3974–3981 *Nucleic Acids Research, 1996, Vol. 24, No. 20*

Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge

Fareed Aboul-ela*, Jonathan Karn and Gabriele Varani

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received July 3, 1996; Revised and Accepted August 22, 1996

PDB no. 1ANR

ABSTRACT

Efficient transcription from the human immunodeficiency virus (HIV) promoter depends on binding of the viral regulatory protein Tat to a *cis*-acting RNA regulatory element, TAR. Tat binds at a trinucleotide bulge located near the apex of the TAR stem–loop structure. An essential feature of Tat–TAR interaction is that the protein induces a conformational change in TAR that repositions the functional groups on the bases and the phosphate backbone that are critical for specific intermolecular recognition of TAR RNA. We have previously determined a high resolution structure for the bound form of TAR RNA using heteronuclear NMR. Here, we describe a high resolution structure of the free TAR RNA based on 871 experimentally determined restraints. In the free TAR RNA, bulged residues U23 and C24 are stacked within the helix, while U25 is looped out. This creates a major distortion of the phosphate backbone between C24 and G26. In contrast, in the bound TAR RNA, each of the three residues from the bulge are looped out of the helix and U23 is drawn into proximity with G26 through contacts with an arginine residue that is inserted between the two bases. Thus, TAR RNA undergoes a transition from a structure with an open and accessible major groove to a much more tightly packed structure that is folded around basic side chains emanating from the Tat protein.

INTRODUCTION

The human immunodeficiency virus type-1 (HIV-1) Tat protein activates transcription from the viral long terminal repeat by stimulating elongation efficiency of RNA polymerase II (1–7). Tat is introduced to the transcription complex following binding to a *cis*-acting RNA regulatory element called TAR, the *trans*-activation response element (8–14). Tat acts analogously to the bacteriophage λ N anti-terminator protein and creates a modified transcription complex by binding to the elongating RNA polymerase together with TAR RNA and cellular co-factors (2,7,15,16).

The TAR RNA element is positioned immediately after the transcription start site (nt +1 to +59) and forms a stable hairpin structure (8,12,17,18). Tat binds in the region of a three base (UCU or UUU) bulge and recognises both the identity of adjacent Watson–Crick base pairs and the positions of surrounding phosphate groups (8,10–12,19–25). The interaction between Tat and TAR is essential for viral growth; mutants in TAR with reduced affinity for Tat are unable to replicate efficiently (26,27).

Tat belongs to the 'basic domain' family of RNA binding proteins (28). Members of this family, which also includes HIV-1 Rev protein and the bacteriophage λ N protein, carry arginine-rich sequences that mediate their interactions with RNA (19,28,29). Upon binding to TAR RNA, the basic region of Tat promotes a conformational rearrangement in TAR which places the functional groups recognised by the protein in a specific spatial arrangement (18,30,31). Evidence that the conformational change in TAR RNA is essential for specific Tat binding comes from the observation that mutations in TAR that result in the largest decreases in Tat affinity are associated with residues G26 and U23, the two bases mediating the conformational change (8).

Using heteronuclear multidimensional NMR, we have previously determined the structure of a 29 nt fragment that contains the top part of the TAR stem–loop bound to a 36 residue peptide which contains the essential basic and core regions of Tat (31). In the present work, we have used 871 NMR-derived experimental restraints to determine the structure of the free HIV-1 TAR fragment. In contrast to the compressed structure adopted by the bound form of TAR RNA, the bulge and adjacent base pairs in the free TAR RNA structure are found in an exceptionally wide and accessible major groove. Comparison of the free and bound structures of TAR also reveals two distinct mechanisms for accommodating bulged residues within an RNA duplex.

MATERIALS AND METHODS

RNA preparation and characterisation

The NMR structure was determined using either a 29 nt RNA oligonucleotide corresponding to the apical region of the wild-type TAR sequence (Fig. 1) or a 27 nt RNA oligonucleotide in which the apical loop sequence of TAR (CUGGGA) was replaced by a stable tetraloop sequence (UUCG). Milligram quantities of TAR RNA were synthesised by *in vitro* transcription of DNA oligonucleotides in a T7 polymerase system and purified by polyacrylamide gel electrophoresis. The RNA concentrations were

* To whom correspondence should be addressed

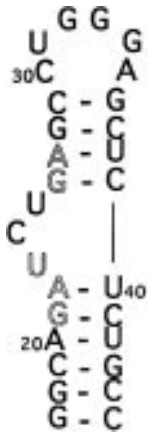

Figure 1. Sequence and secondary structure of the 29 residue TAR RNA used in the present study. Nucleotides critical for Tat recognition are outlined.

determined from the UV absorbance at 260 nm. Preparation of >98% ^{15}N-labelled and ^{15}N-^{13}C-labelled RNA was as described (31–33).

NMR spectroscopy

NMR spectra were recorded either on a Bruker AMX-500 NMR spectrometer operating at 500 MHz (^{1}H) or a 300 MHz Bruker DRX-300 or a Bruker DMX-600 spectrometer operating at 600 MHz and equipped with triple resonance gradient probes. One- and two-dimensional ^{1}H NMR spectra were recorded at 1–2 mM RNA concentrations in 5 mM phosphate buffer, pH 5.5. High quality spectra with the wild-type TAR sequence were obtained only in the presence of ~50 mM NaCl, whereas the tetraloop TAR variant yielded high quality spectra under all conditions.

To determine the structure of the free TAR RNA, essentially the same set of NMR experiments were carried out as in the studies of the TAR RNA–peptide complex (31). Briefly, NMR experiments included a two-dimensional NOESY build-up series, two-dimensional ^{1}H-^{13}C HSQC- and HCCH-TOCSY experiments, three-dimensional ^{13}C-edited HCCH-COSY, HMQC-TOCSY and NOESY-HMQC experiments and a series of ^{31}P-^{1}H correlation spectra. Experimental details concerning NMR data acquisition and processing have been described previously (31).

Hydrogen bonding, interproton distance and dihedral angle constraints

Hydrogen bonding constraints for base pairs in the TAR RNA stem were based on distinctive patterns of NOE interactions involving U and G imino resonances, together with characteristic chemical shift ranges and rates of exchange with solvent (34). Hydrogen bonding constraints were introduced as distance constraints between heavy atoms (2.9 ± 0.3 Å). No hydrogen bonding constraints were introduced for the loop or bulge regions.

Distance restraints between non-exchangeable protons were obtained from the intensities of cross-peaks in NOE build-up series and from a three-dimensional ^{13}C-edited NOESY spectrum acquired at 150 ms mixing time. Briefly, strong peaks were introduced as 0–3 Å distance constraints, medium peaks as 0–4 Å constraints and weak peaks as 0–5 Å constraints. NOE cross-peaks that were only observable at long mixing times (100–200 ms) were treated as very weak constraints (0–6.0 Å) in order to reduce the possibility that inaccurate restraints arising from spin diffusion could distort the final structure. NOE interactions involving exchangeable resonances were only incorporated with upper boundaries of 5.5 or 6.5 Å (31,35), with the exception of the very strong NOE cross-peaks from Watson–Crick base pairs (involving the A-H^{2} and U-NH^{3} resonances or the C-NH^{2} and G-NH^{1} resonances), which were introduced as medium constraints (upper bounds 4 Å).

After completion of a first round of calculations, predicted NOEs which were absent from the restraint list were investigated as described (31). We found no example of predicted close contacts (<3.8 Å) that did not generate clear NOE interactions in the NOESY spectra.

The procedure used to derive dihedral angle constraints has recently been described in great detail (31,34,36). Briefly, α (O3′-P-O5′-C5′) and ζ (C3′-O3′-P-O5′) were constrained only very qualitatively (0 ± 120°) from ^{31}P chemical shifts. ^{31}P chemical shifts are dependent on a number of factors aside from torsion angles (31,34,36–38). Therefore, very qualitative restraints were introduced, and only when standard ^{31}P chemical shift values (between 4 and 5 p.p.m.) were observed. The β (P-O5′-C5′-C4′), ε (C4′-C3′-O3′-P) and γ (O5′-C5′-C4′-C3′) dihedral angles were constrained using semiquantitative estimates of coupling constants (31,36). The sugar puckers, identifying the δ (C5′-C4′-C3′-O3′) dihedral angles, were constrained using a variety of ^{1}H-^{1}H and ^{1}H-^{13}C scalar couplings to C3′-*endo* (δ = 85 ± 30°) conformations or were left unconstrained in cases where significant conformational averaging was present.

No dihedral restraints, other than restraints on δ, were incorporated in the bulge region, due to evidence of averaging in the observed patterns of scalar couplings. This is consistent with the observation of enhanced conformational flexibility in this region of the structure (39).

Structure determination

RNA structures were calculated using restrained molecular dynamics followed by energy minimisation with an all-atom force field using X-PLOR (40). The standard X-PLOR 3.1 angle parameters were modified as described to obtain more realistic sugar puckers (35). RMS deviations between the average structure (calculated using '*clusterpose*'; 41) and the converged structures were calculated for all atoms and for subsets of the structure.

A total of 754 interproton distance restraints derived from NOESY data (448 intranucleotide and 306 internucleotide), corresponding to an average of ~28 constraints/nucleotide, were used in the structure determination. A total of 49 internucleotide restraints were non-sequential, including interstrand restraints involving exchangeable and A-H^{2} protons. Three hundred and twenty three distance restraints were conformationally redundant, meaning that the interproton distances in question are already restrained by covalent bond length and angle restraints to within values less than those of the NMR restraints. Although the presence of those restraints does not affect the final outcome (C.Gubser, personal communication), it is nevertheless important to sift through these restraints carefully to identify important internucleotide NOEs unambiguously. In addition to the interproton distance restraints, 25 hydrogen bonds for the base pairs forming the stem were used. No hydrogen bonding constraints were introduced in the flexible apical loop or within the bulge region.

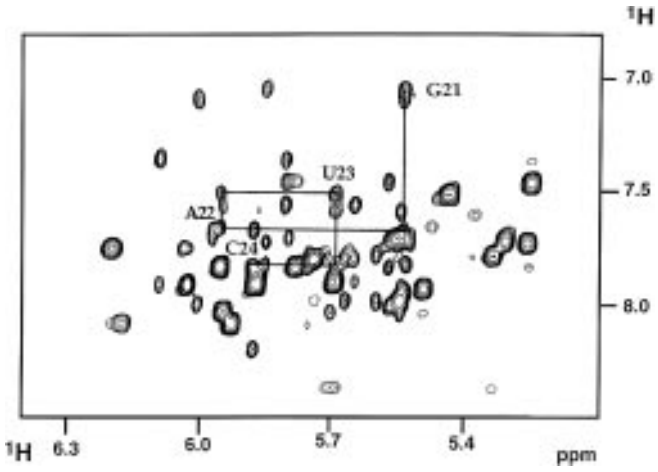

Figure 2. Aromatic to H1′ region of NOESY spectrum obtained in D_2O at 300 ms mixing time. The sequential connectivity pathway for nt 21–24, similar to that seen in A-form RNA, is indicated.

A total of 92 dihedral angles (12 α, 15 β, 17 γ, 19 δ, 17 ε and 12 ζ) were constrained to different degrees of precision as described above. No assumption was made at any stage on the planarity of the base pairs leading to non-experimental 'planarity' restraints.

A total of 60 structures with random backbone torsion angles were generated as initial models for structure determination. Structure calculations were then done in three successive steps (31,34,35,42,43). Converged and non-converged structures were separated from the number and total energy of constraints violations. For the best 20 converged structures, less than five distance violations of between 0.1 and 0.2 were found, whereas violations of >0.3 were observed in the non-converged structures (21 onwards; Fig. 4). No violations of dihedral angle constraints of >2° were observed for the converged structures. The overall energy was also significantly higher (at least by 15 kcal/mol) for the non-converged structures. RMSD values were obtained from energy-ordered RMSD profiles (44).

RESULTS

NMR analysis of free TAR RNA

The binding site for Tat is centred around a three nucleotide bulge found near the apex of the TAR RNA stem–loop (Fig. 1). Extensive mutagenesis and chemical probing studies have defined the key functional groups on TAR RNA that are required for high affinity Tat binding (8,9,11–13,21–25,45,46). The first residue in the bulge must always be a uridine (U23), but the other residues in the bulge, C24 and U25, appear to act predominantly as spacers and may be replaced by other nucleotides, or even by non-nucleotide linkers (8,10,11,13,45). Tat recognises the identity of two base pairs in the stem above the U-rich bulge, G26·C39 and A27·U38 (8,21). Critical phosphate contacts involve phosphates P21, P22 and P40, which are located below the bulge on both strands (19,23–25,47).

The Tat binding site can be presented on short oligoribonucleotides that carry the sequence of the apical portion of TAR RNA, as well as on short duplexes that span the bulge region but lack the apical loop (8,23,25,45). The NMR studies described here used singly (^{15}N) and doubly (^{15}N-^{13}C) labelled 29 residue ribo-oligonucleotides containing the 3 nt UCU bulge and the apical loop (Fig. 1) (31). For ease of assignments and spectral simplifications, data were also obtained from a ^{15}N-labelled TAR RNA sample in which the wild-type apical loop (CUGGA) was replaced by a stable tetraloop sequence (UUCG) (31). The tetraloop substitution leads to only a small decrease in Tat binding affinity (8).

Helical regions in TAR RNA

Qualitative analysis of the free TAR RNA spectra shows that the stems surrounding the UCU bulge form a stable double helical structure. Sharp imino proton resonances for U and G residues within the stem regions are observed at chemical shifts characteristic of Watson–Crick base pairs. Strong evidence for base pairing in the two double helical regions is also provided by prominent U-H^3–A-H^2 and G imino–C amino NOE interactions. However, no such NOE cross-peaks are found corresponding to the A22·U40 base pair immediately below the bulge, suggesting that, in contrast to the bound TAR RNA structure (31), this base pair may be unstable in the free TAR RNA structure.

Across from the bulge strand, residues C39 and U40 produce each of the sequential NOEs found in helical A-form structures. However, the sequential NOE interaction between U40 H6 and C39 H1′ is unusually strong, indicating that this is a region where the helix becomes significantly distorted.

Bulge region

Base–sugar connectivity pathways typical of A-form RNA, including both anomeric and other sugar resonances, are observed in the bulge region up through residue C24. As an example, sequential aromatic–H1′ NOE interactions are shown for residues A22–C24 in Figure 2. Further evidence that base stacking continues from A22 through the first two residues of the loop (U23 and C24) comes from aromatic–aromatic NOE interactions between residues A22 and U23, and sugar–sugar cross-peaks between A22 and U23 and between U23 and C24. Nine sequential NOE cross-peaks were observed between residues U23 and C24 and eight cross-peaks were found between residues A22 and U23. These NOE interactions demonstrate that the first two residues in the bulge are stacked continuously over A22.

In contrast to U23 and C24, the third loop nucleotide in the bulge, U25, is connected by only very few NOE interactions to neighbouring residues and the standard sequential aromatic–sugar NOE cross-peaks were not observed. This is the only part of the entire TAR molecule where the connectivity pattern completely breaks down. NOE cross-peaks between resonances of G26 and the H5′, H5″ sugar resonances of C24 were identified, indicating that U25 is not stacked between C24 and G26. However, in contrast to spectra of the bound TAR RNA, clear NOE cross-peaks between aromatic resonances of G26 and sugar or aromatic resonances of C24 were not found. This suggests that U25 is only partially or transiently looped out of the helix.

No imino or exchangeable resonances could be assigned to any of the three bulge residues and assignments for the U23 and U25-N3 positions obtained from long-range correlation spectra indicate that these nitrogens resonate where unpaired uracyl residues are usually found. Therefore, the U23 and U25 imino groups are probably not involved in hydrogen bonding interactions.

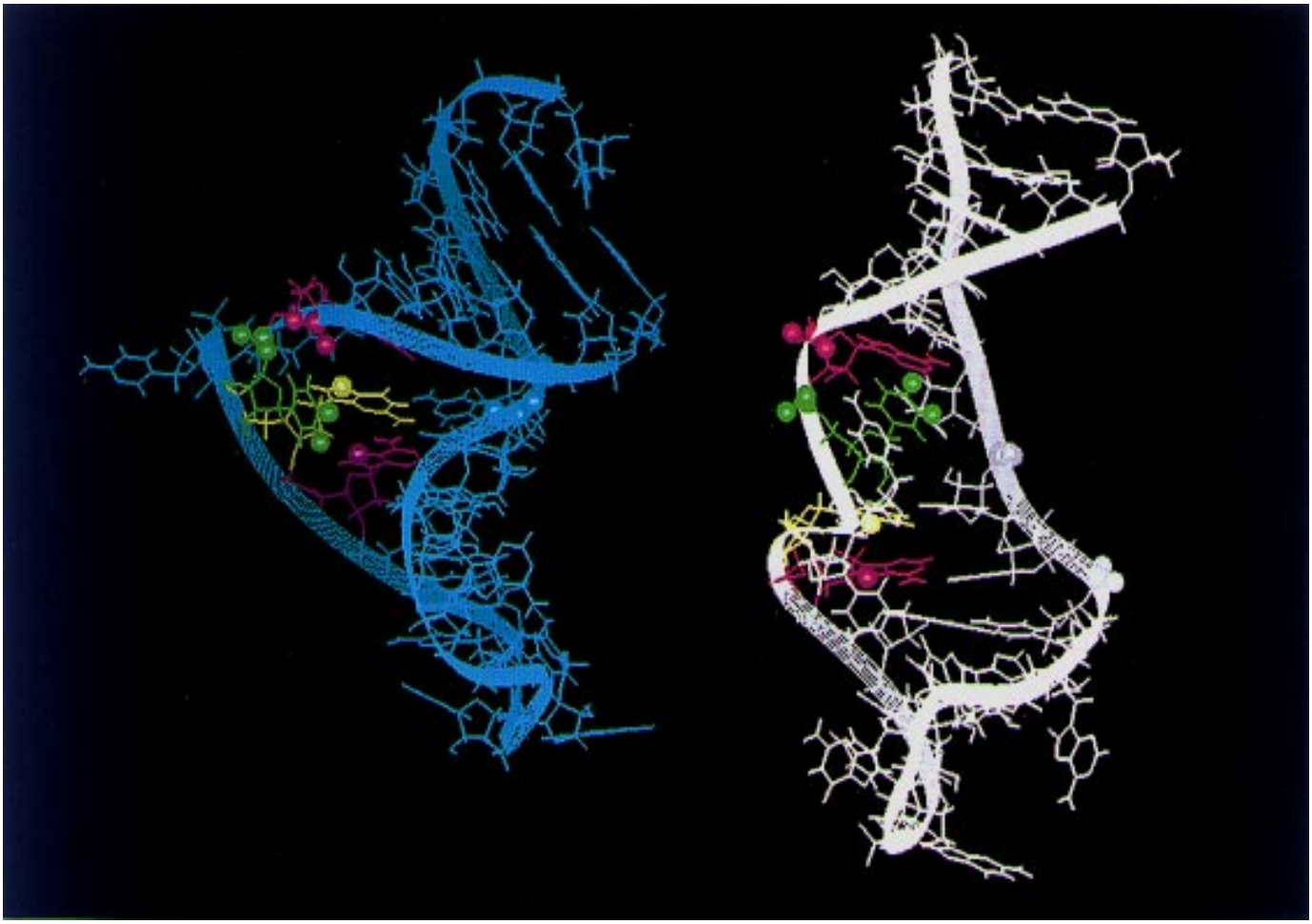

Figure 3. Major groove view of the free TAR RNA (left) and bound TAR RNA (right) structures. These are representative structures with low violations of experimental constraints. Functional groups which have been identified as critical for Tat binding, including the backbone phosphates, are highlighted by van der Waals spheres. These include; phosphates: 21–22, 22–23, 39–40 and 40–41, U23 H1 and O4, G26 and A27 N7. A22 and A27 are highlighted in purple, U23 in green and G26 in yellow.

The structure of TAR RNA in the absence of ligands

The structure of free TAR RNA was calculated using a set of 871 experimental constraints and procedures successfully applied to the determination of the structure of the bound TAR RNA (31). Calculations began with fully randomised starting structures. Extensive restrained molecular dynamics simulations at high temperatures were performed in order to obtain a global fold of the RNA that was consistent with the experimental data. Subsequent steps included restrained molecular dynamics and energy minimisation at decreasing temperatures (31,34,42,43). A final energy minimisation in the presence and absence of electrostatic interactions was carried out in order to evaluate whether calculated structures also had satisfactory van der Waals and electrostatic contacts.

Figure 3 shows a major groove view of a representative structure for the free TAR compared with a similar view of the bound TAR structure. In both structures, the helical regions adjacent to the bulge (residues G18–G21, C41–C44 and G26–C29, G36–C39) show a characteristic A-form geometry with apparent major groove widths within limits typical for A-form structures. However, the three base TAR bulge and the neighbouring residues define a pocket of major groove accessibility in the free TAR RNA structure that is clearly more extensive than the comparable regions in either the bound TAR RNA structure or in A-form RNA.

In the free TAR RNA structure, U23 stacks on A22 and C24 stacks on U23. On the strand opposite the bulge, residues U40–C41 display continuous stacking. To accommodate the displacement in helical twist between the two strands resulting from the presence of the bulged bases, U25 is swung outside the helix and away from G26. This pattern of partial stacking of the bulge residues is similar to that of bulged residues in DNA fragments that carry a kink in the helical axis (48–50), but has not been previously described for an RNA bulge.

The base pairs surrounding the bulge in the free TAR RNA structure also show significant deviations from the bound TAR RNA structure and from A-form RNA. These structural changes appear necessary to relieve the increases in helical twist and rise induced by the continuous helical stacking of the bulged nucleotides. In contrast to the bound structure, where residue G26 is in close proximity to A22, G26 is separated from the lower stem region and has no clear stacking partner in the free TAR RNA structure. Residue U40 stacks continuously on residue C41 from

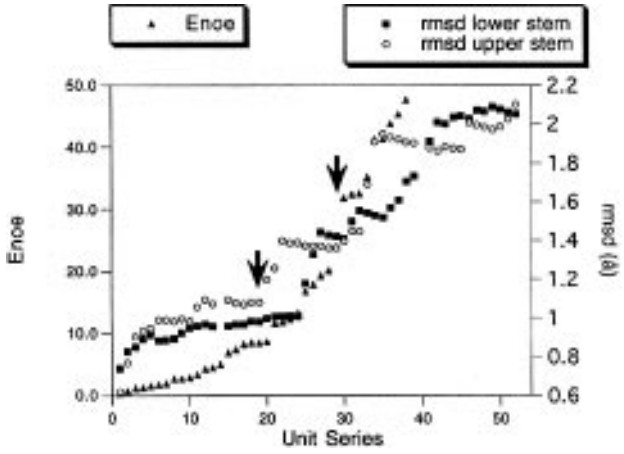

Figure 4. Energies of NOE violations for the 37 lowest energy structures are plotted for each calculated structure in order of increasing energy. Arrows highlight cut-off points at structures 20 and 29, separating structures with significantly differing degrees of agreement with the data.

the lower stem helix, but because of the distortion introduced by the UCU bulge, stacking interactions with the upper stem residue C39 are weakened. Consistent with the high solvent exchange observed for the U40 imino proton, the A22·U40 base pair does not appear to be in a typical Watson–Crick pairing geometry in the free TAR RNA structure. The increased twist across from the bulge may be at least partly responsible for the destabilisation of A22·U40 base pairing. A similar phenomenon was observed for an A·T base pair adjacent to an AAA bulge within a DNA fragment (48).

Precision of the structure

Of 60 calculations initiated with different randomised starting structures, eight failed to reach completion altogether due to unacceptable violations of covalent geometry. The remaining structures were analysed in terms of energies of violations of the NOE restraints and overall energies, including energies associated with constraints on covalent geometry and, after final refinement, electrostatic and van der Waals energies. These 52 structures were ordered according to their agreement with the data as defined by the total energy of NOE violations (E_{NOE}) (Fig. 4). Similar results are obtained when the total energy (E_{tot}) is used to order the structures (34). As shown in Figure 4, E_{NOE} increases dramatically after structure 29. Therefore, structures 30 onwards can be considered to be non-converged. The separation between structures 1–20 and 21–29 is also clear, though less straightforward than for 30 onwards. Within the first 20 structures, there are no violations of NOE constraints >0.2 Å or dihedral angle restraint violations >2°. The next set of nine structures (21–29) show significantly higher energies of NOE violations. The structures in this group each showed one to three NOE constraint violations of 0.2–0.5 Å.

Superpositions of the 20 best structures are shown for the stem regions in Figure 5. RMS deviations relative to the average structure are similar for the two stem regions (~1 Å), a value that is probably close to the current realistic limit for precision of RNA structure determination (34,51). The bulge region (residues G21–C24) is less precisely defined than the stem regions (the overall RMS deviation about the average structure for residues

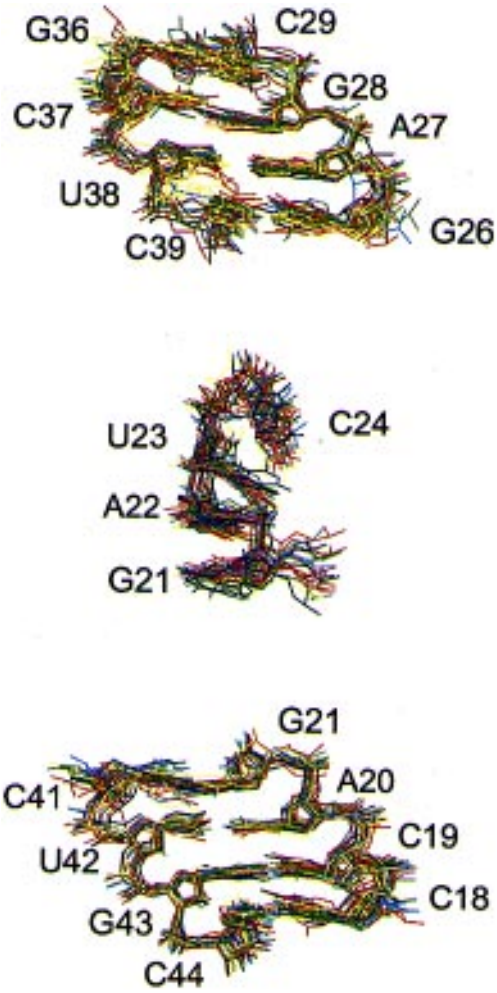

Figure 5. Superpositions of the 20 best calculated structures. (Top) Upper stem residues (26–29 and 36–39). (Centre) Bulge residues (21–24). (Bottom) Lower stem residues (18–21 and 41–44). The superpositions based on residues 21–24 include only the 14 structures which showed the same consistent stacking pattern. The six excluded structures contain small but significant violations of NOE distance restraints within the bulge region.

G21–C24 is just under 2 Å), but continuous helical stacking is observed from residues G21 to C24 in 23 out of the 29 lowest energy structures. Including structures 21–29 in this analysis only slightly increases the RMS deviations from the average structure for the stem regions and did not increase the RMS deviations in the bulge region.

Analysis of torsion angle statistics (see Supplementary material available via NAR Online) reveals that some torsion angles in the bulged nucleotides 23 and 24 were relatively well defined by NOE interactions, in spite of the lack of torsion angle restraints. For all of the 29 lowest energy structures, the glycosidic angle, χ, was confined to between 195° and 230° for residue 24, γ was between 30° and 75° for all but two structures and ε between 180° and 280° for all but one of 29 structures. Similar precision was found for U23, whereas values for U25 were more scattered. This result indicates that the bulge region is less ordered than the stem, but it is nevertheless in a defined conformation for at least the first two residues.

The least well-defined residues in the free TAR RNA structures are U25 and U40. The precise position of U25 is not well defined in the converged structures because of the low number of internucleotide NOE cross-peaks involving this residue. This may reflect genuine conformational flexibility at this position. Although a substantial number of sequential and intranucleotide restraints were observed for U40, no NOE interactions were observed involving the exchangeable imino proton. Therefore, the data did not justify incorporating base pairing or hydrogen bond restraints at this position. The absence of cross-strand restraints rather than intrinsic flexibility of the base explains the lack of definition of the position of U40 in the present study (34). Consistent with this view, recent reports of ^{13}C relaxation measurements for free TAR RNA did not indicate an unusually high degree of flexibility at residue U40 (39).

When converged structures are superimposed based on all heavy atoms, the average RMS deviation from the average structure is ~5 Å. If structures are superimposed based on the lower stem region, the helical axis connected with the upper stem is found to point in different directions for different structures, preventing any definitive conclusion from being reached regarding the degree of helical bending induced by the bulge. This finding reflects the difficulty in defining global conformations of nucleic acids by NMR due to the absence of long-range experimental restraints (51).

Alternate structures in the bulge region

Twenty three of the 29 lowest energy structures showed continuous stacking from residue G21 to C24 and are very similar to the representative structure shown in Figure 3A and in Figure 6. In three structures, residue A22 is unstacked and tilted by >45° relative to the plane of the G21·C41 base pair, while partial stacking of U23 on A22 and of C24 on U23 is still observed. In the remaining three structures, U23 appears sideways and stacking is disrupted throughout this region. These six alternative structures each show small NOE violations within the bulge region. Their presence in the set of converged structures probably reflects the lack of powerful cross-strand NOE restraints in the absence of base pairing within this region, although it is difficult to rule out the possibility that these structures represent transient conformers present in solution.

DISCUSSION

Conformational flexibility of the TAR RNA bulge

Bulge motifs are common building blocks of RNA structures and are often present at sites recognised by RNA binding proteins. NMR studies of the TAR RNA structure reveal two radically different methods for accommodating bulged residues within duplex RNA and provide a detailed understanding of the mechanism of Tat–TAR recognition (18,30,31).

The free TAR RNA structure, reported in this paper, presents a paradigm for the packing of bulged residues into a continuous RNA helix. The presence of several bulged residues stacked continuously within one strand forces a physical separation between the base pairs adjoining the two stem regions on that strand. To accommodate this distortion, while maintaining the energetically favourable continuous stacking interactions of the A-form helices from the adjacent regions, some separation between the residues on the non-bulged strand is also required.

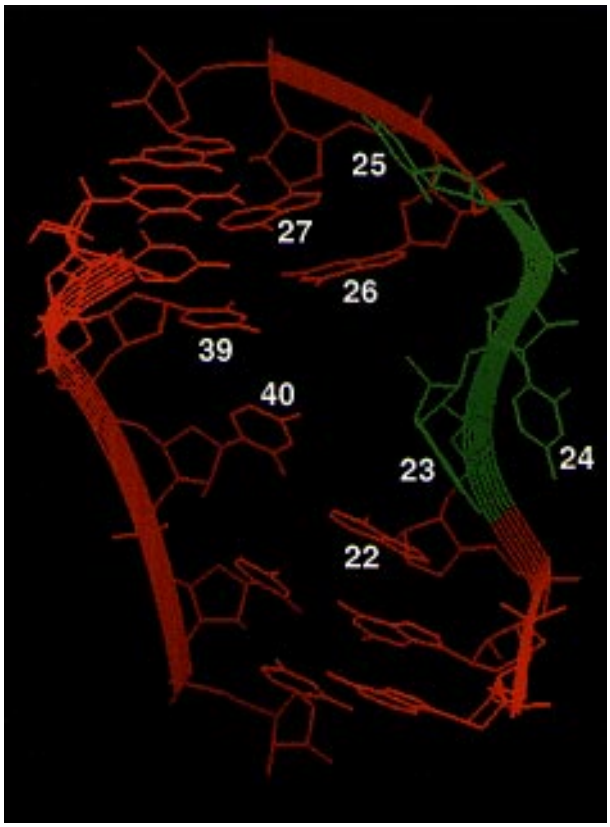

Figure 6. Close-up view of the TAR bulge. In the stacked region of the bulge, the twist continues only through the bulged strand and distortions near the bulge site re-establish A-form helical geometry in the two stem regions. The disparity in twist (and rise) between the two strands is accommodated through a distortion in stacking in the non-bulged strand and by a negative twist between residues C24 and G26. Bending or kinking of the helix helps to accommodate these structural abnormalities.

This may be related to a discontinuity in the direction of the helix axis, which is observed in hydrodynamic and optical experiments for free TAR (52,53), as well as for other bulged nucleic acids (54–58). Thus, the NMR data are consistent with these experiments, even though the global characteristics of TAR RNA are not defined sufficiently precisely by the NMR data to provide an exact measure of the degree of bending induced by the bulged residues.

A second type of distortion results from the difference in twist between the strand carrying the bulged residues and the opposite strand. In the free TAR RNA structure, the disparity in twist between the two strands is accommodated by U25 looping out of the helix and by introducing a negative twist between G26 and C24 (Fig. 6). A slight increase in twist is also observed on the opposite strand, around residue U40, although the precision of the present study does not allow us to determine precisely how this increased twist is distributed between residues C39, U40 and C41.

The bound TAR RNA illustrates a radically different means of adjusting to the presence of the three base bulge. When basic ligands bind TAR (18,31), the bulged residues U23 and C24 loop out of the helix and a binding pocket is created that places the guanidinium and εNH groups of the arginine within hydrogen bonding distance of functional groups on G26 and U23. The

aliphatic side chain of an arginine residue also comes into very close contact with U23 and forms a stacking interaction underneath this bulged residue reminiscent of cation–π interactions observed in proteins and their complexes (59). These interactions, together with a reduction in the energetic strain between the two stem regions above and below the bulge in the bound structure, appear to compensate for the loss of energy associated with disruption of the A22–U23–C24 stacking interactions present in the free TAR structure.

The NMR data are clearly consistent with a major conformer for the free TAR bulge along the lines described above. However, the data also indicate that the bulge is more flexible than the helical regions or the bound bulge. Thus, binding of Tat not only induces a substantial conformational rearrangement in TAR, but also locally produces a more rigid structure. As a consequence of the flexibility in the bulge region of free TAR, the TAR bulge samples a wider, conformational space in the absence of ligands than when bound by Tat-derived peptides. In this situation, a method of structure calculation that assumes one structure rather than multiple conformers does not represent a description of the full conformational range sampled by free TAR. It is possible that minor conformers that differ from the predominant one described here are also present. Nevertheless, we believe that such conformers, if present, must be sparsely populated. If two or more major conformers that differ significantly from the predominant conformer were populated, one would expect to find mutually inconsistent NOE interactions leading to severe violations of experimental restraints. This is clearly not the case. Alternative structures resulting from the calculations represent local minima which happen to fit the data reasonably well but may or may not occur in solution. Furthermore, there are small restraint violations observed for these structures concentrated within the bulge region.

Major groove accessibility of TAR RNA

In an early study, Weeks and Crothers (12) demonstrated that the N^7 groups found on the purines adjacent to the TAR RNA bulge are unusually accessible to chemical modification by DEPC and DMS. In agreement with the chemical modification data, the NMR structures reported here also demonstrate that the TAR RNA bulge region is surrounded by a widened and accessible major groove (Figs 3 and 6).

How does the open structure at the TAR RNA bulge contribute to Tat binding? By analogy with DNA recognition by major groove binding proteins, Weeks and Crothers (12) proposed that the widened major groove created by the bulged residues in TAR RNA facilitates Tat recognition by permitting the insertion of large and comparatively rigid protein structural elements, such as α-helices. However, this model is inconsistent with the NMR data showing that the major groove in the bound TAR RNA structure is not wide enough to allow penetration by such elements (18,31). Instead, TAR RNA appears to fold itself around the basic region of Tat, which exists as an extended and disordered chain in solution (60), though some α-helical tendency appears within the basic region of a HIV–EIAV tat hybrid peptide (61). A similar mechanism occurs when basic peptides derived from Tat bind to TAR RNA (8,12,14,46,62,63). These short, basic peptides are unfolded in solution, but show new intrapeptide NOE cross-peaks after binding to TAR RNA (31).

The conformational change in TAR RNA repositions the $P_{(21–22)}$, $P_{(22–23)}$ and $P_{(40–41)}$ phosphates, which provide energetically important contacts with Tat (25), around the arginine binding pocket on one surface of the TAR RNA structure (Fig. 3). Model building suggests that these phosphates can be easily contacted by other basic residues found in the TAR RNA binding region (data not shown). Contacts between these phosphates and amino acid side chains from Tat contribute not only to the affinity of the interaction, but also to its specificity, by providing discrimination with respect to other bulged RNA structures. The Tat–TAR interaction therefore provides a clear example of the 'indirect readout' of nucleic acid sequences through recognition of backbone phosphates whose positions are uniquely defined by the RNA structure. The importance of the conformational change in TAR for Tat binding is confirmed by the observation that the mutations that produce the most severe reductions in TAR activity involve G26 and U23 and disrupt the intermolecular interactions that are responsible for the folding transition (8,12). Thus, although the widened major groove in the free TAR structure may facilitate initial protein binding events, it is clear that high affinity Tat binding to TAR is due to the refolding of TAR RNA around an extended and unstructured basic domain from Tat, rather than being due to the insertion of a large and stable protein structural element into a large pre-formed RNA binding site.

Design of anti-HIV agents

The interactions between Tat and TAR, as described in detail above, are critical for virus replication. For example, replication of HIV is strongly inhibited by over-expression of TAR RNA 'decoy' sequences which act as competitive inhibitors of Tat binding (64–66). Thus, it seems likely that small molecules that inhibit Tat binding to its recognition site will also have antiviral activity. Two strategies for drug design are suggested by the structural studies on TAR RNA. First, small molecules that target the bound conformation of TAR RNA are expected to behave as competitive inhibitors of Tat binding. An alternative strategy is to target the unbound TAR structure. A small molecule that is able to bind free TAR RNA with high affinity and block the conformational change could also be an effective inhibitor of the Tat–TAR interaction and, consequently, of HIV growth.

See supplementary material available in NAR Online.

ACKNOWLEDGEMENTS

We thank Andres Ramos and Dr Mike Gait for helpful discussions and Brian Wimberly for use of the program 'backwheel' for generating plots of torsion angles. This work was partially supported by a grant from the MRC Aids Directed Program (940 5057).

REFERENCES

1 Laspia,M.F., Rice,A.P. and Mathews,M.B. (1989) *Cell*, **59**, 283–292.
2 Kao,S.-Y., Calman,A.F., Luciw,P.A. and Peterlin,B.M. (1987) *Nature*, **330**, 489–493.
3 Feinberg,M.B., Baltimore,D. and Frankel,A.D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4045–4049.
4 Graeble,M.A., Churcher,M.J., Lowe,A.D., Gait,M.J. and Karn,J. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6184–6188.
5 Rittner,K., Churcher,M.J., Gait,M.J. and Karn,J. (1995) *J. Mol. Biol.*, **248**, 562–580.
6 Marciniak,R.A., Calnan,B.J., Frankel,A.D. and Sharp,P.A. (1990) *Cell*, **63**, 791–802.
7 Marciniak,R.A. and Sharp,P.A. (1991) *EMBO J.*, **10**, 4189–4196.
8 Churcher,M.J., Lamont,C., Hamy,F., Dingwall,C., Green,S.M., Lowe,A.D., Butler,P.J.G., Gait,M.J. and Karn,J. (1993) *J. Mol. Biol.*, **230**, 90–110.

9 Dingwall,C., Ernberg,I., Gait,M.J., Green,S.M., Heaphy,S., Karn,J., Lowe,A.D., Singh,M., Skinner,M.A. and Valerio,R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6925–6929.
10 Dingwall,C., Ernberg,I., Gait,M.J., Green,S.M., Heaphy,S., Karn,J., Lowe,A.D., Singh,M. and Skinner,M.A. (1990) *EMBO J.*, **9**, 4145–4153.
11 Roy,S., Delling,U., Chen,C.-H., Rosen,C.A. and Sonenberg,N. (1990) *Genes Dev.*, **4**, 1365–1373.
12 Weeks,K.M. and Crothers,D.M. (1991) *Cell*, **66**, 577–588.
13 Weeks,K.M., Ampe,C., Schultz,S.C., Steitz,T.A. and Crothers,D.M. (1990) *Science*, **249**, 1281–1285.
14 Calnan,B.J., Biancalana,S., Hudson,D. and Frankel,A.D. (1991) *Genes Dev.*, **5**, 201–210.
15 Churcher,M.J., Lowe,A.D., Gait,M.J. and Karn,J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 2408–2412.
16 Keen,N.J., Gait,M.J. and Karn,J. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 2505–2510.
17 Muesing,M.A., Smith,D.H. and Capon,D.J. (1987) *Cell*, **48**, 691–701.
18 Puglisi,J.D., Tan,R., Canlan,B.J., Frankel,A.D. and Williamson,J.R. (1992) *Science*, **257**, 76–80.
19 Calnan,B.J., Tidor,B., Biancalana,S., Hudson,D. and Frankel,A.D. (1991) *Science*, **252**, 1167–1171.
20 Weeks,K.M. and Crothers,D.M. (1993) *Science*, **261**, 1574–1577.
21 Delling,U., Reid,L.S., Barnett,R.W., Ma,M.Y.-X., Climie,S., Sumner-Smith,M. and Sonenberg,N. (1992) *J. Virol.*, **66**, 3018–3025.
22 Roy,S., Parkin,N.T., Rosen,C.A., Itovitch,J. and Sonenberg,N. (1990) *J. Virol.*, **64**, 1402–1406.
23 Hamy,F., Asseline,V., Grasby,J., Iwai,S., Pritchard,C., Slim,G., Butler,P.J.G., Karn,J. and Gait, M. (1993) *J. Mol. Biol.*, **230**, 111–123.
24 Tao,J. and Frankel,A.D. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2723–2726.
25 Pritchard,C.E., Grasby,J.A., Hamy,F., Zacharech,A.M., Singh,M., Karn,J. and Gait,M.J. (1994) *Nucleic Acids Res.*, **22**, 2592–2600.
26 Harrich,D., Mavankal,G., Mette-Snider,A. and Gaynor,R.B. (1995) *J. Virol.*, **69**, 4906–4913.
27 Klaver,B. and Berkhout,B. (1994) *EMBO J.*, **13**, 2650–2659.
28 Lazinski,D., Grzadienska,E. and Das,A. (1989) *Cell*, **59**, 207–218.
29 Harada,K., Martin,S.S. and Frankel,A.D. (1996) *Nature*, **380**, 175–179.
30 Puglisi,J.D., Chen,L., Frankel,A.D. and Williamson,J.R. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3680–3684.
31 Aboul-ela,F., Karn,J. and Varani,G. (1995) *J. Mol. Biol.*, **253**, 313–332.
32 Nikonowicz,E.P., Sirr,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) *Nucleic Acids Res.*, **20**, 4507–4513.
33 Batey,R.T., Battiste,J.L. and Williamson,J.R. (1995) *Methods Enzymol.*, **261**, 300–322.
34 Varani,G., Aboul-ela,F. and Allain,F.H.-T. (1996) *Prog. NMR Spectrosc.*, **29**, 51–127.
35 Allain,F.H.-T. and Varani,G. (1995) *J. Mol. Biol.*, **250**, 333–353.
36 Varani,G., Aboul-ela,F., Allain,F.H.-T. and Gubser,C.C. (1995) *J. Biomol. NMR*, **5**, 315–320.
37 Jucker,F.M. and Pardi,A. (1995) *Biochemistry*, **34**, 14416–14427.
38 Varani,G. and Tinoco,I.,Jr (1991) *Q. Rev. Biophys.*, **24**, 479–532.
39 King,G.C., Harper,J.W. and Xi,Z. (1995) *Methods Enzymol.*, **261**, 436–450.
40 Brünger,A.T. (1990) *X-PLOR Manual.* Yale UniversityPress, New Haven, CT.
41 Diamond,R. (1992) *Protein Sci.*, **1**, 1279–1287.
42 Wimberly,B.T. (1992) Ph.D. Thesis, University of California, Berkeley, CA.
43 Wimberly,B., Varani,G. and Tinoco,I.,Jr (1993) *Biochemistry*, **32**, 1078–1087.
44 Avis,J., Allain,F.H.T.A., Howe,P.W.A.H., Varani,G., Neuhaus,D. and Nagai,K. (1996) *J. Mol. Biol.*, **257**, 398–411.
45 Sumner-Smith,M., Roy,S., Barnett,R., Reid,L.S., Kuperman,R., Delling,U. and Sonenberg,N. (1991) *J. Virol.*, **65**, 5196–5202.
46 Weeks,K.M. and Crothers,D.M. (1992) *Biochemistry*, **31**, 10281–10287.
47 Tao,J. and Frankel,A.D. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 1571–1575.
48 Aboul-ela,F., Murchie,A.I.H., Homans,S.W. and Lilley,D.M.J. (1993) *J. Mol. Biol.*, **229**, 173–188.
49 Bhattacharyya,A. and Lilley,D.M.J. (1989) *Nucleic Acids Res.*, **17**, 6821–6840.
50 Weisenseel,J.P., Moe,J.G., Reddy,G.R., Marnett,L.J. and Stone,M.P. (1995) *Biochemistry*, **34**, 50–64.
51 Allain,F.H.-T. and Varani,G. (1996) submitted.
52 Riordan,F.A., Bhattacharya,A., McAteer,S. and Lilley,D.M.J. (1992) *J. Mol. Biol.*, **226**, 305–310.
53 Zacharias,M. and Hagerman,P.J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6052–6056.
54 Lilley,D.M.J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7140–7142.
55 Zacharias,M. and Hagerman,P.J. (1995) *J. Mol. Biol.*, **247**, 486–500.
56 Bhattacharya,A., Murchie,A.I.H. and Lilley,D.M.J. (1990) *Nature*, **343**, 484–487.
57 Hsieh,C.-H. and Griffith,J.D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4833–4837.
58 Gohlke,C., Murchie,A.I.H., Lilley,D.M.J. and Clegg,R.M. (1994) *Proc. Natl Acad Sci, USA*, **91**, 11660–11664.
59 Dougherty,D.A. (1996) *Science*, **271**, 163–168.
60 Bayer,P., Kraft,M., Ejchart,A., Westendorp,M., Frank,R. and Rösch,P. (1995) *J. Mol. Biol.*, **247**, 529–535.
61 Mujeeb,A., Parslow,T., Yuan,Y.-C. and James,T.L. (1996) *J. Biomol. Struct.*, **13**, 649–660.
62 Tan,R. and Frankel,A.D. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5282–5286.
63 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. and Colonno,R.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8985–8989.
64 Graham,G.J. and Maio,J.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5817–5821.
65 Sullenger,B.A., Gallardo,H.F., Ungers,G.E. and Gilboa,E. (1990) *Cell*, **63**, 601–608.
66 Sullenger,B.A., Gallardo,H.F., Ungers,G.E. and Gilboa,E. (1991) *J. Virol.*, **65**, 6811–6816.