

Conformational Heterogeneity in Two Regions of TAT Results in Structural Variations of This Protein as a Function of HIV-1 Isolates*

(Received for publication, December 7, 1995, and in revised form, June 14, 1996)

Catherine J. Gregoire and Erwann P. Loret‡

From the Institut de Biologie Structurale et Microbiologie, Laboratoire d'Ingénierie et Dynamique des Systèmes Membranaires, CNRS UPR 9027, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

TAT protein is an essential regulatory protein of the human immunodeficiency virus type 1 (HIV-1). Inhibition of TAT activity blocks the virus cycle, and a drug that blocks TAT is one of the possibilities to cure AIDS. Circular dichroism (CD) was measured for TAT peptides covering the TAT sequence with overlaps. The CD spectrum of each peptide was measured in different solvents to evaluate the ability of each TAT region to form different secondary structures. The most variation or conformational heterogeneity is observed with the two regions adjacent to the TAT basic region. CD data show that the basic region can adopt an extended structure in a full TAT protein, which is not the case for the isolated peptide. TAT sequences from the different HIV-1 isolates were analyzed, and the results showed that the sequences could be gathered into six groups. Molecular modeling was done on the various isolates based on a TAT structure from two-dimensional NMR. After minimization and dynamic steps, the modeled three-dimensional structures were compared. The results showed structural variations of the TAT protein as a function of the HIV-1 isolates. These structural variations were mainly in the two regions adjacent to the basic region, confirming the conformational heterogeneity indicated by the CD measurements. Furthermore, Chou-Fasman analysis shows significant changes in propensities for each secondary structure only for regions III and V. This conformational heterogeneity should be essential for TAT activity and points out that regions III and V are a poor potential target to design a TAT ligand. We propose a target involving TAT structurally conserved regions, accessible whatever the size of the TAT C terminus.

Human immunodeficiency virus type 1 (HIV-1)¹ gene expression is regulated by TAT, which is a viral protein (1). TAT has been described as a trans-activator inducing the elongation of early HIV-1 RNA transcripts (2). This trans-activation necessitates the binding of TAT on an RNA stem-loop structure called TAR, which is located at the 5' end of all HIV-1 mRNA (3–12). Nevertheless, TAT seems to be more than a trans-activator and is involved in cellular disorder connected to AIDS

pathology. TAT modifies the cellular redox potential and amplifies the activity of tumor necrosis factor (13). This cytokine activates the transcription factor NFκB, which binds on the HIV promotor region located in the long terminal repeat and makes possible the transcription process (13). TAT in synergy with basic fibroblast growth factor is involved in the induction of Kaposi's sarcoma lesions (14). TAT is able to repress the transcription of the major histocompatibility complex class I gene, and a decrease in class I molecules provides for the virus a mechanism to evade the host immune response (15). TAT seems to participate in the induction of lymphocyte apoptosis and might contribute to the depletion of the CD4+ T cells in AIDS (16, 17).

TAT is a small protein of 86 to 102 residues with variable C-terminal sizes as a function of HIV-1 isolates (see Ref. 18 for a review). The *tat* gene is composed of two exons, and the first, corresponding to the 72 N-terminal residues, is sufficient for trans-activation (19). TAT amino acid sequence was subdivided into several functional regions (20). Region I is an acidic domain containing several prolines, unable to form an α -helix as previously proposed (21), and could be a motif that prevents exonuclease digestion (22). The cysteine-rich domain (region II), located between residues 21 and 38, contains seven cysteines, and mutation of cysteines completely abolishes TAT trans-activation (20). The presence of disulfide bridges is not yet clearly stated, and contradictory data exist on this issue (23, 24). Region III (residues 38–47) contains a highly conserved motif where lysine 41 was pointed out as essential for TAT activity (20). This region is important for the specific TAT-TAR interaction because a peptide corresponding to regions III, IV, and V was reported to bind to TAR with an affinity similar to the full TAT protein, whereas peptides with region III deleted are still able to bind to TAT but can not discriminate wild type TAR and TAR mutants (25). Region IV (residues 48–58) is called the "basic domain" because of its arginines and is essential for the binding on TAR (10–12). The stem region of TAR, located between the bulge and the loop, adopts an A-form helix (26, 27). TAR keeps its A-form helix structure upon binding, and region IV adopts an extended structure to fit into the TAR major groove (26). Arginines in region IV are directly involved in the binding, creating electrostatic bounds with TAR phosphates (26, 27). Deletion of region V (residues 59–72) affects TAT trans-activation (20), and this region also seems to be involved in the specificity of the TAT-TAR interaction (26, 28). Upon binding on TAR, region V is able to form a helicoidal structure with a glutamine-rich surface and could be involved in a specific interaction with three highly conserved nucleotides located on the TAR loop (26). Region VI, encoded by exon 2, is variable in its length and was involved in the TAT-mediated major histocompatibility complex class I gene repression (15). Region VI is also required to observe the

* This work was supported by the French National Aids Research Agency (ANRS) and CNRS. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. E-mail: loret@ibsm.cnrs-mrs.fr.

¹ The abbreviations used are: HIV-1, human immunodeficiency virus type 1; TAT, HIV-1 trans-acting transcriptional activator; TFE, 2,2,2-trifluoroethanol; TAR, RNA trans-activation response element.

	1	10	20	30	40	50	60	70	80	90	
TAT_HV1Z2	<u>MDPVDENLEPW</u> <u>NHFGSOFK</u> TACNRCHCKKCCYHCQVCFITKGLGISYGRKKRRRRRPSQGGOTHODPIKQPSQPRGDPTGPKE										86
TAT_HV1Z6L.....									N.....	86
TAT_HV1NDL.S.....R.....K.....						K.P.D.A.V.....E.....S.....K				86
TAT_HV1MAL.....R.P..K.Y.....M.....						P.N.A.....L.E.....H.....HPTGPKE				87
TAT_HV1H2	.E.....RL...K.....TN.Y.....F.....A.....						AH.NS.....ASLS...T.S.....				86
TAT_HV112	.E.....RL...K.....TN.Y.....F.....A.....						AP.S.....VSLs...T.S.....				86
TAT_HV1PV	.E.....RL...K.....TN.Y.....F.....A.....						P.S.....VSLs...T.S.....				86
TAT_HV1B1	.E.....RL...K.....TN.Y.....F.....A.....						P.S.....VSLs...T.S.....				86
TAT_HV1BR	.E.....RL...K.....TT.Y.....F.....T..A.....						P.S.....VSLs...T.....				86
TAT_HV1Y2	.E.....L...K.....R.....N.Y.....F.....TK.....						P.DS.....SSLS...T..L.....T.SKKKVERETETDPVH				101
TAT_HV1A2	.E.....L...K.....R.....N.Y.....F.YA..TR.....						AP.DS.....ASLS...A.S.....T.SKKKVERETETDPPD				101
TAT_HV1SCRL...K.....A..TS.Y.....F.....T.....						AP.DS.....VSL..A.A.....SKKKVERETETDPVD				101
TAT_HV1JR	.E.....SL...K.....TN.Y.....L.....T.....						P.DS.....VSL.....Q.....SKKKVERETETDPDN				101
TAT_HV1MN	.E.....RL...K.....TT.Y.....F.....TK.A.....						APEDS.....VSL.....AP.F.....SKKKVERETETHPVD				101
TAT_HV1C4	.E.....RL...K.....TN.Y.....F.....TK.A.....						AH.DS.N..ASLS.....T.....PKKEVEREAETDPLD				101
TAT_HV1S1	.E.....RL...K.....TN.Y.....F.....						APPDSEV..VSL..A..Q.....SKKKVERETETDPVH				101
TAT_HV1OY	.E.....RL...K.....S.N.Y..R..L.....TK.....						AP.DSK...VSLs...A.....SKKKVERETETDPED				101
TAT_HV1RH	.E.....RL...K.....N.Y.....L.....						G.P.S.....VSLs...T.....SKEKVERETETDPAVQ102				
TAT_HV1ELL.....R.P..K.....P.....LN.....						G.P...A..V.....QKKKVESEAEETDP				99

FIG. 1. Amino acid sequences of TAT protein from HIV-1 isolates. Regions I, III, V are underlined. Sequences were aligned using the MULTALIN method (35), with the TAT Z2 sequence as target. From this method, six sequence related groups were pointed out (data not shown). The sequences selected in each group are indicated in bold.

TAT effect on tumor necrosis factor-mediated cytotoxicity (13). Finally, two-dimensional NMR studies were made on TAT, and one described the TAT basic domain included in a chimerical protein as a stable α -helix (29). This is in contradiction to two recent two-dimensional NMR studies describing region IV as an extended structure in the three-dimensional structure of the full HIV-1 TAT protein (30) and in a related peptide-TAR complex from the bovine immunodeficiency virus (31).

The aim of this research is to provide structural information on TAT useful to design an inhibitor of this protein. We used two different approaches to do this structural study. Circular dichroism (CD) was done on six peptides covering with overlaps the full TAT sequence from a HIV-1 isolate found frequently in North America and Europe (Bru isolate). The CD spectrum of each peptide was measured in different solvents to evaluate the conformational heterogeneity of TAT regions. The results showed that the regions III and V had the most variable structures and were required to have region IV in a full extended structure. To confirm this result, molecular modeling was done on TAT sequences from HIV-1 isolates from all over the world. The TAT sequences were gathered in structurally related groups. From the TAT-Z2 two-dimensional NMR structure (30), we built five TAT structures from TAT sequences representative of each group. Our molecular modeling gives evidence for greater structural variation among the different sources for two regions. As could be expected from CD data, these structural changes were located mainly in regions III and V and led to TAT structures with backbones impossible to superimpose. These results point out a target to design a drug able to block the TAT protein whatever the isolates. We propose that an allosteric ligand, making a bridge between region IV and region I or regions IV and VI, should be the most appropriate to inhibit TAT.

MATERIALS AND METHODS

Peptide Synthesis—Peptides were assembled according to the method of Barany and Merrifield (32) on 4-(oxymethyl)-phenylacetamidomethyl resin (0.5 mmol) (Applied Biosystems Inc., Forster City, CA) on a half-automated synthesizer (NPS4000, Neosystem, Strasbourg, France) as described previously (22). Amino acid analyses were performed on a model 6300 Beckman analyzer. Purified peptides (1–4 nmol) were hydrolyzed in 6 M HCl for 20 h at 110 °C according to the method of Sanger and Thompson (34). These analyses were used to determine peptide concentration, which was correlated to the absorption at 190 nm measured with the JOBIN-YVON MARK VI.

Circular Dichroism Measurements—The samples were in 20 mM phosphate buffer (pH 7), in a water solution containing 80% of trifluoroethanol (TFE), or in 2 to 1 mM (nonmicellar) SDS. CD spectra were measured in 50- μ m path length cells from 260 to 178 nm with a JOBIN-YVON (Long-Jumeau, France) UV CD spectrophotometer

(MARK VI). The instrument was calibrated with (+)-10-camphorsulfonic acid. A ratio of 2.2 was found between the positive CD band at 290.5 nm and the negative band at 192.5 nm. All spectra were measured at 25 °C. Data were collected at 0.5-nm intervals with a scan rate of 1 nm/min. CD spectra are reported as $\Delta\epsilon$ per amide. Light transmission of samples in the far UV (260 to 178 nm) were checked with the same machine (MARK VI), which is able to measure the absorption spectrum. The CD data were analyzed to determine the secondary structure content according to the method of Manavalan and Johnson (33) using a set of 32 reference proteins.

Molecular Modeling—Models were built with the Insight II and Discover software from MSI Technologies, Inc. (San Diego, CA), running on a Silicon Graphics R4600 workstation. The structures were optimized with the CVFF force field in term of the internal energies, using the van der Waals energy to monitor each step of the modeling. The pH was set at 7. Minimization was performed with steepest descent and conjugate gradient algorithms with a maximum derivative of 0.001 kcal/Å in the final steps. Dynamic was performed at 300 K for 1.1 ps using 1000 steps. The analysis of the trajectory was done with 110 different structures selected from the 1100 produced from the dynamics.

RESULTS AND DISCUSSION

The TAT two-dimensional NMR structure was determined from a HIV-1 isolated in Zaire. This is an 86-residue-long protein that does not have the long C-terminal sequence observed in other HIV-1 isolates (Fig. 1). These sequences were classified by the MULTALIN software (35) using the TAT Z2 sequence as the target. The dendrogram obtained from MULTALIN points out that the TAT sequences could be gathered in six groups (data not shown), and TAT Z2 has a sequence diverging significantly from HIV-1 isolates found in the United States and Europe. Mutations are observed in every TAT region except region IV, which is rather conserved. The N terminus is partially conserved in most of the HIV-1 isolates, but the TAT Z2 group is one of the exceptions (Fig. 1). TAT is a rather short protein with no possible disulfide bridges to rigidify the structure by binding together different parts of the polypeptidic chain (the cysteine residues are located essentially in region II), and a conformational heterogeneity can be expected in this protein. Although the nature of the environment influences the three-dimensional structure of proteins and particularly their secondary structure (36), the dihedral angles of the α carbons have limits as a function of the nature of the lateral chains (37). Therefore, the conformational heterogeneity could not be equivalent in the six TAT regions.

Circular dichroism (CD) was done on peptides covering the TAT Bru sequence to evaluate if the TAT Bru regions could adopt secondary structures similar to TAT Z2 regions. Since the specific environment present in a full TAT protein could have an effect on the structure, the CD of each peptide was

measured in three different solvents. Fig. 2 shows the CD spectra of six peptides in an aqueous solution at pH 7, in 80% trifluoroethanol (TFE), and in 1–2 mM SDS. These three solvents tend peptides to adopt random coil, α -helix, and extended structures, respectively (36). Measurements were made between 178 and 260 nm and correspond essentially to π - π^* and n - π^* transitions of the amide chromophore located in polypeptide chains (38). The content in secondary structures for each

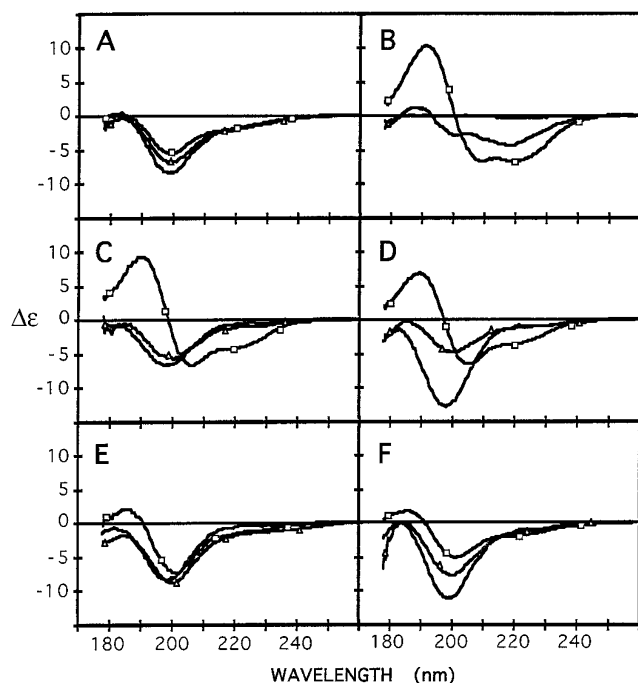


FIG. 2. Circular dichroism spectra in phosphate buffer, pH 7 (solid line), in 80% trifluoroethanol (square), and in 1–2 mM nonmicellar sodium dodecyl sulfate (triangle). Peptide 2–23 is shown in panel A, peptide 13–48 in panel B, peptide 38–60 in panel C, peptide 38–72 in panel D, peptide 47–72 in panel E, and peptide 57–86 in panel F. Most peptides, except peptide 13–48, display an important negative band at 200 nm associated with a random coil structure in an aqueous solvent or SDS. Peptide 13–48 precipitated in SDS. In TFE, peptides 13–48, 38–60, and 38–72 have CD bands typical of helical structures.

peptide was determined from CD spectra according to the SVD method (34) and is summarized in Table I.

There are difficulties in analyzing CD spectra of peptides, because CD cannot tell the difference between static structures and dynamic averages. Peptides in aqueous solution tend to be a random coil, which is a dynamic structure that changes constantly. A CD spectrum of a random coil is characterized by a negative band at 200 nm, which corresponds to an average of conformational space available for the peptide. However, proteins and peptides can have a negative CD band at 200 nm and not be a random coil. In this case, the negative band at 200 nm is due to stable structures that are a mixture of β -strand, β -turns and “other structures.” There are no random coils in a protein, and therefore SVD with a protein basis cannot analyze random coil correctly but will analyze as a mixture of β -strand, β -turn, and other structures.² In this study, we discriminate in aqueous buffer a CD spectrum due to the expected random coil from a CD spectrum due to static structures, by comparing for each peptide the CD spectrum in aqueous buffer with the CD spectra obtained in TFE and SDS. When the CD spectrum changes little in TFE or SDS, this means that the peptide is in a static structure in aqueous buffer, and the SVD analysis is relevant. Random coils are indicated by RC in Table I.

Peptide 2–23 corresponds to region I and has a CD spectrum in aqueous solution that is related to random coil. However, the CD spectra are similar in the three solvents indicating static structures (Fig. 2A). This peptide has a content in secondary structures (Table I) consistent with the presence of two β -turns, which correlates with the TAT Z2 three-dimensional structure. Peptide 13–48 (mainly regions II and III) is the most structured of the peptides in aqueous solvent (Fig. 2B). Precipitation in SDS occurred with peptide 13–48 pointing out the rigidity of this peptide. This rigidity appears to be due to region II since two other peptides, 38–60 and 38–72, having the region III do not precipitate (Fig. 2C). The increase in α -helix observed in TFE is due to region III (22), as the CD of other peptides will confirm. Two β -turns are observed in the region II of TAT Z2. The percentage of β -turn structure observed in the peptide 13–48 is consistent with two β -turns located in region II and not in region III, since the percentage remains similar in

² Professor W. Curtis Johnson, Jr., personal communication.

TABLE I
Secondary structure analyses

Peptides	TAT regions	Solvent	H ^a	B ^a	T ^a	O ^a	Total
A-(2–23)	I	Aqueous ^b	0.08	0.07	0.35	0.49	0.99
		TFE ^c	0.10	0.13	0.29	0.48	1.00
		SDS ^d	0.04	0.15	0.37	0.43	1.00
B-(13–48)	50% I + II + III	Aqueous	0.10	0.34	0.17	0.40	1.01
		TFE	0.48	0.16	0.15	0.21	1.00
		SDS					
C-(38–60)	III + IV	Aqueous	0.03	0.22	0.33	0.42	RC ^e
		TFE	0.43	0.14	0.26	0.16	1.01
		SDS	0.05	0.22	0.28	0.45	1.00
D-(38–72)	III + IV + V	Aqueous	0.00	0.19	0.36	0.45	RC ^e
		TFE	0.31	0.11	0.26	0.31	1.00
		SDS	0.00	0.35	0.25	0.41	1.01
E-(17–72)	IV + V	Aqueous	0.00	0.19	0.36	0.45	1.00
		TFE	0.08	0.22	0.31	0.39	1.01
		SDS	0.00	0.14	0.51	0.36	1.00
F-(57–86)	V + VI	Aqueous	0.00	0.07	0.42	0.50	RC ^e
		TFE	0.14	0.18	0.27	0.40	1.00
		SDS	0.00	0.28	0.33	0.39	1.00

^a The abbreviations used are: H, α helix; B, β -sheet; T, β -turn; O, other structures.

^b Aqueous solvent with 20 mM phosphate buffer, pH 7.

^c Solvent composed of 80% trifluoroethanol and 20% water.

^d Sodium dodecyl sulfate 1–2 mM. The percentage of secondary structures for each peptide was determined from the CD data analysis according to the method of Manavalan and Johnson (33).

^e Random coil.

aqueous and TFE buffer (Table I). Therefore, there is a good correlation between TAT Z2 and TAT BR in region II.

Region IV was described as an extended structure upon binding on TAR (26). However, a peptide corresponding to region IV cannot adopt this structure in aqueous buffer or in TFE without TAR (26). Region IV is described as an extended structure in TAT Z2 (30) and has a sequence rather conserved in the HIV-1 isolates (Fig. 1). To evaluate the influence of regions III and V, three peptides were synthesized, peptide 38–60 (regions III + IV), peptide 38–72 (regions III + IV + V), and peptide 47–72 (regions IV + V). As explained above, peptides 38–60 and 38–72 are random coils in aqueous buffer. Peptide 47–72 (Fig. 2E) satisfies our criterion for a static structure, although a mixture of static and dynamic structure may be present here. The CD data correlate with the two-dimensional NMR data only with peptide 38–72 in SDS, which has a percentage of extended structure consistent with the size of region IV (Table I). Peptides 38–60 and 47–72 have a lower percentage of extended structure in SDS or TFE (Table I). The

CD data suggest that region IV needs at least regions III and V at the N- and C-terminal extremities to adopt an extended structure. Therefore, the influence of regions III and V is determinant for the structure of region IV. This observation is confirmed by a two-dimensional NMR study that described region IV included in a chimerical protein as a stable α -helix (29), while a peptide corresponding strictly to region IV is unable to adopt an α -helix structure in TFE or aqueous buffer (26). A β -turn is observed in the region VI of TAT Z2. Peptide 57–86 satisfies our criterion for a random coil in aqueous solution. In TFE and SDS, the high percentage of β -turn observed (Table I) is consistent with the presence of this structure in the region VI of TAT Bru.

No particular secondary structure is observed for regions III and V in TAT Z2. Fig. 2 shows that peptides with regions III and/or V have changes in their CD spectra as a function of the solvent. These changes can be correlated with the possibility for regions III and V to adopt different secondary structures as a function of the environment. From CD data it can be deduced that regions III and V in TAT Bru have a conformational heterogeneity, whereas regions I–IV are rather rigid. The CD is not decisive for region VI, but the molecular modeling and Chou-Fasman analysis below indicate that region VI is structurally conserved and therefore the change in structure for peptide 57–86 with environment is due to region V. Moreover, the secondary structures predicted from the CD data analysis for the regions I, II, IV, and VI in TAT Bru are similar to the secondary structures observed in the same regions in TAT Z2 (30). We used the Chou-Fasman method (39) on the six TAT sequences selected from Multalin (Fig. 1) to evaluate the potential structural changes due to these mutations. The superimposition of the six plots (Fig. 3) shows that the predictions are dramatically changed as a function of the isolate for the regions III and V but not for the other regions. There is a good correlation between the Chou-Fasman plots and the TAT Z2 structures for regions I, II, and VI that have mainly β -turn structures (Fig. 3C). Nevertheless, region IV is not predicted to be an extended structure (Fig. 3B). There is a good correlation with CD data, which show that region IV as a peptide cannot adopt an extended structure (26). The region IV sequence does not favor an extended structure, and this structure is probably imposed by constraints from regions III and V. It is interesting to note that in regions III and V there are mutations of glycine and proline (Fig. 1), which are residues very important in the structure of a protein (37).

To verify if TAT could adopt a similar three-dimensional structure whatever the sequence, molecular modeling was done on five TAT sequences using the atomic coordinates of TAT Z2 (30) available in the Protein Data Bank (40). The TAT Z2 coordinates were directly used to build the model for TAT BR, TAT MA, TAT JR, TAT OY, and TAT EL when there was a strict sequence homology of three residues or more, whereas only the C α coordinates were used in case of partial homology. For TAT JR, TAT OY, and TAT EL that have a longer sequence than TAT Z2, the C-terminal residue coordinates were arbitrarily set in an extended structure. The measure of the van der

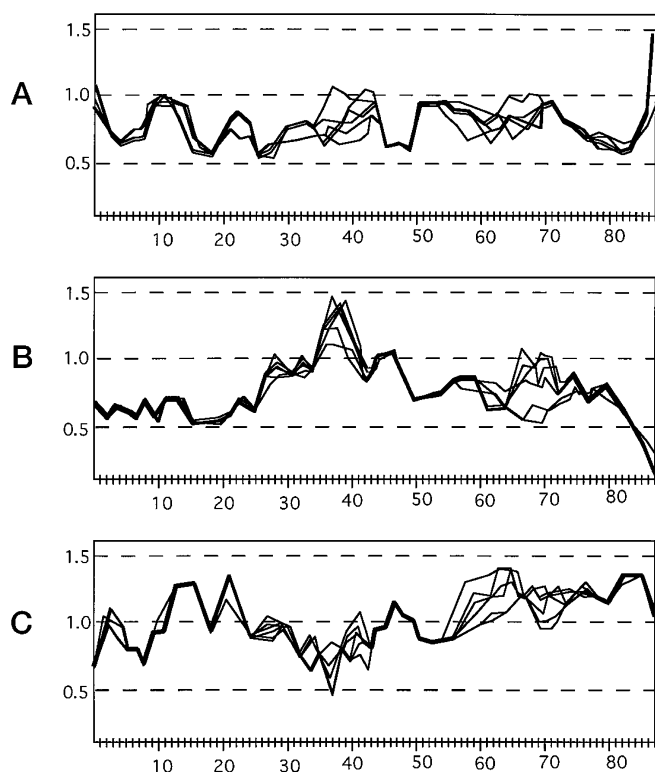


FIG. 3. Secondary structure prediction according to the Chou and Fasman method (39). The six TAT sequences selected from Multalin were analyzed, and the plots were superimposed for each secondary structure. Panel A is the superimposition for α -helix, panel B for β -sheet or extended structure, and panel C for β -turn. The results are similar for the six sequences except for regions III and V where there are significant changes. There is a good correlation with the TAT Z2 two-dimensional NMR structure and CD data except for region IV, which is not predicted to be an extended structure. Region III is predicted to be an α -helix for TAT-BR.

TABLE II
Energies for TAT models

Models were constructed using the structure of the TAT Z2 two-dimensional NMR (30) available in PDB (40) as described in Fig. 4. Energies were minimized with the CVFF force field from MSI Technologies, Inc. (San Diego, CA).

Energies (kcal/mol)	Models					
	TAT Z2	TAT MA	TAT BR	TAT JR	TAT OY	TAT EL
Van der Waals	256	210	183	344	298	238
Coulombic	50	-443	541	-557	-696	-798
Total	307	-233	724	-212	-398	-559

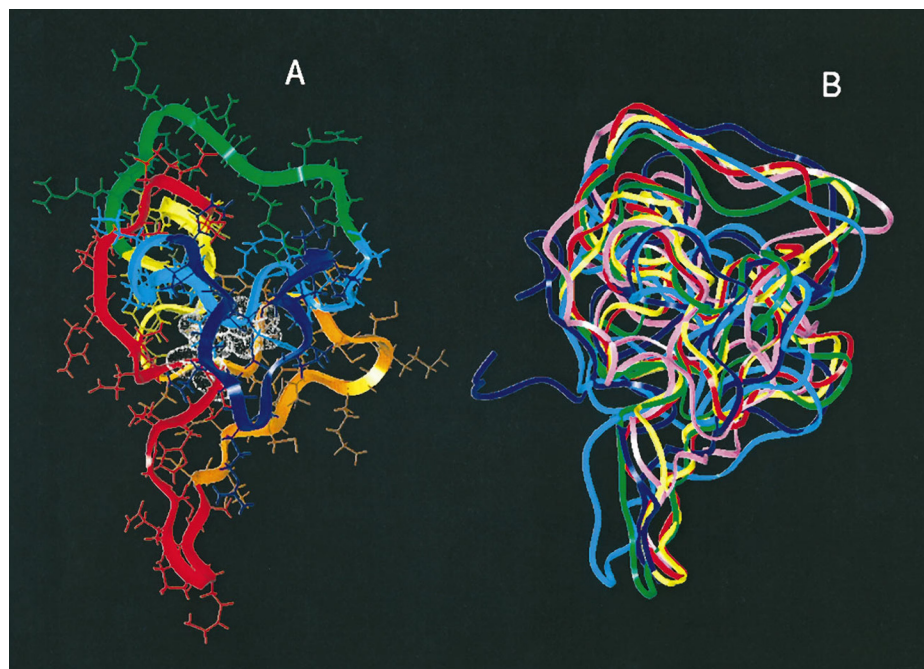


FIG. 4. *Panel A*, TAT BRU structure obtained from molecular modeling using the atomic coordinates of the TAT Z2 two-dimensional NMR structure (30). Region I is colored in *red*, region II in *orange*, region III in *yellow*, region IV in *green*, region V in *light blue*, and region VI in *blue*. The only tryptophan is colored in *white*. *Panel B*, backbone superimposition of the TAT Z2 two-dimensional NMR structure (*yellow*) with the five TAT structures obtained from molecular modeling: TAT MA (*green*), TAT BR (*red*), TAT JR (*blue*), TAT OY (*pink*), and TAT EL (*light blue*). The superimposition was calculated using the backbone C α atomic coordinates of each TAT structure targeted on TAT Z2. Molecular modeling was made with Insight II, Discover, and Homology from MSI Technologies, Inc. (San Diego, CA) running on a R4600 Silicon Graphics workstation. The CVFF force field was used to minimize the structures, and van der Waals energies are shown in Table II. Hydrogens were generated at pH 7. Steepest descent and conjugate gradient were the algorithms used for the minimization. Dynamic was performed at 300 K, and 110 different structures were analyzed in the trajectory.

Waals energy of these models gave values ranging from 10^{10} to 10^{13} kcal/mol, which is not compatible with TAT Z2 in term of minimum energy models (Table II). Using the CVFF force field, the models were minimized with the steepest descent algorithm in a first step. TAT JR, TAT OY, and TAT EL had a supplementary procedure made of a steepest descent minimization associated to a dynamic at 300 K and a constraint of 5 Å imposed between the N- and the C-terminal extremity, since most of the three-dimensional structures available in the Protein Data Bank have their two extremities close together (40). Every model was then submitted to a second step of minimization with a conjugate gradient algorithm associated to a dynamic at 300 K for 1.1 ps (1000 structures obtained per ps). No constraints were used in this second step, and therefore an analysis of 110 structures selected for the trajectory was necessary to determine the structure with the lowest energy obtained from the dynamic. The third and final step was a minimization with the conjugate gradient algorithm. Table II summarizes the van der Waals and Coulombic energies used to compare the models with TAT Z2 and shows that they have similar levels compared with the first models before minimization and dynamics. However, with no constraint all structures will eventually fall apart. The idea here is only to access the relative stability of each sequence and show which regions maintain the TAT Z2 structure, and which regions do not, under these modeling conditions.

The polypeptide chain of the TAT BR structure (Fig. 4 A), as described previously for TAT Z2 (30), has the N terminus sandwiched between region II and region VI, while region IV adopts an extended structure protruding from the core of the protein. Among the seven cysteines located in region II, only cysteines 25 and 34 have distances compatible with the presence of a disulfide bridge. The C-terminal extremity of TAT JR,

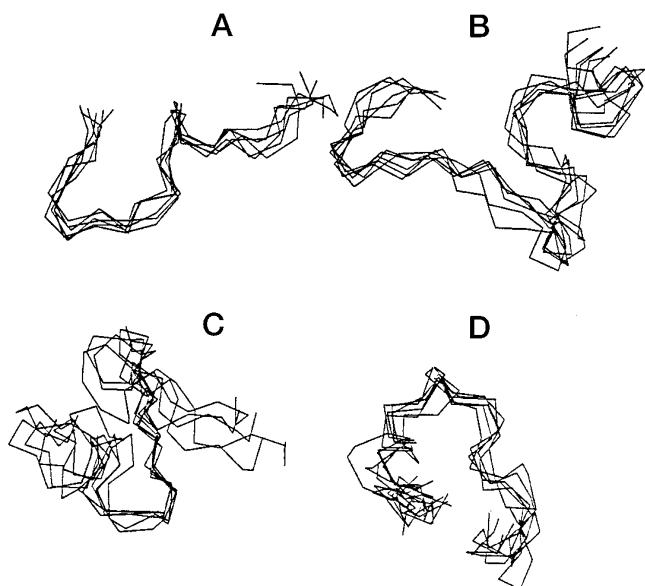


FIG. 5. **Superimposition of TAT regions.** *Panel A*, peptide 2–23 (region I), *panel B*, peptide 13–48 (part of region I, with regions II and III), *panel C*, peptide 38–72 (regions III–V), and *panel D*, peptide 57–86 (regions V and VI). Instead of using the full backbone as represented in Fig. 4B, superimposition was made only with region I (*panel A*), region II (*panel B*), region IV (*panel C*), and region VI (*panel D*). Superimposition with regions III and V (data not shown) did not show structural similarities as observed with regions I–IV and VI.

TAT OY, and TAT EL, which have longer sequences, goes through the loop made by the cysteine-rich region and ends up in a groove made by a part of region I and region III. Table II shows that these three models have a low Coulombic energy,

consistent with stable structures due to intramolecular electrostatic bonds. The superimposition of the six polypeptide chains shows that there are significant structural changes possible among TAT proteins (Fig. 4B). These results confirm that mutations occurring in the *tat* gene of different HIV-1 isolates can be structurally important, and consequently TAT can keep its activity with different possible three-dimensional structures. If the superimposition is made on particular regions of TAT (Fig. 5), structurally related regions can be observed among the six structures, except the two regions adjacent to the basic region (regions III and V), which are impossible to superimpose. Fig. 5C shows that region IV has a structure conserved in the six TAT structures. This result confirms the importance to have region IV in extended structure for the TAT activity.

There is a good correlation between molecular modeling and the CD experiments. Region V has been shown to adopt different conformations in different solvents and when binding on TAR (26). CD data show that regions III and V are affected by solvent; molecular modeling shows that regions III and V have a relatively unstable structure. This is consistent with region III and region V having the highest conformational heterogeneity compared with the other TAT regions. This structural plasticity could be very useful for fitting region IV into the major groove of TAR. The propensity of region III to form an α -helix in our peptides is probably not possible in the full TAT protein. Regions III and V could adapt their structure in their target, in contrast to region IV that must fit into the major groove of TAR and keep its extended structure.

One of the aims of this study was to have a better insight into the TAT structure to be able to do a ligand design. We show here that TAT probably adopts different three-dimensional structures as a function of the HIV-1 isolates. In contrast, region IV always has the same structure. Nevertheless, a ligand that would bind only on region IV should be inefficient because this region has a nanomolar affinity for TAR (10), and it would be very difficult to produce a ligand with a better affinity. Another approach should be to design a molecule as an allosteric inhibitor, binding together region IV and a different region of TAT. This would reduce dramatically the TAT activity in preventing its conformational heterogeneity and should avoid the problem of competition with TAR. From our study, it appears that the conformational heterogeneity of regions III and V would render design of a ligand for these regions unsuccessful. Furthermore, one surface of TAT where region III protrudes and a part of region II changes dramatically with the TAT proteins that have a long C terminus. Although more data should be collected, specially data on structural changes on the full protein, we can propose that a molecule binding together region IV with region I or VI should be the most appropriate. Fluorescence is proposed to evaluate the distance between regions IV and I or VI as a function of TAT structural changes, since there is one tryptophan in TAT located in the middle of the three-dimensional structure (30).

Acknowledgments—We are greatly indebted to Professor W. Curtis Johnson, Jr. for his comments about our data and to Dr. Eric Vives for

the generous gift of TAT peptides. We thank Nicole Zylber for technical assistance and Drs. Arenzana-Seisdedos and Daniel Gautheret for fruitful discussions.

REFERENCES

1. Arya, K. S., Guo, C., Josephs, S. F. & Wong-Staal, F. (1985) *Science* **229**, 69–73
2. Cullen, B. R. (1990) *Cell* **63**, 655–657
3. Muesing, M. A., Smith, D. H. & Capon, D. J. (1987) *Cell* **48**, 691–701
4. Braddock, M., Chambers, A., Wilson, W., Esnouf, M. P., Adams, S. E., Kingsman, A. J. & Kingsman, S. M. (1989) *Cell* **58**, 269–279
5. 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
6. Roy, S., Dellling, U., Chen, C.-H., Rosen, C. A. & Sonenberg, N. (1990) *Genes Dev.* **4**, 1365–1373
7. Marciniak, R. A., Calnan, B. J., Frankel, A. D. & Sharp, P. A. (1990) *Cell* **63**, 791–802
8. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A. & Crothers, D. M. (1990) *Science* **249**, 1281–1285
9. Cordingley, M. G., LaFemina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., Graham, D. J., Nguyen, T. M., LeGrow, K., Gotlib, L., Schlabach, A. J. & Colonno, R. J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8985–8989
10. Calnan, J. B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991) *Genes Dev.* **5**, 210–210
11. Weeks, K. M. & Crothers, D. M. (1991) *Cell* **66**, 577–588
12. Calnan, J. B., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991) *Science* **252**, 1167–1171
13. Westendorp, M. O., Shatrov, V. A., Schulze-Osthoff, K., Frank, R., Kraft, M., Los, M., Krammer, P. H., Dröge, W. & Lehmann, V. (1995) *EMBO J.* **14**, 546–554
14. Ensoli, B., Gendelman, R., Markham, P., Fiorelli, V., Colombini, S., Raffeld, M., Cafaro, A., Chang, H. K., Brady, J. N. & Gallo, R. C. (1994) *Nature* **371**, 674–680
15. Howcroft, T. K., Strebel, K., Martin, M. A. & Singer, D. S. (1993) *Science* **260**, 1320–1322
16. Li, C. J., Friedman, D. J., Wang, C., Metelev, V. & Pardee, A. B. (1995) *Science* **268**, 429–431
17. Westendorp, M. O., Frank, R., Ochsenbauer, C., Stricker, K., Dhein, J., Walczak, H., Debatin, K. M. & Krammer, P. H. (1995) *Nature* **375**, 497–500
18. Jones, A. K. & Peterlin, B. M. (1994) *Annu. Rev. Biochem.* **63**, 717–743
19. Cullen, B. R. (1986) *Cell* **46**, 973–982
20. Kuppuswamy, M., Subramania, T. & Chinnadurai, G. (1989) *Nucleic Acids Res.* **17**, 3551–3561
21. Rappaport, J., Lee, S. J., Khalili, K. & Wong-Staal, F. (1989) *New Biol.* **1**, 101–110
22. Loret, E. P., Vives, E., Ho, P. S., Rochat, H., Van Rietschoten, J. & Johnson, W. C. (1991) *Biochemistry* **30**, 6013–6023
23. Frankel, A. D., Bredt, D. S. & Pabo, C. O. (1988) *Science* **240**, 70–73
24. Koken, S. E. C., Greijer, A. E., Verhoef, K., van Wamel, J., Bukrinskaya, A. G. & Berkhout, B. (1994) *J. Biol. Chem.* **269**, 8366–8375
25. Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. G., Gait, M. J. & Karn, J. (1993) *J. Mol. Biol.* **230**, 90–110
26. Loret, E. P., Georgel, P., Johnson, W. C., Jr. & Ho, P. S. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9734–9738
27. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992) *Science* **257**, 76–80
28. Kamine, J., Loewenstein, P. & Green, M. (1991) *Virology* **182**, 570–577
29. Mujeeb, A., Bishop, K., Peterlin, B. M., Turck, C., Parslow, T. G. & James, T. L. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 8248–8252
30. Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R. & Rosh, P. (1995) *J. Mol. Biol.* **247**, 529–535
31. Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995) *Science* **270**, 1200–1202
32. 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
33. Manavalan, P. & Johnson, W. C., Jr. (1987) *Anal. Biochem.* **167**, 76–85
34. Sanger, F. & Thompson, E. O. P. (1963) *Biochim. Biophys. Acta* **71**, 468–475
35. Corpet, F. (1989) *Nucleic Acids Res.* **22**, 10881–10890
36. Waterhous, D. V. & Johnson, W. C. (1994) *Biochemistry* **33**, 2121–2128
37. Mathews, C. K. & van Holde, K. E. (1990) in *Biochemistry* (Weisberg, S., ed) pp. 176–179, Benjamin/Cummings Publishing Co., Redwood City, CA
38. Johnson, W. C., Jr. (1985) *Methods Biochem. Anal.* **31**, 61–163
39. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251–276
40. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, O. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542