JCTC Journal of Chemical Theory and Computation

Parametrization of Reversible Digitally Filtered Molecular Dynamics Simulations

Adrian P. Wiley, Martin T. Swain, Stephen C. Phillips, and Jonathan W. Essex*

School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Colin M. Edge

GlaxoSmithKline, New Frontiers Science Park (North), Coldharbour Road, Harlow CM19 5AD, U.K.

Received July 16, 2004

Abstract: Reversible Digitally Filtered Molecular Dynamics (RDFMD) is a method of amplifying or suppressing motions in a molecular dynamics simulation, through the application of a digital filter to the simulation velocities. RDFMD and its derivatives have been previously used to promote conformational motions in liquid-phase butane, the Syrian hamster prion protein, alanine dipeptide, and the pentapeptide, YPGDV. The RDFMD method has associated with it a number of parameters that require specification to optimize the desired response. In this paper methods for the systematic analysis of these parameters are presented and applied to YPGDV with the specific emphasis of ensuring a gentle and progressive method that produces maximum conformation change from the energy put into the system. The portability of the new parameter set is then shown with an application to the M20 loop of E-coli dihydrofolate reductase. A conformational change is induced from a closed to an open structure similar to that seen in the DHFR−$NADP^+$ complex.

1. Introduction

Molecular dynamics (MD) simulations have been frequently applied to protein systems to provide atomistic detail that is unavailable to experimental methods.[1−3] The time scale that traditional MD simulations can simulate is typically in the order of tens of nanoseconds and is severely limited by computational resources. However, even with this length of simulation, the large-scale conformational motions of proteins are rare events, due to the significant energy barriers that can lie on the potential energy surface. Methods that enhance the rate of these events so that they occur within the length of typical simulations are therefore of great interest.[4]

To address the sampling issues of molecular dynamics a number of equilibrium and nonequilibrium methods exist. These include generalized-ensemble methods that induce a system's random walk in potential energy space, thereby overcoming conformational energy barriers.[5] Within this collection of algorithms are simulated tempering,[6] the multicanonical algorithm,[7] and replica-exchange molecular dynamics.[8] Other well-studied methods range from self-guided molecular dynamics[9] (and its most recent refinement, self-guided Langevin dynamics[10]), that applies an evolving guiding force to a simulation, to CONCOORD[11] ('from constraints to coordinates'), which predicts protein conformations based on distance constraints.

Reversible Digitally Filtered Molecular Dynamics (RDFMD) has previously been shown to amplify or suppress motions of a specific frequency, and the amplification of low-frequency motions has produced increased conformational sampling in a range of systems.[1,2] RDFMD can be tailored through interdependent parameter selection, and methods to choose appropriate parameters for different applications are therefore required. In this paper, a range of analysis methods are presented to develop an RDFMD

* Corresponding author e-mail: J.W.Essex@soton.ac.uk.

10.1021/ct049970t CCC: $30.25
Published on Web 12/17/2004

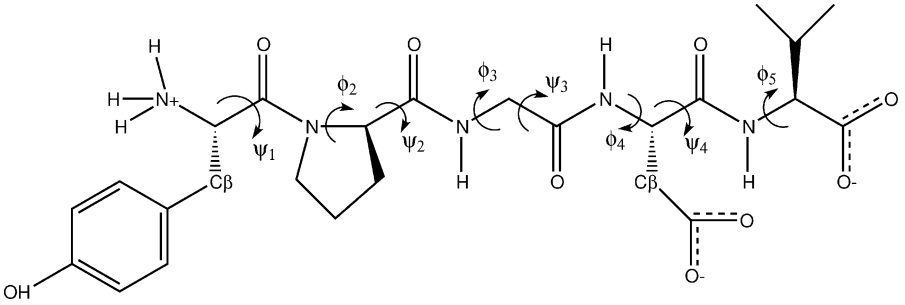

Figure 1. YPGDV with the backbone dihedral angles of interest labeled.

protocol that amplifies the conformational motions of a short peptide chain. The generated parameter set is shown to be suitable for flexible regions of larger protein systems, inducing a conformational change between known forms of the E-coli dihydrofolate reductase M20 loop.

2. RDFMD

A digital filter is a list of coefficients, c_i, that can be used to weight a discrete vector input, $\mathbf{x}_i$, and summed to give a vector output, $\mathbf{y}$ (eq 1). The filter response is the filter's effect on the phase and amplitude of the input signal. The greater the number of coefficients, the closer the filter's response will be to that desired.

$$\mathbf{y} = \sum_{i=-m}^{m} c_i \mathbf{x}_i \tag{1}$$

The RDFMD filter sequence begins by filling a buffer of velocities, $\mathbf{v}$, using a microcanonical (NVE) ensemble MD simulation. This buffer has the same number of steps as the number of filter coefficients ($2m+1$) in eq 2. The external rotational and translational motion of the target system, $\mathbf{v}_{\mathrm{Ext}}$, is removed (eq 3), and a digital filter is then applied to the Cartesian components of the internal velocities, $\mathbf{v}_{\mathrm{Int}}$. Filters used for this work are designed using the *fircls* function in MATLAB[12] and have the property of yielding a frequency response of unity over the frequency range to be amplified and zero elsewhere.[2]

$$\mathbf{v}_{\mathrm{Filt}} = \sum_{i=-m}^{m} c_i \mathbf{v}_{\mathrm{Int},i} \tag{2}$$

$$\mathbf{v}_{\mathrm{Int}} = \mathbf{v} - \mathbf{v}_{\mathrm{Ext}} \tag{3}$$

The filters used are typically designed to extract all components of velocities corresponding to motions most likely to induce conformational change. These filtered velocities, $\mathbf{v}_{\mathrm{Filt}}$, are multiplied by an amplification factor, A, that can be changed to adjust the energy put into the system. The amplified velocities are then summed with the original velocity set for the central buffer point, $\mathbf{v}_0$, producing a new set of velocities, $\mathbf{v}'$, that are in phase with the coordinates at the center of the buffer (eq 4). An amplification factor of 2 would therefore produce a final set of velocities for which the targeted frequency components have been extracted, multiplied by 2, and then summed with the original velocities to give a 3-fold increase in the kinetic energy of targeted motions (as they will also be included in the full set), with all other frequencies left unchanged, in contrast to the self-guided Langevin dynamics method.[10]

$$\mathbf{v}' = \mathbf{v}_0 + A\mathbf{v}_{\mathrm{Filt}} \tag{4}$$

From the new velocities, $\mathbf{v}'$, and the central coordinates of the buffer, conventional simulation continues both backward and forward in time so that a new buffer is filled. Another filter can then be applied, separated from the last by a specified time delay (the filter delay). The purpose of this delay is to allow the system to relax from the effects of the previous filter and to allow some progression over the potential energy surface. Filters are repeatedly applied in this manner until the kinetic energy in the system rises beyond some defined limit (the internal temperature cap) or until a certain number of filters have been applied (the filter cap).

RDFMD is typically run as the combination of periods during which the system velocities are modified by repeated applications of a digital filter and of traditional MD in the canonical (NVT) or the isothermal−isobaric (NPT) ensemble. During the MD for which a thermostat is applied, the temperature can be returned to that desired and new equilibrated velocities and coordinates are generated. From these another set of filter applications can be performed.

In this paper, the parameters associated with the repeated applications of the filter will be examined, their effects on the simulation determined, and methods for their optimization described.

3. RDFMD Parameters

To analyze the response of a system to different RDFMD parameters the YPGDV pentapeptide (tyrosine-proline-glycine-aspartic acid-valine) is used (Figure 1). YPGDV has been used in previous studies as a test case for RDFMD,[2] conformational analysis,[13] the ART-2′ clustering algorithm,[14] and self-guided molecular dynamics.[15] NMR data from *trans*-proline YPGDV in water is known to show an approximately equal proportion of reverse turn and extended conformations at 273 K.[16]

The system was set up from an all-trans Z-matrix of YPGDV, generated and solvated within the MCPRO package[17] using 805 water molecules and 1 sodium ion. Cubic periodic boundary conditions were used throughout. Simulation was performed using the NAMD package[18] with a switching function applied to the Lennard-Jones interactions between 8 and 12 Å, a PME treatment of electrostatics,[19] and SHAKE[20] was applied to all bonds involving a hydrogen atom, with a tolerance of 10^{-8} Å. Explicit water was modeled

by the TIPS3P water model as implemented by CHARMM, and the protein was described by the CHARMM22 force field.[21] NAMD was developed by the Theoretical Biophysics Group in the Beckman Institute at Urbana-Champaign.

Initially, minimization was performed with the conjugate gradient line-search algorithm[18] for a total of 22 000 steps. The system was then gradually heated with a 20 000 step canonical simulation at each temperature between 50 and 300 K at 50 K intervals using a 2 fs time step. A Langevin Thermostat[22] was used with a damping parameter of 10 ps^{-1}. 80 000 steps were then performed at the target 300 K followed by 500 000 steps (1 ns) in the isothermal−isobaric ensemble. A Nosé-Hoover Langevin piston barostat[23] with a pressure target of 1 atm, a piston temperature of 300 K, a damping decay parameter of 200 fs, and an oscillation period of 400 fs was used. A further 200 000 steps (100 ps) of isothermal−isobaric MD were then performed with a 1 ps^{-1} thermostat damping parameter, 300 fs barostat damping decay, and 500 fs piston oscillation period. The simulation state at this point is used when randomising velocities to create a large number of starting points for RDFMD trials. Parameters not specified for later simulations are as above.

A 20 ns isothermal−isobaric MD simulation of YPGDV has been performed from the equilibrated YPGDV system, and coordinate, velocity, and box dimensions were extracted every 2 ns to give 10 largely independent starting points for filter applications.

To investigate the parametrization of RDFMD, parameters are systematically varied from a previously reported protocol[2] and their interdependence discussed. Discussion of the internal temperature cap parameter will be published elsewhere through comparison with parallel tempering[8] and MD simulations. Unless otherwise stated, the RDFMD parameters are as follows: a filter buffer length of 1001 steps, a target frequency of 0−25 cm^{-1}, an amplification factor of 4, a filter delay of 20 steps, an internal temperature cap of 2000 K, and a filter cap of 10 filters. All atoms in the YPGDV peptide are targeted by the digital filter.

3.1. Frequency Target. The frequency target is the range of frequencies that are desired for amplification or suppression by RDFMD. This parameter is the most important and is required before all others can be optimized. It is suggested that some form of frequency analysis on MD simulations is performed to validate the frequency target, or alternatively the response of the system to different filters can be tested. If the range chosen is too broad, energy will be put into or removed from motions that are not intended for manipulation. If the range is too narrow however, the frequencies at which desirable motions occur may be missed.

A 16 678 step microcanonical (NVE) ensemble simulation was performed from the equilibrated YPGDV system for a Fourier analysis, the results of which are reported elsewhere.[2] A 16 678 step simulation with a time step of 2 fs yields approximately a 1 cm^{-1} resolution. A previous suggestion for an RDFMD frequency target of 0−25 cm^{-1} was based upon an amplitude spectra argument that showed the majority of motion in the backbone ψ and ϕ angles (labeling as in Figure 1) occurring at very low frequencies.[2] A further frequency analysis using the Hilbert-Huang Transform[24,25]

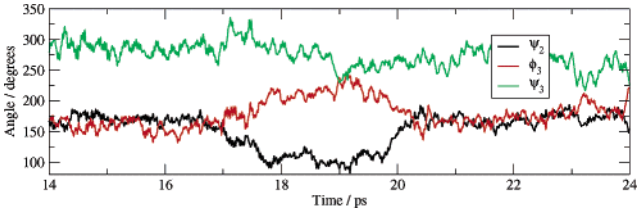

Figure 2. Relevant dihedral angles during the YPGDV conformational change event.

showed the presence of significant energy in the backbone dihedral motions of YPGDV, during conformational changes, with frequencies in the 0−25 cm^{-1} region.[26]

To examine the frequency target further, the Empirical Mode Decomposition[24] Method (EMD) has been applied to the simulation trajectory. EMD is a method of decomposing a signal into a set of intrinsic mode functions (IMFs). IMFs are iteratively determined by following the highest frequency motion and then removing the created IMF from the signal. Thus a set of IMFs are produced, each describing the evolution of a frequency present in the input signal. The last IMFs describe the lowest frequencies present in the system and leave a trend component with no wavelike properties, similar to the dc component of a Fourier analysis. Full details of the algorithm are described elsewhere.[24]

Using EMD, each backbone dihedral angle trajectory from the NVE simulation was split into three components: high frequency 'noise' from interactions with higher frequency degrees of freedom (angle and bond stretching motions), the dihedral angle motion that persists throughout the simulation, and the low-frequency conformational motions resulting from intermittent large scale conformational changes in the system. Fortunately, the YPGDV system is inherently flexible, and a conformational event was captured within the short NVE simulation. This was the brief formation of a β 3-turn between 17.3 and 17.7 ps, as determined by the DSSP (Dictionary of Protein Secondary Structure[27]) algorithm using default options. This secondary structure change was accompanied, and followed, by significant rearrangement in ψ_2, ϕ_3, and ψ_3, as shown in Figure 2.

EMD performed on the ψ_2 signal produces 11 IMFs, the first two (those with the highest frequencies) describe motions over 200 cm^{-1}, as determined by Fourier transform (FT). The last five IMFs describe motions below 25 cm^{-1}. The physical relevance of the IMFs is best described when summed in frequency groups as shown in Figure 3. The Fourier transform of the summed signals are shown in Figure 4, showing the frequency ranges that incorporate the different signals. Although there is some frequency overlap between the signals, it cannot be determined whether this is due to the deficits of the EMD or the FT methods. Limitations of the application of EMD to MD analysis have been discussed elsewhere.[28]

Results show that the vibrations below 25 cm^{-1} are associated with the conformational change event itself. Therefore, the previous frequency target of 0−25 cm^{-1} would only be able to amplify physically relevant signals (those of significant amplitude present in the system) if a large scale conformational event such as that shown is occurring. To

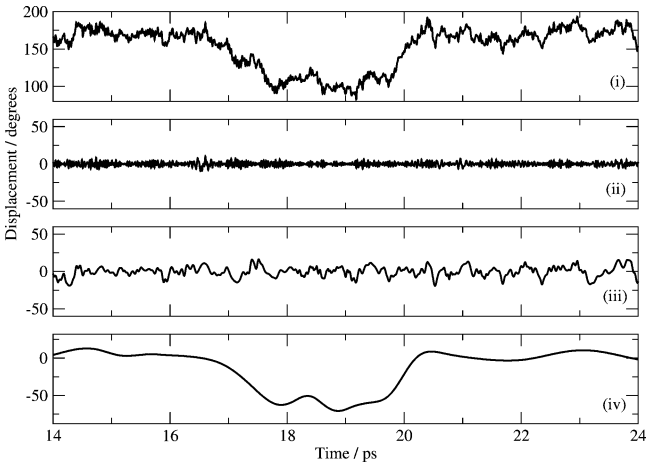

Figure 3. (i) ψ_2 signal during the conformational change event, (ii) sum of IMFs 1 and 2 showing motions above 200 cm^{-1}, (iii) sum of IMFs 3 to 6 showing motions predominantly between 25 and 250 cm^{-1}, and (iv) sum of IMFs 7 to 11 showing motions below 25 cm^{-1}.

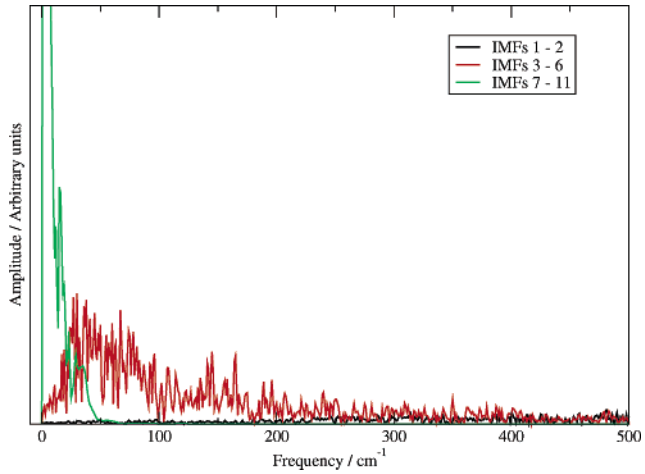

Figure 4. Fourier transforms of summed IMFs derived from the ψ_2 trajectory. A Hanning window was used.

target the dihedral angle motions present in the entire simulation, the frequency range 25−100 cm^{-1} is clearly significant. Amplifying frequencies from 0 cm^{-1} should be considered desirable in case motions are occurring at very low frequencies. It is clear that the inherent dihedral motions, i.e., vibrations within a potential energy well, occur predominately between 25 and 100 cm^{-1} and that these frequencies should be targeted if large-scale conformational change involving escape from local minima is to be induced.

To measure the response of the system to different frequency targets, filters designed to amplify frequencies below a specified value have been used. For each filter tested, 50 simulations of a single filter application to the YPGDV system were performed. The simulation start points were obtained from the equilibrated YPGDV system after a random reassignment of velocities at 300 K and 1000 steps of NPT MD simulation. Filters designed to target broader frequency ranges will see more degrees of freedom, and thus the amount of energy put into the system cannot be controlled with a constant amplification factor as previously used (eq 4). Instead the amplification factor is recalculated for each filter application so that the kinetic energy of the system is always increased by 25 kcal mol^{-1}. This is a comparable energy increase to that produced by a single filter application using the original parameter set.

For analysis, the root-mean-squared deviation (RMSD) is calculated between the trajectories of the eight relevant dihedral angles before and after a filter application. This is done for 100 steps (200 fs) before and after each filter application so that only the effects of the filter application are measured. The sum of the RMSDs for each filter application is considered to be a measure of induced conformational change, and from the 50 simulations an average and standard error can be calculated. Averaged results are shown in Figure 5 (a), showing a drop off in the amount of induced conformational change after 100 cm^{-1}. For higher frequencies, less conformational change is induced, and energy is being less efficiently placed in the dihedral angle motions. To demonstrate the importance of frequency specificity, a control simulation has been performed in which the filter used has a central coefficient of 1, and all other coefficients are zero. 25 kcal mol^{-1} of kinetic energy is therefore put directly into the internal velocities, across all frequencies. The upper and lower error limits of the induced conformational change for the control simulation are shown in Figure 5(a) by dashed lines. The amount of conformational change induced by this simple heating procedure is much less than that obtained using the targeted filters.

Repeated applications of filters have been shown to be more effective than use of a single filter and a greater amplification factor.[2] The different filter targets have been therefore been tested for longer filter sequences, using a delay parameter of 20 steps and a filter cap of ten. Kinetic energy is increased by 25 kcal mol^{-1} by each filter application, and the 10 equilibrated YPGDV states are used as simulation starting points. The induced conformational change is calculated as before for each filter buffer and summed across all 10 buffers for each run. The results are shown in Figure 5(b). A nonfrequency specific energy input has again been applied, and the error limits are shown with dashed lines. As the upper limit on the frequency target is increased, the induced conformational change is reduced. The nonfrequency specific energy input is once again far less efficient than any of the tested frequency targets.

There is a noticeable reduction in the conformational change induced by the 0−25 cm^{-1} filter when repeated filter applications are used. It is believed that this is due to amplifying a frequency range that includes only low amplitude motions (such as 0−25 cm^{-1} when a rare event is not occurring), rather than a range in which significant motions are apparent (such as 25−100 cm^{-1} where dihedral motions persist throughout the simulation). This effect is best investigated by analyzing an individual dihedral trajectory. Figure 6 (a) shows a breakdown of the ψ_2 dihedral angle from the first buffer of one of the simulations. Once again, the signal is separated by EMD into IMFs that can be grouped together as a high-frequency component (> 250 cm^{-1}), a dominant motion of intermediate frequencies (30 cm^{-1}−150 cm^{-1}), and a low-frequency motion (<60 cm^{-1}).

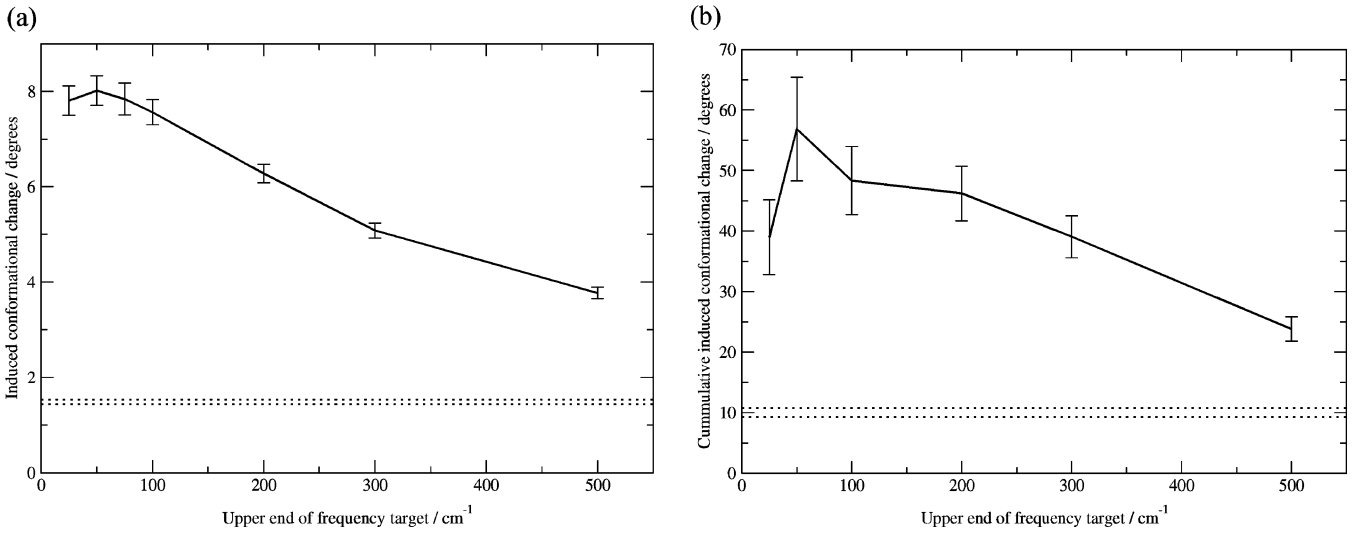

Figure 5. Measure of induced conformational change with different frequency targets. Dashed lines indicate error bounds of conformational change induced by an input of nonfrequency specific energy: (a) 50 applications of a single filter and (b) 10 filter sequences of 10 filters.

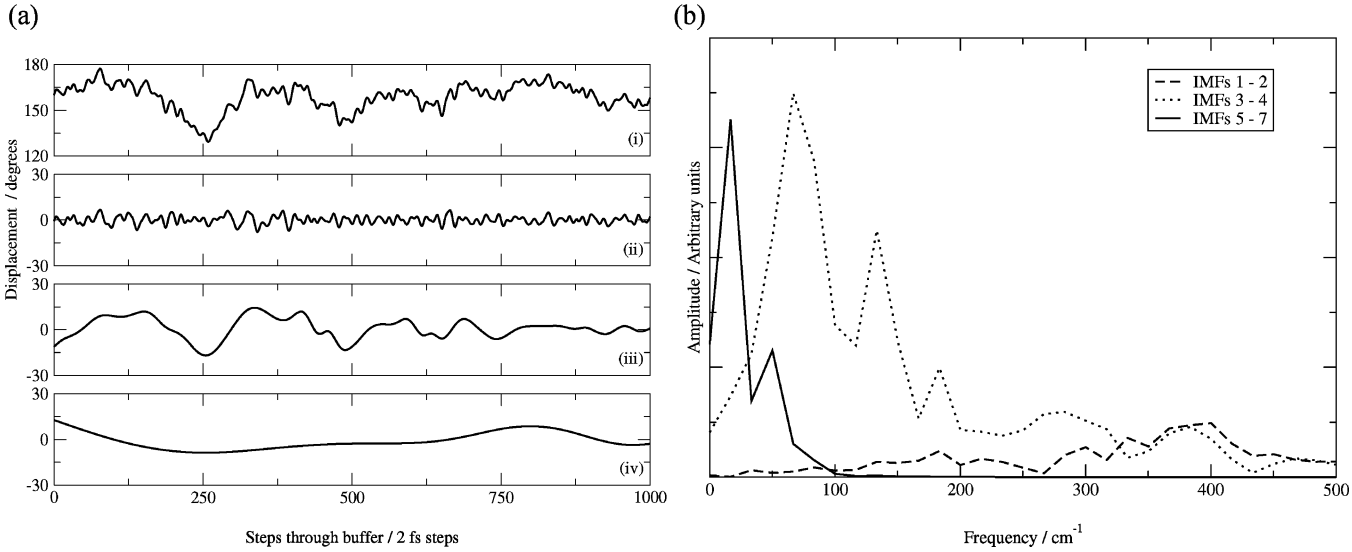

Figure 6. Decomposition of the ψ_2 trajectory from the first buffer of an RDFMD simulation: (a) (i) signal, (ii) sum of IMFs 1 and 2, (iii) sum of IMFs 3 and 4, (iv) sum of IMFs 5 to 7 and (b) Fourier transforms of split signal using a Hanning window.

The Fourier spectra of the grouped IMFs are reported in Figure 6(b). The intermediate frequency motion has significant energy around 70 cm^{-1}, outside the previously proposed frequency target for YPGDV. Figure 7 shows the effect of targeting either the low ($0-25$ cm^{-1}) or the low- and intermediate- ($0-100$ cm^{-1}) frequency regions identified here. Significant conformational motions are induced in both cases. However, amplifying the intermediate frequencies progressively targets the highest amplitude motion, yielding greater conformational change. In this analysis the frequency resolution is limited by the length of the buffer, and a reduced resolution can be expected compared with the 16 678 step NVE simulation analyzed previously.

The filter target can therefore be selected by either measuring the system's response to a range of filters or by predicting the desired frequency range from analysis of a sample trajectory. As has been shown, it is important to target frequencies relevant to the system, and the YPGDV results suggest no benefit in targeting above 100 cm^{-1}. The $0-100$ cm^{-1} range is suggested as optimal from both measurements of the system's response and from analysis of frequencies related to inherent dihedral motions.

3.2. Buffer Length. To apply a digital filter to a buffer of velocities, the buffer must contain at least the number of steps as there are coefficients in the filter. The larger the number of coefficients, the closer the frequency response of the filter will be to that for which it is designed. A shorter buffer requires reduced computational expense, but a filter of insufficient length will not produce a precise response and may result in undesired amplification or suppression of motions.

The previous filter target of $0-25$ cm^{-1} requires 1001 coefficients to achieve a reasonable filter response.[2] A higher cap on the frequency response suggested by the analysis presented here does not require as many coefficients to produce a sufficiently precise response. The responses of a

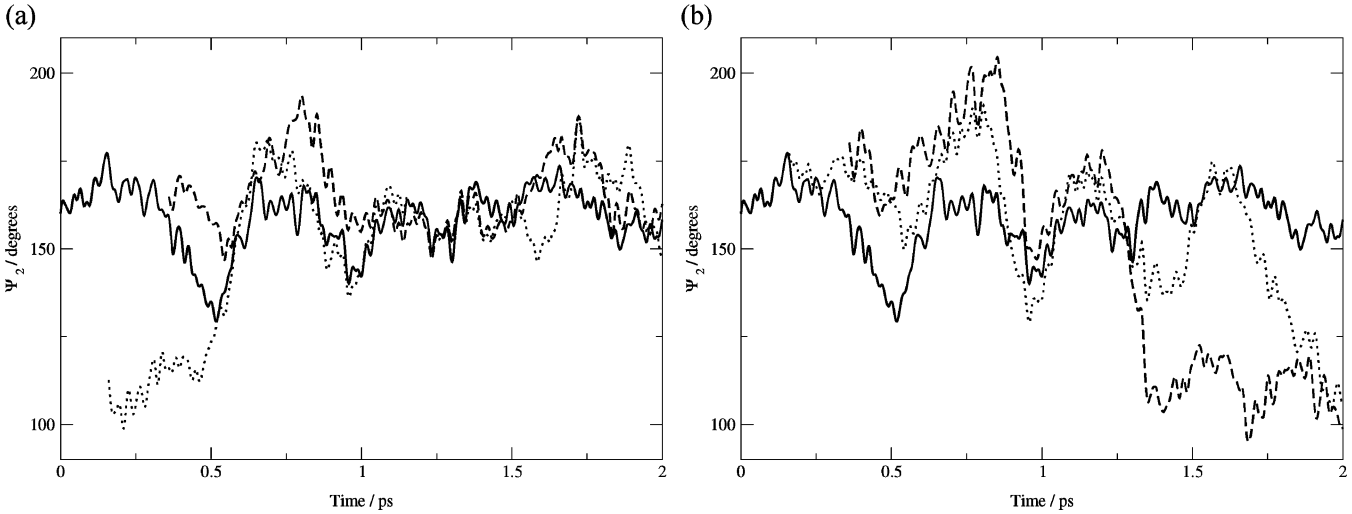

Figure 7. ψ_2 trajectory after 0 (solid), 5 (dotted), and 10 (dashed) filter applications: (a) 0−25 cm^{-1} filter applied and (b) 0−100 cm^{-1} filter applied.

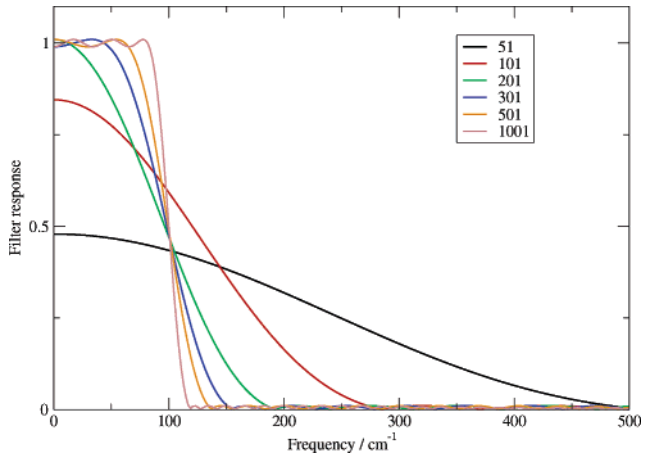

Figure 8. Filter responses of 0−100 cm^{-1} filters using a different number of coefficients (shown in legend).

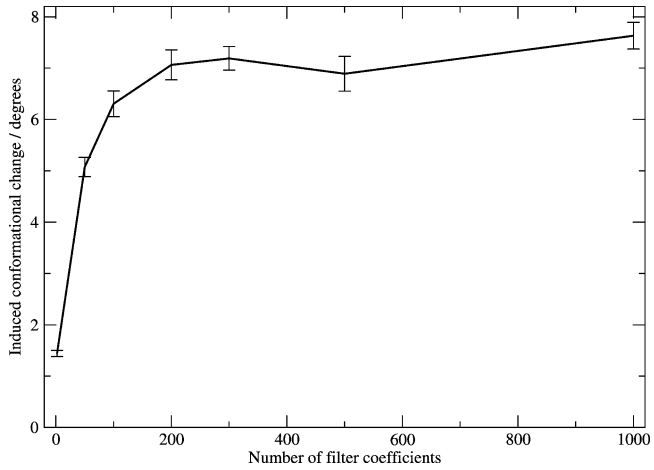

Figure 9. Results averaged from 50 RDFMD simulations using filters with different numbers of coefficients.

number of 0−100 cm^{-1} filters using different numbers of coefficients are shown in Figure 8.

RDFMD simulations using 0−100 cm^{-1} filters with different numbers of coefficients have been performed. 50 starting points for single filter applications were produced from the equilibrated YPGDV system with randomized velocities and 1000 steps (2 ps) of NPT MD. The amplification of each filter is adjusted to increase the system's kinetic energy by 25 kcal mol^{-1}. The induced conformational change is calculated as before, and the results are shown in Figure 9.

No benefit of using a greater number of coefficients than that required to produce an accurate filter response curve (200−300 coefficients) is seen. Fewer coefficients than this amplify higher than desired frequencies, for example, 100 coefficients targets 0−200 cm^{-1} and induces a similar level of conformational change to a filter targeting this region with 1001 coefficients in Figure 5(a).

3.3. Filter Delay. A delay between filters allows energy put into the system by the previous filter to dissipate and thus prevents overheating, keeping the system away from the internal temperature cap. Ideally the delay should be as long as possible so that the energy build up is slow and the simulation advances over the potential energy surface. However the effects of each filter application can quickly dissipate and too long a filter delay will result in a series of essentially independent filter applications.

RDFMD simulations starting from each of the 10 equilibrated YPGDV states were run using different filter delays. Other parameters are as specified in the original protocol, including an amplification factor of 4 and a 0−25 cm^{-1} filter. The induced conformational change produced by each filter application is shown in Figure 10 (a). The shorter the delay, the higher the initial induced conformational change, and the quicker the temperature cap is reached. Longer delays (> 40 steps) do not quickly reach the temperature cap, and more buffers are completed, each yielding progressive amplification of targeted motions. Very long delays (> 150 steps) yield constantly low conformational changes as energy is dissipating between buffers and the filter applications become more independent and less progressive. The energy build up is more clearly seen in Figure 10(b) in which the internal protein temperature after a filter application is plotted against the buffer number. Short delays quickly reach the 2000 K temperature cap, and long delays see no progressive increase in internal temperature.

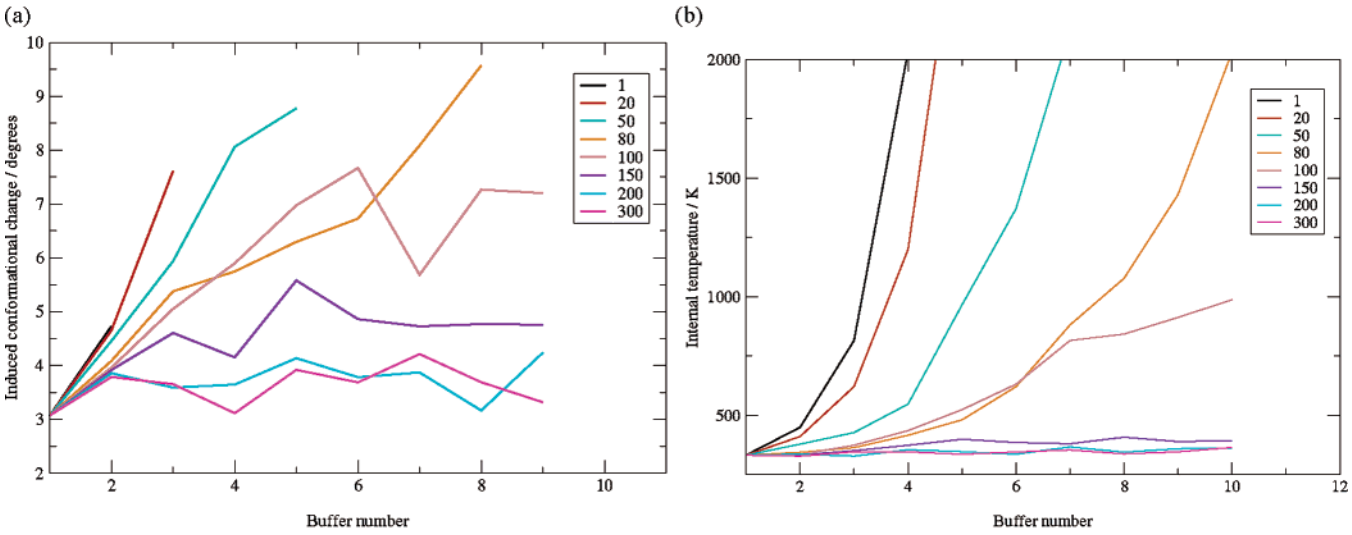

Figure 10. Effects of changing the filter delay parameter. Results averaged over 10 simulations: (a) conformational change induced in filter buffer and (b) internal protein temperature after filter application.

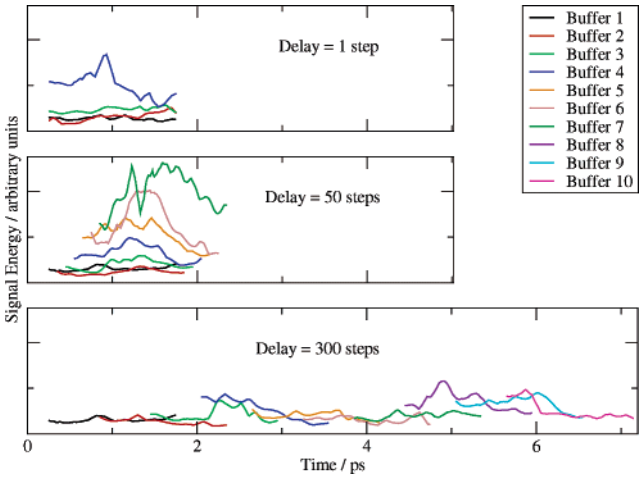

Figure 11. HHT of dihedral angle trajectories for different filter delay parameters. Buffers completed before reaching the internal temperature cap shown. *y*-axes are shown to the same scale for comparison.

Using the Hilbert-Huang Transform, it is possible to see the energy build up in dihedral angle signals (the signal energy) for the targeted frequency range. Figure 11 shows the results for one of the RDFMD simulation starting points showing energy in motions occurring in the 0−50 cm^{-1} region. A running average over 50 steps has been performed on the data and the first and last 100 steps of each buffer removed as frequency information at the edges of HHT data can be unreliable.[26] A delay of 1 step shows the greatest increase in energy for each amplification performed, but only four buffers are completed before reaching the temperature cap. A delay of 50 steps completes seven buffers as the system is able to relax between filter applications. The signal energy in the targeted frequency region reaches a higher level, before breaching the temperature cap, than the simulation with a 1 step delay, an improvement that shows the importance of the filter delay. A long delay of 300 steps shows little amplification, and each buffer does not necessarily reach higher levels of low-frequency energy than the last, suggesting that a long delay produces a nonprogressive protocol.

For the original RDFMD parameter set, a filter delay between 50 and 100 steps is clearly most suitable, showing progressive amplification without overheating the system.

3.4. Amplification Factor. A specified input of kinetic energy (as used to analyze the frequency target) or increase of temperature can be used to adjust the amplification of the velocities in RDFMD. A fixed amplification factor was originally implemented that adjusts the level of kinetic energy put into the system according to the amount that was already there. If there is a small amplitude, low-frequency motion present in the system, the filter will increase the kinetic energy by less than if the motion is of larger amplitude.

Regardless of the method of energy insertion, for a progressive protocol, sufficient energy must be put into the system so the effect of one filter application has not dissipated before the next. Equally, too great an amplification of velocities could overheat the system, a protocol that would risk denaturing larger protein systems.

A range of RDFMD simulations using different amplification factors from each of the 10 equilibrated YPGDV states has been performed. The induced conformational change is measured as previously described. The amount of kinetic energy in the low-frequency range is calculated from the velocities extracted by the applied 0−25 cm^{-1} filter ($\mathbf{v}_{Filt}$ in eq 2). A linear correlation is found between low-frequency energy and induced conformational change, even when results are averaged over several simulations (typically with a correlation coefficient > 0.8). Lines of best fit are plotted in Figure 12 (a) to the edges of the data set. The amount of conformational change induced when a specific quantity of energy is measured in the filter region increases as the amplification factor is raised. An amplification factor of 1 is inefficient, and later buffers do not explore regions of significantly increased low-frequency energy. RDFMD is stopped by reaching the maximum filter cap of ten. An

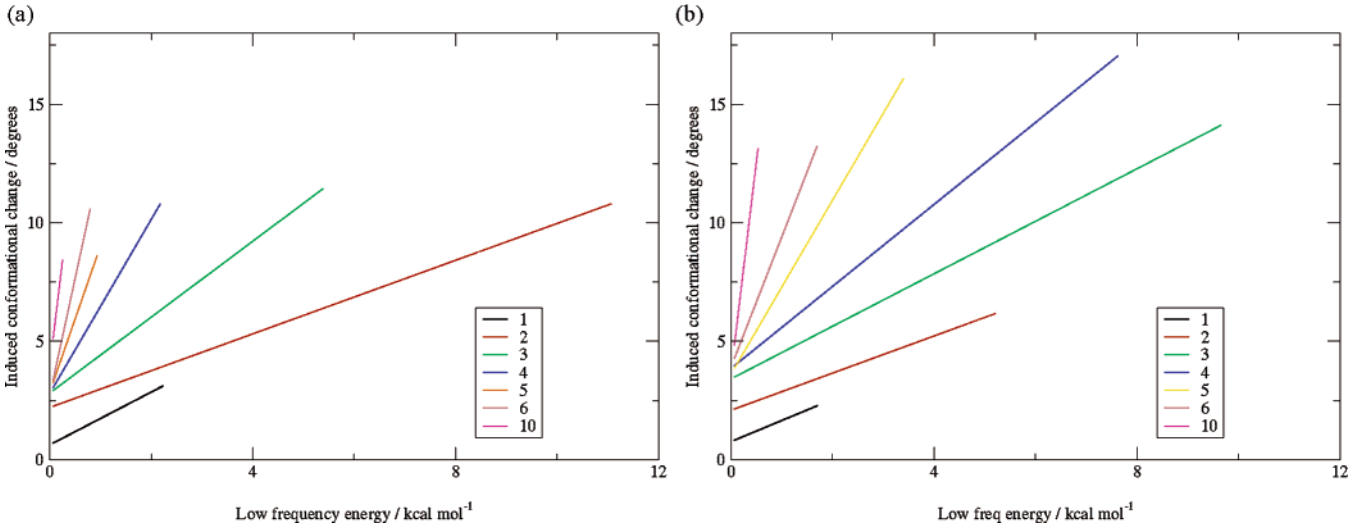

Figure 12. Effects of differing amplification factor (shown in legend): (a) filter delay of 20 steps and (b) filter delay of 50 steps.

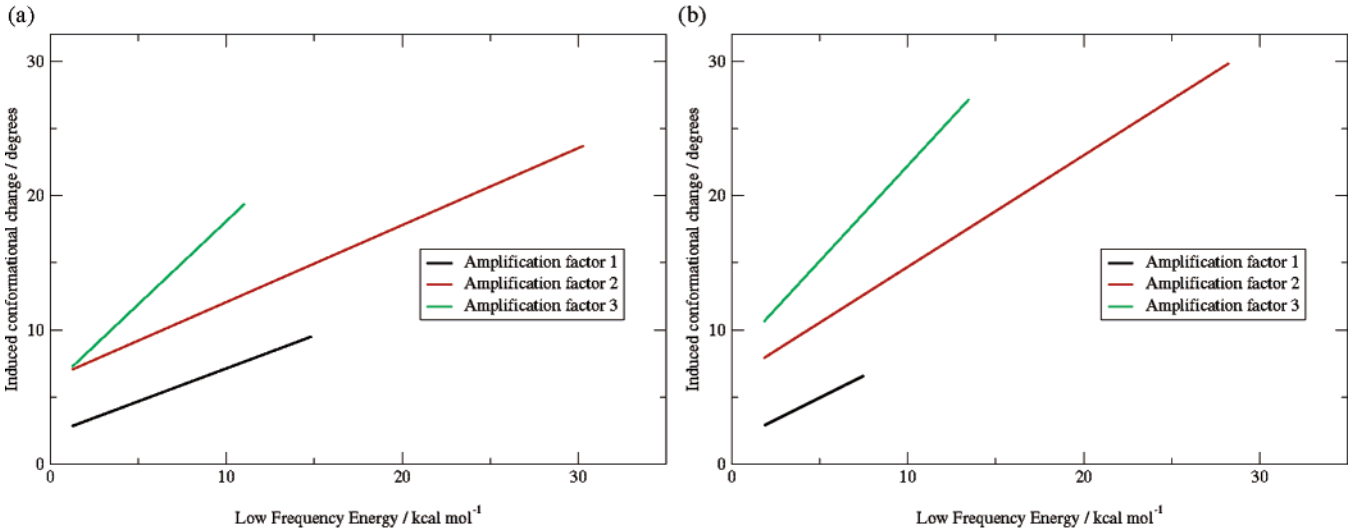

Figure 13. Results for a range of filter delay parameters and amplification factors. Data are averaged over 10 simulations and presented as previously described: (a) filter delay of 50 steps and (b) filter delay of 100 steps.

amplification factor of 2 yields greatly increased energy in the filter region and is clearly the most suitable compliment to the rest of the parameter set. At, and above, an amplification factor of 3, the internal temperatures reach the cap of 2000 K, and simulations are stopped. Although significant conformational motion is induced by high amplification factors, there are minimal increases in low-frequency energy over a small number of buffers before the temperature cap is reached. The protocol is therefore neither gentle nor progressive, as is desired.

The energy input required for gentle but progressive low-frequency energy amplification is heavily dependent on the rest of the parameter set. For example if the delay parameter is increased to 50 steps (a suitable choice as seen from previous analysis), an amplification factor of 2 shows slow energy build up and factors of 3 to 4 are more suitable, as shown in Figure 12(b). The improvement using the longer delay is clear, with 50% more conformational motions induced when comparing the best result to that of the simulations using a 20 step filter delay.

Once a frequency target and filter delay have been chosen for application to a system, analysis similar to that shown here will assist in tailoring the method of amplification.

3.5. Interdependence of Parameters. As previously discussed, the parameters used for RDFMD are heavily interdependent. Once a frequency target has been chosen however, optimizing the other parameters discussed here can be done systematically with a small number of trial simulations.

Results presented so far suggest use of a 0−100 cm^{-1} filter, a delay between 50 and 100 steps (0.1−0.2 ps), and amplification factors less of than 4. An accurate frequency response for a 0−100 cm^{-1} digital filter is produced using 301 coefficients (Figure 8). Averaged results for delay parameters and amplification factors in the suggested regions are presented in Figure 13(a),(b) as previously described. An amplification factor of 2 is shown to be the best choice for a filter delay of 50 or 100 steps. It is worth noting that for both parameter sets the levels of induced conformational

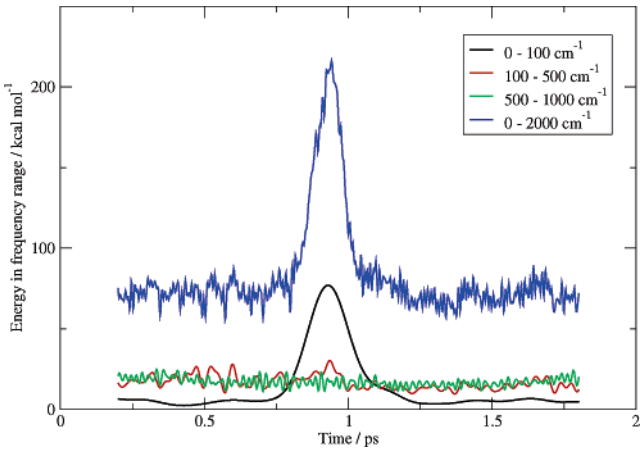

Figure 14. Kinetic energy in different frequency bands (shown in legend) of an RDFMD buffer.

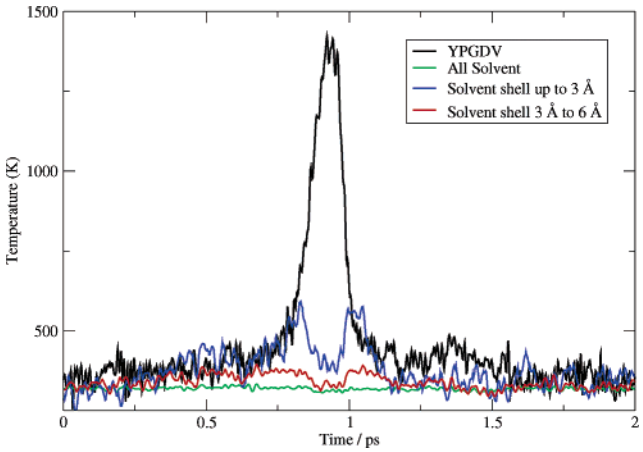

Figure 15. Temperature fluctuations in solvent and solute regions (shown in legend) across an RDFMD buffer.

change and of low-frequency energy are much higher than with previous protocols.

The protocols developed here have been optimized with the goal of promoting maximum dihedral motion while limiting the energy put into the system. The parameters should be applicable to any similar protein, or part of a protein, for which this is desired.

Discussion

It has been shown that energy dissipates between filter applications, and this energy can dissipate into either solute degrees of freedom with higher frequencies of motion than those targeted (i.e. nonselective heating) or into surrounding solvent molecules. To analyze this effect, the kinetic energy in motions of a particular frequency can be calculated across a buffer of velocities using a sliding digital filter. For example, a filter that multiplies all motions with frequencies of 0−100 cm^{-1} by a factor of 1 and all else by a factor of 0 can be produced with reasonable accuracy using 201 coefficients (Figure 8). By applying this filter to the first 201 velocities in an RDFMD buffer, an estimation of the velocities in low-frequency motions can be attained for step 101. The filter can be applied to steps 2−202 to estimate velocities at step 102 and so on across the buffer. Figure 14 shows the kinetic energy seen by a range of filters applied in this manner. The buffer analyzed is the last before reaching the temperature cap from an RDFMD simulation of YPGDV using the previously reported protocol (0−25 cm^{-1} frequency target, filter delay of 50 steps, 1001 coefficient filter, amplification factor of 4, and an internal temperature cap of 2000 K). The energy that has been added is clearly localized under 100 cm^{-1}, as desired.

To analyze the dissipation of energy into surrounding solvent molecules, the average temperature of solvent shells around the YPGDV protein has been calculated across a filter buffer. Figure 15 shows the results of a YPGDV RDFMD simulation using the previously reported protocol. Again the last buffer before reaching the temperature cap is used. The first solvent shell includes all water molecules with the oxygen atom within 3 Å of a protein atom. The number of water molecules included in the first shell fluctuates between 84 and 141 across this buffer. The second shell includes waters for which the oxygen atom lies within 3−6 Å of the closest protein atom and includes between 453 and 549 water molecules. The temperatures of the two shells increase on either side of the filter application, showing energy dissipating into the solvent. This is considered to be desirable, as the conformational flexibility of a protein will be linked to the mobility of surrounding solvent molecules.

Thus it has been shown that the energy dissipation observed after the application of a digital filter occurs predominantly into the solvent and not to higher frequency vibrations in the protein.

5. Application to DHFR

Escherichia coli dihydrofolate reductase (EcDHFR) catalyzes the reduction of 7,8-dihydrofolate (H_2F) to 5,6,7,8-tetrahydrofolate (H_4F) using nicotinamide adenine dinucleotide phosphate hydride (NADPH) as a reducing agent. During this reaction, the M20 loop (residues 15 to 20) adopts various conformations termed “closed”, “open”, and “occluded”.[29] NMR experiments show evidence of motions occurring in this region for apo-EcDHFR.[30,31] X-ray structures of various DHFR complexes can be found in the protein data bank[32] including entry 1RX2[33] for which DHFR exhibits a closed conformation complexed with folate and $NADP^+$, and entry 1RA9[33] with DHFR in an open form complexed with $NADP^+$. Here we show the application of RDFMD to the M20 loop, inducing a conformational change from the closed to open state.

Simulations of the thermal unfolding of EcDHFR have been published[34] revealing that the adenosine-binding domain (ABD, residues 38 to 106) partially unfolds at 400 K, and this region has been used to monitor the affect of RDFMD on protein stability.

An equilibrated solution structure of apo-EcDHFR has been produced from the closed 1RX2 structure. A cubic system with TIPS3P solvent extending at least 10 Å from the protein surface was prepared and neutralized by the addition of sodium ions, using the XLEAP module provided with AMBER 7.0.[35] Polar hydrogens positions and protonation states were determined by WHATIF.[36] The NAMD[18] package was used for further simulation with force field parameters taken from the CHARMM27 force field.[21]

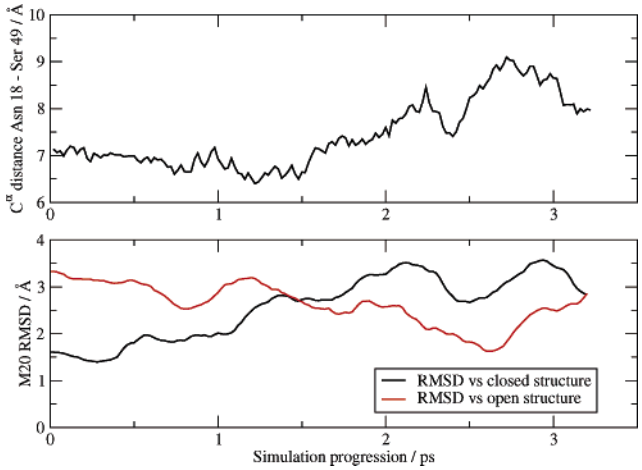

Figure 16. DHFR opening event during RDFMD simulation.

Mimimization was performed with the conjugate gradient line-search algorithm,[18] applying 25 000 steps to the solvent, 200 steps to the ions, 150 000 to solvent and ions, 3000 to the protein, and a further 30 000 steps to the entire system. Annealing was performed using a Langevin thermostat[22] with a damping parameter of 10 ps^{-1}, for 5000 steps at 50 K, 75 000 steps at 100 K, 10 000 steps at 150 K, 125 000 steps at 200 K, and 30 000 steps at 298 K. A Nosé-Hoover Langevin piston barostat[23] with a pressure target of 1 atm, a piston period of 400 fs, and a piston decay of 300 fs was applied for 125 000 steps of NPT simulation, followed by 50 000 steps with a thermostat damping parameter of 1 ps^{-1}. A 2 fs time step is used throughout the protocol with a switching function applied to Lennard-Jones interactions between 9.0 and 10.5 Å, SHAKE[20] applied to all bonds containing a hydrogen, and a PME treatment of electrostatics.[19]

A 4 ns NPT MD simulation was performed using the final equilibration parameters. Root-mean-squared deviations (RMSDs) have been calculated against the known closed (1RX2) and open (1RA9) structures with superposition across all secondary structure units. The RMSD of the protein secondary structure against the starting structure does not rise significantly, with an average of 0.82 Å and standard deviation of 0.10 Å. Motion of the M20 loop fragment is limited with a RMSD reaching 3 Å against the closed structure on three brief occasions, with no accompanying decreases in RMSD against the open structure. The closed conformer is characterized by a short distance between residues 18 (Asn) and 49 (Ser) where favorable polar interactions hold the M20 loop in place. The α-carbon distance between these two residues averages at 6.80 Å with a standard deviation of 0.64 Å for this simulation. The average RMSD of the ABD domain against the starting structure is 0.90 Å for the first 1.5 ns, with a standard deviation of 0.10 Å. It rises after 1.7 ns and remains stable for the remainder of the simulation, with an average of 1.29 Å and standard deviation of 0.17 Å for the final 2 ns. Full analysis of the simulation will be presented elsewhere; however, it does not suggest that DHFR leaves the closed conformation.

A suitable RDFMD protocol as suggested by YPGDV analysis has been chosen for application to DHFR. A frequency analysis of the dihedral angles in the M20 loop of DHFR yields similar results to those observed for YPGDV. A paper reporting this analysis in more detail is in preparation. The protocol includes use of a 301 coefficient filter targeting 0−100 cm^{-1} and a filter delay of 50 steps. The filter is applied to all atoms in DHFR residues 15 to 20. An amplification factor of 2 has been chosen based upon the results shown in Figure 13 and the internal temperature capped at 1500 K. 30 filter buffers have been performed for each simulation, and should an amplification factor of 2 bring the internal temperature above 1500 K, it is lowered so that the temperature cap is not breached and all buffers are completed.

Once again 10 simulations are performed using randomly assigned velocities at 300 K. Each simulation showed significantly increased loop mobility, with several opening events occurring. One of the clearest results measured the M20 loop RMSD against that of the 1RX2 closed structure rising from 1.38 Å to a maximum of 3.54 Å. As this occurs the loop RMSD against the 1RA9 open structure falls from 3.31 Å to 1.60 Å. The interresidue (carbon-α) distance between residues 18 and 49 rises from 7.13 Å to a maximum of 9.10 Å. Figure 16 shows the progression of these factors

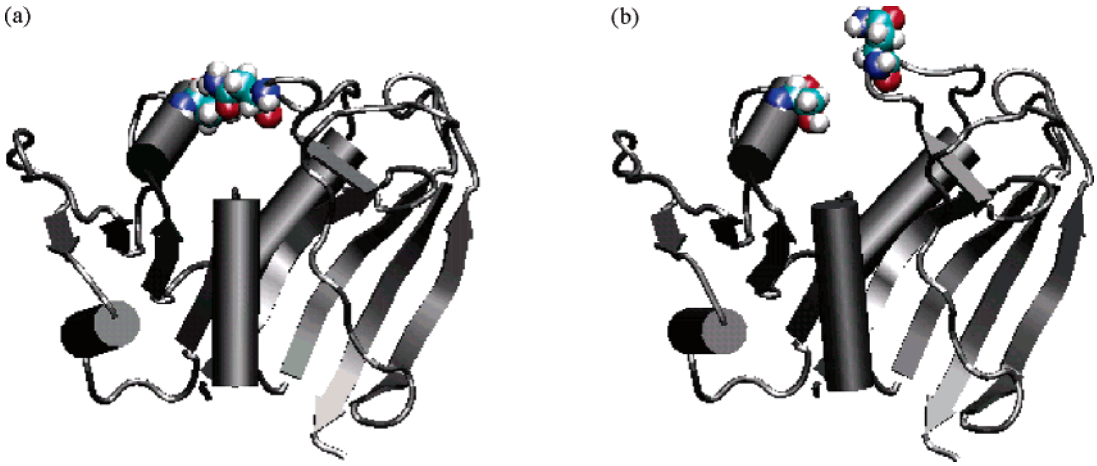

Figure 17. Conformers sampled during RDFMD DHFR simulation. van der Waals radii shown for residues 18 and 49: (a) conformation most similar to known closed structure and (b) conformation most similar to known open structure.

extracted from the forward components of the RDFMD buffers into a single trajectory. During the simulation the RMSD of the ABD domain does not rise above 1 Å. The most closed and open structures from this trajectory are shown in Figure 17.

The parameter set successfully used here has not undergone any further optimization, confirming the suitability of the derived protocol for application to similar systems. Positive results have also been produced using T4 lysozyme and HIV-1 protease (to be presented elsewhere).

Conclusions

In this paper we have discussed methods of assigning RDFMD parameters using the YPGDV pentapeptide as a test case. The goal has been to maximize the induced conformational change and promote the gentle and progressive amplification of low-frequency motions. The analysis methods developed here are intended as a guide for the application of RDFMD to any system and for any purpose. We anticipate, and are investigating, uses in protein folding, ligand binding, and inducing large-scale conformational motion in proteins of significant size.

The first parameter to be chosen is the frequency target. For this, frequency analysis can be performed on signals extracted from a sample trajectory, using Fourier or Hilbert based methods. Alternatively, the response of the system to filters amplifying different frequency regions can be used. Both methods have been presented here, suggesting filters targeting $0-100$ cm^{-1} for promotion of dihedral motions. Empirical Mode Decomposition has been used to separate dihedral signals into high-frequency noise from coupled degrees of freedom, a signal in the region $25-100$ cm^{-1} that persists throughout simulation, and a low-frequency component that only has significant amplitude during rare conformational events (targeted in previous studies). Use of a higher upper limit on the frequency target allows the length of the buffer to be dramatically lowered, thus reducing the computational expense of each filter application. This is particularly important when applying RDFMD to larger systems.

The number of steps between filter applications (the filter delay parameter) has been analyzed in detail, including the use of the Hilbert-Huang Transform, which has only recently been introduced to molecular dynamics simulations. Short delays show rapid energy increases that induce dihedral angle motions but quickly overheat the system, and so few filter applications are performed. Long delays show no progressive increase in low-frequency energy or induced conformational change, the effect of each filter application having dissipated before the next. Intermediate delays of $50-100$ steps show both slow energy increases and significant induced conformational change.

There are several methods of tailoring the amount of energy put into the system. A fixed amplification factor inputs energy according to how much is already present in the target region; this prevents energy being placed into motions of low amplitude that have little relevance to the system. Dynamically adjusting the amplification factor so that a set amount of energy is put into the system is also of use when comparing the system's response to filters that target different frequency ranges. An increase in either kinetic energy (as presented in frequency target section) or an adjustment dependent on the system temperature (used for application to DHFR) can be specified.

The resulting parameter set has been successfully applied to the M20 loop of DHFR: a frequency target of $0-100$ cm^{-1}, a filter delay of 50 steps, and an amplification factor of 2, which is reduced should the internal temperatures reach 1500 K. A conformational change is induced from the closed to open form, without affecting the rest of the molecule.

Further work includes the development of parameter sets for alternative applications, and the use of derived protocols with other systems of significant size, including T4 lysozyme and HIV-1 protease. Long MD and parallel tempering simulations on these systems will be presented in subsequent publications to validate the conformational motions observed using RDFMD.

Acknowledgment. This work has been funded by GlaxoSmithKline, EPSRC, and BBSRC. Many thanks to Dr. R. Gledhill and Prof. N. Huang for their guidance with the EMD method and Hilbert transform and to Dr. C. Woods with his assistance with the preparation of this paper. We should also like to thank Prof. W. L. Jorgensen for his generous provision of the MCPRO program.

References

(1) Phillips, S. C.; Essex, J. W.; Edge, C. M. *J. Chem. Phys.* **2000**, *112*, 2586.

(2) Phillips, S. C.; Swain, M. T.; Wiley, A. P.; Essex, J. W.; Edge, C. M. *J. Phys. Chem. B* **2003**, *107*, 2098.

(3) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.

(4) Tai, K. *Biophys. Chem.* **2004**, *107*, 213.

(5) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.

(6) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsovvelyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776.

(7) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249.

(8) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(9) Wu, X. W.; Wang, S. M. *J. Phys. Chem. B* **1998**, *102*, 7238.

(10) Wu, X. W.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512.

(11) de Groot, B. L.; van Aalten, D. M. F.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 240.

(12) *MATLAB 6.1.0*; The MathWorks Inc., Natick, MA, 2001.

(13) Tobias, D. J.; Mertz, J. E.; Brooks, C. L. *Biochemistry* **1991**, *30*, 6054.

(14) Karpen, M. E.; Tobias, D. J.; Brooks, C. L. *Biochemistry* **1993**, *32*, 412.

(15) Wu, X. W.; Wang, S. M. *J. Phys. Chem. B* **2000**, *104*, 8023.

(16) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161.

(17) Jorgensen, W. L. *MCPRO 1.4*, Yale University, New Haven, CT, 1996.

(18) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283.

(19) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(20) Ryckaert, J. P.; Ciccoti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(21) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(22) Paterlini, M. G.; Ferguson, D. M. *Chem. Phys.* **1998**, *236*, 243.

(23) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.

(24) Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. L. C.; Shih, H. H.; Zheng, Q. N.; Yen, N. C.; Tung, C. C.; Liu, H. H. *Proc. R. Soc. London, Ser. A* **1998**, *454*, 903.

(25) Bendat, J. S. *The Hilbert Transform;* Brüel & Kjær: Nærum, Denmark.

(26) Phillips, S. C.; Gledhill, R. J.; Essex, J. W.; Edge, C. M. *J. Phys. Chem. A* **2003**, *107*, 4869.

(27) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.

(28) Wiley, A. P.; Gledhill, R. J.; Phillips, S. C.; Swain, M. T.; Edge, C. M.; Essex, J. W. *Hilbert-Huang Transform Engineering: The analysis of molecular dynamics simulations by the Hilbert-Huang transform*; Marcel Dekker: New York, in press.

(29) Rod, T. H.; Brooks, C. L. *J. Am. Chem. Soc.* **2003**, *125*, 8718.

(30) Li, L. Y.; Falzone, C. J.; Wright, P. E.; Benkovic, S. J. *Biochemistry* **1992**, *31*, 7826.

(31) Falzone, C. J.; Wright, P. E.; Benkovic, S. J. *Biochemistry* **1994**, *33*, 439.

(32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.

(33) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586.

(34) Sham, Y. Y.; Ma, B. Y.; Tsai, C. J.; Nussinov, R. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 308.

(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.

(36) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52.

CT049970T