Force Field Optimization using Dynamics and Ensemble Averaged Data: Vibrational Spectra and Relaxation in Bound MbCO

Michael Devereux and Markus Meuwly*

Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland

Received November 11, 2009

Force field parameters are ingredients for realistic atomistic simulations of gas- and condensed-phase systems. Here we discuss the effect of including averaged data from explicit MD simulations in optimizing potential energy functions. It is shown that vibrational frequencies (FeC and CO stretch and FeCO bend) and CO vibrational relaxation times (($v = 1$) → ($v = 0$) (T_{10}) and ($v = 2$) → ($v = 1$) (T_{21})) in the active site of CO-bound myoglobin (MbCO) can be well represented with a single set of force field parameters. It is further demonstrated that parameters fitted in a subsystem of MbCO comprising the CO ligand, heme group and proximal histidine, are transferable to investigating the full protein and to providing quantitatively correct results. In particular, it is possible to calculate the CO and FeC stretch and the FeCO bending frequency to within ≈5%; the relaxation time of the first vibrationally excited state including quantum corrections of T_{10} ≈25 ps is calculated close to the experimental value (17 ps), and the ratio T_{10}/T_{21} ≈2 agrees favorably with experimental estimates. In contrast, following the more traditional approach of fitting frequencies from analyzing the Hessian matrix leads to a force field that captures frequencies correctly but not relaxation of vibrations.

INTRODUCTION

Improved parametrization techniques are essential to yield robust force field parameters for realistic atomistic simulations of chemical, biological, and physical processes. While general force fields for common building blocks, such as amino acids, are widely available, polarization and charge transfer alter the electron distribution of a moiety as a function of its chemical environment. A modified set of force field parameters for detailed and quantitative applications may, therefore, be required. Also, parameters that are optimized to accurately represent one measured physical property of general importance may be less accurate when describing other physically observable characteristics. For example, it is often the case that parameters for use in MD simulations are fitted to static data, such as normal-mode frequencies,[1,2] which does not guarantee an accurate representation of energy transfer between modes. Many such dynamic processes can only reliably be included in a fit by explicitly carrying out MD simulations in each optimization cycle.

Parameters for general use should additionally be robust and able to encapsulate as much of the known behavior of a system as possible, so that the model is consistent with all that is known from experiment. A further consideration for atomistic simulations is transferability of parameters between a (small) model compound, often used for fitting, and the full system, such as a protein. A number of methods are currently available to optimize bonded and nonbonded terms of force fields for atomistic simulations. As already mentioned, harmonic bond force constants are often fitted from second derivatives of the energies of internal degrees of freedom by using ab initio calculations[1] and, aditionally, spectroscopic data.[2] Electrostatic interactions are typically modeled using point charges that have been fitted to recreate a molecular electric field. It can be difficult to find a compact set of point charges, each associated with an isotropic electric field, that accurately represents the often strongly anisotropic electric fields of real molecules.[3] The ab initio electrostatic potential calculated around a molecule in different conformations is typically employed in place of electric field strength, for example in the fitting of restrained electrostatic potential (RESP)[4] charges used with the AMBER[5] molecular dynamics (MD) package. Charges may then be further refined to model empirical or ab initio interaction energies with its environment, such as solvent molecules.[2] More detailed force fields are also being developed, fitting extra point charges at electron lone-pair sites and including polarization effects by adjusting point charges in different chemical environments.[6] Remaining nonbonded interaction terms such as the Lennard-Jones potential[7] require parametrization to define the form of the interaction energy arising from stabilizing long-range dispersion interactions and from strong short-range repulsion. A variety of commonly used force fields[5,8,9] fit general sets of atomic Lennard-Jones parameters that give optimal balance between dispersion and repulsion interactions for a wide range of related systems. Realistic Lennard-Jones parameters are particularly important when modeling condensed-phase systems, and the performance of MD simulations using these parameters can be assessed by the accuracy of calculated properties, such as solvation free energy.[10]

In the above examples, the parameters described are usually not experimentally observable properties (partial charges, Lennard-Jones parameters). Instead, they are quantities that can be combined to build a workable model that represents the underlying physics and are validated by

* Corresponding author. Telephone: +41 (0)61 267 38 21. Fax: +41 (0)61 267 38 55. E-mail: m.meuwly@unibas.ch.

10.1021/ci9004404 © 2010 American Chemical Society
Published on Web 02/10/2010

agreement with measurable properties or with what is known from ab initio calculations. As values for nonobservable parameters cannot be obtained directly, methods must be found to fit and refine them using as large a range of experimental data as possible to allow realistic behavior in MD simulations. One such approach is the automated frequency matching method (AFMM),[11] which provides an automated environment to fit force field vibrational parameters, comparing normal modes from a given set of initial force field parameters with ab initio normal mode data to allow fitting of bond, angle, and torsional parameters. A Monte Carlo-like algorithm is used to vary the parameters in the fitting process. However, AFMM can currently only be applied to high-frequency modes. In the region of strongly delocalized modes, the automated assignment is too error prone. A different approach uses an automatic fit of force field parameters to steric, electrostatic, and dynamic properties within a quantum mechanics/molecular mechanics (QM/MM) approach.[12] Another recently developed method[13] automatically fits force field parameters to a given potential energy surface using a Monte Carlo simulated annealing procedure. Finally, commercial software exists that interfaces directly with ab initio software to fit van der Waals and other important force field terms, where standard force field data is missing or inadequate.[14]

Here we discuss a general approach to refine sets of any type of force field parameter using heterogeneous experimental and computational/theoretical data. It distinguishes itself from the methodologies outlined above by including explicit MD simulations in each iteration of the fitting process, ensuring that parameters lead to calculated data that is not only consistent with the known static but also dynamic properties of a system. The approach is related to that used, for example, in the development of the TIP3P potential[15] or to a spectroscopically refined all-atom force field for molecular sieves.[16] For the TIP3P water, the parameters were required to yield not only good agreement with static data for the water dimer but also accurate results for bulk properties, such as radial distribution functions calculated using Monte Carlo simulations. In contrast to the laborious approach of refining parameters by hand to converge to a robust set, however, observables are calculated using MD software and interfaced with the I-NoLLS[17] nonlinear least-squares fitting program. I-NoLLS is used to compute a set of trial parameters for each iteration of an optimization process, based on the difference between calculated and experimental data, and MD software is employed to assess the response of computed observables to each suggested set of force field parameters. Because such fitting problems can be susceptible to small parameter changes, interactive control of the progress is important. This feature is provided by I-NoLLS, which allows users to guide fitting of complex multidimensional problems. Incorporating dynamical data directly into the fitting process ensures that parameters are robust, so that results obtained from simulations are consistent with the full range of experimental data that is available.

In the following, the theoretical background and the computational representation of the systems are first presented. Then, the methods are applied to different typical situations arising in practice, and finally, conclusions are drawn.

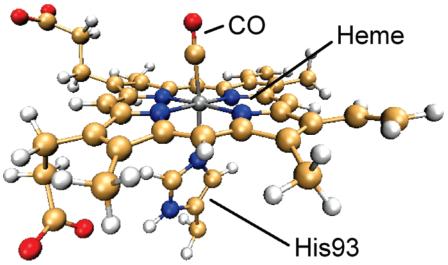

Figure 1. Heme, His93, and CO ligand comprising the model system used to study CO vibrational relaxation time.

METHODS

Computational Setup for the Test Systems. MD simulations were carried out using the CHARMM program[9] and the CHARMM22 force field.[18] The initial structure was bound carbonmonoxy myoglobin (MbCO) from the X-ray study of Kuriyan et al.[19] (Protein Data Bank reference, 1mbc) with hydrogen atoms added within CHARMM. Simulations of the full protein used stochastic boundary potentials[20] with a 16 Å sphere of water molecules around the protein binding site to improve computational efficiency, as the dynamics of the ligand and the active site are the focus of this study. Within a 12 Å inner sphere, centered at the protein heme group, the system was propagated using Newtonian dynamics, while the buffer region between 12 and 16 Å employed Langevin dynamics. For water, a TIP3P potential was used,[15] nonbonded interactions were truncated at 9 Å, and hydrogen atoms were constrained using the SHAKE algorithm.[21] The entire system was equilibrated at 300 K for 90 ps, spectra and relaxation curves were typically averaged over between 12 and 20 trajectories to obtain reliable data. The CO ligand and FeC_{CO} heme−ligand bonds were described using a Morse potential

$$V(r) = D_e(1 - \exp(-\beta(r - r_{eq})))^2 \qquad (1)$$

instead of standard harmonic bonded potentials. The dissociation energy ($D_{e,\,CO}$ and $D_{e,\,FeC}$), equilibrium bond length ($r_{eq,\,CO}$ and $r_{eq,\,FeC}$), and Morse exponent (β_{CO} and β_{FeC}) were fitted to density functional theory (DFT) bond energy profiles of the CO and FeC bonds, respectively. Bond energy profiles were obtained by stretching and compressing a bond at fixed intervals, as described previously.[22]

A subsystem, consisting of the heme, bound CO ligand, and proximal His93 group was used for the fitting procedure and for assessing parameter transferability. The model was constrained with harmonic force constants of 12 kcal/mol Å^{-2} to approximately maintain the conformation close to that in the protein. Constraints were applied to the positions of four atoms at the edge of the heme plane and to the terminal methylene carbon atom of His93 (furthest from the heme, see Figure 1). Additional tests with smaller constraints were also carried out (see below). The model was equilibrated at 300 K for 90 ps. Trajectories were 80 ps in length, with CO excitation (see below) taking place after the first 10 ps of simulation time. From the simulations, vibrational relaxation times and infrared spectra associated with the Fe−C, C−O, and Fe−C−O coordinates were calculated and compared with experiment. Typically, 12 such trajectories were averaged for each parameter set to obtain vibrational relaxation curves, and at least 6 were averaged for the infrared spectra.

Vibrational excitation of the C−O bond was induced by compressing the bond to increase the internal energy by 5.73 kcal/mol (24 kJ/mol), the energy added in pump−probe experiments to populate the first excited state.[23] Excitation to the second excited state was induced similarly by increasing the internal energy to approximately 11 kcal/mol (46 kJ/mol). Energy dissipation away from the excited C−O bond was then assessed by monitoring the bond length as it evolved with time. The vibrational energy $E_v(t)$ was calculated from the bond length by evaluating the potential energy at the maxima of each oscillation and averaged over a number of simulations, which gives the vibrational cooling curve, $\overline{E_v(t)}$, where the overbar represents an average over nonequilibrium trajectories. Averaging of the vibrational energy was carried out in two stages. To avoid instantaneous energy spikes, running averages over 200 fs time windows for each trajectory were first evaluated. The resulting cooling curves were then averaged for all 12 trajectories. The vibrational relaxation time T_{10} is calculated from the averaged curve according to[24]

$$\frac{\overline{E_v(t)} - \overline{E_v(\infty)}}{\overline{E_v(0)} - \overline{E_v(\infty)}} = \exp(-t/T_{10}) \tag{2}$$

$E_v(\infty)$ is the thermal value k_BT, where k_B is the Boltzmann constant, and T is the absolute temperature in Kelvin.[25] T_{10} is then fitted to the calculated CO cooling curve. The T_{21} relaxation time was fitted from at least 12 new trajectories with 11 kcal/mol added to the CO bond, which corresponds to excitation to $v = 2$. $\overline{E_v(t)}$ was again obtained by averaging over all trajectories, and T_{21} was fitted to the sum of two exponential functions (see eq 2). While fitting T_{21}, T_{10} was fixed at the value obtained from simulations of the first excited state with the same parameters and fits to eq 2. It should be noted that the exact amount of energy stored in the bond after the first one or two oscillations after excitation varies slightly, and thus, the averaged initial energies after excitation to the first and second excited states, $\overline{E_{v1}(0)}$ and $\overline{E_{v2}(0)}$, respectively, vary slightly between sets of simulations.

Power spectra $A(\omega)$ are calculated from the Fourier transform of the Fe−C bond length autocorrelation function, $C(t)$.[26] To construct $C(t)$, the Fe−C bond length is recorded at 2^n consecutive timesteps, typically corresponding to the last 65.5 ps of an 80 ps trajectory. Fourier transformation is performed with a Blackman filter[27] to minimize noise. $C(\omega)$ is given its Boltzmann weighting according to

$$A(\omega) = \omega(1 - \exp(-\hbar\omega/(k_BT)))C(\omega) \tag{3}$$

where $\hbar$ is the Planck constant, and T is the temperature. Spectra from 6 or more trajectories are then averaged. Both the Fe−C and C−O stretch frequencies are generally clearly visible in power spectra calculated from the Fe−C bond length. The Fe−C−O bending frequency is obtained by similar Fourier transform of the Fe−C−O angle.

Theoretical Background and Parameter Fitting. Nonlinear least-squares fitting was carried out using the I-NoLLS program,[17] which includes Gauss−Newton,[28] Levenberg−Marquardt,[29,30] and singular value analysis (SVA)[31] algorithms. All methods require calculation of a Jacobian matrix J, which is the matrix of partial derivatives of the computed data, y_i^{calc}, with respect to the parameters p_j

$$J_{ij} = \frac{\partial y_i^{\text{calc}}}{\partial p_j} \tag{4}$$

One element of J could be, for example, the partial derivative of one calculated vibrational frequency with respect to one particular force field bond parameter in the fit. Because the observables y_i are not known as analytical functions of the parameters p_j, elements J_{ij} have to be estimated by a finite difference scheme. For this, separate simulations with perturbed parameter values ($p_j \pm x_j$) are required, where x_j is a small perturbation. Least-squares optimization then takes place by minimizing:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{y_i^{\text{obs}} - y_i^{\text{calc}}(p_1...p_m)}{\sigma_i}\right]^2 \tag{5}$$

where y_i^{obs} is the value of the i^{th} target data, σ_i is the uncertainty assigned to that data point, n is the number of data points, and m the number of adjustable parameters. The uncertainties σ_i can also be used to bias the fit toward particular types of experimental data, for example to give a higher weight to information about intermolecular vibrations than to relaxation data.

Fitting problems based on observables from MD simulations are typically characterized by the large computational effort required to generate data for each iteration. It is, therefore, desirable to minimize the number of iterations to reach a robust set of parameters. In the following, an “iteration” is one accepted fitting round with a particular choice of parameters and data to be fitted, whereas a “cycle” can include several iterations. Between cycles, the number and type of parameters and/or observables can differ. The algorithms included in I-NoLLS are particularly well suited to optimization problems of this type because of the high degree of user interaction that is possible. Users are, for example, able to reject optimization steps that would lead to unphysical parameter values or to exclude parameters that have already converged or that are found to have little impact on calculated observables as the fit progresses. This level of control can significantly reduce the number of iterations in cases where automated gradient-following approaches result in unphysical parameter values and fail. Furthermore, while evaluation of J requires two trajectories to be run for each parameter to be fitted in each iteration (see above), the simulations are independent of each other and can be run in parallel. This considerably reduces the “walltime” for each iteration.

The computational approach pursued here consists of an interface that communicates user input, trial parameter values, and calculated observables from explicit simulations between MD software and I-NoLLS. A summary of the necessary steps is given below:

1. Initially, I-NoLLS reads the starting values of parameters and the list of target data (experimental and/or in silico) to be compared with calculated data from MD simulations. I-NoLLS then calls the user interface to calculate y_i^{obs} from MD simulations.
2. The user selects appropriate conditions for MD simulations (temperature, time step, simulation length, and

others). MD trajectories are run, and the user checks for successful completion of the MD trajectory/trajectories before proceeding.

3. The interface allows extraction of information (e.g., bond lengths or angles) from which observables are calculated using user-supplied scripts for the current parameter set. Essentially any property that can be extracted from an MD simulation can be used as observable data. Calculated data can be viewed and checked, and final values are returned to I-NoLLS
4. The ‘results’ module of I-NoLLS then allows to interact with the parameter optimization process. In addition to default parameter optimization with minimal user input, more control can be exercised by modifying the Jacobian to include or exclude any subset of parameters or data from the next cycle of fitting. Users can reject a given set of trial parameter values in favor of another suggested set. This can be useful, for example, when a given proposed set of x-variables clearly contains unphysical parameter values or is otherwise unlikely to yield meaningful results, so that running unnecessary simulations using these values can be avoided. Proposed steps in parameter space that appear too large or small can be scaled to stabilize the fitting and, again, avoid running costly simulations with inappropriate parameter values. Full information on use of I-NoLLS is available in the documentation.[17]
5. After fitting options have been selected, I-NoLLS again calls the interface to calculate the elements of the Jacobian J. The interface then enables users to calculate the properties associated with each perturbed set of parameters ($p_j \pm x_j$) required to construct J (see eq 4), and the results are used to suggest a new set of parameter values.

The simulation program was left unaltered for this work, and the interface can be easily adapted to link essentially any simulation software to I-NoLLS for a particular fitting problem.

Accurate Potentials for CO-Spectroscopy and Relaxation in Myoglobin. The following section describes a topical application to a problem of recent interest: heme and ligand parameter optimization in bound MbCO. Myoglobin is a protein with an abundance of associated well-resolved structural and spectroscopic data, which in combination with its relatively small size makes it an ideal test for theoretical study. Modeling of vibrational relaxation of the bound ligand in MbCO was the focus of a recent theoretical study[22] based on time-dependent experimental pump−probe data[23,32] and on spectroscopic measurements for the CO ligand and heme Fe vibrations.[33,34] In this study, it was demonstrated that the introduction of Morse potentials to describe the FeC and CO bonds in MD simulations with subsequent fitting of the Morse β parameters and FeCO bending mode angular force constant leads to accurate spectroscopy and realistic CO relaxation times.[22] The Morse potentials enable anharmonic coupling and faster energy transfer to take place between the CO and FeC bonds, while fitting leads to a small reduction in the separation between these vibrational frequencies, which then further increases the coupling efficiency. The sensitivity of vibrational relaxation times to change in vibrational frequency has been additionally demonstrated by isotopic substitution experiments.[35] In the following section, we extend these findings by use of a model in place of the full protein for calculating simulation data, and we further introduce the relaxation time of the second excited state as an additional experimental observable. We, thus, require fitted force field parameters to yield accurate spectroscopy for the FeC ($\approx$507 cm^{-1})[34] and CO (1949 cm^{-1} for the A_1 substate)[33] stretching modes, the FeCO bending mode (561−589 cm^{-1})[34] and to also yield realistic relaxation times for both the 1−0 ($T_{10} = 17$ ps)[23] and 2−1 transitions. Although T_{21} has not been precisely measured, it is generally expected to lie at around half of T_{10} ($T_{21} \approx 9$ ps).[36] Transferability of parameters fitted using MD simulations in the small model, consisting of heme, CO, and His93 (Figure 1), into MD simulations involving the full protein with solvated binding site is also discussed.

Fitting of Force Field Parameters using Static Data. As a first step, it is shown that following the most commonly used approach for parameter fitting of bond force constants leads to acceptable results for the spectroscopy but not to realistic coupling and energy transfer. FeC and CO Morse potential β parameters were fitted using normal-mode analysis of the heme model system to estimate vibrational frequencies. The Hessian matrix was evaluated with respect to each degree of freedom by making small structural perturbations, providing the second derivative. The Morse β parameters were adjusted until close agreement between normal mode and experimental stretch frequencies (within 1 cm^{-1}) was reached. Figure 2 shows the performance of the parameters fitted using normal-mode analysis in simulations of the full protein. Spectroscopy remains surprisingly accurate, although errors are slightly larger than predicted by normal-mode analysis in the model system (FeC stretch is blue-shifted by 23 cm^{-1} from experimental values and by 17 cm^{-1} from CO). Relaxation times, however, are much slower with $T_{10} = 309$ ps and $T_{21} = 114$ ps.

Parameter Transferability between Model System and Full Protein Simulations. Optimizing force field parameters for smaller subsystems using explicit MD simulations is computationally far less demanding than carrying out the optimization in the full protein. However, parameter transferability is not automatically guaranteed. A comparison of the observables obtained from MD simulations of the entire protein[22] with the same observables calculated from simulations of the model system, using the same parameters for both sets of simulations, is shown in Figure 3. CO relaxation after vibrational excitation takes place on a similar time scale in both the full protein (black curve in Figure 3a) and the model (green curve) (50.6 ps in the full protein vs 59.2 ps in the model system). In Figure 3, the initial energy of the excited state $\overline{E_{v1}(0)}$ for the experimental cooling curve is adjusted slightly to coincide with $\overline{E_v(0)}$ of the simulations to allow easy visual comparison of experimental and calculated T_{10} values. Good transferability is found, in this case, as relaxation of CO in the protein takes place mainly via bonded interactions with the adjacent heme, with cooling via electrostatic interactions playing a minor role.[22] Spectroscopy in the heme model, as seen in the power spectra of the FeC bond and FeCO angle (Figure 3b and c, respectively), demonstrates that the FeC and CO stretches as well as the lineshapes are well conserved between the protein and the smaller model (for the FeC stretch 550 vs 540 cm^{-1}; for the CO stretch 1915 vs 1901 cm^{-1} in the full protein and the model, respectively). The small shift in the vibrational frequency is likely to arise from the difference in

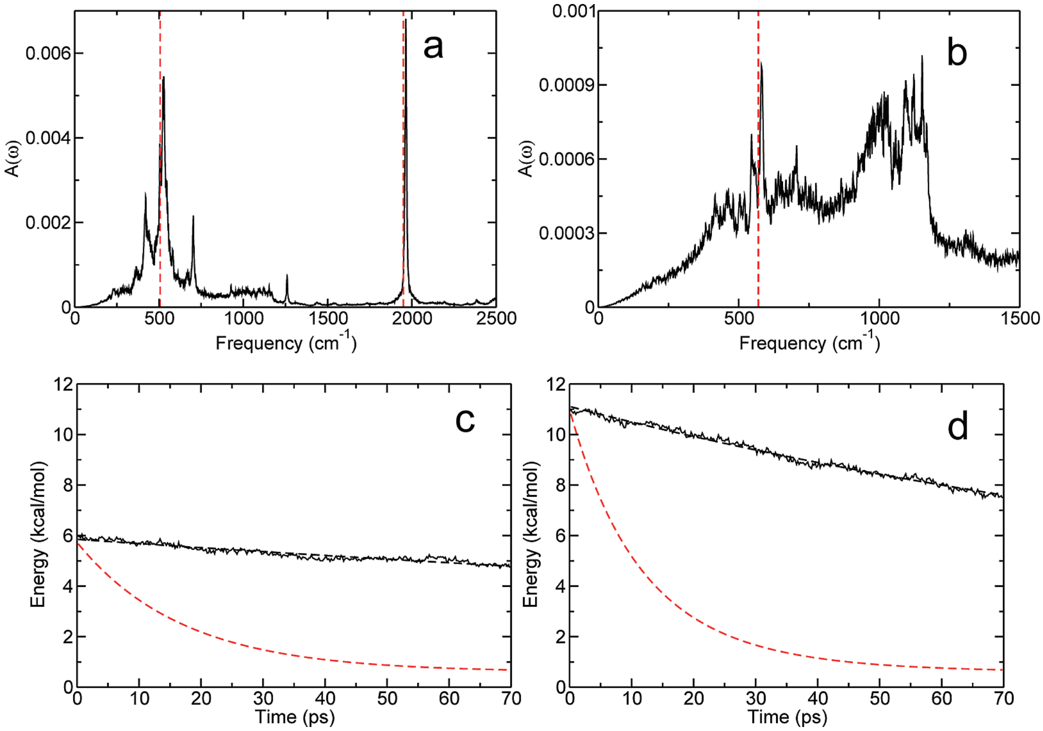

Figure 2. Calculated observables for the full MbCO protein system with parameters (FeCO-bonded Morse parameters and angular force constant) fitted in the heme model system using static data (normal-mode analysis). (a) FeC vibrational power spectrum, (b) FeCO bending mode power spectrum, (c) averaged CO bond energy $\overline{E_v}(t)$ as a function of time after excitation to the first excited state, and (d) averaged CO bond energy $\overline{E_v}(t)$ as a function of time after excitation to the second excited state. In panels a and b, calculated spectra are shown in black, and positions of experimental spectral peaks are shown as red-dashed lines. In panels c and d, CO relaxation curves from the simulations are shown in black (solid line is raw data, dashed line is fit), and relaxation curves corresponding to experimental T_{10} and T_{21} times are red-dashed lines.

electric field experienced by the ligand in the presence of the protein and solvent, which has a larger impact on vibrational frequency than on relaxation times. The FeCO bending mode of the model, however, lacks the clear peak that was present in the protein at around the experimental value. The remainder of the FeCO power spectrum appears similar to that of the full protein, although the broad peaks at around 470 and 750 cm^{-1} are more intense relative to other peaks in the model. The reason for the absence of this peak is not clear. It may relate to the constraints that need to be applied to peripheral atoms of the heme unit to maintain a realistic geometry in the absence of the surrounding protein (see Methods Section). Reducing the strength of the constraints does not, however, affect these results. Another possibility is that the absence of the influential nonbonded interaction to the adjacent His64 in the protein binding site allows the ligand to move more freely in the model, affecting the FeCO bend particularly strongly.

The optimized parameters from the small model system, therefore, serve as a useful starting point for more exhaustive investigations of the solvated protein. The FeCO bending mode is known to be coupled to several other degrees of freedom,[22] and as described above, the relationship between the calculated power spectra and the experimental IR spectrum is ambiguous. Detailed inspection shows that a feature at 470 cm^{-1} in the FeC power spectra of both the model and full protein is particularly susceptible to changes in the FeCO bending force constant and is, thus, most likely related/coupled to the bending mode. Consequently, this feature was followed during subsequent fitting of FeCO parameters in the heme model system, as described below.

Systematic Parameter Fitting Including Dynamics Data. Given the transferability of parameters between the model system and the full protein, parameter optimization from dynamics-based data was performed for the subsystem described in the Methods Section. The overall fitting process is summarized in Table 1, and a more detailed account of the convergence of one cycle is given in the next section. Initial parameters were Morse parameters fitted to DFT bond energy profiles of the CO and FeC bonds and the CHARMM22 FeCO angular force constant. As Table 1 and Figure 4 demonstrate, the initial parameters lie far from their optimized values and do not lead to either accurate spectroscopy or realistic relaxation times in MD simulations.

The first cycle, including two fitting iterations (Cycle 1), included only spectroscopic data. Supervising the fit in this fashion allows rapid improvement of parameters. After two iterations, the spectroscopic data is accurately calculated, and another observable—the relaxation of the first excited state, T_{10}—is added to the fit. After the next cycle with two iterations (Cycle 2), a set of parameters is obtained that is a useful compromise between accuracy of the spectroscopic and relaxation data, with $T_{10} = 65$ and $T_{21} = 29.5$ ps. Additional iterations (not shown) do not significantly improve the value of χ^2 (eq 5) but rather redistribute the total error between the observables. In comparing with experiment, it has to be kept in mind that for both infrared absorption line profiles and vibrational relaxation rates, quantum effects can play a role, and these observables are calculated here within the framework of classical mechanics. Quantum corrections for both observables are available and have been investigated in the past. For the vibrational spectra, quantum corrections

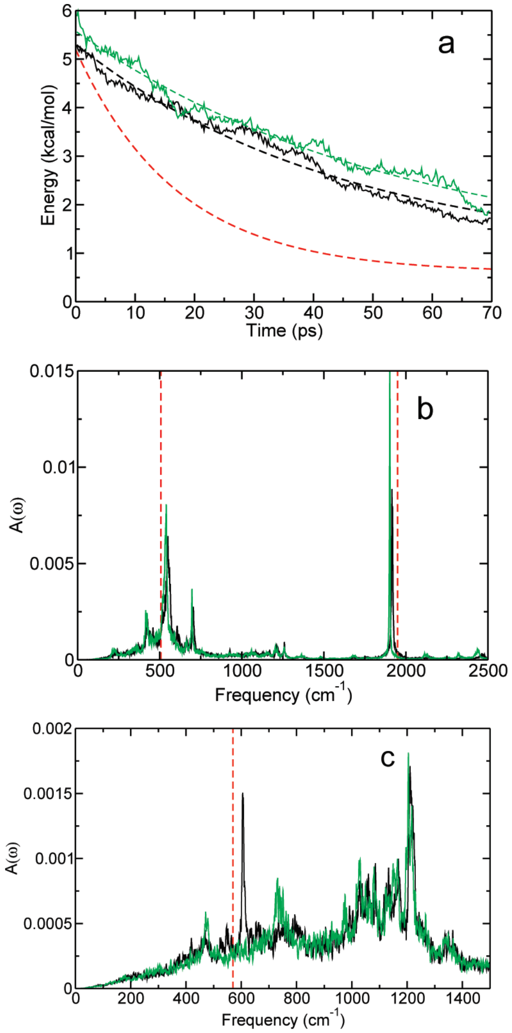

Figure 3. Comparison of calculated observables from the full MbCO protein and the heme model system using parameters optimized in the full protein (FeCO-bonded Morse parameters and angular force constant). (a) Averaged CO bond energy $\overline{E_v(t)}$ as a function of time after excitation to first excited state, (b) FeC vibrational power spectrum, and (c) FeCO bending mode power spectrum. In panel a, relaxation curves in the full protein (black) and the heme model system (green; solid lines represent raw data and dashed lines of corresponding color are fitted). A relaxation curve corresponding to the experimental T_{10} time is included as a red-dashed line for reference. In Panels b and c, calculated spectra from the full protein (black) and spectra from the heme model (green); positions of experimental spectral peaks (red-dashed lines).

to eq 3 primarily influence the band shape and the maximum intensity but not the position of the lines,[37] which is of primary interest here. It was further demonstrated that calculation of vibrational spectra via Fourier transform of the time correlation function of the dipole moment automatically includes dynamical effects, such as motional narrowing, and so leads to excellent line shapes. For relaxation times, a recent study has established that nonequilibrium simulations for the 1 → 0 relaxation of the CO stretch leads to relaxation rates approximately two to three times longer than rates from quantum simulations and that nonequilibrium MD simulations may well account for quantum energy flow, because such simulations treat system and bath in a consistent manner.[38] It was even found that a purely classical approach is superior to a mixed quantum−classical description. Applying such a correction to values of $T_{10} = 65$ and $T_{21} = 29.5$ ps leads to corrected relaxation times of $T_{10} \approx 25$ and $T_{21} \approx 10$ ps, which compare quite well with experiment.[23]

If relaxation times are included at the beginning of the fit (in Cycle 1), then the very slow cooling associated with the initial parameters prevents convergence, as the 80 ps simulation time is not sufficient to extrapolate to large T_{10}. Small fluctuations in the averaged trajectories lead to large random fluctuations in the calculated data when extrapolated to predict T_{10}, which then dominate the fit. The result is that purely automated fitting leads to unphysical parameters and to no convergence, because the fit attempts to reduce the large error in the relaxation time at the cost of the spectroscopy.

To demonstrate the ability to bias the model further toward experimental relaxation times, the T_{21} relaxation time is added and two additional iterations (Cycle 3) are carried out. A comparison between the original and the newly fitted parameters from iteration two of Cycle 3 is shown in Figure 4. Relaxation times of $T_{10} = 33$ and $T_{21} = 9$ ps are found, with very good agreement in the FeCO bending mode frequency of 590 cm^{-1}, while the positions of the FeC and CO peaks in the power spectra are shifted by +43 and −98 cm^{-1} from their experimental positions, respectively. Thus, the decision of which parameter set to use in further work involves an appropriate compromise determined by the given application and by the subjective choice of the user of which observables are more relevant. Further improvement, to achieve more accurate spectroscopy for the same relaxation time, might be facilitated to some extent by including additional parameters for surrounding degrees of freedom in the heme group.

The empirical relationship $T_{10}/T_{21} \approx 2$ inferred from experiments is generally well captured by the parameters determined here.[36] T_{21} is calculated to be a factor of between two and four times faster than T_{10} for all parameter sets taken from the final three fitting iterations (see Table 1). It should be noted that T_{21} is sensitive to and correlated with T_{10}, which makes T_{21} also susceptible to noise and to the number of simulations over which the raw data is averaged. The reason for the improved coupling at higher vibrational energy in the classical MD simulations is likely to be the additional sampling of longer bond lengths, where the bond energy profile is increasingly anharmonic. The density of states available for coupling to the lower frequency FeC stretch is, thus, larger, which improves the coupling efficiency.

As a final refinement, the optimized parameters from the model system were used in one additional fitting cycle with two iterations of solvated MbCO. This leads to only small changes in both parameter values and observables (see final column of Table 1). The deviations in the spectroscopic observables were +71.6 and −58.9 cm^{-1} for the CO and FeC stretches, respectively, but the sum of the absolute errors in CO, FeC, and FeCO vibrational frequency remained essentially constant. Thus, the errors are mainly redistributed, providing further evidence of a high level of transferability of these parameters between the model system and the full protein.

Convergence of Parameter Fitting. It is also important to investigate the ability to converge a fit—characterized by an approximately constant χ^2 as a function of the iteration—for a given set of observables and uncertainties. For this, the parameters of Cycle 1, iteration two in Table 1 with

Table 1. Convergence of Fitted CO and FeC Morse β Parameters (Å $^{-1}$) and Fe−C−O Angular Force Constant (kcal mol^{-1} radian^{-2}) over Three Cycles with Two Fitting Iterations in a Heme Model System[a]

		cycle 1		cycle 2		cycle 3		protein
	iter 0	iter 1	iter 2	iter 1	iter 2	iter 1	iter 2	iter 2
β_{CO}	2.346	2.132	2.114	2.085	2.061	2.045	1.981	2.001
β_{FeC}	2.265	2.743	2.801	2.922	3.093	3.094	3.322	3.460
k_{FeCO}	35.0	37.5	53.8	54.4	54.5	54.5	49.5	43.9
$\delta\nu_{CO}$	−177.9	−2.3	8.8	29.7	41.9	57.7	97.9	58.9
$\delta\nu_{FeC}$	124.3	11.3	5.7	−1.4	−26.4	−18.8	−43.2	−71.6
$\delta\nu_{FeCO}$	11.9	15.5	−21.1	−28.8	−29.3	−34.4	−10.5	−23.2
δT_{10}	−3805.2	−358.4	−365.5	−80.6	−48.3	−34.4	−16.5	−15.2
δT_{21}	−888.2	−797.1	−60.8	−79.8	−20.5	−20.1	0.4	−12.7

[a] A cycle consists of one or more iterations including the same parameters and observables in the fit. In one additional cycle of two fitting iterations (Cycle 3), the fit was biased by increasing the importance of accurate CO cooling times (by decreasing the corresponding uncertainty). Differences between experimental and calculated observables ($\delta\nu$ [cm^{-1}] for frequencies and δT [ps] for relaxation times) are reported for each set of parameters in the lower half of the table. The column labeled ‘protein’ shows the results of one additional cycle of two fitting iterations performed in the full protein, starting from the optimized parameters after three cycles in the model system.

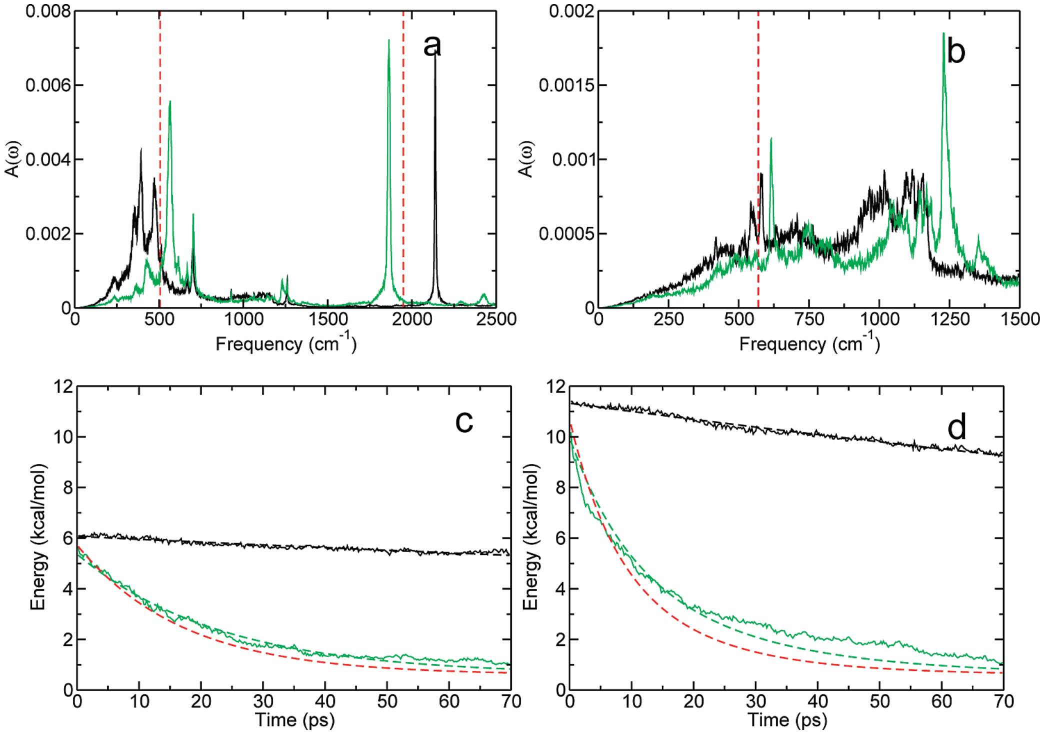

Figure 4. Comparison of calculated observables in the heme model system using fitted and unfitted parameters (FeCO-bonded Morse parameters and angular force constant). (a) FeC vibrational power spectrum, (b) FeCO bending mode power spectrum, (c) averaged CO bond energy $\overline{E_v}(t)$ for $v = 1$, and (d) averaged CO bond energy $\overline{E_v}(t)$ for $v = 2$. In panels a and b, calculated spectra using the unfitted (black) and fitted parameters (green; Cycle 3, iteration 2 of Table 1). Positions of experimental spectral peaks are shown as red-dashed lines. In panels c and d, CO relaxation curves from simulations with unfitted (black) and fitted parameters (green; solid lines represent raw data, dashed lines of corresponding color are fitted). Relaxation curves corresponding to experimental T_{10} and T_{21} times are included as red-dashed lines.

uncertainties σ_i of 1.0 cm^{-1} and 1.0 ps for vibrations and relaxation times, respectively, were used, and the corresponding χ^2 is shown in Figure 5a. The fit was carried out for seven iterations in the heme model system. All suggested trial parameter sets were accepted, and all variables and observables were included for the full seven iterations. Figure 5a shows a rapid reduction in χ^2 by the second iteration, with little improvement by around the fourth iteration. The rapid reduction in χ^2 by the second iteration also justifies the procedure applied in the previous paragraph (see Table 1), where new observables were included after two iterations. Faster improvement might have been achieved by rejecting the first set of trial parameters and selecting an alternative set of parameters and/or observables, thereby removing the need to perform one computationally expensive evaluation of the Jacobian matrix. Because numerical derivatives are used to calculate the Jacobian to suggest each new set of trial parameters, individual simulations are susceptible to fluctuations which can slightly destabilize the fit, as is the case for iterations one and six in Figure 5a.

The evolution of the calculated observables for each iteration is shown in Figure 5b. The change in error in the calculated values (observed − calculated) is plotted, demonstrating a general reduction during the fit. Comparison with Figure 5a illustrates that errors can be redistributed between observables without improving or degrading the overall fit. There is considerable fluctuation in the error associated with

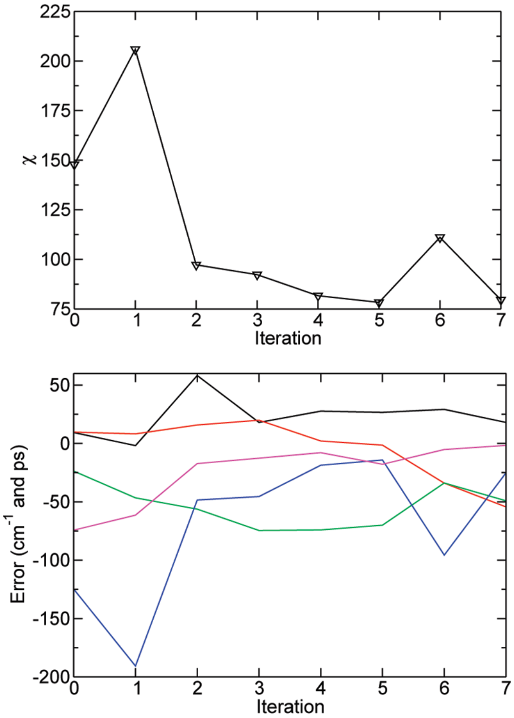

Figure 5. (a) Convergence of the fit for consecutive iterations of the FeCO angular force constant, CO, and FeC Morse β parameters in the model system. Calculated FeC, CO, and FeCO vibrational spectra and CO T_{21} and T_{10} relaxation times are fitted to experimental values. Uncertainties of 1.0 cm^{-1} and 1.0 ps are used in fit. (b) Evolution of error in calculated observables for each fitting iteration. CO stretch (black), FeC stretch (red), FeCO bending frequency (green); relaxation times T_{10} (blue) and T_{21} (magenta) are shown. Errors for spectroscopic observables are in cm^{-1}, those for relaxation times are in ps.

each calculated observable, while χ^2 for the corresponding iteration remains relatively stable.

CONCLUSIONS

In the present work, we demonstrate the feasibility and impact of explicitly including dynamics information in fitting force field parameters. User supervision and control are indispensable in difficult fitting problems with potentially high dimensionality and nonlinear or unknown physical relationships between parameters and calculated dynamical data. The necessary input for each set of parameters was calculated from and averaged over multiple explicit MD simulations and compared with target data from experiment. It should be noted that computed data could equally come from Monte Carlo simulations.

While the dimensionality for the present application is relatively small (three parameters and five experimental observables), so that manual fitting is cumbersome but still possible, the approach pursued here translates naturally to problems with higher dimensionality and with a wider range of observables where manual fitting is no longer feasible. The issues encountered in the example are typical of force field fitting problems using MD data, which is still in its infancy. Automated fitting is often not possible, as relationships between parameters and calculated observables are complex, unknown, or highly nonlinear, and parameter values quickly leave physical bounds if all parameters and data are fitted simultaneously. Random fluctuations, which naturally occur if observables are calculated from averages over multiple simulations, make data inherently noisy, with computational cost often prohibiting more extensive averaging and sampling of available phase space. Supervision is then mandatory to ensure that steps in parameter space are not made on the basis of inferred/erroneous relationships between parameters and noise in the calculated data. The use of model systems to run simulations reduces the computational cost of each iteration, with the main requirement being that parameters fitted in the model are sufficiently transferable to the full system of interest for which explicit simulations are needed for validation.

In the present work, experimental spectroscopic and vibrational relaxation data for the excited ligand in a model heme system and in the solvated protein for bound MbCO are used as a reference. A set of Morse bonded parameters and bending mode force constants was found to give good agreement between calculated and experimental vibrational frequencies and relaxation times simultaneously. Good transferability of parameters between a heme model and MbCO was demonstrated. Relaxation from the second excited state was included as an additional observable and was successfully captured by the model. In contrast, adopting the more conventional route of fitting harmonic frequencies from the Hessian matrix, even based on Morse potentials for the vibrations of interest, was found to be not sufficient to correctly describe vibrational relaxation.

A wide range of other applications is imaginable. For example, classical MD simulations of explicitly solvated proteins can be used to calculate scalar NMR couplings.[39,40] To improve and extend the range of applicability of atomistic simulations, it will be interesting to subject particular parameters (e.g., atomic charges) to fitting of NMR data to averages over long MD simulations. Other dynamics-based data, such as diffusion coefficients, could also be used to improve atomistic force fields.

It is expected that with rapid advances in simulation techniques and with the ever-growing computational power the full range of static and dynamical variables can be included in future parameter fitting for atomistic force fields. It is hoped that the approach discussed here can contribute in this necessary step to further increase the range and applicability of atomistic simulations to realistic problems.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the Swiss National Science Foundation through grant number 200021-117810.

REFERENCES AND NOTES

(1) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. *J. Am. Chem. Soc.* **1979**, *101*, 2550–2560.

(2) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.

(3) Shaik, M.; Devereux, M.; Popelier, P. L. A. The importance of multipole moments when describing water and hydrated amino acid cluster geometry. *Mol. Phys.* **2008**, *106*, 1495–1510.

(4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(5) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. A new force-field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(6) Cieplak, P.; Caldwell, J. W.; Kollman, P. A. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.* **2001**, *22*, 1048–1057.

(7) Lennard-Jones, J. E. Cohesion. *Proc. Phys. Soc.* **1931**, *43*, 461–482.

(8) Levitt, M.; Lifson, S. The Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. *J. Mol. Biol.* **1969**, *46*, 269–279.

(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(10) Yin, D.; MacKerell, A. D., Jr. Combined Ab Initio/Empirical Approach for Optimization of Lennard Jones Parameters. *J. Comput. Chem.* **1997**, *19*, 334–348.

(11) Vaiana, A. C.; Cournia, Z.; Costescu, I. B.; Smith, J. C. AFMM: A molecular mechanics force field vibrational parametrization program. *Comput. Phys. Commun.* **2005**, *167*, 34–42.

(12) Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. Automated Parametrization of Biomolecular Force Fields from Quantum Mechanics/Molecular Mechanics (QM/MM) Simulations through Force Matching. *J. Chem. Theo. Comp.* **2007**, *3*, 628–639.

(13) Guvench, O.; MacKerell, A. D., Jr. Automated Conformational Energy Fitting for Force-Field Development. *J. Mol. Model.* **2008**, *14*, 667–679.

(14) *DirectForceField (DFF)*; Aeon Technology Inc.: San Diego, CA, 2006.

(15) Jorgensen, W. L. Quantum and statistical mechanics studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers - application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(16) Praprotnik, M.; Hocevar, S.; Hodoscek, M.; Penca, M.; Janezic, D. New All-Atom Force Field for Molecular Dynamics Simulation of an $AlPO_4$-34 Molecular Sieve. *J. Comput. Chem.* **2008**, *29*, 122–129.

(17) Law, M.; Hutson, J. I-NoLLS: a program for interactive nonlinear least-squares fitting of the parameters of physical models. *Comput. Phys. Commun.* **1997**, *102*, 252–268.

(18) MacKerell, A. D., Jr.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(19) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. X-ray Structure and Refinement of Carbonmonoxy (Fe II)-Myoglobin at 1.5 Angstrom Resolution. *J. Mol. Biol.* **1986**, *192*, 133–154.

(20) Brooks, C. L.; Karplus, M. Deformable stochastic boundaries in molecular-dynamics. *J. Chem. Phys.* **1983**, *79*, 6312–6325.

(21) Van Gunsteren, W. V.; Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.

(22) Devereux, M.; Meuwly, M. Anharmonic Coupling in Molecular Dynamics Simulations of Ligand Vibrational Relaxation in Bound Carbonmonoxy Myoglobin. *J. Phys. Chem. B* **2009**, 13061–13070.

(23) Owrutsky, J. C.; Li, M.; Locke, B.; Hochstrasser, R. M. Vibrational relaxation of the CO stretch vibration in hemoglobin-CO, myoglobin-CO, and protoheme-CO. *J. Phys. Chem.* **1995**, *99*, 4842–4846.

(24) Whitnell, R. M.; Wilson, K. R.; Hynes, J. T. Fast vibrational relaxation for a dipolar molecule in a polar solvent. *J. Phys. Chem.* **1990**, *94*, 8625–8628.

(25) Bu, L.; Straub, J. E. Vibrational energy frequency shifts and relaxation rates for a selected vibrational mode in cytochrome c. *Biophys. J.* **2003**, *85*, 1429–1439.

(26) McQuarrie, D. A. *Statistical Mechanics*; Harper's Chemistry Series: New York, NY, 1976.

(27) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K., 1987.

(28) Bevington, P. R. *Data reduction and error analysis for the physical sciences*; McGraw-Hill: London, U.K., 1969.

(29) Levenberg, K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Q. Appl. Math.* **1944**, *2*, 164–168.

(30) Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.* **1963**, *11*, 431–441.

(31) Lawson, C. L.; Hanson, R. J. *Solving least squares problems*; Prentice-Hall: Englewood Cliffs, NJ, 1974.

(32) Mizutani, Y.; Kitagawa, T. Ultrafast dynamics of Myoglobin probed by time-resolved Resonance Raman Spectroscopy. *Chem. Rec.* **2001**, *1*, 258–275.

(33) Ansari, A.; Berendzen, J.; Braunstein, D.; Cowen, B. R.; Frauenfelder, H.; Kyung Hong, M.; Iben, I. E. T.; Johnson, J. B.; Ormos, P.; Sauke, T. B.; Scholl, R.; Schulte, A.; Steinbach, P. J.; Vittitow, J. Rebinding and relaxation in the myoglobin pocket. *Biophys. Chem.* **1987**, *26*, 337–355.

(34) Leu, B. M.; Silvernail, N. J.; Zgierski, M. Z.; Wyllie, G. R. A.; Ellison, M. K.; Scheidt, W. R.; Zhao, J.; Sturhahn, W.; Alp, E. E.; Sage, J. T. Quantitative vibrational dynamics of Iron in Carbonyl porphyrins. *Biophys. J.* **2007**, *92*, 3764–3783.

(35) Hamm, P.; Lim, M.; Hochstrasser, R. M. Vibrational energy relaxation of the cyanide ion in water. *J. Chem. Phys.* **1997**, *107*, 10523–10531.

(36) Hamm, P.; Lim, M.; Hochstrasser, R. M. Non-Markovian Dynamics of the Vibrations of Ions in Water from Femtosecond Infrared Three-Pulse Photon Echoes. *Phys. Rev. Lett.* **1998**, *81*, 5326–5329.

(37) Schmitz, M.; Tavan, P. Vibrational spectra from atomic fluctuations in dynamics simulations. II. Solvent-induced frequency fluctuations at femtosecond time resolution. *J. Chem. Phys.* **2004**, *121*, 12247–12258.

(38) Stock, G. Classical Simulation of Quantum Energy Flow in Biomolecules. *Phys. Rev. Lett.* **2009**, *102*, 118301–118304.

(39) Schmid, F. F. F.; Meuwly, M. Direct Comparison of Experimental and Calculated NMR- Scalar Coupling Constants for Force Field Validation and Adaptation. *J. Chem. Theo. Comp.* **2008**, *4*, 1949–1958.

(40) Barfield, M. Structural Dependencies of Interresidue Scalar Coupling $^{h3}J_{NC}$ and Donor ^{1}H Chemical Shifts in the Hydrogen Bonding Regions of Proteins. *J. Am. Chem. Soc.* **2002**, *124*, 4158–4168.

CI9004404