Simulated and NMR-Derived Backbone Dynamics of a Protein with Significant Flexibility: A Comparison of Spectral Densities for the βARK1 PH Domain

Stefania Pfeiffer,[†] David Fushman,[‡] and David Cowburn*

Contribution from The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399

Received August 21, 2000. Revised Manuscript Received January 2, 2001

Abstract: A 7.6 ns molecular dynamics trajectory of the βARK1 PH domain in explicit water with appropriate ions was calculated at 300 K. Spectral densities at $\omega = 0$, ω_N, and $0.87\omega_H$ and the model-free parameters were evaluated from the experimental as well as the simulated data, taking the anisotropic overall motion of the protein into account. Experimental and simulated spectral densities are in reasonable general agreement for NH bond vectors, where the corresponding motions have converged within the simulation time. A sufficient sampling of the motions for NH bonds within flexible parts of the protein requires a longer simulation time. The simulated spectral densities $J(0)$ and $J(\omega_N)$ are, on average, 4.5% and 16% lower than the experimental data; the corresponding numbers for the core residues are about 6%; the high-frequency spectral densities $J(0.87\omega_H)$ are lower by, on average, 16% (21% for the core). The simulated order parameters, S^2, are also lower, although the overall disagreement between the simulation and experiment is less pronounced: 1% for all residues and 6% for the core. The observed systematic decrease of simulated spectral density and the order parameters compared to the experimental data can be partially attributed to the ultrafast librational motion of the NH bonds with respect to their peptide plane, which was analyzed in detail. This systematic difference is most pronounced for $J(0.87\omega_H)$, which appears to be most sensitive to the slow, subnanosecond time scale of internal motion, whereas $J(0)$ and $J(\omega_N)$ are dominated by the overall rotational tumbling of the protein. Similar discrepancies are observed between the experimentally measured ^{15}N relaxation parameters (R_1, R_2, NOE) and their values calculated from the simulated spectral densities. The analysis of spectral densities provides additional information regarding the comparison of the simulated and experimental data, not available from the model-free analysis.

Introduction

It is generally accepted that a more precise and accurate understanding of the dynamic properties of proteins would be of significance in explaining motional contributions to molecular recognition, catalysis, and protein stability and folding. Proteins experience a wide range of dynamic behavior in terms of amplitudes of the motion and the time scale of motion. These motions vary from small-amplitude bond vibrations to conformational changes with an amplitude comparable to the size of the protein itself. The time scale of motions ranges from femtoseconds up to milliseconds and slower. The fast dynamics, on the nano- to picosecond time scale, are partially accessible by experimental methods and are subject to simulation by molecular dynamics calculation. The presence of these motions contributes to temperature factors in X-ray structures and to nuclear magnetic resonance (NMR) spin relaxation: the latter is the most direct way to investigate fast dynamics of proteins in their natural liquid environment. A fuller understanding of how experiment and theoretical simulation agree and disagree is illustrative (i) of the applicability of molecular dynamics simulations on longer time scales and more complex systems, (ii) of systematic issues of the appropriate modeling of molecular motions, solvation, electromagnetic interactions, and magnetic relaxation, and (iii) of the role of atomic scale dynamics contributing to the thermodynamic stability of proteins.

Experimentally, the NMR relaxation is caused by fluctuating magnetic fields surrounding a particular spin. The fluctuation arises from the reorientation of the entire protein molecule in solution as well as from local motions. Both the overall and local motions modulate interactions by changing their orientation with respect to the external magnetic field. Recent advances in NMR spectroscopy and appropriate analysis make high-resolution studies of protein dynamics possible (reviewed in refs 1−5). ^{15}N spin relaxation in ^{15}N-labeled proteins has become common because, first, ^{15}N-labeled proteins can be obtained routinely and inexpensively, and second, ^{15}N relaxes about 76% at 600 MHz by dipolar interaction with the attached amide proton and only 24% by CSA. There are two frequently used approaches for the analysis of ^{15}N spin relaxation rates, $R_1$and R_2, and the steady-state $^{15}N\{^1H\}$NOE: the "model-free" approaches[6,7] and the spectral density mapping[8] approach. The model-free ap-

* To whom correspondence should be addressed. Phone: 212 327 8270. Fax: 212 327 7566. E-mail cowburn@rockefeller.edu.

[†] Present address: Aventis Pharma Deutschland Gmbh, Bldg. G838, D-65926, Frankfurt am Main, Germany.

[‡] Present address: Center of Biomolecular Structure & Organization, Department of Chemistry and Biochemistry, University of Maryland, 1115 Agriculture/Life Science Surge Building, College Park, MD 20742-3360.

(1) Wagner, G. *Curr. Opin. Struct. Biol.* **1993**, *3*, 748−754.
(2) Palmer, A. G., III. *Curr. Opin. Struct. Biol.* **1997**, *7*, 732−737.
(3) Fushman, D.; Cowburn, D. In *Structure, Motion, Interaction and Expression of Biological Macromolecules*; Sarma, R., Sarma, M., Eds.; Adenine Press: Albany, NY, 1998; pp 63−77.
(4) Fischer, M. W. F.; Majumdar, A.; Zuiderweg, E. R. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *33*, 207−272.
(5) Kay, L. *Nature Struct. Biol. (NMR Suppl.)* **1998**, *7*, 513−517.
(6) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

10.1021/ja0031117 CCC: $20.00
Published on Web 03/09/2001

proach assumes that the overall and local motions are separable, and that the corresponding correlation functions decay exponentially. This approach reduces the number of fitting parameters and characterizes protein dynamics with easily interpretable parameters, the overall hydrodynamic correlation time (τ_c) and local parameters, correlation times (τ_{loc}), and squared order parameters (S^2) which describe the time scale and amplitude, respectively, of the local reorientational motion of NH bond vectors. However, interpretation of these microdynamic parameters, τ_{loc} and S^2, as local motions in a protein is not straightforward, because the nature of the underlying structural fluctuations cannot be directly assessed from the relaxation data. The analysis of spectral density components is nominally more straightforward but does not separate overall molecular motion from local contributions directly, and similarly does not directly illustrate the underlying fluctuations.

Molecular dynamics (MD) simulations are capable of providing a detailed atomic-resolution picture of protein motions in a simulation, and therefore might prove to be indispensable for unraveling the complex nature of internal dynamics observed in a real NMR experiment.[9−12] In the simulation, the protein dynamics are obtained in atomic detail by solving Newton's equation of motion. Current limitations of this approach involve the short simulation times, imprecision of force fields, and incomplete representation of electrostatic and multibody effects. MD simulations are preferred which use all-atom force fields calibrated for proteins, an explicit solvent environment, an explicit salt content of the solvent, and no cutoff for electrostatic interactions.[13−18]

NMR relaxation probes protein dynamics in the same subnanosecond and nanosecond time scales as the current MD simulations and therefore provides a unique opportunity to compare experiment and simulation. Recent developments in computers and computational methods have made it possible to extend the MD-simulated coordinate trajectories for a protein in an aqueous environment beyond the nanosecond limit, to approach the NMR-relevant time scale. Some examples of long time scale MD simulations of fully hydrated proteins in their native state can be found in refs 19−21. A combined analysis of NMR- and MD-derived dynamics may provide criteria for the selection of major modes and models of motion that can be checked at longer time scales against other experimental data.

A number of comparisons between NMR-derived and simulated spin relaxation data of proteins in solution have been reported in the past.[9,22−35] In almost all cases, comparisons were restricted to experimental and simulated order parameters and local correlation times. Here we present a comparison of the simulated and experimentally measured protein dynamics in terms of spectral densities. This approach seems more straightforward than the former because the extraction of the microdynamic parameters from relaxation data[31,36,37] relies on assumptions and procedures of limited robustness. Spectral densities, on the other hand, can be directly derived from the experimentally measured relaxation parameters with minimal assumptions, which include the reduced spectral density method,[38,39] and might require knowledge of the strength of the dipolar interaction and CSA (see Methods section below). Calculation of the spectral densities or relaxation rates from the simulated trajectory is performed by using straightforward numerical methods, with simple statistical properties. Therefore, it is desirable to have a comparison between MD and NMR data on a level of directly experimentally measured parameters: relaxation rates and/or spectral densities. Several authors reported such a comparison for short trajectories.[30,40]

Obviously, the most direct way to compare experiment and simulation would be in terms of the experimentally measured relaxation parameters, R_1, R_2, and R_{NOE}. Since these parameters are related to the spectral densities via linear transformations (cf. eqs 1), such a comparison is virtually identical to that for spectral densities. The latter approach is chosen here because it is capable of providing individual spectral density components that could be directly related to various models of motion. This kind of analysis is more straightforward than a comparison of relaxation rates, because there are contributions to the latter from several spectral densities, often of various magnitudes.

In the present study, we provide a comparison of experimental spectral densities, derived from ^{15}N relaxation data, with those derived from a 7.6-ns MD trajectory of the βARK1 PH domain, a 119-residue protein. This length of a trajectory is similar to recent long time scale simulations,[19,20,32,41] as the system size (35 477 atoms) is significantly larger than in the previous MD

(7) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989−4936.
(8) Peng, J.; Wagner, G. *J. Magn. Reson.* **1992**, *94*, 82−100.
(9) Fushman, D.; Ohlenschläger, O.; Rüterjans, H. *J. Biomol. Struct. Dyn.* **1994**, *4*, 61−78.
(10) Bremi, T.; Bruschweiler, R. *J. Am. Chem. Soc.* **1997**, *119*, 6672−6673.
(11) Bremi, T.; Bruschweiler, R.; Ernst, R. R. *J. Am.Chem. Soc*. **1997**, *119*, 4272−4284.
(12) Lienin, S. F.; Bremi, T.; Brutscher, B.; Bruschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.
(13) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557−7561.
(14) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 55−59.
(15) Ibragimova, G. T.; Wade, R. C. *Biophys. J.* **1998**, *74*, 2906−2911.
(16) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001−5002.
(17) Fox, T.; Kollman, P. A. *Proteins: Struct., Funct. Genet.* **1996**, *25*, 315−334.
(18) Pfeiffer, S.; Fushman, D.; Cowburn, D. *Proteins* **1999**, *34*, 206−217.
(19) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.
(20) Radkiewicz, J. L.; Brooks, C. L. I. *J. Am. Chem. Soc.* **2000**, *122*, 225−231.
(21) Li, A.; Daggett, V. *Protein Eng.* **1995**, *8*, 117−128.
(22) Chandrasekhar, I.; Clore, G.; Szabo, A.; Gronenborn, A.; Brooks, B. *J. Mol. Biol.* **1992**, *226*, 239−250.
(23) Kordel, J.; Teleman, O. *J. Am. Chem. Soc*. **1992**, *114*, 4934−4936.
(24) Eriksson, M. A. L.; Berglund, H.; Härd, T.; Nilsson, L. *Proteins: Struct., Funct. Genet.* **1993**, *17*, 375−390.
(25) Balasubramanian, S.; Nirmala, R.; Beveridge, D. L.; Bolton, H. *J. Magn. Reson.* **1994**, *B104*, 240−249.
(26) Yamasaki, K.; Saito, M.; Oobatake, M.; Kanaya, S. *Biochemistry* **1995**, *34*, 6587−6601.
(27) Philippopoulos, M.; Lim, C. *J. Mol. Biol.* **1995**, *254*, 771−792.
(28) Smith, L. J.; Mark, A. E.; Dobson, C. M.; van Gunsteren, W. F. *Biochemistry* **1995**, *34*, 10918−10931.
(29) Smith, P. E.; van Schaik, R. C.; Szyperski, T.; Wüthrich, K.; van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *246*, 356−365.
(30) Ishima, R.; Yamasaki, K.; Saito, M.; Nagayama, K. *J. Biomol. NMR* **1995**, *6*, 217−220.
(31) Fushman, D.; Cahill, S.; Cowburn, D. *J. Mol. Biol.* **1997**, *266*, 173−194.
(32) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.
(33) Gryk., M. R.; Abseher, R.; Simon, B.; Nilges, M.; Oschkinat, H. *J. Mol. Biol.* **1998**, *280*, 879−896.
(34) Wong, K.-B.; Daggett, V. *Biochemistry* **1998**, *37*, 11182−11192.
(35) Buck, M.; Karplus, M. *J. Am. Chem. Soc.* **1999**, *121*, 9645−9658.
(36) Mandel, A. M.; Akke, M.; Palmer, A. G. I. *J. Mol. Biol.* **1995**, *246*, 144−163.
(37) Fushman, D.; Cowburn, D. In *Methods in Enzymology*; James, T., Schmitz, U., Doetsch, V., Eds.; 2000; in press.
(38) Ishima, R.; Nagayama, K. *Biochemistry* **1995**, *34*, 3162−3171.
(39) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. *J. Biomol. NMR* **1995**, *6*, 153−162.
(40) Brunne, R. M.; Berndt, K. D.; Güntert, P.; Wüthrich, K.; van Gunsteren, W. F. *Proteins* **1995**, *23*, 49−62.
(41) Kazmirski, S. L.; Daggett, V. *J. Mol. Biol.* **1998**, *284*, 793−806.

simulations.[42] The analyzed length of the MD trajectory is 6 ns. The protein is the extended pleckstrin homology (PH) domain of the β-adrenergic receptor kinase-1 (βARK1), also known as G-protein-coupled receptor kinase-2 (GRK2). The βARK1 PH domain is an especially suitable system for the study of protein dynamics because of its multiple structural features. The structure has been solved recently by high-resolution NMR and shows the fold and topology of PH domains[43] augmented by other features. It comprises seven β-strands forming a β-sandwich flanked on one side by the C-terminal α-helix. Compared to other reported PH domain structures, the C-terminus of the βARK1 PH domain construct is extended by 19 residues. These additional residues are essential for the affinity of the βARK1 PH domain binding to the $G_{\beta\gamma}$ subunit of the heterotrimeric G-protein family.[44] This C-terminal part appears to be very flexible, and the extended C-terminal α-helix behaves as a molten helix.

Materials and Methods

(A) Protein. The βARK1 PH domain construct used in this study is the same as that used in the NMR structural and dynamic studies.[43] Residue numbering is offset by −551 compared to that of the natural sequence of βARK1;[45] the βARK1 PH domain described in this paper (G1 through S119) corresponds to residues G556 through S670 in our previous study.[43]

(B) Relaxation Data and Analysis. Protein preparation, experimental conditions, and the sets of ^{15}N relaxation experiments performed (R_1, R_2, and steady-state $^{15}N\{^1H\}$ NOE) are from ref 43. The analysis of the ^{15}N relaxation data was performed in two ways: using spectral density mapping and using "model-free" analysis.

(1) Derivation of Spectral Densities from Experimental Data. The spectral densities, $J(0)$, $J(\omega_N)$, and $J(\omega_H)$ were determined directly from the relaxation data, R_1, R_2, and R_{NOE}, by using the standard expressions[46] and the reduced spectral density approach.[38,39] The following expressions were used:

$$J_{NMR}(0.87\omega_H) \equiv d^2\tilde{J}(\omega_H) = \frac{1}{5}\left|\frac{\gamma_N}{\gamma_H}\right|(1 - R_{NOE})R_1 \qquad (1a)$$

$$J_{NMR}(\omega_N) \equiv d^2\tilde{J}(\omega_N) = \frac{(R_1 - 7J(0.921\omega_H))}{3(1 + c^2/d^2)} \qquad (1b)$$

$$J_{NMR}(0) \equiv d^2\tilde{J}(0) = \frac{(2R_2 - R_1 - 6J(\omega_H))}{4(1 + c^2/d^2)} \qquad (1c)$$

where $d = (\mu_0/4\pi)(\gamma_N\gamma_H h)/(4\pi r_{NH}^3)$, $c = -|\omega_N|(\sigma_\parallel - \sigma_\perp)/3$, r_{NH} is the internuclear $^{15}N-^1H$ distance, $(\sigma_\parallel - \sigma_\perp)$ is the anisotropy of the ^{15}N chemical shift tensor (CSA), γ_N, γ_H, ω_N, and ω_H are the gyromagnetic ratios and resonance frequencies of the nuclei, and h is Planck's constant. All calculations assume $\sigma_\parallel - \sigma_\perp = -160$ ppm and $r_{NH} = 1.02$ Å. The spectrometer frequency was 600 MHz. Those residues subject to conformational exchange in a millisecond to microsecond time scale were excluded from the derivation of $J(0)$, eq 1a, because the measured transverse relaxation rates for these residues are affected by this phenomenon. The spectral densities defined here differ from the conventional spectral densities, $\tilde{J}(\omega)$, by the factor d^2. This derivation of $J(\omega)$'s does not directly require knowledge of the strength of the dipolar interaction (NH bond length), except for the c^2/d^2 ratio in eqs 1b,c; the latter is small, about 0.3 at the frequency used, and therefore these spectral densities, $J(0)$ and $J(\omega_N)$, are rather tolerant to small errors in d^2 or c^2 values. According to the reduced spectral density approach,[38,39] the following two alternate methods were used in order to relate the high-frequency components of the spectral density in eqs 1: (1) $J(\omega_H) \approx J(0.87\omega_H) \approx J(0.921\omega_H)$ and (2) $J(\omega) \propto 1/\omega^2$ for $\omega \approx \omega_H$. These assumptions, called methods 1 and 2 in ref 39, yield lower and upper bounds for $J(0)$ and $J(\omega_N)$ but have no effect on the values of $J(0.87\omega_H)$ directly determined from the experimental data according to eq 1a. This derivation of $J(0.87\omega_H)$ does not involve any assumption about the values of d^2 or c^2.

(2) Characteristics of the Overall Motion. The relaxation data for the βARK1 PH domain were analyzed by using the anisotropic overall rotational diffusion model. Axial symmetry was found to be a good approximation for the overall rotational diffusion, as characterized by the principal components, $D_\parallel$ and $D_\perp$. Orientation of the principal axes of the diffusion tensor with respect to the protein coordinate frame is given by the set of Euler angles: Φ, Θ, and Ψ. Ψ is treated as zero for the assumed axial symmetry. The remaining two angles were determined, together with $D_\parallel$ and $D_\perp$, from the analysis of the R_1/R_2 ratio for those residues belonging to the well-defined protein core. The procedure is described in ref 47.

The first snapshot of the analyzed MD trajectory ($t = 1.6$ ns) was taken as the structural model. Of the 55 core residues reliably observed in the relaxation experiments, residues K16 and I73 were excluded from this analysis because of the contribution to relaxation from conformational exchange, as inferred from the R_2/R_1 values for these residues being more than two standard deviations above the mean.

(3) Model-Free Analysis. The model-free parameters characterizing local protein dynamics were derived from the relaxation data for each amide group by using standard protocols.[48] Those residues belonging to the well-defined protein core were treated assuming anisotropic overall rotational diffusion with the parameters described above. Residues in the flexible loops and in the termini, where the backbone structure (hence the NH bond orientation) is ill-defined, were treated by using the isotropic rotational diffusion model.

These analyses were performed by using locally written programs DYNAMICS and R2R1.

(C) MD Simulations. The MD simulations were performed with AMBER version 5.0[49,50] by using the AMBER all-atom force field with the parameters from Cornell et al.[51,52] The TIP3 water model[53] was used. The starting structure of the βARK1 PH domain was immersed into a rectangular water box containing charge-balancing counterions ($6Cl^-$) and explicit salt ions ($6Na^+$, $6Cl^-$) corresponding to a salt concentration of 27 mM NaCl. These salt ions reproduced the number of ions per protein molecule used in the NMR samples of βARK1 PH domain. The system of 35 477 atoms was energetically equilibrated and then simulated for 7.6 ns at 300 K according to the details given elsewhere.[18] Coordinates were stored every 0.1 ps, resulting in 76 000 snapshots. The trajectory was generated on an Origin2000 system (SGI) and required an elapsed time of 8 months (CPU time, ~5 months) using, on average, eight R10000 processors. A second protein MD simulation was performed over a time of 20 ps, storing the coordinates after every integration step (of 2 fs), the starting structure of which was the last snapshot of the 7.6 ns MD trajectory.

(42) Daggett, V. *Curr. Opin. Struct. Biol.* **2000**, *10*, 160−164.
(43) Fushman, D.; Najmabadi-Haske, T.; Cahill, S.; Zheng, J.; LeVine, H., III; Cowburn, D. *J. Biol. Chem.* **1998**, *273*, 2835−2843.
(44) Koch, W. J.; Inglese, J.; Stone, W. C.; Lefkowitz, R. J. *J. Biol. Chem.* **1993**, *268*, 8256−8260.
(45) Benovic, J. L.; Stone, W. C.; Huebner, K.; Croce, C.; Caron, M. G.; Lefkowitz, R. J. *FEBS Lett.* **1991**, *283*, 122−126.
(46) Abragam, A. *The Principles of Nuclear Magnetism*; Clarendon Press: Oxford, 1961.
(47) Fushman, D.; Xu, R.; Cowburn, D. *Biochemistry* **1999**, *38*, 10225−10230.
(48) Fushman, D.; Cahill, S.; Cowburn, D. *J. Mol. Biol.* **1997**, *266*, 173−194.
(49) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(50) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5*; University of California: San Francisco, 1997.
(51) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(52) Kollman, P. A.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Phorille, A. In *Computer simulations of biomolecular systems*; Wilkinson, A., Weiner, P., van Gunsteren, W. F., Eds.; 1997; Vol. 3, pp 83−96.
(53) Jorgensen, W. L.; Chandreskhar, J. D.; Madura, J. D.; Imprey, W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

As a model for the interactions between bond vibrations and bending motion of NH bonds and the influence of the integration step size, a short peptide of six alanyl residues was simulated in water without salt for 500 ps, using the same equilibration procedure and simulation parameters. Three different model MD simulations were carried out: (1) with an integration step of 2 fs using the SHAKE algorithm[54] to keep bond length constant; (2) with an integration step of 1 fs using SHAKE; and (3) with an integration step of 1 fs without SHAKE.

(D) Simulated Correlation Functions. To remove the overall motion, snapshots were superimposed onto the starting structure of the production run ($t = 1.6$ ns) by using the backbone atoms N, C^{α}, and C' of the secondary structure elements as determined from the solution structure of βARK1 PH domain. The autocorrelation function $C_{loc}(t)$ describing the internal motion of the NH bond vectors is defined as

$$C_{loc}(t) = \langle P_2(\mu(t')\mu(t'+t))\rangle = \frac{4\pi}{5}\sum_{|M|\le 2}\langle Y_{2M}^*(\Omega_{t'+t})Y_{2M}(\Omega_{t'})\rangle \quad (2)$$

where $\mu(t)$ is the orientation of the interatomic vector at time t as measured in the molecular frame, and P_2 is the second-rank Legendre polynomial.

(E) Generalized Order Parameters. Assuming that the simulation time has been sufficiently extended to reach the long-term limit of the correlation function, $C_{loc}(t = T_{MD}) = C_{loc}(\infty)$, the values of order parameters can be calculated as the following time (equilibrium) average:

$$S^2 = C_{loc}(\infty) = \frac{4\pi}{5}\sum_{|M|\le 2}|\langle Y_{2M}(\theta,\phi)\rangle|^2 = \frac{4\pi}{5}\sum_{|M|\le 2}\left|\frac{1}{T_{MD}}\int Y_{2M}(\theta_t,\phi_t)\,dt\right|^2 \quad (3)$$

where T_{MD} is the total length of the simulation. Equation 3 assumes a constant length of the NH bond vector, as provided by the SHAKE algorithm. When taking the bond length explicitly into account, eq 3 has to be modified:

$$S^2 = \frac{4\pi}{5}\langle {r_{NH}}^{-3}\rangle^{-2}\sum_{|M|\le 2}\left|\left\langle\frac{Y_{2M}(\theta,\phi)}{{r_{NH}}^3}\right\rangle\right|^2 = \frac{4\pi}{5}\langle {r_{NH}}^{-6}\rangle^{-1}\sum_{|M|\le 2}\left|\frac{1}{T_{MD}}\int\frac{Y_{2M}(\theta_t,\phi_t)}{{r_{NH}}^3}\,dt\right|^2 \quad (3a)$$

Uncertainties of the order parameters S^2 were determined by using the jackknife procedure with blocks of 500 ps length, to allow a block length greater than the expected correlation times.[55] Order parameters for motions on different time scales of the simulated motion were obtained by least-squares fits of the correlation functions as described below.

In general, order parameters can also be obtained by random permutation of the time-ordered values of a quantity (i.e., scrambling) before calculation of the correlation function. A resulting correlation function starts at 1 but drops immediately to its mean value, which corresponds to the order parameter. Values of order parameters for the NH bond vector orientations obtained in this way are virtually the same as those derived from eq 3.

(F) Calculation of Spectral Densities from MD Simulation. The approach described here is directed to making the most effective use of the simulation, robustly deriving values from it, and comparing it to experimental data with minimal modeling assumptions.

The spectral densities are derived here as follows: First, the autocorrelation function of the NH bond motion, calculated by using eq 2, is parameterized (see eq 4 below) to allow subsequent steps in the calculation to be performed analytically. Second, this correlation function describing local motion is multiplied by an analytical expression for the correlation function of the overall motion, and the product is then analytically Fourier transformed to provide values of spectral densities, $J(\omega)$, at relevant frequencies (eq 5a, below). This approach to calculation of spectral densities using parameterization of the correlation function has two advantages over a direct Fourier transformation of the "raw" correlation function derived from a MD trajectory. First, the fitting procedure is less sensitive to noise present in the "raw" data. Second, and probably more importantly, this approach helps circumvent several problems related to the limited length of the trajectory. It accommodates the overall hydrodynamic rotation of the protein, which otherwise cannot be simulated accurately with the finite length of the trajectory and the confines of the solvent box. Moreover, the limited length of the trajectory limits the minimal available frequency for $J(\omega)$ derived by a direct Fourier transformation. With the maximal time range of 600 ps for a robust determination of $C(t)$, the point-to-point resolution in the frequency domain available from the MD data is limited to $\omega_{min} \simeq 1/(600\text{ ps}) \simeq 2\pi(265\text{ MHz})$, which is less than $0.87\omega_H$ but still significantly more than ω_N. Thus, one would expect that only $J(0.87\omega_H)$ could be reliably determined from a direct Fourier transformation of the correlation function. A trajectory at least 4.4 times as long as the current one is then required to achieve the frequency resolution ($\omega_{min} < \omega_N$) necessary for an accurate determination of $J(\omega_N)$ by a direct Fourier transformation of the "raw" correlation function. The parameterization of the correlation function introduced here helps solve this problem and determine $J(0)$ and $J(\omega_N)$ despite the limited size of the trajectory. This then permits calculation of $J(\omega)$ for virtually any ω, appropriate to the time scale of the simulation.

(1) Parameterization of the Correlation Function: Simulated Correlation Times and Local Order Parameters. The correlation function of local motion can be modeled by a weighed sum of exponential decays. The model function used here has the following form:

$$C_{loc}(t) = {S_{MD}}^2 + (1 - {S_u}^2)\exp\left(-\frac{t}{\tau_u}\right) + {S_u}^2(1 - {S_f}^2)\exp\left(-\frac{t}{\tau_f}\right) + {S_u}^2{S_f}^2(1 - {S_s}^2)\exp\left(-\frac{t}{\tau_s}\right) \quad (4)$$

The parameterization was performed by a least-squares fitting procedure based on a constrained minimization, where the relations ${S_{MD}}^2 = {S_u}^2{S_f}^2{S_s}^2$ and $\tau_u < \tau_f < \tau_s$ were used as constraints during the optimization. Here, ${S_u}^2$, ${S_f}^2$, and ${S_s}^2$ are the squared order parameters and τ_u, τ_f, and τ_s are local correlation times corresponding to an ultrafast ("u"), fast ("f"), and slow ("s") time scale of the simulated motion. The three-exponential form for $C_{loc}(t)$ was needed for an accurate fit of the MD-calculated correlation functions. It does not correspond to the classical model-free approach[6] or to the extended model-free approach.[7] The ultrafast time scale of motion was introduced in order to account for the typical initial drop of such correlation functions, and such ultrafast motions ($\tau_u < 0.1$ ps) are beyond the currently available range of sensitivity of NMR relaxation methods. Such processes obviously reduce any observed order parameter, but the time scale is experimentally inaccessible. The values of ${S_{MD}}^2$ were determined by using eq 3 as described above. Uncertainties for $C_{loc}(t)$ were derived by using the standard deviations of the scrambled correlation function except for the time-zero point, which has a value of 1 and no associated uncertainty, by definition. The standard deviations of the scrambled correlation function calculated over the first and the last few picoseconds of the total time interval were used to construct a linearly increasing error estimate for the $C_{loc}(t)$. The error increases by 3 orders of magnitude from $C_{loc}(0)$ to $C_{loc}(T_{MD})$. The time interval of 600 ps used for the parameterization of the correlation function is much shorter than T_{MD}. This error definition does not require a predefined knowledge of the internal correlation times. This somewhat empirically chosen error profile will overestimate, rather than underestimate, the statistical uncertainty from the calculation of $C_{loc}(t)$ because the influences of rare transitions or nonstochastic processes on the profile of $C_{loc}(t)$ are not contained in the error estimates derived from the scrambled correlation function. These estimated uncertainties were used in a

(54) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(55) Quennouille, M. H. *Biometrica* **1956**, *43*, 353−360.

standard error propagation method[56] in order to derive uncertainties of the fitted parameters. The inclusion of more time data points into the fit improves the precision of the correlation time for the slow time scale but decreases the precision of the parameters describing the fast processes. Therefore, the fit was performed in two steps. In the first step, we used only the first 10−100 ps (100−1000 data points) of $C_{loc}(t)$ in order to derive S_u^2, S_f^2, τ_u, and τ_f with the highest possible precision. The obtained parameters together with the uncertainties were used in the second step as additional constraints. In this step, 600 ps of $C_{loc}(t)$ were included in the optimization. This corresponds to 6000 data points for $C_{loc}(t)$. This time range for the correlation function was selected as one-tenth of the total analyzed trajectory, to provide at least 10-fold sampling for the slowest processes depicted by calculated $C_{loc}(t)$.

Uncertainties of the finally fitted correlation times and order parameters were assessed by a Monte Carlo procedure using the inverse covariance matrix approach and χ^2 boundaries.[56] Five hundred statistical events with parameters within the χ^2 boundaries were analyzed for error estimation.

(2) Simulated Spectral Densities. The relevant spectral densities, $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ were determined by using the fitted correlation times (τ_u, τ_f, τ_s) and order parameters (S_u^2, S_f^2, S_s^2) for the local motion. The effect of the overall motion of the protein was taken into account by assuming axially symmetric or isotropic rotational diffusion; the corresponding correlation functions can be found, e.g., in ref 57. In the case of an axially symmetric overall tumbling, this resulted in the following functional form for the spectral density function:

$$J_{MD}(\omega) = 0.1d^2[(3\cos^2\theta - 1)^2 j(\omega,\tau_1) + 3\sin^2\theta\, j(\omega,\tau_2) + 3\sin^4\theta\, j(\omega,\tau_3)] \quad (5)$$

Here, $\tau_1^{-1} = 6D_\perp$, $\tau_2^{-1} = 5D_\perp + D_\parallel$, $\tau_3^{-1} = 2D_\perp + 4D_\parallel$, θ is the angle between a given NH vector and the unique principal axis of the diffusion tensor, and

$$j(\omega,\tau_k) = S_u^2 S_f^2 S_s^2 \frac{\tau_k}{1+(\omega\tau_k)^2} + (1-S_u^2)\frac{\tau'_u}{1+(\omega\tau'_u)^2} + S_u^2(1-S_f^2)\frac{\tau'_f}{1+(\omega\tau'_f)^2} + S_u^2 S_f^2(1-S_s^2)\frac{\tau'_s}{1+(\omega\tau'_s)^2} \quad (5a)$$

with $\tau'^{-1}_{u,f,s} \equiv \tau^{-1}_{u,f,s} + \tau_k^{-1}$ ($k = 1$, 2, or 3). The principal values and the orientation of the unique principal axis of the overall rotational diffusion tensor were derived from the analysis of the ^{15}N spin relaxation data, as described above. In the case of isotropic rotational diffusion, $D_\parallel = D_\perp = D$, $\tau_1 = \tau_2 = \tau_3 = \tau_c = 1/(6D)$, eq 5 reduces to

$$J_{MD}(\omega) = (2/5)d^2\, j(\omega,\tau_c)$$

Uncertainties in the spectral densities were assessed by using 500 synthetic data sets obtained from eq 5, using randomly distributed values for S_u^2, S_f^2, S_s^2, τ_u, τ_f, and τ_s within their standard deviations as determined from the Monte Carlo simulations.

(G) Statistical Comparison of the Experiment and Simulation. The agreement between the experimental ("exp") and simulated ("sim") data sets for a particular parameter A ($A = S^2$, $J(0)$, $J(\omega_N)$, or $J(0.87\omega_H)$) is analyzed here by addressing the following issues: (1) statistical significance of the observed differences $\Delta A_i = A_i^{exp} - A_i^{cal}$ on a per residue basis and (2) correlation between site-specific variations in the experimental and simulated data along the backbone.

The central tendency of the differences ΔA_i was characterized by the median and its 95% confidence interval, calculated by considering a distribution of N ranked values[56] (see also ref 58). In addition, the width of the distribution of the differences ΔA_i around the mean was characterized by the mean absolute deviation, $A\mathrm{Dev} = 1/N\sum_{i=1}^{N}|\Delta A_i - \overline{\Delta A}|$.[56] This (first moment) approach is less sensitive to the tails of the observed distribution than the approaches based on the higher moments.

The linear correlation between the experimental and simulated data sets was calculated. The statistical significance of the correlation is estimated as that of the observed slope in the correlation, by testing the null hypothesis that there is no correlation (i.e., zero slope). Both the experimental and simulated data sets are subject to uncertainties in data determination, so this approach is expected to provide a more robust estimate of the significance of the correlation than the conventional one based on Pearson's correlation coefficient. Practically, the statistics are computed as the probability of a nonzero, characterized by the t-statistics determined as t =slope/error(slope), and computed by using standard techniques designed (e.g., ref 56).

Further quantification of the agreement on a per residue basis was performed by using z-score values, $z_i = (A_i^{exp} - A_i^{cal})/\sigma_i$, where $\sigma_i = [(\sigma_i^{exp})^2 + (\sigma_i^{sim})^2]^{1/2}$ is the standard deviation due to both experimental and simulation uncertainties[59] in A_i. Here we assume that the observed values of A_i^{exp} and A_i^{sim} are drawn from normal distribution functions with the mean values corresponding to the true values, A_{io}^{exp} and A_{io}^{sim}, of these parameters for a particular residue i, and the standard deviations determined by the uncertainties σ_i^{exp} and σ_i^{sim}. This approach tests the null hypothesis that for residue i, the true experimental and simulated values of A_i are indistinguishable versus the alternative hypothesis that they are different. These distributions, assumed normal, could have different standard deviations due to differences in the methods of determination (experiment vs simulation). Moreover, the parameters of these distribution functions (mean and standard deviation) could vary from residue to residue; therefore, the comparison is meaningful on a per residue basis. According to the null hypothesis, the z-score value z_i for each residue is then normally distributed with zero mean and unit standard deviation. The hypothesis can then be tested for each residue by using standard methods by computing the probability $P(z_i)$ that a value of z_i could occur by chance: $P(z_i) = P(z \geq |z_i|) + P(z \leq -|z_i|) = 2\,\mathrm{erfc}(|z_i|/2^{1/2})$. In addition, this null hypothesis could also be tested on the overall level (for all residues simultaneously), since the overall z-score, $Z = \sum_{i=1}^{N} z_i/N^{1/2}$, should then be normally distributed with zero mean and unit standard deviation. For this latter test, to reduce the effect of outliers, we also computed $P(Z)$ by using a 10% trimmed data set.

Results and Discussion

(A) Quality of the Simulated Trajectory. (1) Equilibration of the MD Trajectory. The energetic and thermodynamic equilibrium of the MD trajectory was considered by monitoring the energy terms of the force field as well as the root-mean-square deviation (rmsd) values and displacements of coordinates with respect to the starting structure at $t = 0$. The potential and kinetic energy terms remained essentially constant. However, an inspection of single energy terms, namely the electrostatic and van der Waals energy between the protein and ions, as well as between the protein and the solvent, showed that a thermodynamic equilibrium had not yet been reached. The equilibration is retarded by the presence of numerous charged side chains of the protein and their interaction with the ions in the solvent. Details of the equilibration and the behavior of the ions during the first 4 ns of the simulation are described in detail elsewhere.[18] The average backbone structure of the MD trajectory, calculated over the last 6 ns of the simulation time, differs by up to 3 Å from the starting structure in terms of displacements. The thermodynamic equilibration undergone during the first 1.5 ns is also visible from the atomic rmsd values and displacements with respect to the starting structure in Figure 1. The rmsd values for the core backbone atoms rapidly plateau

(56) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: New York, 1992.
(57) Woessner, D. *J. Chem. Phys.* **1962**, *37*, 647−654.
(58) Philippopoulos, M.; Mandel, A. M.; Palmer, A. G., III; Lim, C. *Proteins* **1997**, *28*, 481−493.

(59) The estimate of the mutual standard deviation σ_i assumes that the experimental and simulated values of A_i are independent. The presence of any correlation between them will affect the actual σ_i values and thus the z-scores.

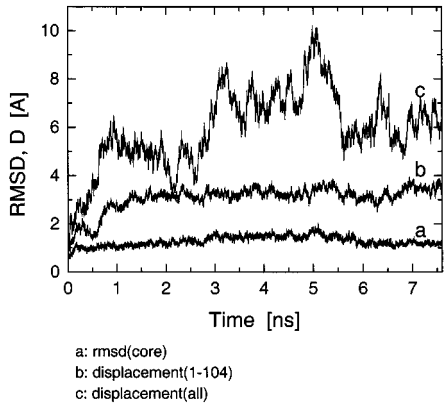

Figure 1. Backbone RMSD values (a) with respect to the starting structure and displacements, *D*, (b, c) as a function of simulation time. The RMSD values were obtained for the core residues only, while the displacements represent data calculated (b) for residues 1−104, excluding the flexible C-terminus, and (c) for the entire backbone (residues 1−119). The structures were superimposed by using the backbone atoms N, C^{α}, and C′ of the residues belonging to the protein core, as defined for the βARK1 PH domain in ref 43.

at approximately 1 Å. Beyond 1.5 ns, the rmsd values increase to 1.5 Å on average and decrease to 1.2 Å after 6 ns simulation time. The displacements for the fragment including the loops but excluding residues 105−119 in the flexible C-terminus beyond the C-terminal α-helix reach a stable plateau around 3.2 Å after approximately 1.5 ns. The region of relatively stable rmsd from 1.6 to 7.6 ns was used for analysis. The displacements including all residues of the protein increase to 5 Å during the first nanosecond but show strong oscillations over the entire simulation. While these characteristics are different from previous simulations (e.g., of compact globular proteins), they are not unexpected for a protein with an experimentally determined structure which is partially disordered. The maximum displacement, including all residues is nearly 10 Å. The flexibility of the C- and N-termini of the βARK1 PH domain renders the deconvolution of the overall reorientation of the molecule using *all* residues inaccurate. Therefore, only the protein core was used for the alignment.

Figure 1 shows the relatively slow rearrangements of the loops and the termini of the βARK1 PH domain during the thermodynamic equilibration in a very flat energy landscape. The starting structure for the simulation was derived from NMR data by using a distance geometry algorithm that included neither an explicit force field nor charges. Thus, adjustments of a few angstroms are expected, especially when a protein sequence comprises numerous charged residues in flexible segments. In addition, the counterions placed in the solvent near the charged residues of the βARK1 PH domain also require a certain amount of time to equilibrate as an ionic solution.[18] The slight increase and subsequent decrease of rmsd values for the backbone of the protein core between 1.5 and 6 ns visible in Figure 1 can be attributed to a rearrangement of the β-strands $\beta4-\beta7$ with respect to each other. This global conformational change was revealed by evaluating rmsd values of different subsets of residues of the protein core. The time-dependent rmsd values calculated for $\beta1-\beta3$ and the α-helix, as well as those calculated for $\beta1-\beta3$ alone, did not show the increase between 1.5 and 6 ns. However, time-dependent rmsd values calculated for $\beta4-\beta7$ including or excluding the α-helix, as well as rmsd values for $\beta1-\beta7$, did show a profile similar to that in Figure 1. The profile is caused exclusively by changes within $\beta4-\beta7$. The amplitude of the rearrangement is small (only 0.5 Å increase for the entire protein core) but correlated for a number of residues.

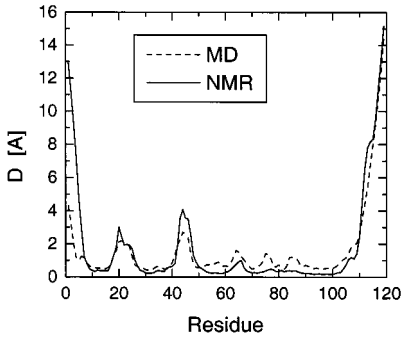

Figure 2. Displacements of the backbone N atoms with respect to the average structure (after superposition of the backbone N, C^{α}, and C′ atoms of the core residues) observed in the present MD simulation (solid line) and in the ensemble of the 20 lowest-target-function NMR structures (PDB access code 1bak) (dashed line). In the case of the NMR solution structure, the ensemble includes 20 structures, while the structure ensemble of the MD simulation comprised 1200 structures taken in equidistant time steps over the last 6 ns of the simulation time.

(2) Comparison with the Ensemble of NMR Structures. Figure 2 provides a comparison between the structural diversity within the ensemble of the NMR solution structures of the βARK1 PH domain and the simulated atomic fluctuations. Qualitatively, the structural diversity of the NMR structure set corresponds quite well to the flexibility of the backbone of the βARK1 PH domain that was simulated without any experimental constraints. Exceptions are the termini and the $\beta3/\beta4$ loop (residues in the vicinity of A45), where the MD simulation shows less flexibility than the NMR conformers suggested. On the other hand, the secondary structure elements of the protein appear slightly more rigid in the NMR structure than in the MD simulation. The corresponding displacements of the NMR structural ensemble in Figure 2 are on average 0.25 Å smaller than those for the simulated conformations.

(3) Effect of SHAKE on Simulated Order Parameters. The NH bond vectors in the 500 ps MD simulations of the alanine heptapeptide were transformed into the residue-fixed frame, which will be described later in detail. Subsequently, the order parameters S_{lib}^2 were determined by using eq 3. The analysis of 500 ps MD simulations with and without SHAKE revealed that the bending and torsion motional modes of amide NH bonds are influenced by the inhibition of the bond stretching using SHAKE. The order parameters S_{lib}^2 simulated without SHAKE are on average 0.005 smaller than values derived using SHAKE. This difference is significant, with uniform uncertainties of S_{lib}^2 of 0.002. Uncertainties in S_{lib}^2 were obtained by the jackknife procedure using a block size of 10 ps. When taking the NH bond length explicitly into consideration by using eq 3a for the calculation of S_{lib}^2, the difference is even more obvious, reaching 0.01. Without SHAKE, the average bond length increases by 0.5%, caused by the anharmonic stretching vibration. In this case, averaging over different powers of r_{NH} leads to the additional decrease of S_{lib}^2 calculated by using eq 3a compared to the values obtained by using eq 3. However, considering uncertainties in the NMR-derived order parameters, the effect of SHAKE on the simulated order parameters of NH bond vector orientations is negligible. As expected, there was no systematic difference in simulated order parameters S_{lib}^2 obtained with integration step sizes of 2 and 1 fs using SHAKE. The randomly distributed differences in S_{lib}^2 are on average $5 \times 10^{-4} \pm 0.004$. For comparison, the average S_{lib}^2 values obtained by using

SHAKE for the alanine heptapeptide and for the βARK1 PH domain (over 6 ns) are 0.923 ±0.002 and 0.926 ± 0.005, respectively. These results suggest that the calculated librational motions of the NH vector are intrinsic to the peptide plane itself, independent of structure and size of the polypeptide chain in the AMBER simulations.

(4) Statistics of Motional Events. For sufficiently long simulation time, the calculated values of order parameters S_{MD}^2 should not depend on the time window of the calculated MD trajectory. To test this, we calculated S_{MD}^2 values by dividing the trajectory into equal time windows of various lengths, ranging from 100 ps to 1 ns. The order parameters were calculated separately for all time windows and subsequently averaged over the windows of equal length. These order parameters, S_{win}^2, are depicted in Figure 6d (below). In general, the calculated values decrease with increasing time-window size, especially in the loops and at the flexible part C-terminal to the α-helix. For residues within secondary structure elements, the standard deviation of the different values of S_{win}^2 is basically within the uncertainties of the order parameters S_{MD}^2 determined over the entire MD trajectory, indicating that the trajectory length was long enough to provide sufficient sampling for these parts of the protein. This does not hold, however, for the less structured parts of the protein, in particular the loops β1/β2, β3/β4, β5/β6, and β7/α, as well as for the C-terminal segment of the βARK1 PH domain.

(B) Overall Characteristics of the Simulated Backbone Dynamics. We describe the overall characteristics of the simulated backbone dynamics of the βARK1 PH domain in terms of dihedral angle fluctuations and reorientation of the NH bonds resulting from the dihedral angle fluctuations.

(1) Dihedral Angle Fluctuations from MD Simulation. A large amount of positional fluctuation is attributed to changes in dihedral angles. Changes in the ω angles of the βARK1 PH domain are restricted to narrow intervals: the fluctuation amplitudes for individual residues range from 6.0° to 11.7°, with an average value of 8.8°. The peptide units undergoing the largest fluctuations in ω are M14, L37, R40, L50, I63, and K64. The mean values of the ω angle for individual peptide planes are characterized by the mean of 179.0° and the standard deviation of 4.6°.

The fluctuations in the φ and ψ dihedral angles are much larger. The amplitudes in terms of standard deviations are given in Figure 3. Fluctuation amplitudes are lower in elements of secondary structure (<30°) and increase in loops and termini (to 80°). In general, the ψ dihedral angles exhibit a greater variation than the φ angles within flexible regions of the protein. The correlation coefficients $C(\psi_{i-1},\varphi_i)$ and $C(\varphi_i,\psi_i)$ are depicted in Figure 3c and d. Most residues show a striking anticorrelation for fluctuations of the ψ and φ angle adjacent to a peptide plane (ψ_{i-1},φ_i). This has already been observed and discussed earlier for other proteins.[9,22,60−62] A few residues in this simulation of the βARK1 PH domain are exceptions to anticorrelation; all are located in flexible loops or at the C-terminal extension of the βARK1 PH domain and correspond to residues with large amplitudes of dihedral angle fluctuations. Their correlated fluctuations of ψ_{i-1} and φ_i with large amplitude facilitate the hinge-bending motion of loops. In contrast, the anticorrelation for smaller fluctuations of the ψ_{i-1} and φ_i angles within secondary structure elements prevents the protein from unfolding due to single dihedral angle transitions. This largely maintains

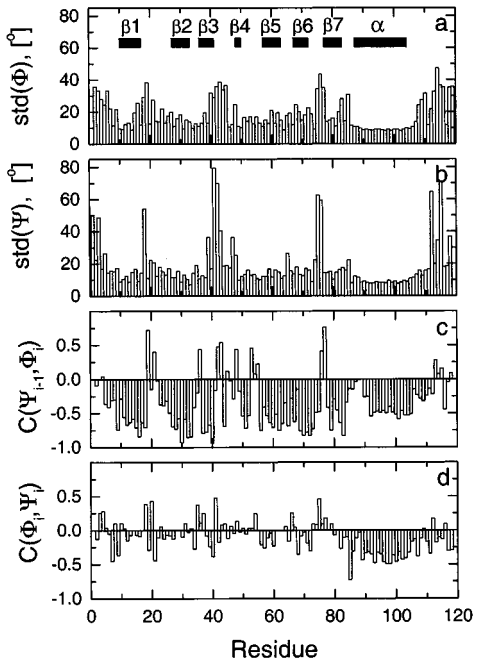

Figure 3. Amplitudes and correlation coefficients of the backbone dihedral angle fluctuations: (a,b) standard deviations of φ and ψ; correlation coefficient (c) for the adjacent residues, ψ_{i-1} and φ_i belonging to the same peptide plane, and (d) for the same residue, φ_i and ψ_i.

the characteristic hydrogen bonds. The correlation coefficient for the fluctuations of the ψ and φ angles of the same residue (φ_i,ψ_i) shows no trend except for the α-helix and the beginning of the C-terminal extension of the βARK1 PH domain. Here, the backbone dihedral angle fluctuations for a particular residue are also anticorrelated. The additional anticorrelation of the φ_i and ψ_i angles reflects the overall rigidity of an α-helical secondary structure compared to a β-sheet structure.

In this context, rare dihedral angle transitions are frequently discussed, but a generally accepted definition of what is rare does not exist. In addition, an observed transition (1) can be fast or continuing, (2) can be single, or there and back, (3) can have a small or large amplitude, and (4) can be correlated or anticorrelated with other transitions. These characteristics effect how a "rare" transition is reflected in simulated parameters, such as order parameters S^2 or correlation functions of NH bond vectors. Due to this lack of a clear definition of rare transitions, we do not report or discuss this issue for the βARK1 PH domain in this paper.

(2) Librational Motion of NH Bonds. It may be assumed that NH bond vectors undergo at least two different types of motion in a simulation: (i) the librational motion with respect to the peptide plane (in- and out-of-plane bending as well as bond vibration) associated with the force field parameters for the peptide plane geometry, and (ii) rocking motion of the relatively rigid peptide plane mainly due to dihedral angle fluctuations. The stretching of NH bonds was inhibited during the 7.6 ns MD trajectory using the SHAKE algorithm.

(a) Selection of a Residue-Fixed Frame. The bending motion of the NH bond vectors can be visualized by transforming the vector coordinates from the protein-fixed frame into a residue-fixed frame. The choice of a residue-fixed frame is not readily apparent, for the following reasons. In general, the *i*th peptide plane can be defined by the $C'_{i-1}N_i$ vector and a vector connecting any other pair of atoms in the peptide plane. It seems obvious to choose the second vector along one of the bonds

(60) McCammon, J. A.; Celin, B. R.; Karplus, M.; Wolynes, P. G. *Nature* **1976**, *262*, 325−326.
(61) Levitt, M. *J. Mol. Biol.* **1983**, *168*, 621−657.
(62) Wasserman, Z. R.; Salemme, F. R. *Biopolymers* **1990**, 1613−1631.

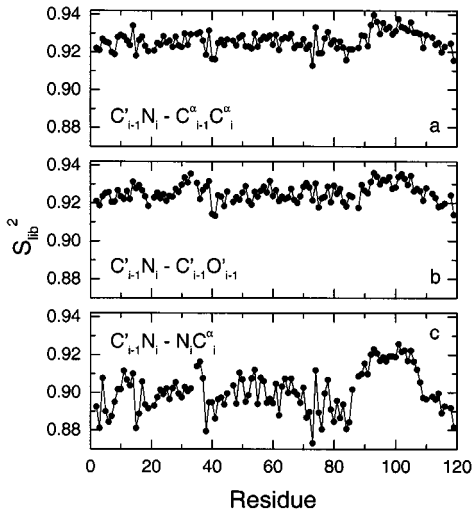

Figure 4. Comparison of various choices of a residue-fixed coordinate frame. The order parameters for librational motion shown are in the coordinate frames determined by the (a) $C'_{i-1}N_i$ and $C^{\alpha}_{i-1}C^{\alpha}_i$, (b) $C'_{i-1}N_i$ and $C'_{i-1}O_{i-1}$, and (c) $C'_{i-1}N_i$ and $N_iC^{\alpha}_i$ bond vectors, as described in the text.

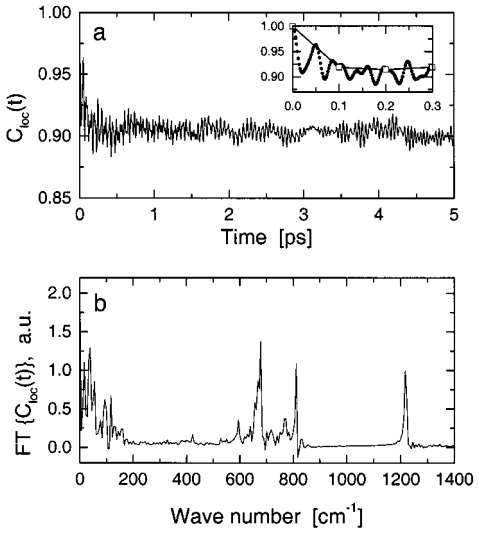

Figure 5. Typical initial decay of the correlation function (a) and its Fourier transform, the spectral density (b). The data shown here correspond to M10 and were obtained from a 20 ps interval of trajectory, with protein coordinates being saved every integration step (2 fs). The inset in (a) shows the first 300 fs of the time course of $C_{loc}(t)$ demonstrate that this initial drop is caused by coarse sampling of the data. Solid circles in the inset correspond to a storage step of 2 fs (20 ps interval), while the $C_{loc}(t)$ data shown in open squares, connected by a line, were obtained with a 0.1 ps storage step (6 ns interval). The apparent initial drop in $C_{loc}(t)$ from 1 at $t = 0$ to a value of ~0.92 at the next time point, $t = 0.1$ ps, depicted by open squares in the inset, is typically observed for the correlation functions derived from MD simulations.[9,22]

attached to the ends of the $C'_{i-1}-N_i$ bond, but these bonds undergo bending motions as well. We investigated three different residue-fixed frames defined by the following pairs of vectors: (1) $C'_{i-1}N_i$ and $C^{\alpha}_{i-1}C^{\alpha}_i$; (2) $C'_{i-1}N_i$ and $C'_{i-1}O_{i-1}$; and (3) $C'_{i-1}N_i$ and $N_iC^{\alpha}_i$. Figure 4 shows order parameters S_{lib}^2 of NH vector motions with respect to the three different residue-fixed frames calculated from the 6 ns MD simulation by using eq 3. The S_{lib}^2 values derived in frames 1 and 2 are less sequence dependent than those for frame 3. The observed sequence dependence is caused by the librational motion of the $N_iC^{\alpha}_i$ bond that is reduced within regular secondary structure, in particular in central parts of the α-helix. Furthermore, fluctuations of the ω angle contribute, to some extent, to the observed displacement of the NH bond within the chosen residue-fixed frames because these fluctuations are independent of the librational motion of the NH bonds.[35] The fluctuations in ω for the βARK1 PH domain are reduced within the α-helix (6.6° on average) by about 30% compared to all other residues. While the librational motion of the $N_iC^{\alpha}_i$ bond has a considerable effect in frame 3, the combined librational motion of the $C^{\alpha}_{i-1}C'_{i-1}$ and $N_iC^{\alpha}_i$ bonds in frame 1 seems to compensate this effect.

(b) Order Parameters for Librational Motion. The values of S_{lib}^2 are largest when frame 1 is used. The mean values and standard deviations for frames 1, 2, and 3 are 0.926 ± 0.005, 0.925 ± 0.005, and 0.90 ± 0.01, respectively. The S_{lib}^2 values derived in frame 1 were used for further interpretation of the data. The *z*-axis of this frame was chosen along the $C'_{i-1}N_i$ bond of residue *i*. The *x*-axis is lying in the peptide plane and pointing toward the oxygen. In addition, the *y*-axis is perpendicular to both, thus completing a right-handed coordinate frame. Order parameters S_{lib}^2 reflect only the amplitude of motion relative to the peptide plane to which a NH bond vector is attached. The uniform profile of S_{lib}^2 (Figure 4a) suggests that the bending motion of the NH bond vector is essentially sequence independent. Considering the observed diffusion-like character of the librational motion within a cone, the mean values of S_{lib}^2 correspond to a cone semiangle of about 13°. The average standard deviation of the in-plane angular motion of the N_iH_i bond vectors is $4.1 \pm 0.6°$, whereas the average standard deviation of the out-of-plane angular motion is $7.9 \pm 0.6°$. For comparison, the average standard deviations of the in-plane and out-of-plane angular motions for the $C'_{i-1}O_{i-1}$ bond vectors are $2.4 \pm 0.1°$ and $5.2 \pm 0.3°$, respectively. The latter values are also reflected in higher average librational order parameters S_{lib}^2 of 0.97 ± 0.003 for the CO bond vectors. Order parameters S_{lib}^2 and fluctuation amplitudes of the NH bond vectors calculated for the βARK1 PH domain are in good agreement with those reported for lysozyme using the CHARMM force field.[35]

(c) The Nature of the Initial Drop. Autocorrelation functions of NH bond vectors calculated from MD trajectories typically show an initial drop, i.e., a significant difference in $C(t)$ between $t = 0$ and the next time point. It was assumed that the initial drop is caused by librational motion (vibration and bending) of the bonds[22,63] and can be attributed to the finite storage step size.[9] In this case, processes faster than the storage step get lost. To elucidate the cause of this drop in detail, internal correlation functions $C_{loc}(t)$ were calculated from the 20 ps MD trajectory sampled every integration step (2 fs). The initial drop in $C_{loc}(t)$, observed for a storage step of 0.1 ps, completely disappeared when the finest possible storage step of 2 fs was used (Figure 5a), consistent with the expected nature of this phenomenon. In addition, these finely sampled correlation functions also reveal pronounced high-frequency oscillations on a subpicosecond time scale. The superposition of an exponential decay and an oscillation clearly indicates that the underlying motional process, namely, the in-plane and out-of-plane librations of the NH bond vector within this time scale, is not entirely stochastic. These strong oscillations preclude accurate characterization of the ultrafast time scale behavior of the correlation function by a direct fit to an exponential function, although a storage step of 2 fs provided a sufficient sampling of this time

(63) Bruschweiler, R.; Case, D. *Prog. NMR Spectrosc.* **1994**, *26*, 27−58.

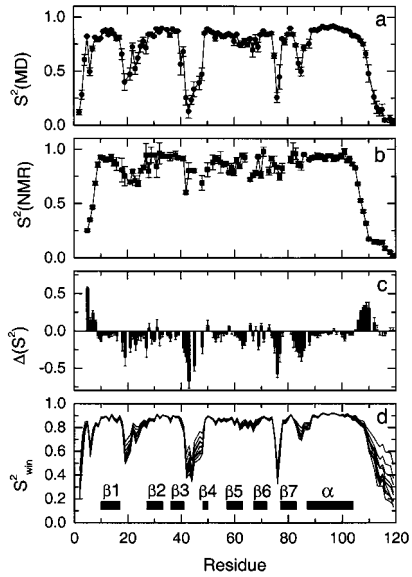

Figure 6. Comparison of the (a) simulated and (b) experimentally derived squared order parameters for the βARK1 PH domain, and (c) the difference between S_{MD}^2 and S_{NMR}^2. Panel (d) depicts time window dependence of the simulated order parameters. Shown in this panel are 10 sets of order parameters, S_{win}^2, calculated by using time intervals (windows) of 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 ps. For each time window, the whole trajectory was divided into the corresponding number of time intervals of equal length. The order parameters were calculated separately for each interval and subsequently averaged over all intervals of a given length.

scale of motion. However, the correlation time for the initial drop could safely be estimated as being shorter than 0.1 ps. For lysozyme, relaxation times of 0.5−2 ps were calculated.[35]

To characterize in detail the motional process leading to the initial drop, the Fourier transformation (spectral density in terms of wavenumbers) of the oscillating correlation function was examined with respect to the frequency components that show increased values of the spectral density. As an example, the correlation function and the spectral density for the NH bond vector of M10 are shown in Figure 5. In general, all correlation functions of NH bond vectors derived from the 20 ps MD trajectory show increased spectral densities at wavenumbers around 700 and 1200 cm^{-1} which correspond excellently to the torsional (γ(NH), motion perpendicular to the peptide plane) and bending (δ(NH), motion within the peptide plane) modes of an amide NH bond.[64] This reflects the use of these motional modes for the parameterization of the force field.[51] The average value of the order parameter S_{lib}^2 determined in the $\{C'_{i-1}N_i, C^{\alpha}_{i-1}C^{\alpha}_i\}$instant frame from the 20 ps MD trajectory was 0.929 ± 0.009, in agreement with that for the full trajectory. The out-of-plane motion of the NH and CO vectors of a particular peptide plane is slightly anticorrelated, with an average correlation coefficient of −0.32 ± 0.08. No correlation was observed between the amplitudes of librational motion of NH and CO vectors and secondary structure, in agreement with refs 9 and 35. These high-frequency oscillations are not present in the correlation functions for the backbone dihedral angles.

(C) Comparison of Experimental and Simulated Dynamics. (1) Experimental and Simulated Generalized Order Parameters. A comparison of the simulated and experimental squared order parameters, S_{MD}^2 and S_{NMR}^2, for NH bond vectors is shown in Figure 6. Although it is difficult to compare the estimates of precision between two disparate methods, for the simulated and experimental data for this case, large values of order parameters have on average higher precision in the MD simulation, whereas smaller values of order parameters have, on average, higher apparent precision when derived from ^{15}N spin relaxation experiments. This observation is consistent with the results of previous studies.[58] Values of S_{NMR}^2 for the βARK1 PH domain range from 0.02 to 0.98, with a mean value of 0.77 (the standard error of the mean is ±0.03). Values of S_{MD}^2 are distributed between 0.04 and 0.92, and the mean value is 0.71 ± 0.02. The experimental and simulated order parameters are greater than 0.8 for the secondary structure elements and drop in the loop regions connecting them. In the case of the βARK1 PH domain, the order parameters decrease rapidly toward the N- and C-termini, reaching values of almost zero, indicating high local flexibility.

With only a few exceptions, the MD simulated order parameters are lower than the values derived from NMR, especially in the loop regions (Figure 6). The simulated order parameters for the protein core are on average 6% lower than the NMR-derived values (Table 1). The average S_{MD}^2 within secondary structure elements, 0.83 ± 0.01, is lower than the corresponding average S_{NMR}^2, 0.89 ± 0.01. It is not readily apparent what combination of factors is responsible for the systematic discrepancy between the simulation and apparent S_{NMR}^2; these include the inaccuracy of the S^2 derivation from experimental data, the inaccuracy of calculation of parameters from simulation, and limitations of the force field and the representation of interatomic forces. This discrepancy could also be attributed to the contribution of the NH bond libration to the simulated order parameters. While the time scale of the librational motions of the NH vectors is accurate, because the force field perfectly reproduces the experimental IR frequencies,[64] it is possible that this parameterization does not represent the magnitude of this motion accurately. The amplitude of librational motion is determined by the kinetic energy, a function of the atomic mass and the velocity. While the mass of a hydrogen atom is accurately known, the average velocity is temperature dependent. A comparison of librational order parameters S_{lib}^2 (Figure 4a) and maximal values of S_{NMR}^2 (Figure 6b) suggests that the amplitude of librational motion is slightly overestimated in the MD simulation compared to the experimental results: a uniform scaling of the simulated order parameters, S_{MD}^2, by a factor between $1/S_{lib}^2$ and 1 significantly improves the comparison (Table 1). The SHAKE algorithm might be expected to affect the libration. Larger values of the simulated amplitude of librational motion were observed in MD simulation for the model six-alanyl peptide without using SHAKE (see above), so SHAKE is certainly not responsible for the apparent overestimate in the simulation.

As expected, the largest deviations between NMR-derived and simulated order parameters S_{MD}^2 occur in the flexible parts of the protein, namely the termini and the loops. Here, simulated order parameters are generally lower than the NMR-derived values. This has been observed for many other protein MD simulations.[9,22,23,28,32,34,58] The order parameters S_{win}^2 calculated over slices of the trajectory (Figure 6d) indicate that the low values of S_{MD}^2 in certain loop regions (residues N19−Q24, E42−S48, G76−K77, D84−D86) and at the termini (residues S2−M4, K6, D7, Q108−S119) belong to a time scale of internal motions which is probably around 500 ps or slower. The analyzed MD simulation of 6 ns length has not yet reached a full convergence for these motions, i.e., sampling of all possible conformations according to the Boltzmann distribution. The low

(64) Rey-Lafon, M.; Forel, M. T.; Garrigou-Lagrange, G. *Spectrochim. Acta* **1973**, *29A*, 471−486.

Table 1. Statistics of the Comparison of the Simulated and Experimental Spectral Densities, Order Parameters, and ^{15}N Relaxation Parameters at 600 MHz

	$A^{exp} - A^{sim}$ (%)[a]	median[b] (confidence interval)	ADev[c]	Z[d]	$P(Z)$[e]	slope[f]	t[g]	N[h]
$J(0)$ all[i,n]	4.5	0.22 (0.16, 0.32)	0.51	17.5	10^{-68}	0.85 (0.11)	7.7	82
$J(0)$ core[j,n]	6.0	0.22 (0.16, 0.31)	0.26	22	10^{-111}	0.68 (0.15)	4.6	44
$J(0)$ core, u-scale = 0.926 [j,k,n]	−1.6	−0.09 (−0.14, −0.001)	0.26	−0.003	1.00	0.63 (0.14)	4.6	44
$J(0)$ core, scale = S_{lib}^2 [j,l,n]	−1.3	−0.06 (−0.12, 0.03)	0.26	0.75	0.45	0.65 (0.14)	4.6	44
$J(\omega_N)$ all[i,n]	15.6	0.030 (0.026, 0.042)	0.042	54	0	0.70 (0.05)	13.8	97
$J(\omega_N)$ core[j,n]	6.2	0.022 (0.016, 0.027)	0.015	17.5	10^{-35}	0.60 (0.13)	4.8	52
$J(\omega_N)$ core, u-scale = 0.936 [j,k,n]	−0.9	−0.002 (−0.007, 0.002)	0.016	-9×10^{-4}	1.00	0.56 (0.12)	4.9	52
$J(\omega_N)$ core, scale = S_{lib}^2 [j,l,n]	−1.1	−0.005 (−0.009, −0.001)	0.015	−2.2	0.03	0.57 (0.11)	4.9	52
$J(0.87\omega_H)$ all[i]	16.0	0.003 (0.002, 0.004)	0.004	34	10^{-250}	0.72 (0.10)	7.4	97
$J(0.87\omega_H)$ core[j]	21.4	0.003 (0.002, 0.004)	0.002	23	10^{-116}	0.18 (0.11)	1.6	52
$J(0.87\omega_H)$ core, u-scale = 0.748[j,k]	−4.9	9.0×10^{-4} (−0.0002, 0.0016)	0.003	8×10^{-4}	1.00	0.14 (0.08)	1.7	52
$J(0.87\omega_H)$ core, scale = S_{lib}^2 [j,l]	15.3	0.003 (0.002, 0.003)	0.002	18	10^{-69}	0.17 (0.10)	1.7	52
$J(0.87\omega_H)$ core, S_u^2, S_f^2, S_s^2 scaled[j,m]	−2.5	0.001 (−0.0005, 0.002)	0.002	2.2	0.02	0.18 (0.11)	1.6	52
S^2 all[l]	1.0	0.048 (0.036, 0.063)	0.11	7.2	10^{-12}	0.94 (0.10)	9.2	97
S^2 core[j]	6.0	0.042 (0.035, 0.059)	0.05	13.9	10^{-43}	0.05 (0.08)	0.6	52
S^2 core, u-scale = 0.953[j,k]	1.4	-7×10^{-4} (−0.007, 0.018)	0.055	0.004	1.00	0.05 (0.08)	0.6	52
S^2 core, scale = S_{lib}^2[j,l]	−1.3	−0.022 (−0.030, −0.010)	0.054	−7.6	10^{-14}	0.04 (0.08)	0.49	52
R_1 all[i,o]	15.5	0.15 (0.13, 0.17)	0.16	23	10^{-116}	0.56 (0.06)	9.0	97
R_1 core[j,o]	7.4	0.11 (0.10, 0.13)	0.06	16	10^{-55}	0.72 (0.15)	4.9	52
R_2 all[i,o]	7.0	0.71 (0.51, 0.97)	1.36	12	10^{-31}	0.82 (0.10)	8.1	82
R_2 core[j,o]	6.2	0.64 (0.51, 0.84)	0.68	19	10^{-77}	0.68 (0.15)	4.6	44
NOE, all[i,o]	34.6	−0.045 (−0.08, −0.02)	0.30	−18.7	10^{-77}	0.44 (0.03)	14.1	96
NOE, core[j,o]	−12	−0.09 (−0.11, −0.05)	0.07	−19	10^{-80}	0.22 (0.11)	2.0	51

[a] Average relative deviation, $(A^{exp} - A^{sim})/|A^{exp}|$, in percent, between the measured and simulated values. [b,c] Median value and mean absolute deviation of the differences between the simulated and measured values of the corresponding parameters. Numbers in the parentheses indicate the 95% confidence interval for the median. [d] Overall z-score value. [e] The probability that the observed value of the overall z-score could occur by chance. [f] The slope of the linear correlation between the simulated and observed values of the parameters: A^{sim} = offset + slope(A^{exp}). Indicated in the parentheses are standard errors in this parameter. [g] t-statistics determined as t = slope/error(slope). [h] Number of residues included in the analysis. [i] All resolved residues were included in the analysis. [j] Only residues belonging to the protein core were used for the analysis. [k] Uniform (residue-independent) scaling was applied to simulated values, $(A^{sim})_{sc} = A^{sim}$/scale. The indicated values of a uniform scaling parameter were optimized to yield the lowest absolute value of the overall z-score, Z. The optimal uniform scaling for a 10%-trimmed data set was 0.943, 0.946, 0.69, and 0.952 for $J(0)$, $J(\omega_N)$, $J(0.87\omega_H)$, and S^2, respectively. [l] The simulated values were divided by the specified scaling parameter: A^{sim}/scale. [m] Simulated values of S_u^2 were divided by S_{lib}^2, while those of S_f^2 and S_s^2 were multiplied by S_{lib} prior to calculating the spectral densities; the values of S_{MD}^2 were not altered. This modification compensates approximately evenly for the effect of librations on both S_f^2 and S_s^2. [n] Values of $J(0)$ and $J(\omega_N)$ presented here were derived by using method 1 of the reduced spectral density approach[39] (see Materials and Methods). Similar results were obtained by using method 2 (Supporting Information). [o] The simulated values of R_1, R_2, and heteronuclear NOE were derived from simulated spectral densities according to standard expressions:[46] a full treatment of the high-frequency spectral densities was applied, distinguishing between $J(\omega_H)$, $J(\omega_H + \omega_N)$, and $J(\omega_H - \omega_N)$.

values of order parameters S_{MD}^2 for these regions of the βARK1 PH domain are then dominated by transitions of backbone dihedral angles rather than by fluctuations of the dihedral angles within one particular conformation. A MD simulation so short as not to sample a transition would result in larger than realistic simulated order parameters. Sampling only a few of these transitions during a simulation of intermediate length could result in significantly reduced simulated order parameters S_{MD}^2 in these parts of the protein compared to the experimentally derived values of S_{NMR}^2 (Figure 6). The equilibrium value from a converged MD trajectory might be larger when the transitions occur frequently on the time scale of the simulation, as was

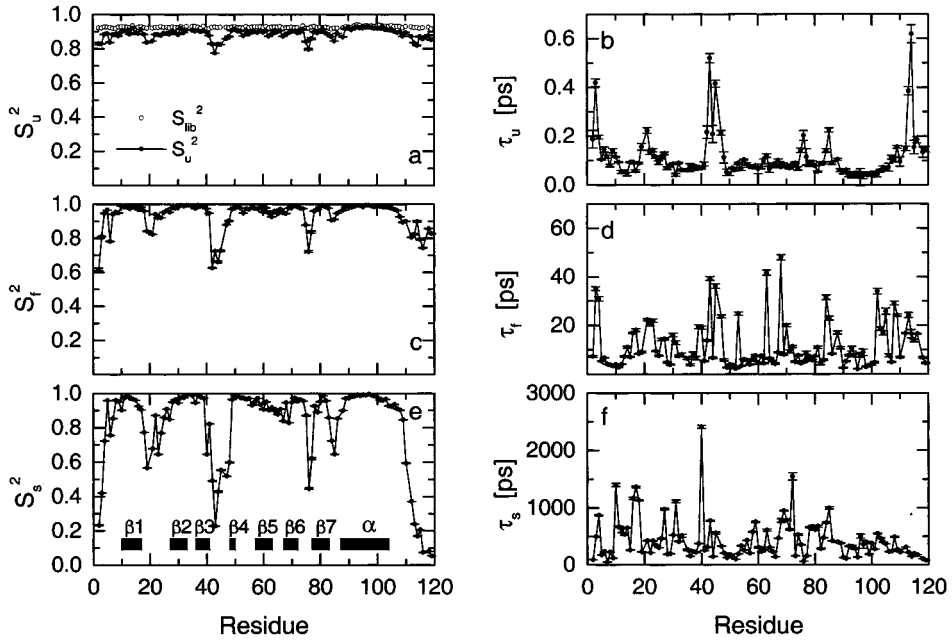

Figure 7. Microdynamic parameters, S^2 and τ_{loc}, for the (a,b) ultrafast, (c,d) fast, and (e,f) slow motions of the backbone NH vectors observed in the βARK1 PH domain MD simulation. These results were obtained by fitting the correlation functions of vector reorientations, obtained from the simulated trajectory, to eq 4, as described in the Materials and Methods section. Also shown in panel (a) are librational squared order parameters (open circles), for comparison.

also seen by Smith et al.[28] Furthermore, these dihedral angle transitions influence several NH bond vectors in the neighborhood. This becomes obvious when comparing correlation functions of dihedral angles φ and ψ with correlation functions of NH bond vectors in the neighborhood. The characteristic time course of the correlation function for a dihedral angle undergoing a rare transition is reflected in correlation functions $C_{loc}(t)$ of several NH bond vectors at the same time t (data not shown).

The C-terminal extension of the α-helix of the βARK1 PH domain appears stiffer than that derived from NMR spin relaxation experiments. We assume that the repulsion of the five positively charged side chains R109, K112, K114, K116, and R118 hinders the local motion. The distribution of charges, their compensation by shielding ions, and the "sample concentration" (box size when using periodic boundary conditions) will influence the details of protein backbone dynamics. The ionization states of the side chains in the MD simulation were described by assuming unit charges corresponding to pH 7 because experimental data were not available. One would not expect that side chains in less compact and solvent accessible parts of a protein would differ extremely from their intrinsic pK_a values measured for free amino acids. But, the less restricted motion of the C-terminus of the βARK1 PH domain could be easily influenced by a change in the ionization state. Unfortunately, accurate simulation of a specific pH value and intermediate side chain ionization states in MD simulations is still not yet practical for these time scales.

This comparison of the simulated and experimental order parameters does not allow an assessment of the relative role of motions on different time scales. The analysis of spectral densities presented below resolves this problem, in part.

(2) Librational and Dihedral Angle Contributions to Simulated Order Parameters. As already pointed out in ref 9, the values of S_{lib}^2 could be considered as the upper bound for the simulated order parameters, corresponding to the case of a rigid backbone, when dihedral angle fluctuations are ignored. The observed variation between S_{lib}^2 and S_{MD}^2 along the protein backbone (Figure 7a) indicates different contributions of the librational motion and the dihedral angle fluctuations for residues belonging to the rigid and flexible parts of the protein backbone. The values of S_{MD}^2 in the α-helix and most residues of the β-strands are dominated by the librational motion of NH bond vectors. This is consistent with the earlier observation[9] that the dihedral angle fluctuations are strongly suppressed in the secondary structure elements, and so their contribution to the simulated order parameters is minor. The contribution of dihedral angle fluctuations to the simulated order parameters within β-sheets is somewhat larger than that for the α-helix. These observations are confirmed by the calculated standard deviations of simulated dihedral angles, φ and ψ, which are on average 8.8° for the α-helix, 14.8° for the β-sheets, and 23.5° for the loops and termini.

(3) Parameterization of the Correlation Functions Provides a Glimpse into Local Dynamics on Different Time Scales. In addition to the overall S^2 reported above, time-scale-dependent microdynamic parameters were derived from the parameterization of $C_{loc}(t)$, according to eq 4. An F-test[65] was used to assess the statistical significance in the improvement of fit by using six adjustable parameters (three time scales of motion) versus four parameters (two time scales of motion). For all amides, the six-parameter fit was significantly better, with a confidence level greater than 99%. The order parameters (S_u^2, S_f^2, S_s^2) and local correlation times (τ_u, τ_f, τ_s) derived from the fitting procedure correspond to three time scales of motion. The data are depicted in Figure 7, and the statistics are presented in Table 2. Although this model function is just an analytical approximation of the simulated correlation function, the three time scales of motion may have physical meaning.

(a) Ultrafast Motion. The order parameters S_u^2 for the ultrafast ($\tau_u \leq 0.1$ ps) processes are expected to be influenced mainly by the librational motion of the NH bond vectors. Indeed, the calculated mean values of S_u^2 and S_{lib}^2 are close, 0.888 ± 0.029 and 0.926 ± 0.005, respectively. The individual values of S_u^2 are systematically lower than the librational order

(65) Draper, N. R.; Smith, H. *Applied regression analysis*; John Wiley & Sons: New York, 1981.

Table 2. Average Values and Standard Deviations (in Parentheses) of the Derived Parameters for Simulated Backbone Dynamics in the βARK1 PH Domain

	S_u^2	S_f^2	S_s^2	τ_u, ps	τ_f, ps	τ_s, ps
secondary structure	0.903 (0.015)	0.978 (0.022)	0.941 (0.069)	0.071 (0.023)	9.44 (7.69)	517.0 (408.4)
loops and termini	0.869 (0.032)	0.883 (0.098)	0.679 (0.285)	0.162 (0.118)	14.10 (11.02)	329.6 (236.4)
total	0.888 (0.029)	0.936 (0.082)	0.824 (0.236)	0.111 (0.092)	11.52 (9.57)	433.3 (353.7)

parameters, although the relative deviations are small (Figure 7), on average 4.1% (2.6% for the secondary structure). The difference is minor for residues in the α-helix and more pronounced for residues in the loops $\beta1/\beta2$, $\beta3/\beta4$, $\beta6/\beta7$, and $\beta7/\alpha$ as well as termini. The calculation of S_{lib}^2 excludes contributions from reorientational motion of the peptide plane (e.g., due to dihedral angle fluctuations) whereas that of S_u^2 does not; therefore, S_{lib}^2 should be considered as an upper bound of S_u^2. The reduced values of S_u^2 could be due to fluctuations in the backbone dihedral angles. The amplitudes of the φ and ψ fluctuations in the 0.1 ps time frame, averaged over 10 000 such frames from the 20 ps trajectory with fine storage step, indeed, show some increase concomitant with the decrease in S_u^2. However, this increase is marginal, in most cases of the order of 10−15%, and cannot account for the observed difference between S_u^2 and S_{lib}^2 in the flexible parts of the protein. Besides the possible influence of the dihedral angle fluctuations, the decreased order parameters S_u^2 could also result from the influence of the subsequent fast time scale of motion, as an artifact of the fitting procedure. Obviously, the ultrafast and the fast time scales of motion cannot be separated accurately when the time resolution of $C_{loc}(t)$ is greater than τ_{lib}. Constraining S_u^2 to the value of S_{lib}^2 derived from the residue-fixed frame did not improve the fit. The oscillatory behavior of the correlation functions calculated with 2 fs resolution (Figure 5) rendered derivation of S_u^2 from these high-resolution data impossible.

The fit provides only a rough estimate of the τ_u values, since the storage step size of 0.1 ps renders the fitted values of $\tau_u \leq 0.1$ ps overestimated/imprecise. In this regard, the increased values of τ_u (>0.2 ps) and a concomitant decrease in S_u^2 in the $\beta3/\beta4$ loop and in the termini could be caused by dihedral angle fluctuations. This inaccuracy/imprecision in τ_u, however, is not critical for the calculated values of $J(\omega)$, as the corresponding contributions to relevant spectral densities are negligible, see eq 5a.

(b) Fast Motion. The fast time scale motions have an average local correlation time of about 10 ps, while the slow motions are characterized by average local correlation times of 400 ps. The different time scales of motion are separated by at least 1 order of magnitude and are therefore likely to be of some physical significance. With the exception of the ultrafast time scale, low order parameters for a particular time scale are not correlated with an increase in the corresponding correlation times and vice versa. The fast time scale of motion of the NH bond vectors is caused by the fluctuations of dihedral angles φ and ψ in a particular conformation. This becomes obvious from correlation functions for the dihedrals for residues in which dihedral angle transitions are absent or calculated for intervals which are free of dihedral angle transitions (data not shown). These correlation functions show a decay on a comparable time scale, typically in the range of 10 ps. The corresponding order parameters S_f^2 reflect the amplitudes of dihedral angle fluctuations in conjunction with their correlation coefficients (cf. Figure 3). S_f^2 is close to 1 for residues where ψ_{i-1} and φ_i angles change in an anticorrelated manner and exhibit low amplitudes of fluctuations. The order parameters S_f^2 drop around residues P20, E44, G76, and C83 as well as at the termini where ψ_{i-1} and φ_i angles fluctuate in a correlated manner. The amplitudes of dihedral angle fluctuations are similar in the loops and termini, as indicated by both the order parameters S_f^2 (Figure 7) and the standard deviations of dihedral angles in Figure 3. In general, larger values of τ_f were obtained for loops and termini compared to those for secondary structure elements. The local correlation times τ_f could be somewhat underestimated because of the overestimation of τ_u, as discussed above. Similar to τ_u, this should not significantly affect the spectral densities, since the corresponding contributions to relevant spectral densities are small, see eq 5a.

(c) Slow Time Scale. The slow time scale motion is caused mostly by dihedral angle transitions. These include either rare transitions or slow but constant changes in dihedral angles. The majority of low order parameters S_s^2 belong to fragments in which the MD simulation is not yet fully converged with respect to the time scale of motion. The obtained local correlation times τ_s are statistically not very precise, as their values are close to or beyond 600 ps, which is the maximum time we included in the fitting procedure for reasons of statistical significance. This is also reflected in the uncertainties derived from Monte Carlo simulations. The local correlation times τ_s seem to correspond to τ_{loc} derived from the ^{15}N spin relaxation data. It is striking that whenever τ_{loc} is large and close to 1 ns, the fitted correlation times τ_s are small, especially at the C-terminus of the βARK1 PH domain. For residues involved in a nonconverged motional mode, this is a result of the underestimated overall order parameters S_{MD}^2 used in the optimization as a constraint. A converged MD trajectory would likely lead to higher equilibrium order parameters and therefore larger values of fitted correlation times τ_s. The quantitative agreement between fitted values τ_s and NMR-derived local correlation times is poor. Again, the model function used was chosen to obtain an analytical expression of the correlation function but was not *a priori* anticipated to display accurate time scales of motion in the physical reality.

(4) Simulated Spectral Densities. Simulated values of the spectral density were derived from the fitted order parameters S_u^2, S_f^2, S_s^2 and local correlation times τ_u, τ_f, τ_s by using eq 5. It is impractical to fully reproduce the overall hydrodynamic rotation of a protein the size of the βARK1 PH domain ($\tau_c \approx 8$ ns) from a 6 ns MD simulation. Therefore, we used the characteristics of the overall motion derived from the NMR relaxation measurements. The overall rotation of the protein was assumed to be anisotropic and was characterized by an axially symmetric diffusion tensor derived from the analysis of ^{15}N spin relaxation rates R_1 and R_2: $\tau_c = 7.94 \pm 0.21$ ps; $D_\parallel/D_\perp = 1.34 \pm 0.07$; $\Phi = 179.0° \pm 8.5°$; $\Theta = 94.0° \pm 6.2°$. Values of the spectral density obtained for isotropic overall motion did not differ significantly from those for the anisotropic model, with the exception of residues in the α-helix. Here, there are systematic differences, namely $\sum_{i=1,n}|J_{aniso}(\omega) - J_{iso}(\omega)|/n = 0.35 \pm 0.07$ for $J(0)$ and 0.025 ± 0.005 for $J(\omega_N)$, consistent with the α-helix orientation along the unique principal axis ($D_\parallel$) of the overall rotational diffusion tensor. Being well structured and quite rigid, the α-helix has to be treated with the anisotropic model for the overall motion. Therefore, all the simulated values depicted in Figure 8 were calculated assuming anisotropic overall motion.

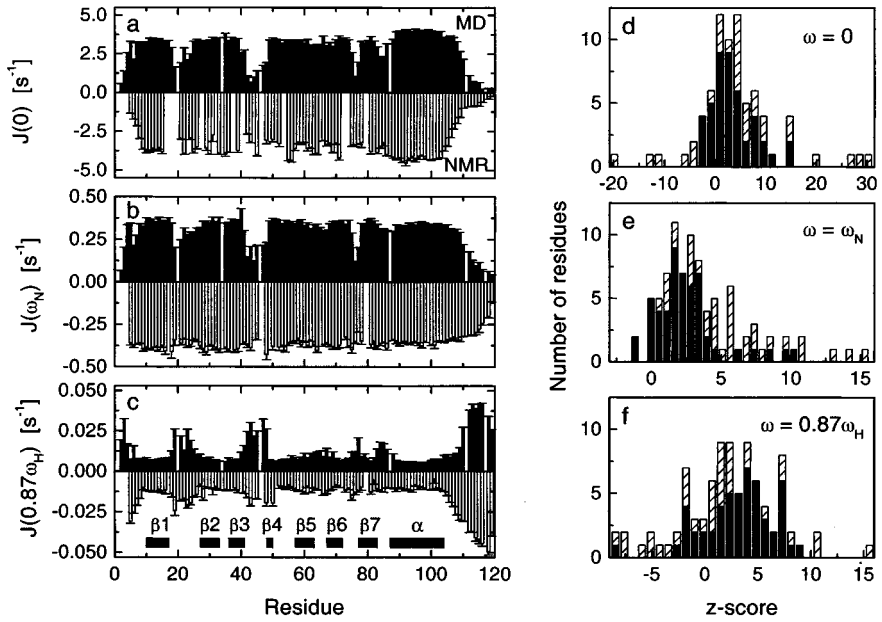

Figure 8. Comparison of the simulated (solid bars) and experimental (open bars) spectral density functions, (a) $J(0)$, (b) $J(\omega_N)$, and (c) $J(0.87\omega_H)$. The experimental values were multiplied by -1, for better comparison. Panels (d)−(f) on the right represent histograms of the *z*-scores characterizing deviations between the experimental and simulated spectral densities: hatched bars correspond to all residues (10%-trimmed) and solid bars to core residues in the βARK1 PH domain.

(5) Contribution to Spectral Densities from the Motions in Different Time Scales. Various time scales of motion considered here, represented by various terms in eq 5, have different effects on the calculated spectral densities, as illustrated in Figure 9. The spectral densities at the NMR-relevant frequencies are sensitive to the amplitudes but not the correlation times of the ultrafast and fast motions. In general, due to their Lorentzian functional form (eqs 5 and 5a), the relevant spectral densities $J(\omega)$ in eq 1 are sensitive to those correlation times that approach the inverse of the corresponding frequencies, $1/(0.87\omega_H)$ (305 ps) or $1/\omega_N$ (2.62 ns). As can be seen from a comparison of the contributions to spectral density from various terms in eq 5 presented in Figure 9, the relevant spectral densities are largely determined by the interplay between the first and the last terms in eq 5a, related to contributions from the overall and slow local motions, respectively. Neglecting small contributions from the ultrafast and fast time scale motions (see Figure 9), the relevant spectral densities can be approximated as follows:

$$J_{MD}(0) \approx \frac{2}{5}d^2 S_{MD}{}^2 \tau_c \left[1 + \frac{1 - S_s{}^2}{S_s{}^2}\frac{\tau_s}{\tau_c}\right] \approx \frac{2}{5}d^2 \tau_c S_{MD}{}^2 \quad (6a)$$

$$J_{MD}(\omega_N) \approx \frac{2}{5}d^2 S_{MD}{}^2 \tau_c{}^{-1}\omega_N{}^{-2}\left[1 + \frac{1 - S_s{}^2}{S_s{}^2}\frac{\tau_s\tau_c\omega_N{}^2}{1 + (\tau_s\omega_N)^2}\right] \approx \frac{2}{5}d^2 \tau_c{}^{-1}\omega_N{}^{-2} S_{MD}{}^2 \quad (6b)$$

$$J_{MD}(0.87\omega_H) \approx \frac{2}{5}d^2 S_{MD}{}^2 \tau_c{}^{-1}(0.87\omega_H)^{-2} \times \left[1 + \frac{1 - S_s{}^2}{S_s{}^2}\frac{\tau_c\tau_s(0.87\omega_H)^2}{1 + (\tau_s(0.87\omega_H))^2}\right] \quad (6c)$$

Here we assume $\tau_c \gg \omega_N{}^{-1}$ and consider the isotropic overall tumbling, for simplicity. The spectral densities at $\omega = 0$ and ω_N are practically controlled by the correlation time of the overall motion; this contribution is proportional to $S_{MD}{}^2 = S_u{}^2 S_f{}^2 S_s{}^2$. The effect of protein dynamics on the high-frequency component, $J(0.87\omega_H)$, is more complex: the contributions from slow internal motions are comparable to or greater than those from the overall motion. In this case, both the amplitude ($S_s{}^2$) and the correlation time (τ_s) of the slow internal motion become important for the calculation of spectral density. The slow motion contribution scales as $(1 - S_s{}^2)/S_s{}^2$ and therefore becomes significant for lower values of the related order parameter.

(6) Comparison of Experimental and Simulated Spectral Densities. Experimentally derived values of the spectral density function $J_{NMR}(\omega)$ at $\omega = 0$, ω_N, and $0.87\omega_H$ were calculated by using the measured ^{15}N spin relaxation rates, R_1 and R_2, and heteronuclear NOE as described in the Materials and Methods section, eqs 1a−c. The data are shown in Figure 8. Several residues were excluded from the comparison of the simulated and experimental $J(0)$ values: K16, G18, N19, T23, R27, R28, L32, E38, G41, A45, E53, E54, Q56, I73, R74, L81, and R101; all these residues revealed a significant conformational exchange contribution to R_2.

One would expect a good agreement between the simulated and experimental spectral densities dominated by the overall tumbling for those parts of the protein that have high order parameters and converged motional modes. Experimental and simulated values of $J(0)$ in Figure 8 show a similar profile over the sequence of the protein. They disagree obviously for residues with unconverged motions for the NH bonds, namely for the loops $\beta1/\beta2$, $\beta3/\beta4$, $\beta6/\beta7$, and $\beta7/\alpha$ and the termini. Here, the simulated order parameters $S_{MD}{}^2$ and therefore the simulated spectral densities $J(0)$ are underestimated due to the unconverged motions. The overall *z*-score values are high (Table 1). A similar picture is observed for $J(\omega_N)$, although the overall ratio between the simulated and the experimental values is much lower than that for $J(0)$. Overall, the simulated $J(0)$ values and $J(\omega_N)$ values are slightly but systematically underestimated (by 6% on average for the core residues) compared to the experimental values (Figure 8, Table 1). This holds for both the lower and upper bound estimates of the experimental spectral densities (Sup-

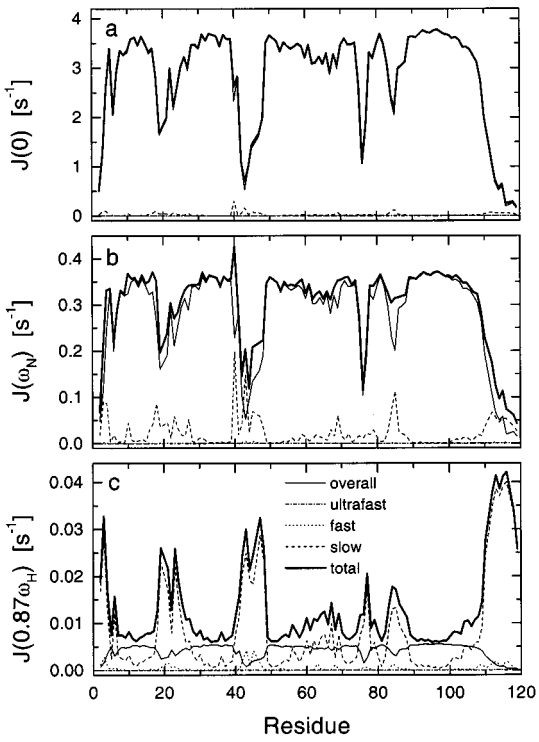

Figure 9. Comparison of the contributions from various terms in eq 5a to the spectral densities (a) $J(0)$, (b) $J(\omega_N)$, and (c) $J(0.87\omega_H)$. The thin solid, dashed, dotted, and dash−dotted lines represent contributions from the overall (macromolecular), slow, fast, and ultrafast motions, respectively, calculated for each amide group in the βARK1 PH domain. The thick solid line corresponds to the total spectral density, calculated by using eq 5. These calculations assume isotropic overall rotational diffusion with $\tau_c = 7.94$ ns, for simplicity. The contributions from the ultrafast and fast motions are practically negligible for all three spectral densities. The contribution to $J(0)$ and $J(\omega_N)$ from the slow-motion term can also be neglected for most residues but becomes comparable to the overall motion term or even dominant in the case of $J(0.87\omega_H)$. In panel (a), the macromolecular contribution overlaps the total for nearly all residues.

porting Information) derived by using methods 1 and 2 of the spectral density approach[39] (see Materials and Methods). The disagreement between the experimental and simulated values of $J(0.87\omega_H)$ is more pronounced: most simulated values for the protein core are underestimated by 10−50% (16% on average for all residues, 21% for the core).

A direct comparison of the relaxation parameters, simulated versus experimentally measured (Table 1), indicates a similar trend in these data: the simulated R_1 and R_2 values are underestimated by ~6−7%. This is expected because the major contributions to these relaxation rates come from $J(0)$ and $J(\omega_N)$, respectively. The discrepancy between measured and simulated heteronuclear NOE values is more pronounced (Table 1): the simulated values are overestimated by 59% (10% for the core); this difference, however, is not easily interpretable, because NOE values are determined by the ratio of a combination of the high-frequency spectral densities, $6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)$, to R_1. This comparison did not involve the reduced spectral density assumption: the simulated values of relaxation parameters were derived from simulated spectral densities by using standard expressions[46] relating R_1, R_2, and NOE to all relevant $J(\omega)$ values.

Since spectral densities represent the amount of motion at a particular frequency, comparison of the experimental and simulated spectral densities suggests that there is more internal motion present in the simulation than in the real protein. There are several possible explanations for this.

(a) Low-Frequency Spectral Densities, $J(0)$ and $J(\omega_N)$. The calculated values of the simulated spectral densities depend on the hydrodynamic characteristics of the protein, which were derived directly from the experimental NMR data. This derivation uses a well-established procedure based on the R_2/R_1 ratio,[43,47,66−68] so the errors in the hydrodynamic parameters are expected to be minimal. The experimentally determined overall rotational diffusion characteristics of the protein are in good agreement with the results of hydrodynamic calculations using the "bead-model"[69] and assuming a hydration shell of 3.5 Å around the protein (data not shown). Although small errors in the overall correlation time cannot be excluded, these would have different effects on $J(0)$ and $J(\omega_N)$, due to differences in the τ_c dependence of the low-frequency spectral densities, as follows from eqs 6a,b. For example, an underestimation of τ_c would result in a decrease in $J(0)$ and, at the same time, an increase in $J(\omega_N)$. Therefore, it is very unlikely that the observed systematic overall underestimation of both simulated $J(0)$ and $J(\omega_N)$ could be caused by an error in the derived hydrodynamic characteristics of the protein.

The observed difference between the experimental and simulated values of $J(0)$ and $J(\omega_N)$ could be attributed to a slight underestimation of the generalized order parameters S_{MD}^2 (eqs 6a,b). The underestimated values of S_{MD}^2 for the protein core cannot be explained by the effect of an unconverged trajectory (Figure 6d). S_{MD}^2 values are derived directly from the whole trajectory (eq 3) and therefore are independent of the parametrization of the correlation function. It is possible that some of the motional processes indirectly contributing to the resulting amplitude of motion (S_{MD}^2) are less constrained in MD simulation compared to the case in a real protein. For example, the agreement for $J(0)$ and $J(\omega_N)$ improves when the contribution from librational motion is taken out (e.g., by dividing the resulting spectral densities by S_{lib}^2) (Table 1). An optimal uniform scaling parameter of 0.93 for $J(0)$ is similar to that for $J(\omega_N)$, 0.94, as well as to the uniform scaling of 0.96 for the generalized order parameters. These numbers are also close to the overall level of S_{lib}^2. The MD simulation reproduced the time scale of the librational motion accurately. Therefore, one could speculate that the amplitude of librational motion might not be reproduced accurately by the used force field, a conclusion similar to that made on the basis of the analysis of S^2 in rigid parts of the molecule.

An alternative, spectroscopic explanation is that the $J_{MD}(\omega)$ values, eqs 6, are determined from the product $d^2S_{MD}^2$, so the simulated spectral densities $J_{MD}(\omega)$ also directly depend on the strength of the dipolar interaction. The dipolar coupling was calculated here by assuming the nominal NH bond length of 1.02 Å. The ultrafast librational motions observed in this MD simulation are at least 3 orders of magnitude faster than any characteristic frequency of spin transitions in the system. One might expect that, on the time scale relevant for spin relaxation, the effective $^{15}N-^{1}H$ dipolar interaction is already somewhat reduced by these ultrafast librations of the NH bond; therefore, the nominal bond length of 1.02 Å used in the calculations will

(66) Bruschweiler, R.; Liao, X.; Wright, P. E. *Science* **1995**, *268*, 886−889.
(67) Lee, L. K.; Rance, M.; Chazin, W. J.; Palmer, A. G., III. *J. Biomol. NMR* **1997**, *9*, 287−298.
(68) Copie, V.; Tomita, Y.; Akiyama, S. K.; Aota, S.; Yamada, K. M.; Venable, R. M.; Pastor, R. W.; Krueger, S.; Torchia, D. A. *J. Mol. Biol.* **1998**, *277*, 663−682.
(69) de la Torre, J. G.; Bloomfield, V. A. *Q. Rev. Biophys.* **1981**, *14*, 81−139.

already be increased compared to the actual bond length, to account for the effect of averaging. It is possible that the effect of the ultrafast motions is accounted for in our calculations twice: in the effective (scaled) bond length and in the order parameter S_{MD}^2, containing the contribution from the bond librations. To correct this, one could either remove the librational motion contribution from S_{MD}^2 or use a shorter (generic, unscaled) NH bond length. As mentioned above, rescaling of the MD spectral densities or order parameters by ~6−7% (S_{lib}^2) significantly improves the agreement for $J(0)$ and $J(\omega_N)$; similar results were obtained for generalized order parameters. The same improvement is obtained when the NH bond length is set to a shorter value of 1.008 Å, which is the actual bond length in the AMBER force field. Interestingly, if we scale this bond length by $(S_{lib}^2)^{-1/6}$ to account for the effect of librational motion, we get 1.02 Å, in excellent agreement with the conventionally used value of r_{NH}. A similar averaging effect, due to librational motions although on a larger time scale, could account for an increased effective r_{NH} derived from residual dipolar coupling measurements.[70]

A somewhat greater ratio of the experimental to simulated values for $J(0)$ as compared to that for $J(\omega_N)$ (hence greater *z*-scores) could be due to small, undetectable, conformational exchange contributions to R_2, affecting the experimental values of $J(0)$. Site-specific variations in the amplitude and orientation of ^{15}N CSA[71−73] neglected here, could also contribute to the apparent experimental values of $J(0)$ and $J(\omega_N)$. Experimental approaches to address this issue are currently being developed. However, given the distribution of CSA values observed, it is unlikely that this effect could lead to a systematic overestimation of the experimental values.

In conclusion, the underestimation of low-frequency correlation functions is likely related to systematic discrepancies in apparent bond lengths and should be subject to further experimental investigation.

(b) High-Frequency Spectral Density, $J(0.87\omega_H)$. The nature of the disagreement between the experimental and simulated values of $J(0.87\omega_H)$ is more complex. The agreement was only marginally improved after dividing the total amplitude of motion by S_{lib}^2 (Table 1). The optimal uniform scaling of S_{MD}^2 (0.748) is significantly lower than that for the other spectral densities. This spectral density component is expected to be more sensitive to slow local motions than $J(0)$ and $J(\omega_N)$, see, e.g., eq 6. Therefore, the observed disagreement for $J(0.87\omega_H)$ is likely due to some combination of both overestimated S_s^2 and underestimated τ_s. This effect is not observed in the other spectral density components, probably because of the negligible contribution from the corresponding slow motion terms in eq 5a (cf. eqs 6). The disagreement between measured and simulated values of $J(0.87\omega_H)$ could then be related to either (1) inadequate representation of the slow motion processes by the force field or (2) errors in the parameterization of the correlation function, likely due to insufficient length of the trajectory that limits the accuracy of fitting, in particular, for the slow motional processes. According to the model-free analysis of experimental data (Supporting Information), many residues in the βARK1 PH domain are characterized by $\tau_s \approx 1$ ns, which exceeds the time range for $C_{loc}(t)$ analyzed here. In addition, the underestimation of S_u^2 during the parameterization, as mentioned above, might contribute to this difference, since the product $S_{MD}^2 = S_u^2 S_f^2 S_s^2$ has to be constant. If S_u^2 is underestimated, the other order parameters will be overestimated, hence the reduced contribution of the corresponding terms in eq 5.

To shed light on possible sources of the disagreement for $J(0.87\omega_H)$, various microdynamic parameters in eq 5a were varied. We found that some improvement was obtained when the contribution from librational motions was removed from S_u^2 and included in the apparent order parameters for both fast and slow motions (as $S_f^2 S_{lib}$ and $S_s^2 S_{lib}$). It is currently not clear whether this rescaling of S_s^2 has any physical significance, although no improvement was obtained by applying a similar procedure to S_f^2 alone or by a uniform scaling of all τ_s values within ±50%.

Underestimation of the errors in the derived experimental and simulated spectral densities cannot be excluded and could also lead to higher *z*-score values. However, to account for the large observed *z*-score values (Figure 8f), one would need to assume ~5−10-fold underestimation of the uncertainties in $J(0.87\omega_H)$, which seems unlikely.

The effect of an unconverged trajectory is another possible source of errors. Unlike $J(0)$ and $J(\omega_N)$, the disagreement between the experimental and simulated values of $J(0.87\omega_H)$ is more pronounced in the protein core (Figure 8f), where the trajectory is most close to convergence, than in some of the loops (β1/β2, β3/β4, β5/β6) and in the C-terminus.

The finite length of the trajectory, which inevitably limits the values of τ_s derived here, is expected to result in the underestimation of the correlation time for slow motions. One would expect that a longer trajectory could result in greater values of τ_s. As follows from eqs 6, this in itself might somewhat increase $J(\omega_N)$ (for $\tau_s < 1/\omega_N \approx 2.65$ ns) but not the other spectral density components, as $J(0)$ is practically insensitive to these motions (for the observed S_s^2 values) and $J(0.87\omega_H)$ can only decrease when τ_s increases above $1/(0.87\omega_H) \approx 305$ ps.

(D) What Additional Information Is Available from Spectral Densities That Could Not Be Derived from the Order Parameters? The generalized order parameter provides a rather global characterization of protein dynamics, which usually does not permit dissection of various motional contributions. The comparison of the order parameter values here revealed differences in the simulated and measured amplitudes of motion for the βARK1 PH domain but did not allow an assessment of possible contributions from motions in various time scales. A more detailed understanding of motional contributions can be obtained from comparison of the spectral densities, due to their differential sensitivity to motions in various time scales. For example, as demonstrated here, the spectral densities $J(0)$ and $J(\omega_N)$ are most sensitive to the overall motion, whereas $J(0.87\omega_H)$ can also sense slow (hundreds of picoseconds) local motions. The generalized order parameters, simulated and experimental, show a disagreement similar to that observed for the spectral densities. In addition, the detailed analysis of spectral densities performed here revealed new features that were not available from the conventional comparison. In particular, it turned out that the observed disagreement for the spectral densities $J(0)$ and $J(\omega_N)$ is controlled by the underestimated simulated generalized order parameters, probably due to overestimated contribution from the (ultrafast) librational motion. The disagreement for the high-frequency component, on the other hand, is likely due to the overestimated order parameters for slow internal motion, which is still much

(70) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12341.
(71) Fushman, D.; Tjandra, N.; Cowburn, D. *J. Am. Chem. Soc.* **1998**, *120*, 10947−10952.
(72) Fushman, D.; Cowburn, D. *J. Biomol. NMR* **1999**, *13*, 139−147.
(73) Fushman, D.; Tjandra, N.; Cowburn, D. *J. Am. Chem. Soc.* **1999**, *121*, 8577−8582.

faster than the time scale of the overall rotation. While beyond the scope of this paper, it is possible that fitting *J*'s could provide useful additional parameterization of the force field used for simulations.

In addition, the approach used here allows control over the accuracy and precision of the spectral density determination from MD data. The derivation of spectral densities from the experimental data involves minimal assumptions, as opposed to order parameter determination using the model-free approach.

The analysis of dynamics of protein with extensive flexibility, using multiple approaches, is likely to aid significantly in understanding the contributions of flexibility to structure and the role of ions and solvation.

Conclusions

We have analyzed the simulated motion of the fully solvated βARK1 PH domain during an MD equilibrium simulation of 6 ns and subsequently compared the results with experimental data derived from ^{15}N spin relaxation data. Despite the long simulation time, the motions underlying the NH bond vector fluctuation almost converged for residues within secondary structure elements but did not converge for many residues within the flexible loops and at the termini. The NMR- and MD-derived microdynamic parameters suggest local correlation times of a few hundred picoseconds for secondary structure elements and up to 1 ns or more for some large-amplitude motions within loops and at the termini of the protein. Thus, MD simulations performed on more complex and larger flexible systems such as the βARK1 PH domain require an even longer simulation time for convergence, more than 10-fold the maximal local correlation time derived by NMR relaxation. This has been already pointed out[29] for BPTI and confirmed by our results. The slow internal motions on a time scale of several hundred picoseconds contribute significantly to $J(0.87\omega_H)$ and partially to $J(\omega_N)$. Here, the frequency and the amplitude of these motions become important in determining their contribution to $J(0.87\omega_H)$. The simulated order parameters and spectral densities are lower than those experimentally measured. Some differences between experimental and simulated spectral densities can be explained by the still poor sampling of the slow motions. The systematic decrease in the simulated spectral densities can be partially explained by the amplitude of the ultrafast librational motion of the NH bond vectors with respect to the peptide plane which seems to be overestimated by the force field used. The inhibition of the bond stretching using the SHAKE algorithm during the MD simulation decreases the amplitude of the bending motion. However, the effect is below the precision with which squared order parameters can be derived from NMR relaxation and MD simulation. Analysis of simulated spectral densities might be helpful in the future for the improvement of force field parameterizations. The bending and stretching motion of NH bonds is also responsible for the initial drop of the autocorrelation functions of NH bond vectors when using a conventional, coarse-grained data-storage step in MD simulation. Autocorrelation functions derived from a trajectory with very small storage steps of 2 fs show strong oscillations due to the nonstochastic character of the bond libration. Both the initial drop and the oscillations prevent an accurate characterization of the ultrafast motion of the NH bond vectors.

Acknowledgment. S.P. gratefully acknowledges a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft. The computing resources were purchased with NSF Grant BIR-9601845. D.C. and D.F. are supported by NIH Grant GM-47021. S.P. thanks R. Abseher and C. Francart for advice and comment.

Supporting Information Available: One table containing experimentally derived model-free parameters for the backbone NH vectors in the βARK1 PH domain and one table presenting statistics of the comparison between the simulated and measured spectral densities (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0031117