

# Analysis of the Backbone Dynamics of the Ribonuclease H Domain of the Human Immunodeficiency Virus Reverse Transcriptase Using $^{15}\text{N}$ Relaxation Measurements<sup>†</sup>

Robert Powers,<sup>‡</sup> G. Marius Clore,<sup>\*‡</sup> Stephen J. Stahl,<sup>§</sup> Paul T. Wingfield,<sup>§</sup> and Angela Gronenborn<sup>\*‡</sup>

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Building 2, and Protein Expression Laboratory, Office of the Director, Building 6B, National Institutes of Health, Bethesda, Maryland 20892

Received June 8, 1992; Revised Manuscript Received July 17, 1992

**ABSTRACT:** The backbone dynamics of the uniformly  $^{15}\text{N}$ -labeled ribonuclease H (RNase H) domain of human immunodeficiency virus (HIV-1) reverse transcriptase have been investigated using two-dimensional inverse-detected heteronuclear  $^{15}\text{N}$ - $^1\text{H}$  NMR spectroscopy.  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and nuclear Overhauser enhancement (NOE) data were obtained for 107 out of a total of 134 backbone amide groups. The overall rotational correlation time ( $\tau_R$ ) for the protein at 26 °C is 10.4 ns. The backbone N-H vectors for all the measurable residues exhibit very fast motions on a time scale of  $\leq 20$  ps. The  $^{15}\text{N}$  relaxation data for only 14 residues can be explained by this single internal motion alone. A further 39 residues display a second motion on a time scale ranging from 28.8 ps to 3.9 ns, while another 15 residues are characterized by an additional motion on the 170-ns to 2.25-ms time scale resulting in  $^{15}\text{N}$   $T_2$  exchange line broadening. There are 39 residues that exhibit both the additional  $^{15}\text{N}$   $T_2$  exchange line broadening and the slow (28.8 ps–3.9 ns) internal motion. Thus, the RNase H domain experiences extensive mobility throughout its structure as evidenced by the 93 residues which exhibit multiple modes of motion. Distinctly mobile regions of the protein are identified by large decreases in the overall order parameter ( $S^2$ ) and correspond to the C-terminal residues and the loop regions between  $\beta$ -strands  $\beta_1$  and  $\beta_2$  and between  $\alpha$ -helix  $\alpha_B$  and  $\beta$ -strand  $\beta_4$ . The high mobility of the C-terminus is of particular interest since one stretch of the sequence in this region of the protein constitutes part of the proposed substrate binding site. Thus, a highly flexible or partially folded binding pocket could explain the lack of enzymatic activity observed for this particular HIV-1 RNase H domain.

The human immunodeficiency virus (HIV-1)<sup>1</sup> reverse transcriptase is a multifunctional enzyme which plays a crucial role in the life cycle of the virus. The enzyme exhibits both DNA- and RNA-dependent DNA polymerase and ribonuclease H (RNase H) activity which is essential for reverse transcription of the viral genomic RNA (Goff, 1990; Gilboa et al., 1979). Unlike in *Escherichia coli*, RNase H activity is required for viral replication, as evidenced by the failure of mutant provirus defective for RNase H function to produce infective virus particles (Schatz et al., 1989). The RNase H domain catalyzes the cleavage of the RNA portion of a DNA/RNA hybrid, a process which is required at several stages during reverse transcription, and displays both endonuclease and 3'→5' exonuclease activity (Krug & Berger, 1989; Mizrahi, 1989; Schatz et al., 1990). The native form of the enzyme is a heterodimer consisting of p66 and p51 subunits. This arises from asymmetric processing of the p66 homodimer by the HIV protease (di Marzo Veronese et al., 1986; Lightfoote et al., 1986). The RNase H activity resides in the 15-kDa C-terminal region of the p66 subunit as shown by deletion experiments, linker insertions, and point mutagenesis (Kotewicz et al., 1988; Tanese et al., 1988; Repaske et al., 1989; Schatz et al., 1989). There exists a significant degree of structural and sequence homology between *E. coli* RNase H and the C-terminal region of reverse transcriptase from

HIV and other retroviruses (Johnson et al., 1986; Yang et al., 1990; Katayanagi et al., 1990; Yamazaki et al., 1991; Davies et al., 1991). The Moloney murine leukemia virus (MoMuLV) C-terminal segment of reverse transcriptase has been expressed separately and shown to retain its RNase H activity (Tanese et al., 1988). In contrast, the 15-kDa C-terminal segment of HIV reverse transcriptase does not exhibit RNase H activity when expressed separately (Becerra et al., 1990; Hostomsky et al., 1991), although some controversy exists on this matter (Schatz et al., 1990; Evans et al., 1991; Restle et al., 1992). The full HIV RNase H activity, however, can be reconstituted by recombining the RNase H and DNA polymerase domains (Hostomsky et al., 1991).

In a recent paper, we presented the  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ , and  $^{13}\text{CO}$  backbone resonance assignments of the polypeptide backbone of the HIV-1 RNase H domain (Trp-113 → Ala mutant) and the determination of its secondary structure by 3D heteronuclear NMR spectroscopy (Powers et al., 1991). The structural conclusions derived from the NMR data were found to be in agreement with the high-resolution X-ray crystal structure (Davies et al., 1991). The principal difference between the two structures is located at the C-terminus of the protein. The C-terminal  $\alpha$ -helix observed in the crystal structure is absent, and residues Val-114 to Gly-133 are ill-defined in the NMR structure. Some of this disorder is also observed in the crystal structure, where a putative loop from residues 116 to 120 is not visible in the electron density map.

The present paper presents an analysis of the backbone dynamics of the RNase H domain of HIV-1 reverse transcriptase using 2D inverse-detected  $^1\text{H}$ - $^{15}\text{N}$  NMR methods.  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE data were obtained for 107 out of a total of 134 backbone amide groups. We show that the RNase H

<sup>†</sup> This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.)

<sup>‡</sup> Laboratory of Chemical Physics, NIDDK.

<sup>§</sup> Protein Expression Laboratory, Office of the Director.

<sup>1</sup> Abbreviations: HIV-1, human immunodeficiency virus 1; RNase H, ribonuclease H; NOE, nuclear Overhauser enhancement; SD, standard deviation; 2D, two dimensional.

domain experiences extensive motion throughout its structure, as manifested by the large number of residues which exhibit multiple modes of motion. In addition, there are distinct regions of the protein which exhibit an increased overall mobility relative to the remainder of the protein, indicated by large decreases in the overall order parameter ( $S^2$ ). These regions comprise the C-terminal residues from Tyr-110 to Leu-138 and the loops between  $\beta$ -strands  $\beta_1$  and  $\beta_2$  and between  $\alpha$ -helix  $\alpha_B$  and  $\beta$ -strand  $\beta_4$ . The high mobility of the C-terminus is of particular interest since this region contains the proposed substrate binding site and could explain the lack of enzymatic activity of the isolated HIV-1 RNase H domain.

## EXPERIMENTAL PROCEDURES

**Sample Preparation.** The preparation and purification of the RNase H domain comprising residues 427–560 of the 66-kDa reverse transcriptase with an additional four-residue sequence at the N-terminus was as described previously (Becerra et al., 1990; Powers et al., 1991). The Trp-113  $\rightarrow$  Ala mutant, on which all relaxation measurements were performed, was constructed by primer-directed mutagenesis (Oostra et al., 1983). Uniform  $^{15}\text{N}$  labeling (>95%) was carried out by growing cells in minimal medium containing  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source. The NMR sample contained 1.1 mM protein in 100 mM sodium acetate- $d_3$  buffer, pH 5.4, in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$ .

**NMR Spectroscopy.** All NMR experiments were carried out at 26 °C on a 600-MHz Bruker AM600 spectrometer operating in "reverse" mode. The pulse sequences used to measure  $^{15}\text{N}$   $T_1$  and  $T_2$  relaxation rates were based on those described previously (Kay et al., 1989a; Clore et al., 1990b), appropriately modified to eliminate cross-correlation between dipolar and chemical shift anisotropy relaxation mechanisms (Boyd et al., 1990; Kay et al., 1992; Palmer et al., 1992). The  $^1\text{H}$ - $^{15}\text{N}$  NOEs were measured exactly as described previously (Kay et al., 1989a; Clore et al., 1990b).

All experiments were recorded with a sweep width of 4166.67 Hz in the  $F_2$  dimension with the carrier set to the center of the amide NH region of the spectrum and the water signal on the right-hand edge. Low-power DANTE-style off-resonance irradiation was used to suppress the solvent resonance (Kay et al., 1989b) and, except for the NOE experiment, was applied for 1.2 s. A total of 512 increments, each of 1K data points, were recorded, giving a total acquisition time of 72.2 ms and a sweep width of 1773 Hz in the  $F_1$  dimension. Thirty-two transients per  $t_1$  increment were recorded for the  $T_1$  and  $T_2$  experiments and 136 transients for the NOE experiments.  $T_1$  values were obtained by using seven  $T_1$  delays of 14.49, 43.55, 101.67, 198.53, 392.25, 634.39, and 876.55 ms, while  $T_2$  values were obtained by using eight  $T_2$  delays of 0, 14.36, 29.19, 43.68, 72.80, 87.36, 101.92, and 116.48 ms. The  $T_1$ ,  $T_2$ , and NOE data sets were processed with the NMR2 software (New Methods Research Inc., Syracuse, NY) running on a Sun Sparc 1+ workstation. Intensities and volumes of cross peaks were measured by peak-picking routines provided in the software package. A linear baseline correction was applied in the  $F_2$  dimension for the  $T_1$  and  $T_2$  data. A polynomial baseline correction of order 3 was applied in the  $F_2$  dimension for the NOE data.

The decay of cross-peak intensity with time was found to be exponential within experimental error for both the  $T_1$  and  $T_2$  experiments. The experimental decays were fit to a single exponential using conjugate gradient minimization with errors calculated on the basis of a Monte Carlo analysis (Kamath et al., 1989). The FACSIMILE program (Chance et al.,

1979; Clore et al., 1983), running on a MicroVAX III, was used to analyze the  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE data as described previously (Clore et al., 1990a,b). The fitting procedure takes into account the errors in the  $T_1$ ,  $T_2$ , and NOE values. The errors in the values of the optimized parameters were obtained from conventional analysis of the variance-covariance matrix generated by the nonlinear least-squares optimization routine (Chance et al., 1979; Clore et al., 1983).

## RESULTS

**$^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE Relaxation Data.** Figure 1 illustrates a set of three spectra each from the  $^{15}\text{N}$   $T_1$  and  $T_2$  data recorded at 600 MHz. We were able to obtain quantitative data for 107 out of a total of 134 (i.e., ~80%) backbone amide groups. The majority of the remaining 25 backbone amide groups did not yield quantitative data owing to either the absence or very weak intensity of the cross peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum and/or lack of assignment from the 3D triple-resonance experiments (Powers et al., 1991). The decay of cross-peak intensity with time for both  $T_1$  and  $T_2$  experiments is shown in Figure 2 for selected residues covering the range of observed  $T_1$  and  $T_2$  values. The standard deviations of the best-fit  $T_1$  and  $T_2$  values were all less than 10%, with the majority being less than 5%.

The  $^1\text{H}$ - $^{15}\text{N}$  NOE correlation spectrum is shown in Figure 3. There are two residues, Val-137 and Leu-138, which have a negative NOE, and one residue, Glu-124, has a near zero NOE. The estimated errors in the measured values of the experimental NOEs are of the order of  $\pm 0.1$ . In addition, there is a potential systematic error in the measurement of the NOE due to chemical exchange between water and labile NH protons, which gives rise to an apparent NOE if the recycle time is short relative to the  $T_1$  of water (Smith et al., 1987; Kay et al., 1989a). For a  $T_1$  of  $\sim 2.5$  s and a recycle time of 3 s we can estimate a systematic increase of at most 20% in the relative magnitude of the experimental NOE (Clore et al., 1990b). These errors can account for the observation that some of the experimental NOEs have values greater than the theoretical maximum of 0.82 (cf. Figure 4 and the supplementary material). Consequently, in the few cases where the observed NOE was greater than 0.82, a value of  $0.82 \pm 0.1$  was assumed for fitting the data to various spectral density functions (see below).

The  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE values are plotted as a function of residue number in Figure 4, and their values and errors are included in the supplementary material.

The observed mean  $^{15}\text{N}$   $T_1/T_2$  ratio for the RNase H domain was  $11.73 \pm 2.96$  with a maximum ratio of 28.64 for Thr-48 and a minimum ratio of 4.80 for Glu-124. The overall correlation time ( $\tau_R$ ) can be directly determined from the  $T_1/T_2$  ratio under conditions where the internal correlation time ( $\tau_e$ ) is less than 100 ps,  $\tau_R > 1$  ns, and  $T_2$  is not significantly shortened by chemical exchange (Kay et al., 1989a; Clore et al., 1990b). The presence of a significant contribution to the relaxation from either chemical exchange or internal motion ( $\tau_e$ ) is implied by a  $T_1/T_2$  ratio which lies either 1 SD above or below the mean, respectively (Clore et al., 1990b). Since  $\tau_R$  must clearly be the same for all residues, we therefore excluded 24  $T_1/T_2$  ratios which were outside  $\pm 1$  SD from the mean. The  $T_1/T_2$  ratios for the remaining 85 residues had an average value of  $11.45 \pm 1.49$ . These 85  $T_1/T_2$  ratios were then fitted simultaneously by optimizing a single value of  $\tau_R$ , yielding a correlation time of  $10.4 \pm 0.08$  ns which was used in all subsequent calculations. (This compares to a value of  $10.5 \pm 0.30$  ns for  $\tau_R$  when the  $T_1/T_2$  ratios for all 107 are

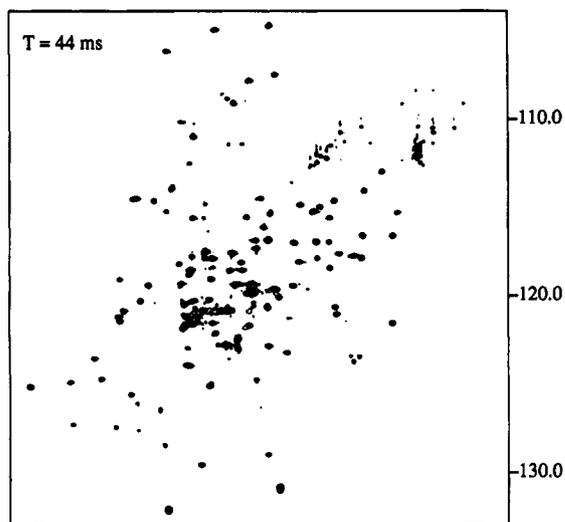
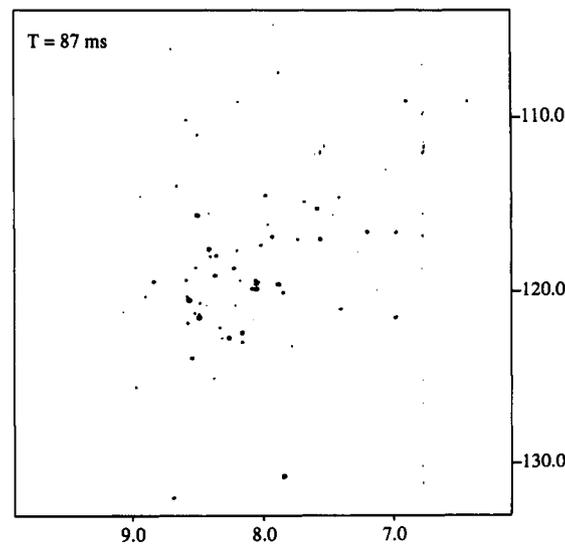
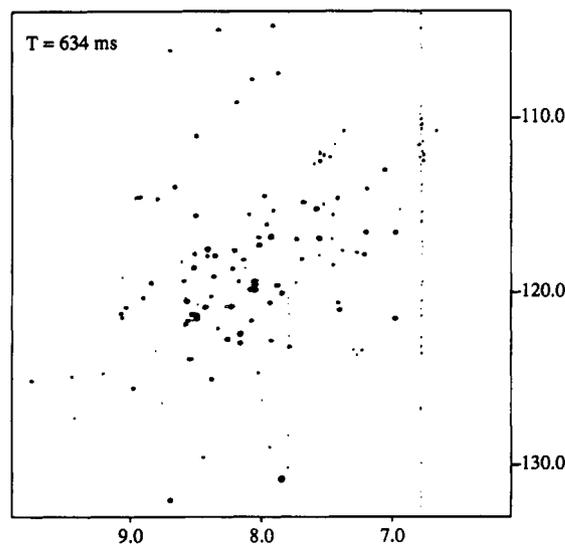
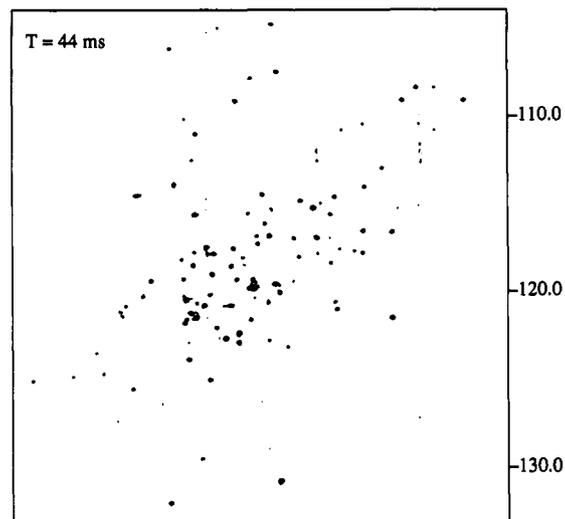
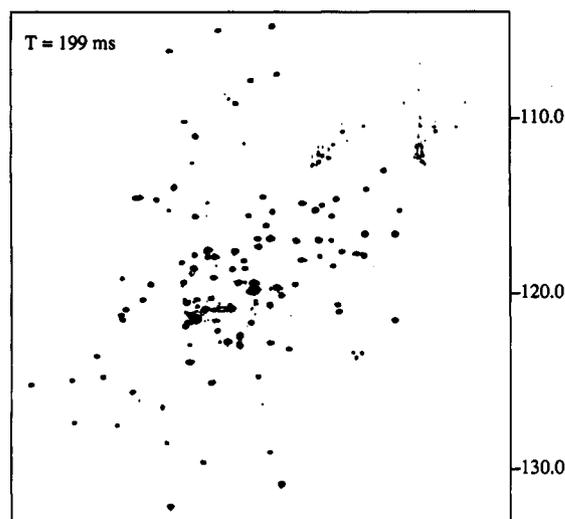
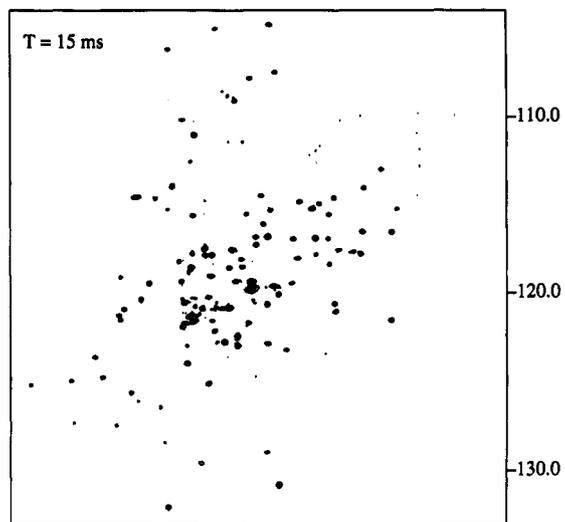
A  $T_1$ B  $T_2$  $^1\text{H F}_2$  (ppm) $^1\text{H F}_1$  (ppm)

FIGURE 1: Contour plots of the 600-MHz  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra of RNase H obtained for the  $^{15}\text{N}$   $T_1$  (A) and  $T_2$  (B) experiments at different interval times,  $T$ .

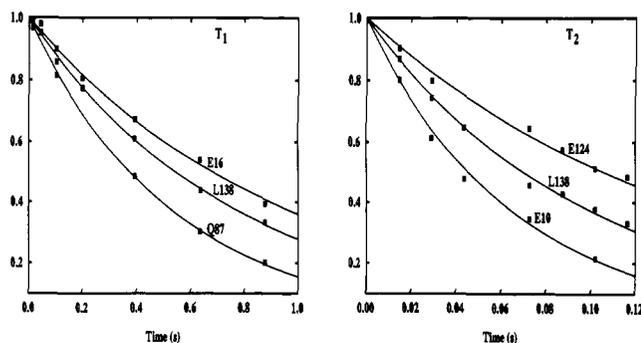


FIGURE 2: Comparison of the experimental  $^{15}\text{N}$   $T_1$  and  $T_2$  data (■) of selected residues of RNase H spanning the full range of  $T_1$  and  $T_2$  values with the single-exponential least-squares best-fit curves (—).

used.) In an attempt to further validate the accuracy of this  $\tau_R$  determination, we constructed histogram plots of the  $T_1$  and  $T_2$  data for residues known to be part of well-defined secondary structure in solution (Powers et al., 1991). The median  $T_1/T_2$  ratio for these residues is 11.55 and corresponds to a calculated correlation time of 10.4 ns, which is identical to the results obtained using data that fall within  $\pm 1$  SD of the mean  $T_1/T_2$  ratio. Subsequent analysis (see below) indicated that, of the 85 residues whose  $T_1/T_2$  ratios were within  $\pm 1$  SD of the mean, 56 required a small contribution from chemical-exchange line broadening to  $T_2$ . The correlation time calculated from the  $T_1/T_2$  ratios of the 29 remaining residues was  $10.2 \pm 0.07$  ns, in agreement with the above results.

It is also instructive to compare the correlation time obtained for the RNase H domain with those of other proteins of similar size. Values of 8.3 and 9.1 ns were obtained for interleukin- $\beta$  (153 residues; Clore et al., 1990b) and staphylococcal nuclease (149 residues; Kay et al., 1989a) at 36 °C. Assuming that the relationship between the solvent viscosity and the rotational correlation time is linear and that the viscosity of the protein solution follows the same temperature dependence as that of water, these values translate to values of 10.4 and 11.4 ns, respectively, at 26 °C (taking the viscosity of water at 26 and 36 °C to be 880 and 700  $\mu\text{Pa s}$ , respectively; Lyde, 1990), which is consistent with the value of 10.4 ns obtained for the RNase H domain at 26 °C.

There is an inherent problem with determining the true rotational correlation time for the RNase H domain. The underlying assumption in using  $T_1/T_2$  ratios in the determination of  $\tau_R$  is that the protein behaves in the main as a rigid tumbling sphere. Such an assumption would be perfectly reasonable on the basis of the X-ray structure of the RNase H domain (Davies et al., 1991) which has a radius of gyration of 14 Å and a calculated ratio of 1.00:0.81:0.59 for the three principal components of the inertia tensor. This may not be the case for the RNase H domain in solution (see below). It is also impossible to know a priori which residues are likely to lack any significant internal motion. The ones chosen in the  $\tau_R$  calculation are those which do not deviate significantly (i.e., by more than  $\pm 1$  SD) from the mean  $T_1/T_2$  ratio, based on the assumption that these residues most likely represent the stable protein core with little additional internal mobility. This is necessarily an oversimplification and may well be flawed in the present case, since a large degree of flexibility and motion is apparent for the majority of residues in the RNase H domain (see below). Thus, even for residues close to the  $T_1/T_2$  mean there may exist a systematic contribution from both chemical exchange and internal motion. Consequently,

the absolute value of  $\tau_R$  could contain a systematic error which would obviously propagate itself throughout the further analysis of the  $^{15}\text{N}$  relaxation data when the order parameters ( $S^2$ ) and internal correlation times ( $\tau_e$ ) are calculated. We therefore stress that the absolute values that are calculated in this particular case may be unreliable and that the results presented here should therefore be viewed as a qualitative analysis of the internal dynamics of the RNase H domain of HIV-1.

**Analysis of the Relaxation Data.** The approach employed to analyze the relaxation data involved fitting the  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE values simultaneously for each residue using a value of 10.4 ns for the rotational correlation time  $\tau_R$  and optimizing the value of the generalized order parameter  $S^2$  and, where appropriate, the values for other terms relating to internal motion and/or chemical exchange (Clore et al., 1990b). The data for each residue were fitted to six different functions, which are summarized in Table I. These functions provide a model-free description of the dynamics and do not yield a description of the underlying motions. Model 1 comprises a simplified spectral density function in which the second term of the model-free Lippari and Szabo (1982) spectral density function, given by model 2, is assumed to make a negligible contribution to  $T_1$  and  $T_2$  (i.e.,  $\tau_e \leq 20$  ps). Model 1 can only be applied when the experimental NOE value is greater than 0.7, since it predicts that the value of the NOE is at its theoretical maximum of 0.824. Two different fitting procedures were used for model 2: in the first (model 2a), only the  $T_1$  and  $T_2$  data were fitted, whereas in the second (model 2b), the  $T_1$ ,  $T_2$ , and NOE data were used in the fitting procedure. Model 3 is the expanded spectral density function of Clore et al., (1990a,b) in which there are two internal motions which differ in time scale by at least 1 order of magnitude and where the slower of the two internal motions is no longer in the extreme narrowing limit (i.e.,  $100 \text{ ps} < \tau_s < \tau_R$ ). The faster of the two motions has to have a lifetime ( $\tau_f$ ) of less than about 30 ps, and for the purposes of the present analysis we assumed that  $\tau_f$  was sufficiently small ( $< 10$  ps) so as to make a negligible contribution to the relaxation parameters (Clore et al., 1990b). It should also be noted that the overall order parameter  $S^2$  remains constant up to  $\tau_f \sim 30$  ps (Clore et al., 1990b). Models 4–6 correspond to models 1–3, respectively, but include an additional chemical exchange term for  $T_2$ .

The results of the fits of each of the six models in Table I were compared to determine which model best described the experimental relaxation data while minimizing the complexity of the model and ensuring that the optimized parameters were well determined by the data. The results of these computations are plotted in Figure 4 on a per residue basis along with the experimental  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE data.

Only 14 residues could be fitted to model 1 which yielded values of  $S^2 > 0.7$ . Only a single residue, namely, Gln-53, could be fitted to model 2a. In all other cases this resulted in severe underestimation of the experimental NOE. For example, the experimental NOE for Glu-8 is 0.7, but a fit of model 2 to the  $T_1$  and  $T_2$  data yields a calculated NOE of  $-0.067$ . Even though there are inherent errors in the measurement of the experimental NOE, there is no ambiguity in the distinction between positive and negative peaks, and the potential 20% error due to chemical exchange between NH protons and water (see above) is not sufficient to account for the observed discrepancy between the experimental data and the fits obtained using model 2.

When the NOE data, however, are taken into account during the optimization, 15 residues could be fitted to model 2b. Of

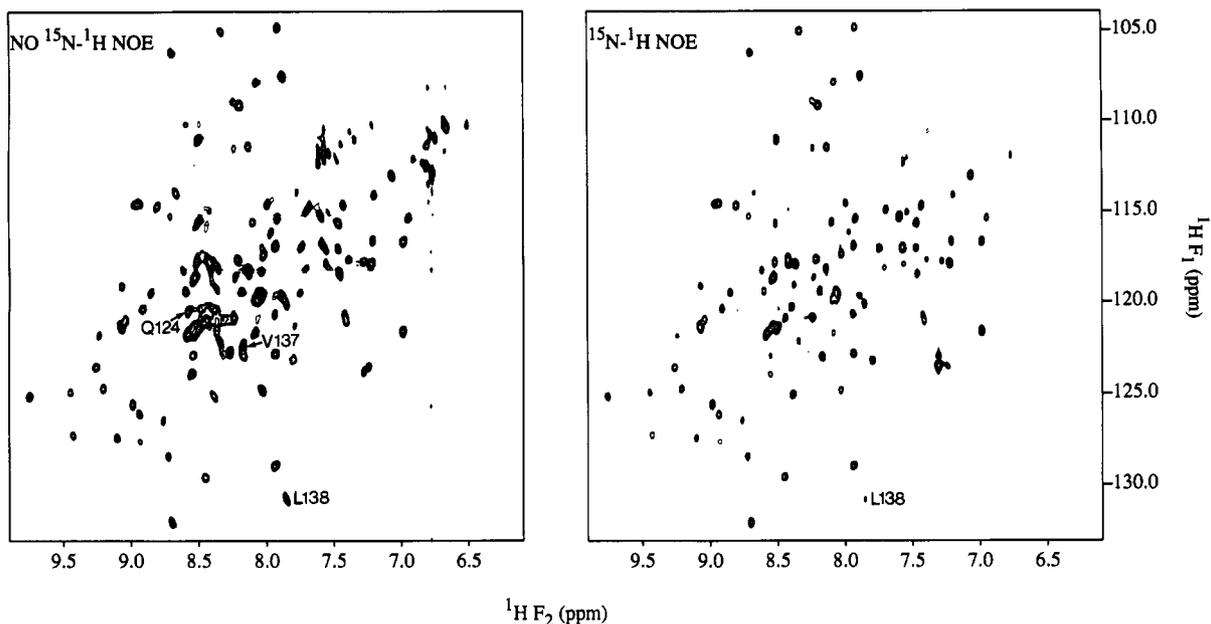


FIGURE 3: 600-MHz  $^1\text{H}$ - $^{15}\text{N}$  correlation and  $^1\text{H}$ - $^{15}\text{N}$  NOE correlation spectra of HIV-1 RNase H. The locations of the cross peaks for Gln-124, Val-137, and Leu-138, which have negative and near-zero NOEs, are indicated in the  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum, as they cannot be seen in the  $^1\text{H}$ - $^{15}\text{N}$  NOE correlation spectrum at the contour level plotted in the figure. The contours for Leu-138, which has a negative NOE, are displayed as dashed lines in the  $^1\text{H}$ - $^{15}\text{N}$  NOE correlation spectrum. Both spectra are plotted on the same contour intensity scale.

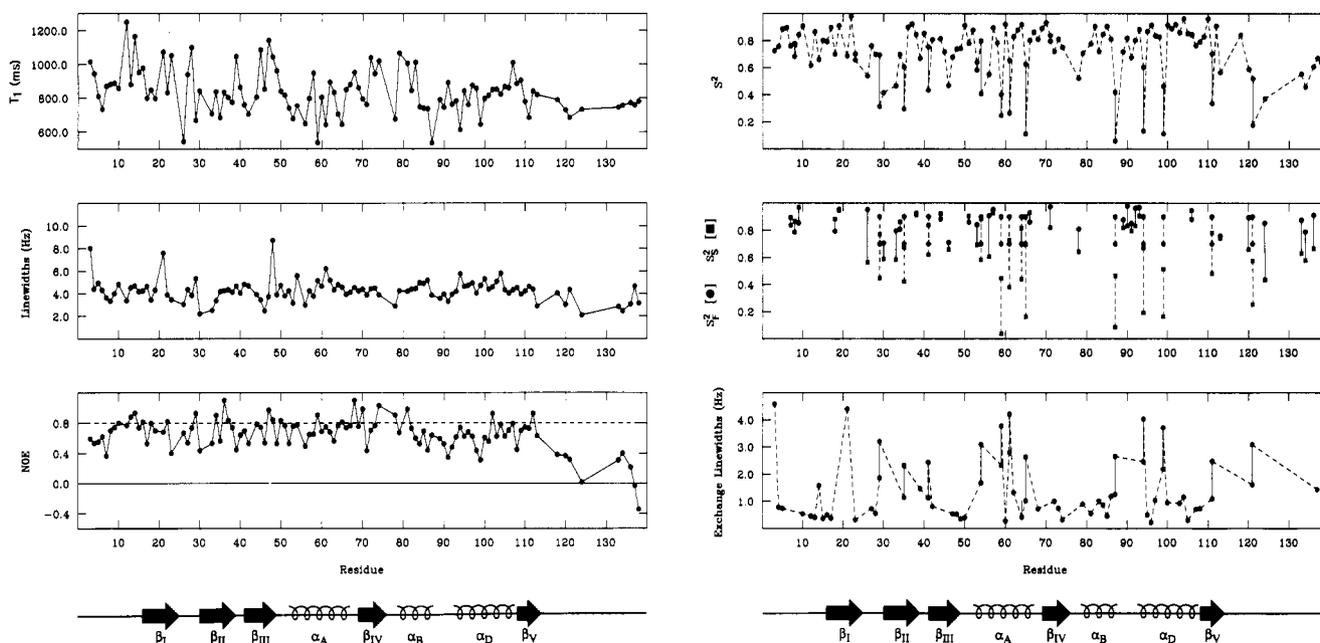


FIGURE 4: Plots as a function of residue number of  $^{15}\text{N}$   $T_1$ ,  $T_2$  [displayed as line width =  $1/(\pi T_2)$ ], and NOE values measured at 600 MHz, together with the overall order parameter ( $S^2$ ), fast-motion order parameter ( $S^2$ ), slow-motion order parameter ( $S^2$ ), and exchange line width calculated from the data. The solid lines in the plot of  $S^2$  and exchange line widths connect the range of values for those residues whose data were fit to model 6 with a grid search, assuming a value of  $\tau_f \leq 20$  ps and a range of  $S^2$  of 0.9–0.7. The dashed lines connect the range of values for those residues whose data were fit to model 3. The locations of the five  $\beta$ -strands and three  $\alpha$ -helices, as determined by NMR (Powers et al., 1991), are indicated below the figure.

these 15 residues, 9 could also be fitted to model 3 (see below). The results of fitting the  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE data to model 2b predict values for the effective internal correlation time  $\tau_c$  ranging from 28.7 to 839.7 ps. This time scale is significantly slower than that predicted for model 1 ( $\leq 20$  ps). Although the very fast thermal motions described by model 1 must be present for all residues, the decay of the internal correlation function can still be described by a single exponential providing the additional motions are in the extreme narrowing limit. The time constant for this decay, however, will be a weighted average of the various internal motions present and, conse-

quently, cannot be related to the actual time scale of any particular motion.

When the time scale of the slower motion is no longer in the extreme narrowing limit, the relaxation data can no longer be fitted to model 2, and the internal correlation function has to be expanded to include two distinct internal motions (model 3). A total of 32 residues could be fitted to model 3, including the 9 residues which could also be fitted to model 2b. Comparison of the results obtained for those residues which fit models 2b and 3 equally well reveals that the overall order parameter  $S^2$  is generally the same in both cases.

Table I: Model-Free Functions Used To Analyze the  $^{15}\text{N}$  Relaxation Data of the RNase H Domain of HIV-1<sup>a</sup>

model	spectral density functions	data used in fitting	optimized parameters
1	$J(\omega_i) = S^2\tau_R/(1 + \omega_i^2\tau_R^2)$	$T_1, T_2$	$S^2$
2 <sup>b</sup>	$J(\omega_i) = S^2\tau_R/(1 + \omega_i^2\tau_R^2) + (1 - S^2)\tau_c/(1 + \omega_i^2\tau_c^2)$	(a) $T_1, T_2$ (b) $T_1, T_2, \text{NOE}$	$S^2, \tau_c$ $S^2, \tau_c$
3 <sup>c</sup>	$J(\omega_i) = S^2S^2\tau_R/(1 + \omega_i^2\tau_R^2) + S^2(1 - S_s^2)\tau_s/(1 + \omega_i^2\tau_s^2)$	$T_1, T_2, \text{NOE}$	$S^2, S_s^2, \tau_s$
4	$J(\omega_i) = S^2\tau_R/(1 + \omega_i^2\tau_R^2)$ $1/T_2(\text{obs}) = 1/T_2 + \pi\Delta\text{ex}$	$T_1, T_2$	$S^2, \Delta\text{ex}$
5 <sup>b</sup>	$J(\omega_i) = S^2\tau_R/(1 + \omega_i^2\tau_R^2) + (1 - S^2)\tau_c/(1 + \omega_i^2\tau_c^2)$ $1/T_2(\text{obs}) = 1/T_2 + \pi\Delta\text{ex}$	$T_1, T_2, \text{NOE}$	$S^2, \Delta\text{ex}, \tau_c$
6 <sup>c,d</sup>	$J(\omega_i) = S^2S^2\tau_R/(1 + \omega_i^2\tau_R^2) + S^2(1 - S_s^2)\tau_s/(1 + \omega_i^2\tau_s^2)$ $1/T_2(\text{obs}) = 1/T_2 + \pi\Delta\text{ex}$	$T_1, T_2, \text{NOE}$	$S_s^2, \tau_s, \Delta\text{ex}$

<sup>a</sup> The equations to calculate  $T_1$ ,  $T_2$ , and NOE from the spectral density functions are given by eqs 1–5 in Clore et al. (1990b). <sup>b</sup>  $\tau_c' = \tau_R\tau_c/(\tau_R + \tau_c)$ . <sup>c</sup>  $\tau_s' = \tau_R\tau_s/(\tau_R + \tau_s)$ ;  $S^2 = S_f^2S_s^2$ ;  $\tau_f \leq 20$  ps. <sup>d</sup> A grid search with  $S_f^2$  values of 0.9–0.7 was used to evaluate this model.

Thus, of the 38 residues which were fitted to either model 2b or model 3, there were 6 residues which could only be fitted to model 2b and 23 which would only be fitted to model 3. Failure to fit a model was usually manifested by an overall order parameter greater than 1, very large errors in the optimized parameters, or a failure of the program to converge. Since the underlying assumption in model 2b is that the internal motions are in the extreme narrowing limit, this model begins to fail when the contribution of motions slower than the extreme narrowing limit become significant. Conversely, model 3 only describes the data when the internal motions inside (fast) and outside (slow) the extreme narrowing limit are distinct. There is obviously an overlap region where both models describe the data equally well. This overlap region probably occurs when the two rates differ by a factor of 5–10.

The  $^{15}\text{N}$  relaxation data for the remaining 54 residues could only be fitted by including an additional term to account for chemical-exchange contributions to  $T_2$  such that  $1/T_2(\text{obs}) = 1/T_2 + \pi\Delta\text{ex}$ , where  $\Delta\text{ex}$  is the increase in line width due to a chemical-exchange process. Fifteen of these residues could be fitted to model 4 (i.e., model 1 plus chemical exchange), indicating that they experience both a very fast internal motion ( $<20$  ps) and a motion longer than the overall rotational correlation time. As with the data that were fitted with model 1, these residues all have NOE values  $\geq 0.7$  and are characterized by values  $\geq 0.7$  for the overall order parameter  $S^2$ . The distinguishing feature for these residues is the large  $T_1/T_2$  ratio relative to the mean which arises from the shortening of  $T_2$  as a result of chemical exchange. The observed increase in line width is directly proportional to the lifetime  $\tau_{\text{ex}}$  of the exchange process. In the case of two-site exchange  $\Delta\text{ex} = (1/2)\pi(\nu_A - \nu_B)^2\tau_{\text{ex}}$ , where  $\nu_A$  and  $\nu_B$  are the chemical shifts of states A and B. Since the chemical shift difference between the two states is unknown, we can only estimate the range of rates for this process. The upper limit for the lifetime of the chemical-exchange process is determined by the time interval between the refocusing pulses of the Carr–Purcell–Meinboom–Gill sequence used in the  $T_2$  experiment (in this case set to 400  $\mu\text{s}$ ). This sets the upper limit of  $\sim 2.25$  ms for the exchange lifetime  $\tau_{\text{ex}}$ . We can estimate the lower limit by assuming a maximum  $^{15}\text{N}$  chemical shift difference of 10 ppm and a minimum detectable increase in line width of 0.1 Hz. The minimum value of the exchange lifetime  $\tau_{\text{ex}}$  is then calculated to be about 170 ns.

An additional 27 residues were fitted to model 5 (i.e., model 2b plus chemical exchange). These residues differ from the previous 15 residues, which were fitted to model 4, by the values of the observed NOEs which are all  $\leq 0.7$ . This indicates that, in addition to the fast motion ( $\leq 20$  ps) observed for all residues and the presence of chemical exchange indicated by the large  $T_1/T_2$  ratio, these residues also experience other

slower motions which are still in the extreme narrowing limit.

Finally, there were 12 residues which could not be fitted to models 1–5. These residues are similar to the previous 27 residues, which were fitted to model 5, in that they all have NOE values  $\leq 0.7$  and large  $T_1/T_2$  ratios. This implies that the difference in the rates for the slow and fast internal motions is too great to fit the Lipari and Szabo (1982) model and requires the expanded two-exponential internal correlation function (Clore et al., 1990a,b). Thus, these residues are similar to the 23 residues which could only be fitted to model 3, but with a required additional chemical-exchange term. The parameters describing such a model cannot be determined explicitly, since this would require solving four unknowns with three experimental data points. We approximated the results for these residues by doing a grid search for a range of values for  $S_f^2$  between 0.9 and 0.7. This range was chosen on the basis of the average value of  $0.88 \pm 0.06$  for  $S^2$  determined for the 14 residues of RNase H which could be fitted to model 1 and the average value of  $0.82 \pm 0.05$  and  $0.75$  for  $S_f^2$  found for interleukin-1 $\beta$  from experimental  $^{15}\text{N}$  relaxation studies (Clore et al., 1990b) and a 500-ps molecular dynamics simulation (Chandrasekhar et al., 1992), respectively. This approach allowed us to calculate a range of  $\Delta\text{ex}$ ,  $\tau_s$ , and  $S_s^2$  values. The resulting range for these residues is plotted as a solid line connecting the maximum and minimum points in Figure 4.

## DISCUSSION

We have demonstrated that the RNase H domain of HIV-1 is a very dynamic and mobile protein exhibiting multiple rates of motion on three different time scales. Each residue experiences a very fast motion on the time scale of  $\leq 20$  ps, with most residues exhibiting additional motions on a slower time scale. Only 14 residues out of a total of 107 residues experience exclusively this fast local motion, compared to the results for IL-1 $\beta$  where approximately 40% of the residues exhibited only this fast internal motion (Clore et al., 1990b). The mobility of RNase H appears even more dramatic when compared to the results for staphylococcal nuclease in which all but three residues can be accurately described by the presence of only this fast motion (Kay et al., 1989a; Clore et al., 1990a). The additional slow motions present can be grouped in the following manner. A slow motion on the order of 28.8 ps–3.9 ns was present for 39 residues, while another 15 residues are characterized by an additional chemical exchange which correlates to a motion on the time scale of 170 ns–2.25 ms. More strikingly, the remaining 39 residues (36% of the observed residues) can only be described by a model that contains three motional regimes, namely, a fast, a slow, and a chemical-exchange motion. The location of these different groups of residues with regard to the secondary

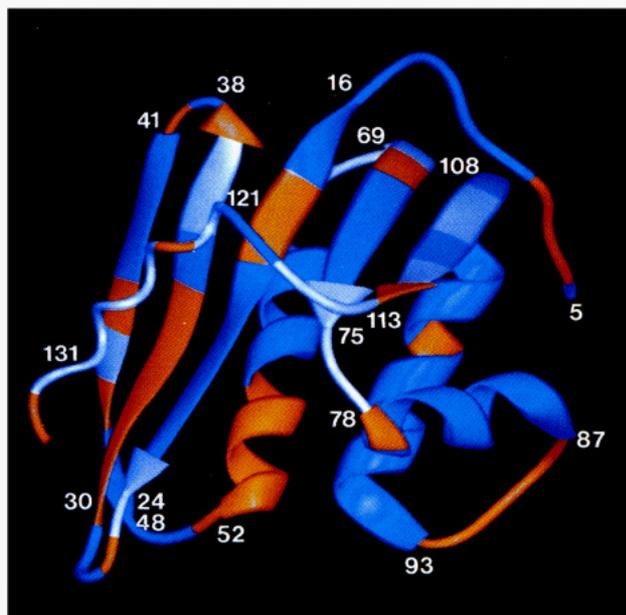


FIGURE 5: Summary of the results of the relaxation data analysis superimposed on the crystal structure of the RNase H domain of HIV-1. Regions in white represent either residues which can be fit to model 1 or residues for which there are no data; regions in orange represent residues which can be fit to models 2 or 3; regions in blue represent residues which exhibit exchange line broadening (i.e., models 4–6). Helix  $\alpha_E$  (residues 121–131) in the crystal is shown as a coil as it is ill-defined in solution (Powers et al., 1991, and this work). The crystal structure is from Davies et al. (1991), and the schematic drawing was produced using the program Ribbon 2.0 (Carson, 1987).

and tertiary structure of the RNase H domain is summarized in Figures 4 and 5, respectively. A further observation which supports the extensive mobility of the RNase H domain was our inability to quantify the results for the remaining 25 residues of the RNase H structure. The NH and  $^{15}\text{N}$  chemical shifts for 19 of these 25 residues could not be assigned using 3D double- and triple-resonance experiments as reported previously (Powers et al., 1991). We attribute this to chemical exchange resulting in line widths greater than the relevant coupling constants such that they are no longer detectable in the 3D heteronuclear experiments. In addition, a majority of these residues do not show cross peaks in the 2D  $^{15}\text{N}$ -NH correlation experiments. The other six residues which were unassigned also give very weak or nonexistent cross peaks in the 2D  $^{15}\text{N}$ -NH correlation experiments. This is also indicative of extensive mobility on the part of these residues since the lack of a cross peak can only be explained by chemical exchange between conformational states resulting in broadening of the NH and/or  $^{15}\text{N}$  resonances beyond detection.

The overall order parameter  $S^2$  is a good indicator of the local motion experienced by each residue. The average value of  $S^2$  for the 14 residues whose data could be fitted to model 1 is  $0.88 \pm 0.06$ , which corresponds to a mean semiangle of  $17 \pm 5^\circ$  for free diffusion in a cone. If this motion is interpreted to represent the fast random thermal fluctuations experienced by all atoms in the protein (Clare et al., 1990b; Chandrasekhar et al., 1992), then any significant decrease of  $S^2$  below 0.88 is indicative of large amplitude fluctuations of slower internal motions. On this basis, analysis of the relative magnitude of the overall order parameters for the RNase H domain should provide some insight into the dynamics that are occurring in this protein.

Examination of the plot of  $S^2$  versus residue number clearly shows some distinct regions of relative high mobility and flexibility (Figures 4 and 5). Two major dips observed in the

plot of the overall order parameter occur in loop regions between  $\beta$ -strands  $\beta_1$  and  $\beta_2$ , centered around residue 30, and between  $\beta$ -strand  $\beta_4$  and  $\alpha$ -helix B, centered around residue 79 (Figure 4). The most striking and possibly functionally relevant variation in  $S^2$  occurs at the C-terminus, where the value of  $S^2$  decreases dramatically beginning at residue 110 (Figure 4). This reduction in the value of  $S^2$  is paralleled by a large decrease in the values of the  $^{15}\text{N}$ -NH NOEs. Comparing the relative magnitude of the order parameters and  $^{15}\text{N}$ -NH NOEs between the two mobile loops and the C-terminal regions, it appears that the motion observed in the C-terminus is greater than that in the loop regions. In addition, all the residues (except Ala-112) in the C-terminal region can only be described by models with multiple motions (model 2b, 3, or 5). This clearly indicates that the C-terminal end of RNase H, beginning at residue 110, is very dynamic, undergoing multiple and large amplitude motions, and should probably be characterized as a disordered region of the protein.

It is also interesting to note that, of the 12 residues that could only be fitted to model 6 in which there is both chemical-exchange line broadening and two internal motions differing by at least an order of magnitude, 9 are located within regions of secondary structure: namely, Tyr-35 in strand  $\beta_2$ , Lys-54, Ala-59, Tyr-61, Leu-64, and Gln-65 in helix  $\alpha_A$ , Glu-94 and Ile-99 in helix  $\alpha_D$ , and Leu-111 in strand  $\beta_5$ . Although the underlying motional basis for this phenomenon is unknown, these observations provide yet further clear indication of the highly unusual dynamic properties of the RNase H domain.

The observation of extensive mobility at the C-terminus from the  $^{15}\text{N}$  relaxation data is consistent with the NMR secondary structure previously presented where the C-terminus was ill-defined (Powers et al., 1991). It is also hinted at in the crystal structure (Davies et al., 1991) where residues Ala-116 to Ile-120 could not be placed in the model due to weak or absent electron density.

The observation of extreme flexibility at the C-terminus may shed some light on the curious observation that the enzymatic activity is lost when the RNase H domain is cleaved from the intact reverse transcriptase heterodimer (p51/p61) but restored following reconstitution with a p51 homodimer (Hostomsky et al., 1991). There exists extensive structural and sequence similarity between the HIV RNase H domain and the *E. coli* RNase H protein (Johnson et al., 1986). The X-ray crystal structure of *E. coli* RNase H has implicated a well-defined protruding loop between strand  $\beta_5$  and helix  $\alpha_E$  in substrate binding (Yang et al., 1990; Katayanagi et al., 1990). This loop contains His-118, which is one of seven invariant residues identified in all the known RNase H sequences (Doolittle et al., 1989), and is clearly located in a highly mobile region of the NMR structure. In addition, residues Ala-113 to Asn-123 comprise the most highly conserved stretch of 12 amino acids in five lentivirus RNase H sequences (Davies et al., 1991). The point mutation of His-118  $\rightarrow$  Phe in HIV-1 reverse transcriptase results in a significantly reduced RNase H activity while maintaining reverse transcriptase activity (Schatz et al., 1989; Tisdale et al., 1990). Also, deletions of more than 16 amino acids from the C-terminus or more than 23 amino acids from the N-terminus result in loss of RNase H activity (Hizi et al., 1990). These results suggest that the C-terminal region of RNase H plays a crucial role in its function, and partial unfolding of this region, as exemplified by its extreme flexibility in solution, may account for the observed loss of enzymatic activity. This implies that the presence of p51 provides substantial stabilization of the RNase H C-terminus, either by indirect means or by a direct interaction. This notion derives

support from two sources. First, mutational studies show that changes in the polymerase domain of reverse transcriptase can cause dramatic effects on RNase H activity (Hizi et al., 1990). Second, the presence of helix  $\alpha_E$  (residues 121–131) in the crystal structure (Davies et al., 1991) indicates that the C-terminus can be stabilized by external interactions, in this case presumably crystal packing. These observations have been further substantiated by the publication of the 3.5-Å resolution crystal structure of HIV-1 reverse transcriptase (Kohlstadt et al., 1992) which appeared following submission of this paper. Thus in the crystal structure, the C-terminal region of the RNase H domain, comprising helix  $\alpha_E$  and the loop between strand  $\beta_5$  and helix  $\alpha_E$ , interacts with the "thumb" region of the p51 domain. This interaction stabilizes the short polypeptide loop which is ill-defined in both the NMR and X-ray structures of the isolated RNase H domain and supports the required interaction between the p51 domain and the C-terminal helix  $\alpha_E$  to stabilize this helix in solution. The crystal structure of reverse transcriptase also supports the observation that the activity of the RNase H domain depends upon an ordered structure at its C-terminus.

#### SUPPLEMENTARY MATERIAL AVAILABLE

One table giving the values and errors of the experimental  $^{15}\text{N}$   $T_1$ ,  $T_2$ , and NOE relaxation data together with the best-fit values for the calculated relaxation data and the optimized values of the order parameters, effective correlation times, and  $T_2$  line-broadening parameters (22 pages). Ordering information is given on any current masthead page.

#### REFERENCES

- Becerra, S. B., Clore, G. M., Gronenborn, A. M., Karlstrom, A. R., Stahl, S. J., Wilson, S. H., & Wingfield, P. T. (1990) *FEBS Lett.* 270, 76–80.
- Carson, M. (1987) *J. Mol. Graphics* 5, 103–110.
- Chance, E. M., Curtis, A. R., Jones, I. P., & Kirby, C. R. (1979) *U.K. At. Energy Auth., Harwell Lab., [Rep.] AERE-R 8775*.
- Chandrasekhar, I., Clore, G. M., Gronenborn, A. M., & Brooks, B. R. (1992) *J. Mol. Biol.* (in press).
- Clore, G. M. (1983) in *Computing in Biological Science* (Geisow, M. J., & Barrett, M., Eds.) pp 313–348, Elsevier North-Holland, Amsterdam, The Netherlands.
- Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., & Gronenborn, A. M. (1990a) *J. Am. Chem. Soc.* 112, 4989–4991.
- Clore, G. M., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990b) *Biochemistry* 29, 7387–7401.
- Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R., & Matthews, D. A. (1991) *Science* 252, 88–95.
- di Marzo Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Orozlan, S., Gallo, R. C., & Sarngadharan, M. G. (1986) *Science* 231, 1289–1291.
- Doolittle, R. F., Feng, D.-F., Johnson, M. S., & McClure, M. A. (1989) *Q. Rev. Biol.* 64, 1–30.
- Evans, D. B., Brown, K., Deibel, M. R., Tarpley, W. G., & Sharma, S. K. (1991) *J. Biol. Chem.* 266, 20583–20585.
- Gilboa, E., Mitra, S. W., Goff, S., & Baltimore, D. (1979) *Cell* 18, 93–100.
- Goff, S. (1990) *J. AIDS* 3, 817–831.
- Hizi, A., Hughes, S. H., & Shaharabany, M. (1990) *Virology* 175, 575–580.
- Hostomsky, Z., Hostomska, Z., Hudson, G. O., Moomaw, E. W., & Nodes, B. R. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 1148–1152.
- Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J., & Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 7648–7652.
- Kamath, U., & Shriver, J. W. (1989) *J. Biol. Chem.* 264, 5586–5592.
- Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., & Morikawa, K., (1990) *Nature* 347, 306–309.
- Kay, L. E., Torchia, D. A., & Bax, A. (1989a) *Biochemistry* 28, 8972–8979.
- Kay, L. E., Marion, D., & Bax, A. (1989b) *J. Magn. Reson.* 83, 72–84.
- Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., & Torchia, D. A. (1992) *J. Magn. Reson.* 97, 359–375.
- Kohlstadt, L. A., Wang, J., Friedman, J. M., Rice, P. A., & Steitz, T. A. (1992) *Science* 256, 1783–1790.
- Kotewicz, M. L., Sampson, C. M., d'Alessio, J. M., & Gerard, G. F. (1988) *Nucleic Acids Res.* 16, 265–277.
- Krug, M. S., & Berger, S. L. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 3539–3543.
- Lide, D. R. (1990) *CRC Handbook of Chemistry and Physics*, pp 6–8, CRC Press, Boca Raton, FL.
- Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A., & Venkatesan, S. (1986) *J. Virol.* 60, 771–775.
- Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc.* 104, 4546–4559.
- Mizrahi, V. (1989) *Biochemistry* 28, 9088–9094.
- Oostra, B. M., Harvey, R., Ely, B. K., Markham, A. F., & Smith, A. E. (1983) *Nature* 304, 456–459.
- Palmer, A. G., Skelton, N. J., Chazin, W. J., Wright, P. E., & Rance, M. (1992) *Mol. Phys.* 75, 699–711.
- Powers, R., Clore, G. M., Bax, A., Garrett, D. S., Stahl, S. J., Wingfield, P. T., & Gronenborn, A. M. (1991) *J. Mol. Biol.* 221, 1081–1090.
- Repaske, R., Hartley, J. W., Karlick, M. F., O'Neill, R. R., & Austin, J. B. (1989) *J. Virol.* 63, 1460–1465.
- Restle, T., Müller, B., & Goody, R. S. (1992) *FEBS Lett.* 300, 97–100.
- Schatz, O., Cromme, F. V., Gruninger-Leitch, F., & Le Grice, S. F. J. (1989) *FEBS Lett.* 257, 311–314.
- Schatz, O., Mous, J., & Le Grice, S. F. (1990) *EMBO J.* 9, 1171–1176.
- Smith, G. M., Yu, P. L., & Domingues, D. J. (1987) *Biochemistry* 26, 2202–2207.
- Tanese, N., & Goff, S. P. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 1777–1781.
- Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H., & Nagayama, K. (1991) *Biochemistry* 30, 6036–6047.
- Yang, W., Hendrickson, W. A., Crouch, J., & Satow, Y. (1990) *Science* 249, 1398–1405.