From this free energy, the expected acceleration is obtained via the Boltzmann factor $e^{\Delta F_0/k_B T}$. The coupling is achieved by updating the flooding strength at each timestep i,

$$E_{fl}^{(i+1)} := E_{fl}^{(i)} + \frac{\Delta t}{\tau_1}\left[\Delta F_0 - \Delta F^{(i)}\right],$$

and taking a sliding average for $\langle V_{fl}\rangle$,

$$\Delta F^{(i+1)} := \left(1 - \frac{\Delta t}{\tau_2}\right)\Delta F^{(i)} + \frac{\Delta t}{\tau_2} V_{fl}^{(i)}.$$

Here, Δt is the integration step size used for the MD simulation, and the time constants τ_1 and $\tau_2 = f\tau_1 (f \geq 1)$ define the coupling strengths.

Implementation

Overview

We implemented the flooding method into the program `mdrun` (v3.3), which is the main MD engine of the MD package `GROMACS`.[9] To take the additional forces due to V_{fl}, $F_i = -\nabla_i V_{fl}$, into account, the force routine of `mdrun` calls our flooding implementation within the `edsam`[24] module of `GROMACS`. All necessary information is condensed in a single input file, `sam.edi`, which is an extension of the old `edsam` input file. `sam.edi` is generated by means of the program `make_edi` (see next Section). Since it is an ascii file, it can also be edited manually. To perform flooding simulations with `GROMACS`, `sam.edi` is passed to `mdrun` via the option `-ei`. The option `-eo` chooses a filename for the ascii output, e.g., `sam.edo`, which contains, in that order: the MD time step, E_{fl}, V_{fl}, and ΔF.

Input Generation

The flooding input file, `sam.edi`, which contains all necessary information (such as the flooding matrix $\mathbf{A}$ and parameters) is generated by the program `make_edi`. The required collective coordinates and the motional amplitudes can be obtained, e.g., by PCA, NMA, or FCA.

To use PCA, the program `g_covar` computes the covariance matrix of fluctuational motion from an MD trajectory $\mathbf{x}(t)$. `g_covar` removes rotational and translational motion by least squares fitting to a reference structure $\mathbf{x}_{ref}$, yielding a corrected trajectory $\mathbf{x}'(t)$. The covariance matrix $\mathbf{C} = \langle(\mathbf{x}' - \langle\mathbf{x}'\rangle)(\mathbf{x}' - \langle\mathbf{x}'\rangle)^T\rangle$, where the brackets denote the time average, is diagonalized and its eigenvectors and eigenvalues are stored together with $\mathbf{x}_{ref}$ and the average structure $\mathbf{x}_{av} = \langle\mathbf{x}'\rangle$ in the files `eigenvec.trr` and `eigenval.xvg`. `make_edi` reads these files and stores m selected eigenvectors a_{ij} and the corresponding reciprocal eigenvalues $\sigma_j, \lambda_j = \sigma_j^{-1}$, as well as the reference and average structure in the flooding input file `sam.edi`.

NMA can be performed with `mdrun`, yielding `hessian.mtx`, which is subsequently diagonalized by `g_nmeig`. Note that for the application of NMA within flooding, no mass weighting should be applied.[8] The normal modes and their frequencies are also stored in files `eigenvec.trr` and `eigenval.xvg`. To treat the two cases NMA and PCA in a common framework, as shown below, `make_edi` transforms the frequencies ω_j to curvatures according to $\lambda_j = \omega_j^2/k_B T$. This allows for one common format of the flooding input file.

Table 1. Command line options for make_edi.

Parameter	Command line option
$E_{fl}^{(0)}$	-Eflnull
α	-alpha
τ	-tau
ΔF_0	-deltaF0

In the following, we give a short overview of the command line interface of `make_edi`. The default behavior is to assume that the input files, provided with options `-f` for the vectors and `-eig` for the eigenvalues or frequencies, refer to PCA modes, created with `g_covar`. If the input files refer to normal modes, this has to be flagged by `-hesse`. The option `-flood <list>` selects m principal or normal modes. Note that the first six normal modes correspond to rotational and translational motion. These are (within numerical accuracy) zero and should not be used. Additional command line options are used to set parameters such as those which control flooding strength and adaptive flooding (cf. Table 1).

Coordinate Transformation

Within each flooding simulation time step, the forces due to the flooding potential V_{fl} need to be computed and added to the forces derived from the original atomistic potential. Computing the forces by directly evaluating $\partial V_{fl}/\partial\mathbf{x}$ is inefficient, however, mainly due to the $O(N^2)$ operations required for computing the matrix product $\mathbf{x}^T\mathbf{A}^T\Lambda\mathbf{A}\mathbf{x}$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$. Instead, we apply the chain rule and compute

$$\mathbf{f}_{fl} = \frac{\partial V_{fl}}{\partial \mathbf{x}} = \frac{\partial V_{fl}}{\partial \mathbf{c}}\,\frac{\partial \mathbf{c}}{\partial \mathbf{x}}. \tag{2}$$

The benefit of this is that the gradient can be computed efficiently by evaluating the m equations

$$\frac{\partial V_{fl}}{\partial c_j} = \frac{k_B T}{E_{fl}}\lambda_j V_{fl}(c_1, \ldots, c_m)\left(c_j - c_j^{(0)}\right)$$

plus a single evaluation of eq. (1) to gain $V_{fl}(c_1, \ldots, c_m)$. Carried out in this way, the complete force evaluation scales with $O(Nm)$.

To project the atomic coordinates $\mathbf{x}$ to collective coordinates, first the rotational and translational motion are removed. Consequently, the center of mass $\mathbf{x}_{cm}$ is subtracted and a rotation $\mathbf{R}$ is applied, which is obtained by least squares fitting to the reference structure $\mathbf{x}_{ref}$: $\mathbf{x}' = \mathbf{R}(\mathbf{x} - \mathbf{x}_{cm})$. The m collective coordinates are obtained by

$$\mathbf{c} = \mathbf{A}^T\left(\mathbf{x}' - \mathbf{x}'_{av}\right),$$

where the average structure $\mathbf{x}'_{\mathrm{av}}$ and the column vectors a_{ij} of the matrix $\mathbf{A}$ are read from the input file `sam.edi` during initialization of the MD program `mdrun`. Accordingly, the second part of the chain rule eq. (2) reads

$$\frac{\partial \mathbf{c}}{\partial \mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{R}.$$

Restraining Potentials

The original idea of flooding is to destabilize conformations by applying the Gaussian-shaped flooding potential V_{fl}. Additionally, a simple inversion of its sign allows to stabilize, i.e., restrain the system at an arbitrary position $\mathbf{c}^{(0)}$ of the collective coordinates. This sign inversion is switched by the option `-restrain` of the input file generator `make_edi`. The effect is accomplished by inverting the signs of the parameters α^2 and E_{fl} in the input file `sam.edi`. Note that the adaptive coupling to a target energy ΔF_0 can still be applied using inverted potentials, and might be judiciously used to gradually switch the restraints on or off. Furthermore, the switch `-harmonic` chooses a harmonic potential (rather than a Gaussian function)

$$V_{\mathrm{fl}} = -\frac{E_{\mathrm{fl}}}{2}\sum_{j}^{m}\frac{1}{\lambda_j}\left(c_j - c_j^{(0)}\right)^2, \tag{3}$$

where the parameters E_{fl} and λ_i of the input file are reinterpreted as force constants. Note, that to obtain a restraining harmonic potential, the sign of eq. (3) has to be inverted by also setting the option `-restrain`. The harmonic restraining potential can be used to perform umbrella sampling[36] along collective coordinates. Free energy profiles can then be obtained via the weighted histogram analysis method (WHAM),[37] e.g., using `g_wham`.

Multiple Flooding Potentials

More than one flooding potential defined by different sets of eigenvectors $a_{ij}^{(k)}$ and inverted eigenvalues $\lambda_j^{(k)}$ can be applied and combined by concatenation of several `*.edi` files. The input data sets within this concatenated input file are read sequentially and lead to independent sequential execution of the aforementioned steps for the respective flooding or restraining potentials $V_{\mathrm{fl}}^{(k)}$; the overall flooding force is given by $\mathbf{f}_{\mathrm{fl}} = \sum_k \mathbf{f}_{\mathrm{fl}}^{(k)}$. Therefore, any combination of parameters for adaption and constant flooding as well as different origin structures $\mathbf{c}^{(0)}$ is possible. For a discussion of typical cases where multiple flooding potentials are particularly useful, see General Remarks section.

Examples

All simulations for the illustration of the flooding method were carried out using a beta version of `GROMACS 3.3`, which includes a QM/MM interface[25] to `GAUSSIAN03`,[26] and the flooding implementation described in this paper. Two examples shall illustrate the flooding method and its implementation. In the first example, *n*-butane, the trans-gauche conformational transition was accelerated. The second example shows the breakage of a chemical bond in a dialkoxymethylene-cyclopropane in aqueous solution. This chemical reaction is thus predicted in a QM/MM framework.

Trans-Gauche Transition of n-Butane

The trans-gauche dihedral transition of *n*-butane is governed by a high-energy barrier of 15.9 kJ/mol,[27] and thus is unlikely to occur at accessible MD simulation time scales. This renders *n*-butane a suitable model system for our illustrative purposes.

Setup

The `GROMACS` united atom force field (ffgmx) was used. First, the trans conformation of *n*-butane was energy minimized with steepest descent until the maximum force was smaller than 10^{-4} kJ/mol/nm. All MD simulations, starting from this energy minimized structure, and using the same initial velocities, were carried out with an integration time step of 1 fs and temperature coupling to a 300 K heat bath ($\tau_c = 0.1$ ps).[28]

Two different flooding potentials V_{fl} were generated, the first from PCA, and the second from NMA. For PCA, the covariance matrix $\mathbf{C}$ was computed from a 100 ps trajectory, during which *n*-butane stayed in the trans conformation, as expected. For NMA, the Hessian matrix $\mathbf{H}$ was computed from the energy minimized structure. Subsequently, four flooding MD simulations were performed. First, two constant flooding simulations with V_{fl} based on PCA coordinates (MD1) or based on NMA coordinates (MD2), respectively. Second, two adaptive flooding simulations with V_{fl} based on PCA coordinates (MD3) or NMA coordinates (MD4). The flooding subspace was chosen to be spanned by all $m = 6$ internal degrees of freedom of united atom *n*-butane, i.e., the first six PCA or last six NMA modes, respectively, and the energy minimized structure of the trans conformer was used as the reference structure for the rotation/translation correction.

Constant Flooding

To accelerate the trans-gauche dihedral transition of *n*-butane, a flooding potential V_{fl} with constant flooding strength $E_{\mathrm{fl}} = 12.0$ kJ/mol was applied throughout the simulations. Figure 2 shows the time evolution of the dihedral angle (A) and the flooding energy V_{fl} (B) for the two simulations MD1 and MD2, based on PCA and on NMA coordinates, respectively. In both simulations, the applied flooding potential induced a trans-gauche dihedral transition. As expected for a small molecule such as *n*-butane *in vacuo*, the PCA- and NMA-based flooding simulations behave very similarly.

Starting with the trans conformation (dihedral$= -180°$), the dihedral angle displayed a high-motional amplitude and, correspondingly, a low frequency in the broadened minimum (Fig. 2A). Within the first few picoseconds of the simulations, the fluctuational amplitude of the dihedral angle was slightly higher in MD2 compared to MD1, resulting in a faster escape to the gauche minimum (dihedral $= -60°$ or $-300°$), which occurred after 5.9 ps in MD2 compared to 12.4 ps in MD1. The system escaped from the influence of the flooding potential via the dihedral transition. Therefore, V_{fl} dropped to (nearly) zero concomitant with the structural transition.

MD1 and MD2 were started from the same initial configuration, i.e., same structure and velocities. Nonetheless, the slight differences of the applied PCA- or NMA-based flooding potentials yielded

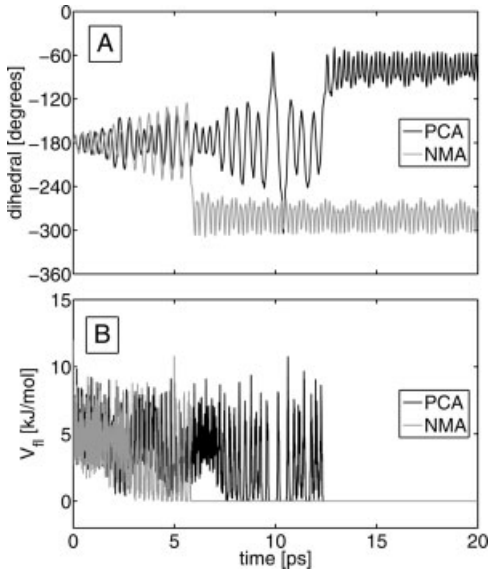

Figure 2. Time evolution of the dihedral angle (A) and V_{fl} (B) in the course of the constant flooding simulations MD1 (black) and MD2 (gray).

two different gauche conformers, which are mirror images and thus equally likely. Before the final transition at 12.4 ps, MD1 transiently visited the gauche conformer at 9.8 and 10.3 ps, respectively.

Note that for monoexponential kinetics, the escape times obtained from otherwise identical flooding simulations, started with various sets of initial velocities randomly chosen from a Maxwell distribution at 300 K, scatter by about one order of magnitude. This is seen by considering a single exponential decay

$$p(t) = p_0 e^{-t/\tau},$$

with relaxation time τ. A change of variables to a logarithmic escape time $u = \ln(t/\tau)$ reveals that individual escape events are distributed according to

$$f(u) = e^{u-\exp(u)}, \tag{4}$$

as shown in Figure 3. The difference by a factor of two for the escape times of MD1 and MD2 falls completely within this expected scatter.

The choice of E_{fl} strongly affects the mean escape time, because the destabilization free energy, ΔF, which determines the acceleration due to the flooding potential, scales with the flooding strength as

$$\Delta F = E_{fl} \exp(-mk_B T/2E_{fl}), \tag{5}$$

(eq. (33) in ref. 5). To demonstrate this, MD2 was repeated with $E_{fl} = 11.5$ and 12.5 kJ/mol, respectively, starting from identical initial configurations. In the first case, the transition occurred after 19.3 ps, whereas in the latter, the escape time was only 1.5 ps (data not shown). This is in line with eq. (5), considering the scatter of escape times (eq. (4)). Therefore, the parameter E_{fl} has to be chosen very carefully. If E_{fl} is chosen too small, one does not achieve a sufficiently large acceleration to observe a transition within the limited simulation time span. In contrast, too strong flooding might alter the free energy landscape also within the crucial transition state region, and might thus lead to artefacts. To allow a more direct control of the expected acceleration, one would wish to adjust the destabilization free energy ΔF directly rather than E_{fl}. This is achieved by the adaptive flooding scheme, in which E_{fl} is updated in every MD time step such as to reach a desired target destabilization free energy ΔF_0 (cf. Theory section).

Adaptive Flooding

To illustrate its benefits, adaptive flooding with a target destabilization free energy $\Delta F_0 = 3$ kJ/mol, $E_{fl}^{(0)} = 0$, and a time constant $\tau_1 = 2$ ps was employed on *n*-butane. Figure 4 shows the time evolutions of the dihedral angle (A) and the flooding energy V_{fl} as well as the flooding strength E_{fl} (B) for the two adaptive flooding simulations MD3 and MD4. Also here, the results of the simulations based on PCA (MD3) and on NMA (MD4) are very similar. The trans-gauche transition occurred after 15 ps (MD3) or 55 ps (MD4), respectively, after which the gauche conformation was adopted for the rest of the simulation time. Similar to the constant flooding simulations discussed earlier, the fluctuational amplitude of the dihedral angle within the first few picoseconds of the simulations was slightly higher in the NMA based flooding simulation MD4 compared to the PCA based MD3, resulting in a faster escape to the gauche minimum.

Starting at zero, E_{fl} increased at the beginning of the simulations, until it reached a plateau at about 11–12 kJ/mol (Fig. 4B). V_{fl} increased accordingly, before it dropped to nearly zero concomitant with the dihedral transition, and should have been switched off at this point. However, for the sake of illustration this was not done here, with the effect that at longer time scales contrary to constant

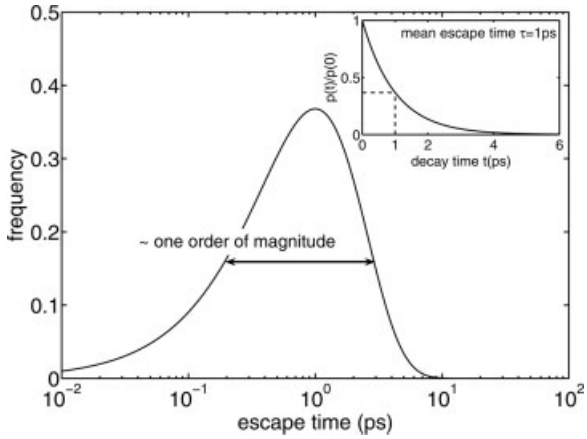

Figure 3. Distribution of escape times corresponding to a single exponential relaxation process with relaxation time $\tau = 1$. Note that the mean escape time according to this distribution is also τ. The width at half maximum of this about one order of magnitude. The resulting population decay is shown in the inset.

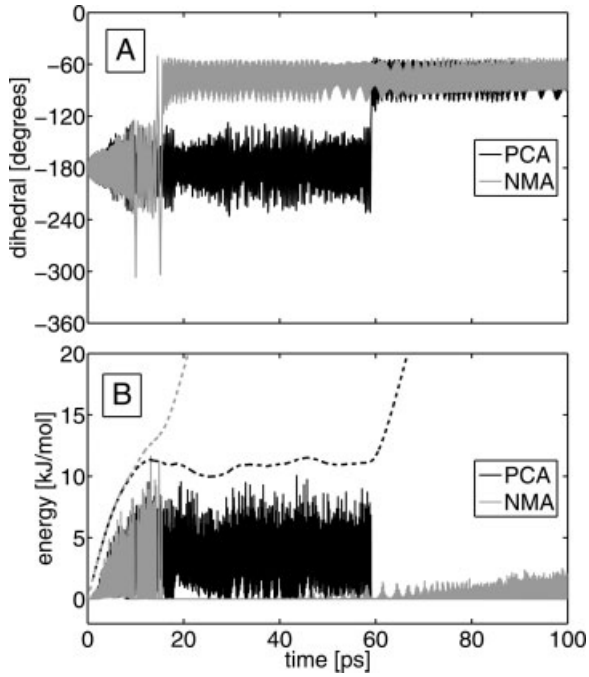

Figure 4. Time evolutions of dihedral angles (A) and V_{fl} (B, solid line) as well as E_{fl} (B, dashed line) for MD3 (black) and MD4 (gray).

flooding, the adaptive scheme caused V_{fl} to increase again. After about 50–60 ps in MD4, the adaptively growing flooding potential started to affect the final gauche conformer as well. Hence, to allow for an unperturbed relaxation, flooding should be switched off after the structural transition.

Figure 5 shows the potential energy (with V_{fl} excluded) in the course of the flooding simulation MD3. Here, the flooding potential V_{fl} was switched off manually after the structural transition occurred. To explain Figure 5 note that, according to the equipartition theorem, each of the six internal degrees of freedom of *n*-butane has an average thermal energy of $k_{\mathrm{B}}T/2$, resulting in an average thermal energy of 7.5 kJ/mol at $T = 300$ K. Therefore, setting the potential energy of the trans minimum to zero, the potential energy of *n*-butane

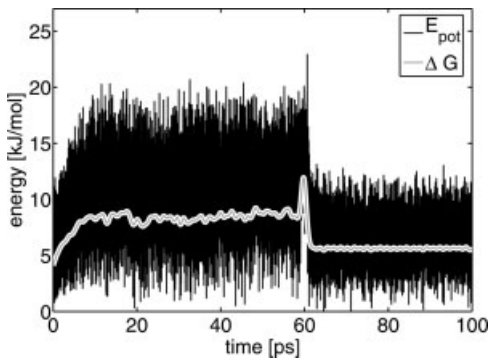

Figure 5. Potential energy (black) in the course of MD3. The E_{pot} of the initial structure was set to zero. For the free energy estimate (gray), a width $\sigma = 0.4$ ps was chosen for the Gaussian function g_σ (cf. eq. (6)).

fluctuated around this value at the start of the simulation. The gradually increasing flooding potential caused a respective increase of the mean potential energy during the first few picoseconds. Upon the dihedral transition, the system escaped the region in conformational space influenced by the flooding potential (cf. Fig. 4B) and rapidly relaxed into thermal equilibrium of the gauche conformer. Note that the maximum potential energy in Figure 5 of 23 kJ/mol surely overestimates the true transition state energy, since the minimum energy path is very unlikely to be followed in a flooding MD simulation at finite temperature. A more reliable estimate of about 12 kJ/mol for the free energy barrier is provided by a sliding Gaussian-averaged time-dependent partition function (eq. (3.3) in ref. 8),

$$Z(t) = \int_{-\infty}^{\infty} dt' g_\sigma(t - t') \exp\left(E_{\mathrm{pot}}(t')/k_{\mathrm{B}}T\right), \qquad (6)$$

see Figure 5.

Methylene-Cyclopropane

To demonstrate the flooding method for the more challenging case of a chemical reaction in solution, we investigated the rearrangement of the dialkoxy-substituted methylene-cyclopropane (MCP) **1** to the dialkoxy-trimethylenemethane (TMM) intermediate **2** in water, see Figure 6. The electronic configuration of TMM **2** can be described as a superposition of the diradicalic structure **2a** and the zwitterion **2b**. The balance between these two configurations is critically influenced by the solvent polarity. Apolar solvents prefer the diradical, whereas increasing solvent polarity stabilizes the zwitterion.

Since MM force fields are unable to describe chemical reactions, we applied a hybrid QM/MM[15] approach. The flooding potential destabilizes the MCP **1**, and triggers the ring-opening reaction towards TMM **2**. Since this bond-breaking comprises a high-energy barrier, it is not accessible to unperturbed QM/MM simulations.

Setup

The QM/MM flooding simulations involved the following steps. First, the system was properly prepared, including the generation of an equilibrated solvent configuration. Then, a free QM/MM equilibration run was carried out to sample data for a PCA. Finally, the QM/MM flooding MD simulation was performed.

Initially, the geometry of MCP was optimized in the gas phase at the HF/3-21G* level and then solvated in a box of 1182 SPC water molecules. Positionally restraining all MCP atoms, the solvent was first energy minimized (steepest descent, 1000 steps) and subsequently equilibrated for 250 ps at 300 K. During these simulations,

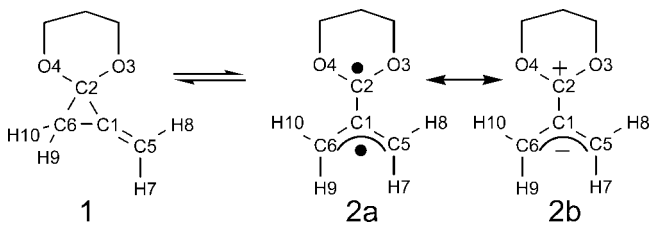

Figure 6. MCP **1** and TMM **2**.

the whole system was described classically, using the GROMOS96 force field (ffG43a2) for MCP. Application of the SETTLE[29] method allowed for an integration time step of 2 fs. The system was coupled to a 300 K heat bath ($\tau_c = 0.1$ ps) and to a 1-bar pressure bath ($\tau_p = 1$ ps).[28] For the evaluation of the nonbonded forces, a cutoff of 1.4 nm was applied.

The resulting structure was taken as a starting point for a subsequent 50 ps QM/MM free MD simulation, with MCP treated at the HF/3-21G* level. All QM/MM simulations were carried out without constraints on the QM-QM bonds, and an integration time step of 1 fs was used (all other simulation parameters as described earlier). In the applied QM/MM scheme,[25] the van- der- Waals interactions between the QM and the MM atoms were described at the force field level, whereas the point charges of the SPC water molecules were included into the QM Hamiltonian and thus polarize the solute wavefunction. For PCA, the covariance matrix **C** of the motion of the eight atoms of the hydrocarbon methylene-cyclopropane ring only (C1, C2, C5, C6, H7–H10, see Fig. 6) was computed from the last 40 ps of this trajectory, using the gas phase optimized MCP structure as the reference.

Adaptive flooding with a target destabilization free energy $\Delta F_0 = 75$ kJ/mol, a time constant $\tau_1 = 0.1$ ps, and $\alpha = 1$ was employed. The flooding subspace was chosen to be spanned by all $3N - 6 = 18$ eigenvectors of the constructed covariance matrix **C**. In the QM/MM flooding simulation, the MCP solute was treated at the UB3LYP/6-31G* level. UB3LYP was chosen, since unrestricted density functional theory, in addition to its capability to accurately describe heterolytic bond cleavage, turns out to work exceptionally well also for diradicalic singlet states resulting from homolytic bond cleavage, at comparably low computational cost.[30,31] To relax putative strain due to the previously applied Hartree-Fock level, 200 fs free density functional QM/MM MD were performed before switching on the flooding potential.

Results

As expected, the applied flooding potential induced the ring-opening reaction from **1** to **2**, as can be seen from Figure 7A, which shows the distance between the C2 and C6 carbon atoms in the course of the flooding simulation. The C2—C6 bond broke, and subsequently, the initially perpendicular C1—C2—C6 and C2—C3—C4 planes became coplanar (Fig. 7B). The resulting TMM was then stable for the rest of the simulation time. Through the bond breaking, the system escaped from the influence of the flooding potential, which was switched off manually after 1.5 ps to allow for an unperturbed relaxation (Fig. 7C).

In Figure 7D, the potential energy of the solute in the course of the flooding simulation is shown (gray) along with a free energy estimate (black). For comparison, the two different widths of $\sigma = 10$ fs and $\sigma = 20$ fs were chosen for $g_\sigma(t - t')$ in eq. (6) to obtain the free energy from the potential energy $E_{\mathrm{pot}}(t)$, shown as solid and dashed line, respectively. From the former, an activation free energy barrier of ca. 125 kJ/mol is estimated, whereas the latter yields about 100 kJ/mol. These computed values are close to the experimentally determined activation energy of 107 kJ/mol in C_6D_6[32]. As already stated above for *n*-butane, the minimum energy pathway cannot be expected to be followed in a flooding simulation, and therefore, the obtained energy barrier represents an upper bound. In contrast to the unsubstituted trimethylenemethane in the gas phase,[8] the TMM **2** is a stable intermediate in water and persists throughout the rest of the simulation time. The free energy of **2** is not only significantly lower than that of the transition state but also, after full relaxation, even

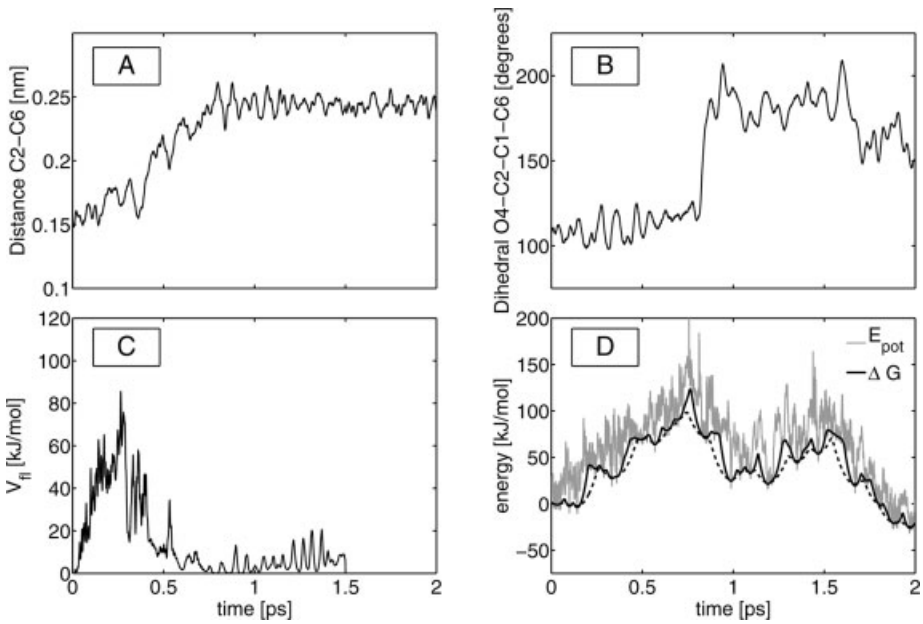

Figure 7. Time evolutions of the C2—C6 bond breakage (A), the dihedral angle between the C1—C2—C6 and C2—O3—O4 planes (B), V_{fl} (C), and potential (gray, D) as well as free energy estimate (black, D) in the course of the adaptive flooding simulation.

lower than that of MCP **1** (Figure 7D). The stability of the resulting TMM **2** compared to the unsubstituted trimethylenemethane in the gas phase due to the polar water solvent and the dialkoxy substituents clearly advocates the formation of the zwitterion **2b**, which is also underlined by the distribution of Mulliken atomic partial charges (−0.63 for the allylic moiety (atoms C1, C5, C6, H7—H10) and +0.63 for the remaining atoms, values computed for snapshot after 2 ps).

General Remarks

In the following, some aspects will be discussed that go beyond the presented examples and might provide some additional help to the user to successfully apply the flooding method.

From our experience, suitable flooding potentials V_{fl} can be generated from relatively short, i.e., 1–10 ns free MD trajectories of the initial educt conformation. This is backed up by a recent study, which has shown that for the small protein crambin, convergence of a medium sized ($m = 100$) conformational subspace is remarkably fast [O.L. and H.G., submitted to JPC B]. In particular, the *directions* of conformational motion in the conformational subspace are already correctly described by a 5 ns MD trajectory. Furthermore, the method does not require fully converged ensembles. Accordingly, in our flooding implementation, the fluctuation *amplitudes*, i.e., eigenvalues of the covariance matrix, which need much longer to converge,[33,34] are adjustable via an empirical parameter α. From our experience, scaling the fluctuation amplitudes obtained from short MD simulations by $\alpha^2 = 2$ is a good choice for proteins.

Multiple flooding potentials are useful for the following cases (see earlier sec.). First, to examine multistep pathways. To this end, the number of applied flooding potentials can be increased successively, i.e., after the escape from the initial conformation and proper relaxation of the accessed product state, a second flooding potential to additionally destabilize this new state might be added to the first one, etc. This procedure is required, since, by construction, V_{fl} acts on the energy landscape only in the vicinity of the initial structure. To find alternative transitions and to avoid revisiting already known (and therefore uninteresting) conformations, additional flooding potentials to destabilize the known states can be applied simultaneously.

Second, transitions from *anharmonic* minima can be accelerated by multiple flooding potentials, appropriately centered at different origin structures $\mathbf{c}^{(0)}$. Third, degenerate stereochemical rearrangements of the educt (i.e., leading to enantiomers of the educt) can be obviated by simultaneous application of multiple flooding potentials, which can be generated using symmetry.

We have described both, constant flooding as well as adaptive flooding. Which of these two techniques shall be applied for a particular case? In a constant flooding simulation, the system propagates on a stationary free energy landscape, and properly chosen parameters accelerate spontaneous transitions after an initial equilibration phase. In contrast, adaptive flooding implies a time-dependent free energy landscape. The advantage of the latter approach is that a transition is guaranteed to occur within a controllable simulation time, as long as a sufficiently large target destabilization free energy ΔF_0 and small flooding time constant τ are chosen. On the other hand, if ΔF_0 is too large or τ is too small (tight coupling), too little time is provided to sufficiently equilibrate in the rapidly changing free energy landscape. This might lead to artefacts, i.e., the system might be driven towards high energy configurations before an exit channel from the altered region of the free energy landscape is found. Thereby, artificially large activation energies and possibly wrong pathways and product states would be predicted. Slow degrees of freedom, e.g., in proteins, require long equilibration time.[35] Therefore, to induce conformational transitions in proteins, adaptive flooding with sufficiently small ΔF_0 and large τ or constant flooding should be applied. On the other hand, for systems displaying fast degrees of freedom only, e.g., the MCP example discussed earlier, adaptive flooding with tight coupling is well suited and the technique of choice.

In both cases, the flooding strength E_{fl} critically determines the mean escape time (see earlier section). The proper choice of the parameter E_{fl} is therefore very important, since too weak flooding potentials will not sufficiently accelerate conformational transitions, whereas too strong flooding potentials heavily distort the energy landscape and lead to artefacts. However, in the majority of cases, the heights of the energy wells that have to be surmounted are unknown. We therefore suggest to first apply adaptive flooding with large destabilization free energy for an initial exploration of the energy landscape, and to extract appropriate parameters from this for subsequent more gentle flooding simulations, which will then yield improved reaction pathways. From our experience, to accelerate conformational transitions in proteins by means of a constant flooding potential, flooding strengths of 0.2–3 $k_{\mathrm{B}}T$ per degree of freedom influenced by the flooding potential are suitable.

We finally re-emphasize that although flooding also provides energy profiles, its primary purpose is to predict reaction pathways. Once the pathway is identified, established methods, which require the pathway as input, will generally provide more accurate barrier heights and energy profiles.

The current `GROMACS` (v3.3) available at the official gromacs website http://www.gromacs.org contains the flooding implementation described in this report.

Appendix: Used `GROMACS` Commands to Perform Flooding Simulations

n-Butance

PCA and NMA

`g_covar` performed the PCA over the 100 ps equilibration trajectory using the energy minimized trans *n*-butane as the reference:

- `g_covar -f traj.xtc -s minimized.pdb -o eigenval.xvg -v eigenvec.trr`

For NMA, trans *n*-butane was first energy minimized:

- `grompp_d -f em.mdp -c minimized.pdb`
- `mdrun_d -o minimized_d.trr -c minimized_d.pdb`

The Hessian matrix `nm.mtx` was then computed. Note that for generating the run input file one should use the minimized conformation from the full precision trajectory file, and not from confout.gro:

- `grompp_d -f md.mdp -t minimized_d.trr -c minimized_d.pdb`
- `mdrun_d -mtx nm.mtx`

Finally, the Hessian matrix was diagonalized by `g_nmeig`. Thereby, no mass weighting was applied:

- `g_nmeig_d -f nm.mtx -ol eigenval.xvg -v eigenvec.trr -nom`

Note that `grompp`, `mdrun`, and `g_nmeig` should be run in double precision when NMA is performed with `GROMACS`.

Generation of Flooding Input Files `sam.edi`

The files `eigenval.xvg` and `eigenvec.trr` were then passed to `make_edi` to produce `sam.edi`. For the constant flooding simulation MD1:

- `make_edi -f eigenvec.trr -eig eigenval.xvg -s minimized.pdb -o sam.edi -flood 1-6 -tau 0 -Eflnull 12.0`

For the constant flooding simulation MD2:

- `make_edi -f eigenvec.trr -eig eigenval.xvg -s minimized.pdb -o sam.edi -flood 7-12 -tau 0 -Eflnull 12.0 -hesse`

For the adaptive flooding simulation MD3:

- `make_edi -f eigenvec.trr -eig eigenval.xvg -s minimized.pdb -o sam.edi -flood 1-6 -tau 2 -deltaFO 3`

For the adaptive flooding simulation MD4:

- `make_edi -f eigenvec.trr -eig eigenval.xvg -s minimized.pdb -o sam.edi -flood 7-12 -tau 2 -deltaFO 3 -hesse`

Flooding Simulations

`sam.edi` is passed to `mdrun` to perform the flooding simulation:

- `mdrun -ei sam.edi -eo sam.edo`

Methylene-Cyclopropane

Principal Component Analysis

`g_covar` performed the PCA over the last 40 ps of the 50 ps equilibration trajectory using the HF/3-21G* geometry optimized structure as the reference:

- `g_covar -f traj.xtc -s minimized.pdb -n index.ndx -o eigenval.xvg -v eigenvec.trr -b 10`

From the index file, the group containing the 8 atoms Cl, C2, C5, C6, H7–H10 was choosen.

Generation of Flooding Input File sam.edi

The first 18 eigenvectors were subjected to the flooding potential, and data was written to a disk every step:

- `make_edi -f eigenvec.trr -eig eigenval.xvg -s minimized.pdb -o sam.edi -flood 1-18 -tau 0.1 -deltaFO 75 -logfrq 1 -outfrq 1`

Flooding Simulation

- `mdrun -ei sam.edi -eo sam.edo`

Acknowledgments

We thank Dr. Bert L de Groot for helpful discussions and Dr. Gerrit Groenhof for help with the QM/MM interface. LVS thanks the Boehringer Ingelheim Fonds for a doctoral fellowship.

References

1. Karplus, M.; McCammon, J. A. Nat Struct Biol 2002, 9, 646.
2. de Groot, B. L.; Grubmüller, H. Curr Opin Struct Biol 2005, 15, 176.
3. Isralewitz, B.; Gao, M.; Schulten, K. Curr Opin Struct Biol 2001, 11, 224.
4. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glaettli, A.; Huenenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Angew Chem 2006, 45, 4064.
5. Grubmüller, H. Phys Rev E 1995, 52, 2893.
6. Tai, K. Biophys Chem 2004, 107, 213.
7. Schlegel, H. B. J Comput Chem 2003, 24, 1514.
8. Müller, E. M.; de Meijere, A.; Grubmüller, H. J Chem Phys 2002, 116, 897.
9. van der Spoel, D.; Hess, B.; Lindahl, E.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. J Comput Chem 2005, 26, 1701.
10. Huber, T.; Torda, A. E.; van Gunsteren, W. F. J Comput Aided Mol Des 1994, 8, 695.
11. Voter, A. F. Phys Rev Lett 1997, 78, 3908.
12. Voter, A. F. J Chem Phys 1997, 106, 4665.
13. Hamelberg, D.; Mongan, J.; McCammon, J. A. J Chem Phys 2004, 120, 11919.
14. Laio, A.; Parrinello, M. Proc Natl Acad Sci USA 2002, 99, 12562.
15. Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227.
16. Kitao, A.; Hirata, F.; Go, N. Chem Phys 1991, 158, 447.
17. Garcia, A. E. Phys Rev Lett 1992, 68, 2696.
18. Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Proteins 1993, 17, 412.
19. Levitt, M.; Sander, C.; Stern, P. S. Int J Quantum Chem 1983, Suppl. 10, 181.
20. Brooks, B.; Karplus, M. Proc Natl Acad Sci USA, 1983, 80, 6571.
21. Go, N.; Noguti, T.; Nishikawa, T. Proc Natl Acad Sci USA, 1983, 80, 3696.
22. Amadei, A.; de Groot, B. L.; Ceruso, M. A.; Paci, M.; Nola, A. D.; Berendsen, H. J. C. Proteins: Struct Funct Genet 1999, 35, 283.
23. Hayward, S.; Kitao, A.; Hirata, F.; Go, N. J Mol Biol 1993, 234, 1207.
24. de Groot, B. L.; Vriend, G.; Berendsen, H. J. C. J Mol Biol 1999, 286, 1241.
25. Groenhof, G.; Bouxin-Cademartory, M.; Hess, B.; de Visser, S. P.; Berendsen, H. J. C.; Olivucci, M.; Mark, A. E.; Robb, M. A. J Am Chem Soc 2004, 126, 4228.
26. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin,

K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Wallingford, CT, 2004.
27. Vollhardt, K. P.; Schore, N. E. Organic Chemistry. W H Freeman: New York, 1998.
28. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
29. Miyamoto, S.; Kollman, P. A. J Comp Chem 1992, 13, 952.
30. Nendel, M.; Sperling, D.; Wiest, O.; Houk, K. N. J Org Chem 2000, 65, 3259.
31. Engels, B.; Hanrath, M.; Lennartz, C. Computers and Chemistry 2001, 25, 15.
32. Nakamura, E.; Yamago, S.; Ejiri, S.; Dorigo, A. E.; Morokuma, K. J Am Chem Soc 1991, 113, 3183.
33. Hess, B. Phys Rev E 2002, 65, 031910.
34. Faraldo-Gomez, J. D.; Forrest, L. R.; Baaden, M.; Bond, P. J.; Domene, C.; Patargias, G.; Cuthbertson, J.; Sansom, M. S. P. Proteins 2004, 57, 783.
35. Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. J Phys Chem 1996, 100, 2567.
36. Torrie, G. M.; Valle, J. P. J Comput Phys 1977, 23, 187.
37. Souaille, M.; Roux, B. Comput Phys Commun 2001, 135, 40.