The Solution Structure of HIV-1 Tat Protein

Markus Boehm, Heinrich Sticht, Gabi Seidel, and Paul Rösch

Lehrstuhl für Struktur und Chemie der Biopolymere, Universität Bayreuth, D-95440 Bayreuth

Phone: ++49-921-553545; Fax: ++49-921-553544; Email: Markus.Boehm@uni-bayreuth.de

Abstract

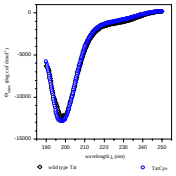
The trans-activator protein (Tat) controls the expression of early genes of the human immunodeficiency virus type 1 (HIV-1). It exerts its effects by binding to the Tat-responsive region (TAR), an RNA stem-loop structure located at the 5'-end of all lentiviral RNAs. While the structure of the RNA element has been determined in solution in its free and ligand bound forms (Aboul-ela *et al.*, 1996, 1995), a high resolution structure of wild type HIV-1 Tat could never be determined, although several structural features of this protein have been elucidated (Bayer *et al.*, 1995).
Here, we describe the determination of the structure of mutant HIV-1 Tat protein by NMR-spectroscopy and simulated annealing calculations. This variant of the HIV-1 Tat protein is indistinguishable from the wild type protein with respect to RNA-binding as judged from CD-spectroscopy and UV-melting studies. The protein exhibits a rigid structure, mainly stabilized by turns in most parts, but a higher degree of conformational variability in secondary structure for the cysteine rich and the basic domain. The NH_2-terminus is positioned in the center of the molecule, interacting with the core, the basic, and the glutamine rich domain. An RGD-stretch at the COOH-terminus of the protein is solvent exposed, enabling the interaction with other proteins such as the integrin receptor. The present structure of HIV-1 Tat is compared with structural data obtained for several Tat peptides, and indications for structural rearrangements upon RNA binding are discussed.

Introduction

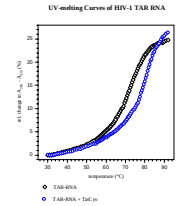
The 86 amino acid Tat protein of HIV-1 is essential for viral replication and only short transcripts of viral RNA are found in its absence, whereas full length transcripts of viral RNA are produced and the transcription rate is increased in its presence. Tat is located mainly in the nucleus and nucleolus of infected cells, and acts through binding to the Tat-responsive region (TAR), an RNA stem-loop structure located at the 5'-end of all HIV transcripts. The solution structure of TAR is known both in its ligand bound and free form.
Tat has pleiotropic effects on the host cell as well as on gene expression in uninfected cells. Genetic analysis has revealed the existence of several functionally important stretches of amino acids within the protein, referred to as sequence domains. Residues 1 to 19 comprise the acidic and Pro-rich domain, the next 12 residues containing five cysteines define the Cys-rich domain, residues 49 to 57 represent the basic domain that is essential for RNA binding and nuclear/nucleolar targeting of the protein, amino acids 60 to 72 form the Gln- rich domain, and residues 32 to 47 make up the core domain. Both, Gln-rich and core domain contribute to TAR binding specificity (Churcher *et al.*, 1993). Residues 73 to 86 are encoded by the second exon and are not necessary for intracellular HIV-1 Tat function *in vivo*. There is, however, some evidence that these residues are important for the uptake of extracellular HIV-1 Tat, and a conserved RGD motif, amino acids 78 to 80, may mediate this function. General structural features of HIV-1 Tat are known, but the wealth of biochemical data accessible for this protein can only be explained on the basis of a highly resolved structure.

Methods

In order to be able to obtain highly concentrated samples of HIV-1 Tat protein in solution, we devised a mutant protein in which all seven cysteine residues were replaced by a cassette of alanine and serine residues. The TatCys$^-$ protein was expressed in *E. coli BL21/DE3* and purified to homogeneity by a two-step metal affinity chromatography procedure followed by C_{18}-reversed phase HPLC.
In order to test whether the constructed variant of HIV-1 Tat was behaving similar to the wild type protein, we employed several functional and structural assays, such as CD-spectroscopy and UV-melting.



CD-Spectra were used to show that the wild type protein and the cysteine free mutant of Tat were identical with respect to their secondary structure. The protein lacks regular secondary structural elements such as α-helices or β-sheets. However, a certain percentage of turn elements can be deduced from the dichroism of the samples.



In UV-melting experiments the mutant protein also proved to be indistinguishable from the wild type protein. The midpoint of melting, as judged by the relative increase in A_{269} - A_{320}, of TAR-RNA was increased by about 9 °C upon the addition of the mutant protein. This effect could not be seen after the addition of arginine at sevenfold excess (Metzger *et al.*, 1996), and can be attributed to the specific binding of TatCys- protein to its *in vivo* RNA target TAR.

From the data obtained we concluded that TatCys$^-$ was similar to the wild type protein with respect to structure and functionality and, therefore, went on to solve the three dimensional structure of the protein with the help of NMR-spectroscopy. As there prove to be spectral overlap simililarly as seen in the spectra of the wild type protein before (Bayer *et al.*, 1996), we introduced ^{15}N nitrogen in order to allow for the complete assignment of the spin systems.

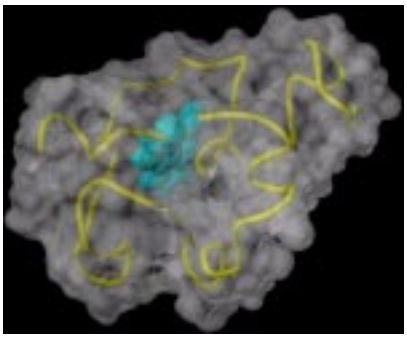
The following spectra were recorded:
homonuclear $^1H,^1H$-Clean-TOCSY (MLEV-17 spin lock) spectra, mixing times 40 and 80 ms
$^1H,^1H$-NOESY spectra, mixing times 100 and 200 ms
$^1H,^1H$-DQF-COSY spectra
heteronuclear $^1H,^{15}N$-HSQC spectra
$^1H,^1H,^{15}N$-HMQC-TOCSY (MLEV-17 spin lock) spectra, mixing time 80 ms
$^1H,^1H,^{15}N$-HMQC-NOESY spectra, mixing time 200 ms

References

Aboul-ela, F. *et al.*, J. Mol. Biol. 253: 313 (1995)
Aboul-ela, F. *et al.*, Nucleic Acids Res. 24: 3974 (1996)
Bayer, P. *et al.*, J. Mol. Biol. 247: 529 (1995)
Brünger, A.T., X-PLOR 3.851 online (Yale University, New Haven, CT., 1996)
Churcher, M.J., *et al.*, J. Mol. Biol. 230: 90 (1993)
Garret, D.S., et al., J. Magn. Res. Series B 104: 99 (1994)
Kuszewski J., *et al.*, Prot. Sci. 5: 1067 (1996)
Metzger, A. *et al.*, FEBS Lett. 384: 255 (1996)
Varani, G., *et al.*, Prog. Nucl. Magn. Res. Spec. 29: 51 (1996)
Wei, P., *et al.*, Cell 92: 451 (1998)

A total of 1721 NOE cross peaks and 23 $^3J_{HNA}$ coupling constants could be identified in the NOESY and COSY spectra, and structures were calculated with a modified X-PLOR 3.851 *ab initio* simulated annealing protocol (Brünger, 1996) using Ramachandran database refinement (Kuszewski *et al.*, 1996), and direct refinement against $^3J_{HNA}$ coupling constants (Garret *et al.*, 1994). After that, the wild type sequence was reestablished *in silico* and 500 steps of energy minimization were performed, including an attractive Lenard-Jones potential (Varani *et al.*, 1996).
The reintroduction of the cysteine residues served the purpose to reveal whether the introduction of alanines and serines had brought about structural features which were different from the wild type protein. In that case, the placement of a cysteine residue with its larger side chain would cause structural violations, for example by steric hindrance. The following energy minimization would then result in a structure which partly violated the previously observed distance restraints.
However, we could not observe any structural violations upon the introduction of the cysteines, and therefore conclude on the determined structure to be similar to the structure of wild type HIV-1 Tat protein.

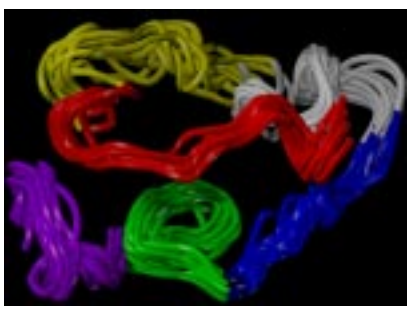
The structure of HIV-1 Tat protein



The picture shows the protein backbone in tube representation with its Connolly surface in grey. The NH_2-terminus of the protein is positioned to the left, and the COOH-terminus to the upper right corner of the picture. The central position of Trp11 is highlighted.
The protein shows no elements of regular secondary structure besides bends (B) and turn (T) elements which are found for amino acids 3-8 (B), 15-22 (B+T), 26-33 (B), 36/37 (B), 39-47 (B+T), 57/58 (T), 61/62 (B), 66-68 (B), 70/71 (T), and 75-80 (B+T). Structural statistics are given in the following table:

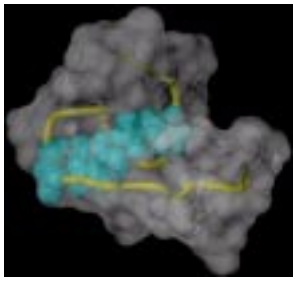
Experimental restraints for the final structure calculation	
Total number of internesidual NOEs	939
sequential NOEs |i - j| = 1	409
medium-range NOEs |i - j| = 2, 3, 4, 5	430
long-range NOEs |i- j| > 5	100
Total number of intaresidual NOEs	782
Dihedral angle restraints	23
Molecular dynamics statistics	
(Average energy (kJ/mol))	
E_{tot}	1595.76
$E_{covalent}$	78.87
E_{vdw}	154.25
E_{elec}	1094.21
$E_{restraints}$	137.94
E_{vdw}	-149.81
E_{elec}	279.46
E_{rms}	0.84
R.m.s.d. from ideal distances (nm)	
NOE	0.01
Bond length	0.03
R.m.s.d. from ideal angles (°)	
Bond angles	1.63
Improper angles	0.96
Dihedrals	0.19
R.m.s.d. of 10 calculated structures (nm)	
Backbone (1-86)	0.16
Heavy atoms (1-86)	0.20

Overlay of the 10 least energy structures after minimization

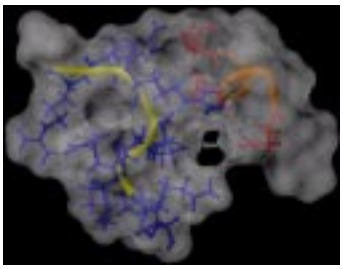


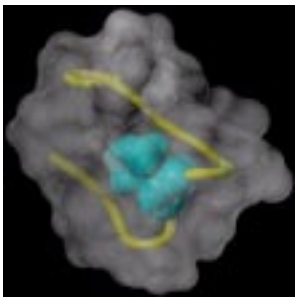
Sequence domains of HIV-1 Tat are colored as follows:
Pro-rich domain (1-19): red; Cys-rich domain (20-31): yellow; core domain (32-48): white; basic domain (49-58): blue; Gln-rich domain (59-72): green; 2nd exon domain (73-86): violet

To the right, the core domain and the interaction of its two central residues Phe38 and Tyr47 which causes the formation of a hydrophobic core is shown. These two residues are thought to rearrange upon the formation of the specific Tat-TAR complex which causes the disruption of the hydrophobic core of the protein.



The lower drawing depicts the interaction of arginines from the basic domain with acidic residues from the NH_2-terminus. This contributes to the stability of the protein and fixes the basic domain relatively to the rest of the protein.





Ile69 is the structural anchor of the Gln-rich domain, as demonstrated by its central position in the domain. The backbone wraps around this residue in three extended stretches from Pro58 to Gly61, Gly62 to Gln66, and Lys71 to Ser75. These are connected by turn regions at Gly61/Gly62 and Asp67 to Pro70.

Free molecular dynamics simulations uncover internal flexibility in HIV-1 Tat protein

The stability of the obtained wild type structure has been tested in free molecular dynamics simulations with a simulation time of 350 ps. The results show a considerable degree of flexibility, albeit the preservation of the overall structure. The conformational variability can be attributed mainly to residues in the Cys-rich and Gln-rich domains. This correlates with r.m.s.d. values >0.2 nm observed in the initial structure calculations for residues Thr20 to Cys27 and, therefore, represents an intrinsic feature of Tat protein. Possibly, binding of cyclin T by HIV-1 Tat (Wei et al., 1998) would impose conformational restraints on this particular stretch of amino acids. Furthermore, a variety of longrange H-bonds, e.g. between residues Gln35, Arg53, and Asp67 and salt bridges between residues of the basic domain and the NH_2-terminus contribute to the stability of the tertiary structure of HIV-1 Tat.