NMR structure of a biologically active peptide containing the RNA-binding domain of human immunodeficiency virus type 1 Tat
(protein structure/RNA-binding proteins/gene regulation/AIDS)

ANWER MUJEEB*, KARL BISHOP*, B. MATUA PETERLIN†‡§, CHRISTOPH TURCK§, TRISTRAM G. PARSLOW‡¶||, AND THOMAS L. JAMES*

Departments of *Pharmaceutical Chemistry, †Internal Medicine, ‡Microbiology and Immunology, and §Pathology, and ||Howard Hughes Medical Institute, University of California, San Francisco, CA 94143

Communicated by Christine Guthrie, May 17, 1994 (received for review March 9, 1994)

ABSTRACT The Tat protein of human immunodeficiency virus type 1 enhances transcription by binding to a specific RNA element on nascent viral transcripts. Binding is mediated by a 10-amino acid basic domain that is rich in arginines and lysines. Here we report the three-dimensional peptide backbone structure of a biologically active 25-mer peptide that contains the human immunodeficiency virus type 1 Tat basic domain linked to the core regulatory domain of another lentiviral Tat—i.e., that from equine infectious anemia virus. Circular dichroism and two-dimensional proton NMR studies of this hybrid peptide indicate that the Tat basic domain forms a stable α-helix, whereas the adjacent regulatory sequence is mostly in extended form. These findings suggest that the tendency to form stable α-helices may be a common property of arginine- and lysine-rich RNA-binding domains.

Human immunodeficiency virus type 1 (HIV-1), like other lentiviruses, encodes a trans-activating regulatory protein, called Tat, that is needed for efficient transcription of the viral genome (1). HIV-1 Tat acts by binding to an RNA stem-loop structure, the trans-activation response element (TAR), found at the 5′ ends of all nascent HIV-1 transcripts. When bound to TAR, Tat alters the properties of the transcription complex, enhancing transcriptional initiation and processivity so that production of full-length viral RNA markedly increases (2–4). The minimal sequences in HIV-1 Tat that can mediate specific TAR binding in vitro have been mapped to a 10-amino acid region (residues 48–57) called the basic domain, composed almost entirely of arginines and lysines (5). Regulatory activity also requires the N-terminal 47 residues of Tat, which can interact with components of the transcription complex and function as a transcriptional activation domain (6–9).

Tat is indispensable for HIV-1 replication, making it an attractive target for antiviral drug development. Efforts to design or optimize inhibitors of HIV-1 Tat would be aided by knowledge of its three-dimensional structure, particularly of those peptide regions most critical for function. However, no detailed structural information has yet been reported for this protein. As an approach to that goal, we have been studying a 25-amino acid hybrid peptide that contains the basic RNA-binding domain of HIV-1 Tat fused to a 15-residue activation domain from the Tat protein of equine infectious anemia virus (EIAV), an unrelated lentivirus. Owing to the short length of the EIAV activation domain, this hybrid peptide is only one-third the size of HIV-1 Tat, yet it can specifically bind and transactivate transcription through the HIV-1 TAR (8). It is the smallest Tat derivative that has been found to retain biological activity.

Here we report studies based on circular dichroism (CD) and two-dimensional proton NMR spectroscopy, which reveal the peptide backbone conformation of this hybrid Tat peptide. Our findings include the unexpected observation that the unliganded HIV-1 Tat basic domain forms a stable α-helix in solution.

MATERIALS AND METHODS
Peptide Synthesis and Purification. The 25-residue hybrid Tat peptide (Fig. 1A) was synthesized using standard 1-fluorenylmethoxycarbonyl (FMOC) chemistry and was purified by preparative reverse-phase HPLC. The N and C termini were unprotected. Its purity was confirmed by analytical reverse-phase chromatography and amino acid analysis. Mass spectral analysis confirmed the proper size of the peptide to be 3209 Da.

CD Spectroscopy. Samples containing ~113 μM peptide were prepared by diluting aliquots of a 1.5 mM aqueous peptide stock with appropriate amounts of water and 1 M NaCl. Samples were assayed at pH 3.5–3.6 under conditions ranging from 0 to 100 mM NaCl and from 5 to 35°C also at pH 7.2 with no added salt at 23°C. Spectra were recorded on Jasco (Easton, MD) J500A and J720 CD spectropolarimeters with either 1- or 5-mm path length cells; each was the average of 20 scans over the wavelength range 184–260 nm with a step resolution of 0.2 nm.

NMR Spectroscopy. Samples for NMR were prepared in 90% H2O/10% 2H2O/0.1 mM EDTA, pH 3.1. We found that sample aggregation could be avoided, while maintaining excellent spectral linewidths, by using 350 μM peptide samples in a 10-mm NMR tube. Two-dimensional nuclear Overhauser effect (NOE) (10) and total coherence transfer (TOCSY) (11,12) spectra were collected on a 600-MHz NMR spectrometer (Varian Unity 600). All experiments were done at 15°C. Two-dimensional NOE spectra were recorded at mixing times of 100, 150, and 250 ms, with a spectral width of 7000.4 Hz, 512 transients, and 2000 data points in the t2 dimension. Data were processed on a Sun Microsystems (Mountain View, CA) Sparcstation using locally written NMR processing software packages (STRIKER version 1.1.1 and SPARKY version 2.0.4). During processing, data were multiplied by a phase-shifted sine-squared function and were zero-filled in both dimensions to achieve a final 2000 x 2000 spectrum. Sequence-specific assignments were obtained much as described by Wüthrich (13). Results of total coherence transfer spectroscopy experiments were used to estab-

Abbreviations: HIV-1, human immunodeficiency virus type 1; TAR, trans-activation response element; EIAV, equine infectious anemia virus; NOE, nuclear Overhauser effect.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
lish amino acid spin networks, and two-dimensional NOE results provided sequential connectivities.

**Distance-Geometry Calculations.** Of the 273 NOE cross-peaks we identified, only 154 observed at the shortest mixing time were selected for analysis, to minimize multispin cross-relaxation effects. Intensities of these peaks were used to estimate the upper bounds for interproton distances as described (14): each peak was semiquantitatively classified as weak, medium, or strong, and the corresponding interproton distance was accordingly assigned an upper-bound value of 5.0, 3.0, or 2.5 Å. These constraints were then used in the distance-geometry program DIANA (15) to generate ensembles of peptide conformations that satisfied our experimental constraints. DIANA uses a variable-target function algorithm to generate structures in accord with the NMR distance constraints (16). Standard minimization parameters for the objective function were applied in DIANA calculations in which 10 levels were used to minimize the variable-target function with weighting factors for all distance constraints set to 1.0. Throughout the calculations, the van der Waals weightings were modulated from 0.2 (for initial levels) to 2.0 (for level 10). Elements from the resulting ensembles of conformations that exceeded the cut-off target function value of 1.5 at level 8 were not accepted for further analysis. A total of 500 structures were generated. The final ensemble discussed in this report was composed of the 27 structures with the lowest values of final target function and the lowest distance constraint violations.

**RESULTS**

As depicted in Fig. 1A, the peptide we examined consisted of the 15-amino-acid minimal-activation domain of ELAV Tat linked through its C terminus to the 10-residue basic domain from HIV-1 Tat. CD spectra were obtained for this peptide over the wavelength range of 184–260 nm at a concentration of 113 µM and under various salt concentrations (0–100 mM NaCl) and temperatures (5–35°C) at pH 3.5–3.6. CD analysis of the Tat peptide (113 µM) was also done at pH 7.2 at 23°C with no added salt. Under all of these conditions, the spectra exhibited two minima (see Fig. 1B), one centered at 222 nm and another at 208 nm. These features suggest the presence of α-helical structure (17–19). Quantitation of the degree of helicity from these spectra is compromised by the two tyrosines in our sequence because tyrosine shows a positive band at ~225 nm (17). Therefore, only a qualitative estimate of α-helical content can be obtained (20). The CD spectra suggest that approximately half of the peptide has α-helical character in aqueous solution, being similar at pH 3.5 or pH 7.2. Addition of 20% (vol/vol) trifluoroethanol to the solution enhanced the α-helical characteristics (spectra not shown).

Analysis of two-dimensional total coherence transfer spectroscopy spectra from the peptide enabled us to identify the spin networks corresponding to each of the 25 amino acid residues, beginning with the distinctive signals from I13. From this starting point, inter-residue peaks involving CaH and NH protons in the fingerprint region of the two-dimensional NOE spectra (Fig. 2 A) provided a sequential walk in both directions along the rest of the peptide backbone. Sequential dNN(i + 1) cross-peaks (Fig. 2B) were identified from residues Q4 to S10 and from L11 to K18. Due to almost complete overlap of resonances from K18 and K19 with R20, dNN(i + 1) cross-peaks could not be observed for these residues. Additional dNN(i + 1) connectivities could be traced from residues R20–R23 but could not be traced between R23 and R24 due to signal overlap. A weak NH–NH cross-peak was visible in 250-ms two-dimensional NOE spectra for residues R24 and R25.

The assignments of side-chain proton resonances were mainly achieved by using either backbone amide-to-sidechain or CaH-to-side-chain cross-peaks. Resonances from arginine NH2 protons of guanidine groups showed almost complete overlap at ~7.20 ppm and 6.61 ppm and were not individually distinguishable.

To obtain an estimate of peptide secondary structure from these NMR data, we applied the chemical-shift index strategy of Wishart et al. (21), which is based on the observation that the chemical shifts of CaH resonances strongly depend on protein secondary structure. Empirical testing indicates that stretches of four or more residues having Ca proton resonances shifted upfield by 0.1 ppm or more relative to the corresponding random-coil Ca proton chemical shifts are indicative of an α-helix (21, 22). We found that CaH resonances from the N-terminal half of our peptide failed to meet this criterion, so that its structure could not be deduced by this method. By contrast, resonances from residues 14–25 were all shifted significantly upfield, as indicated by a down-
ward bar in the graph of Fig. 3. This result strongly suggests that the HIV-1-derived portion of the peptide has a stable α-helical structure, which likely accounts for the partial helical character seen in the CD spectra.

We then estimated upper bounds on interproton distances corresponding to each of the two-dimensional NOE cross-peaks (vide supra) and used these to determine the three-dimensional structure of the peptide backbone through dis-
FIG. 3. Chemical-shift indices of the Ca proton resonances from each of the 25 residues in the peptide, derived from NMR spectra. By convention (21), resonances shifted upfield by 0.1 ppm or more with respect to the values expected for a random coil are assigned an index of -1, whereas downfield shifts of similar magnitude are indexed as 1.

tance–geometry calculations. This approach yielded an ensemble of 27 highly convergent structures, which are shown superimposed in Fig. 4. All coincide closely with one another, indicating that residues 14–25 are α-helical and that residues 1–13 are in a relatively extended conformation with a tendency for turn formation, so that the peptide is roughly L-shaped overall. The global folding topology of the peptide structures resulting from distance–geometry calculations was in excellent agreement with our experimental constraints. It should be noted that we have observed specific medium–long-range NOEs in our NMR data that account for the relative orientation of C and N termini and also lead to the surprisingly well-defined structure of the extended loop region. For example, NOEs between the backbone amide proton of Q4 and CaH of D14, between CaH of G12 and the D14 amide proton, and also between residue Y1 to residues F7 and L8 were observed. In conjunction with these, amide–amide contacts between residues C6 and G12 and also several \( i - (i + 3) \) contacts in the N-terminal region of the peptide were apparent. For any of the distance–geometry structures, only six to seven NMR constraints were violated by \( >0.25 \) Å, and no individual constraint was consistently violated in the whole structure ensemble. Global backbone pairwise atomic root-mean-squared deviation for the 27 best-converged structures was 0.37 Å. The pairwise root-mean-squared deviation value in the helical portion of the peptide (residues 14–25) was 0.21 Å.

DISCUSSION

We have used two-dimensional NMR spectroscopy in concert with distance geometry to determine the backbone solution structure of a peptide consisting of the EIAV Tat core domain fused to the RNA-binding basic domain of HIV-1 Tat. The EIAV sequence in this peptide is similar to its HIV-1 counterpart, so the peptide as a whole bears a resemblance to residues 32–57 of HIV-1 Tat (Fig. 1A). Hence, its structure may well approximate that of a significant portion of the native HIV-1 protein. Unlike the corresponding HIV-1 sequence, however, the hybrid peptide has been reported to be biologically active (8). Consequently, it may offer a uniquely tractable model for studying the structural requirements for both TAR binding and transcriptional regulation by Tat.

FIG. 4. Stereoviews of an ensemble of the 27 peptide backbone structures for the hybrid Tat peptide, generated using the distance-geometry program DIANA, which exhibited the best target function and NOE constraints violation values. Average pairwise root-mean-squared deviation for all backbone atoms in these structures is 0.37 Å. (A) Side view. (B) Top view. The locations of Ca atoms for selected residues are indicated.
The backbone structure of the peptide is depicted in Fig. 4 A and B. The most notable feature of the structure is an α-helix formed by residues 14–25, a region that encompasses the entire HIV-1 basic domain and whose sequence differs at only one position from residues 44–57 of HIV-1 Tat. It is this region that mediates specific HIV-1 TAR recognition by both the peptide and Tat protein (5, 8). The EIAV activation sequence at the N-terminal end of the peptide has a relatively extended structure, but it is best represented by a loose reverse turn whose long axis is disposed roughly perpendicularly to that of the helix, imparting an overall L-shaped conformation to the peptide.

The results we obtained from CD spectroscopy, NMR chemical-shift index, and distance-geometry calculations all support the view that the basic domain in the hybrid peptide forms an α-helix, at least under the conditions we studied. It appears unlikely that the helicity is imposed on this region by the adjacent nonhelical EIAV sequence. However, it is possible that the EIAV sequence stabilizes the helix. This stabilization may be aided, for example, by an aspartate residue at position 14 of the hybrid peptide in place of the serine at the corresponding position in the HIV-1 core domain. Although we cannot be certain that the basic domain adopts this same conformation in the context of native HIV-1 Tat, that possibility is compatible with the limited data available. In particular, earlier CD analyses have revealed an overall helix content of 15–20% for full-length HIV-1 Tat (23) and a marked helix propensity in a peptide representing HIV-1 Tat residues 38–60 (24); shorter basic-region peptides were reported to be unstructured in the absence of ligand but to acquire an undetermined structure upon binding TAR (25).

Our data do not address the possible role that helicity might play in defining the affinity and specificity with which Tat binds to TAR. However, chemical footprinting and NMR-based modeling studies suggest that the primary Tat-binding site lies within a wide major groove induced by local distortion of the double-stranded stem in TAR RNA (26, 27); such a cleft could be of sufficient width to accommodate the putative basic helix of Tat. This suggests a model in which HIV-1 Tat binds its target by inserting the positively charged recognition helix into the RNA major groove, in a manner resembling that used by many DNA-binding proteins (28). HIV-1 Tat is one of several basic DNA-binding proteins whose recognition domains include 10- to 15-amino acid sequences rich in arginines and lysines (29, 30). Virtually all lentiviral Tat, Rev, or Rex proteins are of this type, as are the bacteriophage N proteins. The basic domain in each of these proteins appears critical for RNA binding. Recently, short peptides representing the basic domain from HIV-1 Rev were found to have a strong propensity for α-helix formation, as judged by CD, and helicity correlated strongly with their ability to bind the specific Rev target RNA (31, 34). NMR studies of the EIAV Tat protein further suggest that the basic domain in this protein, too, has a helical conformation (22, 32). Together with our findings, these results suggest that the tendency for helix formation may be an attribute shared by many basic RNA-binding domains. It remains to be determined whether and how this conformation changes as a consequence of specific RNA binding (33).