Articles

NMR Solution Structure of the RNA-Binding Peptide from Human Immunodeficiency Virus (Type 1) Rev[†,‡]

Martin J. Scanlon,* David P. Fairlie, David J. Craik, Darren R. Englebretsen, and Michael L. West

Centre for Drug Design and Development, University of Queensland, St. Lucia, Queensland 4072, Australia

Received November 22, 1994; Revised Manuscript Received March 27, 1995[⊗]

ABSTRACT: NMR spectroscopy has been used to solve the three-dimensional solution structure of a minimal RNA-binding domain of the Rev protein from the human immunodeficiency virus (type 1), an essential regulatory protein for viral replication. The presence of 10 arginine residues in the 17-residue peptide Rev_{34-50} caused significant problems in assignment of the NMR spectra. To improve spectral resolution, the peptide was synthesized with an alanine replacing a nonessential arginine and with selectively ^{15}N-labeled residues. Contrary to Chou−Fasman modeling predictions an α-helix was detected in both water and 20% trifluoroethanol (TFE) and was found to span residues that constitute the RNA-binding and nuclear-localizing domains of Rev. The sequence-specific information provided by the NMR data gives a full description of the solution conformation of Rev_{34-50} which serves as a template for investigating binding of the peptide to RNA from the Rev response element (RRE). Preliminary modeling suggests that the helix can fit neatly into the expanded major groove of the RRE where interactions between the peptide side chains and the RNA can be identified. These data may aid the construction of a suitable pharmacophore model for the rational design of molecules that block Rev−RNA binding and inhibit HIV replication.

One approach to AIDS[1] therapy is to develop drugs that inhibit protein function at different stages of the replication of the human immunodeficiency virus (HIV) (Ratner, 1993). A promising target for therapeutic intervention is the protein HIV-1 Rev (formerly art/trs) (Rosen, 1992). Although only a few inhibitors of Rev function have been reported to date, such inhibition is known to prevent replication of HIV-1 in cells (Zapp et al., 1993). Rev is essential for viral replication, regulating splicing of *gag-pol* and *env*-encoding mRNA in the nucleolus of infected immune cells and facilitating export of singly spliced or unspliced mRNA to the cytoplasm (Sodroski et al., 1986; Feinberg et al., 1986). Rev acts by binding to a *cis*-acting component of the *env* region of the mRNA known as the Rev responsive element (RRE) (Rosen, 1992). The Rev−RRE complex is then activated by an unknown protein (Fankhauser et al., 1991), which is a shuttle receptor for the nuclear import of ribosomal proteins. Export of the mRNA−Rev protein complex from the nucleolus to the cytoplasm, release of the HIV−mRNA, and import of the Rev protein back into the nucleolus complete the transport cycle (Rosen, 1992; Fankhauser et al., 1991).

The Rev protein is a 116 amino acid phosphoprotein of unknown structure possessing at least two functionally important regions, an RRE-binding domain and a nuclear protein binding or effector domain (Stutz & Rosbash, 1994). The specific RRE-binding domain of Rev is localized within the basic amino acid sequence Arg_{38}−Arg_{46} contained within the nucleolar-localizing motif that spans amino acids Arg_{35}−His_{53} (Malim et al., 1990; Olsen et al., 1990). The 17-residue peptide Rev_{34-50} has been the subject of a number of biochemical investigations, as it exhibits specific binding to the RRE, with an apparent dissociation constant ($K_d \sim 20$ nM) comparable to that ($K_d \sim 10$ nM) for the Rev protein itself (Kjems et al., 1992). Furthermore, as a series of Tat fusion proteins of Rev_{34-50} have been shown to bind the RRE *in vivo* (Tan et al., 1993), Rev_{34-50} is a good functional model for the RRE-binding domain of Rev. The determination of its three-dimensional solution structure is an important prerequisite for studying interactions with RRE and may provide sufficient information to enable the rational design of inhibitors of Rev binding to RNA.

[†] This work was financed in part by a grant from the Commonwealth Aids Research Grant Scheme.

[‡] Coordinates have been deposited in the Brookhaven Protein Data Bank under the filename 1RPV.

* To whom correspondence should be addressed (Fax, +61-7-3365-1990; telephone, +61-7-3365-4944; Email, M.Scanlon@mailbox.uq.oz.au).

[⊗] Abstract published in *Advance ACS Abstracts,* June 1, 1995.

[1] Abbreviations: NMR, nuclear magnetic resonance, RNA, ribonucleic acid; Rev, regulator of virion expression; TFE, trifluoroethanol; HIV, human immunodeficiency virus; AIDS, acquired immune deficiency syndrome; *gag-pol*, group-specific antigen polymerase; *env*, envelope; RRE, Rev response element; Tat, transactivating protein; SPPS, solid-phase peptide synthesis; Boc, *tert*-butoxycarbonyl; MBHA, *p*-methylbenzhydrylamine; HBTU, 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; DIEA, diisopropylethylamine; DMF, dimethylformamide; Trp(CHO), *N*-indolylformyl-L-tryptophan; Arg(Pmc), N^G-(2,2,5,7,8-pentamethylchroman-6-sulfonyl)-L-arginine; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; WATERGATE, water suppression by gradient-tailored excitation; DQF-COSY, double-quantum-filtered correlated spectroscopy; ROESY, rotating frame Overhauser enhancement spectroscopy; HMQC, heteronuclear multiple-quantum coherence; DEPT, distortionless enhancement by polarization transfer; NOE, nuclear Overhauser enhancement; RMS, root mean squared.

0006-2960/95/0434-8242$09.00/0

To date the only structural information on this peptide has come from circular dichroism measurements which suggest that Rev_{34-50} has a partially helical conformation in aqueous solution (Tan et al., 1993). However, details of the extent and localization of the helicity cannot be provided by this technique. Interestingly, molecular modeling programs, based on the algorithm of Chou and Fasman (1977), do not predict α-helicity in the important arginine-rich region of Rev_{34-50}. Thus a detailed and accurate experimental determination of the structure of Rev_{34-50} is needed to establish the extent and location of any helicity and so clarify possible interactions between the RRE-binding domain of Rev and RNA. Such results are timely in view of recent NMR and modeling studies of RRE which indicate that RRE forms a widened major groove that is large enough to accommodate a helical peptide (Battiste et al., 1994; Leclerc et al., 1994).

In this work we have used two-dimensional NMR spectroscopy to determine the three-dimensional structure of Rev_{34-50}. Assignment of resonances to specific amino acids of this peptide is complicated by the presence of multiple contiguous arginine residues which produce overlapping signals that are difficult to assign unambiguously. By substituting amino acids in the sequence and incorporating specific ^{15}N labels at strategic sites, these problems have been solved. We report the extent and localization of the α-helicity in this peptide and the orientation of side chains of the amino acids that are essential for binding to RNA. Our results constitute the first stage in developing a novel pharmacophore model to facilitate rational design of drugs that block Rev function and hence replication of HIV.

MATERIALS AND METHODS

Materials. Boc-amino acids were obtained from the Peptide Institute Inc. (Osaka, Japan). NMR solvents were obtained from Cambridge Isotope Laboratories (Woburn, MA).

Peptide Synthesis. Three peptides were synthesized (Figure 1): Rev_{34-50} corresponding to residues 34−50 of the native Rev sequence, $Rev_{34-50}R_{42}A$ in which one arginine residue is replaced by an alanine, and $Rev_{34-50}R_{42}A$ in which all three alanine residues were labeled with ^{15}N. Peptides were synthesized on a 0.5 mM scale using MBHA resin (0.93 mmol g^{-1}; Peninsula Labs, Belmont, CA) by manual Boc solid-phase peptide synthesis using HBTU/DIEA activation and neutralization *in situ* (Schnölzer et al., 1992). All residues were coupled with a 4-fold excess of activated Boc-amino acid, with the exception of [^{15}N]Ala for which a 1.5-fold excess was used. Coupling yields (>99.6%) were determined by quantitative ninhydrin assay. The succinylation was performed by addition of succinic anhydride (2 mmol) and DIEA (2.6 mmol) in DMF (4 mL). Following removal of the Trp(CHO) protecting group with DMF/ethanolamine/H_2O (15:1:1, 10 mL, room temperature, 2 × 30 min), the peptides were cleaved from the resin using either *p*-cresol/HF (1:9, 10 mL, −5 °C, 120 min) or *p*-thiocresol/*p*-cresol/HF (1:1:18, 10 mL, −5 °C, 90 min). HF was removed in vacuo, and the peptides were precipitated with diethyl ether (10 mL), purified by HPLC (Vydac C4 column, 25 × 250 mm; 0−60% B over 60 min at 8 mL min^{-1}; A = 0.1% TFA, B = 0.09% TFA, 10% H_2O, 90% MeCN), and characterized by HPLC (Vydac C4 column, 4.6 × 150 mm, 5−67% B over 30 min at 1 mL min^{-1}) and ion spray mass spectrometry.

NMR Spectroscopy. NMR samples contained ~3 mM peptide in either 90% H_2O/10% ^2H_2O, 80% ^2H_2O/20% TFE, or 72% H_2O/8% ^2H_2O/20% TFE at pH 4.7. All NMR spectra were recorded on a Bruker ARX500 spectrometer equipped with a shielded gradient unit. Except where otherwise noted, the probe temperature was maintained at 4 °C using a BT V2000 control unit and a Bruker BCU refrigeration unit. In all ^1H NMR experiments the carrier was set at the center of the spectrum on the water resonance frequency, and quadrature detection was used in both dimensions. Spectra were recorded in the phase-sensitive mode using the time-proportional phase incrementation method (Marion & Wüthrich, 1983). TOCSY experiments (Braunschweiler & Ernst, 1983) using the MLEV-17 (Bax & Davis, 1985a) spin lock sequence were recorded with mixing times of 50 and 100 ms. Trim pulses of 2.5-ms duration were employed before and after the MLEV-17 sequence. Two-dimensional nuclear Overhauser enhancement (NOESY) spectra (Anil-Kumar et al., 1980) were recorded with mixing times of 50−200 ms. In the TOCSY and NOESY experiments suppression of the residual water resonance was achieved using a modified WATERGATE (Piotto et al., 1992) sequence. Two gradient pulses of 2-ms duration and a gradient strength of 6 G cm^{-1} were applied on either side of a binomial 3−9−19 pulse which employed a field strength of 10 kHz. In all other experiments solvent suppression was achieved using selective low-power irradiation of the water resonance during the relaxation delay of 1.8 s. DQF-COSY spectra were acquired as described by Rance et al. (1983). Rotating frame Overhauser enhancement (ROESY) spectra (Bax & Davis, 1985b) were recorded with a continuous wave 2.5-kHz field strength spin locking pulse of 200-ms duration. $^3J_{HNC^\alpha H}$ coupling constants were measured from either one-dimensional spectra acquired over the temperature range 4−14 °C or the DQF-COSY spectrum. After processing, the appropriate rows were selected from the 2D matrix. These rows were inverse Fourier transformed and zero filled to 32K to give a final resolution of 0.18 Hz/point. $^3J_{HNC^\alpha H}$ coupling constants were then derived from a simulation of the t_2 cross sections of the HNC$^\alpha$H cross peaks.

To detect slowly exchanging amide protons, a series of one-dimensional and TOCSY spectra were recorded for a fully protonated sample of $Rev_{34-50}R_{42}A$ immediately following dissolution in 80% ^2H_2O/20% TFE at pH 4.7. The TOCSY spectra were recorded with 2048 data points and 256 t_1 increments. HMQC experiments (Bax et al., 1983) were carried out in 72% H_2O/8% ^2H_2O/20% TFE with a polarization transfer time of 3.45 ms for ^{13}C and 5.3 ms for ^{15}N. Spectral widths were 25 kHz in the ^{13}C dimension and 5066 Hz for ^{15}N. Unless otherwise noted, ^1H NMR experiments were recorded with a spectral width of 6024 Hz over 4096 data points. Between 400 and 512 t_1 increments were acquired with 16−64 scans per increment.

^{15}N spin−lattice relaxation times (T_1's) were measured using a one-dimensional ^1H-detected double DEPT sequence (Sklenár et al., 1987). Raw data were analyzed using standard Bruker software by fitting the data to the three-parameter exponential:

$$S_\tau = A(1 - Be^{-\tau/T_1})$$

where S_τ is the signal intensity at time τ and A, B, and T_1 are variable parameters. A total of 12 τ values (0.007, 0.022,

0.082, 0.102, 0.152, 0.202, 0.252, 0.402, 0.502, 1.002, 1.502, and 2.502 s) gave a satisfactory curve fit, and 128 scans were recorded per experiment.

Spectra were processed on a Silicon Graphics Indigo work station using UXNMR (Bruker). The data were multiplied with squared sine-bell window functions using shifts of between 45° and 90°. Using a fifth-order polynomial function, baseline corrections were applied across the spectrum to within 150 Hz on either side of the water resonance frequency in ω_2 and, for the ROESY spectrum, across the whole spectral width in ω_1. Matrix sizes were typically 2048 × 2048 real points. Processed data were analyzed using FELIX 2.30 (Biosym Technologies, San Diego, CA). Matrices were byte-swapped and converted into FELIX format using an in-house program (supplied by D. Smith). Chemical shifts were measured relative to the water frequency at 4.98 ppm downfield from DSS for the aqueous sample and 5.00 ppm downfield from DSS in 20% TFE. ^{13}C shifts were referenced to the TFE signal at 61.9 ppm downfield of DSS. The pH values are uncorrected room temperature measurements.

Structure Determination. Distance constraints were calculated from experimental NOESY cross-peak intensities using FELIX 2.30 and classified as strong, medium, and weak. The corresponding interproton distances were assigned upper bounds of 2.7, 3.5, and 5.0 Å, respectively. Distances to degenerate methylene and methyl groups were multiplied by 1.1 and 1.2, respectively, and a further 0.9 Å was added to distances involving methylene and methyl pseudoatoms. Backbone dihedral restraints of $-60° \pm 30°$ were applied for $^3J_{HNC^\alpha H} < 5$ Hz (Pardi et al., 1984).

Structures were calculated using a dynamic simulated annealing procedure within the program X-PLOR 3.1 (Brünger, 1992). The simulated annealing was performed using a geometric force field, starting from a linear template structure, and initially, random velocities were assigned to the starting coordinates. A soft square well potential (Nilges et al., 1988) for NOE and dihedral constraints was used, and bond lengths were fixed during the high-temperature and cooling stages. A total of 12 000 steps were performed at 1000 K and a further 6000 steps as the molecule was cooled to 100 K, using a time step of 2.5 fs.

The final 20 structures with the lowest overall energies and least violations of NOE and dihedral restraints were energy minimized using the CHARMm force field (Brooks et al., 1983). These structures were retained for structural analysis. The structures were displayed in InsightII (Biosym) and analyzed using the program SSTRUC, based on the methods of Kabsch and Sander (1983).

RESULTS

Sample Preparation. Synthesis of Rev_{34-50}, containing 10 arginine residues of which only 4 may be critical for RNA binding, has been previously reported using Fmoc chemistry (Calnan et al., 1991). In order to avoid possible problems associated with Trp alkylation by Arg(Pmc)-derived cations during cleavage (Choi & Aldrich, 1993), an alternative approach involving Boc chemistry was used to prepare the sequences for this study (Figure 1) as the *N*-succinyl-carboxamide in ~25% overall yield. A pH titration of Rev_{34-50} in 90% H_2O/10% ^2H_2O at 4 °C showed that optimal amide proton chemical shift dispersion was obtained at pH

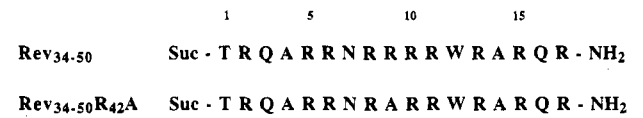

FIGURE 1: Sequences of peptides corresponding to the RNA-binding region of Rev.

4.7. This pH was used for the following NMR investigations of the conformation of Rev_{34-50} and the mutant $Rev_{34-50}R_{42}A$. No concentration-dependent change in the chemical shifts was observed over the range 0.5−4.0 mM.

Resonance Assignments. Resonances were assigned to specific protons using the sequential assignment procedure of Wüthrich (1986). ^1H-NMR spectra of Rev_{34-50} in 90% H_2O/10% ^2H_2O at pH 4.7 and 4 °C were extensively overlapped, and only partial assignments could be made. In particular, it was not possible to distinguish a number of backbone resonances for the arginine residues in the region Arg_5-Arg_{11}. While Arg_5 and Arg_8 could be partially assigned, the remaining four arginines in this putative binding region could not be resolved. In an attempt to overcome this problem, a second peptide was synthesized containing a single Arg → Ala substitution. It has been shown that this substitution has little effect on the binding affinity of the peptide for the minimal sequence of RRE−RNA (Tan et al., 1993). In this peptide the partial sequence $RNRR_9RR$ was replaced by $RNRA_9RR$, and this led to some simplification of the spectra. The ready identification of Ala_9, together with an NOE between $Ala_9C^\beta H-Arg_{10}NH$, enabled unequivocal assignment of the backbone resonances of all residues.

Even though all backbone resonances could be assigned for the peptide, there remained some ambiguities in side-chain assignments. In an attempt to overcome this problem, the synthesis of the peptide $Rev_{34-50}R_{42}A$ was repeated with ^{15}N labels incorporated in the three Ala residues. It was possible to obtain an almost complete set of resonance assignments for this peptide using a combination of DQF-COSY, TOCSY, and ROESY spectra measured for both unlabeled peptide and, in nondecoupled spectra, the ^{15}N-labeled material. Figure 2 shows a region from the 100-ms mixing-time TOCSY of $Rev_{34-50}R_{42}A$, and Table 1 contains chemical shifts of the assigned proton resonances in 90% H_2O/10% ^2H_2O.

The NMR data provide strong evidence for the presence of helicity in water. First, analysis of the $C^\alpha H$ resonance positions using the chemical shift index strategy (Wishart et al., 1992), which predicts secondary structure on the basis of deviations of the $C^\alpha H$ chemical shifts from their random coil values, supports the existence of a helical conformation. This can be inferred from the fact that every $C^\alpha H$ in the peptide has a shift index of −1. Second, despite the problems of degeneracy, a number of cross peaks can be observed in the ROESY spectrum which are typical of an α-helix. Due to the poor chemical shift dispersion of the amide protons, few cross peaks could be resolved in the NH−NH or the NH−$C^\alpha H$ region of the spectrum. However, a total of seven $d_{\alpha\beta(i,i+3)}$ cross peaks were observed. In fact, except in cases where overlap prevented observation of the cross peak, $d_{\alpha\beta(i,i+3)}$ connectivity was present for all but one residue pair. Lastly, the $^3J_{HCC^\alpha H}$ coupling constants are indicative of a helical conformation. A total of 12 $^3J_{HNC^\alpha H}$ coupling constants were determined, either from 1D spectra

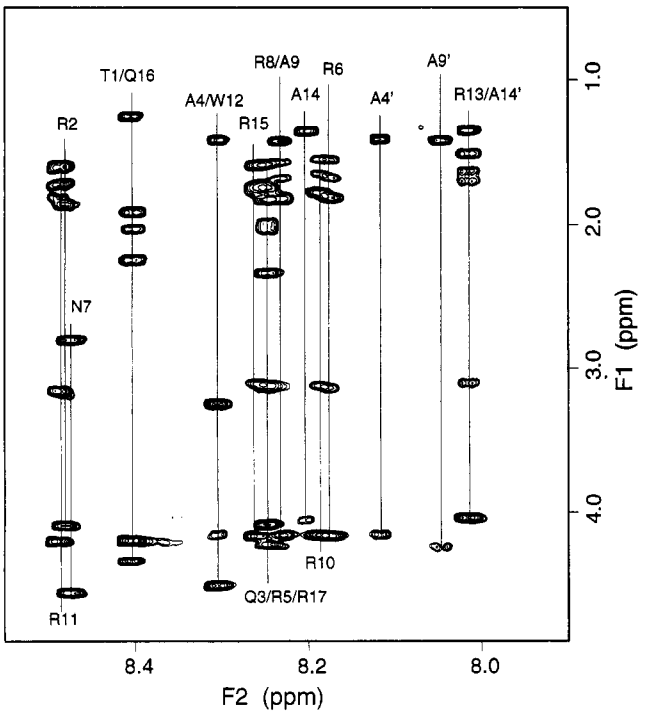

FIGURE 2: Fingerprint region of the 100-ms mixing-time TOCSY spectrum of $Rev_{34-50}R_{42}A$ in 90% H_2O/10% ^2H_2O, showing assignment of the spin systems. The one-letter notation for amino acids is used, and the number represents the position of the residue in the peptide sequence. Two pairs of cross peaks are observed for each of the alanine residues due to coupling of the amide proton with the adjacent ^{15}N nucleus.

Table 1: Chemical Shifts of the ^1H Resonances of $Rev_{34-50}R_{42}A$ in 90% H_2O/10% ^2H_2O at 4 °C and pH 4.7

residue	NH	αH	βH	others
Suc				2.62, 2.46
Thr-1	8.41	4.20	4.34	γH 1.25
Arg-2	8.48	4.09	1.86	γH 1.70, 1.59; δH 3.20, 3.16; εH 7.53
Gln-3	8.25	4.08	1.99	γH 2.33, 2.04; εH 7.64, 6.88
Ala-4	8.21	4.16	1.41	
Arg-5	8.25	4.10	1.82	γH nd;[a] δH nd; εH 7.22
Arg-6	8.18	4.17	1.81, 1.68	γH 1.55; δH 3.14; εH 7.21
Asn-7	8.48	4.56	2.83	δH 7.70, 7.07
Arg-8	8.23	4.16	1.81, 1.68	γH 1.57; δH 3.12; εH 7.23
Ala-9	8.13	4.24	1.42	
Arg-10	8.19	4.16	1.78, 1.65	γH 1.54; δH 3.11; εH 7.17
Arg-11	8.49	4.21	1.81, 1.72	γH 1.60; δH 3.16; εH 7.22
Trp-12	8.30	4.50	3.25	N(1)H 10.16; 2H 7.22; 4H 7.43; 5H 7.08; 6H 7.18; 7H 7.53
Arg-13	8.02	4.04	1.70, 1.63	γH 1.50; δH 3.10; εH 7.19
Ala-14	8.11	4.06	1.35	
Arg-15	8.26	4.17	1.77, 1.72	γH 1.58; δH 3.10; εH 7.19
Gln-16	8.40	4.22	1.92	γH 2.24, 2.03; εH 7.48, 6.88
Arg-17	8.25	4.19	1.83	γH nd; δH nd; εH nd; NH_2 7.68, 7.22

[a] nd, not determined.

directly or from the 2D DQF-COSY. These values were used to calculate the average helicity over the entire molecule, which was found to be 70%. This compares favorably with the percentage helicity determined for a similar peptide in aqueous solution from CD measurements (Tan et al., 1993). In principle, the next step in the analysis would be the calculation of a three-dimensional structure based on distance constraints derived from the ROESY spectrum. However, due to the substantial overlap in these spectra it was not possible to resolve sufficient cross peaks

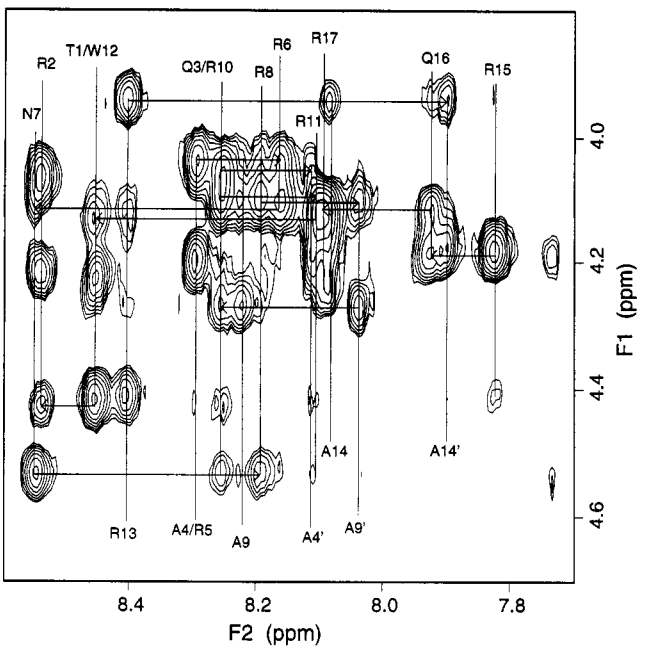

FIGURE 3: NH−$C^\alpha H$ region of the 150-ms mixing-time NOESY spectrum of $Rev_{34-50}R_{42}A$ in 72% H_2O/8% ^2H_2O/20% TFE showing sequential connectivities.

Table 2: Chemical Shifts of the ^1H Resonances of $Rev_{34-50}R_{42}A$ in 72% H_2O/8% ^2H_2O/20% TFE at 4 °C and pH 4.7

residue	NH	αH	βH	others
Suc				2.69, 2.51
Thr-1	8.44	4.21	4.41	γH 1.33
Arg-2	8.52	4.05	2.00, 1.89	γH 1.83, 1.67; δH 3.29, 3.20; εH 7.73
Gln-3	8.23	4.03	2.11	γH 2.41; εH 7.67, 6.96
Ala-4	8.19	4.18	1.53	
Arg-5	8.27	4.01	1.96, 1.88	γH 1.65; δH 3.17; εH 7.24
Arg-6	8.18	4.09	1.94, 1.85	γH 1.60; δH 3.19; εH 7.30
Asn-7	8.53	4.52	2.98, 2.85	δH 7.71, 7.12
Arg-8	8.18	4.08	1.96, 1.82	δH 1.58; δH 3.19; εH 7.29
Ala-9	8.11	4.25	1.57	
Arg-10	8.24	4.10	1.99, 1.90	γH 1.68; δH 3.23; εH 7.26
Arg-11	8.09	4.13	1.94, 1.80	γH 1.63; δH 3.21; εH 7.33
Trp-12	8.44	4.40	3.47, 3.37	N(1)H 10.14; 2H 7.22; 4H 7.46; 5H 7.12; 6H 7.23; 7H 7.60
Arg-13	8.39	3.93	1.93, 1.84	γH 1.71; δH 3.23; εH 7.37
Ala-14	7.98	4.16	1.50	
Arg-15	7.82	4.18	1.88, 1.80	γH 1.66; δH 3.11, 3.04; εH 7.23
Gln-16	7.91	4.10	1.89	γH 2.12, 2.03; εH 7.06, 6.58
Arg-17	8.08	4.23	1.89, 1.78	γH 1.70; δH 3.19; εH 7.28; NH_2 7.49, 7.18

to justify such a structure calculation.

In an attempt to better resolve the spectra by stabilizing the helical conformation of the peptide, TFE was added to a concentration of 20% (v/v). This improved the chemical shift dispersion (Figure 3), and a full set of resonance assignments was obtained using the sequential assignment procedure (Wüthrich, 1986). Table 2 contains the chemical shifts of the assigned proton resonances of $Rev_{34-50}R_{42}A$ in 20% TFE, and Figure 4 summarizes the sequential (i,i+1) NOE connectivities on which these assignments were based. Of those residues which have been identified as essential for RNA binding (Tan et al., 1993), only Arg_{11} and Arg_{13} have backbone chemical shifts which change by more than 0.1 ppm upon addition of 20% TFE. Given that these residues are both adjacent to Trp_{12} in the sequence, it is possible that a small change in the conformation of the Trp

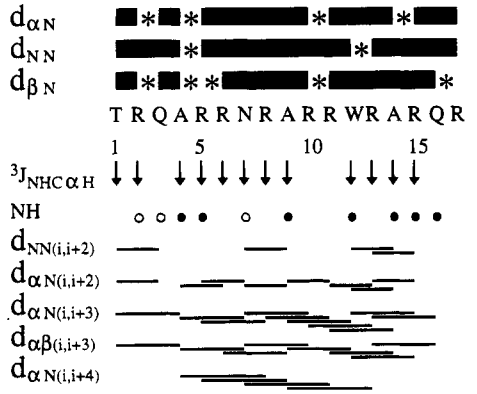

FIGURE 4: Summary of the NMR data used for sequence-specific assignments and identification of secondary structure in $Rev_{34-50}R_{42}A$ in 72% H_2O/8% ^2H_2O/20% TFE. $d(i,i+1)$ connectivities which could not be observed due to overlap are indicated with asterisks (*). Values of $^3J_{HNC^\alpha H}$ are indicated by ↓ if <5 Hz. Filled circles in the NH row denote slowly exchanging amide protons (as denoted by their presence in TOCSY experiments of $Rev_{34-50}R_{42}A$ recorded immediately following dissolution in 80% ^2H_2O/20% TFE). Open circles represent amide protons in intermediate exchange.

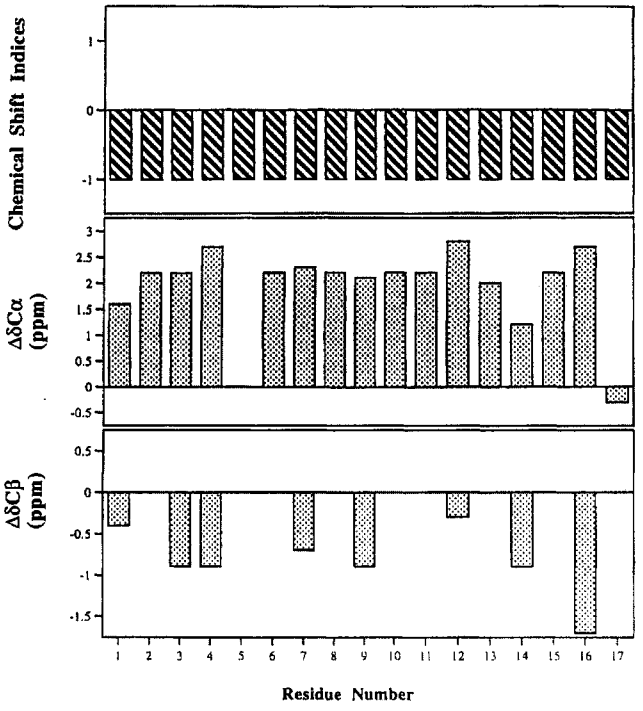

FIGURE 5: Chemical shift indices of the $C^\alpha H$ resonances and deviation of the $^{13}C^\alpha$ and $^{13}C^\beta$ resonances from random coils for $Rev_{34-50}R_{42}A$ in 72% H_2O/8% ^2H_2O/20% TFE. Gaps in the ^{13}C chemical shift deviation graphs represent cross peaks in the HMQC spectrum which could not be resolved due to extensive overlap in the ^1H dimension.

side chain could cause these large deviations in chemical shift. Therefore, it is likely that the structure of the peptide in 20% TFE provides a good model for that in aqueous solution. Also shown in Figure 4 are estimates of the $^3J_{HNC^\alpha H}$ coupling constants, the positions of the slowly exchanging amide protons obtained from a series of TOCSY spectra measured immediately following dissolution of the peptide in 80% ^2H_2O/20% TFE, and the location of the medium-range NOE connectivities. Figure 5 shows the $C^\alpha H$ chemical shift indices and the deviations of the $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts from their random coil values for the peptide in 20% TFE (Wishart & Sykes, 1994).

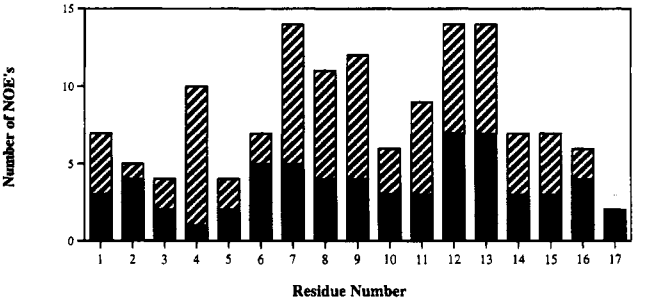

FIGURE 6: Distribution of short- and medium-range connectivities observed in the 150-ms mixing-time NOESY spectrum of $Rev_{34-50}R_{42}A$ in 72% H_2O/8% ^2H_2O/20% TFE.

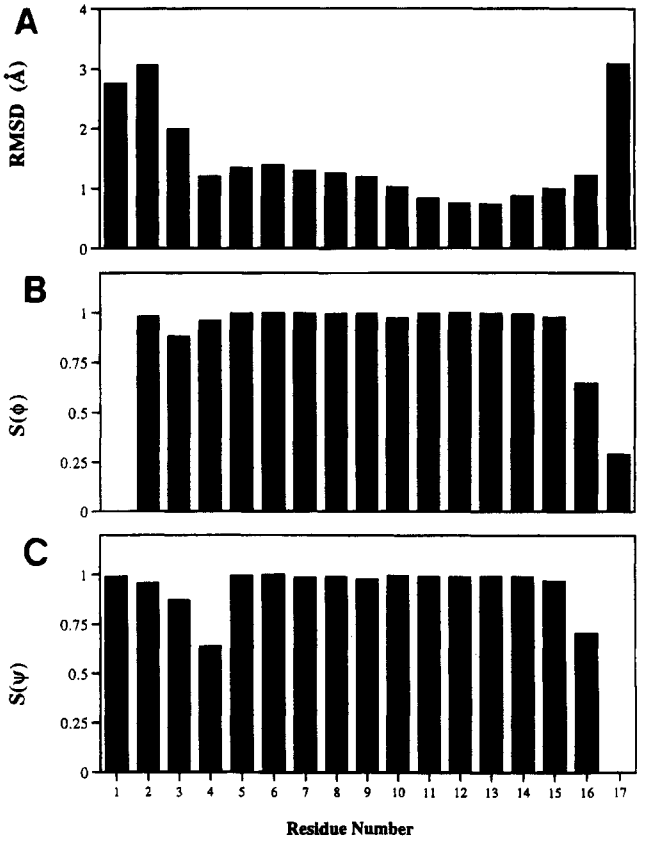

FIGURE 7: Structural parameters for the 20 final structures of $Rev_{34-50}R_{42}A$. (A) RMS differences from the average structure of the backbone heavy atoms (N, C^α, C, O) when the structures are superimposed over the entire length of the molecule. (B, C) Angular order parameters (Hyberts et al., 1992; Pallaghy et al., 1993) for the backbone dihedral angles ϕ and ψ.

The presence of several low $^3J_{HNC^\alpha H}$ coupling constants, the sequence of residues with chemical shift indices of −1, the deviations of the $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts from their random coil values (Wishart et al., 1991; Spera & Bax, 1991), and the large number of slowly exchanging amide protons present strong evidence that the peptide adopts an α-helix in 20% TFE. In order to quantify the extent and location of this helix, a structure calculation was performed using experimental distance constraints in a simulated annealing protocol.

Structure Calculations. An initial family of 50 structures was calculated by simulated annealing in X-PLOR using a constraint set comprising a total of 75 interresidue distances and 12 backbone dihedral angle constraints inferred from spin−spin coupling constants. The distance constraint set consisted of 33 sequential and 42 medium-range ($1 < |i +$

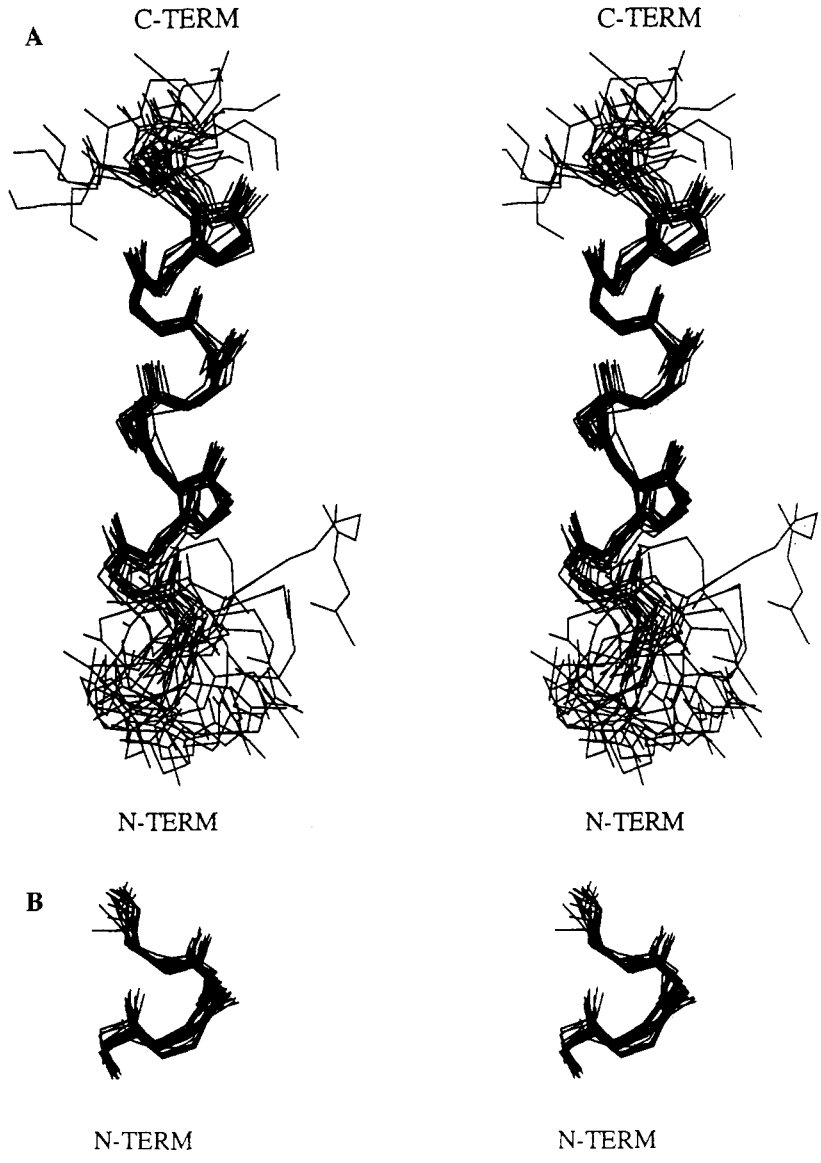

FIGURE 8: Stereoviews of the 20 final structures of $Rev_{34-50}R_{42}A$ superimposed over (A) the backbone heavy atoms (N, C^{α}, C, O) of residues Arg_5-Arg_{15} and (B) the backbone heavy atoms (N, C^{α}, C, O) of Thr_1-Ala_4.

$j| < 5$) constraints. A summary of the distribution of these constraints throughout the sequence of $Rev_{34-50}R_{42}A$ is presented in Figure 6. A total of 20 of these structures had no violations of either dihedral or distance constraints greater than 30° or 0.3 Å, respectively, and these were energy minimized in X-PLOR using the CHARMm force field to give the final 20 structures. The mean pairwise RMS differences for the backbone heavy atoms when the final 20 structures were superimposed over the entire sequence are shown in Figure 7. These data indicate that the structure is well defined over residues Arg_5-Arg_{15} but less so at the N- and C-termini. Stereoviews of the 20 structures superimposed over the backbone atoms of the well-defined region are shown in Figure 8A. The mean pairwise RMS difference over the backbone heavy atoms of Arg_5-Arg_{15} is 0.57 ± 0.18 Å. The backbone angular order parameters (Figure 7) suggest that the peptide is well defined except for residues $Gln_{16}-Arg_{17}$ and the region around Gln_3-Ala_4. While the NOE data support the observation that the helix is frayed at its C-terminus, overlap of a number of resonances at the N-terminus results in very few NOE's being resolved, particularly for residues Arg_2-Gln_3 (Figure 6). The possibility exists, therefore, that the apparent disorder at the N-terminus reflects the lack of NOE data rather than helical fraying. Indeed, the coupling constants, slow exchange data for residues Thr_1-Ala_4, and the backbone angular order parameters all suggest that the peptide may be ordered at the N-terminus. Figure 8B shows stereoviews of the final 20 structures superimposed over the backbone atoms of residues Thr_1-Ala_4, which have a mean pairwise RMS difference over the backbone heavy atoms of 0.61 ± 0.21 Å.

An alternative possibility for the apparent disorder at the N-terminus could be the presence of significant internal motions in this region. This was investigated by measuring ^{15}N T_1's for the three labeled Ala residues in $Rev_{34-50}R_{42}A$. If the molecule was disordered at the N-terminus and undergoing more rapid motion than the well-defined helical region of the peptide, this would be reflected in the T_1 values which are sensitive to internal motions in peptides and proteins. A $^{15}N/^{1}H$ HMQC experiment recorded for the peptide showed that the ^{15}N shifts of the three alanines were almost coincident, thereby precluding measurement of ^{15}N T_1's for the individual sites using conventional direct

detection experiments. However, the greater dispersion of the ^1H chemical shift meant that the measurements could be made by the method described by Sklenár et al. (1987) using inverse detection of the ^{15}N T_1's via their attached protons. The derived T_1 values for Ala_4, Ala_9, and Ala_{14} were 0.55, 0.55, and 0.54 s, respectively. These data are consistent with no additional motion at the *N*-terminus (sampled by Ala_4) relative to the rest of the peptide (sampled by Ala_9 and Ala_{14}). Taken in conjunction with the experimental constraints derived from the NMR data, these results strongly suggest that the helix extends further toward the N-terminus than the overlaid structures initially suggested and that the divergence of the structures is simply a result of fewer resolved NOE's in this region. In a Ramachandran plot of the angular average of the final 20 structures, all residues lie within the allowed α-region of the plot, consistent with a fully helical conformation.

DISCUSSION

We have determined the first high-resolution solution structure of the putative RNA-binding domain of HIV-1 Rev by NMR spectroscopy. The only previous information on this peptide has come from molecular modeling predictions (Leclerc et al., 1994) and CD measurements (Tan et al., 1993; Daly et al., 1990). The NMR structure provides experimental support for those studies but also reveals more detailed information on the conformation of $Rev_{34-50}R_{42}A$ and the orientation of residues known to be essential for binding to the RRE.

Preliminary investigations of Rev_{34-50} in aqueous solution at 4 °C revealed the presence of a helical conformation which was stabilized by addition of 20% TFE. Despite the small size of Rev_{34-50}, the presence of so many adjacent arginine residues complicated both the assignment of the proton resonances and the calculation of the structure due to problems of overlapping resonances, especially in the aliphatic region of the spectrum. In particular, many of the NOESY cross peaks which are diagnostic of helix formation could not be resolved. This problem was partly overcome by synthesizing a mutant peptide, $Rev_{34-50}R_{42}A$, in which alanine replaced an arginine that is nonessential for binding to RRE. A further modification, involving strategic replacement of three amino acids with their ^{15}N-isotope-enriched analogues, was sufficient to reduce signal overlap and resolve most of the NMR anomalies that hindered the final determination of the peptide structure. These studies have highlighted the utility of isotope labeling, a technique more commonly applied to larger proteins, in the solution of highly overlapped spectra of smaller molecules. Residue substitution and selective isotope labeling in conjunction with SPPS have been shown to provide powerful tools in structure determination by NMR [e.g., Cowburn et al. (1983)].

The results of structure calculations, using NOE-restrained simulated annealing, have shown that in 20% TFE/water solution (and probably for water as well) $Rev_{34-50}R_{42}A$ adopts a well-formed helix. This extends at least from Arg_5 to Arg_{15}, and much of the NMR evidence is consistent with extension of the helix to the N-terminus of the peptide. The α-helix certainly spans all residues thought to be essential for high-affinity binding to RRE. The dimensions of the peptide helix are ~22 Å long by ~14 Å in diameter, which could be accommodated within the major groove of the high-affinity binding region of the RRE that has been reported to be ~20 Å in diameter (Leclerc et al., 1994).

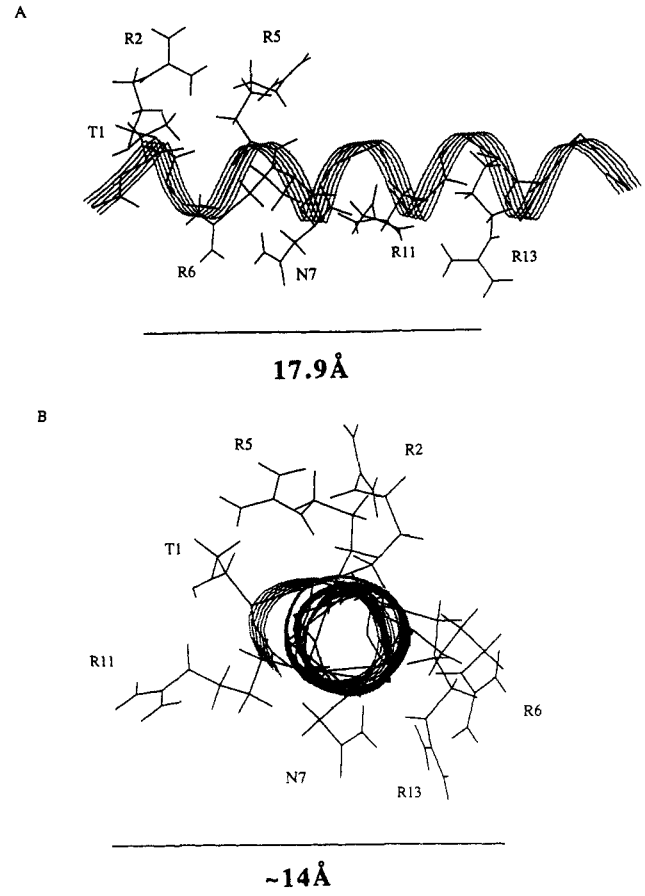

FIGURE 9: Ribbon diagram of the single structure of $Rev_{34-50}R_{42}A$ which is closest to the average over the entire backbone showing those residues known to be essential for binding to RRE (Tan et al., 1993). (A) The approximate $C^\alpha-C^\alpha$ distance is shown for Thr_1-Arg_{13}. (B) End-on view of the peptide showing the approximate dimensions (in Å) of the widest part of the helix.

A recent modeling study (Leclerc et al., 1994) has shown that Rev_{34-50}, constructed in an α-helical conformation with side chains in an extended conformation, can be energy minimized to fit the binding site of RRE. Our solution structure of $Rev_{34-50}R_{42}A$ is similar to this model of the bound peptide. Indeed, when the backbone C^αs of our solution structure were superimposed on the peptide used in the predictive model, a good fit (RMS 0.51 Å) was found for residues spanning Arg_6-Arg_{13}, precisely those residues known to be critical for binding of Rev to RRE. Preliminary analysis of our structure superimposed onto the modeled RRE-binding site (Leclerc et al., 1994) indicates that the amino acid side chains which have been identified as being required for RNA binding (Figure 9) are suitably placed to form hydrogen bonds with the phosphate backbone of the RNA. However, our solution structure more accurately defines the position of those residues essential for binding (Figure 9), and there are significant differences in the locations of side chains between our structure and the model. A notable feature of the model structure (Leclerc et al., 1994) was that the side chain of Asn_7, which is the residue identified as being most crucial for RRE-binding affinity (Tan et al., 1993), makes no contacts with the RNA. When our structure is superimposed on the model peptide in the RRE-binding site, this residue is in a position to form several hydrogen

bonds with the RNA. This highlights the difficulty of obtaining accurate information on the interaction of biological macromolecules in the absence of direct experimental data. Our structure provides information on the spatial distribution of those functional groups which are involved in binding to RRE. These data allow development of a preliminary pharmacophore model that may help in the development of small molecules which inhibit Rev–RRE binding.

ACKNOWLEDGMENT

We thank Dr. J. McKee (University of Queensland) for assistance with modeling the RNA.

SUPPORTING INFORMATION AVAILABLE

A table of NOE connectivities observed for $Rev_{34-50}R_{42}A$ in 90% H_2O/10% ^2H_2O (1 page). Ordering information is given on any current masthead page.

REFERENCES

Anil-Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.
Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994) *Biochemistry 33*, 2741–2747.
Bax, A., & Davis, D. G. (1985a) *J. Magn. Reson. 65*, 355–360.
Bax, A., & Davis, D. G. (1985b) *J. Magn. Reson. 63*, 207–213.
Bax, A., Griffey, R. H., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301–315.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.
Brünger, A. T. (1992) *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT.
Calnan, B. J., Bianclana, S., Hudson, D., & Frankel, A. D. (1991) *Genes Dev. 5*, 201–210.
Choi, H., & Aldrich, J. V. (1993) *Int. J. Pept. Protein Res. 42*, 58–63.
Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135–175.
Cowburn, D., Live, D. H., Fischman, A. J., & Agosta, W. C. (1983) *J. Am. Chem. Soc. 105*, 7435–7442.
Daly, T. J., Rusche, J. R., Maione, T. E., & Frankel, A. D. (1990) *Biochemistry 29*, 9791–9795.
Fankhauser, C., Izaurralde, E., Adachi, Y., Wingfield, P., & Laemmli, U. (1991) *Mol. Cell. Biol. 11*, 2567–2575.
Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C., & Wong-Staal, F. (1986) *Cell 46*, 807–817.
Hyberts, S. G., Goldberg, M. S., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736–751.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Kjems, J., Calnan, B. J., Frankel, A. D., & Sharp, P. A. (1992) *EMBO J. 11*, 1119–1129.
Leclerc, F., Cedergren, R., & Ellington, A. D. (1994) *Nature Struct. Biol. 1*, 293–300.
Malim, M. H., Tiley, L. S., McCarn, D. F., Rusche, J. R., Hauber, J., & Cullen, B. R. (1990) *Cell 60*, 675–683.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 239*, 129–136.
Olsen, H. S., Cochrane, A. W., Dillon, P. J., Nalin, C. M., & Rosen, C. A. (1990) *Genes Dev. 4*, 1357–1365.
Pallaghy, P. K., Duggan, B. M., Pennington, M. W., & Norton, R. S. (1993) *J. Mol. Biol. 234*, 405–420.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Piotto, M., Suadek, V., & Sklenár, V. (1992) *J. Biomol. NMR 2*, 661–665.
Rance, M., Sørensen, O., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Ratner, L. (1993) *Perspect. Drug Discovery Des. 1*, 3–22.
Rosen, C. A. (1992) *AIDS Res. Hum. Retroviruses 8*, 175–181.
Schnölzer, M., Alewood, P., Jones, A., Alewood, D., & Kent, S. B. H. (1992) *Int. J. Pept. Protein Res. 40*, 180–193.
Sklenár, V., Torchia, D., & Bax, A. (1987) *J. Magn. Reson. 73*, 375–379.
Sodroski, J., Goh, W. C., Rosen, C., Dayton, A., Terwillger, E., & Haseltine, W. A. (1986) *Nature 321*, 412–417.
Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490–5492.
Stutz, F., & Rosbash, M. (1994) *EMBO J. 13*, 4096–4104.
Tan, R., Chen, L., Buettner, J. A., Hudson, D., & Frankel, A. D. (1993) *Cell 73*, 1031–1040.
Wishart, D. S., & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171–180.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647–1651.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley & Sons, New York.
Zapp, M. L., Stern, S., & Greene, M. R. (1993) *Cell 74*, 969–978.

BI942703X