Articles

Characterization of the Internal Motions of *Escherichia coli* Ribonuclease HI by a Combination of ^{15}N-NMR Relaxation Analysis and Molecular Dynamics Simulation: Examination of Dynamic Models

Kazuhiko Yamasaki,* Minoru Saito,‡ Motohisa Oobatake, and Shigenori Kanaya

Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan

Received December 28, 1994; Revised Manuscript Received March 10, 1995⊗

ABSTRACT: The backbone dynamics of *Escherichia coli* ribonuclease HI (RNase HI) in the picosecond to nanosecond time scale were characterized by a combination of measurements of ^{15}N-NMR relaxation (T_1, T_2, and NOE), analyzed by a model-free approach, and molecular dynamics (MD) simulation in water. The MD simulations in water were carried out with long-range Coulomb interactions to avoid the artificial fluctuation caused by the cutoff approximation. The model-free analysis of the ^{15}N-NMR relaxation indicated that RNase HI has a rotational correlation time of 10.9 ns at 27 °C. The generalized order parameter (S^2) for the internal motions varied from 0.15 to 1.0, with an average value of 0.85, which is much larger than that of the RNase H domain of HIV-1 reverse transcriptase (0.78). Large internal motions (small order parameters) were observed in the N-terminal region (Leu2−Lys3), the loop between β-strands A and B (Cys13−Gly15), the turn between α-helix I and β-strand D (Glu61, His62), the loop between β-strand D and α-helix II (Asp70−Tyr71), the loop between α-helices III and IV (Ala93−Lys96), the loop between β-strand E and α-helix V (Gly123−His127), and the C-terminal region (Gln152−Val155). The effective correlation time observed in these regions varied from 0.45 ns (Glu61, Lys96) to 2.2 ns (Leu14). The order parameters calculated from the MD agreed well with those from the NMR experiment, with a few exceptions. The distributions of most of the backbone N−H vectors obtained by MD are approximately consistent with the diffusion-in-a-cone model. These distributions, however, were elliptic, with a long axis perpendicular to the plane defined by the N−H and N−C_α vectors. Distributions supporting the axial fluctuation model or the jump-between-two-cones model were also observed in the MD simulation.

NMR provides detailed information on the dynamic properties of molecules in a wide time range, from picoseconds to days (Wagner & Wüthrich, 1986). The measurement of ^{15}N-NMR relaxation corresponds to the very short time scale of picoseconds to nanoseconds. Two-dimensional ^{1}H-detected pulse schemes have been developed to measure relaxation times with higher resolution (Nirmala & Wagner, 1988, 1989; Kay et al., 1989, 1992). When analyzed by a model-free approach (Lipari & Szabo, 1982), the relaxation times provide information about the spatial restriction of the N−H vector orientation (generalized order parameter: S^2),[1] the effective correlation time of the internal motion (τ_e; picosecond to nanosecond time scale) and rotational tumbling (τ_R; nanosecond time scale), and the contribution of chemical exchange in the millisecond time scale (Δex) (Kay et al., 1989; Clore et al., 1990).

For interpretation of the NMR relaxation, several dynamic models have been reported, such as Woessner's reorientation-about-an-axis model (Woessner, 1962), the diffusion-in-a-cone model (Kinoshita et al., 1977; Lipari & Szabo, 1980), the N-site-jump model (Wittebort & Szabo, 1978), the jump-between-two-cones model (Clore et al., 1990), and the Gaussian axial fluctuation model (Brüschweiler & Wright, 1994). The generalized order parameter (S^2) can be calculated if the dynamic model and a few parameters that describe the model are given. For example, in Woessner's model, $S^2 = [P_2(\cos\theta)]^2$, where $P_2 = (3x^2 - 1)/2$ is the second Legendre polynomial and θ is the cone semiangle (Lipari & Szabo, 1982). However, NMR does not provide an actual description of the motions, and therefore, NMR relaxation data alone are not satisfactory to judge the adequacy of the dynamic models. In contrast, molecular dynamics (MD) simulation describes the unaveraged internal motions in the time scale of picoseconds to nanoseconds. Comparisons between the results obtained by NMR relaxation analysis and MD have been reported for several proteins (Lipari et al., 1982; Chandrasekhar et al., 1992; Eriksson et al., 1993; Fushman et al., 1994a; Balasubramanian et al., 1994).

* To whom correspondence should be addressed (telephone, +81-6-872-8204; Telefax, +81-6-872-8210; E-mail, yamasaki@pes4.peri.co.jp).

‡ To whom correspondence on molecular dynamics simulation should be addressed (telephone, +81-6-872-8202; Telefax, +81-6-872-8210).

⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1995.

[1] Abbreviations: ASA, accessible surface area; Δex, exchange contribution to line shape; HIV, human immunodeficiency virus; MD, molecular dynamics; NOE, nuclear Overhauser effect; RNase HI, ribonuclease HI; RMSD, root mean square deviation; RMSF, root mean square fluctuation; S^2, generalized order parameter; S_f^2, order parameter for the faster phase; S_s^2, order parameter for the slower phase; SD, standard deviation; τ_e, effective correlation time for internal motions; τ_R, correlation time for rotational tumbling; T_1, spin−lattice relaxation; T_2, spin−spin relaxation.

0006-2960/95/0434-6587$09.00/0

Approximate agreements between order parameters from NMR and MD have been obtained, which prompts us to employ MD to describe the motion observed by NMR and to examine the dynamic models. Very recently, Fushman et al. (1994b) have demonstrated an MD study classifying the distribution of the NH vectors to three types: very high restriction, a few jumps, and very low restriction. They have concluded that the first type corresponds to the diffusion-in-a-cone model (Kinoshita et al., 1977; Lipari & Szabo, 1980).

These MD simulations, however, have been carried out in water, with the truncation of long-range Coulomb interactions to decrease the computational burden. Recently, Saito (1994) has demonstrated that this cutoff approximation caused artificial structural deviations and fluctuations in a protein. These artifacts may cause a discrepancy between the MD simulation and the NMR experiments.

Escherichia coli ribonuclease HI (RNase HI) is an enzyme that cleaves the RNA moiety in a DNA−RNA hybrid duplex (Crouch & Dirksen, 1982) and that has a molecular mass of 17.5 kDa (Kanaya & Crouch, 1983). X-ray crystallography has revealed its globular structure, consisting of five α-helices and five β-strands (Katayanagi et al., 1990, 1992; Yang et al., 1990). A related protein is the RNase H domain of HIV-1 reverse transcriptase, which alone is enzymatically inactive. A ^{15}N NMR analysis of the internal motion of this domain, using a model-free approach, has been carried out by Powers et al. (1992). This analysis revealed the mobile dynamic features of the entire protein.

In the present study, we analyzed the NMR relaxation times of *E. coli* RNase HI to determine the dynamic features of this active enzyme. To obtain an actual description of the motions, the MD simulation in water was carried out without the cutoff of the long-range Coulomb interactions, to avoid the artifacts caused by the cutoff approximation. The MD provided information for examination of the dynamic models of the internal motions, which is necessary for optimal interpretation of the NMR results.

MATERIALS AND METHODS

Sample Preparation. *E. coli* RNase HI was overproduced and purified as described previously (Kanaya et al., 1993), except that *E. coli* JM109 (Maniatis et al., 1982), instead of *E. coli* HB101, was used as the host strain for the overproduction and that M9 medium (Maniatis et al., 1982) containing 0.5 g/L $^{15}NH_4Cl$ (Shoko Co. Ltd.), instead of LB medium, was used for the growth of cells. From a 1-L culture, 15−25 mg of ^{15}N-labeled RNase HI was obtained. Four milligrams of the purified protein was dissolved in 200 μL of a solution for NMR measurement [100 mM CD_3COONa (pH 5.5), prepared from CD_3COOD (CIL) and NaOH (Nacalai Pure Chemicals Co. Ltd.)] using a Centricon 10 filter unit (Amicon). A 10% solution of D_2O (Isotec Inc.) was added to the protein solution to make a final concentration of roughly 1 mM RNase HI. Before the measurement, the sample was degassed for 5 min in order to minimize the dissolved O_2.

Measurement of NMR. All NMR spectra were measured on a Bruker AMX-600 spectrometer (600.14 MHz for ^1H and 60.82 MHz for ^{15}N) with a probe temperature of 27 °C. Proton and nitrogen chemical shifts were determined relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) and liquid NH_3, respectively. The pulse sequences for the measurements of ^{15}N T_1 and T_2 were the same as those described by Kay et al. (1992). The pulse sequence for the measurement of $^1H-^{15}N$ NOE was also described by Kay et al. (1989). The sweep width for ^1H was 8065 Hz, and that for ^{15}N was 3250 Hz. The delay after acquisition was 3.0 s for T_1 or 1.5 s for T_2 measurement. The ^1H saturation by 120° pulses for the NOE experiments was done for a duration of 3.0 s. The delays for the relaxation measurements were 50, 100, 200, 400, and 600 ms for T_1 and 14.6, 29.3, 58.6, 87.8, and 117.0 ms for T_2. The measurements for different delays were done simultaneously, in an interleaved way. GARP1 ^{15}N decoupling (Shaka et al., 1985) was employed during the detection period of 2K data points. Free induction decays of 32 (T_1), 64 (T_2), or 64 (NOE) scans were collected for 512 points in the t_1 domain, using the TPPI−States method (Marion et al., 1989). Data processing was done by using the FELIX 2.30 software (Biosym) on an IRIS indigo II computer (Silicon Graphics). With squared-sine-bell windows of 60° for the t_2 and 90° for t_1 dimensions, followed by zero-filling and Fourier transformation, the spectra of 2048 × 512 data points were obtained.

Analysis of Relaxation Times. Peak heights of the cross-peaks were measured using FELIX 2.30. The standard deviations of the peak heights were estimated by the root mean square of the baseline noise. Five data points for T_1 or T_2 were fitted to a single exponential function, using a homemade FORTRAN77 program employing the nonlinear least-squares minimization algorithm of the Levenberg−Marquardt method (Levenberg, 1944; Marquardt, 1963) in an IMSL library (IMSL Inc.) on a VAX computer (DEC). The initial guess values were obtained by a linear least-squares minimization of the logarithms of the peak heights. To estimate the uncertainty of the relaxation times, a Monte Carlo simulation (Palmer et al., 1991) was done as follows. Randomizing the peak heights of the time points as Gaussian distributions, using an algorithm in the IMSL library

$$\chi^2 = \sum_{j=1}^{5} \frac{(A \exp(-t_j/T_i) - I_j)^2}{\sigma_j^2} \qquad (i = 1 \text{ or } 2) \qquad (1)$$

was minimized for 500 times, where I_j is the peak height at the time point t_j and σ_j is the standard deviation of the peak heights. T_1 and T_2 were obtained as the average values of the fitted values. The uncertainties of the relaxation times were estimated as the standard deviations. The $^1H-^{15}N$ NOE was determined as the cross-peak ratios of the spectra with and without NOE. The uncertainties of the NOEs were also estimated from the baseline noise as

$$\sigma_{NOE} = \frac{\sigma_{sat} I_{unsat} + \sigma_{unsat} I_{sat}}{I_{unsat}^2} \qquad (2)$$

where σ_{sat} and σ_{unsat} are the root mean square deviations of the baseline noises for the spectra with and without saturation, respectively, and I_{sat} and I_{unsat} are the heights of the peaks in the spectra with and without saturation, respectively.

Model-free analysis (Lipari & Szabo, 1982) was performed with FORTRAN77 programs using the nonlinear least-squares minimization algorithm on the VAX computer, basically by the strategy described by Kay et al. (1989), Clore et al. (1990), Palmer et al. (1991), and Powers et al. (1992).

T_1, T_2, and NOE are expressed using the spectral density function $J(\omega)$ as described by Abragam (1961)

$$1/T_1 = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N) \quad (3)$$

$$1/T_2 = 0.5d^2[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + (1/6)c^2[3J(\omega_N) + 4J(0)] \quad (4)$$

$$\text{NOE} = 1 + (\gamma_H/\gamma_N)d^2[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]T_1 \quad (5)$$

with the constants

$$d^2 = 0.1\gamma_H^2\gamma_N^2(h/2\pi)^2(1/r^3_{NH})^2 \quad (6)$$

$$c^2 = (2/15)\omega_N^2(\sigma_{||} - \sigma_\perp) \quad (7)$$

where ω_H and ω_N are the Larmor frequencies of ^1H ($2\pi \times$ 600.14 MHz) and ^{15}N ($2\pi \times$ 60.82 MHz), respectively, γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N (2.6753×10^4 and -2.71×10^3 rad·s^{-1}·G^{-1}, respectively), h is Planck's constant ($6.6260755 \times 10^{-27}$ erg·s), r_{NH} is the distance of the ^1H and ^{15}N atoms in the amide groups (1.02 Å), and $\sigma_{||}$ and $\sigma_\perp$ are the parallel and perpendicular components, respectively, of the axial symmetric ^{15}N chemical shift tensor. $(\sigma_{||} - \sigma_\perp)$ is assumed to be -160 ppm (Hiyama et al., 1988).

The rotational correlation time (τ_R) was calculated by fitting of the T_1/T_2 ratios of the rigid residues that show values within the standard deviation from the average. The uncertainty of τ_R was estimated by Monte Carlo simulations randomizing T_1 and T_2 as Gaussian distributions for 500 times, using the algorithm in the IMSL library. To obtain the model-free parameters, such as the generalized order parameters (S^2), the effective correlation time (τ_e), the order parameters of the faster and slower phases (S_f^2 and S_s^2), and the exchange contribution (Δex), five alternative models of spectral density functions (function models), which are listed in Table 1, were tested for the fitting. Function model I is a simplified spectral density function that describes an internal motion with a very short, unanalyzable correlation time as compared with τ_R (Lipari & Szabo, 1982). In this function model, T_1 and T_2 were used as the data to fit this function to obtain S^2. Function model II is the standard Lipari−Szabo function, which includes τ_e (Lipari & Szabo, 1982). T_1, T_2, and NOE were used to obtain S^2 and τ_e. Function model III was introduced by Clore et al. (1990) and contains two internal motions, a faster one and a slower one. The faster internal motion corresponds to the very fast one, like the motion in function model I. The slower one can be analyzed like the motion in function model II. T_1, T_2, and NOE were used to obtain S_f^2, S_s^2, and τ_e (Clore et al., 1990). The product of S_f^2 and S_s^2 corresponds to S^2. Function model IV is identical with function model I, except that it includes Δex, which shortens T_2. Likewise, function model V is identical with function model II, except that it includes Δex. For the function model I, T_1 and T_2 values were randomized as Gaussian distributions for 200 times. Using each T_1 and T_2 value

$$\chi_2^2 = \frac{(R_1^{calc} - R_1^{meas})^2}{\sigma_{R_1}^2} + \frac{(R_2^{calc} - R_2^{meas})^2}{\sigma_{R_2}^2} \quad (8)$$

was minimized, where R_i^{calc} and R_i^{meas} ($i = 1$ or 2) are the $1/T_i$ values obtained by calculation of the spectral density function and by experimental data, respectively, and σ_{R_i} is the standard deviation for R_i^{meas}. The average and the standard deviation were assigned to the fitted S^2 and its uncertainty, respectively. For function models II, III, IV, and V, the T_1, T_2, and NOE values were randomized as Gaussian distributions for 200 times. Using each value

$$\chi_3^2 = \frac{(R_1^{calc} - R_1^{meas})^2}{\sigma_{R_1}^2} + \frac{(R_2^{calc} - R_2^{meas})^2}{\sigma_{R_2}^2} + \frac{(\text{NOE}^{calc} - \text{NOE}^{meas})^2}{\sigma_{\text{NOE}}^2} \quad (9)$$

was minimized, where R_i^{calc}, R_i^{meas} ($i = 1$ or 2), and σ_{R_i} are the same as those for χ_2^2, NOE^{calc} and NOE^{meas} are the NOE values obtained by calculation of the spectral density functions and by experimental data, respectively, and σ_{NOE} is the standard deviation for the experimental NOE data. The average and the standard deviation of the model-free parameters were assigned to its values and uncertainties. To select the most appropriate function model, values of χ_3^2 for each function model resulting from the obtained model-free parameters and the experimental data were used. It is assumed that χ_3^2 distributes as a χ^2 distribution with three degrees of freedom, if an ideal function model is selected and the ratio of the two χ_3^2 values is an F distribution with degrees of freedom of 3 and 3, unless one of the functions is meaningfully preferable to the others (Hoel, 1971). The strategy is as follows. If the χ_3^2 for function model I (χ_{3I}^2) is less than 20% critical value of the χ^2 distribution (4.642), then function model I is selected. If not, function model II is then considered. If χ_3^2 (χ_{3II}^2) is less than 20% critical value, and if the ratio of χ_3^2 of function models I and II (χ_{3I}^2/χ_{3II}^2) is more than 20% critical value of the F distribution (2.936), then function model II is selected. If χ_{3II}^2 is less than 20% critical value, but if χ_{3I}^2/χ_{3II}^2 is less than 20% critical value, then function model I is selected. If neither χ^{3I2} or χ_{3II}^2 is less than 20% critical value, then χ_3^2 of function model III (χ_{3III}^2) is examined next. If χ_{3III}^2 is less than 20% critical value, and if both $\chi_{3I}^2/\chi_{3III}^2$ and $\chi_{3II}^2/\chi_{3III}^2$ are more than 20% critical value of the F distribution, then function model III is selected. If χ_{3III}^2 is less than 20% critical value, and if $\chi_{3I}^2/\chi_{3III}^2$ (or $\chi_{3II}^2/\chi_{3III}^2$, next) is less than 20% critical value, then function model I (or II, next) is selected. In this way, function models IV and V are considered in this order. If none of these function models gives χ_3^2 values less than 20% critical values, then 10% critical values of the χ^2 distribution (6.251) becomes the next standard for which the function models are considered from I to V. Likewise, the critical value is changed to 5% (7.815) and 1% (9.837) in this order. If all of the function models give χ_3^2 values more than 1% critical values, the function model that gives the smallest χ_3^2 value is selected.

Molecular Dynamics Simulation. The protein in solution was prepared by immersing RNase HI, the structure of which has been determined by X-ray crystallography (Katayanagi

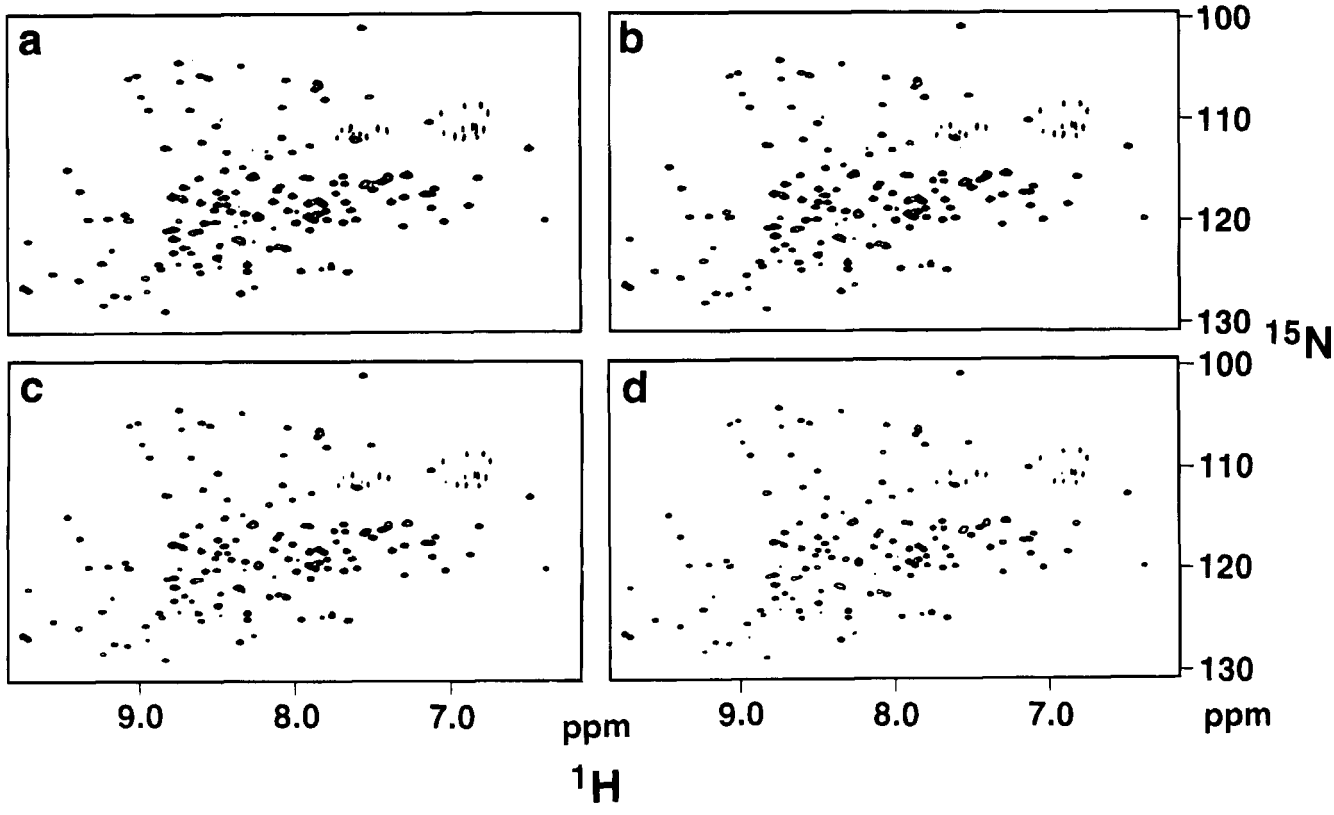

FIGURE 1: Two-dimensional NMR spectra for the measurement of ^{15}N T_1 with different delay times: (a) 100 ms, (b) 200 ms, (c) 400 ms, and (d) 600 ms.

et al., 1992; Protein Data Bank entry 2RN2), into a 34 Å water sphere with a density of 1.0 g/cm^3, where the water molecules overlapping the volume occupied by the protein were removed. The force-field parameters used in the present simulations were AMBER/OPLS (Weiner et al., 1984; Jorgensen & Tirado-Rives, 1988; Singh et al., 1986) for RNase HI and SPC (Berendsen et al., 1981) for the water molecules, respectively. The protein molecule, consisting of 1538 atoms, was surrounded by 4566 water molecules. Long-range Coulomb interactions were efficiently calculated using the PPPC method (Saito, 1992) without the truncation. The superiority of this method has been confirmed by performing MD simulations for human lysozyme (Saito, 1994).

The initial conformation of RNase HI in solution was equilibrated using the following protocol. First, the water molecules surrounding RNase HI were equilibrated by carrying out 10-ps MD simulations, where the heavy atoms of the protein and the crystal waters were restrained to their initial positions. The restraining harmonic force constant was 10 kcal mol^{-1} Å^{-2}. Next, the whole system was equilibrated by performing the simulations without any restraints for 420 ps. The time step was 0.5 fs. The temperature was kept at 300 K during all the simulations (Brown & Clarke, 1984). The translational momentum of the center of mass and the angular momentum around it were kept to zero during the simulations. Therefore, we see only the internal motions by the MD simulation.

All MD simulations were carried out using COSMOS90 (Saito, 1992) on a Fujitsu VP2600 at the Protein Engineering Research Institute.

After an equilibration of 100 ps, the N−H vectors in the trajectories of every 1 ps were extracted for analyses. Analyses of the distributions of the N−H vectors were done using programs written in FORTRAN77 on an IRIS indigo II computer (Silicon graphics).

Calculation of Accessible Surface Area (ASA). Hydrogen atoms were attached to the heavy atoms of the crystal structure of RNase HI (Katayanagi et al., 1992) derived from the Protein Data Bank (entry 2RN2) using the geometry of the amino acid residue in AMBER (Weiner et al., 1986). The energy of the protein was mildly minimized using the PRESTO software (Morikami et al., 1992) with position restraints from the original coordinates in a vacuum, to eliminate the small van der Waals overlap of the atoms. The RMSD from the starting structure was only 0.03 Å. The accessible surface areas were calculated by the method of Connolly (1983, 1993). Van der Waals radii for H, C, N, O, and S were settled to 1.0, 1.8, 1.7, 1.6, and 1.9 Å, respectively. A probe radius of 1.25 Å was employed since it was the optimal probe radius to detect cavity waters (Hubbard et al., 1994). Because the ASA values for the amide N atoms are almost zero for all residues, except the N-terminus, those of the amide H atoms were analyzed.

RESULTS AND DISCUSSION

T_1, T_2, and NOE. Figure 1 shows the 2D spectra for determining the ^{15}N T_1, with delays of 100 ms (a), 200 ms (b), 400 ms (c), and 600 ms (d). The assignments of the cross-peaks by Yamazaki et al. (1991, 1993) were easily adapted to these spectra. We could analyze 124 separated cross-peaks out of 149 backbone amides (the N-terminus and five Pro residues were excluded from the 155 residues). For measurements of T_2 and ^{1}H−^{15}N NOE, spectra with the same qualities were obtained (data not shown). It can be seen in the spectra that the intensity of each cross-peak decreases

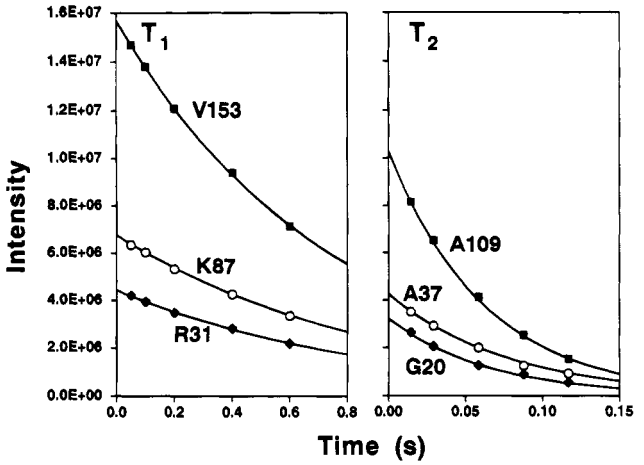

FIGURE 2: Cross-peak intensities of selected residues as functions of the delay times for the measurements of T_1 (left) or T_2 (right) with the single-exponential least-squares best-fit curves.

as the delay time increases (Figure 1). This dependence is shown in Figure 2 with the best-fit single-exponential curves for T_1 and T_2.

Figure 3 shows the T_1, T_2, and NOE values as functions of the residue number. The values and fitting errors are also listed in Table 1S in the supplementary material. Negative or very small NOE values were obtained for Leu2, Glu154, and Val155 in the N- and C-termini (Figure 3c). Small NOEs were also observed for the loop between β-strands A and B (Gly15), the turn between α-helix I and β-strand D (Glu61), the loop between α-helices III and IV (Ala93−Asn100), and the loop between β-strand E and α-helix V (Lys122−Glu129) (Figure 3c). Large T_2 values were observed in these regions, especially in the N- and C-termini, the loop between β-strands A and B (Cys13−Gly15), the loop between α-helices III and IV (Ala93−Asp94), and the loop between β-strand E and α-helix V (Gly123−His127) (Figure 3b). Relatively small T_1 values, however, were obtained for the amino acid residues in these regions (Figure 3a). The internal motions are expected to be large in all the regions listed above.

Model-Free Analysis. From the measured T_1, T_2, and NOE, parameters for the description of motions were calculated by the model-free analysis (Lipari & Szabo, 1982). First, the overall rotational correlation time (τ_R) was calculated using the T_1/T_2 ratio by the method described by Kay et al. (1989). The average and the standard deviation (SD) of the T_1/T_2 ratio are 12.2 ± 2.4. The T_1/T_2 ratios of 96 residues are within ±1 SD from this value. The calculated τ_R was 10.92 ± 0.03 ns at 27 °C, which is comparable to 10.4 ns for the 138 residues of the RNase H domain of HIV-1 reverse transcriptase measured at 26 °C (Powers et al., 1992). If the T_1/T_2 ratios of all residues are used for the calculation, the τ_R becomes 10.75 ± 0.02 ns. We used the value of 10.92 for the τ_R in the present study.

Appropriate models of spectral density functions (function models) (Table 1) were selected for the model-free analysis, according to the definition described in the Materials and Methods section. Consequently, the simplified spectral density function (function model I) (Lipari & Szabo, 1982) could be applied for 47 residues (class I residues), Val5 (βA), Asp10 (βA), Gly11 (βA), Gly18 (c), Gly20 (βB), Gly21 (βB), Tyr22 (βB), Gly23 (βB), Ala24 (βB), Ile25 (βB), Arg27 (βB), Lys33 (βC), Thr34 (βC), Tyr39 (βC), Arg41 (c), Thr42 (c), Thr43 (c), Asn44 (αI), Met47 (αI), Glu48 (αI), Ile53 (αI), Ala58 (αI), Val65 (βD), Ser68 (βD), Thr69 (c), Tyr73 (αII), Val74 (αII), Arg75 (αII), Gly77 (αII), Thr79 (αII), Lys86 (αIII), Gln105 (αIV), Gly112 (c), Glu119 (βE), Trp120 (βE), Glu129 (αV), Ala131 (αV), Arg132 (αV), Cys133 (αV), Asp134 (αV), Leu136 (αV), Ala139 (αV), Ala140 (αV), Ala141 (αV), Met142 (αV), Asp148 (c), and Tyr151 (c). The secondary structure units determined in solution (Yamazaki et al., 1991) are indicated in parentheses (α, β, and c represent α-helix, β-strand, and coil, respectively). The standard Lipari−Szabo function that includes τ_e (function model II) (Lipari & Szabo, 1982) could be applied for 19 residues (class II residues), Leu2 (c), Phe8 (βA), Leu26 (βB), Ile66 (βD), Asp70 (c), Trp81 (c), His83 (αIII), Asn84 (αIII), Lys87 (αIII), Gly89 (αIII), Val98 (c), Lys99 (c), Arg106 (αIV), Ile116 (βE), Val121 (βE), Lys122 (βE), Asn130 (αV), Asn143 (c), and Gly150 (c). Function model III, containing two internal motions (Clore et al., 1990), could be applied for 37 residues (class III residues), Lys3 (c), Glu6 (βA), Thr9 (βA), Cys13 (βA), Leu14 (c), Gly15 (c), Tyr28 (βB), Arg29 (c), Gly30 (c), Arg31 (βC), Phe35 (βC), Ser36 (βC), Ala37 (βC), Gly38 (βC), Glu61 (c), His62 (c), Ser71 (c), Gln72 (αII), Thr92 (c), Ala93 (c), Asp94 (c), Lys95 (c), Lys96 (c), Gln115 (βE), Lys117 (βE), Gly123 (c), Ala125 (c), Gly126 (c), His127 (c), Thr145 (c), Leu146 (c), Glu147 (c), Thr149 (c), Gln152 (c), Val153 (c), Glu154 (c), and Val155 (c). A simplified function including the exchange contribution (Δex) (function model IV) could be applied for 20 residues (class IV residues), Met50 (αI), Ala51 (αI), Ala52 (αI), Ala55 (αI), Leu56 (αI), Leu59 (c), Ile78 (αII), Ile82 (αIII), Lys91 (c), Asn100 (c), Val101 (αIV), Asp102 (αIV), Asp103 (αIV), Trp104 (αIV), Leu107 (αIV), Ala109 (αIV), Ala110 (αIV), Leu111 (c), Trp118 (βE), and Glu135 (αV). The standard Lipari−Szabo function including Δex (function model V) could be applied for only

Table 1: Spectral Density Function Models Used for Fitting

	spectral density function	data used	parameters
model I	$J(\omega) = S^2\tau_R/(1 + \omega^2\tau_R^2)$	T_1, T_2	S^2
model II	$J(\omega) = S^2\tau_R/(1 + \omega^2\tau_R^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2)$ $\tau = \tau_R\tau_e/(\tau_R + \tau_e)$	T_1, T_2, NOE	S^2, τ_e
model III	$J(\omega) = S_f^2S_s^2\tau_R/(1 + \omega^2\tau_R^2) + S_f^2(1 - S_s^2)\tau/(1 + \omega^2\tau^2)$ $\tau = \tau_R\tau_e/(\tau_R + \tau_e)$	T_1, T_2, NOE	S_f^2, S_s^2, τ_e
model IV	$J(\omega) = S^2\tau_R/(1 + \omega^2\tau_R^2)$ $1/T_2(\text{obs}) = 1/T_2 + \pi\Delta\text{ex}$	T_1, T_2	S^2, Δex
model V	$J(\omega) = S^2\tau_R/(1 + \omega^2\tau_R^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2)$ $\tau = \tau_R\tau_e/(\tau_R + \tau_e)$ $1/T_2(\text{obs}) = 1/T_2 + \pi\Delta\text{ex}$	T_1, T_2, NOE	S^2, τ_e, Δex

Table 2: Model-Free Parameters and ASA

	function models					
	I	II	III	IV	V	total
no. of residues	47	19	37	20	1	124
α-helix	22	7	1	15	0	45
β-sheet	17	6	11	1	0	35
coil	8	6	25	4	1	44
S^2	0.93	0.85	0.72	0.89	0.94	0.85
τ_e (ns)		0.33	1.05		0.23	0.80
S_f^2			0.88			0.88
S_s^2			0.81			0.81
Δex (Hz)				0.71	2.61	0.80
ASA (Å^2)	0.31	1.20	1.46	0.58	0.00	0.83

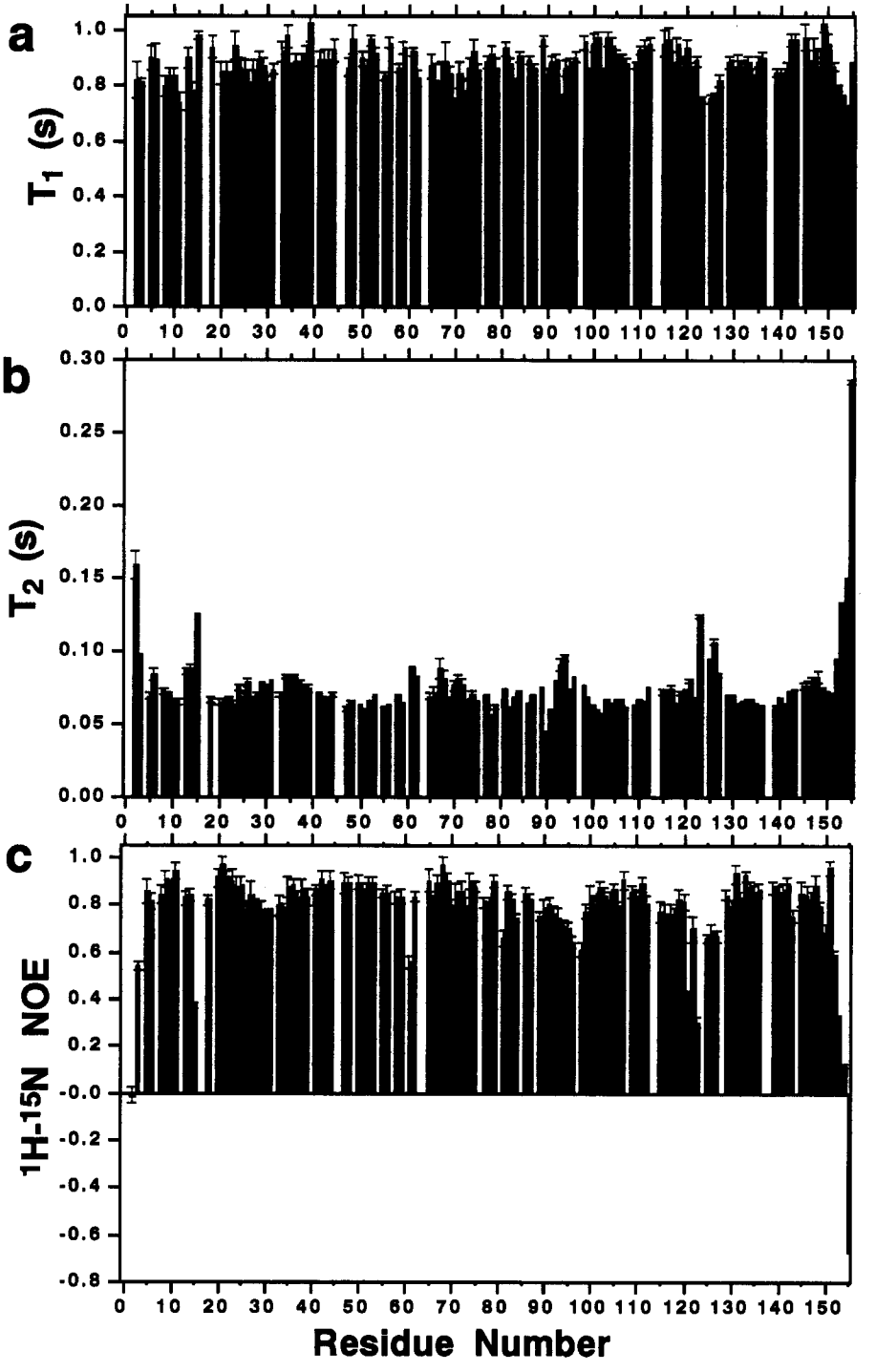

FIGURE 3: T_1 (a), T_2 (b), and NOE (c) as functions of residue numbers.

one residue (class V residue), Trp90 (c). The selected function models, the model-free parameters, and the χ_3^2 values are listed in Table 1S in the supplementary material.

The average value of each parameter describing each function model, as well as the accessible surface area (ASA) of the amide H atoms, is listed in Table 2. The residues in α-helices are mainly included in class I or IV, which suggests that helical residues tend to show only very fast motion, below the picosecond time scale. Almost all of the residues in β-sheets are included in classes I, II, and III, which means that they do not show very slow motion in the millisecond time scale; only one exception (Trp118) gives a high χ^2 value (Table 1S). More than half of the residues in loops are included in the class III residues. These residues tend to show the two-phased motion. The class I and IV residues tend to have relatively large S^2 and small ASA values. In contrast, the class II and III residues tend to have relatively small S^2 and large ASA values. Thus, upon these classifications, the correlation coefficient between S^2 and the ASA of each class became as high as −0.91. It seems likely that the motions of the residues in α-helices are highly restricted so that they have large S^2 values in the picosecond to nanosecond time scale. Furthermore, because the ASA values of the amide protons in helices tend to be small (most of them are zero), the negative correlation between S^2 and the ASA was observed. Actually, 45 residues in the α-helices have average S^2 and ASA values of 0.92 and 0.25, respectively. Thirty-five residues in the β-sheet have S^2 and

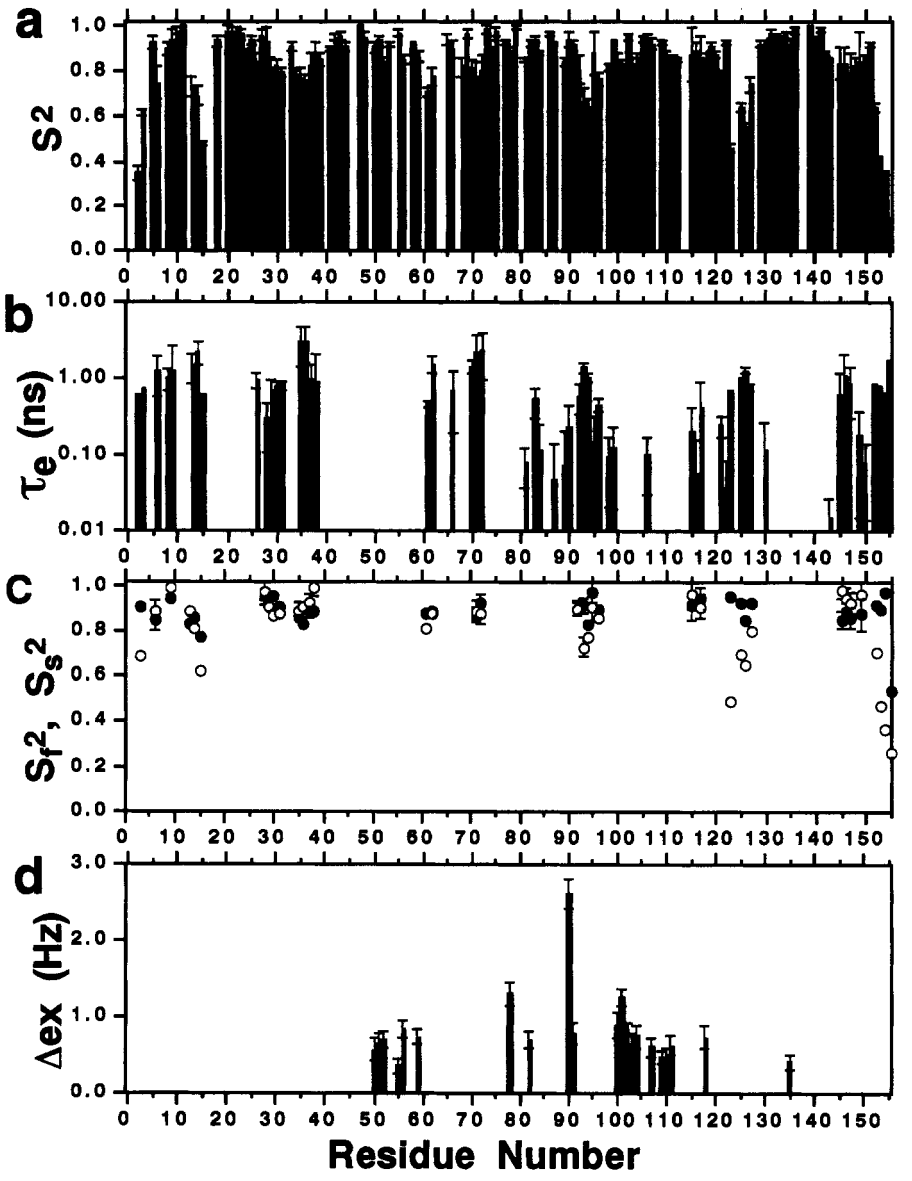

FIGURE 4: Model-free parameters as functions of residue numbers: (a) generalized order parameters, (b) efficient correlation times for internal motions, (c) order parameters of the faster and slower phases (● and ○, respectively), and (d) exchange contributions.

ASA values of 0.88 and 0.87, respectively. Forty-four residues in loops have S^2 and ASA values of 0.75 and 1.40, respectively. Therefore, the negative correlation between S^2 and the ASA upon classification may be explained through the average values in each secondary structure type. The amide H atoms with large ASA values tend to be located at the surface of the protein so that steric hindrances are weakened and water molecules collide frequently. However, the direct correlation between S^2 and the ASA was much weaker (−0.26), when the residues were not classified. The correlation may be masked by large deviations of S^2 and ASA within the classes or secondary structure types.

It should be noted, however, that this classification is not absolute. There are several residues on the border of the classification, which show small τ_e (less than 50 ps) or large S_f^2 or S_s^2 (more than 0.95) (Figure 4). Therefore, for each residue, the values of the parameters, rather than the classification, should be considered for the description of the internal motions.

Implications in Structures and Functions. Figure 4 shows the generalized order parameters (S^2) for all five function models (a), the effective correlation times of the internal motion (τ_e) for the class II, III, and V residues (b), the order parameters of the fast and small phases (S_f^2 and S_s^2) for the class III residues (c), and the exchange contribution (Δex) for the class IV and V residues (d). Large internal motions were observed in the N-terminal region (Leu2−Lys3), the loop between β-strands A and B (Cys13−Gly15), the turn between α-helix I and β-strand D (Glu61, His62), the loop between β-strand D and α-helix II (Asp70−Tyr71), the loop between α-helices III and IV (Ala93−Lys96), the loop between β-strand E and α-helix V (Gly123−His127), and the C-terminal region (Gln152−Val155). The results are consistent with those expected using the T_1, T_2, and NOE values (Figure 3). The locations of these regions in the tertiary structure are shown in Figure 5. The effective correlation time observed in these regions varied from 0.45 ns (Glu61, Lys96) to 2.2 ns (Leu14) (Figure 4b). S_f^2 tends to be larger than S_s^2, and it shows less deviation (Figure 4c). Very fast motion might be restricted in its orientation, regardless of the environment of each residue.

Exchange contributions (Δex) were predominantly observed in α-helices I (Asn44−Ala58) and IV (Val101−Ala110), including their C-terminal cap residues (Leu59 and

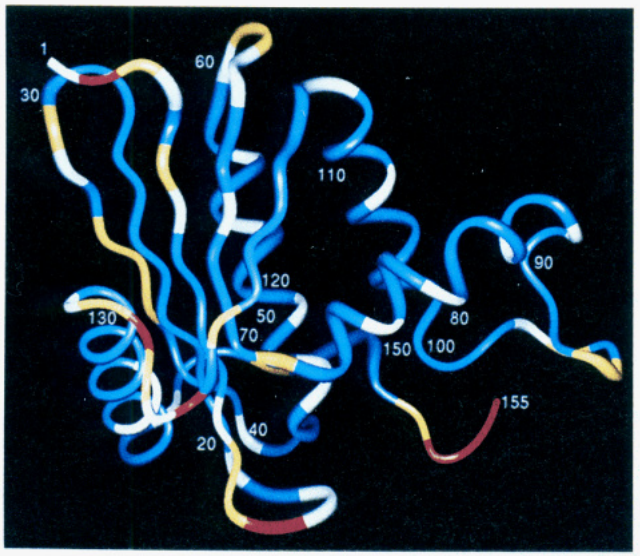

FIGURE 5: Generalized order parameters (S^2) on the tertiary structure of *E. coli* RNase HI. The residues colored red, yellow, and blue have S^2 values of less than 0.6, between 0.6 and 0.8, and more than 0.8, respectively. The residues colored white could not be analyzed by NMR. The crystal structure (Katayanagi et al., 1992) was obtained from the Protein Data Bank (entry 2RN2). The molecular display was done by Insight II (Biosym) on a Silicon Graphics IRIS-4D/25TG computer. Residues colored red are Leu2, Gly15, Gly123, Gly126, and Val153−Val155. Those colored yellow are Lys3, Glu6, Cys13, Leu14, Arg31, Thr34−Ser36, Glu61, His62, Ser71, Ala93, Asp94, Lys96, Val121, Ala125, His127, Glu147, and Gln152. The colored blue are Val5, Phe8−Gly11, Gly18, Gly20−Gly30, Lys33, Ala37−Tyr39, Arg41−Asn44, Met47, Glu48, Met50−Ile53, Ala55, Leu56, Ala58, Leu59, Val65, Ile66, Ser68−Asp70, Gln72−Arg75, Gly77−Thr79, Trp81−Asn84, Lys86, Lys87, Gly89−Thr92, Lys95, Val98−Leu107, Ala109−Gly112, Gln115−Trp120, Lys122, Glu129−Leu136, Ala139−Asn143, Thr145, Leu146, and Asp148−Tyr151.

Leu111), which are adjacent to each other and form a hydrophobic core (Katayanagi et al., 1990). These two helices may move cooperatively. It should be noted that the Δex values of the residues in these helices are not necessarily the same, because Δex depends on the deviation of the microenvironment (chemical shift) caused by the movement. In addition, these two helices are stable in the acid-denatured structure and are also formed in the early refolding intermediate(s) (Yamasaki et al., unpublished results).

For the C-terminal region, a series of deletion mutants have been constructed by Haruki et al. (1994). The protein stability has not been influenced when the C-terminal three residues (Val155, Glu154, and Val153) were deleted. However, the additional deletion of Gln152 results in a decrease in the melting temperature by 3 °C. In the present study, the S^2 values of Tyr151, Gln152, Val153, Glu154, and Val155 were 0.91, 0.64, 0.42, 0.36, and 0.14, respectively (Figures 4a and 5). It is possible that the elimination of all of the mobile residues at the C-terminus results in a perturbation in the conformation of Tyr151 and thereby results in a reduction in the thermal stability. It is also possible that Gln152 forms a transient noncovalent bond with the rest of the protein, which contributes to the stability (Sharp & Englander, 1994).

Mutational studies have suggested that the active site of this enzyme consists of three acidic residues (Asp10, Glu48, and Asp70) (Kanaya et al., 1990). In the present study, the order parameters of these residues were 0.98, 0.95, and 0.81 for Asp10, Glu48, and Asp70, respectively (Figure 4a). The observation of the relatively large mobility in Asp70, with a correlation time of 1.4 ns, may support the previous proposal that this residue is directly involved in the catalytic reaction (Oda et al., 1993), because it should have a flexible conformation. The other two backbone amides are less mobile than that of Asp70, and therefore, the conformations of these residues may not change in the catalytic reaction.

The region between α-helices III and IV is designated as the basic protrusion (Katayanagi et al., 1992) or the handle region (Yang et al., 1990). This region consists of a positively charged cluster and has been suggested to be involved in the substrate binding (Kanaya et al., 1991). In the present study, relatively large mobility was observed in a part of this region (Ala93−Lys96) (Figures 4a and 5). The S^2 values for Ala93, Asp94, Lys95, and Lys96 were 0.66, 0.64, 0.88, and 0.77, respectively. It is possible that these residues change their conformation upon substrate binding (induced fit). This should be tested in the future by an experiment with the complex of the enzyme and the substrate (analog). Although the S^2 value of Lys95 is high, this does not necessarily imply that this residue is fixed. The MD analysis suggested that the internal motion of this loop has less influence on the N−H vector orientations than the positions, as described later. The mobility in this region has not been clearly observed by an isotropic thermal factor (*B*-factor) in the X-ray crystal analysis, due to the crystal packing (Katayanagi et al., 1992) (Figure 7b).

In the loop from Cys13 to Gly15 and the loop from Gly123 to His127, very large internal motions were observed (Figures 4a and 5). The S^2 values for Cys13, Leu14, Gly15, Gly123, Ala125, Gly126, and His127 were 0.73, 0.69, 0.48, 0.46, 0.64, 0.55, and 0.74, respectively. These two loops are located relatively close to the active site of the enzyme and contribute to form a cleft-like depression, in which the substrate should bind (Katayanagi et al., 1992) (Figure 5). Therefore, it seems likely that, like a part of the basic protrusion, these two loops change their conformations upon the substrate binding, with a large induced fit. In addition, the loop from Gly123 to His127 contains His124, which has been previously shown to be involved in the catalytic function by Oda et al. (1993). They have proposed that the conformation of the loop with His124 is flexible, and therefore His124 changes its conformation to enhance the catalytic efficiency by removing a proton from a catalytically essential carboxylate as a proton pump. The observation of the large internal motion in this loop supports this proposal.

The backbone dynamics of a related protein, the RNase H domain of HIV-1 reverse transcriptase, have been characterized by Powers et al. (1992). The function models I and IV have been applied for only 29 residues, less than half of those for *E. coli* RNase HI. In contrast, function models II, III, and V have been applied for 75 residues, more than those for *E. coli* RNase HI. The average S^2 value is 0.78, which is less than that of *E. coli* RNase HI by 0.07. Although the magnitude of order parameter depends strongly on τ_R (Jones et al., 1994), the difference is significant; a trial perturbation of τ_R by 0.09 ns for *E. coli* RNase HI changed the average of S^2 by only 0.002; the τ_R values for *E. coli* RNase HI and the RNase H domain of HIV-1 reverse transcriptase were 10.92 ± 0.03 and 10.4 ± 0.08 ns, respectively. Therefore, the RNase H domain of HIV-1 reverse transcriptase is much more mobile than *E. coli* RNase

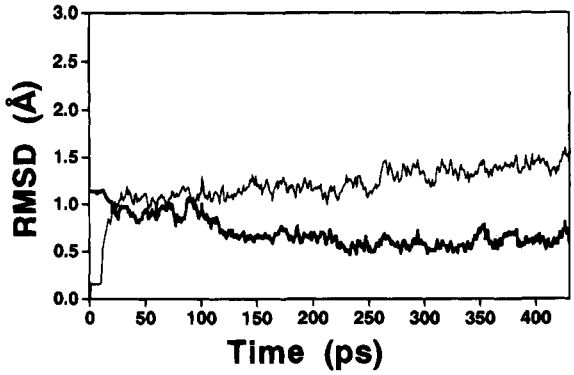

FIGURE 6: Root mean square deviations (RMSD) during MD simulation: deviation from the average structure of the last 330 ps of the MD simulation (thick line) or from the initial crystal structure (thin line).

HI, with slow motions within an analyzable time scale. This may explain, in part, the fact that the RNase H domain alone has no enzymatic activity, because the equilibrium in the conformation of a protein in solution between the active and inactive states may change unfavorably as the mobility in the protein molecule increases.

Molecular Dynamics Simulation. Figure 6 shows the root mean square deviation (RMSD) value from the initial crystal structure (Katayanagi et al., 1992) (thin line) and from the averaged structure during the last 330 ps (thick line). The results show that there is no large conformational change after the initial 100 ps. Therefore, we excluded the initial 100 ps as an equilibration time and used the data obtained at the last 330 ps (actually 300 ps for simplification) for the following procedures.

Comparison of Internal Motions Obtained by NMR, MD, and Crystallography. The generalized order parameter (S^2) for MD was calculated according to Lipari and Szabo (1982)

$$S^2 = \sum_{i \neq j}^{N} P_2(\mu_i \mu_j)/(N(N-1)/2) \qquad (10)$$

where $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial, μ_i and μ_j are the N−H unit vectors extracted from the molecular dynamics trajectory for every 1 ps, and N is the total number of the extracted N−H vectors (300).

Figure 7a shows the order parameters obtained by MD, together with those by NMR, as a function of residue number. For most of the residues, a simulation time of 300 ps was sufficient, judging from the time dependence of the order parameter (data not shown). Apparently, the residue profiles of the two order parameters are similar, overall. The values of the basic levels of $(1 - S^2)$ were about 0.1 for the two parameters. The MD also indicates that there are large internal motions in the N-terminal region (Leu2−Lys3), the loop between β-strands A and B (Cys13−Gly18), the turn between α-helix I and β-strand D (Glu61, His62), the loop between α-helices III and IV (Ala93−Lys96), the loop between β-strand E and α-helix V (Gly123−His127), and the C-terminal region (Gln152−Val155). The coincidence encouraged us to employ MD for the description of the motions observed by NMR. However, large differences between the order parameters obtained from NMR and MD were observed for several residues. For these residues, such as Gly123 and Val155, the NMR showed higher mobility than the MD. In these cases, the simulation time in MD might not be enough to characterize the motions of these residues. On the other hand, for other residues, such as Gly30 and Thr43, the MD showed higher mobility than the NMR. Although the simulation time of MD is limited, it could collect data revealing a rare, large conformational change. In that case, MD shows more mobility than NMR, which deals with a population of molecules as a whole. These problems may be solved when the MD simulation is performed longer or many times with different starting structures. It is also possible, however, that the discrepancies between NMR and MD are due to the force-field imperfection.

The order parameter describes the deviation of the orientation of the N−H vectors but not of the positions of the N or H atoms. Figure 7b shows the root mean square fluctuation (RMSF) of the N atoms during a MD simulation and the fluctuation based on the X-ray isotropic thermal factors (*B*-factors) (Katayanagi et al., 1992) of the N atoms. These two parameters reflect deviations in the position. Although the time scales of the two parameters were quite different, the pattern of their residue profiles agreed approximately (Figure 7b). However, the fluctuations from the *B*-factor tend to have much larger values than those from MD (Figure 7b), which has been explained as the contributions of the translational and rotational motions of the protein and the errors induced by the refinement process, which assumes isotropic and harmonic motion (Brooks et al., 1988). Large differences were observed in the basic protrusion (Asp92−Lys96) and loop between β-strand E and α-helix V (Gly123−His127), where the MD fluctuation is larger than that from the *B*-factor or is so large that it has a similar value (Figure 7b). It is probably due to the crystal packing, since these regions are involved in the interaction with the adjacent molecule in the crystal lattice (Katayanagi et al., 1992), which is absent in the present MD simulation.

Differences were observed between the order parameter and the fluctuation by MD, especially in the basic protrusion (Asp92−Lys95) (Figure 7a,b), where the positions of the N and H atoms move largely, but in parallel, with relatively little change of the N−H vector orientations.

Correlation Function by MD. Figure 8 shows the correlation function calculated by MD, as introduced by Chandrasekher et al. (1992)

$$\langle P_2(\mu(0)\mu(t))\rangle = (1/N)\sum_{i=1}^{N} P_2(\mu(\tau_i)\mu(t+\tau_i)) \qquad (11)$$

where $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial, $\mu(t)$ is the N−H unit vector at time t, and N is the number of data points used for the averaging. For each time point, t, τ_i was varied from 0 to 99 ps ($N = 100$ points). Apparently, the functions can be divided in two types: One has only a very fast motion within 1 ps (burst phase) and is flat after that, such as the function for Ile53 (Figure 8). The other has a burst phase and a following decay of correlation, such as the functions for Lys96, His127, and Val155 in Figure 8. The former corresponds to the function models I and IV, which were used for NMR analysis. The latter corresponds to the function model III. The burst phases were observed for all residues and tended to have very similar values (about 0.1). This is consistent with the present NMR results showing a small deviation in S_f^2 and an average value of 0.88 (Figure 4c, Table 2). The burst phase has also been

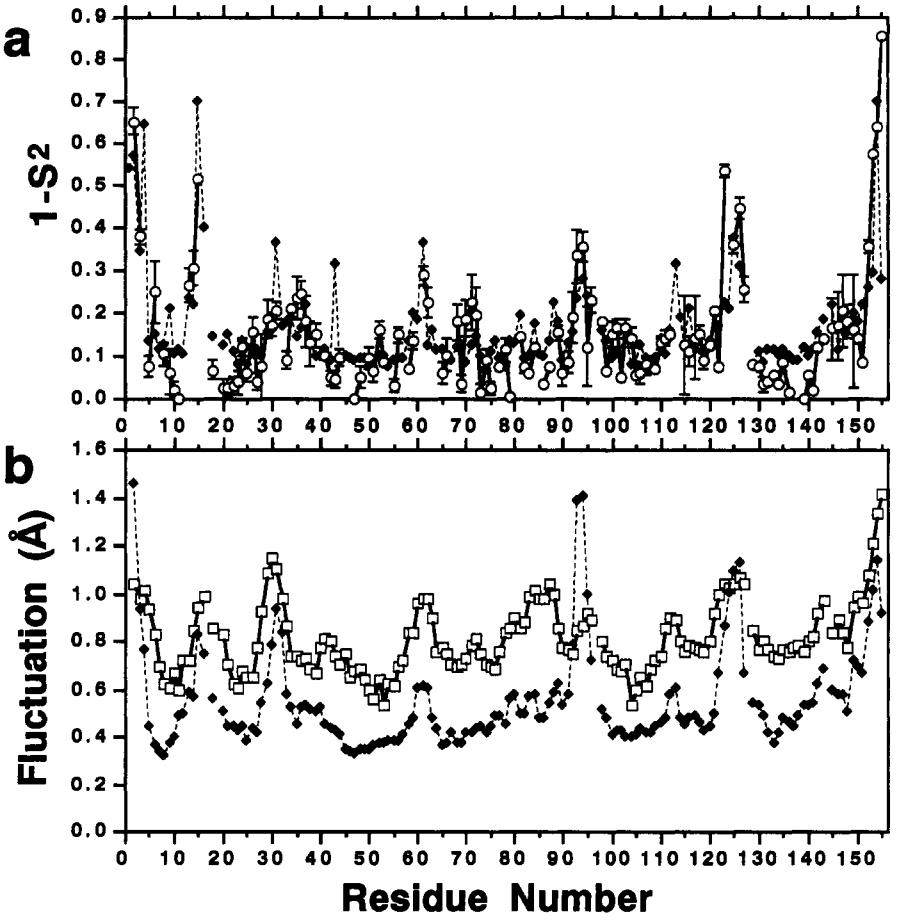

FIGURE 7: Internal motions of backbone amides obtained by NMR, MD, and X-ray crystallography. (a) Order parameters were obtained by NMR (○ with solid line) and MD (◆ with broken line). (b) Root mean square fluctuation (RMSF) of the N atoms during MD simulation (◆ with broken line) and fluctuation were based on the isotropic thermal factors (*B*-factors) of the N atoms in the X-ray crystal analysis (□ with solid line), calculated as $(3B/8\pi^2)^{1/2}$ (Brooks et al., 1988). The *B*-factors were obtained from the Protein Data Bank, entry 2RN2 (Katayanagi et al., 1992).

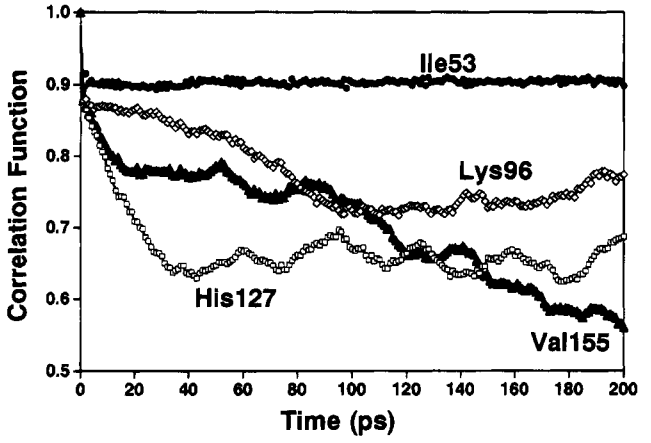

FIGURE 8: Correlation functions of the N−H vectors of selected residues during the MD simulation.

observed for all residues in interleukin-1β (Chandrasekhar et al., 1992) and RNase T1 (Fushman et al., 1994b). Because the function models II and V for the model-free analysis (Table 1) do not assume this unanalyzable fast motion, below the picosecond time scale, these models of the dynamics do not seem to be applicable for MD simulation. For NMR, the function models II and V were applied for 20 residues (Table 2). However, these residues showed relatively short τ_e values (Table 2), which possibly masks the existence of the burst phase.

Exponential decays after the burst phase were observed for some residues. For most of them, however, the decay curves were too complex to be fitted to a few exponential functions. Furthermore, some residues, such as His127, show oscillation. The complicated decay may turn into more simple exponential decays when a longer simulation time and many more samples are used to average. In contrast, the oscillation will not vanish by an averaging of more points if there is no random perturbation in the frequency. Because the Fourier transformation of oscillation results in a Δ function, this motion cannot be observed by NMR unless the frequency matches either ω_H (Larmor frequency of ^{1}H), ω_N (Larmor frequency of ^{15}N), or $\omega_H \pm \omega_N$.

Apparently, the duration of the MD simulation is not sufficient to describe the motions for several residues with correlation times greater than a nanosecond. Actually, for Val155, the correlation function does not reach its plateau level (Figure 8), which results in the smaller motion by MD than by NMR, as described above.

Chandrasekhar et al. (1992) have calculated S^2 as the plateau level ($t = \infty$) of the correlation function. Since the simulation time of MD is limited, they have assigned the average value of the function to S^2. This method is applicable to the residues with a correlation function that rapidly reaches to its plateau level, such as Ile53 in Figure 8. However, for the residues showing exponential decay, such as Lys96, it emphasizes the values of the early times

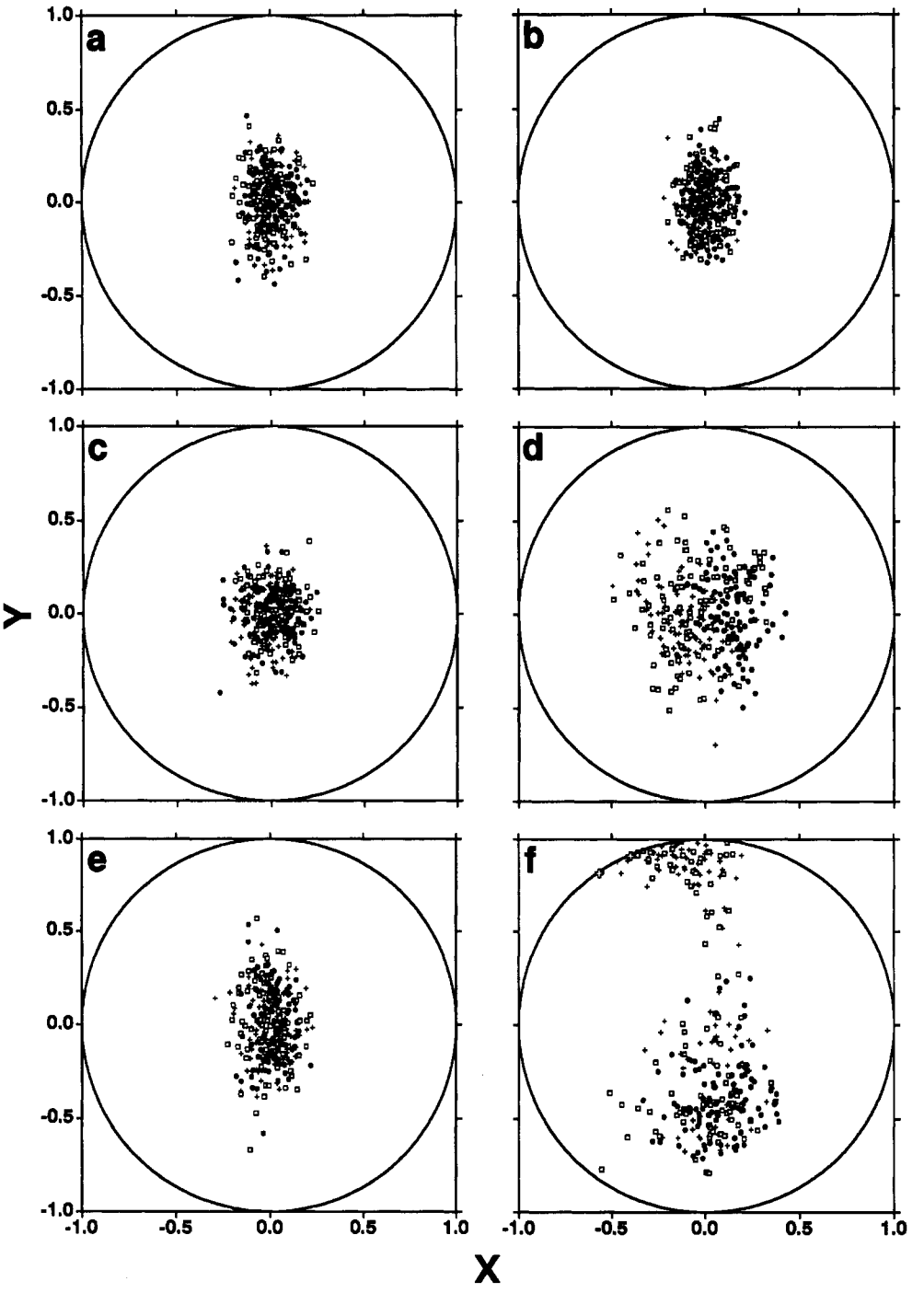

FIGURE 9: Distribution of the N−H unit vectors extracted for every 1 ps of 300 ps of the MD simulation. The origin represents the position of the N atom. The z-axis is the average of the N−H vectors, which is perpendicular to the paper plane and is pointing toward this side. The x-axis is the projection of the average of the N−C_α vectors to the paper plane. Panels: (a) Ile53, (b) Glu131, (c) Asp70, (d) Lys96, (e) Ala58, and (f) Gly15. ○, □, and + represent the N−H vectors in simulation times of 1−100, 101−200, and 201−300 ps, respectively. Open crosses in (f) show the N−H vectors pointing toward the opposite side of the paper and in a simulation time of 201−300 ps.

of the correlation and makes S^2 higher than the real value. Alternatively, it is possible to assign the minimum value during the duration to S^2. However, in this way, the oscillations observed for residues such as His127 make S^2 lower than the real value. In the present study, S^2 was calculated without considering the time dimension, which is relatively free from artifacts.

Distributions of N−H Vectors Observed by the Molecular Dynamics Simulation. Figure 9 shows the distribution of the N−H vectors of several residues during 300 ps of the MD calculation. The z-axis corresponds to the average of the orientations of the N−H vectors. The x-axis is defined as the projection of the average N−C_α vectors to the plane perpendicular to the z-axis. For most of the residues, this presentation makes an ellipse (Figure 9a,b). Two examples (Ile53 and Glu131) are shown in Figure 9a,b. The "ellipticity" of the distribution is defined as

$$\max\left[\sum_{i=1}^{N}\sin^{-1}|\mu_i\mu_\phi|\right]/\min\left[\sum_{i=1}^{N}\sin^{-1}|\mu_i\mu_\phi|\right] \qquad (12)$$

where μ_i is each unit N−H vector ($N = 300$) and μ_ϕ is the unit vector in the x−y plane with the ϕ angle in polar coordinates ($0 \leq \phi \leq \pi$). The ϕ angles that give the

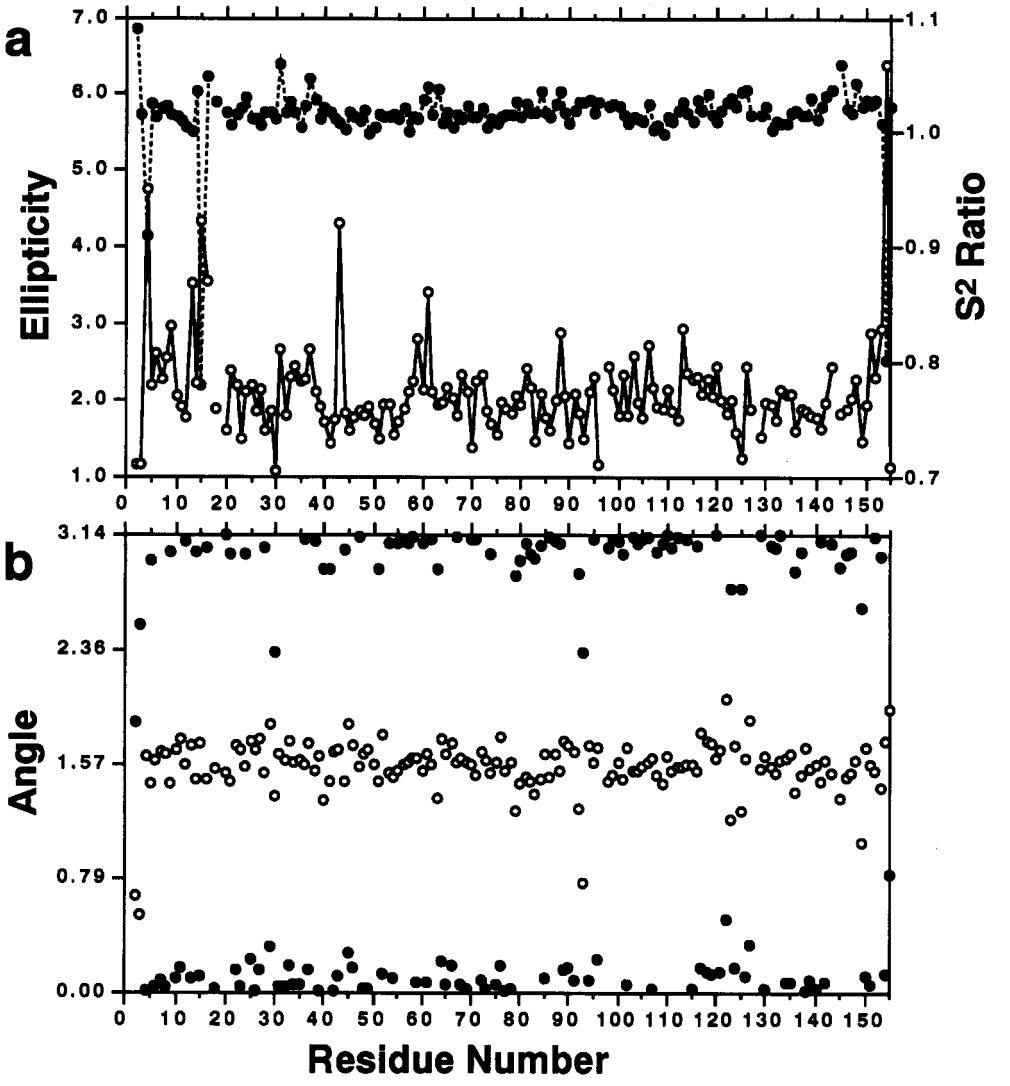

FIGURE 10: Ellipticities of distributions (a, ○), ratios of the order parameters calculated with and without the diffusion-in-an-elliptic-cone model (a, ●) (see text), and angles of the long and short axes (b, ○ and ●, respectively) relative to the x-axis in Figure 9 (the projection of the N−C_α vector).

maximum and the minimum of $\sum_{i=1}^{N}\sin^{-1}|\mu_i\mu_\phi|$) were defined as the long and short axes, respectively.

For Ile53 and Glu131, the ellipticities were calculated as 1.93 and 1.94, respectively. Figure 10 a (open circles) shows the ellipticity as a function of residue number. Low ellipticity corresponds to a typical cone. Very high ellipticity, as discussed below, corresponds to shapes other than a standard ellipse. The long axis of the ellipse tends to be perpendicular to the plane defined by the N−H vector and the N−C_α vector (Figure 10b, open circles). The short axis tends to be within the plane (Figure 10b, closed circles). The result indicates the tendency of the movement along the N−CO or the N−C_α bond.

These observations are largely consistent with a recent MD work on RNase T1 by Fushman et al. (1994b). They have reported that all the NH vectors experience a librational motion in the subpicosecond time scale and that the diffusion-in-a-cone model (Kinoshita et al., 1977; Lipari & Szabo, 1980) is the proper model to mimic the distribution of the vectors. This type of motion corresponds to the major one in the present study. They have also observed that this librational motion has angular amplitudes of 10−12° for out-of-peptide-plane displacements of the N−H bond and 3−5° for in-plane displacements (Fushman et al., 1994b), which is also consistent with the elliptic distribution in the present study.

Thus, the diffusion-in-a-cone model (Kinoshita et al., 1977; Lipari & Szabo, 1980) is approximately applicable. However, it seems to be better to modify it as a diffusion-in-an-elliptic-cone, where the limit of the θ angle depends on the ϕ angle as an ellipse. The description of the plateau distribution ($t = \infty$) of this model needs only two θ limits, for the long and short axes. To examine the adequacy of this model, we compared the generalized order parameters obtained by MD, as described above, with those obtained by assuming the model as follows. The generalized order parameter (Lipari & Szabo, 1982) can be calculated as

$$S^2 = \int_0^{\theta_B(\phi_1)} d\theta_1 \int_0^{2\pi} d\phi_1 \int_0^{\theta_B(\phi_2)} d\theta_2 \int_0^{2\pi} d\phi_2 \, P_2(\cos\theta_{12})\, p(\theta_1, \phi_1)\, p(\theta_2, \phi_2) \quad (13)$$

where $P_2(x) = (3x^2 - 1)/2$, θ_{12} is the angle between $[\theta_1, \phi_1]$ and $[\theta_2, \phi_2]$, and $p(\theta_i, \phi_i)$ represents the normalized probability of $[\theta_i, \phi_i]$ ($i = 1$ or 2). The dependence of the

boundary of θ ($=\theta_B$) on ϕ is given by

$$\theta_B(\phi) = 1/\sqrt{\cos^2 \phi/{\theta_S}^2 + \sin^2 \phi/{\theta_L}^2} \quad (14)$$

where θ_L is the long axis limit and θ_S is the short axis limit. Without a loss of generality, the short axis was set to the x-axis. Instead of solving this equation, we simulated the integral as

$$S^2 = \frac{\sum\sum\sum\sum P_2(\cos \theta_{12}) \sin \theta_1 \sin \theta_2}{\sum\sum\sum\sum \sin \theta_1 \sin \theta_2} \quad (15)$$

where the summation was done for θ_1, θ_2, ϕ_1, and ϕ_2, while gradually changing the values (by 1° for θ_i and 5° for ϕ_i) for the ranges in the equation. The θ_L and θ_S values were obtained from the average values of $\sin^{-1}|\mu_i\mu_x|$ along the axes by fitting the equation

$$\langle \sin^{-1}|\mu_i\mu_x| \rangle = \frac{\sum\sum \sin^{-1}|\sin \theta \cos \phi| \sin \theta}{\sum\sum \sin \theta} \quad (16)$$

where μ_x is the unit vector of the x-axis, and the summation was done for θ and ϕ, while gradually changing the values (by 1° for θ and 5° for ϕ) over the ranges [0, θ_L (or θ_S)] for θ and [0, $\pi/2$] for ϕ.

Thus, S^2, with an assumption of the diffusion-in-an-elliptic-cone model, was calculated and is shown in Figure 10a as the ratios to the S^2 by 300 ps of MD simulation, without assuming any model (same as Figure 7a). This shows that the distribution of most residues can be explained by the diffusion-in-an-elliptic-cone model, but the distribution of the residues with large ellipticity values cannot.

Distributions of Diffusion-in-a-Cone. Low ellipticity corresponds to a circle-shaped distribution, to which the diffusion-in-a-cone model (Kinoshita et al., 1977; Lipari & Szabo, 1980) is applicable. Low ellipticity values (less than 1.4) were obtained for Leu2, Lys3, Gly30, Asp70, Lys96, Ala125, and Val155 (Figure 9c,d). The S^2 ratios in Figure 10a indicate, however, that the motions of Leu2 and Gly30 are not appropriately described as diffusion-in-a-cone, due to the heterogeneity in the density of the vectors in the cone (data not shown). All of these residues showed large internal motions (Figures 4a and 7a). The average of the S^2 values for these residues was 0.59 for NMR and 0.60 for MD. Furthermore, the average of the ASA values for the amide H atoms of these residues was 1.30 Å^2, which is larger than the average of all the analyzed residues (0.83 Å^2). Therefore, it is suggested that the diffusion-in-a-cone model is suitable for the protein surface residues that show large internal motions, with small steric hindrance.

Distributions of Gaussian Axial Fluctuation. Distributions with crescent-like shapes were observed for some residues, such as Asn16, Arg31, Asn43, Ala58, Glu61, Asp94, Ile116, Gly126, and Asp148 (Figure 9e). For these residues, the Gaussian axial fluctuation model (Brüschweiler & Wright, 1994) may be applicable, because it represents a restricted motion along a single axis. The orientations of the crescents are apparently reasonable when the axial rotation is considered to be along the N−C$_\alpha$ bond, but not the N−CO bond, which has the nature of a double bond. When the rotation along the N−C$_\alpha$ bond is considered, the axis vector in the polar coordinates [θ, ϕ] in Figure 9 should be [1.08 (62°), 3.14 (180°)], with a cone semiangle of 1.08 (62°). A fitting of the distribution of Ala58 and Ile116, for example, to the axial fluctuation model was performed, with minimization of the deviations of the cone θ angle. It gave cone axis vectors of [0.89 (51°), 3.18 (182°)] for Ala58 and [1.07 (61°), 3.07 (176°)] for Ile116, with cone semiangles of 1.17 (67°) and 1.08 (62°), respectively. Therefore, these distributions could be explained by axial fluctuation along the N−C$_\alpha$ vector. The standard deviations of the ϕ angles in the cone for these residues were 0.194 (11°) and 0.252 (14°), respectively. Assuming this model, the order parameters were calculated using the equation (Brüshweilar & Wright, 1994)

$$S^2 = 1 - 3 \sin^2 \theta(\cos^2 \theta(1 - \exp(-{\sigma_\phi}^2)) + (1/4) \sin^2 \theta(1 - \exp(-r{\sigma_\phi}^2))) \quad (17)$$

where θ is the cone semiangle and σ_ϕ is the standard deviation in the ϕ angle along the cone axis. The order parameters calculated according to this model for Ala58 and Ile116 were 0.93 and 0.87, respectively, which are higher than those calculated without assuming this model (0.87 and 0.79, respectively). This is natural because the fluctuation of the cone axis itself is not considered in this model. In addition, the distribution of the ϕ angles along the cone axis may indicate that the N−H vectors diffuse freely in a certain ϕ limit rather than with a Gaussian distribution. Therefore, the Gaussian-axial-fluctuation model seems to be applicable to some residues but not so ideally as to allow an explamation of the order parameter.

The average of the S^2 values for the residues listed above was 0.79 for NMR and 0.72 for MD. These values are smaller than the average for all the analyzed residues but are larger than those with the diffusion-in-a-cone distribution described above. The average of the ASA values was 2.18 Å^2, which is still larger than that for the residues with the diffusion-in-a-cone distribution (1.30 Å^2). Therefore, compared with those with the diffusion-in-a-cone distribution, the motions of the residues with the axial fluctuation distributions were restricted but not by steric hindrance of the amide H atom itself. Probably, the reorientation of the N−C$_\alpha$ vector is restricted, which gave the axial fluctuation of the N−H vector.

Distributions of Jump-between-Two-Cones. For residues such as Gln4, Gly15, and Glu154, the distribution seems to be two cones (Figure 9f). For these residues, the jump-between-two-cones model (Clore et al., 1990) seems to be applicable. These residues have very high ellipticity value, which could not be explained by the diffusion-in-an-elliptic-cone model (Figure 10a). The "long axes" for the three residues were perpendicular to the plane defined by the N−H vector and the N−C$_\alpha$ vector (Figure 10b). Therefore, the jump between the cones is probably caused by the rotation along the N−C$_\alpha$ bond. For Gly15 (Figure 9f), all of the N−H vectors in the initial 100 ps (closed circle) were included in one cone, and then they jump to the other one in the next 100 ps (open circles). They come back again in the last 100 ps (cross). This movement is shown in Figure 11 as the time dependence of the θ angle in the plane defined by the average of the N−H vectors and the "long axis" (thick solid line) or the "short axis" (thin dotted line). This was calculated as $[\sin^{-1}\mu_i\mu_x]$, where μ_i is each N−H unit vector and μ_x is the unit vector parallel to the "long" or "short" axis. This clearly shows that the motions can be divided into two types: a smaller noise-like motion and a larger

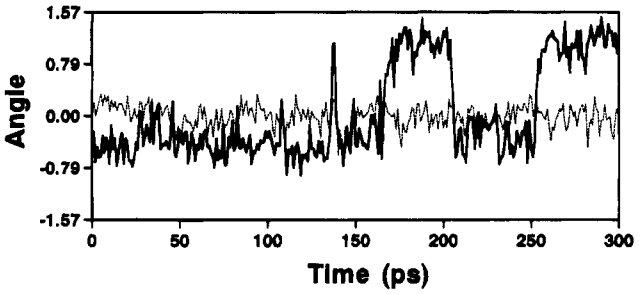

FIGURE 11: Time dependence of the θ angle of the N−H vector of Gly15 in a plane defined by the average of the N−H vectors and the long axis (thick solid line) or the short axis (thin broken line); see text.

motion along the long axis, which corresponds to the jump. The smaller one is equivalently observed along the long and short axes and is also equivalently observed for the two cones. The jump between the cone is a rare event, which occurred only five times during 300 ps of calculation.

These three residues show very small S^2 values (the average is 0.42 for NMR and 0.32 for MD). However, the ASA values for these three residues were 0. Therefore, this large jump-between-two-cones model seems to be applicable to the residues with amide H atoms largely hindered by other surrounding atoms, but in very mobile regions.

NMR and MD. In the present study, the backbone dynamics of *E. coli* RNase HI were characterized by a combination of ^{15}N-NMR relaxation analysis and MD simulation. The results by these two methods largely agreed with each other, judging from the order parameters. The two methods are complementary. NMR relaxation is a powerful experimental tool to obtain information about the internal motions of proteins on a picosecond to nanosecond time scale, with an atomic resolution. It provides amplitudes and correlation times. However, it gives little information on the actual motion. In addition, because the NMR relaxation analysis employed in the present study has only three kinds of data, it can describe only a simple correlation with one or two exponential terms. In contrast, the MD analysis provides detailed information on the internal motion, on almost the same time scale, with a clear image. It can describe more complex motions. Thus, we could use MD for the examination of dynamic models that have been employed for the interpretation of NMR relaxation. The problem with MD is its shorter limits in the simulation time. We hope that the limits will be extended so that it will be possible to use a larger time scale, such as microseconds, in the future.

ACKNOWLEDGMENT

We benefited greatly from numerous helpful discussions with Dr. R. Ishima and Prof. K. Nagayama of the University of Tokyo on the relaxation analyses. We also thank Dr. Y. Oda of Kyowa Hakko Kogyo and Dr. M. Haruki of our Institute for helpful advice on protein preparation, Dr. T. Ohkubo of our institute for advice and numerous discussions on NMR measurements and computation, and Drs. Y. Kuroda, T. Hirai, S. Hayashi, and F. Inoue of our Institute for helpful advice on computation.

SUPPLEMENTARY MATERIAL AVAILABLE

Table 1S, containing the T_1, T_2, and NOE values and the model-free parameters together with the ASA values of the amide H atoms (12 pages). Ordering information is given on any current masthead page.

REFERENCES

Abragam, A. (1961) *Principles of Nuclear Magnetism*, pp 264−353, Claredon Press, Oxford.

Balasubramanian, S., Nirmala, R., Beveridge, D. L., & Bolton, P. H. (1994) *J. Magn. Reson., Ser. B 104*, 240−249.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981) in *Intermolecular forces* (Pullman, B., Ed.) pp 331−342, Reidel, Dordrecht.

Brooks, C. L., III, Karplus, M., & Pettitt, B. M. (1988) *Proteins—A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, pp 194−195, Wiley-Interscience, New York, NY.

Brown, D., & Clarke, J. H. R. (1984) *Mol. Phys. 51*, 1243−1252.

Brüschweiler, R., & Wright, P. E. (1994) *J. Am. Chem. Soc. 116*, 8426−8427.

Chandrasekhar, I., Clore, G. M., Szabo, A., Gronenborn, A. M., & Brooks, B. R. (1992) *J. Mol. Biol. 226*, 239−250.

Clore, G. M., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.

Connolly, M. L. (1983) *Science 221*, 709−713.

Connolly, M. L. (1993) *J. Mol. Graphics 11*, 139−141.

Crouch, R. J., & Dirksen, M. L. (1982) in *Nucleases* (Linn S. M., & Roberts, R. J., Eds.) pp 211−241, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Eriksson, M. A. L., Berglund, H., Härd, T., & Nilsson, L. (1993) *Proteins: Struct., Funct., Genet. 17*, 375−390.

Fushman, D., Weisemann, R., Thüring, H., & Rüterjans, H. (1994a) *J. Biomol. NMR 4*, 61−78.

Fushman, D., Ohlenschläger, O., & Rüterjans, H. (1994b) *J. Biomol. Struct. Dyn. 11*, 1377−1402.

Haruki, M., Noguchi, E., Akasako, A., Oobatake, M., Itaya, M., & Kanaya, S. (1994) *J. Biol. Chem. 269*, 26904−26911.

Hiyama, Y., Niu, C., Silverton, J. V., Bavaso, A., & Torchia, D. A. (1988) *J. Am. Chem. Soc. 110*, 2378−2383.

Hoel, P. G. (1971) *Introduction to Mathematical Statistics*, 4th ed., John Wiley & Sons Inc., New York, NY.

Hubbard, S. J., Gross, K.-H., & Argos, P. (1994) *Protein Eng. 7*, 613−626.

Jones, D. N. M., Searles, M. A., Shaw, G. L., Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A., & Neuhaus, D. (1994) *Structure 15*, 609−627.

Jorgensen, W. L., & Tirado-Rives, J. (1988) *J. Am. Chem. Soc. 110*, 1657−1666.

Kanaya, S., & Crouch, R. J. (1983) *J. Biol. Chem. 258*, 1276−1281.

Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A. Iwai, S., Inoue, H., Ohtsuka, E., & Ikehara, M. (1990) *J. Biol. Chem. 264*, 4615−4621.

Kanaya, S., Katsuda-Nakai, C., & Ikehara, M. (1991) *J. Biol. Chem. 266*, 11621−11627.

Kanaya, S., Oobatake, M., Nakamura, H., & Ikehara, M. (1993) *J. Biotechnol. 28*, 117−136.

Katayanagi, K., Miyagawa, K., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1990) *Nature 347*, 306−309.

Katayanagi, K., Miyagawa, M., Matushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1992) *J. Mol. Biol. 223*, 1029−1052.

Kay, L. E., Torchia, D. A., & Bax, A. (1989) *Biochemistry 28*, 8972−8979.

Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., & Torchia, D. A. (1992) *J. Magn. Reson. 97*, 359−375.

Kinoshita, K., Jr., Kawato, S., & Ikegami, S. (1977) *Biophys. J. 20*, 289.

Levenberg, K. (1944) *Q. Appl. Math. 2*, 164−168.

Lipari, G., & Szabo, A. (1980) *Biophys. J. 30*, 489.

Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.

Lipari, G., Szabo, A., & Levy, R. M. (1982) *Nature 300*, 197−198.

Maniatis, T., Fritsh, E. F., & Sambrook, J. (1982) *Molecular Cloning—A Laboratory Manual,* 1st ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
Marquardt, D. (1963) *SIAM J. Appl. Math. 11*, 431−441.
Morikami, K., Nakai, T., Kidera, A., Saito, M., & Nakamura, H. (1992) *Comput. Chem. 16*, 243−248.
Nirmala, N. R., & Wagner, G. (1988) *J. Am. Chem. Soc. 110*, 7557−7558.
Nirmala, N. R., & Wagner, G. (1989) *J. Magn. Reson. 82*, 659−661.
Oda, Y., Yoshida, M., & Kanaya, S. (1993) *J. Biol. Chem. 268*, 88−92.
Palmer, A. G., III, Rance, M., & Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T., & Gronenborn, A. (1992) *Biochemistry 31*, 9150−9157.
Saito, M. (1992) *Mol. Simul. 8*, 321−333.
Saito, M. (1994) *J. Chem. Phys. 101*, 4055−4061.
Shaka, A. J., Barker, P. D., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
Sharp, K. A., & Englander, S. (1994) *Trends Biochem. Sci. 19*, 526−529.
Singh, U. C., Weiner, P. K., Caldwell, J. W., & Kollman, P. A. (1986) *AMBER (UCSF), version 3.0*, Department of Pharmaceutical Chemistry, University of California, San Francisco.
Wagner, G., & Wüthrich, K. (1986) *Methods Enzymol. 131*, 307−326.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765−784.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
Wittebort, R. J., & Szabo, A. (1978) *J. Chem. Phys. 69*, 1722.
Woessner, D. E. (1962) *J. Chem. Phys. 36*, 1−4.
Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H., & Nagayama, K. (1991) *Biochemistry 30*, 6036−6047.
Yamazaki, T., Yoshida, M., & Nagayama, K. (1993) *Biochemistry 32*, 5656−5669.
Yang, W., Hendrickson, W. A., Crouch, R. J., & Satow, Y. (1990) *Science 249*, 1398−1405.

BI942983G