Sequence-Specific Resonance Assignments of the ^{1}H-NMR Spectra of a Synthetic, Biologically Active EIAV Tat Protein†

D. Willbold,‡ U. Krüger,§ R. Frank,§ R. Rosin-Arbesfeld,‖ A. Gazit,‖ A. Yaniv,‖ and P. Rösch*,‡

Lehrstuhl für Struktur und Chemie der Biopolymere, Universität Bayreuth, Postfach 10 12 51, 95440 Bayreuth, Germany, Zentrum für Molekulare Biologie Heidelberg (ZMBH), INF, W-6900 Heidelberg 1, Germany, and Sackler Faculty of Medicine, Department of Human Microbiology, Tel Aviv University, Ramat-Aviv, 69978 Tel Aviv, Israel

Received February 15, 1993; Revised Manuscript Received May 18, 1993

ABSTRACT: The equine infectious anemia virus (EIAV) *trans*-activating (Tat) protein is a close homologue of the human immunodeficiency virus (HIV) Tat protein. Both of these proteins bind to an RNA *trans*-activation responsive element (TAR). We synthesized chemically a protein with the sequence of the 75 amino acid Tat protein from EIAV. The chemically synthesized protein was shown to be biologically active. Circular dichroism (CD) and ^{1}H nuclear magnetic resonance (NMR) spectroscopy were used to structurally characterize the synthetic protein. We obtained nearly complete resonance assignments in the 2D-NMR spectra of the protein at pH 3.0. There is at least some evidence from the experimental data that the basic TAR binding domain of the synthetic protein has a tendency to form a helix, but our experiments also indicate that the protein probably does not have an overall stable tertiary structure in aqueous solution at this pH. CD spectroscopy suggested that the protein adopts a more stable, predominantly α-helical structure in a trifluoroethanol/water solution.

Retroviruses are eukaryotic viruses which carry their genetic information as RNA sequences rather than DNA sequences. On infection of a host cell, the viral RNA is transcribed into host-cell DNA by means of an enzyme called reverse transcriptase. Retroviruses remain dormant in their host cell for a period which can typically last up to several years. Then, the viral DNA is transcribed back to viral RNA, and the viruses multiply within the host cell.

On the basis of their pathogenic potential, the family of retroviruses can be divided into three different subfamilies, namely, the oncoviruses, the lentiviruses, and the spumaviruses. Oncoviruses, such as the Rous sarcoma virus (RSV)[1] and the human T-cell leukemia viruses (HTLV), are cancer-causing agents. Lentiviruses, or slow viruses, such as the equine infectious anemia virus (EIAV) and the human immunodeficiency virus (HIV), are non-neoplastic pathogens. Spumaviruses do not have any known pathogenic potential.

The lentivirus subfamily not only includes HIV and EIAV but extends to several HIV- and EIAV-related viruses, such as the simian immunodeficiency virus (SIV), the bovine immunodeficiency virus (BIV), the feline immunodeficiency virus (FIV), the sheep visna virus (VILV), the caprine arthritis encephalitis virus (CAEV), and the ovine lentivirus (OMVV).

Retroviruses use a complex machinery of RNA-binding proteins to regulate transcription of DNA to viral RNA and thus replication of the viruses. The best known lentiviral transcriptional regulators are the *tat*, *rev*, and *nef* gene products. The *tat* gene product activates DNA transcription in *trans* (*tat* = *trans*-activator). It is a key regulatory protein, as it positively regulates expression of all viral genes. Thus, it is of utmost importance to understand the mechanism of transcriptional activation by the *tat* gene product in atomic detail in order to be able to design Tat protein inhibiting drugs for the prevention of the proliferation of lentiviruses and thus, ultimately, for the prevention of the outbreak of lentivirus-caused diseases. In addition, because of their size of around 100 amino acids, Tat proteins may well be suited to serve as model compounds for general biophysical studies of protein–RNA interactions.

From a comparison of Tat protein sequences it can be inferred that Tat proteins of HIV-1 and EIAV contain several characteristic sequence elements. Both proteins show homology in sequence [17% identity according to the program "align" from Myers and Miller (1988)] and domain structure. The N-terminal domain is expected to develop an amphipathic helix comparable to the acidic activator domain of transcription factors. The core domain is highly conserved among Tat proteins from different lentiviruses (Dorn et al., 1990).

All known immunodeficiency viruses seem to contain a cysteine-rich sequence element with four or five cysteines in a stretch of eleven amino acids. This sequence element is responsible for metal-mediated dimerization of the protein observed in vitro (Frankel et al., 1988). The Tat proteins from the OMVV, the VILV, the CAEV, and the EIAV do not contain this sequence, although the OMVV, VILV, and CAEV Tat proteins all contain a short C-terminal stretch of six amino acids with three cysteines.

† D.W. was supported by a grant from the Fonds der Chemischen Industrie (FCI) and a MINERVA travel fund.

* To whom correspondence should be addressed.
‡ Universität Bayreuth.
§ Zentrum für Molekularbiologie Heidelberg (ZMBH).
‖ Tel Aviv University.

[1] Abbreviations: 2D, two-dimensional; CAEV, caprine arthritis encephalitis virus; CAT, chloramphenicol acetyltransferase; CD, circular dichroism; clean-TOCSY, TOCSY with suppression of NOESY-type cross peaks; COSY, correlated spectroscopy; DQF-COSY, double quantum filtered COSY; DSS, 2,2-dimethyl-2- silapentane-5-sulfonate EIAV, equine infectious anemia virus; FIV, feline infectious virus; Fmoc, 9-fluorenylmethyloxycarbonyl; HIV, human immunodeficiency virus; HPLC, high-performance liquid chromatography; LTR, long terminal repeat; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; OMVV, ovine lentivirus; PyBop, (benzotriazol-1-yl)-*N*-oxytripyrrolidinophosphonium hexafluorophosphate; RSV, Rous sarcoma virus; HTLV, human T-cell leukemia viruses; SIV, simian immunodeficiency virus; SIVM, macaque SIV; Tat, *trans*-activator; TFE, trifluoroethanol; TAR, *trans*-activation response element; TOCSY, total coherence transfer spectroscopy; TPPI, time proportional phase incrementation; VILV, sheep visna virus.

0006-2960/93/0432-8439$04.00/0

Common to all known lentiviral Tat proteins is a sequence of about 10–13 amino acids containing mostly basic residues. This basic domain is essential for binding to the RNA recognition sequences (TAR, for Tat response element), and the glutamine-rich domain was suggested to be important for RNA binding at least in the case of EIAV Tat (Derse et al., 1991). However, data by Noiman et al. (1991) demonstrated that this sequence was not essential for Tat activity.

Another peculiarity of at least the human immunodeficiency virus *tat* gene product is its RNA recognition sequence, which contains a characteristic bulge region not present in the RNA recognition sequence of, e.g., EIAV Tat protein. It has been shown recently that a synthetic peptide identical to the basic region of HIV Tat protein is able to bind to the HIV TAR sequence with a dissociation constant in the nanomolar range (Calnan et al., 1991).

Binding to the TAR sequence was also shown for arginine as a free amino acid and argininamide with a dissociation constant of the order of several millimolar (Tao & Frankel, 1992; Puglisi et al., 1992). The basic region is most probably also responsible for nucleolar localization of the Tat proteins (Dang & Lee, 1989; Endo et al., 1989; Hauber & McCullen, 1989; Ruben et al., 1989; Siomi et al., 1990).

As a first step to understanding the structural features of lentiviral Tat proteins, we synthesized a peptide of the EIAV Tat protein sequence and began a preliminary structural characterization of this protein by CD as well as by one- and two-dimensional NMR spectroscopy in acidic aqueous solution.

MATERIALS AND METHODS

Peptide Synthesis and Activity Test. The EIAV Tat protein (assuming initiation at the first CUG codon in the open reading frame; Dorn et al., 1990; Noiman et al., 1990) was synthesized on polystyrene–polyethylene glycol copolymer (Tentagel) in an automated continuous flow peptide synthesis instrument using Fmoc (9-fluorenylmethyloxycarbonyl) chemistry and PyBop [(benzotriazol-1-yl)-*N*-oxytripyrrolidinophosphonium hexafluorophosphate] as activator (Frank & Gausepohl, 1988). The product was purified by reversed-phase high-performance liquid chromatography (HPLC) to a purity of more than 95%. The synthetic protein was freeze-dried for storage after purification.

Semiconfluent canine thymus cells, Cf2Th (ATCC CRL 1430), in 35-mm dishes were transfected (Sherman et al., 1988) with 5 mg of the reporter plasmid, pEIAV LTR–CAT, in which the bacterial chloramphenicol gene is governed by the EIAV long terminal repeat (LTR) (Sherman et al., 1988). Forty-eight hours later, the medium was removed, and various amounts of the synthetic Tat protein in phosphate-buffered saline were added to the cells. Cells were immediately scraped with a rubber policeman and resuspended in medium containing 10% fetal calf serum. After 24 h of incubation, the chloramphenicol acetyltransferase (CAT) assay was performed as previously described (Sherman et al., 1988).

Sample Preparation. The product of the protein synthesis procedure could be readily dissolved in aqueous 25 mM phosphate buffer, pH 3.0. Unfortunately, very broad lines in the ^{1}H-NMR spectrum indicated that the protein formed oligomers in the pH 3.0 solution. Consequently, the resulting sample was unusable for further NMR studies under these conditions. We thus had to apply a denaturation/renaturation procedure: The freeze-dried product was dissolved in 0.5 mL of aqueous 25 mM phosphate buffer, pH 3.0, to form a 3 mM protein solution. This solution was diluted to a volume of 7.5 mL and incubated overnight to allow for folding of the protein. The solution was subsequently concentrated in a SpeedVac apparatus under vacuum to a concentration of 2 mM. The ^{1}H-NMR spectrum of the resulting protein solution was well resolved, as expected for a protein this size. A rationale for this procedure was the observation with other proteins that the formation of oligomers was fast compared to the refolding process (Kiefhaber et al., 1991).

CD Spectroscopy. CD spectra were recorded from 260 to 180 nm at 20 nm/min on a Jasco J 600 CD spectrometer. The sample contained 3.3 mg/mL protein in 10 mM sodium phosphate, pH 3.0. The required sample volume was 30 μL. The reference sample contained buffer without protein. Eight scans were accumulated from sample and reference, respectively.

NMR Spectroscopy. A series of NMR experiments in H_2O/D_2O (9:1) solution, pH 3.0, was performed on Bruker AM 500 and Bruker AMX 600 instruments: double quantum filtered correlated spectroscopy (DQF-COSY; Rance et al., 1983), nuclear Overhauser effect spectroscopy (NOESY; Bodenhausen et al., 1984), total coherence transfer spectroscopy with MLEV-17 mixing cycles (TOCSY; Braunschweiler & Ernst, 1983; Davies & Bax, 1985), and clean-TOCSY (Griesinger et al., 1988) with a spin-lock field of about 12 kHz. The spectra were acquired in the phase-sensitive mode with quadrature detection in both dimensions using the time-proportional phase incrementation technique (TPPI; Marion & Wüthrich, 1983). Solvent suppression was performed by continuous coherent irradiation prior to the first excitation pulse and during the mixing time in the NOESY experiment.

Acquisition and Processing Parameters. Typical spectral parameters: frequency width, 6024 Hz; data size, 2K × 1K data points; a sinebell-squared filter with a phase shift of $\pi/3$ was applied before Fourier transformation for the DQF-COSY experiment; mixing times were 100, 200, and 300 ms for the NOESY experiments and 66 ms for the TOCSY and CLEAN-TOCSY experiments; sample temperature was kept at 18 °C with a standard VT1000 (AM 500) and VT2000 (AMX 600) Bruker temperature control unit using a stream of precooled air. In addition to the standard Bruker software, the commercial NDee software package (Herrmann et al., unpublished; Biostructure, Strassbourg) for data processing on Evans & Sutherland workstations, DECstations, and SPARCstations was used. Chemical shift values are reported in ppm from 2,2-dimethyl-2-silapentanesulfonic acid (DSS).

Secondary Structure Predictions. Secondary structure predictions have been carried out using three algorithms as supplied with the program package SYBYL 5.4 (Tripos Associates, Inc.), namely, Bayes statistics (Maxfield & Scheraga, 1976), information theory (Gibrat et al., 1987), and a neural net (Qian & Sejnowski, 1987). In addition, the Chou–Fasman algorithm (Chou & Fasman, 1974) as supplied with the FASTA program package (Pearson and Lipman, 1988) and the PHD neural network structure prediction method (Rost & Sander, 1993) as obtained via the European Molecular Biology Laboratory (EMBL) mail server were applied.

RESULTS AND DISCUSSION

Biological Activity of Tat Protein. The biological activity of the chemically synthesized protein was assessed by its ability to activate the EIAV long terminal repeat. The "scrape loading" approach (Gentz et al., 1989) was adopted to introduce the synthetic, purified Tat protein into canine cells containing the EIAV LTR–CAT. Slight *trans*-activation of as little as 200 ng of Tat protein could be detected, and

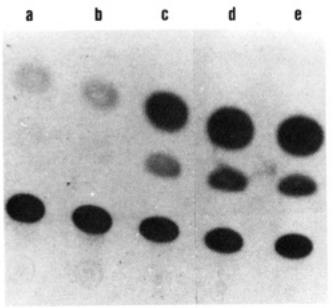

FIGURE 1: Biological activity of the chemically synthesized Tat protein. Semiconfluent canine thymus cells were transfected with pEIAV LTR-CAT (5 μg). Forty-eight hours later, the cells were mock-introduced (a) or introduced with 10 μg of a bacterially synthesized heterologous retroviral matrix peptide (b) or with various amounts (c, 200 ng; d, 1 μg; e, 10 μg) of the chemically synthesized Tat. After 24 h, CAT activity was assayed. Percent conversions, representing the amount of [^{14}C]chloramphenicol converted to acetylated derivatives, are as follows: a, 6; b, 6.3; c, 39; d, 68; e, 70.

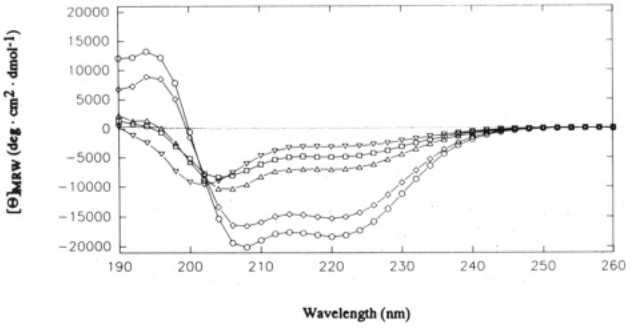

FIGURE 2: Far-UV circular dichroism spectrum of EIAV Tat protein in H_2O and in a TFE/H_2O mixture. The percentages of TFE (v/v) were 0% (▽), 10% (□), 20% (△), 40% (▲), and 80% (○). Other experimental conditions were as in Material and Methods.

increasing concentrations stimulated CAT expression in a dose-dependent manner (Figure 1). The samples used for the activity tests were prepared in a manner identical to the NMR samples.

CD Spectra. A series of CD and one-dimensional proton NMR measurements under various temperature and pH conditions led to the definition of the optimal parameters for the NMR experiments in H_2O. The CD spectra do not necessarily lead to the conclusion that the protein possesses a clear-cut overall secondary structure in H_2O and phosphate buffer, pH 3.0. In contrast, it can be seen from the CD spectra obtained in H_2O/phosphate buffer solution with varying amounts of trifluoroethanol (TFE) (Figure 2) that EIAV Tat assumes a predominantly helical structure under these conditions. The degree of helicity is dependent on the concentration of TFE and increases with increasing TFE concentration as judged from the CD spectra. As a starting point for further structural studies, we decided to proceed with the resonance assignments at pH 3.0.

Assignment Procedure. The 1D-NMR spectrum was well resolved after the unfolding/refolding procedure we employed (Figure 3). Sequence-specific assignments of resonances in H_2O/D_2O solution were performed using standard 2D-NMR procedures (Englander & Wand, 1987; Wüthrich, 1986). All assignments were made on the basis of DQF-COSY, clean-TOCSY (66 ms), and NOESY (100, 200, and 300 ms) spectra in an H_2O/D_2O (9:1) mixture. Spin systems within an amino acid residue were defined by the DQF-COSY and the clean-TOCSY spectra, and the NOESY spectra were used to determine the sequence-specific resonance assignments (Figures 4 and 5).

We began the sequence-specific assignment procedure with an A_3B_3MX spin system, which turned out to represent V21, as the protein backbone resonances could easily be traced down to K16 via a complete set of H_α–HN NOEs. Additional confirmation of this assignment was achieved by observation of HN–HN NOEs from G19 to V21. The assignment of P22 could be achieved by NOE cross peaks leading from H_α(V21) to H_δ(P22). The following amino acids up to Q29 could be assigned by using H_α-HN NOEs of sequential residues. The assignment could be confirmed by observation of HN–HN NOEs from G23 to G28 and NOEs between HN(N25) and H_β/H_γ(Q24) as well as between HN(T26) and H_β(N25). Assignment of E30 was not possible. A31 and R32 were assigned using H_α/H_β(A31)–HN(R32) NOEs. P33 showed NOEs between H_δ and H_α(R32). A2 could be identified by an NOE cross peak to an H_α that did not show any connection to an amide proton but to the H_β of a leucine spin system (L1). H_α–HN NOEs could be found for residues A2 to I6. An additional cross peak indicating an H_β(R4)–HN(R5) NOE was found.

H_α–HN NOEs of sequential residues lead from P7 to A10. Partial confirmation was achieved by observation of H_γ(T9)–HN(A10). E11 was assigned using the H_β(A10)-HN(E11) NOE. N13 and L14 were assigned by H_α(N13)–HN(L14). Assignment of E12 and Q15 was not possible. From N34 to Q66 the whole chain could be traced via H_α–HN NOEs. Sequential HN–HN NOEs were found for the pieces Q38 to C40, F41 to R43, I47 to Y49, L50 to D51, S53 to R55, K56 to N58, and I65 to Q66. Additional confirmation was obtained

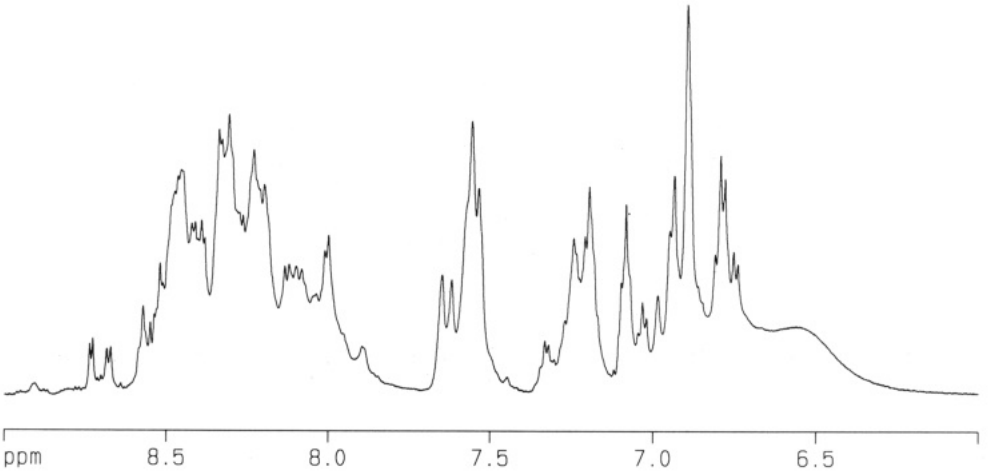

FIGURE 3: 1D-NMR spectrum of the EIAV Tat protein recorded at 291 K on a Bruker AMX 600. The sample contained 450 μL of a 2 mM solution of Tat in sodium phosphate, pH 3.0, and 90% H_2O/10% D_2O.

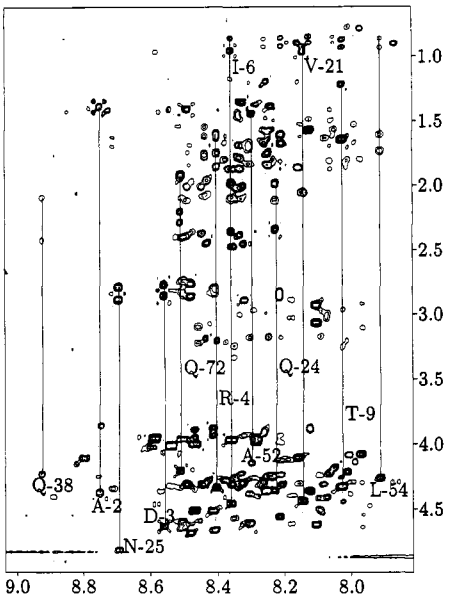

FIGURE 4: Clean-TOCSY (Griesinger et al., 1988) spectrum recorded at 291 K on a Bruker AMX 600 spectrometer with a spin-lock mixing time of 66 ms. The sample contained 450 μL of a 2 mM solution of Tat in sodium phosphate, pH 3.0, and 90% H_2O/10% D_2O.

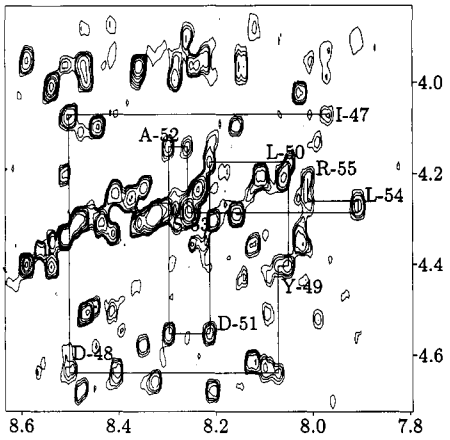

FIGURE 5: Fingerprint region of a NOESY spectrum recorded on a Bruker AMX 600 spectrometer at 291 K with a mixing time of 300 ms. The lines demonstrate the procedure for the sequential assignment from I47 to R55, i.e., part of the core region/basic domain of the Tat sequence.

using the following NOEs between side-chain protons and amide protons of sequential amino acids: H_β(F41)–HN(L42), H_γ(I47)–HN(D48), H_β(D48)–HN(Y49), H_β(Y49)–HN(L50), H_β(D51)–HN(A52), H_β(A52)–HN(S53), H_β(S53)–HN(L54), H_β(N58)–HN(K59), H_β(A64)–HN(I65), and $CH_{3\gamma}$(I65)–HN(Q66).

Q67 could not be assigned. By the observation of sequential H_α–HN NOEs and an HN–HN NOE between R69 and Q70, residues 68–70 were assigned. P71 could be assigned by an H_α(Q70)–H_δ(P71) NOE, and sequential H_α–HN NOEs identified the C-terminal sequence of the protein up to L75. Remaining ambiguities could be resolved by the observation of H_δ(P71)–HN(Q72), HN(Q72)–HN(Y73), and H_β(Y73)–HN(L74). For the results of the assignment procedure, see Table I. A summary of the backbone NOEs is given in Figure 6.

Structural Information. The EIAV Tat protein is highly homologous to the HIV Tat protein, except for the cysteine-rich region of 16 amino acids which is present as a sequence insert in the HIV Tat protein. This can be seen in the sequence alignment of Tat proteins from four different sources according to the ClustalV program by Higgins et al. (1991) in Table II

Table I: Sequence-Specific Resonance Assignments

		HN	C_αH	C_βH	C_γH	others
Leu	1		4.01	1.69/1.74	1.68	0.95/0.97
Ala	2	8.75	4.37	1.39		
Asp	3	8.56	4.63	2.77/2.85		
Arg	4	8.41	4.32	1.75/1.86	1.61	3.20(δ), 7.23(δHN)
Arg	5	8.44	4.32	1.77	1.54	3.19(δ), 7.28(δHN)
Ile	6	8.36	4.45	1.87	1.17/1.49	0.95(δ), 0.86(δ)
Pro	7		4.44	n.a.	n.a.	3.65(δ), 3.83(δ)
Gly	8	8.54	4.02			
Thr	9	8.03	4.33	4.25	1.22	
Ala	10	8.49	4.31	1.41		
Glu	11	8.36	4.30	1.99/2.11	2.36/2.47	
Glu	12	n.a.[a]	n.a.	n.a.	n.a.	
Asp	13	8.48	4.68	2.76/2.86		7.63/6.95
Leu	14	8.21	4.31	1.67	1.60	0.86/0.92
Gln	15	n.a.	n.a.	n.a.	n.a.	n.a.
Lys	16[b]	8.25	4.24	n.a.	n.a.	2.99(ε), 7.6(εHN)
Ser	17	8.42	4.51	3.87/3.93		
Ser	18	8.47	4.51	3.89/3.95		
Gly	19	8.46	4.00			
Gly	20	8.28	3.94/3.99			
Val	21	8.14	4.44	2.06	0.92/0.97	
Pro	22		4.41	1.94/2.31	1.97/2.07	3.70/3.92
Gly	23	8.59	3.98/3.93			
Gln	24	8.22	4.35	1.98/2.11	2.34	7.59/6.91
Asn	25	8.70	4.82	2.79/2.89		7.66/6.97
Thr	26	8.26	4.36	4.33	1.19	
Gly	27	8.54	4.00			
Gly	28	8.36	3.96			
Gln	29	8.49	4.30	2.01/2.12	2.39	6.95/7.60
Glu	30	n.a.	n.a.	n.a.	n.a.	
Ala	31	8.33	4.29	1.35		
Arg	32	8.31	4.61	1.72/1.84	1.68	3.17(δ), 7.20(δHN)
Pro	33		4.33	1.68/2.18	1.97	3.62/3.77
Asn	34	8.52	4.64	2.82/2.40		
Tyr	35	8.10	4.50	2.90/2.96		7.04(C2/6H), 6.76(C3/5H)
His	36	8.46	4.59	3.22/3.10		7.21(C2H), 8.59(C4H)
Cys	37	8.35	4.57	3.24/3.33		
Gln	38	8.92	4.22	2.10	2.44	n.a.
Leu	39	8.71	4.34	1.71	1.61	0.93
Cys	40	8.02	4.44	2.96/3.25		
Phe	41	8.09	4.52	3.08/3.17		7.22(C2/6H), 7.35(C3/5H), 7.29(C4H)
Leu	42	7.99	4.13	1.66	1.54	0.88/0.91
Arg	43	8.21	4.24	1.81	1.60	3.18(δ), 7.25(δHN)
Ser	44	8.12	4.37	3.88		
Leu	45	8.03	4.31	1.64	1.53	0.87/0.93
Gly	46	8.20	3.96/3.99			
Ile	47	7.97	4.07	1.79	1.13/1.41	0.77(γ), 0.85(δ)
Asp	48	8.51	4.64	2.74/2.85		
Tyr	49	8.07	4.41	3.00/3.03		7.09(C2/6H), 6.82(C3/5H)
Leu	50	8.05	4.17	1.57	1.49	0.85/0.89
Asp	51	8.21	4.56	2.83/2.87		
Ala	52	8.30	4.14	1.44		
Ser	53	8.26	4.29	3.90		
Leu	54	7.90	4.25	1.60	1.74	0.86/0.93
Arg	55	8.01	4.21	1.81	1.60	3.20(δ), 7.27(δHN)
Lys	56	8.07	4.23	n.a.	n.a.	2.99(ε), 7.60(εHN)
Lys	57	8.23	4.23	n.a.	n.a.	2.98(ε), 7.60(εHN)
Asn	58	8.41	4.66	2.79/2.82		7.67/7.00
Lys	59	8.33	4.22	n.a.	n.a.	3.00(ε), 7.60(εHN)
Gln	60	8.43	4.27	1.96/2.07	2.44	n.a.
Arg	61[b]	8.38	4.27	1.80	1.61	3.21(δ), 7.26(δHN)
Leu	62[b]	8.26	4.33	n.a.	n.a.	0.86/0.93
Lys	63	8.35	4.25	n.a.	n.a.	2.99(ε), 7.60(δ)
Ala	64	8.24	4.29	1.38		
Ile	65	8.16	4.09	1.86	1.20/1.51	0.90(γ), 0.86(δ)
Gln	66	8.44	4.30	2.01/2.09	2.37	6.91/7.57
Gln	67	n.a.	n.a.	n.a.	n.a.	n.a.
Gly	68	8.47	3.96			
Arg	69	8.25	4.35	1.74/1.84	1.58/1.64	3.19(δ), 7.22(δHN)
Gln	70	8.51	4.59	1.92/2.10	2.40	6.92/7.55
Pro	71		4.35	1.72/2.22	1.99	3.65/3.79
Gln	72	8.51	4.20	1.92	2.21/2.29	6.91/7.54
Tyr	73	8.11	4.61	2.93/3.07		7.11(C2/6H), 6.80(C3/5H)
Leu	74	8.12	4.36	1.57	1.49	0.85/0.89
Leu	75	8.01	4.08	1.70	1.56/1.81	0.77

[a] Not assigned. [b] Assignment is preliminary.

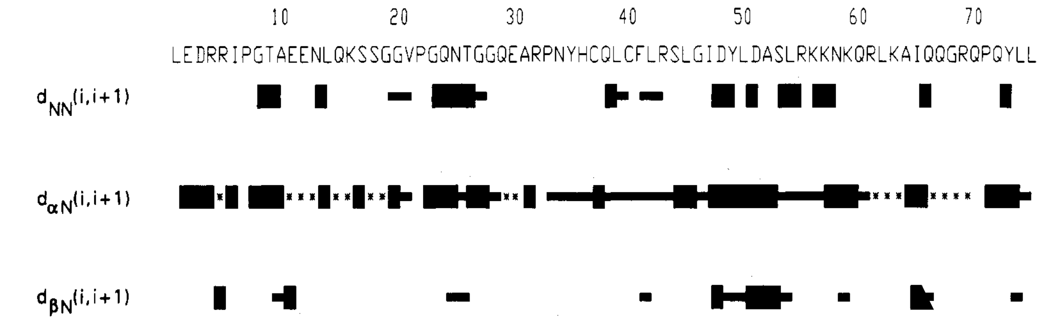

FIGURE 6: NOE backbone connectivities NH–NH, C_αH–NH, and C_βH–NH. Medium-range NOEs could not be observed.

Table II: Sequence Alignment of Tat Proteins[a]

```
                10         20         30         40         50         60         70
HIV10 MEPV----DP RLEPWKHPGS ---------- ---------- ----QPKTAC TN-CYCKKCC FHCQVCFITK
HIV21 METPLKEPES SLESYNEPSS CTSERDVTAQ ERAKQGEELL AQLHRPLEAC TNSCYCKQCS YHCQLCFLKK
SIVMK METPLREQEN SLESSNERSS CILEADATTP ESANLGEEIL SQLYRPLEAC YNTCYCKKCC YHCQFCFLKK
EIAV  LADR-RIPGT AEENLQKSSG GVPGQNTGGQ EA-------- ----RPN--- ---------- YHCQLCFL-R
         .          *   .  ..                         .•                .*** **. .
                80         90        100        110        120        130
HIV10 ALGISYGRKK RRQRRRPPQG SQTHQVSLSK QPTSQSRGDP TGPKE----- ---------
HIV21 GLGIWYARQG RR--RRTPRK TKTHPPPASD KS-ISTRTGD SQPTKKQKKT PETTVVSAC
SIVMK GLGICYEQSR KR--RRTPKK AKANTSSASN NRLIPNRTRH CQPEKAKKET VEKAVATAP
EIAV  SLGIDYLDAS LRKKNKQRLK AIQQGRQ--- ---------- ---------- --------P
      .*** *          *    .     .  .
```

[a] Simultaneous alignment of Tat proteins from different sources according to the ClustalV program using the standard set of parameters (Higgins et al., 1991). Bold and underlined regions are of special interest, namely, the cysteine-rich region in immunodeficiency virus Tat proteins (aa 50–60), the highly homologous core region (aa 61–76), and the C-terminal basic domain (80–92), which is most probably the RNA-binding region. Dots under the sequences (·) indicate conservative amino acid replacements; asterisks (*) indicate identical amino acids. The sequences shown are as follows: HIV10, human immunodeficiency virus type 1 (BH10 isolate; HIV-1) (Ratner et al., 1985); HIV21, human immunodeficiency virus type 2 (D194 isolate; HIV-2) (Kuehnel et al., 1989); SIVMK, simian immunodeficiency virus (K6W isolate; SIV-MAC) (Franchini et al., 1987); and EIAV, equine infectious anemia virus (Isolate Wyoming; EIAV) (Noiman et al., 1990).

Table III: Secondary Structure of the EIAV Tat Protein[a]

```
           10         20         30         40         50         60         70
  LADRRIPGTA EENLQKSSGG VPGQNTGGQE ARPNYHCQLC FLRSLGIDYL DASLRKKNKQ RLKAIQQGRQ PQYLL
1 AA-A------ AAAAB----- ---------- ------AA-A B--A---AAA AAAAAAAAAA AAAAAA---- -----
2 BBTTB----A AAAATTTTT- --T------- --TTTTTBBB BBTTTTBBBA AAAAAAAAAA ABBBBBTTT- -TBBB
3 ---------- -AAA------ ---------- ------AAAA AA-----AAA AAAAAAAA-- AAAAA----- -----
4 AAAT---T-A AAAAAT--T- --TT---TT- --T-BBB--- ---------A AAAAAAAAAA A--------- -----
5 ---------- ---------- ---------- ----BBBBBB BB--BBBBB- ---AAAAAAA A--------- -BBB-
6 ---------- --------- ---------- ---------A AAAA---AAA AAAAAAAAAA A--------- -----
```

[a] Secondary structure predictions were carried out using four algorithms: the Bayes statistics (1; Maxfield & Scheraga, 1976), information theory (2; Gibrat et al., 1987), and neural net (3; Qian & Sejnowski, 1987) algorithms supplied with the program package SYBYL 5.4 (Tripos Associates, Inc.) and the Chou–Fasman (4; Chou & Fasman, 1974) algorithms of the FASTA program package (Pearson & Lipman, 1988). In addition, results from the PHD neural network structure prediction method (5; Rost & Sander, 1993) are shown (5). The strategy of Wishart et al. (1992) and the present set of data were used to arrive at the NMR secondary structure estimate for the protein at pH 3.0 (6). Symbols: A, α-helix; B, β-sheet; T, turn; –, random coil or no clear decision. Basic and core domains are underlined.

and in similar alignments [e.g., Dorn et al. (1990)]. From this multiple sequence alignment it is clear that several regions are well conserved among the immunodeficiency virus Tat proteins, namely, the Cys-rich region between amino acids 50 and 60 in the numbering scheme of Table II, the so-called core region, which is highly homologous in all Tat proteins, from amino acid 61 to 76 (according to the numbering scheme in Table II), and the C-terminal basic domain from amino acid 80 to 92. The basic region is supposedly the RNA-binding region.

We performed computational secondary structure predictions using standard algorithms and protocols, namely, Bayes statistics (Maxfield & Scheraga, 1976), information theory (Gibrat et al., 1987), neural net (Qian & Sejnowski, 1987), Chou–Fasman algorithm (Chou & Fasman, 1974), and PHD neural network (Rost & Sander, 1993). It can be seen easily from the summary of the results (Table III) that the only major local secondary structure in the EIAV Tat protein agreed upon by all the methods is an α-helix formed by the residues comprising the basic region.

Subsequently, we used the chemical shift data available from our experiments to perform a secondary structure estimation according to the chemical shift index strategy by Wishart et al. (1992). The procedure depends on a simple correlation between chemical shifts of C_α proton resonances of consecutive amino acids and local secondary structure.

C_α proton resonances shifted upfield relative to the corresponding random coil C_α proton resonances indicate local α-helical structure; C_α proton resonances shifted downfield compared to the corresponding resonances in a random coil structure indicate local β-sheet structure. According to Wishart et al. (1992), the criterion for taking a particular shift into account in secondary structure estimates is its deviation from the random coil resonance shift by more than

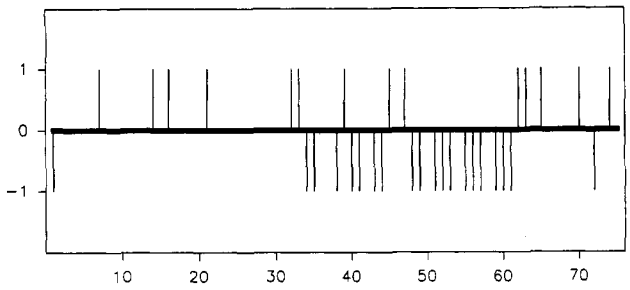

FIGURE 7: Chemical shifts of the C_α proton resonances derived from the pH 3.0 protein spectra relative to their random coil positions for use in the chemical shift index procedure. A value of −1 is assigned to the resonances for which the difference between the observed chemical shift and the random coil chemical shift is less than −0.1 ppm; a value of 1 is assigned if this difference is larger than 0.1 ppm.

0.1 ppm. Also, to get a more reliable picture, it was suggested that only resonances with the same sense chemical shift deviation for a stretch of more than three sequential residues should be taken into account.. It was shown recently that this procedure resulted in surprisingly accurate results [e.g., Archer et al., (1993)].

In the case of the EIAV Tat protein in aqueous solution at pH 3.0, the above criteria for secondary structure assignments with the chemical shift index procedure were met for only 19 out of 75 amino acid residues (Figure 7). In accord with all the secondary structure predictions, a local secondary structure element which can be deduced from the chemical shift data is a helix in the basic domain and the short stretch of amino acids between the core domain and the basic domain (D48–R61). In addition, formation of a short helical region is indicated within the core domain (C40–S44). The structure of the protein could not be further defined since no long-range NOEs could be observed.

Thus, in total, it seems to be safe to conclude that the EIAV Tat protein contains no elements of regular secondary structure in aqueous solution at pH 3.0 with the exception of the RNA-binding basic region, part of the core domain, and the linker region between the basic domain and the core domain, which all form helical structures. It should be noted that the chemical shift values for the C_α protons even in these regions are not very much above the values judged to be indicative of helical secondary structure. Therefore, it may well be that, on average, a large percentage of the EIAV Tat protein is unstructured even in the basic and core domains under our present experimental conditions.

Calnan et al. (1991) studied synthetic peptides with the sequence of the basic domain of the HIV Tat protein and mutant peptides. They observe in their experiments that the basic domain is predominantly unstructured and obtains a structure only after binding to the recognition RNA sequence. Thus, it may well be that in the whole protein the short peptide studied by Calnan et al. (1991) is pushed toward forming a helical structure.

Outlook. Presently we are finishing the determination of the structure of the EIAV Tat protein in phosphate buffer, pH 6.3 (Willbold, 1993) and in TFE (40% v/v) solution (Sticht, 1993), as the protein seems to have a more stable tertiary structure in either solvent. We also began studies of the structural transitions involved in going from pH 6.3 to pH 3.0 solution. Work on the solution structure of a synthetic HIV Tat protein is also in progress.

ACKNOWLEDGMENT

Prof. A. Ejchart helped with the NMR experiments, and Prof. F. X. Schmid and Mr. L. Mayr helped with the CD spectra. The computer program NDee was written by Mr. F. Herrmann and Dr. J. Weber.

REFERENCES

Archer, S. J., Bax, A., Roberts, A. B., Sporn, M. B., Ogawa, Y., Piez, K. A., Weatherbee, J., Tsang, M., Lucas, R., Zheng, B., Wenker, J., & Torchia, D. A. (1993) *Biochemistry 32*, 1164–1171.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Calnan, B. J., Biancalana, S., Hudson, D., & Frankel, A. D. (1991) *Genes Dev. 5*, 201–210.

Chou, P. Y., & Fasman, G. (1974) *Biochemistry 13*, 211–222.

Dang, C. V., & Lee, W. M. F. (1989) *J. Biol. Chem. 264*, 18019–18023.

Davies, D. G., & Bax, A. (1985) *J. Magn. Reson. 65*, 355–360.

Derse, D., Carvalho, M., Carroll, R., & Peterlin, M. (1991) *J. Virol. 65*, 7012–7015.

Dorn, P., DaSilva, L., Martarano, L., & Derse, D. (1990) *J. Virol. 64*, 1616–1624.

Endo, S., Kubata, S., Siomi, H., Adachi, A., Oroszlan, S., Maki, M., & Hatanaka, M. (1989) *Virus Genes 3*, 99–110.

Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953–5958.

Franchini, G., Gallo, R. C., Guo, H.-G., Gurgo, C., Callati, E., Fargnoli, K., Hall, L., Wong-Staal, F., & Reitz, M. S., Jr. (1987) *Nature 328*, 539–543.

Frank, R., & Gausepohl, H. (1988) in *Modern Methods in Protein Chemistry* (Tschesche, H., Ed.) Vol. 3, pp 41–60, de Gruyter, Berlin.

Frankel, A. D., Bredt, D. S., & Pabo, C. O. (1988) *Science 240*, 70–73.

Gentz, R., Chen, C.-H., & Rosen, C. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 86*, 821–824.

Gibrat, J. F., Garnier, J., & Robson, B. (1987) *J. Mol. Biol. 198*, 425.

Griesinger, C., Otting, G., Wnthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Hauber, J., & McCullen, B. R. (1989) *J. Virol. 63*, 1181–1187.

Higgins, D. G., Bleasby, A. J., & Fuchs, R. (1992) *Comput. Appl. Biosci. 8*, 189–191.

Kiefhaber, T., Rudolph, R., Kohler, H.-H., & Buchner, J. (1991) *Biotechnology 9*, 825–829.

Kuehnel, H., von Briesen, H., Dietrich, U., Adamski, M., Mix, D., Biesert, L., Kreutz, R., Immelmann, A., Henco, K., Meichsner, C., Andreesen, R., Gelderblom, H., & Ruebsamen-Waigmann, H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2383–2387.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Maxfield, F. R., & Scheraga, H. A. (1976) *Biochemistry 15*, 5138.

Myers, E., & Miller, W. (1988), *Comput. Appl. Biosci. 4*, 11–17.

Noiman, S., Gazit, A., Tori, O., Sherman, L., Miki, T., Tronick, S. R., & Yaniv, A. (1990) *Virology 176*, 280–288.

Noiman, S., Yaniv, A., Tsach, T., Miki, T., Tronick, S. R., & Gazit, A. (1991) *Virology 184*, 521–530.

Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444–2448.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76–80.

Qian, N., & Sejnowski, T. (1988) *J. Mol. Biol. 202*, 865.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B. R., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn,

E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., & Wong-Staal, F. (1985) *Nature 313*, 277–284.

Rost, B., & Sander, C. (1993) *J. Mol. Biol.* (in press).

Ruben, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W. A., & Rosen, C. A. (1989) *J. Virol. 63*, 1–8.

Sherman, L., Gazit, A., Yaniv, A., Kawakami, T., Dahlberg, J. E., & Tronick, S. R. (1988) *J. Virol. 62*, 120–126.

Siomi, H., Shida, H., Maki, M., & Hatanaka, M. (1990) *J. Virol. 64*, 1803–1807.

Sticht, H. (1993) M.S. Thesis, Bayreuth University.

Tao, J., & Frankel, A. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2723.

Willbold, D. (1993) Ph.D. Thesis, Bayreuth University.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647–1651.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley Interscience, New York.