Eur Biophys J (2004) 33: 344–351
DOI 10.1007/s00249-003-0358-z

ARTICLE

Sergio Pantano · Mudit Tyagi · Mauro Giacca
Paolo Carloni

Molecular dynamics simulations on HIV-1 Tat

Received: 1 November 2002 / Accepted: 6 February 2003 / Published online: 8 November 2003

Abstract Molecular dynamics simulations are used to investigate dynamics and intramolecular interactions of the HIV-1 transactivator (Tat) in aqueous solution. The calculations are based on the AMBER force field with particle mesh Ewald treatment for long-range electrostatics. The Tat structure exhibits a large flexibility, consistent with its absence of secondary structure elements. From an analysis of the correlation matrix and of electrostatic interactions we suggest that segments expressed by the two exons (amino acids 1–72 and 73–86, respectively) exhibit rather separated dynamic and energetic properties. We also identify intramolecular interactions of importance for structure stabilization. In particular, significant electrostatic interactions are recognized between the N-terminus and the basic domain of the protein, consistent with site-directed mutagenesis performed in this work.

Keywords Correlation matrix · Electrostatic interactions · HIV-1 transactivator · Molecular dynamics · Site-directed mutagenesis

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00249-003-0358-z

S. Pantano · P. Carloni (✉)
International School for Advanced Studies (ISAS) and INFM, DEMOCRITOS Modeling Center for Research in Atomistic Simulation, Via Beirut 2–4, 34014 Trieste, Italy
E-mail: carloni@sissa.it
Tel.: +39-040-3787407
Fax: +39-040-3787528

M. Tyagi · M. Giacca
Molecular Medicine Laboratory, International Center for Genetic Engineering and Biotechnology, Padriciano 99, 34012 Trieste, Italy

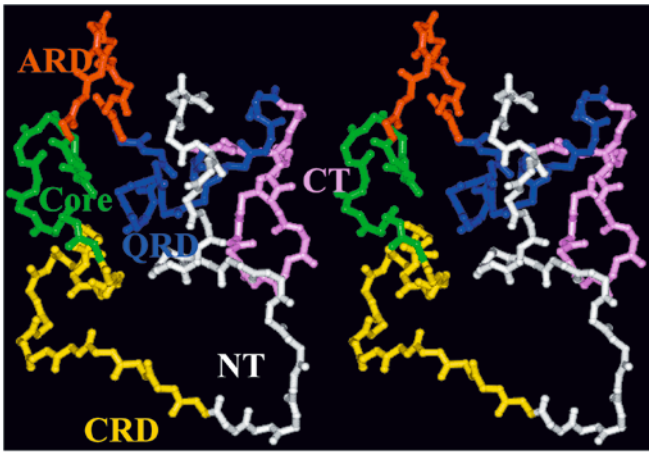
M. Tyagi · M. Giacca
Scuola Normale Superiore, Piazza dei Cavalieri 7, 56100 Pisa, Italy

S. Pantano
IMASL, CONICET, National University of San Luis, Ejercito de los Andes 950, CP 5700, San Luis, Argentina

Introduction

The transactivator of the human immunodeficiency virus type 1 (HIV-1 Tat) is a small protein essential for viral replication (Sodroski et al. 1985). In the host cells, Tat activates transcription of the viral genome by recruiting the human positive transcription elongation factor PTEFb to the C-terminal domain of the RNA polymerase II (see Karn 1999 for a review), specifically binding the nascent viral mRNA transactivation-responsive (TAR) region (Berkhout and Jeang 1989; Berkhout et al. 1989).

Although Tat does not present secondary structure elements, several regions of the protein can be defined on the basis of the type of amino acids (aa) in the primary structure and by homology with other lentiviral transactivators (Dorn et al. 1990; Carroll et al. 1991; Derse et al. 1991). Scheme 1 provides a representation of the spatial disposition of the protein regions and its correspondence with the primary sequence: (1) the N-teminal, rich in negative residues and prolines (NT, aa 1–21, light gray); (2) the cysteine-rich domain (CRD, aa 22–37, yellow); (3) the core region (aa 38–48, green), rich in hydrophobic residues; this region is crucial for Tat transactivation (Kuppuswamy et al. 1989; Churcher et al. 1993) and is highly conserved among other lentiviruses; (4) the arginine-rich domain (ARD, aa 49–57, red), which contains the nuclear localization signal of the protein (Friedler et al. 2000) and specifically interacts with a uridine-rich bulge motif in the RNA TAR (Berkhout et al. 1989); (5) the glutamine-rich domain (QRD, aa 58–72, blue), which has been proposed to provide structural stabilization to the protein (Kuppuswamy et al. 1989); (6) the C-terminal region (CT, aa 73–86, violet), which appears to be dispensable for



MDPVDPNIEP[10]WNHPGSQPKT[20]ACNRCHCKKC[30]CYHCQV
CFIK[40]KGLGISYGRK[50]KRRQRRRPSQ[60]GGQTHQDPIP[70]KQ
PSSQPRGD[80]PTGPKE[86]

Scheme 1

transactivation (Frankel et al. 1989; Verhoef et al. 1998) and has a relatively more variable sequence among isolates than the other domains. Regions (1) to (5) are expressed by a first exon on the viral genome, while region (6), the C-terminal, is expressed by a second exon.

Here we used molecular dynamics (MD) simulations to investigate the dynamics and intramolecular interactions among the different regions of the HV1Z2 Tat protein (Bayer et al. 1995). Our calculations establish a correlation between some of the above-described regions and motional domains. Furthermore, they allow to identify a previously unrecognized interaction between Asp2 (from the NT domain) and Lys50–Arg53 (from ARD). This may be consistent with mutagenesis experiments performed here on the variant HXB2, which is expected to be structurally similar to the Z2 variant (Pantano et al. 2002). Finally, our results suggest that the structure of aa 1–72 is not modified by the presence of CT, consistent with a variety of biological data (Vives et al. 1994; Brown et al. 1998; Verhoef et al. 1998; Weissman et al. 1998).

Methods

Molecular dynamics calculations

Our calculations are based on the NMR spectroscopy determined structure of HV1Z2 Tat (human immunodeficiency virus type 1, Zaire isolate 2, aa 1–86), PDB entry 1TIV (Bayer et al. 1995). Ten structures have been deposited. The first structure, which reproduces the NOE constraints most accurately (Bayer et al. 1995), was chosen as the initial structure. The protonation state of the whole protein was kept as found in the PDB reported structure, e.g. His and Cys residues neutral, Glu and Asp negative and Arg and Lys positive. The model was immersed in a box of ~70 Å×55 Å×45 Å edges containing ~5000 water molecules. Eleven Cl^- counterions (Smith and Deng 1994) were added so as to neutralize the highly positive charge of the protein (+11). The anions were located at least 5 Å away from any protein atom in random positions. It should be noted that during the dynamics these ions experienced very large mobility. Rectangular periodic boundary conditions were used.

The particle mesh Ewald summation method (Darden and York 1993) was used for long-range electrostatic interactions. A 10 Å cutoff was adopted for short-range non-bonded interactions. A permittivity equal to 1 was assumed.

All chemical bonds were constrained using the SHAKE algorithm (Ryckaert et al. 1977). The equations of motion were integrated using a time step of 1 fs. Temperature and pressure were kept constant at 300 K and 1 atm by coupling the system to external baths (Berendsen et al. 1984) with coupling constants of $\tau_t = 0.05$ and $\tau_p = 0.5$ ps, respectively.

For the protein and the counterions, the AMBER (Cornell et al. 1995) force field parameters were used. For water, the TIP3P model (Jorgensen et al. 1983) was chosen.

The system underwent energy minimization. Then, solvent and counterions were allowed to move for 0.02 ns. Subsequently, the system was gently heated up from 0 K to 300 K in 0.12 ns. Finally, 3 ns of MD simulations at 300 K and 1 atm were performed. MD calculations were performed with the SANDER module of AMBER 5.0 (Case et al. 1995) suite of programs.

The following properties were calculated:

1. The root mean square displacements (RMSD) (Leach 2001) of the backbone atoms as a function of time:

$$\mathrm{RMSD}_{(t)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\chi_i^{\mathrm{ref}} - \chi_{i(t)}\right|^2} \tag{1}$$

where N is the number of atoms on which the RMSD is calculated and χ_i^{ref} and $\chi_{i(t)}$ are the position vectors of the reference structure (which corresponds to the NMR-derived structure after energy minimization in the presence of solvent and counterions) and the structure at time t.

2. The radius of gyration (Daune 1999) as a function of time was calculated on the positions of the Cα atoms as:

$$G_{\mathrm{R}(t)} = \frac{1}{(N+1)}\sqrt{\sum_{j=1}^{N}\sum_{i=1}^{N}\left|\chi_{i(t)} - \chi_{j(t)}\right|^2} \tag{2}$$

where N is the number of atoms.

3. The correlation matrix calculated on the instantaneous position of the Cα atoms, provides useful information about the interrelation of the motion of non-closely linked parts of the proteins during the dynamics (Radkiewicz and Brooks 2000). The correlation matrix reads:

$$\frac{\left\langle\left[\chi_{i(t)} - \left\langle\chi_{i(t)}\right\rangle\right]\left[\chi_{j(t)} - \left\langle\chi_{j(t)}\right\rangle\right]^T\right\rangle}{\sqrt{\left\langle\chi_{i(t)} - \left\langle\chi_{i(t)}\right\rangle\right\rangle^2\left\langle\left[\chi_{j(t)} - \left\langle\chi_{j(t)}\right\rangle\right]^T\right\rangle^2}} = \frac{C_{ij}}{N_{ij}} \tag{3}$$

where the $\langle\ \rangle$ brackets indicate an MD average and $\chi_{i(t)}$, $\chi_{j(t)}$ are the position vectors of the Cα atoms of two residues i and j, and N_{ij} is the normalization factor. The normalized matrix ranges from −1.0 for completely anticorrelated motions to 1.0 for completely correlated motions.

4. Large-scale displacements (Garcia 1992; Amadei et al. 1993). The eigenvectors of the above-defined C_{ij} matrix represent the large-scale motions of the protein. It results that just a few eigenvectors already cover most of the large-scale motion in proteins (Amadei et al. 1993). Translational and rotational motions were eliminated by performing a RMSD fit on the Cα atoms between the MD snapshots and the first MD conformer. The comparison among the eigenvalues' spectrum in the ranges $t = 0$–2.5 ns and $t = 0$–3.0 ns (Fig. S1, Supplementary material) suggests that the timescale investigated is long enough for a proper dynamical description, in agreement with previous findings (Radkiewicz and Brooks 2000).

The projections or contributions of the eigenvectors δ_n to the real space motion of the n^{th} protein domain (CD_n) may be obtained as a sum of dot products between δ_n and the displacement vectors:

$$v_i = \chi_{i(t)} - \chi_i^{ref} \quad (4)$$

of the Cα atoms belonging to the domain:

$$CD_n = \frac{1}{N_{AD}} \sum_{i}^{N_{AD}} v_i \cdot \delta_n \quad (5)$$

where N_{AD} is the number of atoms belonging to the corresponding domain.

5. Energy maps. The electrostatic inter-residue energy matrix was calculated as:

$$E_{i-j} = NBE_{R_i-R_j} + BE_{R_i-R_j} \quad (6)$$

where NBE and NB correspond to non-bonded and bonded (1–4) electrostatic interactions respectively for all the residues (R_i) in the protein. Energy maps were calculated on corresponding minimized MD snapshots using a cutoff radius of 50 Å, which is larger that the protein diameter. Test calculations with larger cutoffs (60 and 70 Å) provided essentially the same results. Because the energies are calculated based on an effective force field approximation, they may not be quantitatively accurate; here, they are used to identify, at the qualitative level, the largest interactions between domains of the protein.

6. Hydrogen bonds between a donor (X–H) and an acceptor (Y) were assumed to take place if: (1) $d(X–Y) < 3.5$ Å; (2) $\angle(X–H–Y) > 120°$; (3) the existence time during the dynamics is $> 20\%$. MD average distances were calculated over the last 1.5 ns.

7. The number of water molecules interacting with a given protein atom was evaluated as the integral of the radial distribution function ($g_{(r)}$) (Allen and Tildesley 1987) between the water oxygen atoms and this atom:

$$N = \int_0^{R_{cut}} g_{(r)} dr \quad (7)$$

where R_{cut} has been set at 8 Å. The calculations were carried out for the last 1.5 ns of the dynamics.

8. Hydrophobic contacts between hydrophobic amino acids with $d(Cα–Cα) < 6.5$ Å. MD average distances were calculated over the last 1.5 ns.

Biochemical experiments

E2Q and E2A Tat mutants were obtained starting from plasmid pcDNA3-Tat86, expressing the wt Tat protein of HIV-1, strain HXB2 (86 aa), from the cytomegalovirus (CMV) promoter in the pcDNA3 (Invitrogen, Carlsbad, Calif., USA) vector backbone. Mutations were introduced by the recombinant polymerase chain reaction (PCR) techniques using appropriate primer pairs. Direct DNA sequencing was used to verify the nucleotide sequence of the final constructs.

HIV-1 long-terminal repeat (LTR) transactivation assays were performed by transfection of HL3T1 cells, kindly donated by B. Felber (Felber and Pavlakis 1988), a human epithelial HeLa derivative cell line containing an integrated LTR-chloramphenicol acetyltransferase (CAT) construct. HL3T1 cells were grown in Dulbecco's modified eagle's medium (DMEM) and HAM'S F10 medium, respectively, supplemented with 10% fetal calf serum, 2 mM glutamine and 50 μg/mL gentamicin. Cells were transfected by the standard calcium phosphate procedure (Sambrook et al. 1989). Transfections were adjusted to the same content of transfected plasmid (and of CMV promoter sequences) by addition of the appropriate amounts of pcDNA3. CAT assays were performed 48 h after transfections. Reported transactivation results represent the average values obtained in at least three independent transfections.

Note that a conservative mutation (Glu2 → Asp) occurs on passing from HXB2 to the HV1Z2 variant.

Results

Molecular dynamics calculations

During the dynamics, Tat exhibits a high degree of mobility. Our MD simulation was first checked against the raw experimental NOE constraints, as provided in (Bayer et al. 1995). Our MD trajectory exhibited only 3% more violations than those exhibited in the structures deposited in the PDB (Table S1, Supplementary material), suggesting that our computational setup may be adequate to study the dynamics and energetics of this protein.

The radius of gyration along the simulation remains nearly constant along the time scale explored, with an average value of 12.6 Å. The RMSD fluctuates around an average value of 2.8 Å after ~1.5 ns (Fig. 1a). The correlation matrix provides information on the correlation between the motions of the six domains of the protein (Fig. 2) (Radkiewicz and Brooks 2000). Inspection of the matrix suggests that the protein does not exhibit significant anticorrelated motion and that, in general, the motion is not highly correlated. Indeed, only the QRD and NT exhibit correlated motion with almost all the other parts of the protein. These interdomain correlated motions are ascribed to electrostatic interactions between QRD and NT, QRD and ARD, QRD and CT, and NT and ARD (Table 1). To a less extent, they are due to hydrophobic interactions between NT and CRD, NT and core, NT and QRD, and CRD and QRD (Table 2).

Within the N-terminal region, the motion of aa 1–14 is significantly self-correlated; in contrast, the motion of the last 7 aa is not correlated. The latter are also more exposed to the solvent than aa 1–14. These facts suggest that the NT domain can be divided into two regions displaying clearly different motional behavior: a first part (aa 1–14) showing intradomain correlation motion and relatively low mobility, and a second part (aa 15–21) which is mobile and exhibits a highly uncorrelated motion. This observation leads to a redefinition of the NT-CRD "border" from aa 21–22 to aa 14–15.

CT exhibits correlated motion only with its neighboring region QRD and, less, with ARD owing to electrostatic interactions between the ARD and the terminal Glu86 (Fig. S2, Supplementary material).

Diagonalization of the covariance matrix provides useful information on the large-scale motions of the protein (Amadei et al. 1993). Indeed, the calculated eigenvectors individuate collective displacements.

Figure 3a (inner frame) shows that the three largest eigenvectors span about 90% to the total motion. Thus, we focus here only on these three modes. Figure 3b plots the contribution of the domains to the first eigenvector. The largest motional contributions are provided by the

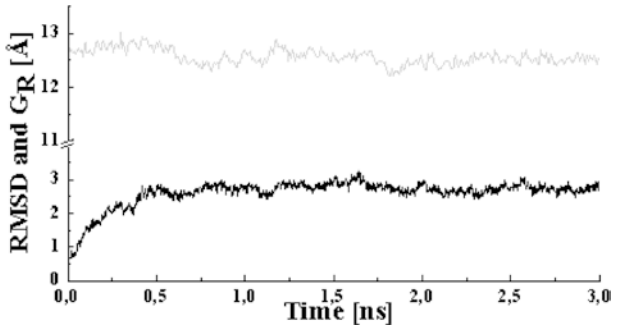


Fig. 1 Root mean square deviation (RMSD, *black line*) and radius of gyration (G_R, *gray line*) plotted as a function of simulation time

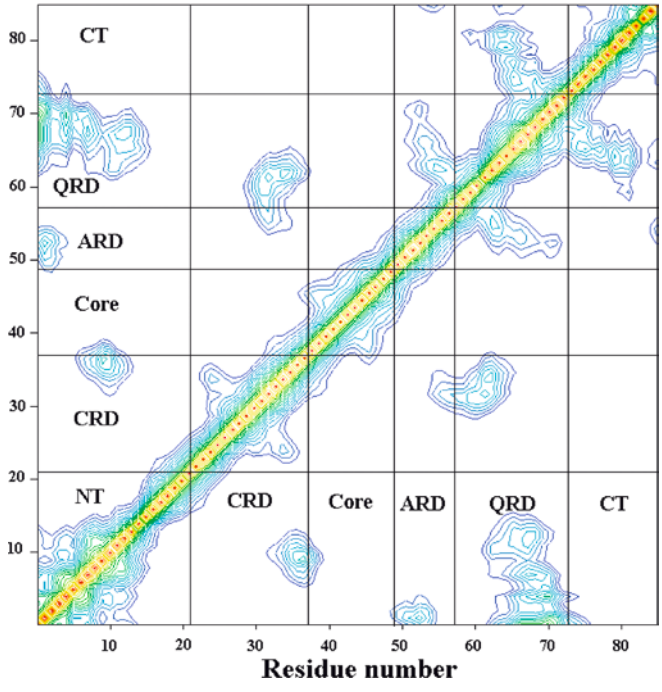


Fig. 2 Correlation matrix of the C_α atoms (Radkiewicz and Brooks 2000). Color code ranges from *blue* (0.3) to *red* (1). The diagonal elements of the matrix represent the self-correlation motion, which are unitary. The off-diagonal elements provide information on correlations between residues which are not directly connected. Anticorrelated motion, represented by negative matrix elements, is absent

CRD (redefined according to the results above). During the first part of the dynamics also the core domain provides an important contribution to the large-scale motions of the protein (Fig. 3b). This may be due to the presence of three Gly residues (42, 44 and 48) the small size of these residues may increase the mobility of the system (Bayer et al. 1995) and allow for an initial small rearrangement of this domain.

In contrast, the smaller contributions to the first eigenvalue arise from NT (redefined as a domain between aa 1–14) and ARD, possibly since, owing to its high charge density, the basic residues tend to minimize the repulsive interaction by staying as far as possible from each other and by being highly solvated, limiting the degree of motional freedom of the backbone. The largest

Table 1 Selected H-bond interactions between residues belonging to different domains. Standard deviations are reported in parentheses (see Methods)

H donor–H acceptor	Av. dist. (Å)	Occupancy (%)
Met1 (N)–Asp67 (O)	2.9 (0.3)	73
Met1 (N)–Ile69 (O)	2.8 (0.2)	95
Met1 (N)–Lys71 (O)	2.8 (0.2)	93
Arg49 (Nη1)–Tyr47 (O)	3.0 (0.3)	34
Lys51(Nζ)–Asp2 (O)	3.1 (0.4)	33
Lys51 (Nζ)–Asp2 (Oδ1-2)	3.0 (0.2)	98
Arg53 (Nη1-2)–Asp2 (Oδ1-2)	3.2 (0.3)	48
Arg53 (Nη1-2)–Asp67 (Oδ1-2)	3.3 (0.5)	67
Arg53 (Nη1)–Gln72 (O)	3.1 (0.3)	42
Arg55 (N)-Arg53 (O)	3.1 (0.4)	44
Arg55 (Nε)–Gln60 (Oε1)	3.1 (0.4)	21
Arg55 (Nη1-2)–Gln60 (Oε1)	3.1 (0.4)	41
Arg56 (N)–Gln64 (O)	3.2 (0.4)	21
Arg56 (Nη1-2)–His33 (Nε2)	3.1 (0.4)	42
Arg56 (Nη1-2)–Gln54 (Oε)	2.9 (0.3)	19
Arg57 (Nη1-2)–Gln60 (Oε)	3.1 (0.3)	40

Table 2 Selected distances between hydrophobic residues belonging to different protein domains (note the correspondence with Fig. 2). MD-averaged Cα–Cα distances calculated over the last 1.5 ns. Standard deviations are reported in parentheses

Residues	Distance (Å)
Pro6–Pro70	6.4 (0.6)
Ile8–Pro68	4.8 (0.5)
Trp11–Val36	4.6 (0.4)
Trp11–Pro68	6.0 (0.7)
Val36–Phe38	4.7 (0.5)
Val36–Pro68	6.1 (0.7)

contributors of the total motion for eigenvectors 2 and 3 are both CRD and CT (Fig. 3c and d).

We now turn our attention to the energetics. Electrostatics, which is expected to be the dominant interaction in this highly charged protein, is here analyzed in terms of a force-field-based electrostatic interaction energy map (Fig. 4). Diagonal and off-diagonal elements provide the intra- and inter-residue electrostatic stabilization. Comparison of the maps associated with the initial and final MD structures (Fig. 4) indicate that the dominant interactions are well conserved during the dynamics despite the high flexibility of the protein. Furthermore, a correspondence might be individuated between the correlated motions (Fig. 2) and electrostatic interactions between the NT and ARD, NT and QRD, and ARD and CT domains (Fig. S2).

The region that presents the most important intra- and interdomain interactions is ARD, as can be expected from its elevated charge density. The repulsion among the positively charged basic residues is reduced by the presence of the aqueous solvent. In particular, Arg52, located in the middle of the ARD, turns out to be the most stabilized residue. The calculations of water radial distribution functions on Lys and Arg confirm this picture. The most hydrated residue in this domain is Arg52, which is found always solvent-exposed. Lys51, Arg53, Arg55 and Arg56 are

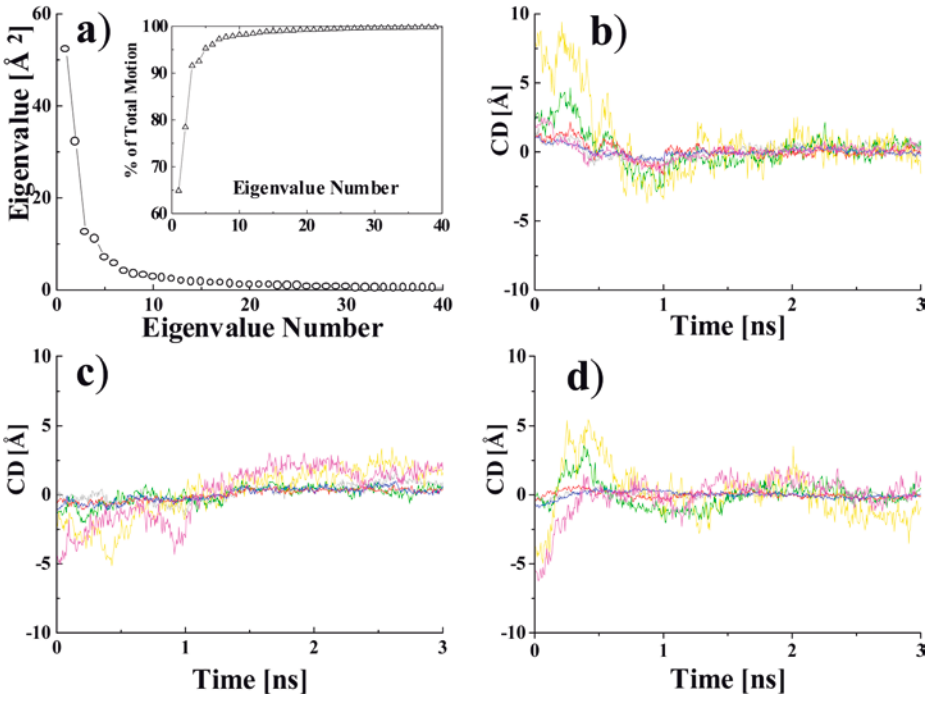


Fig. 3a–d Large-scale dynamics analysis (Garcia 1992; Amadei et al. 1993). (**a**) Eigenvalues and percentage of total motion spanned by the corresponding eigenvectors (*inner frame*). (**b–d**) Projected contributions per domains of the first, second and third essential eigenvectors, respectively. Colors according to Scheme 1

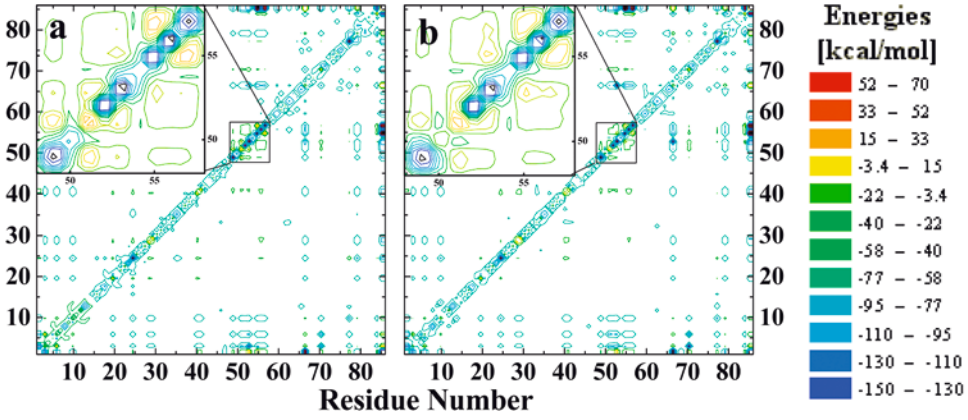


Fig. 4 Electrostatic interaction energy maps per residue at time = 0 (**a**) and at time = 3 ns (**b**). *Inset*: close-up on the interactions among ARD residues. Strong stabilizing intradomain interactions are found between: (1) Asp2 (NT) and Lys51, Arg53 (ARD); (2) Asp67 (QRD) and Arg55, Arg57 (ARD); (3) Glu86 (CT) and Arg55, Arg57 (ARD); (4) Lys71 (QRD) and Asp5 (NT); (5) Met1 (NT) and Asp67 (QRD). The most depleted areas in the map correspond to interactions of the CRD and core domain with themselves and with other domains owing to their low charge content. As the values were obtained with a force-field-based model, they should be interpreted only at the qualitative level

less hydrated (Table 2), as they form strong H-bond interactions (Table 1).

Site-directed mutagenesis experiments

Our calculations suggest that Asp2 (unlike Asp5 and Glu9 that remain solvated) is crucial for NT–ARD binding (Fig. 5). In HVIZ2 Tat Asp2 is the only negative residue within the NT region to interact directly with both Lys51 and Arg53 in the ARD. Thus, mutations at this position, which extinguish the charge, may cause a partial destabilization of the structure, which in turn might correlate with an impaired biological activity (Pantano et al. 2002).

To test this proposal, we have expressed and measured the biological activity of two mutants at position 2, Glu → Ala and Glu → Gln on wt HXB2 Tat, which is expected to be structurally and energetically similar to HV1Z2 (Pantano et al. 2002).

E2A HXB2 Tat turns out to show a dramatically reduced biological activity (Fig. 6). The transactivation level supported by E2Q is, instead, partially abolished (Fig. 6).

Discussion

The protein exhibits a high overall flexibility. Structural verification against experimental NOE connectivities (Bayer et al. 1995) suggests that our computational setup

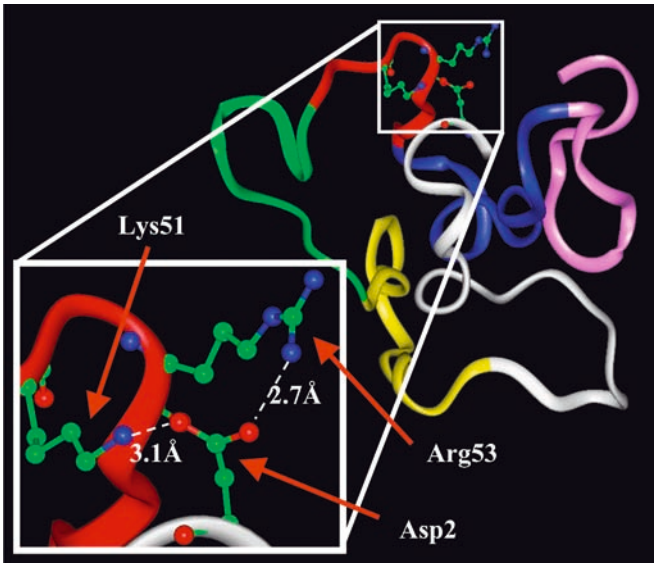


Fig. 5 Lys51– and Arg53–Asp2 interactions. Representative MD snapshot (after 1.5 ns) showing the H-bond pattern. Hydrogen bonds are represented as *dashed lines*. *Ribbon colors* according to Scheme 1

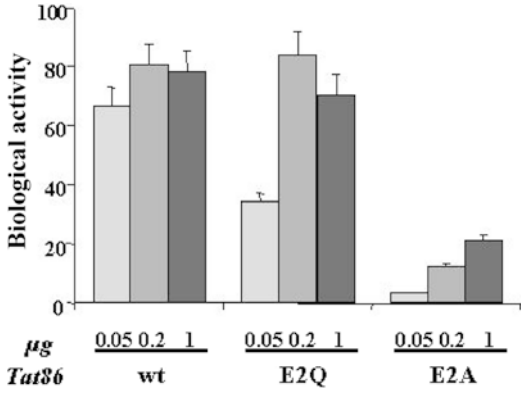


Fig. 6 Biological activity of Tat mutants at position 2. HL3T1 cells (containing an integrated LTR-CAT construct) were transfected with the indicated amounts of plasmid constructs expressing wt Tat and Tat mutants at position 2. CAT assays were performed 48 h after transfection. Shown are means and standard deviations of three independent transfections

may be suitable to describe the dynamics and energetics of this protein.

In spite of the relatively high flexibility of the protein, our calculations suggest that the intra-residue electrostatic interactions do not vary significantly during the dynamics (Fig. 4).

Owing to the lack of secondary structure elements in Tat, it may be difficult to classify the polypeptide segments in clearly defined motional units (Rana and Jeang 1999). However, our calculations suggest that a correlation between some regions of the protein and motional domains does exist. Indeed, if the NT and CRD domains are redefined as regions from aa 1 to 14 and from aa 15 to 37, respectively, the motional behavior of these domains are well separated (Fig. 3b–d). The separation of the CRD into two domains might be consistent with the fact that the terminal residues (namely, Pro14 and Gly15) are highly conserved in the isolates so far identified (http://www.sissa.it/sbp/bc/research/research_hiv.html).

The high degree of flexibility of the CRD (Fig. 3) might play a role for Zn^{2+} binding to this region (Frankel et al. 1988a, 1988b; Slice et al. 1992; Huang and Wang 1996; Garber et al. 1998).

The motion of the QRD turns out to be correlated to almost all the other biochemical domains (Fig. 2). This might support the proposal that a key function of the QRD is to provide structural stability to the protein (Kuppuswamy et al. 1989).

The residues belonging to the ARD are highly hydrated, reducing in this way their mutual electrostatic repulsion (Fig. 4). Arg52, which is the residue mostly stabilized by hydration (Table 3), does not interact with the protein frame. This supports the hypothesis that Arg52 may be involved in the Tat/TAR binding (Calnan et al. 1991; Aboul-ela et al. 1995).

Several stabilizing interactions are present between the NT domain and the last aa of the polypeptide segment expressed by the first exon (aa 1–72, Tables 1, 2 and Fig. 4). On the other hand, the CT domain lacks correlated motion with any other protein domain except the QRD; indeed, the CT domain is in physical contact only with the QRD. The CT domain also presents large contributions to the large-scale motions, indicating that it is highly mobile. Additionally, the part expressed by the second exon of the protein is less conserved than that expressed by the first. Taken together, all these data suggest that the protein conformation should not be dramatically modified by the presence of the second exon (aa 72–86). This might be consistent with the fact that the second exon part of the protein is unnecessary for transactivation (Vives et al. 1994; Verhoef et al. 1998) and, furthermore, that the activities of both parts of the protein may be separable (Brown et al. 1998; Weissman et al. 1998).

Significant electrostatic interactions are formed between the ARD and NT negatively charged residues (Fig. 4, Fig. S2 and Table 1). These interactions are accompanied by correlated motion of the two domains, suggesting that they play an important role for protein stability, consistent with the experimental observation that NT is absolutely essential for protein function (Kuppuswamy et al. 1989). To substantiate these findings we performed site-directed mutagenesis experiments on one of the three negatively charged residues in NT, Asp2. This residue appears to be more important than

Table 3 ARD hydration. Number of water molecules interacting with Lys ammonium and Arg guanidinum nitrogen atoms calculated as integral of radial distribution functions (see Methods)

Residue	Arg49	Lys50	Lys51	Arg52	Arg53	Arg55	Arg56	Arg57
Hydration	5.6	5.5	4.1	6.1	4.6	4.3	4.7	5.6

the other two (Asp5 and Glu9) for the stability of the protein. Indeed: (1) Asp2, unlike Asp5 and Glu9, exhibits significantly correlated motion with other protein domains; (2) Asp2 forms the largest stabilizing electrostatic interactions with the rest of the protein; (3) Asp2 is either conserved or conservatively substituted by Glu, Asp5 and Glu9 are more often mutated in neutral residues in all the isolates so far individuated. In contrast, Asp5 and Glu9 have been shown to be more important than Asp2 for molecular recognition processes (Wrenger et al. 2000).

Here we expressed and measured the biological activity of E2A and E2Q HXB2 Tat. We find that the biological activity of the E2A mutant was drastically impaired, whereas the Gln2 variant shows significantly reduced transactivation capabilities. Although this decrease of activity may be due to a variety of factors (Pantano et al. 2002), these findings may be consistent with the relevance of the strong electrostatic interaction between the ARD and the negatively charged residue at position 2 (Fig. 5) for the structural stability emerging from the calculations[1] Indeed, replacement of Glu with Ala disrupts such interaction, whereas in the E2Q the salt bridge interaction might be replaced by a weaker charge–dipole interaction, resulting in a partially distorted structure of the NT that could be responsible for the observed phenotypes. In this respect, it is worth mentioning that a remarkable correlation has been recently established between the degree of structure conservation and the transactivation capability of several Tat mutants (Pantano et al. 2002).

In conclusion, our combined theoretical and experimental investigations suggest that an essential function of the N-terminal portion of Tat is to maintain the proper conformation of the protein by forming a favorable interaction with the basic domain.

Acknowledgements COFIN-MURST is acknowledged for financial support. We thank U. Rothlisberger and L. Guidoni for useful discussions.

References

Aboul-ela F, Karn J, Varani G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. J Mol Biol 253:313–332

Allen MP, Tildesley DJ (1987) Computer simulation of liquids. Oxford University Press, New York, pp 54–55

Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. Proteins 17:412–425

Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rosch P (1995) Structural studies of HIV-1 Tat protein. J Mol Biol 247:529–535

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Berkhout B, Jeang KT (1989) Trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis. J Virol 63:5501–5504

Berkhout B, Silverman RH, Jeang KT (1989) Tat trans-activates the human immunodeficiency virus through a nascent RNA target. Cell 59:273–282

Brown JA, Howcroft TK, Singer DS (1998) HIV Tat protein requirements for transactivation and repression of transcription are separable. J Acquir Immune Defic Syndr Hum Retrovirol 17:9–16

Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD (1991) Arginine-mediated RNA recognition: the arginine fork. Science 252:1167–1171

Carroll R, Martarano L, Derse D (1991) Identification of lentivirus tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimeras. J Virol 65:3460–3467

Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley MF, Ferguson DM, Radmer RJ, Singh UC, Weiner PK, Kollman PA (1995) AMBER 5. University of California, San Francisco

Churcher MJ, Lamont C, Hamy F, Dingwall C, Green SM, Lowe AD, Butler JG, Gait MJ, Karn J (1993) High affinity binding of TAR RNA by the human immunodeficiency virus type-1 tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. J Mol Biol 230:90–110

Cornell WD, Cieplak P, Bayly CI (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197

Darden TA, York D (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10094

Daune M (1999) Molecular biophysics. Oxford University Press, New York, p 10

Derse D, Carvalho M, Carroll R, Peterlin BM (1991) A minimal lentivirus Tat. J Virol 65:7012–7015

Dorn P, DaSilva L, Martarano L, Derse D (1990) Equine infectious anemia virus tat: insights into the structure, function, and evolution of lentivirus trans-activator proteins. J Virol 64:1616–1624

Felber BK, Pavlakis GN (1988) A quantitative bioassay for HIV-1 based on trans-activation. Science 239:184–187

Frankel AD, Bredt DS, Pabo CO (1988a) Tat protein from human immunodeficiency virus forms a metal-linked dimer. Science 240:70–73

Frankel AD, Chen L, Cotter RJ, Pabo CO (1988b) Dimerization of the tat protein from human immunodeficiency virus: a cysteine-rich peptide mimics the normal metal-linked dimer interface. Proc Natl Acad Sci USA 85:6297–6300

Frankel AD, Biancalana S, Hudson D (1989) Activity of synthetic peptides from the Tat protein of human immunodeficiency virus type 1. Proc Natl Acad Sci USA 86:7397–7401

Friedler A, Friedler D, Luedtke NW, Tor Y, Loyter A, Gilon C (2000) Development of a functional backbone cyclic mimetic of the HIV-1 Tat arginine-rich motif. J Biol Chem 275:23783–23789

Garber ME, Wei P, KewalRamani VN, Mayall TP, Herrmann CH, Rice AP, Littman DR, Jones KA (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. Genes Dev 12:3512–3527

Garcia AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699

Huang HW, Wang KT. (1996) Structural characterization of the metal binding site in the cysteine- rich region of HIV-1 Tat protein. Biochem Biophys Res Commun 227:615–621

Jorgensen WL, Chandrasekhar J, Madura JD (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

[1] Since Glu has the same charge as Asp and it has a longer side chain, its interactions with the ARD residues are expected to be at least as strong as those formed by Asp2 in the HV1Z2 variant.

Karn J (1999) Tackling Tat. J Mol Biol 293:235–254
Kuppuswamy M, Subramanian T, Srinivasan A, Chinnadurai G (1989) Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. Nucleic Acids Res 17:3551–3561
Leach AR (2001) Molecular modelling, 2nd edn. Pearson, Harlow, Essex, UK, p 491
Pantano S, Tyagi M, Giacca M, Carloni P (2002) Amino acid modification in the HIV-1 Tat basic domain: insights from molecular dynamics and *in vivo* functional studies. J Mol Biol 318:1331–1339
Radkiewicz J, Brooks C (2000) Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. J Am Chem Soc 122:225–231
Rana TM, Jeang KT (1999) Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. Arch Biochem Biophys 365:175–185
Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical Integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
Sambrook J, Fritsch EF, Aniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, New York
Slice LW, Codner E, Antelman D, Holly M, Wegrzynski B, Wang J, Toome V, Hsu MC, Nalin CM (1992) Characterization of recombinant HIV-1 Tat and its interaction with TAR RNA. Biochemistry 31:12062–12068
Smith DE, Deng L (1994) Computer simulations of NaCl association in polarizable water. J Chem Phys 100:3757–3766
Sodroski J, Rosen C, Wong-Staal F, Salahuddin SZ, Popovic M, Arya S, Gallo RC, Haseltine WA (1985) Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science 227:171–173
Verhoef K, Bauer M, Meyerhans A, Berkhout B (1998) On the role of the second coding exon of the HIV-1 Tat protein in virus replication and MHC class I downregulation. AIDS Res Hum Retroviruses 14:1553–1559
Vives E, Charneau P, van Rietschoten J, Rochat H, Bahraoui E (1994) Effects of the Tat basic domain on human immunodeficiency virus type 1 transactivation, using chemically synthesized Tat protein and Tat peptides. J Virol 68:3343–3353
Weissman JD, Brown JA, Howcroft TK, Hwang J, Chawla A, Roche PA, Schiltz L, Nakatani Y, Singer DS (1998) HIV-1 tat binds TAFII250 and represses TAFII250-dependent transcription of major histocompatibility class I genes. Proc Natl Acad Sci USA 95:11601–11606
Wrenger S, Faust J, Mrestani-Klaus C, Fengler A, Stockel-Maschek A, Lorey S, Kahne T, Brandt W, Neubert K, Ansorge S, Reinhold D (2000) Down-regulation of T cell activation following inhibition of dipeptidyl peptidase IV/CD26 by the N-terminal part of the thromboxane A2 receptor. J Biol Chem 275:22180–22186