Solution Structure of the RNase H Domain of the HIV-1 Reverse Transcriptase in the Presence of Magnesium†

Koteppa Pari, Geoffrey A. Mueller, Eugene F. DeRose, Thomas W. Kirby, and Robert E. London*

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, P.O. Box 12233, Research Triangle Park, North Carolina 27709

Received July 23, 2002; Revised Manuscript Received October 22, 2002

ABSTRACT: This paper presents the first solution structure of the RNase H domain of HIV-1 reverse transcriptase (RT) determined by NMR methods. The solution conditions in this study were at physiological pH in the presence of Mg^{2+}. An investigation of the dependence of the $^{1}H-^{15}N$ HSQC spectrum of the RNase H domain on [Mg^{2+}] indicates that Mg^{2+} produces significant, global effects on the amide chemical shifts, implying that divalent metal ion binding is important for stabilizing the structure of the isolated domain in solution. Analysis of amide shift data as a function of $MgCl_2$ concentration using either a single- or two-site binding model indicated that the latter provided a significantly improved fit, with the K_D for site A = 2.7−3.2 mM and K_D for site B ∼ 35 mM, calculated on the assumption that site A is already occupied. Resonances of the [U-^{13}C,^{15}N]RNase H domain, measured at pH 6.8, in 80 mM $MgCl_2$, were assigned and NOESY data collected in order to determine the structure. Assignment of the NOESY spectra using the ARIA program resulted in a high-resolution structure for residues 6−114 which was similar to the crystal structure of the isolated domain, 1HRH. The data were insufficient to define a compact structure for the C-terminal residues after 114. Residues I134−L138 located at the C-terminus are highly disordered and give rise to relatively sharp and intense amide resonances, while the amide resonances for the segment from E124 to A132 appear to be largely absent and are presumably subject to significant exchange broadening between different conformational states. Comparisons with crystal structure data for the full reverse transcriptase molecule indicate that the corresponding region is absent in nearly all of the crystal structures determined for the $P2_12_12_1$ space group, while these residues adopt an α-helix in structures determined for other symmetry groups. This structural heterogeneity indicates that significant conformational variability exists for this segment of the full reverse transcriptase enzyme as well, and the structure of the C-terminal peptide can be selected or deselected, depending on crystallization conditions. This analysis, along with the structural characterization contained herein, challenges the previous paradigm that the dynamic behavior of the isolated RNase H domain differs substantially from the behavior in the intact enzyme. The poor Mg^{2+} binding and conformational flexibility of residues located near the active site indicate that substrate binding is a precondition for metal ion binding and for selecting the active site conformation of the RNase H domain.

The reverse transciption of viral genomic RNA into DNA is an essential step in the life cycle of the human immunodeficiency virus (HIV)[1] (*1*−*3*). This process is catalyzed by a bifunctional enzyme, HIV reverse transcriptase (RT), which consists of a DNA polymerase domain and a ribonuclease H domain (RNase H). The RNase H domain degrades the RNA template and specifically removes the RNA primer. The domain can be isolated and expressed as a stable protein which, however, is inactive (*4*, *5*) or very weakly active (*6*, *7*).

Despite extensive crystallographic data for the full reverse transcriptase molecule as well as two studies of the isolated RNase H domain, the structural characterization of the RNase H domain remains incomplete. Crystallographic studies of the full reverse transcriptase enzyme from HIV reveal a surprising heterogeneity of the RNase H domain, which we have found to be strongly correlated with the crystal space group. In a majority of the structures reported for crystals in the $P2_12_12_1$ space group, the C-terminal helix containing Asp549 is not present, while this helix can be observed in crystals corresponding to other space groups (for example, refs *8*−*11*). These results indicate that the stability of this structural element is marginal, so that its observation depends on the crystallization conditions. There is also some variability in the loop connecting the first two β-strands of the RNase H domain, for example, 1VRU (*10*). In addition to crystallographic studies of the structure of the entire reverse transcriptase enzyme, there have been two published crystal-

† The authors gratefully acknowledge support from the NIEHS intramural AIDS program. K.P. is the recipient of the National Institutes of Health Visiting Fellowship.

* To whom correspondence should be addressed. Phone: (919) 541-4879. Fax: (919) 541-5707. E-mail: london@niehs.nih.gov.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; RNase H, ribonuclease H; RT, reverse transcriptase; RMSD, root mean squared deviation; NOE, nuclear Overhauser enhancement.

10.1021/bi0204894
Published on Web 01/01/2003

lographic studies of the isolated RNase H domain (*12*, *13*). In the structure of Davies et al. (*12*) residues 538−542, located in a loop near the active site, are missing. Similarly, Chattopadhyay et al. (*13*) concluded that the electron density corresponding to those residues does not represent a single, discrete conformation but arises from a combination of conformational states. Several NMR studies of the protein at acidic pH have also been performed, and relaxation measurements have indicated considerable dynamic heterogeneity for the isolated domain (*14*, *15*). In addition, the backbone resonances for only two of the four acidic residues in the active site of the enzyme, Asp443 and Glu478, were assigned (*14*−*16*) (numbering based on the full RT sequence). Hence, the structural characterization of the RNase H domain of HIV RT is incomplete at this point.

There exist several conflicting proposals as to the role and number of Mg^{2+} ions in the activity of RNase H (*17*−*21*). RNase H is known to require divalent metal ion(s) for activity, with the highest activity typically observed with Mg^{2+} ions (*22*). A mechanism involving two divalent metal ions, analogous to the mechanism of the Klenow exonuclease domain (*19*), has been proposed for the RT RNase H domain (*12*), while a mechanism involving a single Mg^{2+} ion has been proposed for the *Escherichia coli* enzyme (*18*, *23*). Crystallographic studies of both the *E. coli* RNase HI and the RT RNase H domain have shown that the active site can bind two Mn^{2+} ions separated by approximately 4 Å (*12*, *24*). However, only a single bound Mg^{2+} ion was identified in the crystal structure of the *E. coli* enzyme, even though it was crystallized in 90−100 mM $MgSO_4$ (*23*). Calorimetric analysis has indicated that, for the intact HIV-1 reverse transcriptase, only a single Mg^{2+} ion is bound in the RNase H catalytic center, leading to the proposal that Mg^{2+} binding to the two sites observed for the Mn^{2+} structures may be mutually exclusive (*21*). NMR titration studies of the *E. coli* RNase HI as a function of added Mg^{2+} yielded a dissociation constant of 4.8 ± 1.0 mM (*17*). Previous NMR studies of the RT RNase H domain have been performed at low pH and without added Mg^{2+} (*14*−*16*). One goal of this paper is to carefully analyze the influence of magnesium on the structure at physiological pH in order to better understand the reaction mechanism.

Since the RNase H activity is an absolute requirement for HIV-1 replication, it has been proposed as a potential target for antiviral chemotherapy (*25*). At present, there is no solution structure for the RNase H domain and no reported inhibitor complex with this domain. Given the potential chemotherapeutic significance of this domain of reverse transcriptase, we have initiated an NMR study of its solution structure.

MATERIALS AND METHODS

Numbering. In general, residues of the RT RNase H domain are referred to by numbering them 1−138, which includes four N-terminal MNEL residues followed by residues 427−560 of the HXB2 strain of HIV. Hence, adding 422 to our residue number gives the corresponding residue in the reverse transcriptase P66 subunit for strain HXB2. This facilitates comparisons with previous NMR studies (*14*−*16*). Where comparisons to the full-length RT are needed, it is carefully noted in the text which numbering scheme is being used.

Expression of HIV-1 RNase H. The RNase H expression plasmid was obtained from David Gorenstein at the University of Texas Medical Branch, Galveston, TX. The plasmid contains the coding region of HIV-1 RNase H (strain HXB2) plus four additional N-terminal residues: MNEL (*4*) cloned into the commercially available pET30a(+) vector (Novagen). The plasmid was transformed into *E. coli* strain BLR(DE3). Cells containing the plasmid were grown to mid-log phase ($A_{600} \sim 0.6$) at 37 °C in M9 minimal medium containing 50 μg/mL kanamycin, ^{15}N-labeled ammonium chloride, and either ^{13}C-labeled glucose or unlabeled glucose. RNase H protein expression was induced by addition of isopropyl thio-β-D-galactoside (IPTG) to 1 mM, and the growth was continued for 3 h. The cells were harvested by centrifugation and stored at −20 °C.

Purification. The harvested cells were resuspended in a 30 mL solution of 50 mM Tris-HCl (pH 6.8) and 1 mM NaN_3, with protease inhibitors added (100 μL Calbiochem protease inhibitor cocktail set III), and lysed by sonication in a Branson sonifier model 200 using a microtip probe at an output level of 6 for 10 × 30 s with 30 s cooling each cycle. The lysate was centrifuged at 30000*g* for 30 min, and then solid ammonium sulfate was added to the supernatant to 40% saturation. This was kept on ice for 30 min and then centrifuged at 30000*g* for 10 min. The supernatant was brought to 60% saturation with addition of solid ammonium sulfate, and again it was kept on ice for 30 min and centrifuged as before. The 60% ammonium sulfate precipitate, containing RNase H, was resuspended in a minimal amount of 50 mM Tris-HCl, pH 8.0, and the protein solution was applied to a 2.6 × 63 cm column of Sephacryl S-100 which was eluted with 50 mM Tris-HCl, pH 8.0. On the basis of SDS−PAGE analysis, fractions containing RNase H were pooled and loaded onto an FPLC column (2.6 × 15 cm) containing Q-Sepharose (Amersham). The column was washed with 50 mM Tris-HCl, pH 8.0, and RNase H was eluted with a linear gradient of 0−1 M NaCl in 50 mM Tris-HCl, pH 8.0. Fractions containing the pure protein were pooled and concentrated using a Centricon YM-3 filter unit (Millipore).

NMR Spectroscopy. After purification, the buffer was exchanged to 10 mM Tris-d_{11}, pH 6.8, by repeated dilution and concentration. The final NMR sample contained 1.1 mM RNase H in 10 mM Tris-d_{11}, pH 6.8, 0.1 mM AEBSF, a protease inhibitor, 50 μM DSS, as an internal chemical shift standard, and 10% D_2O for the lock.

All NMR experiments were performed at 25 °C on Varian INOVA 600 and 500 MHz NMR spectrometers. The spectra were processed with NMRPIPE (*26*) and analyzed with NMRVIEW (*27*) software on LINUX workstations running Red Hat 7.1. Molecules were visualized and aligned with MOLMOL (*28*).

Sequential assignments were made primarily from the HNCACB and CBCA(CO)NH experiments (*29*). In some cases, where the intensity was weak, analysis of the CN NOESY-HSQC with the scalar coupling experiments confirmed the assignments (*30*). $^1H-^{15}N$ and $^1H-^{13}C$ correlation spectra were acquired with Varian's gNhsqc and gChsqc (in constant time mode) experiments, respectively. Side chains were assigned primarily from the H(CCO)NH and (H)C(CO)NH TOCSY experiments (*31*, *32*), again with occasional assistance from analysis of the NOESY spectra. To assign

the aromatic side chains, (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE (*33*) and the aromatic region HSQC spectra were used. The carbonyl chemical shift assignments were obtained from the HNCO experiment (*29*, *34*). NOEs were analyzed from the CN NOESY-HSQC spectra and the 4D $^{13}C/^{13}C$-separated HMQC-NOESY-HMQC (*35*). The assignments were submitted to the BMRB, accession number 5347.

Metal Titration: Mg^{2+} Binding to RNase H. For titration experiments, the salt $MgCl_2 \cdot 6H_2O$ (Aldrich) was dissolved in 10 mM Tris-d_{11}, pH 6.8. The initial protein concentration was 1.1 mM. $^1H-^{15}N$ HSQC experiments were performed to monitor the changes that occur in the course of the titration. At each step of the titration, dissolved $MgCl_2$ was added, which diluted the concentration of protein at the final Mg^{2+} titration step to ~1 mM. The chemical shifts were measured relative to an internal DSS standard. Titration curves were fit to the equations indicated in the text for either the single- or two-site model using the nonlinear least-squares fitting routine provided in the statistics package of Mathematica (Wolfram Reseaerch, Champaign, IL). For each fit, a normalized root mean square deviation (RMSD) was calculated using the equation:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\left(\frac{\delta_{\mathrm{obs}} - \delta_{\mathrm{calc}}}{\delta_{80}}\right)^2} \qquad (1)$$

where N is the number of titration steps, δ_{obs} and δ_{calc} are the observed and calculated shift values, and δ_{80} is the measured shift at the maximum magnesium concentration of 80 mM used in the study. The use of δ_{80} as a normalization factor reduces the extent of variations among resonances that exhibit significantly different chemical shift sensitivity to the Mg^{2+}, so that the set of RMSD values is comparable throughout the data set.

Structure Calculations. The restraints used to calculate the RNase H structures included dihedral angles, hydrogen bonds, and NOEs. Results of the CSI predictions and TALOS were analyzed for secondary structure information (*36*−*38*). Dihedral angle restraints were based on the predictions of TALOS, which matches sequence and chemical shifts with a database of structures and chemical shifts. Using the criteria established by the authors of TALOS, "good" matches of dihedral angles were used for restraints, and errors were set ±2 times the standard deviation or at least ±20°. The hydrogen bonds were assigned on the basis of an analysis of manually assigned NOEs and similarly with the predicted regions of secondary structure by TALOS. In total 47 H-bonds were assigned, 25 in β-sheets and 22 in α-helices.

The program ARIA (version 1.0) was used to calculate structures as described below (*39*−*42*). Input data for ARIA included the previously mentioned dihedral and H-bond restraints and the CN NOESY peak lists. The peak lists were culled of artifacts due to water and due to apodization in the case of very intense methyl groups. The peak lists were unassigned and uncalibrated with respect to distance. While the Mg^{2+} stabilized the protein enough to collect triple resonance experiments, the NOESY spectra had relatively poor signal to noise compared to expectations for a protein with similar size and at similar concentrations. To improve the noise discrimination by ARIA, an initial model structure was provided, and the protocols were modified from the default parameters as described below.

The SWISS-PDB modeler generated the initial structure for assigning NOEs in the first iteration of the ARIA calculations. Given the large number of crystal structures of RNase H from various strains of HIV-1 either isolated or as part of the complete reverse transcriptase, we anticipated the model would be a very good initial structure. The model found 61 structures with 91−100% sequence identity, although none completely spanned our specific construct. We used the program MOLMOL to add residues to the N- and C-termini (*28*). Hydrogen atoms were added using the program CNS (*43*).

The default parameters for ARIA produced structures with poor convergence (data not shown). We optimized several parameters to provide better noise discrimination including the ambiguous cutoff (ρ), the violation tolerance (vtol), and maximum number of ambiguities per peak (maxn). We found the following scheme led to the best convergence over eight iterations: ρ = (0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.85, 0.8), vtol = (2, 1, 0.5, 0.5, 0.12, 0.1, 0, 0), maxn = 5 for all iterations. For calibrating the NOE distances the rotational correlation time, τ_c, was set to 10.2 ns as reported by Powers et al. (*14*). In some of the calculations residues directly involved in the metal binding site were restrained with pseudo-NOEs as detailed in the results. Four side-chain carbonyl distances were chosen for their similarity to the Mn^{2+}-bound states of *E. coli* RNase HI (1GI5) and RT RNase H (1HRH) (*12*, *24*). Using our NMR numbering, the distances were as follows: D21 Cγ to E56 Cδ, 4.8 Å; E21 Cγ to E76 Cγ, 4.7 Å; D21 Cγ to D127 Cγ, 5.2 Å; D76 Cγ to E56 Cδ, 5.4 Å. The errors on all these pseudo-NOE restraints were ±0.3 Å. To test the influence of these restraints on the convergence of the structures, we calculated structures with and without these pseudo-NOEs, as well as excluding only the D21 to D127 pseudo-NOE, which is part of the second binding site. The structure ensembles calculated using all four pseudo-NOE restraints have been submitted to the PDB, code number 1LKU.

RESULTS

Assignment of the NMR Spectra. Previous NMR studies of the RNase H RT domain were performed at low pH and in the absence of magnesium (*14*−*16*). It seemed likely that the Mg^{2+} binding interaction would stabilize the enzyme as a result of the interaction with the coordinating acidic residues. We thus began our studies by performing a Mg^{2+} titration at pH 6.8. An assigned $^1H-^{15}N$ HSQC spectrum of the [U-^{15}N]RNase H and a summary of the effects of Mg^{2+} on this spectrum are shown in Figure 1. The $^1H-^{15}N$ HSQC spectra showed a significant reduction of broad, overlapping resonances and a parallel improvement in resolution and uniformity of peak intensity as the Mg^{2+} concentration is increased (spectra obtained at 0 and 80 mM $MgCl_2$ are shown in Figure 1a). Although the shift perturbations resulting from Mg^{2+} binding are most pronounced for residues located near the Mg^{2+} binding site(s), significant shifts are observed throughout the protein (vide infra), suggesting that Mg^{2+} binding plays a significant role in stabilizing the isolated RT RNase H domain.

Reasonable spectral dispersion was achieved at 80 mM $MgCl_2$, and further NMR studies were conducted at this concentration. At this point the assignment was straightfor-

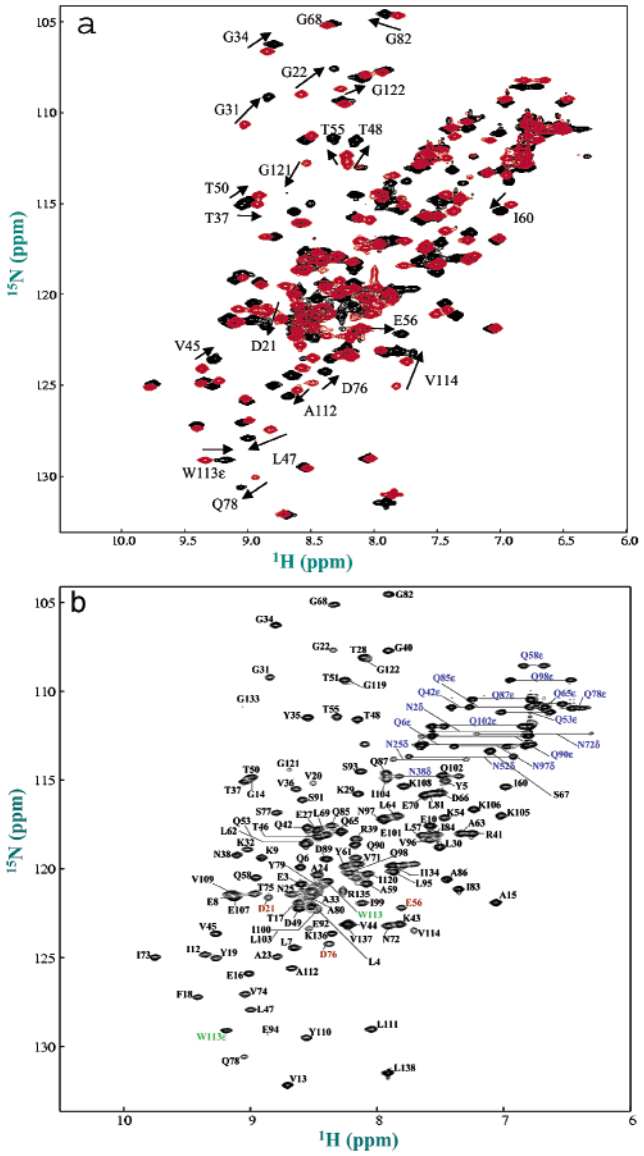

FIGURE 1: ^{1}H−^{15}N HSQC spectra of RNase H. (A) Two HSQC spectra at different concentrations of $MgCl_2$ are displayed, with red = 0 mM and black = 80 mM $MgCl_2$. Arrows indicate the direction of chemical shifts for a given residue with the head of the arrow closest to 80 mM $MgCl_2$. (B) Assignments of the backbone amide proton resonances are indicated on the spectrum (600 MHz, 25 °C, 10 mM Tris-HCl, pH 6.8, 80 mM $MgCl_2$, 10% D_2O, 90% H_2O). Pairs of signals corresponding to side-chain amino groups of asparagines and glutamines are connected by solid horizontal lines and labeled with blue text. The labels of metal binding residues are in red and W113 is in green.

ward, as described in Materials and Methods. Figure 1b also shows the ^{1}H−^{15}N HSQC spectrum of RNase H in the presence of 80 mM $MgCl_2$. As can be seen in this spectrum, most backbone amide resonances are resolved. Of the number of signals expected for the full range of sequence in this protein, we confidently assigned 121 of the 136 backbone amide resonances, the indole Nε-H of Trp113, and 18 of the 19 expected side-chain NH_2 protons for 7 Asn and 12 Gln residues. Other than the proline residues, the only contiguous unassigned segments were residues Ala116−His117, corresponding to two of the active site residues missing in the crystal structure (*12*), and residues Glu124 through Ala132, corresponding approximately to the C-terminal α-helix in 1HRH. The amide resonance for Gly133 is weak but observable (Figure 1b).

Titration. HSQC spectra were collected at various $MgCl_2$ concentrations up to 80 mM. The chemical shift data were analyzed for information on the binding of Mg^{2+}. Residues were chosen for analysis on the basis of good spectral dispersion and clear assignments at 0 mM $MgCl_2$. Titration curves for four residues, Gly34, Gly68, Val45, and Asp21, are shown in Figure 2. The theoretical curves for the binding of Mg^{2+} were derived on the assumption of fast exchange between Mg^{2+}-complexed and uncomplexed species:

$$\delta_{obs} = p_F\delta_F + p_B\delta_B \quad (2)$$

where δ_{obs} is the observed chemical shift, δ_B and δ_F refer to the chemical shifts, and p_B and p_F refer to the fractional concentrations of the bound and free species, respectively. For a simple 1:1 enzyme:magnesium binding stoichiometry defined by a dissociation constant K_D

$$K_D = \frac{[Mg][E]}{[MgE]} = \frac{[Mg_T - MgE][E_0 - MgE]}{[MgE]} \quad (3)$$

we obtain for the relative shift perturbation

$$\delta_{obs} - \delta_F = (\delta_B - \delta_F)[Mg_T + K_D + E_0 - \sqrt{(Mg_T + K_D + E_0)^2 - 4Mg_TE_0}]/2E_0 \quad (4)$$

where E_0 is the total RNase H concentration present and Mg_T is the total added Mg^{2+} concentration. Although it is apparent from Figure 2 that we can obtain reasonable fits for the shifts of the four residues shown using a model with a single Mg^{2+} bound, it also is apparent that (1) the K_D values (Table 1, first column) show a significant variation among these residues, for example, the K_D value obtained for Gly82 is three times larger than the value for Gly68, and (2) in some cases, such as the curve shown for Gly68, the shifts seem to reverse at higher Mg^{2+} concentrations. For nearly all of the residues, the single-site binding curves show an error pattern similar to that for Gly34, with the observed shift values initially falling below the calculated curve, then above the curve, and, at the highest Mg^{2+} concentration, below the curve again.

We subsequently investigated the analysis of the Mg^{2+} titration data on the basis of a model that allows Mg^{2+} binding to two sites, A and B, according to

$$\delta_{obs} = p_F\delta_F + p_A\delta_A + p_B\delta_B + p_{AB}\delta_{AB} \approx p_F\delta_F + p_A\delta_A + p_{AB}\delta_{AB} \quad (5)$$

where the second approximation is based on the assumption that binding to site B is sufficiently weak so that the concentration of enzyme with Mg^{2+} bound only to site B is negligible. Then, using the equilibria

$$K_D^A = \frac{[Mg][E]}{[MgE_A]} \qquad K_D^{BA} = \frac{[MgE_A][Mg]}{[Mg_2E_{AB}]} \quad (6)$$

where K_D^A is the dissociation constant for binding to site A and K_D^{BA} is the conditional dissociation constant for Mg^{2+} binding to site B if site A is already occupied. The observed

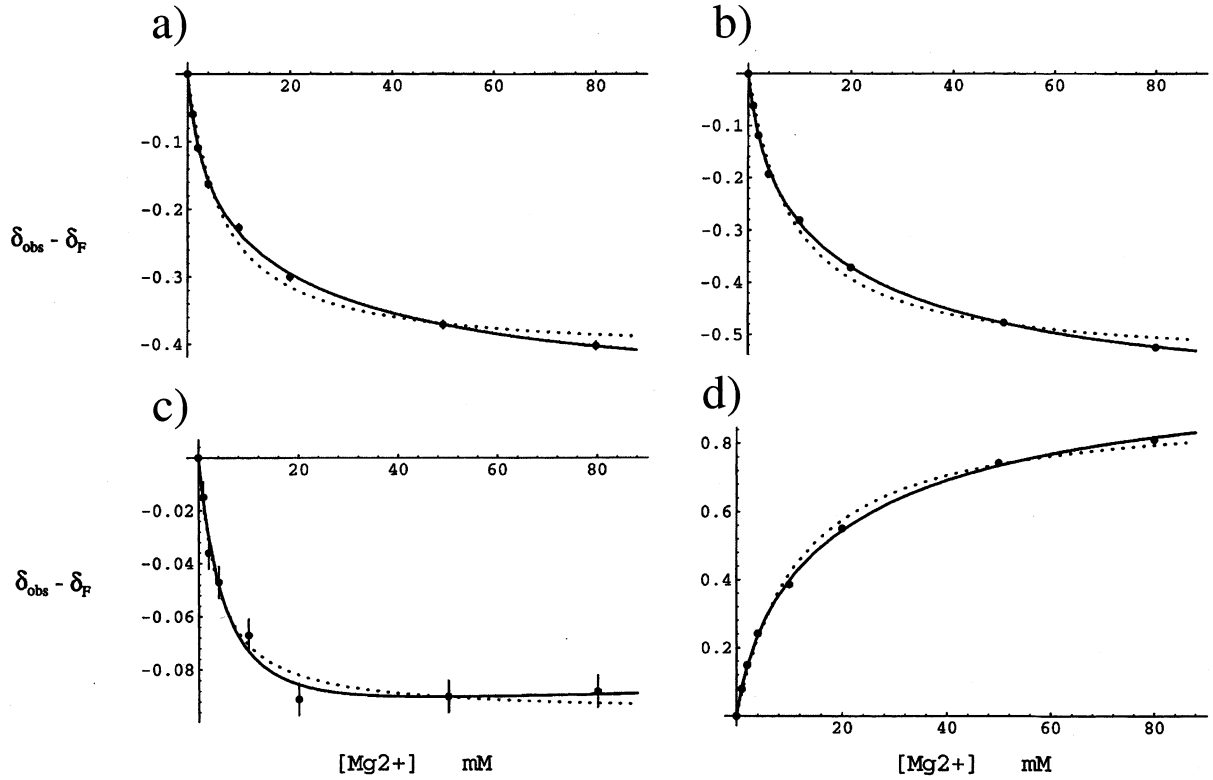

FIGURE 2: Nitrogen-15 shifts of the amide resonances indicated as a function of added $MgCl_2$. Panels a, b, c, and d correspond to the chemical shift differences during the titration for residues Gly34, Val45, Gly68, and Asp21, respectively. The dotted curve was fit using a single-site model (eq 4), while the smooth curve was fit using a two-site model (eq 7). The error bars correspond to the digital resolution in the ^{15}N dimension of ±6 ppb, which is generally too small to be seen.

shifts, referenced to the uncomplexed (δ_F) values, are then given by

$$\delta_{obs} - \delta_F = \frac{\frac{\Delta_A[Mg]}{K_D^A} + \frac{\Delta_{AB}[Mg]^2}{K_D^A K_D^{BA}}}{1 + \frac{[Mg]}{K_D^A} + \frac{[Mg]^2}{K_D^A K_D^{BA}}} \quad (7)$$

where $\Delta_A = \delta_A - \delta_F$ and $\Delta_{AB} = \delta_{AB} - \delta_F$. In general, the data are insufficiently precise to allow simultaneous fitting of the four parameters: K_D^A, Δ_A, K_D^{BA}, and Δ_{AB}. Thus, the data were initially fit using large values of K_D^{BA} in the range 25−400 mM. In general, there was significant covariation of the parameters with K_D^{BA}. Noting that the fitted shifts of Gly82 and Asp21 tended to be more significantly affected by the (assumed) weaker binding to site B, i.e., $|\Delta_{AB}|/|\Delta_A| \sim 3$, these data were fit setting K_D^A at the initially calculated mean value and allowing K_D^{BA} to vary. The best-fit value then corresponded to $K_D^{BA} = 35$ mM. A reasonably self-consistent set of results for K_D^A was obtained by fixing K_D^{BA} to 35 mM and refitting K_D^A. Under these conditions $K_D^A = 3.68 \pm 0.77$, with Δ_A and Δ_{AB} varying for each residue depending on the shift sensitivity of the particular resonances to Mg^{2+} binding to sites A and B. We note that it is generally not possible to reduce K_D^{BA} significantly, say by a factor of 2, without losing the improvement that results from having a second, weaker binding site. Alternatively, equally good fits can be obtained by increasing K_D^{BA} by an order of magnitude; however, the values of Δ_{AB} obtained also increase by large amounts since the shift contribution of the second binding site declines with its lower occupancy. Values of $\Delta_{AB} \gg \Delta_A$ for all residues are physically unreasonable. Thus, although the curves are not too sensitive to the exact value of K_D^{BA}, the value of 35 mM represents a reasonable estimate. In general, the fits for the titration shifts are significantly improved relative to the single binding site model. This improvement is indicated, for example, by the reduction in the RMSD values (Table 1) and by a comparison of the fits shown in Figure 2.

The values of K_D^A and K_D^{BA} obtained above using the two-site model are subject to a systematic overestimation since in fitting the data with eq 6 we set $[Mg] = Mg_T$, the total added Mg^{2+} concentration. To correct for this effect, we recalculated the free Mg^{2+} concentration using the initially obtained K_D^A and K_D^{BA} values according to

$$Mg^{2+} = Mg_T - [MgE_A] - [Mg_2E_{AB}] \approx Mg_T - (1/2)[Mg_T + K_D^A + E_0 - \sqrt{(Mg_T + K_D^A + E_0)^2 - 4Mg_T E_0} + Mg_T + K_D^{BA} + E_0 - \sqrt{(Mg_T + K_D^{BA} + E_0)^2 - 4Mg_T E_0}] \quad (8)$$

The above relation provides an approximate correction for the Mg^{2+} complexed to the enzyme, resulting in a limiting decrease in the free Mg^{2+} concentration of 2 mM Mg^{2+} for $Mg_T \gg K_D^A$, K_D^{BA}. The data were then fit again using the calculated, free Mg^{2+} values instead of the Mg_T values. For this fit, we again set $K_D^{BA} = 35$ mM and obtained $K_D^A = 2.71 \pm 0.66$ mM. In general, the RMSD values were similar to those obtained using the Mg_T concentrations (Table 1).

Table 1: Calculated Parameters for Mg^{2+} Binding Models

	model I: single Mg^{2+} site[a]			model II: two Mg^{2+} sites[b]				model II: recalculated Mg^{2+} K_Ds[c]	
residue	K_D (mM)	Δ_A (ppm)	RMSD $\times 10^{-2}$	K_D^A (mM)	Δ_A (ppm)	Δ_{AB} (ppm)	RMSD $\times 10^{-2}$	K_D^A (mM)	RMSD $\times 10^{-2}$
				(I) ^{15}N Titration Data					
G34	6.11	−0.415	3.35	2.68	−0.226	−0.486	1.05	1.93	1.10
G68	3.34	−0.096	4.90	5.55	−0.108	−0.083	4.19	4.13	4.38
G82	11.36	−0.103	2.89	4.01	−0.039	−0.119	1.39	3.13	1.38
G31	5.51	−1.570	2.57	3.27	−1.032	−1.746	1.15	2.49	1.10
G22	5.88	−1.487	1.87	4.05	−1.05	−1.600	1.06	3.05	1.08
T55	4.86	−1.413	2.61	2.99	−0.957	−1.577	0.90	2.27	0.91
T48	5.74	−0.876	2.65	3.55	−0.585	−0.966	1.68	2.69	1.65
I60	5.55	0.360	2.35	3.65	0.242	0.399	1.12	2.46	1.28
				(II) ^{1}H Titration Data					
W113ϵ	5.08	−0.172	2.85	3.07	−0.110	−0.196	1.21	2.20	1.24
Q78	4.24	0.116	2.24	3.69	0.092	0.123	1.72	2.55	1.78
L47	3.27	0.169	5.15	3.67	0.129	0.195	1.20	2.56	1.28
V114	7.40	0.142	3.04	2.79	−0.071	−0.166	0.67	1.94	0.84
E56	6.82	−0.349	5.45	5.01	−0.247	−0.369	5.35	3.81	5.26
D76	2.83	−0.110	2.27	3.17	−0.102	−0.111	2.19	2.27	2.23
A112	3.68	0.066	6.04	4.47	0.049	0.072	2.22	3.34	2.22
mean ± SD	5.93 ± 2.46			3.68 ± 0.77				2.71 ± 0.66	

[a] Values calculated from fitting titration data to eq 4. [b] Values calculated from fitting titration data to eq 7. K_D^{BA} was set at 35 mM for these fits. [c] Values recalculated using eq 7 and free $[Mg^{2+}]$ values corrected using eq 8. The Δ_A and Δ_{AB} values for this model are very similar to those determined without the correction for free $[Mg^{2+}]$ and hence have been omitted from the table.

Structure Determination. The structure determination proceeded as described in Materials and Methods. After assignment of the ^{1}H, ^{13}C, and ^{15}N resonances as described in Materials and Methods, manually assigned NOEs and the results of TALOS were analyzed to determine where hydrogen bonds would be in the β-sheets and α-helices. Dihedral restraints were based on the results of TALOS. NOE peak lists from the CN NOESY experiment were provided without assignments for structure calculations with ARIA. An initial model structure from the SWISS-PDB modeler was provided to aid in the assignments in the first round of structure calculations. In all subsequent rounds of structure calculations the input data included hydrogen bonds, dihedral restraints, the chemical shift assignments, and the unassigned NOE peak lists.

On the basis of our titration studies with Mg^{2+}, we calculated that in our solution of 80 mM Mg^{2+} there are three forms of the protein. On the basis of the K_D values determined herein, the percentage of RNase H with two Mg^{2+} bound to sites A and B is 68%, one Mg^{2+} bound to site A is 31%, and no Mg^{2+} bound is 1%. To simulate these conditions, we calculated structures with and without pseudo-NOEs restraining the side chains of the binding site residues. The four residues that are important for coordinating the divalent ions in our number scheme are Asp21, Glu56, Asp76, and Asp127. The first three residues coordinate site A. Three pseudo-NOEs restrained site A. The position of Asp127 was restrained using a single pseudo-NOE to Asp76, effectively creating the B ion binding site. Distances between the side chains were based on the distances observed for *E. coli* and HIV-1 RT RNase H with Mn^{2+} bound (*12*, *24*). As a control we took out all four pseudo-NOEs and recalculated the structures.

Residues 6−114 comprise a well-structured domain that was determined to high precision. Qualitatively, the structures calculated for these residues in each case are very similar to each other and have similar precision, regardless of the inclusion of pseudo-NOE restraints on the active site residues (Table 2). Table 2 shows the RMSD of secondary structure elements to be very low for all of the calculations. This can be visualized in Figure 3a, where the backbone traces of the seven lowest energy structures are shown for the case of all pseudo-NOEs included. The high precision of core residues 6−114 can be seen in Figure 3b when all of the heavy atoms, including the side chains, are visualized. The high quality of the structures is evident from the high percentage of residues in the most favored region of Ramachandran space and from the low number of NOE, H-bond, and dihedral angle violations and the low experimental RMSD (Table 2). We note that, in calculated structures for the full protein, the C-terminal residues are fairly disordered; however, this disorder is limited by NOE restraints for several C-terminal residues which bring them into closer proximity with the central β-sheet (Figure 3a). Manual examination indicates that many if not all of these NOE restraints are questionable, so the degree of disorder of the C-terminal segment may be greater than the calculations suggest. Figure 3c shows the ribbon diagram of the average minimized structure from Figure 3a aligned with the crystal structure 1HRH shown separately in Figure 3d. The precision among the average structures in Figure 3a (1.6 Å mean RMSD) is lower than the alignments with 1HRH (2.3 Å mean RMSD). Figure 4 shows how the RMSD of these alignments varies by residue. Plotted are the precisions among the average structures by residue, as well as a comparison between the average structures and 1HRH. The trend in the precision among the average structures is similar to the trend in the comparison with 1HRH. Concomitantly, the largest differences between the crystal structure and the NMR structures are primarily in loop regions and in segments where assignments are sparse, which is where the NMR structures are not well determined.

Effect of Magnesium on the Structures. A plot of the chemical shift change for each residue in the presence and absence of Mg^{2+} can assess the effect of magnesium binding on the structures. Figure 5 shows the absolute value of the

Table 2: Statistics of Structural Ensembles for the RNase H Domain of HIV-1 Reverse Transcriptase

	with all metal restraints	without D127 restraints	without all metal restraints
NOEs[a]			
H−CH	2648	2642	2638
H−NH	707	712	706
HC−CH	93	89	86
unambiguous	1890	1970	1962
ambiguous	1037	987	931
total	2927	2904	2983
ensemble RMSD (Å)			
secondary structure (backbone)[a]	0.40 ± 0.16	0.54 ± 0.22	0.69 ± 0.30
secondary structure (heavy)[a]	0.87 ± 0.38	1.14 ± 0.47	1.20 ± 0.49
backbone (residue 6−114)[b]	0.60 ± 0.15	0.77 ± 0.16	0.82 ± 0.16
heavy atoms (residues 6−114)[b]	1.25 ± 0.17	1.45 ± 0.18	1.56 ± 0.17
all backbone[a]	1.41 ± 0.61	2.78 ± 1.56	2.00 ± 0.93
all heavy[a]	1.92 ± 0.80	3.16 ± 1.63	2.56 ± 1.11
violations (experimental restraints)[a]			
NOEs and H-bonds	1.3	3.0	4.0
dihedral angles	2.0	1.7	2.7
RMSD (experimental restraints)[a]			
NOEs (Å)	0.035 ± 0.001	0.057 ± 0.043	0.042 ± 0.012
H-bonds (Å)	0.051 ± 0.006	0.145 ± 0.117	0.139 ± 0.110
dihedral angles (deg)	1.0 ± 0.2	1.3 ± 0.9	1.5 ± 0.4
RMSD (covalent geometry)[a]			
bonds (Å)	0.0030 ± 0.0002	0.0028 ± 0.0003	0.0029 ± 0.0003
angles (deg)	0.4615 ± 0.0258	0.4411 ± 0.0339	0.4632 ± 0.0338
impropers	0.3744 ± 0.0176	0.3647 ± 0.0650	0.3814 ± 0.0357
Ramachandran space[c]			
most favored (%)	67.2 ± 1.3	68.6 ± 1.3	66.7 ± 3.1
additionally allowed (%)	27.0 ± 1.9	23.8 ± 2.1	27.4 ± 2.9
generously allowed (%)	4.8 ± 2.0	4.3 ± 1.2	4.0 ± 0.5
disallowed (%)	0.9 ± 0.1	3.3 ± 1.5	1.9 ± 1.1

[a] Output by ARIA (*39*), calculated by CNS (*43*) using the ensemble of the 14 lowest energy structures. [b] Calculated with the program MOLMOL (*28*). [c] Calculated with PROCHECK (*52*).

chemical shift change versus residue for the ^{15}N shifts (Figure 5a) and the ^{1}H shifts (Figure 5b). Recall that a number of assignments are missing in the C-terminus. In Figure 5 there are clusters of large shift changes near the metal binding residues 21, 56, and 76. Unfortunately, there are no data for residues near 127. There are also large shift changes near the N-terminus and elsewhere. The magnitudes of the ^{15}N shift changes are color coded on the lowest energy RNase H structure calculated with all the pseudo-NOEs in Figure 6. Most of the large changes occur for residues located near the metal binding residues or near the N-terminus. However, at the top of molecule in this orientation there are many residues with very little change. Specifically this includes residues 10−20 and 60−75. In conclusion, there are significant global effects on the structure in the presence of magnesium as evidenced by the many residues exhibiting large chemical shift changes and the disperse positioning of residues with chemical shift changes on the structure.

DISCUSSION

Sequential Assignments and Relations to Other Structural Studies. Previously reported NMR studies of the RNase H domain of RT have been performed in the absence of Mg^{2+} at low pH (*14*−*16*). Only backbone resonances were assigned, and no solution structure was reported. Assignments for only two of the four acidic residues in the active site have been determined. In the present studies, we were interested in the behavior of the protein under more physiological conditions. We first titrated the sample with $MgCl_2$ at pH 6.8. The ^{1}H−^{15}N HSQC spectra show a global, general improvement as the Mg^{2+} concentration is increased, with no asymptotic limit apparent at $[Mg^{2+}] = 80$ mM. Although the shift perturbations resulting from Mg^{2+} binding are most pronounced for residues located near the Mg^{2+} binding site(s), significant shifts are observed throughout the protein (Figures 5 and 6), suggesting that Mg^{2+} binding plays a significant role in stabilizing the isolated RT RNase H domain. Powers et al. (*14*, *15*) reported difficulty in working with a construct derived from the same HXB2 strain used in the present work and reported that a W113A mutation resulted in substantial improvement of spectral parameters. Kern et al. (*16*) studied the RNase H domain derived from strain BH10 at low pH and did not specifically report this problem. They still did not assign the backbone resonances of Trp110 and Val111, although they did provide an assignment for the Nϵ-H of Trp110 (corresponding to Trp113 in our numbering scheme). The RNase H domain of BH10 has only six amino acid substitutions relative to the domain from HXB2. At neutral pH and in the presence of added Mg^{2+}, we did not encounter problems with Trp113. It is possible that the slow exchange behavior of this residue affects the shifts of nearby residues in the active site and, in particular, is one of the factors that prevented assignment of residues near His117. However, the corresponding resonances have been assigned in the BH10 construct (*16*). Interestingly, the backbone resonances for only two of the four acidic residues in the active site were assigned in these earlier studies (*14*−*16*), which lacked assignments for the third and fourth active site aspartyl residues in the sequence, Asp76 and Asp127. However, Asp76 was assigned in the

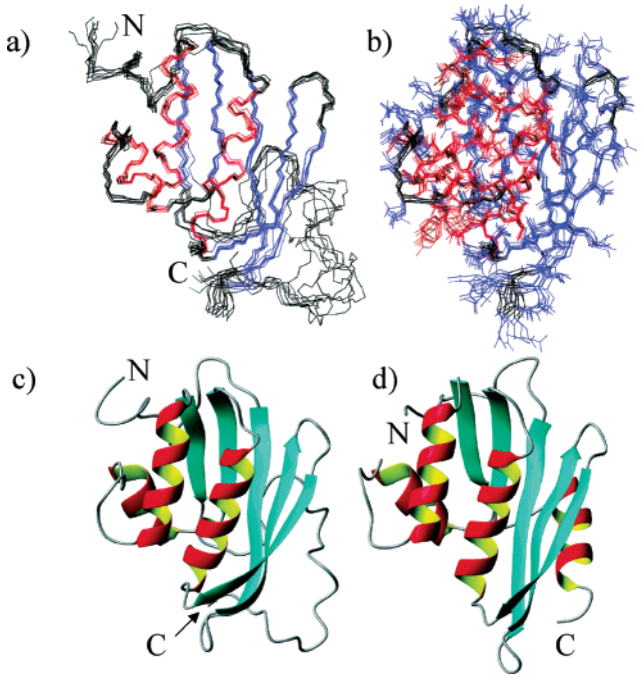

FIGURE 3: RNase H structures. Panel a shows a backbone atom trace for the seven lowest energy structures for the case where all the pseudo-NOEs restraining the metal binding sites are included. β-Strands are colored blue and α-helices are colored red. Panel b shows the heavy atoms, including the side chains, of residues 6−114 aligned similarly to those in (a). Panels c and d show ribbon diagrams of the average minimized NMR structure and the crystal structure 1HRH (*12*), respectively. The structures in (c) and (d) are aligned but drawn separately for ease of comparison. In (a), (c), and (d) the N- and C-termini are labeled.

present study. Since Asp76 is one of the Mg^{2+} binding residues, it is likely that the conformational heterogeneity near this residue is reduced in the presence of Mg^{2+}, making a more complete assignment possible.

The primary region with missing sequential assignments corresponds to residues Glu124 through Ala132 located in the C-terminal helix. The amide peak for Gly133 is very weak, but observable. Residues Ala116 and His117 were not assigned in the present study, and a longer sequence, Ala113−His117, was similarly not assigned in the study by Powers et al. (*15*). Importantly, residues I134−L138 have been assigned, and the amide resonances are relatively sharp and exhibit shifts typical of unfolded proteins. Thus, these residues located at the C-terminus of the protein appear to be relatively unstructured. In contrast, the apparent absence of resonances for Glu124 through Ala132 is not consistent with this type of disorder but rather with a slower conformational exchange process which probably samples the C-terminal helix observed in the crystal structures (*12*) as well as other conformational states with sufficient lifetimes (microsecond to millisecond) to significantly broaden the resonances.

To further understand the results of the NMR structural data, we have examined the crystallographic data reported for both the isolated RNase H domain and the domain in the full RT structures. We found two reported studies of the isolated domain by Davies et al. (1HRH; *12*) and Chattopadhyay et al. (1RDH; *13*). Active site residues 538−542 are missing in the 1HRH, as are data for many of the side chains. The 1RDH structure provides only α-carbon coordinates; however, the authors similarly were unable to construct a model for residues 538−542 that fit the electron density and conformed with all of the geometric constraints imposed. They concluded that the observed density does not represent a single discrete conformation but arises from a combination of conformational states (*13*).

Analysis of 55 reported crystal structures of HIV RT revealed an unanticipated heterogeneity in the RNase H domain. It was found that this heterogeneity was strongly correlated with the space group of the crystal. In particular, in nearly all of the $P2_12_12_1$ crystals, significant portions of the C-terminal segment containing active site residue Asp549 are missing, while this region was observed in most of the crystals corresponding to other space groups, although the *B*-factors for this region are often very high, greater than 100. Exceptions to this generalization occur for $P2_12_12_1$ crystals with two RT molecules in the unit cell. Analysis of intermolecular interactions revealed that the side chains from Lys82, Arg83, Met16, and other residues would interfere with the C-terminal helix of the RNase H domain, so that $P2_12_12_1$ crystals are sterically incompatible with this helix. In contrast, crystals from the *C*2, $P3_212$, and other space groups were characterized by intermolecular interactions in this region that potentially would stabilize the C-terminal helix. These interactions are similar to those present in the crystal of the isolated RNase H domain (*12*). In particular, the C-terminal helices of the RNase H domain from pairs of RT molecules or from the two RNase H domains in the unit cell of structure 1HRH are located in a close, antiparallel arrangement with some potentially stabilizing intermolecular H-bonding involving Gln464. Further, the β-strands of residues 465−467 on two molecules appear to form an extended β-sheet, with H-bonding interactions between Val466 and Gln464 on different monomers. This interaction is also observed between the two RNase H domain molecules in the unit cell in the structure of the isolated RNase H domain (1HRH; *12*). In addition to the intermolecular steric conflicts with the C-terminal helix, intermolecular steric conflicts with residues in the loop connecting β-strands 1 and 2 (nomenclature from ref *12*) are observed in some structures. For example, in 1VRU (*10*), residues 444−454 are not observed. Analysis of the structure indicates that Tyr354, Arg356, and Trp337 on a second RT molecule interfere with residues Lys451, Asn447, and Glu449 in this loop of the RNase H domain. We note a large difference in the RMSD values for residues 25−29 in this loop in our calculated structures with respect to the 1HRH crystal structure (Figure 4).

We interpret the above results to indicate that residues located in the C-terminal helix of the RNase H domain of RT are subject to significant conformational heterogeneity. The ensemble of crystal structures that have been reported reflects this disorder. Solution conditions that favor structures with a disordered C-terminus tend to result in crystals in the $P2_12_12_1$ space group, while conditions which favor the C-terminal helix tend to produce crystals in the *C*2 space group (for examples, see refs *8*−*11*). This is illustrated schematically in Figure 7. Furthermore, as noted above, even for crystals in the *C*2 space group in which the C-terminal helix can be observed, this region of the protein frequently exhibits *B*-factors greater than 100. These comparisons indicate that the equilibria of the C-terminal segment of the

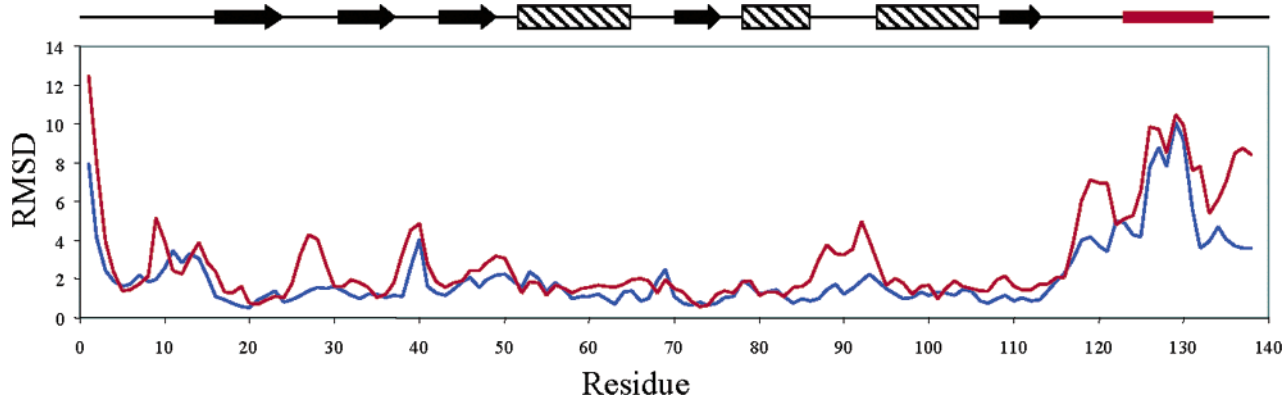

FIGURE 4: RMSD by residue. The precision among the average calculated structures for each pseudo-NOE case is shown in blue. The mean RMSD of each of the average structures to 1HRH is shown in red. The secondary structure of 1HRH is diagramed above, where arrows represent β-strands and diagonally shaded rectangles represent α-helices. The line is colored red where assignments are missing.

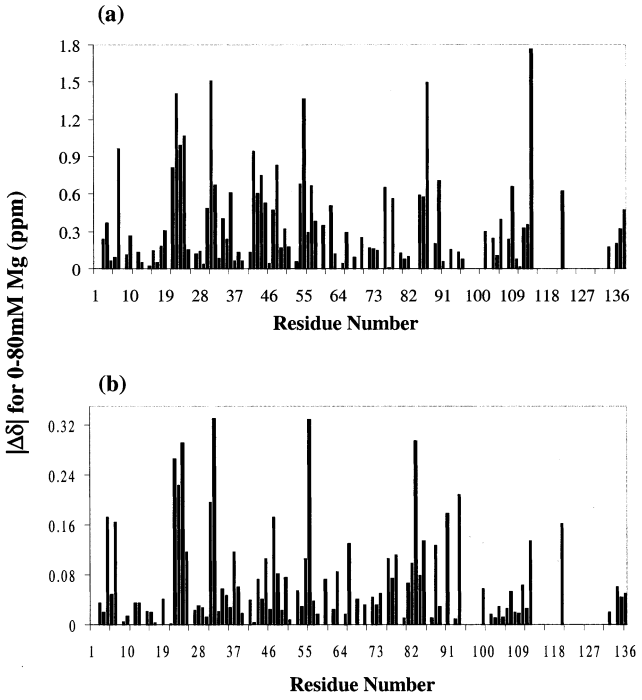

FIGURE 5: Chemical shift changes due to Mg^{2+} binding. The absolute values of the chemical shift changes going from $[Mg^{2+}] = 0$ to 80 mM are plotted for ^{15}N (a) and ^{1}H (b).

RNase H domain among various conformations are not a unique feature of the isolated RNase H domain but also characterize the domain when it is part of the full reverse transcriptase molecule. The equilibrium reflected in the ensemble of crystal structures is mirrored in the conformational exchange broadening of the resonances of the corresponding residues. This conclusion is consistent with the fact that there appear to be no significant intramolecular contacts between the C-terminal helix and the remainder of the RT molecule. Hence, the only factors likely to stabilize this domain are the binding of the active site Mg^{2+} ions and the interaction with substrate. Some of these interactions are apparent in structures with various substrate analogues. For example, in the crystal structure 1RTD showing the full RT complexed with a double-stranded DNA template-primer (*8*), invariant His539 is located in reasonable proximity to a phosphate group on the DNA backbone; i.e., His539 Nϵ2 and Mg602 are within 4 Å of a nonbridging oxygen on the phosphate backbone (A23 O1P).

Role of Magnesium. The nature of the interaction of RNase H with divalent metal ions and the role of these ions in catalysis are unclear. Crystallographic studies of both the *E. coli* RNase HI and the RT RNase H have shown that two Mn^{2+} ions separated by about 4 Å are bound at the active site (*12*, *24*). However, in the absence of additional ligands, the *E. coli* RNase HI binds only a single Mg^{2+} ion (*23*). As discussed above, we can obtain moderate fits of the individual resonance titrations, but the resulting K_D^{Mg} values show a large dispersion (mean = 5.8 ± 2.6 mM). Alternatively, the use of a two-site model results in a significant improvement of the fits (Figure 2), with the lower K_D^{Mg} value corresponding to binding to site A showing a significantly lower standard deviation (Table 1). Hence, the present titration data support the conclusion that, for RT RNase H, two Mg^{2+} ions can bind at the active site in the absence of nucleic acid substrate. Since the shift perturbations involve most of the same residues, it is concluded that the two Mg^{2+} sites are similar to although possibly not identical with the two Mn^{2+} binding sites. Recent studies of RT and the E478Q and D549N mutants found evidence for three distinct Mg^{2+} binding modes (*44*). This result would be consistent with two binding sites in the RNase H domain as found in the present study and an additional binding site in the polymerase domain.

While the measured K_D^{Mg} for RT RNase H seems very high and would require supraphysiological concentrations of Mg, it is similar to measured values for *E. coli* RNase HI. Oda et al. (*17*) have studied the $^{1}H-^{15}N$ HSQC spectrum of *E. coli* RNase HI and found that the chemical shift changes are largely localized to residues near the Mg^{2+} binding site, although significant shifts were also observed for Gln152 and Val153 (see Figure 5 of ref *17*). The latter shifts were proposed to result from chloride binding near these residues. As in this study, Oda et al. observed significant shifts out to high (100 mM) Mg^{2+} concentrations but obtained a relatively low K_D^{Mg} = 4.8 ± 1.0 mM. We suggest that the K_D^{Mg} calculated in this case is too high, as in our single-site model, because the data correspond to a weighted average for two separate binding sites. The analysis may reconcile the differing K_D^{Mg}, but the problem remains that even for two-site binding the calculated dissociation constants indicate that site A is partially occupied ($p_A = 0.12$) and site B ($p_{B/A} = 0.014$) essentially unoccupied at physiological levels of Mg^{2+} of about 0.5 mM (*45*−*47*). The most likely interpretation of

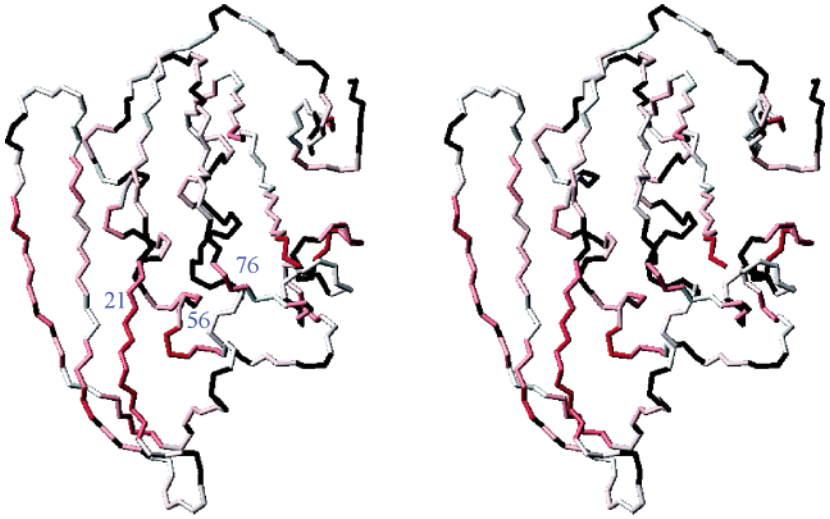

FIGURE 6: ^{15}N chemical shift changes due to Mg^{2+} mapped to the structure of residues 6−114, shown as a stereopair. Shift changes were color-coded on the basis of the difference between 0 and 80 mM $MgCl_2$, where red is the maximum change which fades to white when there is no change. Black indicates no data. Metal binding site residues are labeled.

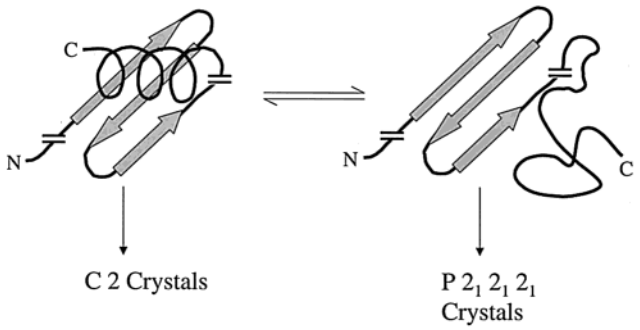

FIGURE 7: Schematic illustration of the equilibrium representing HIV reverse transcriptase structures in which the C-terminal segment of P66 containing the RNase H domain is either disordered or folded into an α-helix. Crystallization conditions which favor the disordered structure lead to the formation of $P2_12_12_1$ crystals, while conditions favoring the α-helical conformation adopt other space groups. When formed, the α-helix lays across β-strands 1, 2, and 3. The double bars indicate a break in the sequence of secondary structure elements.

these results is that the enzyme relies on the substrate to bring along the catalytic Mg^{2+} ion(s) or, at least, to complete the Mg^{2+} binding site so that Mg^{2+} can bind after substrate binding.

Therefore, we suggest that, in many respects, the present study and other data reported on the Mg^{2+} dependence of the RNase H activity indicate that this enzyme may behave similarly to inositol monophosphatase, an enzyme that catalyzes the hydrolysis of inositol monophosphates as well as ribonucleoside 2′-monophosphates and also binds two Mg^{2+} ions with four acidic residues. This enzyme is reported to have an absolute requirement for two divalent metal cations such as Mg^{2+} or Mn^{2+} (*48*, *49*), and it has been established that one metal ion binds to the enzyme prior to substrate binding and the second ion binds after substrate binding (*49*, *50*). The enzyme is also inhibited by high Mg^{2+} concentrations (*51*), as recently demonstrated for *E. coli* RNase H (*24*). This inhibition has been interpreted as resulting from the binding of the second Mg^{2+} to the product phosphate, slowing down the dissociation reaction. The high K_D^{Mg} for RT RNase H is consistent with this reaction mechanism.

Structural Analysis Using ARIA. The performance and utility of the program ARIA can be seen in the high quality of the structures that we were able to calculate. The NOE data were relatively noisy, and yet with the help of an initial model, ARIA was able to assign almost 3000 NOEs and produce structures with good convergence. The effect of the pseudo-NOEs was relatively minor as in each case the precision of the ensemble of structures was similar. In fact, the standard deviations in precision imply that the ensembles are equally likely (Table 2). Specifically, there are some differences in the positioning of the side chains Asp21, Glu56, and Asp76 in the different structures. The side chains are restrained to similar positions when the pseudo-NOEs are present but are positioned randomly when they are not restrained. This is not surprising when one considers that these are pseudo-NOEs of approximately 5 Å that connect carboxylic groups. In the absence of the restraints there are few protons in the vicinity that could provide NOEs to restrain these side chains. However, by examining the precision among the average structures in Figure 4 and the RMSD values in Table 2, one can see that the introduction of the pseudo-NOEs appears not to affect the global structure of residues 6−114.

It was surprising to us to find ARIA had assigned long-range NOEs from the C-terminus. In this case, the program attempted to connect the largely unstructured C-terminal region with the more structured N-terminal region because of the degeneracy in assignments. Upon inspection we found that these NOE assignments were likely in error. On the basis of chemical shift alone there are peaks that could correspond to C-terminal residues interacting with the core domain. However, we found very few symmetry-related cross-peaks that could verify these interactions. This is an inherent problem with the ARIA approach when stretches of the protein are presumably unstructured, yet assigned.

Conclusions. The solution structure of the HIV-1 RT RNase H domain was determined to high precision in the presence of Mg^{2+}. Although addition of high concentrations of $MgCl_2$ leads to significant spectral improvements, the amide resonances for residues Glu124 through Ala132 are not observed. Gly133 gives a very weak resonance, while

residues Ile134 through Leu138 give rise to sharp resonances with shifts similar to those expected for a random coil. These data indicate that residues Glu124 through Ala132 are subject to a slow conformational exchange process which significantly broadens the corresponding resonances. Comparison with crystal structure data indicates that the corresponding conformational ensemble probably includes a significant proportion of α-helix, which is sometimes observed in the RNase H domain of RT. The variable appearance of this helix in the reported reverse transcriptase crystal structures and the large *B*-factors which characterize this helix when it is observed indicate that this region of the intact enzyme is subject to an analogous conformational heterogeneity. There is a similar parallel between the failure to observe resonances Ala116−His117 in the active site and crystallographic data for the isolated RNase H domain which could not define a single structure for residues 538−542 (*12*, *13*). In contrast, the remaining regions of the enzyme do not exhibit significant exchange broadening and appear to be dynamically homogeneous, making possible the determination of a solution structure with reasonably low RMSD values of 0.60 ± 0.15 for backbone residues from 6 to 114. We are thus led to conclude that these dynamic features, present in both the full enzyme and the isolated RNase H domain, are probably not responsible for the inactivity of the latter. A $MgCl_2$ titration study revealed chemical shift changes that could best be fit to a two-site model with dissociation constants of $K_D^A = 2.7-3.2$ mM (site A) and $K_D^{BA} = \sim 35$ mM (site B when site A is already occupied). These values indicate that the Mg^{2+} binding sites on the enzyme are largely unoccupied under physiological conditions and in the absence of substrate, so that substrate binding is a precondition for Mg^{2+} binding. It may be that the RNase H domain requires the tighter binding characteristic of the full reverse transcriptase molecule to achieve the necessary selectivity. We are currently investigating other ligands and solution conditions to determine the extent to which these can stabilize the C-terminal α-helix.

ACKNOWLEDGMENT

We are grateful to Dr. David Gorenstein of the University of Texas Medical Branch in Galveston for providing us with the plasmid for the RT RNase H domain and to Dr. Joseph Krahn for help with the interpretation of the crystal structure data.

REFERENCES

1. Schatz, O., Cromme, F. V., Guniger-Leitch, F., and Le Grice, S. F. J. (1989) *FEBS Lett. 257*, 311−314.
2. Schatz, O., Mous, J., and Le Grice, S. F. J. (1990) *EMBO J. 9*, 1171−1176.
3. Tisdale, M. S., T., Larder, B. A., and Moelling, K. (1991) *J. Gen. Virol. 72*, 59−66.
4. Becerra, S. P. (1990) *FEBS Lett. 270*, 76−80.
5. Hostomsky, Z. H., Z., Hudson, G. O., Moomaw, E. W., and Nodes, B. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1148−1152.
6. Stammers, D. K., and Skinner, R. H. (1991) *FEBS Lett. 283*, 298−302.
7. Evans, D. B., Brawn, K., Deibel, M. R., Jr., Tarpley, W. G., and Sharma, S. K. (1991) *J. Biol. Chem. 266*, 20583−20585.
8. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) *Science 282*, 1669−1675.
9. Ren, J., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D., and Stuart, D. (1995) *Structure 3*, 915−926.
10. Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D., et al. (1995) *Nat. Struct. Biol. 2*, 293−302.
11. Sarafianos, S. G., Das, K., Ding, J., Boyer, P. L., Hughes, S. H., and Arnold, E. (1999) *Chem. Biol. 6*, R137−R146.
12. Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991) *Science 252*, 88−95.
13. Chattopadhyay, D., Finzel, B. C., Munson, S. H., Evans, D. B., Sharma, S. K., Strakalaitis, N. A., Brunner, D. P., Eckenrode, F. M., Dauter, Z., Betzel, C., and Einspahr, H. M. (1993) *Acta Crystallogr., Sect. D 49*, 423−427.
14. Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. (1992) *Biochemistry 31*, 9150−9157.
15. Powers, R., Clore, G. M., Bax, A., Garrett, D. S., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. M. (1991) *J. Mol. Biol. 221*, 1081−1090.
16. Kern, G., Handel, T., and Marqusee, S. (1998) *Protein Sci. 7*, 2164−2174.
17. Oda, Y., Nakamura, H., Kanaya, S., and Ikehara, M. (1991) *J. Biomol. NMR 1*, 247−255.
18. Nakamura, H., Oda, Y., Iwai, S., Inoue, H., Ohtsuka, E., Kanaya, S., Kimura, S., Katsuda, C., Katayanagi, K., and Morikawa, K. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 11535−11539.
19. Beese, L. S., and Steitz, T. A. (1991) *EMBO J. 10*, 25−33.
20. Yang, W., and Steitz, T. A. (1995) *Structure 3*, 131−134.
21. Cowan, J. A., Ohyama, T., Howard, K., Rausch, J. W., Cowan, S. M., and Le Grice, S. F. (2000) *J. Biol. Inorg. Chem. 5*, 67−74.
22. Cirino, N. M., Cameron, C. E., Smith, J. S., Rausch, J. W., Roth, M. J., Benkovic, S. J., and Le Grice, S. F. (1995) *Biochemistry 34*, 9936−9943.
23. Katayanagi, K., Okumura, M., and Morikawa, K. (1993) *Proteins 17*, 337−346.
24. Goedken, E. R., and Marqusee, S. (2001) *J. Biol. Chem. 276*, 7266−7271.
25. Arts, E. J., and Le Grice, S. F. (1998) *Prog. Nucleic Acid Res. Mol. Biol. 58*, 339−393.
26. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
27. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR 4*, 603−614.
28. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
29. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 203−216.
30. Pascal, S. M., Muhandiram, R. D., Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 197−201.
31. Grzesiek, S., Anglister, J., and Bax, A. (1993) *J. Magn. Reson., Ser. B 101*, 114−119.
32. Logan, T. M., Olejniczak, E. T., Xu, R. X., and Fesik, S. W. (1993) *J. Biomol. NMR 3*, 225−231.
33. Yamazaki, T., Forman-Kay, J. D., and Kay L. E. (1993) *J. Am. Chem. Soc. 115*, 11054−11055.
34. Kay, L. E., Xu, G. Y., and Yamazaki, T. (1994) *J. Magn. Reson., Ser. A 109*, 129−133.
35. Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R., and Bax, A. (1993) *J. Magn. Reson., Ser. B 101*, 210−213.
36. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR 13*, 289−302.
37. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
38. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
39. Nilges, M. (1995) *J. Mol. Biol. 245*, 645−660.
40. Nilges, M. (1997) *Folding Des. 2*, S53−S57.
41. Nilges, M., and O'Donoghue, S. (1998) *Prog. NMR Spectrosc. 32*, 107−139.
42. Linge, J. P., and Nilges, M. (1999) *J. Biomol. NMR 13*, 51−59.
43. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
44. Cristofaro, J. V., Rausch, J. W., Le Grice, S. F., and DeStefano, J. J. (2002) *Biochemistry 41*, 10968−10975.
45. Levy, L. A., Murphy, E., Raju, B., and London, R. E. (1988) *Biochemistry 27*, 4041−4048.

46. Heinonen, E., and Akerman, K. E. (1987) *Biochim. Biophys. Acta 898*, 331−337.
47. Grubbs, R. D., and Maguire, M. E. (1987) *Magnesium 6*, 113−127.
48. Takimoto, K., Okada, M., Matsuda, Y., and Nakagawa, H. (1985) *J. Biochem.* (*Tokyo*) *98*, 363−370.
49. Leech, A. P., Baker, G. R., Shute, J. K., Cohen, M. A., and Gani, D. (1993) *Eur. J. Biochem. 212*, 693−704.
50. Wilkie, J., Cole, A. G., and Gani, D. (1995) *J. Chem. Soc., Perkin Trans. 1*, 2709−2727.
51. Nigou, J., Dover, L. G., and Besra, G. S. (2002) *Biochemistry 41*, 4392−4398.
52. Laskowski, R. A., MacArthur, M. W., and Thornton, J. M. (1998) *Curr. Opin. Struct. Biol. 8*, 631−639.

BI0204894