Raschdorf, F., Dahinden, R., Maerki, W., Richter, W. J., & Merryweather, J. P. (1988) *Biomed. Environ. Mass Spectrom. 16*, 3–8.
Ryle, A. P., Sanger, F., Smith, L. F., & Kitai, R. (1955) *Biochem. J. 60*, 541–556.
Sanborn, B. M., Kuo, H. S., Weisbrodt, N. W., & Sherwood, O. D. (1980) *Endocrinology 106*, 1210–1215.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368–379.
Schwabe, C., & McDonald, J. K. (1977) *Science 197*, 914–915.
Sherwood, C. D., & O'Byrne, E. M. (1974) *Arch. Biochem. Biophys. 160*, 185–196.
Shih, A., Goldsmith, L. T., Weiss, G., Bourell, J., & Winslow, J. (1989) *Endocrine Soc.*, 1543.
Shire, S. J. (1983) *Biochemistry 22*, 2664–2671.
Sieber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Marki, F., & deGasparo, M. (1978) *Hoppe-Seyler's Z. Physiol. Chem. 359*, 113–123.
Smith, M. C., Cook, J. A., Furman, T. C., & Occolowitz, J. L. (1989) *J. Biol. Chem. 264*, 9314–9321.
Tregear, G. W., Du, Y.-C., Wang, K.-Z., Southwell, C., Jones, P., John, M., Gorman, J., Kemp, B., & Niall, H. D. (1983) in *Biology of Relaxin and Its Role in the Human* (Bigazzi, M., Greenwood, F. C., & Gasparri, F., Eds.) pp 42–55, Excerpta Medica, Amsterdam.
Winslow, J., Shih, A., Laramee, G., Bourell, J., Stults, J., & Johnston, P. (1989) *Endocrine Soc.*, 889.
Zhang, R., & Snyder, G. H. (1989) *J. Biol. Chem. 264*, 18472–18479.

Activating Region of HIV-1 Tat Protein: Vacuum UV Circular Dichroism and Energy Minimization†

Erwann P. Loret,*,‡ Eric Vives,§ Pui Shing Ho,‡ Hervé Rochat,§ Jurphaas Van Rietschoten,§ and W. Curtis Johnson, Jr.‡

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331-6503, and Laboratoire de Biochimie, CNRS URA 1179, Faculté de Médecine, Secteur Nord, Boulevard Pierre Dramard, 13326 Marseille Cedex 15, France

Received March 21, 1991

ABSTRACT: Tat protein is a trans-acting transcriptional activator of the human immunodeficiency virus type 1 and is essential for viral transcription. By homology with other transcriptional activators, Tat is expected to possess a nucleic acid binding region and a separate adjacent activating region. In order to localize the activating region of Tat, we have synthesized the sequences 2–23 and 38–60 of the protein. These two peptides contain the two candidates for the activating regions proposed from mutation experiments in previous studies: sequence 1–13 and sequence 38–45. The argument advanced to justify the location of the activating region within the sequence 1–13 was the periodicity of acidic, polar, and hydrophobic residues consistent with that of an amphipathic α-helix, similar to the activating region of many eukaryotic transcriptional activators. We have monitored by vacuum UV circular dichroism the ability of each peptide to adopt an α-helical conformation under conditions that strongly favor the formation of secondary structures. Only peptide 38–60 adopts an α-helical conformation in these conditions, in keeping with Chou–Fasman prediction. Energy minimization and molecular dynamics were carried out for several possible conformations of sequences 1–14 and 38–60. Our results indicate that only the sequence 38–45 is able to form an α-helix with amphipathic characteristics.

The protein Tat is a trans activator of the human immunodeficiency virus type 1 (HIV-1)[1] in vivo (Arya et al., 1985; Sodroski et al., 1985). The major action of Tat is to increase the steady-state level of transcription from the HIV-1 long terminal repeat (LTR), although the mechanism of this trans activation remains unclear [see Rosen and Pavlakis (1990) and Cullen and Greene (1990) for reviews]. The Tat gene is composed of two exons predicting a protein of 86 amino acid residues (Figure 1); however, the first 72 N-terminal residues, corresponding to the first exon, are sufficient for the full activity (Cullen, 1986). Moreover, the residues 58–72 can also be deleted without impairing the activation function of Tat (Siegels et al., 1986; Wright et al., 1986).

Mutational analyses of Tat (Garcia et al., 1988, 1989; Sadaie et al., 1989; Hauber et al., 1989; Kuppuswamy et al., 1989; Green et al., 1989; Ruben et al., 1989; Rappaport et al., 1989) made it possible to propose a functional organization of this protein (Figure 1). Two regions have been functionally characterized. Region II, a cysteine-rich cluster containing

† This work was supported by grants from the Association pour la Recherche sur le Cancer, the American Foundation for AIDS Research, the ANRS, the American Cancer Society (NP-740 and JFRA306), the CNRS, and the National Institutes of Health (Grant GM 21479).
* To whom correspondence should be addressed.
‡ Oregon State University.
§ Laboratoire de Biochimie, CNRS URA 1179.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; Tat, trans-acting transcriptional activator of HIV-1; LTR, long terminal repeat sequence located at the two extremities of the HIV-1 DNA; C18 HPLC, reverse-phase high-performance liquid chromatography on C18 columns; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; CD, circular dichroism.

0006-2960/91/0430-6013$02.50/0

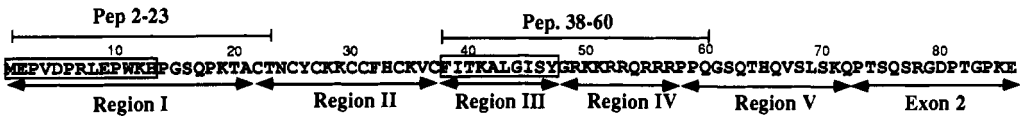

FIGURE 1: Amino acid sequence of the HIV-1 Tat protein with the two candidate activating regions enclosed in boxes. The different functional regions of Tat are indicated as previously proposed by Kuppuswamy et al. (1989). The first candidate region is located in the peptide 2–23 and corresponds to the sequence 1–13, which has been considered as a potential amphipathic α-helix (Rappaport et al., 1989). The second candidate region is located in the peptide 38–60 and corresponds to the sequence 38–47 (Green et al., 1989).

seven cysteines, is located between residues 21 and 38. The mutation of six of the seven cysteines (cysteine 31 is the exception) completely abolishes Tat function. This region possesses a sequence motif similar to that of several other proteins that play a role in DNA binding [see Mitchell and Tjian (1989) for a review]. The Tat cysteine-rich cluster was shown to be essential for binding metal ligands, causing Tat to form metal-linked dimers (Frankel et al., 1988). However, recent experiments show that metal binding is not required for Tat function at the transcription level (Jeyapaul et al., 1990). A second important region is the highly basic sequence GRKKR at position 48–52 within region IV. Tat is localized mainly in the nucleus, and mutations within this sequence yield a cytoplasmic Tat protein that is nonfunctional (Hauber et al., 1989).

Many eukaryotic transcriptional activators studied to date act by using a scheme that implies the presence of two surfaces: a DNA-binding region positions the protein on the DNA so that the other activating region can interact with the target protein to promote initiation of the transcription (Ptashne, 1988). It is not clear whether Tat acts through a DNA or RNA target nucleotide sequence, although there is evidence that Tat recognizes RNAs and functions posttranscriptionally (Rosen et al., 1985, 1986; Ratner et al., 1985; Muesing et al., 1987; Dingwall et al., 1989; Southgate et al., 1990; Braddock et al., 1990; Weeks, 1990). However, there are also data indicating that Tat functions transcriptionally (Rice & Matthews, 1988; Jeang et al., 1988, Garcia et al., 1989), either at initiation (Gentz et al., 1989), at elongation (Kao et al., 1987), or at both initiation and elongation (Laspia et al., 1989). The fact remains that the binding of Tat to a target nucleotide sequence is probably necessary for trans activation, and this implies the presence of a binding region and an activating region.

Attempts have been made to localize the activating region in Tat by using mutation experiments. Green and Lowenstein (1988) proposed sequence 37–48 and 49–57 (region III and IV) as the activating region and binding region, respectively, and reported that a peptide corresponding to the sequence 37–62 was able to trans activate in their assay. However, the same peptide synthesized by another group was completely inactive in their assay (Frankel et al., 1989). Region III contains a highly conserved motif Lys-X-Leu-Gly-Ile-X-Tyr, and mutation of lysine 41 severely reduces Tat function (Kuppuswamy et al., 1989). Rappaport et al. (1989) proposed the 13 amino acid sequence adjacent to the N-terminus (Figure 1) as being the activating region from mutation experiments of glutamic acid 2, aspartic acid 5, and glutamic acid 9. The argument advanced to justify the location of the activating region within the N-terminal part was the periodicity of acidic, polar, and hydrophobic residues consistent with that of an amphipathic α-helix, similar to the activating region of many eukaryotic transcriptional activators (Ptashne, 1988). However, in a recent study Tiley et al. (1990) show that this 13 amino acid sequence adjacent to the N-terminus contributes in vivo but is not absolutely essential for Tat function, and the integrity of the sequence centered on lysine 41 is essential. They conclude that neither region fulfills the criteria for a discrete activation domain found with other transcription factors.

These conflicting views inspired the present research. We have attempted to induce an α-helix in sequences 2–23 and 38–60 of Tat, which contain the two candidates for the activating regions. We show that (1) under conditions that strongly favor the α-helix conformation, only peptide 38–60 is able to adopt an α-helix; (2) in this sequence residues 38–45 are predicted to form an α-helix with amphipathic characteristics; (3) the presence of an α-helix in the 13 amino acid sequence adjacent to the N-terminus, as previously proposed, is not likely.

Experimental Procedures

Peptide Synthesis. Peptides were assembled according to the method of Barany and Merrifield (1980) on 4-(oxymethyl)phenylacetamidomethyl (PAM) resin (0.5 mmol) (Applied Biosystems Inc., Forster City, CA) on a semiautomated synthesizer (NPS4000, Neosystem, Strasbourg, France). Side-chain protection of the butyloxycarbonyl amino acids (Boc amino acid) (Neosystem) was as follows: Asp and Glu, cyclohexyl; Ser and Thr, benzyl; Lys, 2-chlorobenzyloxycarbonyl; Tyr, bromobenzyloxycarbonyl; Arg, tosyl; His, benzyloxycarbonyl; Trp, formyl; Cys, acetamidomethyl.

The synthesis cycle used for each Boc amino acid incorporation, according to the method of Coste et al. (1990), was (1) dichloromethane (DCM) wash, 1 min; (2) 65% trifluoroacetic acid (TFA) in DCM wash, 1 min; (3) 65% TFA-DCM wash, 15 min; (4) isopropanol wash, 1 min; (5) two DCM wash, 1 min; (6) *N*-methylpyrrolidone (NMP) wash: 1 min; (7) Boc-amino acid (4 equiv) and benzotriazole-1-yloxy-tris(pyrrolidino)phosphonium hexafluorophosphate (PyBOP) (Novabiochem, Laufelfingen, Switzerland) (4 equiv) in NMP for the amino acid addition; (8) diisopropylethylamine (DIEA) (8 equiv) in situ for the coupling step (mean reaction time 20 min); (9) two NMP washes, 1 min; and (10) DCM washes, 1 min. Each coupling step was monitored by using the ninhydrin test.

The peptides were deprotected and removed from the resin by a high hydrogen fluoride (HF) procedure using 10% by volume of *p*-cresol as a scavenger. However, acetamidomethylcysteine remains stable in HF, and thus cysteine remains protected. The tryptophan residue of crude peptide 2–23 was deformylated by treatment with 0.04 N sodium hydroxide, pH 11.5, for 3 min and then neutralized with glacial acetic acid, according to the method of Li et al. (1978).

The crude peptides were purified by medium-pressure liquid chromatography (MPLC) on an apparatus composed of the following elements: a model 590 solvent delivery system from Waters (Milford, MA), a labomat VS200 solvent programmer from Labomatic (Sinsheim, Germany), a UV spectrophotometer (Model 2138), and a recorder (Model 2210) from LKB (Bromma, Sweden). The preparative MPLC column was a Eurosil Bioselect 100-20-C18 (13 × 304 mm) from Knauer (Berlin, Germany). The solvent system was buffer A, 0.1% TFA in water; buffer B, 0.08% TFA in acetonitrile. Following

the MPLC run, the peptide fractions were monitored by analytical C18 HPLC, and homogeneous fractions were pooled and lyophilized.

Amino Acid Analyses. Amino acid analyses were performed on a model 6300 Beckman analyzer. Purified peptides (1–4 nmol) were hydrolyzed in 6 M HCl for 20 or 72 h at 110 °C. These analyses were used to determine peptide concentration and subsequently the extinction coefficient at 190 nm on a per amide basis, 8500 for peptide 2–23 in buffer or 90% TFE, 13 400 for peptide 38–60 in buffer, and 12 500 for peptide 38–60 in 90% TFE.

Circular Dichroism Spectra. Circular dichroism (CD) spectra were measured with a vacuum ultraviolet (VUV) CD spectrophotometer described elsewhere (Johnson, 1971). The instrument was calibrated with (+)-10-camphorsulfonic acid, with the assumption that $\Delta\epsilon(290.5) = +2.36$ and $\Delta\epsilon(192.5) = -4.9$. The cell path length was 50 or 100 μm, and to avoid absorption artifacts the total optical density of the cell, solvent, and sample never exceeded 1.0 over the range of the CD spectrum. All spectra were measured at 20 °C. Data were collected at 0.5-nm intervals with a scan rate of 1 nm per minute, and the spectral slit was 2 nm. $\Delta\epsilon$ is on a per amide basis.

Analysis of VUV CD Spectra. The VUV CD data were analyzed for secondary structure according to the method of Manavalan and Johnson (1983). Variable selection removes unimportant variables from an underdetermined system of equations. For instance, there are five independent equations in the CD spectrum of a protein measured into the VUV region to 178 nm. However, the CD spectrum of a protein depends upon many parameters, e.g., α-helix, parallel and antiparallel β-sheet, the various types of β-turns, aromatic side chains, prosthetic groups, nonrepetitive structures, etc. Because the CD spectrum of a protein is analyzed by using the CD spectra of a set of reference proteins with known secondary structures, it is reasonable to remove reference proteins that have contributions to their CD spectra from parameters not found in the protein to be analyzed with the set. The goal of removing reference proteins with extraneous contributions to the CD is to reduce the number of variables to five, which is consistent with the information content of our data. It is not known a priori which reference proteins have extraneous contributions in their CD that invalidate the analysis. Therefore, calculations were performed for all possible subsets. The fractions of secondary structures are not constrained to be positive, nor is the sum of fractions of secondary structures constrained to 1.0. The following criteria were used to select solutions from among the subsets: (1) the sum of fractions for predicted secondary structures must be in the range 0.95–1.05; (2) no fraction should have a value less than −0.05; (3) the average sum of squares of the residuals between the reconstructed CD and observed CD should be less than 0.25; (4) the fraction of α-helix should not change significantly on removing reference proteins; (5) of the subsets meeting these criteria, the analyses with the largest number of reference proteins will be used. For our work, each subset contained at least 19 of the 22 reference proteins.

Computer Modeling. Structures of sequence 1–13 and 38–60 of the Tat protein were generated by computer modeling to study the possible conformations that could be adopted by the two peptide fragments. Initial models were built around the conformations of analogous sequences located in structurally defined regions of proteins in the Brookhaven crystallographic data base. These were then refined with the AMBER molecular mechanics force field (Weiner et al., 1984, 1986) under conditions that mimic a nonaqueous environment to minimize the interaction energies of side-chain and backbone atoms. To determine which of the structural elements of these models were most stable and would therefore contribute significantly to the structure of each peptide, the lowest energy models were subjected to molecular dynamics calculations. The models were allowed to relax for up to 12 ps at room temperature (300 K) and at elevated temperatures (up to 600 K).

All calculations were performed on a Silicon Graphics personal Iris workstation using the DISCOVER molecular mechanics package from BioSym Technologies. The final derivatives for the final energy-minimized structure were less than 0.1 δkcal/mol per interaction.

Results and Discussion

Peptide Synthesis. Two peptides of similar size corresponding respectively to the Tat sequences 2–23 and 38–60 (Figure 1) were chemically synthesized following the method of Barany and Merrifield (1980). The protocol used for assembly of the peptides was inspired by Coste et al. (1990) using the newly discovered coupling reagent PyBOP. The total time for one synthesis cycle was about 45 min, including the ninhydrin test, with a coupling time of about 20 min. This cycle lacks the neutralization step, thus avoiding some of the side reactions that take place in the very basic conditions. It was very efficient, and only three double-coupling reactions had to be performed to obtain complete incorporation of Boc amino acids in a total of 45 couplings. The peptides could be purified to better than 96% homogeneity as monitored by analytical C18 HPLC with a gradient of acetonitrile. Amino acid analyses were in good agreement with the expected ratios.

VUV CD Spectra. Each peptide possesses one of the two sequences proposed as an activating region: the 13 amino acid sequence adjacent to the N-terminus for the peptide 2–23 (Rappaport et al., 1989) and the sequence 38–48 for the peptide 38–60 (Green & Lowenstein, 1988). Mutation and sequence comparisons were used to identify the activating region in those previous works. No structural arguments were provided, except for a Chou–Fasman prediction of secondary structure, to justify the suggestion of an α-helix in the N-terminal part of the Tat protein (Arya et al., 1985). Moreover, the possibility that the N-terminal region of the Tat protein could adopt an acidic amphipathic α-helix, similar to a known consensus transcriptional activating region, prompted the proposal that this part of Tat was a discrete activating region (Rappaport et al., 1989). Thus we have measured by vacuum UV CD the ability of this candidate activating region to adopt an α-helix under conditions that favor the structure, and we have compared these results with those obtained for another peptide of the same size from the Tat sequence, which contains the second candidate activating region.

The two peptides were studied in water at pH 7 (phosphate buffer, 20 mM) and in 90% TFE. TFE is a solvent that mimics the interior of a protein and induces secondary structures in peptides in solution (Urry, 1972). The presence of an α-helix, in solution, is easily identified by CD spectra measured between 178 and 260 nm, due to a well-characterized positive band at 190 nm, a negative band at 205–207 nm related to a π–π^* transition, and a negative band at 215 nm related to a n–π^* transition [see Johnson (1985) and Woody (1985) for reviews].

Of the peptides studied, only the CD spectrum of peptide 38–60 in 90% TFE shows the three CD bands typical of an α-helix (Figure 2). In aqueous solvent, the CD spectrum of this peptide shows only a negative band at 197 nm associated

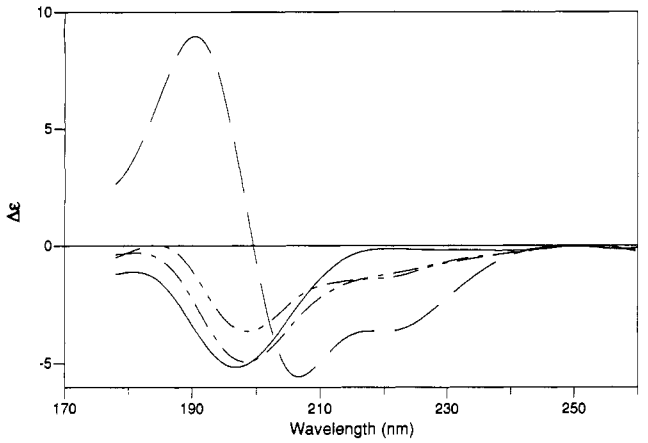

FIGURE 2: Vacuum UV CD spectra of potential Tat activating regions. Peptides 2–23 (–·–) and 38–60 (—) in 20 mM phosphate buffer, pH 7. Peptides 2–23 (– · · –) and 38–60 (– – –) in 90% trifluoroethanol and 10% water. $\Delta\epsilon$ is on a per amide basis expressed as M^{-1} cm^{-1}.

Table I: Secondary Structure Analyses[a]

peptide	solvent	H	A	P	T	O	total
38–60	aqueous[b]	0.05	0.28	(–0.01)	0.27	0.38	0.97
38–60	90% TFE[c]	0.44	0.16	0.01	0.18	0.19	0.97
2–23	aqueous[b]	0.13	0.21	(–0.02)	0.30	0.39	1.03
2–23	90% TFE[c]	0.13	0.21	0.00	0.25	0.42	1.00

[a] Abbreviations: H, α-helix; A, antiparallel β-sheet; P, parallel β-sheet; T, β-turn; O, other structures. [b] Aqueous solvent with 20 mM phosphate buffer, pH 7. [c] Solvent composed of 90% trifluoroethanol and 10% water.

with the random-coil structure. In contrast, the CD spectrum of peptide 2–23 is almost unchanged in solvents ranging from water to 90% TFE. The two spectra show the negative band at 197 nm typical of random coil and a low negative band at 215 nm, which could be due to the prolines, as in random-coil collagen (Jenness et al., 1976), or to the aromatic ring of the tryptophan (Woody, 1985).

The CD spectrum of the peptide 2–23 was also studied with different hydrophobic solvents, in aqueous solution at different pHs and at different temperatures (spectra not shown). A mixture of 70% ethanol and 30% water as well as a mixture of 30% ethanol, 30% cyclohexane, and 40% octanol give a random-like spectrum similar to the one observed in 90% TFE of Figure 2. At 20 °C, the spectrum at pH 3 and 9 is a random-like spectrum similar to the one observed at pH 7 in Figure 2 (called A in Table II). Interestingly, temperature has a different effect at different pHs. From 2 to 65 °C, the CD spectrum remains unchanged at pH 3 and 9 (A-type), while above 45 °C a spectrum similar to the one in 90% TFE appears at pH 7 (B-type). The change occurs between 37 and 45 °C. These experiments indicate that in aqueous solution the uncharged amino acids block the temperature effect. The spectrum in 70% ethanol, pH 9 and 45 °C is similar to the one in 90% TFE (B-type). The CD spectrum of peptide 2–23 was also measured after different incubation times (2, 6, 24, and 48 h) because the slow trans–cis isomerization of the many prolines might influence the secondary structure. However, the spectra obtained were independent of the time.

The VUV CD data were analyzed for the percentage of secondary structure (Table I). A high percentage of α-helix (44%) was observed for peptide 38–60 in 90% TFE, while there was only 5% in aqueous solution. The analyses for peptides 2–23 were similar, the percentage of α-helix remaining unchanged (13%). CD is a function of the amide chromophores located between amino acid residues, and the percentage of α-helix calculated from CD data makes it possible to calculate

Table II: Type of Random-Like CD Observed at Different Temperatures and pHs for the Peptide 2–23[a]

pH	2 °C	37 °C	45 °C	65 °C
3	A	A	A	A
7	A	A	B	B
9	A	A	A	A

[a] A-type CD is similar to that observed in 20 mM phosphate buffer at pH 7. B-type CD is similar to that observed in 90% TFE (see Figure 2).

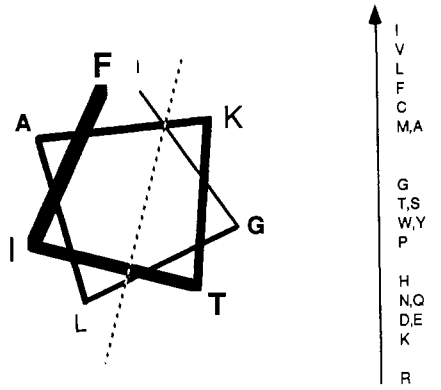

FIGURE 3: Helical wheel representation of the α-helix for the sequence 38–45. Hydrophobic residues are located to the left of the dotted line and the hydrophilic and neutral residues to the right of the dotted line, showing an amphipathic characteristic for this α-helix. Hydrophobicity of the amino acid residues using the scale from Kyte and Doolittle (1982) is showed on the right.

the number of peptide bonds in an α-helix conformation, assuming a single structure in solution. For peptide 2–23 there are 21 amide chromophores, so 13% α-helix corresponds roughly to three amides (four residues), which is just enough to form just one turn of an α-helix. Analysis of the CD data shows that peptide 38–60 in 90% TFE has 9–10 amides (10–11 residues) in an α-helix.

Sequence Prediction of Secondary Structure. The Tat sequence was analyzed for prediction of secondary structure according to the method of Chou and Fasman (1978). The results show that the highest probability of α-helix is for the sequence 37–43, which corresponds to a part of the activating region proposed by Green and Lowenstein (1988). There is a lower but significant probability of α-helix for two other sequences (6–9 and 48–55). However, sequence 6–9 is too short for the prediction of an α-helix by this method, contrary to what was indicated by Rappaport et al. (1989). Indeed, there are three prolines in positions 3, 6, 10, and, according to the prediction method, proline cannot occur in the inner helix or at the C-terminal end but can occupy the first turn in the N-terminal helix (Previlege & Fasman, 1989).

Helical wheel representation (Schiffer & Edmundson, 1967) of the α-helix for sequence 38–45 (Figure 3) shows a gathering of hydrophobic residues on a side of the helix, while hydrophilic and neutral residues are located on the other side. This feature provides an amphipathic characteristic to this α-helix. Inspection of activating sequences suggests that many of them could, in principle, form amphipathic α-helices. The presence of negatively charged residues (Asp and Glu) on the hydrophilic surface of the α-helix was retained originally as a common feature for activating regions (Ptashne, 1988). However, two other types of activating regions have been described (Mitchell & Tjian, 1989): an α-helix rich in glutamine and very few charged amino acid residues, and a proline-rich motif. This enlarged the criterion to describe a discrete activating region and perhaps means that other types of activating regions exist.

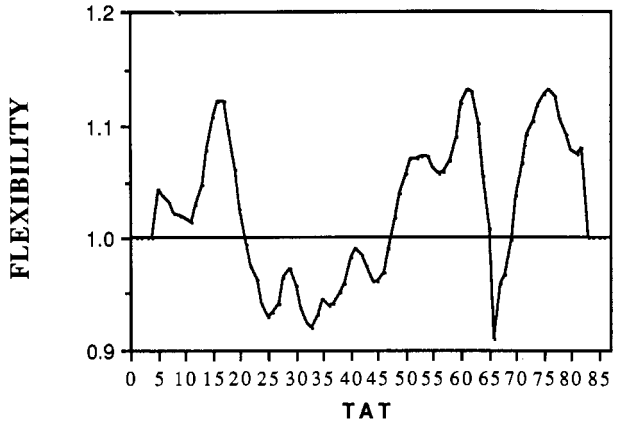

FIGURE 4: Flexibility and hydrophobicity of Tat. The flexibility plot was calculated according to the method of Karplus and Schulz (1985) by using a sliding average of seven amino acid residues. Positions having a value less than 1 are considered as rather flexible.

Tat Flexibility. The flexibility of Tat was calculated according to the method of Karplus and Schulz (1985). The results (Figure 4) show that the sequence 23–48 in region II and III (Figure 1) is the most flexible part of Tat while region I is rather rigid, which is confirmed by the similarity of the CD spectrum in the aqueous solution and 90% TFE. The flexibility of region III agrees with the proposed function of the activating region. Indeed, it has been observed in the case of transcriptional activator that the binding region exerts control on the activating region in allowing activation only if the activator is complexed with its target nucleotidic sequence (Ptashne, 1988). This suggests that the activating region should adopt its fully active conformation only after the binding region is complexed, implying that this region should be flexible.

Computer Modeling of Peptide Fragments. We used two approaches to construct structural models for the Tat peptide fragments, in those studies. First, the Brookhaven data base of protein structures was searched for sequences that were homologous to the Tat peptide fragments. This allowed us to determine the range of conformations that are adopted by the peptide sequences within the context of crystallographically defined protein structures. Sequences that were highly homologous to the 1–14 and 38–60 fragments of the Tat protein were found in several structures in the protein data base. From this, it was clear that each sequence of these fragments falls into limited set of conformations. Sequences that are similar to the 1–14 fragment did not generally adopt α-helical conformations in the proteins of the data base. There was some evidence from this homology search, however, to suggest that the first 9–10 residues of the 38–60 fragment would form an α-helix, while the remainder of the sequence would be extended or dynamic.

Several conformations were constructed for each fragment by using the structures for the homologous sequences from the crystallographic data base. Each structure was subjected to 1 ps of molecular dynamics to relieve the inherent stress and was then energy minimized with the AMBER force field (Weiner et al., 1984, 1986). Analysis of the final energies of each minimized conformation allowed us to compare the relative intrinsic stabilities of the structures for each fragment.

We then asked the question of which, if any, residues within each fragment could adopt an α-helical conformation. There have been a number of arguments suggesting an α-helical structure for both fragments. Models were therefore constructed for each fragment that was entirely α-helical. These models were first energy minimized, then subjected to molecular dynamics at elevated temperatures to determine which amino acids in the chains were most stable in the helical conformation. For both fragments, a 12-ps dynamics run was found to be sufficient to distinguish between domains that would be stable as a helix from those that were not as stable. For the 38–60 fragment, the external energy at 450 K was sufficient to extend the cationic C-terminal tail out of the helical conformation. For the 1–14 sequence, the dynamics was run at a slightly higher temperature (600 K) to provide the prolines with enough energy to undergo a trans-to-cis conversion and vice versa. Although *cis*-prolines are rare, they have been found in the left-handed helix of polyproline I and may help to stabilize a helical structure for this fragment. The detailed results from these studies for each sequence are discussed below.

Computer Modeling of Sequence 1–14. A search through the Brookhaven protein data base did not yield a match identical with the Tat protein sequence MEPVDPRLEPWKHP at position 1–14. Two sequences were found in the data base that conformed to the general formula XPXXPXXXPXXXXP, where X is any amino acid residue. One was in the M chain of the photosynthetic reaction center (Deisenhofer et al., 1989), while the second was found in the structure of human rhinovirus coat protein (Arnold & Rosmann, 1988). The positions of the first three prolines are identical with that of the Tat sequence, while one additional amino acid separated the third and fourth prolines. Both of these sequences from the crystallographic data base are very hydrophobic and neither are in helical structures. This is consistent with the expectation that the high proline content would not favor formation of an α-helix.

Residues 302–315 in the M chain of the photosynthetic reaction center (Ala-**Pro**-Asp-Tyr-**Pro**-Ala-Tyr-Leu-**Pro**-Ala-Thr-**Pro**-Asp-**Pro**, where the residues in boldface represent exact matches) form a bent loop structure with the bend centered between residues positions Tyr 308 and Leu 309. The sequence Ile-**Pro**-Ile-Ala-**Pro**-Leu-Thr-Val-**Pro**-Thr-Gly-Ala-Thr-**Pro** is a compact loop structure at positions 223–236 of subunit 2 in the coat protein of human rhinovirus. Two bends centered at residues Ile 224 and Thr 232 and a hydrophobic core formed by the side chains of Leu 228, Val 230, Ala 234, and Pro 236 contribute to the compactness of this latter structure. The side chains of the hydrophilic threonine residues at 229, 232, and 234 extend out from the exterior surface of the loop.

To test the possibility that the N-terminal 14 amino acids may form an α-helix, we searched the data base for this secondary structure for sequences having multiple prolines spaced three to four residues apart. Only a single such sequence was found: taka-amylase contains two prolines separated by four amino acid residues within an α-helical conformation.

Structural models were constructed around each of the conformations located within the proteins of the crystallographic data base. In all cases, the side chains were initially generated in their extended conformations. Two compact loop models (A1-Rhino and A2-Rhino) were constructed by using the rhinovirus sequences as templates. In the first (A1-Rhino), the two bends from the rhinovirus loop were maintained at their respective amino acid positions relative to the prolines. This resulted in a structure in which the majority of the charged residues were placed in the interior core of the loop, and the hydrophobic residues protruded out from the exterior surface. A second structure (A2-Rhino) was built where each proline was shifted-1 toward the N-terminus relative to the

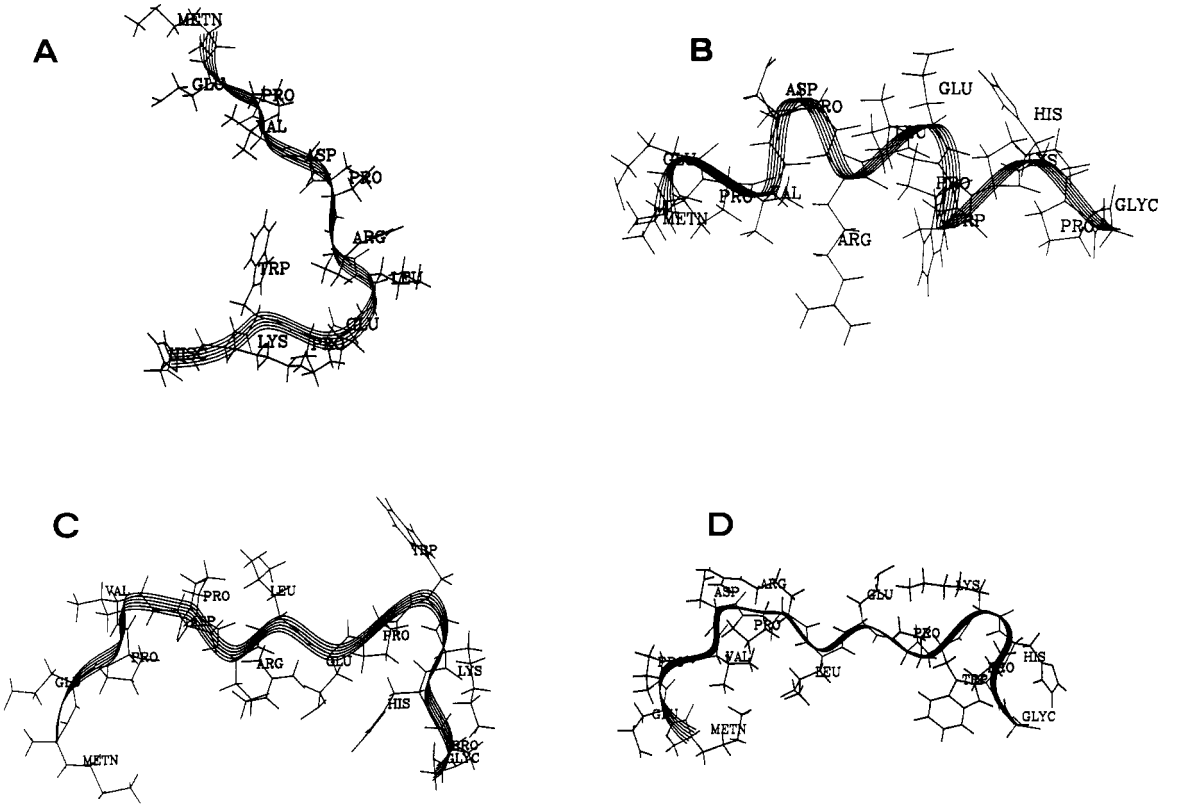

FIGURE 5: Initial models for amino acids 1–14 of the Tat protein. The backbone is traced as a ribbon to emphasize the secondary structure of the peptide. A is model A-PRC, B is A-H, C is A1-Rhino, and D is A2-Rhino (see the legend to Table III for the key to model designations).

bends. This placed a hydrophobic core at the interior, while all the charged side chains protruded out and away from the backbone.

A bent loop model (A-PRC) was constructed by using the sequence from the photosynthetic reaction center as a template. In the initial structure, the Arg 7 side chain was rotated around the C_α–C_β bond to remove unfavorable contacts with the backbone and ring atoms of Trp 11. Finally, a model (A-H) was built around the α-helix of the taka-amylase protein (Matsuura et al., 1984), with the helix extended towards the N- and C-termini with standard ϕ and ψ angles for a right-handed α-helix.

The energy-minimized structures for each model are shown in Figure 5. In general, the backbone conformations did not distort significantly from the initial structures. The models built around the rhinovirus conformation showed some compaction of the loop as a result of interactions within the interior of the loop structure that are analogous to the hydrophobic interactions of the rhinovirus structure.

The final energies for each model are summarized in Table III. The absolute numbers for the energies are not meaningful, except to compare the stabilities of the individual models. The bond energies for lengths, angles, and dihedral angles were all very similar for all the models. The primary differences in the overall energies of each structure resulted from variations in hydrogen bonding, coulombic repulsion, and van der Waals interactions. The conformation with the least favorable final free energy was the structure built from the photosynthetic reaction center (−142 kcal/mol final total energy). Because it was the most extended structure, this showed the lowest electrostatic repulsion of all the models (325 kcal/mol) but also had the lowest potential for hydrogen bond formation (−529 kcal/mol). In the final structure, the tryptophan side chain extended into the interior of the core, while the majority of the remaining residues protruded out and away from the exterior surface. In general, this structure was not significantly different from the photosynthetic reaction center template from which it was built.

Table III: Energies for Sequence 1–14 of the Tat Protein[a]

energies (kcal/mol)	models[b]					
	A1-Rhino	A2-Rhino	A-PRC	A-H	ARD	AHD
bonding	94	75	63	72	90	90
H-bonding	-823	-735	-529	-755	-935	-940
nonbonding	1	-5	-1	-25	-13	-7
coulombic	503	436	325	496	595	595
total	-225	-229	-142	-212	-261	-259

[a] Energies were minimized by using the AMBER force fields (Weiner & Kollman, 1986) to a maximum derivative of 0.1 kcal/mol. The energies of the four initial models (A1-Rhino, A2-Rhino, A-PRC, and A-H) as well as the lowest energy structures extracted from the molecular dynamics calculations on A2-Rhino and A-H models (ARD and AHD, respectively) are compared after energy minimization of their structures. [b] Models A1-Rhino and A2-Rhino were constructed by using the structure of the coat protein from human rhinovirus as a template. A1-Rhino has the bends of the loop positioned at residues V4 and W11 of the Tat protein, while A2-Rhino places the bends at residues D5 and K12. Model A-PRC was built by using the structure of the photosynthetic reaction center as a template. A-H was constructed as a helical model by using an α-helix in taka-amylase as a template, and it was extended by using standard ϕ and ψ angles for an α-helix. Models ARD and AHD are the lowest energy conformations after molecular dynamics calculations on the A2-Rhino (after 4 ps) and A-H (after 6 ps) models, respectively.

The two structures built from the human rhinovirus coat protein and the helical model constructed around the taka-amylase showed nearly identical energies. Model A1-Rhino started with the two bends located at the same positions as in the rhinovirus protein. The resulting energy-minimized structure had the hydrophobic residues extending out along the exterior surface. The interior of the loop is lined primarily with longer charged residues, giving slightly greater coulombic

and van der Waals repulsion in this structure. These repulsive forces, however, are more than compensated for by the −68 kcal/mol more favorable hydrogen-bonding energy, again associated with packing the polar groups into the compact interior of the loop.

The A2-Rhino model was built with the proline residues displaced by one position toward the N-terminus relative to the two bends in the loop structure. This conformation is essentially an inverted version of the A1-Rhino model, with all the charged side chains extending out along the exterior surface and the hydrophobic groups lining the interior surface of the loop structure. This model has a hydrophobicity pattern similar to that of the original loop in the rhinovirus protein. The N-terminus of the loop becomes more compact primarily as a result of the dense packing of hydrophobic side chains from residues Met 1, Val 4, Pro 6, and Leu 8. This is associated with a reduction in the van der Waals energy by −6.4 kcal/mol relative to the A1-Rhino model. The center of the exterior surface alternates between positive and negative charges from residues Asp 5, Arg 7, Glu 9, and Lys 12. By placing the larger charged side chains along the exterior surface, the coulombic repulsion is reduced by −67 kcal/mol compared to the first rhinovirus model. The hydrogen-bonding energy, however, becomes less favorable by 88 kcal/mol relative to A1-Rhino. In an aqueous environment, however, we would not expect the enthalpy contributions of the hydrogen bonds to be significantly different for the A2-Rhino versus A1-Rhino models. The entropic contribution may favor the A2-Rhino model if this is an external loop in the Tat protein. Finally, the overall bonding energy is −19 kcal/mol lower in this latter model, resulting from reduced strain on the backbone, primarily at the proline residues.

The primary features of the α-helical model A-H are the two salt bridges resulting from the attractive charge–charge interactions between Asp 5 and Arg 7 and between Glu 9 and His 13. These coulombic energies dominate the overall energies of the helical model. To determine how stable each of these helical elements are for a four proline sequence, this model was subjected to molecular dynamics calculations for 12 ps. The structure was initially equilibrated at room temperature for 1 ps and then allowed to relax at 600 K for 12 ps (structure AHD). The conformations at approximately 1-ps intervals were saved and subsequently reequilibrated at 300 K for 1 ps and then energy minimized. The plot of the final minimized energies shows three distinct minima at 2, 6, and 10 ps (Figure 6). Diagonal contact plots were generated to correlate each energy minimum with a structural transition associated with the relaxation of specific domains of the helix (Figure 6 insets). The dots at the intersection of the residue–residue plot connect amino acids that are separated by 4.5–6.5 Å in their C_α carbons. Patterns showing contacts within this range of residues separated by three to four amino acids along the chain are indicative of α-helical structures.

Plot A at 0 ps shows that the α-helix conformation of A-H for the starting energy resides primarily between residues 3 and 13, with a small intervening turn centered at Arg 7. As the dynamics run proceeds at 600 K, the α-helical region between residues 3 and 7 is extended into a standard turn. The N-terminal residues 2–9 persists as a series of helical turns, with approximately a 2.5-residue repeat, through 5 ps of the dynamics calculation. This series of turns unravel beyond the 5-ps time frame, leaving the N-terminus as a series of extended loops. Thus the energy minimum at 2 ps is associated with melting of one α-helical turn at residues 3–7. The second minimum at 5–6 ps appears to be primarily associated with

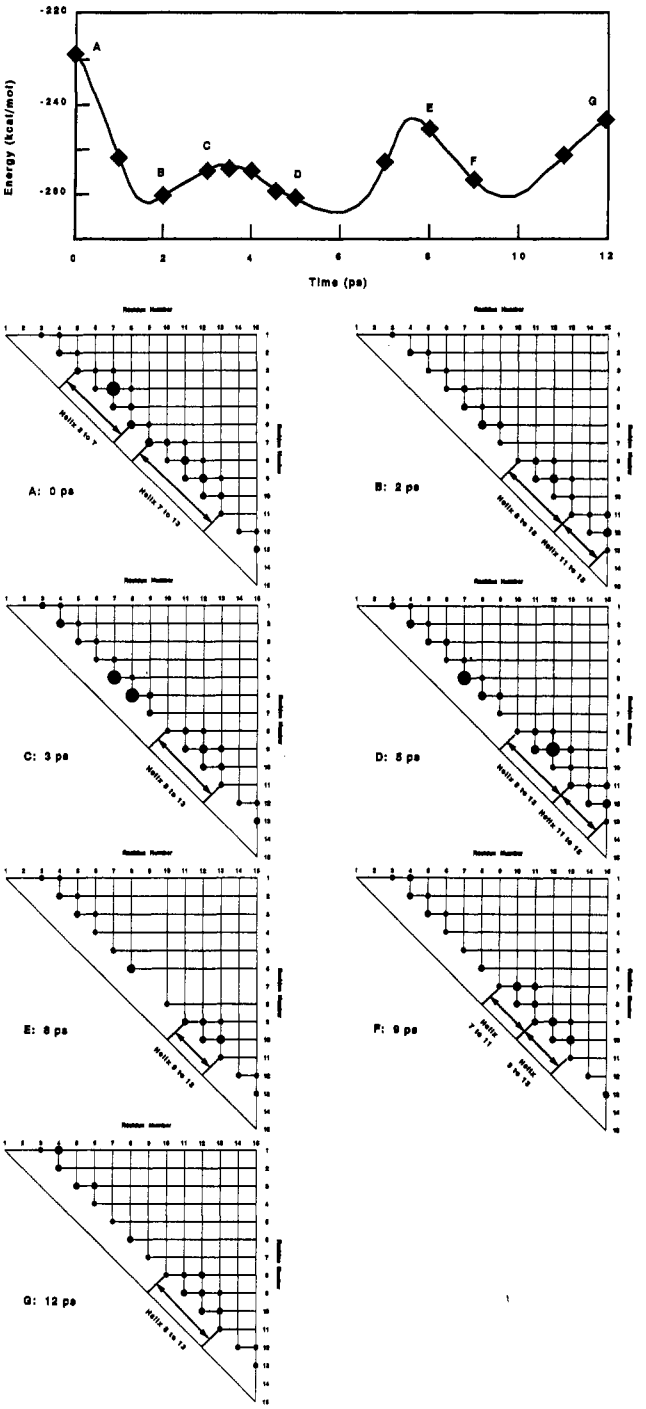

FIGURE 6: Energies and diagonal contact plots for structures along the molecular dynamics path for residues 1–15 of the Tat protein. The top panel shows the energy after minimization of conformations extracted at 0, 1, 2, 3, 3.5, 4, 4.5, 5, 7, 8, 9, and 12 ps into the dynamics calculation. The curve was drawn as an interpolation of these points. Panels A–G are the diagonal contact plots associated with these conformations. The residues are plotted along the x and y axes. Each point at the intersections indicates that the C_α carbons of the two residues are separated by 4.5–6.5 Å. The size of the points are indicative of the contact distance, with the larger points designating a closer contact.

side-chain rearrangements, leaving the secondary structural elements intact. The third minimum at 9–10 ps occurs after unraveling of the turns at the N-terminus. This requires overcoming a relatively large energy barrier. The single structural element that persists throughout this dynamics calculation is the α-helix residing between residues Leu 8 and His 13. This region is held intact primarily by the salt bridge

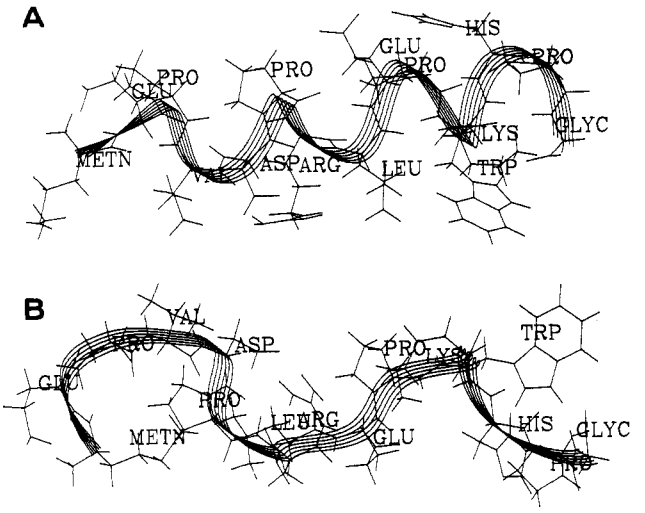

FIGURE 7: Comparison of the lowest energy conformations extracted from the molecular dynamics pathway for model A-H and A2-Rhino models. The backbone atoms are traced with ribbons to emphasize the secondary structures within each conformation. (A) Energy-minimized conformation of model A-H (designated AHD) after 6 ps of molecular dynamics and subsequent energy minimization by molecular mechanics using the AMBER force fields. (B) Energy-minimized conformation of model A2-Rhino (designated ARD) after 4 ps of molecular dynamics and subsequent energy minimization by molecular mechanics using the AMBER force fields.

between Glu 9 and His 13. Assuming that the imidazole ring of His 13 is protonated in the peptide, this α-helical turn appears to be the most stable element of the entire 14 amino acid stretch. The low energy structure at 6 ps is designated AHD.

A similar dynamics calculation was performed on the model A2-Rhino constructed from the rhinovirus coat protein. No distinguishing characteristics were observed in the diagonal plots of this model, showing that no helical structures formed within this entire sequence during the dynamics calculation. The turn at the N-terminus that forms a hydrophobic core was the only structure that persisted through the dynamics calculations. The middle eight residues of the peptide formed extended loops similar to those found in the helical model after 12 ps of dynamics calculations. The residues from position 8 to the C-terminus started as a bend at residues 10–15. This bend completely melted out by 3 ps and was replaced by a more stable bend centered between Pro 10 and Trp 11. The salt bridge between Glu 9 and His 13 in the helical model was thus replaced by an analogous interaction between Glu 9 and Lys 12. The resulting secondary structure for these residues was not helical.

A comparison of the results from the two dynamics calculations show that a helical structure may be stabilized by attractive coulombic forces that bridge the side chains of residues separated by three to four amino acids along the chain. However, the Glu 9 to His 13 interaction in the helical loop will not be stable in the peptide, if the imidazole ring of the His residue is not protonated. The Glu 9 to Lys 12 salt bridge observed in the rhinovirus model is, therefore, likely to be a more prevalent interaction, arguing strongly against the existence of any helical structure in the 1–14 sequence.

The two lowest energy conformations starting from the α-helix and the A2-Rhino models are compared in Figure 7. They both have final minimized energies of approximately −260 kcal/mol, which are at least 30 kcal/mol lower in energy than any of the starting models. The final structure starting from the α-helical model maintains the helix from residues 8 to 13, while the remainder of the chain is in extended helical loops. Starting from the rhinovirus protein, the resulting peptide maintains the hydrophobic core at the N-terminal amino acids, but the C-terminal end forms a bend rather than a helical turn. The energetics of these structures suggest that both are viable models, with the actual structure perhaps combining features from both (Table III).

Computer Modeling of Sequence 38–60. A search through the Brookhaven protein data base yielded three classes of sequences that were highly homologous with residues 38–47, residues 45–49, and residues 49–60 of the Tat protein. The results of this search lead to a model for this region of the Tat protein where the N-terminus forms two turns of an α-helix, followed by a five-residue S-shaped bend, and an extended chain at the highly basic C-terminus tail.

A nine-residue sequence (**Phe**-Gly-Val-Ser-Arg-**Ala-Leu-Gly**-Val) was found in the middle of a seven-turn α-helix chain in citrate synthase at residues 397–405 of both the A and B subunits (Wiegand et al., 1984). This shows four exact amino acid matches with the Tat protein (Phe 397 with 38, Ala 402 with 42, Leu 403 with 43, and Gly 404 with 44 when comparing the citrate synthase to Tat sequences, respectively, as indicated by the residue in boldface). Of the remaining amino acids, four are analogous replacements (Val 399 for Ile 39, Ser 400 for Thr 40, Arg 401 for Lys 41, and Val 405 for Ile 45 in the citrate synthase versus Tat sequences, respectively, as indicated by the underlined residues); one position (Gly 398) is an insertion within the citrate synthase structure that is absent in the Tat sequence. This strongly suggests that these residues within the Tat protein also form an α-helix, consistent with the results from the Chou–Fasman predictions.

The sequence **Gly-Ile**-Thr-**Ser**-Trp-**Gly**-His was found between positions 116–122 in the kallikrein B subunit of porcine serine protease (Bode et al., 1983). In addition, an analogous sequence (**Gly-Ile**-Val-**Ser-Tyr-Gly**-His) was located at 194–200 of a serine protease from rat mast cells (Remington et al., 1988). All these sequences form S-shaped bends at the N-terminus that lead into a β-strand. The latter sequence showed the highest degree of homology with the Tat sequence, with five exact residue matches (Gly 194 to 44, Ile 95 to 45, Ser 197 to 46, Tyr 198 to 47, and Gly 199 to 48, protease to Tat, respectively), one analogous substitution (His 200 for Arg 49 comparing the protease to Tat), and a single insertion of a valine residue at position 196 of the protease that is not present in the Tat sequence. This peptide from the rat serine protease was used to model the region from 44 to 48 in the Tat protein.

Finally, the highly basic region from 49 to 60 (with the sequence Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg) of the Tat protein was found to be highly homologous to the basic region from residues 22 to 31 (with the sequence **Arg**-Arg-**Lys-Arg-Arg**-Ala-Lys-**Arg-Arg**) of the southern mosaic bean virus coat protein (Namba et al., 1989). This stretch forms part of the N-terminus domain of the coat protein that is located toward the interior of the virus. Unfortunately, the first 61 residues at the N-terminus were crystallographically disordered in the structure and therefore could not provide a template with which to build a model. The long stretch of basic residues, however, suggests that this stretch of amino acids is extended due to electrostatic repulsion of the positively charged side chains and does not form a helical structure. Indeed, a search through the crystallographic data base for stretches of sequences that were composed of at least five out six basic residues showed that nearly all were in loop structures. The exceptions include residues 407–412 of tRNA synthetase (Brick et al., 1989), which showed only weak electron density

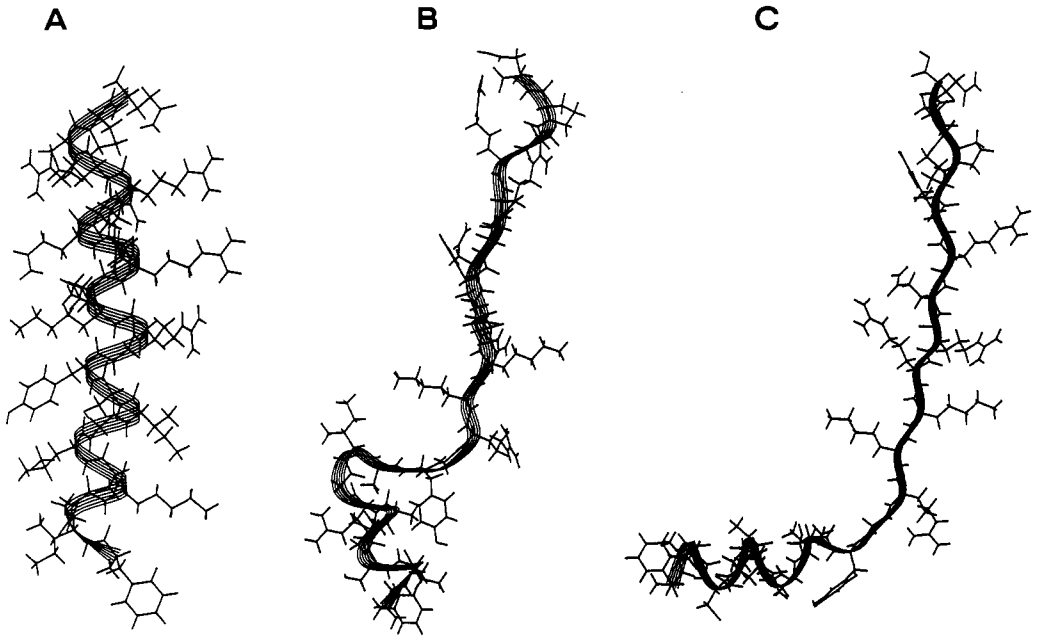

FIGURE 8: Models for amino acids 38–60 of the Tat protein. The backbone is traced as a ribbon to emphasize the secondary structure of the peptide. (A) Model TATH; (B) model TAT38a; and (C) model TAT38b (see the legend to Table IV for the key to model designations).

and therefore could not be assigned, and a short stretch of Arg-Leu-Lys-Lys-Arg-Arg (residues 7–12) imbedded in middle of a four-turn α-helix of a corepressor (Mondragon et al., 1989). In this latter sequence, the arginine residue at position 7 was preceded by an acidic glutamic acid. This negatively charged residue may facilitate propagation of the helix through the five positively charged residues by interacting with the two lysine side chains at positions 9 and 10. The stretch of basic residues in the Tat protein is neither preceded nor followed by any acidic residues, which therefore precludes any electrostatic stabilization of a helix through this region. This stretch of basic amino acids was thus modeled as an extended chain.

The final three amino acids residues (Pro-Pro-Gln) were simply modeled by using the conformations of two *trans*-prolines and a glutamine in an extended conformation. The complete structure of residues 38–60 was thus modeled by patching a helix to an S-shaped bend, followed by a long stretch of positively charged amino acids and prolines in an extended chain conformation.

Three unique models were built from for the sequence 38–60 of the Tat protein. The first two (TAT38a and TAT38b) were built from the protein templates with a two-turn α-helix extending from Phe 38 to Leu 43 and an S-shaped bend at residues Gly 44 to Tyr 47 connecting this helix to the extended basic tail. These two models were constructed with the glycine conforming either in the helix conformation or in the S-bend conformation for the models TAT38a and TAT38b, respectively. A third model (TATH) was constructed with the entire 38–60 sequence of Tat in an α-helix. This latter model allows us to test the prediction that the long stretch of basic residues would not favor an α-helix.

After energy minimization, the TATH model that was entirely helical (Figure 8) had the highest (least favorable) overall energy (Table IV). This resulted from an average +124 kcal/mol increase in the electrostatic repulsion of the basic side chains. This more than compensated for the −63 kcal/mol hydrogen-bonding stabilization resulting from extension of the helix. Extending the helix to residue 60 had an average overall destabilizing effect of +18 kcal/mol on the Tat protein. We therefore predict that only the N-terminal part of this sequence from the Tat protein would adopt a helical conformation.

Table IV: Energies for Sequence 38–60 of the Tat Protein[a]

energies (kcal/mol)	models[b]				
	TATH	TAT38a	TAT38b	TATHD	TAT38D
bonding	51	71	64	94	84
H-bonding	-1349	-1321	-1249	-1722	-1593
nonbonding	-45	-14	-24	-29	-4
coulombic	1197	1097	1049	1405	1248
total	-146	-167	-160	-251	-265

[a] Energies were minimized by using the AMBER force fields (Weiner & Kollman, 1986) to a maximum derivative of 0.1 kcal/mol. The final energies after minimization of the three starting models (TATH, TAT38a, and TAT38b) are compared along with the lowest energy conformations (TATHD and TAT38D) extracted from molecular dynamics calculations on TATH and TAT38a, respectively. [b] Model TATH was constructed as a helical model by using standard ϕ and ψ angles for an α-helix, with the proline residues in their trans conformations. Models TAT38a and TAT38b were constructed by using the crystal structures of citrate synthetase and rat serine protease as templates for helical region. TAT38a has the residue G44 adopting the conformation of the preceding helix, while, in TAT38b, this residue adopts the conformation of the following S-bend. Model TATHD and TAT38D are the lowest energy conformations extracted from molecular dynamics calculations on TATH and TAT38a, respectively.

The two structures constructed from the crystallographic templates were both similar in overall energies, but for different structural reasons (Figure 8). The model TAT38a, built with G44 assuming the conformation of the preceding helix, formed a more compact bend structure connecting the helix and basic chain domains. The second model TAT38b built with G44 in the S-bend, was more extended. This resulted in a 48 kcal/mol difference in the coulombic terms of the energies, with the compact TAT38a model showing more electrostatic repulsion. This repulsion is from an interaction of the positively charged basic tail with the helical dipole (which points toward the C-terminus, positive to negative). The extended TAT38b structure places the acidic tail such that the positive charges help to reinforce the helical dipole, while the orientation of the tail is nearly perpendicular to the helix in the compact TAT38a structure. In contrast, the more compact TAT38a structure can form more hydrogen bonds between the loop and

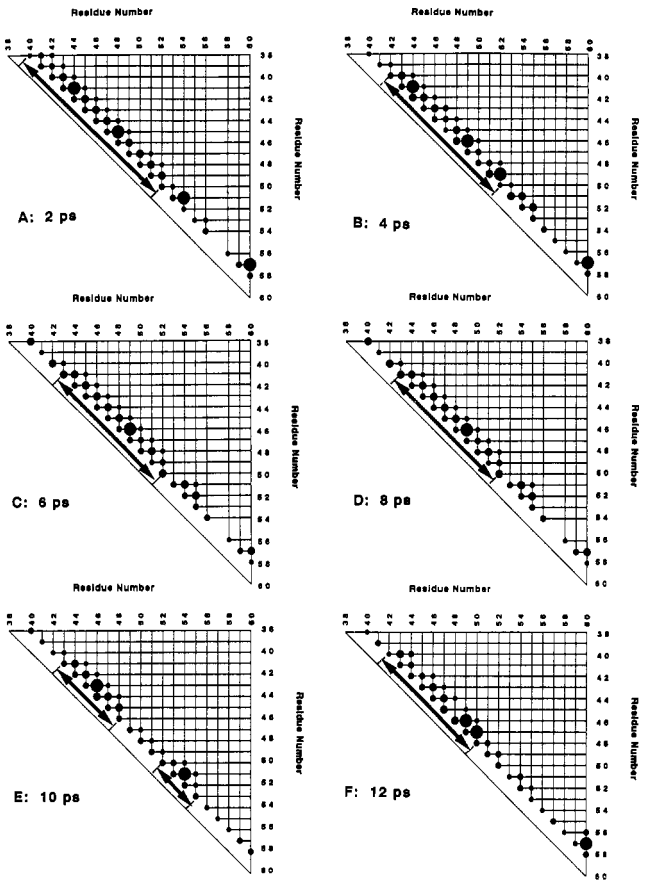

FIGURE 9: Diagonal contact plots of conformations extracted along the molecular dynamics path for model TATH. Contact plots were generated from conformations extracted at 2–12 ps during the molecular dynamics calculation on the TATH model. The residues are numbered according to where each residue occurs in the native Tat protein (38–60). The residues are plotted along the x and y axes. Each point at the intersections indicates that the C_α carbons of the two residues are separated by 4.5–6.5 Å. The size of the points are indicative of the contact distance, with the larger points designating a closer contact.

the helix than can the extended form. The two effects are compensatory, leaving only 6.4 kcal/mol separating the two conformations. Thus both or either structure may exist in solution. These models place an α-helix at the seven N-terminal residues of this domain in the Tat protein, which is consistent with the percent of helical residues determined from the CD spectra for this peptide in 90% TFE.

Molecular dynamics calculations were performed on the TAT38a model and the fully α-helical TATH model to determine which, if any, of the structural elements are stable. These calculations were run for 13 ps, starting with a 1-ps equilibration at room temperature, followed by 12 ps at 450 K. The model starting from the TAT38a model became entirely unwound after 5 ps of dynamics. The helix constructed at the 10 N-terminal residues unwound starting from residue 1 and worked its way toward the C-terminus. It was no longer helical after 3 ps. The bend at residues Ser 46 to Lys 50 persisted through 5 ps of dynamics. The polybasic C-terminal tail remained extended throughout the calculation.

When the fully helical TATH model was subjected to the energetics of the dynamics calculations, the structure was found to be more resilient to melting. The last 14 amino acids at the C-terminus lost its helicity during the initial equilibration steps. The melting of the helix proceeded from both the N- and the C-termini through 6 ps of dynamics. At that point, the N-terminus stabilized, showing no additional melting.

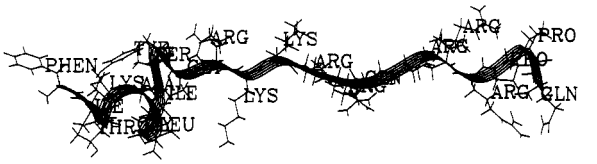

FIGURE 10: Molecular model of the lowest energy conformation for residues 38–60 of the Tat protein. The sequence is drawn from the N- to C-terminus. A ribbon traces the backbone atoms to emphasize the secondary structure along the chain.

After 12 ps of dynamics, the helical region stabilized somewhere between residue Thr 40 and Lys 50, existing as either a continuous two turn helix or a three turn helix that is interrupted intermittently by helical loops. Residue–residue contact plots are given in Figure 9.

These molecular dynamics calculations provided us with the best model for this region of the Tat peptide. Of all the conformations searched during this process, the lowest minimized energy structure resulted after 1 ps of dynamics starting with the TAT38a model, designated TAT38D. Its final minimized energy (−265 kcal/mol) was over 100 kcal/mol lower than any of the starting structures (see Table IV for comparisons). The final energy-minimized structure for this peptide (shown in Figure 10) shows approximately 1.5 turns of an α-helix starting at residue Lys 41 and ending at Tyr 47. The dynamics calculations on the TATH model shows that the coulombic repulsion at the basic tail tends to favor the extended conformation. The compact S-shaped bend in TAT38a was found to be reasonably stable over the course of the dynamics calculation. The four residues at the N-terminus of the TAT38a peptide melted out from the helical structure. This is probably an end effect peculiar to the isolated peptide. The dynamics results from the helical starting structure suggests that, in the presence of other stabilizing factors, the melting of the helix from the N-terminus stops at residue Lys 41. This suggests that, in the intact Tat protein, the α-helix extends back to residue Phe 38, as suggested from the Chou–Fasman predictions.

Conclusions. In this study, we have shown by CD that peptide 38–60, but not peptide 2–23, is able to adopt some α-helical structure in 90% TFE (Figure 2). The presence of an α-helix for sequence 38–45 (region III) is predicted by energy minimization with molecular dynamics and the Chou–Fasman method and possesses an amphipathic characteristic (Figure 3). Region IV, containing a nuclear localization signal, is described as an extended structure by energy minimization with molecular dynamics and as an α-helix by the Chou–Fasman prediction. Assuming a single structure in solution, CD data analysis (Table I) gives 9 or 10 residues in an α-helix for peptide 38–60 and thus supports the models TAT38a and TAT38b obtained by computer modeling (Figure 10). An α-helical conformation for the 13 amino acid sequence adjacent to the N-terminus, as previously proposed, is not likely, on the basis of CD spectra and computer modeling. Activating regions are known that are proline rich (Mitchell & Tjian, 1989) and sequence 1–14 could be an activating region for this reason but not because it is an α-helix. Sequence 38–45 has the amphipathic characteristics of known α-helical activating region and is adjacent to the expected nuclear localization signal.

ACKNOWLEDGMENTS

We are greatly indepted to Thêrèse Brando, Barbara Hanson, Jeannine Lawrence, and Barbara Robbins for technical assistance. Drs. Robert R. Becker, Charles Robert, and Arazdordi Tourmadje are acknowledged for fruitful discus-

sions. Erwann P. Loret gratefully acknowledges the Franco-American Commission for Educational Exchange, Dr. El Mostapha Bahraoui, Prof. Jean Claude Gluckman, and Prof. Luc Montagnier for their support of his research.

REFERENCES

Arnold, E., & Rossmann, M. G. (1988) *Acta Crystallogr., Sect. A 44*, 270–282.

Arya, S. K., Guo, C., Josephs, S. F., & Wong-Staal, F. (1985) *Science 229*, 69–73.

Barany, G., & Merrifield, R. B. (1980) in *The Peptides: Analysis, Synthesis, Biology* (Gross, E., & Meinhofer, J., Eds.) Vol. 2, pp 1–284, Academic Press, New York

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., & Bartunik, H. (1983) *J. Mol. 164*, 237–251.

Braddock, M., Thorburn, A. M., Chambers, A., Elliot, G. D., Anderson, G. J. Kingsman, A. J., & Kingsman, S. M. (1990) *Cell 62*, 1123–1133.

Brick, P. Bhat, T. N., & Blow, D. M. (1989) *J. Mol. Biol. 208*, 83–95.

Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.

Coste, J., Le Nguyen, D., & Castro, B. (1990) *Tetrahedron Lett. 31*, 205–208.

Cullen, B. R. (1986) *Cell 46*, 973–982.

Cullen, B. R., & Greene, W. C. (1990) *Virology 178*, 1–5.

Deisenhofer, J., & Michel, H. (1989) *Science 245*, 1463–1466.

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., & Valerio, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6925–6929.

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

Frankel, A. D., Biancalana, S., & Hudson, D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7397–7401.

Garcia, J. A., Harrich, D., Pearson, L., Mitsuyasu, R., & Gaynor, R. B. (1988) *EMBO J. 7*, 3143–3147.

Garcia, J. A., Harrich, D., Soultanakis, E., Wu, F., Mitsuyasu, R., & Gaynor, R. B. (1989) *EMBO J. 8*, 765–778.

Gentz, R., Chen, C. H., & Rosen, C. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 821–824.

Green, M., & Lowenstein, P. M. (1988) *Cell 55*, 1179–1188.

Green, M., Ishino, M., & Lowenstein, P. M. (1989) *Cell 58*, 215–223.

Hauber, J., Perkins, A., Heimer, E. P., & Cullen, B. R. (1989) *J. Virol. 63*, 1181–1187.

Jeang, K. T., Shank, P. R., & Kumar, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8291–8295.

Jenness, D. D., Sprecher, C., & Johnson, W. C. (1976) *Biopolymers 15*, 513–521.

Jeyapaul, J., Reddy, M. R., & Khan, S. K. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 7030–7034.

Johnson, W. C., Jr. (1971) *Rev. Sci. Instrum. 42*, 1283–1286.

Johnson, W. C., Jr. (1985) *Methods Biochem. Anal. 31*, 61–163.

Kao, S. Y., Calman, A., Luciw, P. A., & Peterlin, B. M. (1987) *Nature 330*, 489–493.

Karplus, P. A., & Schulz, G. E. (1985) *Naturwissenschaften 72*, 212–213.

Kuppuswamy, M., Subramanian, T., Srinavasan, A., & Chinnadurai, G. (1989) *Nucleic Acids Res. 17*, 3551–3561.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Laspia, M. F., Rice, A. P., & Peterlin, B. M. (1989) *Cell 59*, 283–292.

Li, C. H., Chung, D., Yamashiro, D., & Lee, C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4306–4309.

Manavalan, P., & Johnson, W. C., Jr. (1987) *Anal. Biochem. 167*, 76–85.

Matsuura, Y., Kusunoki, M., Harada, W., & Kakudo, M. (1984), *J. Mol. Biol. 174*, 205–216.

Mitchell, P. J., & Tjian, R. (1989) *Science 245*, 371–378.

Mondragon, A., Wolberger, C., & Harrison, S. C. (1989) *J. Mol. Biol. 205*, 179–190.

Muesing, M. A., Smith, D. H., & Capon, D. J. (1987) *Cell 48*, 691–701.

Namba, K., Pattanayek, R., & Stubbs, G. (1989) *J. Mol. Biol. 208*, 307–318.

Previlege, P., Jr., & Fasman, G. D. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 391–417, Plenum Press, New York.

Ptashne (1988) *Nature 335*, 683–689.

Rappaport, J., Lee, S. J., Khalili, K., & Wong-Staal, F. (1989) *New Biol. 1*, 101–110.

Ratner, L., Haseltine, W. A., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whithorn, 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. A., & Wong-Staal, F. (1985) *Nature 313*, 277–284.

Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthiews, B. W., & Neurath, H. (1988) *Biochemistry 27*, 8097–9005.

Rice, A. P., & Matthews, M. B. (1988) *Cell 52*, 5–6.

Rosen, C. A., & Pavlakis, G. N. (1990) *AIDS (London) 4*, 499–509.

Rosen, C. A., Sodroski, J. G., & Haseltine, W. A. (1985) *Cell 41*, 813–825.

Rosen, C. A., Sodroski, J. G., Goh, W. C., Dayton, A. I., Loppke, J., & Haseltine, W. A. (1986) *Nature 319*, 555–559.

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

Sadaie, M. R., Benaissa, Z. N., Cullen, B. R., & Wong-Staal, F. (1989) *DNA 8*, 669–674.

Schiffer, M., & Edmundson, A. B. (1967) *Biophys. J. 8*, 29–39.

Siegels, L. J., Ratner, L., Josephs, S. F., Derse, F., Feinberg, M. B., Reyes, G. R., O'Brien, S. J., & Wong-Staal, F. (1986) *Virology 148*, 226–231.

Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F., & Haselthine, W. (1985) *Science 229*, 74–77.

Southgate, C., Zapp, M. L., & Green, M. C. (1990) *Nature 345*, 640–642.

Tiley, L. S., Brown, P. H., & Cullen B. R. (1990) *Virology 178*, 560–567.

Urry, D. W. (1972) *Biochim. Biophys. Acta 265*, 115–168.

Weeks, K. M., Ampe, C., Shultz, S. C., Steitz, T. A., & Crothers, D. M. (1990) *Science 249*, 1281–1285.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr. & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.

Weiner, S. J., Kollman, P. A., & Nguyen, D. T. (1986) *J. Comput. Chem. 7*, 230–252.

Wiegand, G., Remington, S., Deisenhofer, J., & Huber, R. (1984) *J. Mol. Biol. 174*, 205–213.

Woody, R. W. (1985) in *The Peptides* (Hruby, V., Ed.) Vol. 7, pp 15–114, Academic Press Inc., New York.

Wright, C. M., Felber, B. K., & Paskalis, G. N. (1986) *Science 234*, 988–992.