Specific HIV-1 TAR RNA Loop Sequence and Functional Groups Are Required for Human Cyclin T1−Tat−TAR Ternary Complex Formation†

Sara Richter, Hong Cao, and Tariq M. Rana*

Chemical Biology Program, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605-2324

Received November 19, 2001; Revised Manuscript Received March 26, 2002

ABSTRACT: Replication of human immunodeficiency virus requires Tat protein which activates elongation of RNA polymerase II transcription at the HIV-1 promoter through interaction with the cyclin T1 (CycT1) subunit of the positive transcription elongation factor complex (P-TEFb). Tat binds directly through its transactivation domain to the CycT1 subunit of the P-TEFb and induces loop sequence specific binding of the P-TEFb onto nascent HIV-1 TAR RNA. By using a gel electrophoresis method and a comprehensive set of TAR loop mutants, we have identified the sequence and structural determinants for high-affinity CycT1−Tat−TAR ternary complex formation. Our results show that CycT1 and Tat binding to TAR RNA is highly cooperative, and a capacity of 85%, a Hill coefficient of 2.7, and a dissociation constant (K_D) of 2.45 nM were observed. These results indicate that there are three binding sites on TAR RNA. CycT1 does not bind TAR RNA in the absence of Tat, and Tat binding to TAR, while detectable, is very inefficient in the absence of CycT1. It is conceivable that the CycT1−Tat heterodimer directly binds to TAR RNA in the U-rich RNA bulge region and this binding facilitates the interactions of the CycT1−Tat heterodimer at the other two sites in the RNA loop region. On the basis of our results, we suggest a model where CycT1 interacts with Tat protein and positions the protein complex to make contacts with the G34 region of the loop sequence; G34 is critical for CycT1−Tat binding and forms a C30·G34 base pair. Two functional groups, O6 and N7, at nucleotide positions 32 and 34 in the TAR loop are essential for CycT1−Tat interactions with TAR RNA. The identity of two nucleotides, U31 and G33, is not critical, but they contribute to the stabilization of the RNA−protein complex. The presence of a single-nucleotide bulge of A35 or C35 is essential for distortion of the backbone RNA structure as well as the accessibility of functional groups in the major groove of the double-helical region. CycT1−Tat interaction with TAR RNA represents another example of the flexibility and complexity of RNA structure involved in protein recognition.

The human immunodeficiency virus (HIV-1)[1] encodes a transcriptional activator protein, Tat, which is expressed early in the viral life cycle and is essential for viral gene expression, replication, and pathogenesis (for reviews, see refs *1*−*6*). Tat enhances the processivity of RNA Pol II elongation complexes that initiate in the HIV long terminal repeat (LTR) region. In nuclear extracts, HIV-1 Tat associates tightly with the CDK9-containing positive transcription elongation factor complex, P-TEFb (*7*−*9*). Recent studies indicate that Tat binds directly through its transactivation domain to the cyclin subunit (CycT1) of the P-TEFb and induces loop sequence specific binding of the P-TEFb to TAR RNA (*10*−*12*). Recruitment of the P-TEFb to TAR has been proposed to be both necessary and sufficient for activation of transcription elongation from the HIV-1 long terminal repeat promoter (*13*). Neither CycT1 nor the P-TEFb binds TAR RNA in the absence of Tat, and thus, the binding is highly cooperative for both Tat and the P-TEFb (*10*, *12*). Tat appears to contact residues in the carboxyl-terminal boundary of the CycT1 cyclin domain which are not critical for binding of cyclin T1 to CDK9 (*11*, *13*−*20*), and basic residues in CycT1 (R251 and R254) further stabilize the Tat−P-TEFb−TAR RNA complex (*11*). Mutagenesis studies showed that the CycT1 sequence containing amino acids 1−303 was sufficient to form complexes with the Tat−TAR heterodimer and CDK9 (Figure 1) (*11*, *13*−*20*). Thus, the assembly of this complex appears to involve a series of adaptive interactions between the transactivation and arginine-rich motif (RNA-binding) domains of Tat and their respective protein (CycT1) and nucleic acid (TAR) partners during transcription.

It is well established that Tat protein binds the TAR (transactivation responsive) RNA element, a 59-base stem−loop structure located at the 5′ end of all nascent HIV-1 transcripts (*21*). TAR RNA was originally localized to nucleotides 1−80 within the viral long terminal repeat (LTR) (*22*). Subsequent deletion studies have established that the region from nucleotide 19 to 42 incorporates the minimal domain that is both necessary and sufficient for Tat responsiveness in vivo (*23*−*25*). TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge

† This work was supported by National Institutes of Health Grant AI 41404.

* To whom correspondence should be addressed. Phone: (508) 856-6216. Fax: (508) 856-8015. E-mail: tariq.rana@umassmed.edu.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; TAR, transactivation response element; CycT1, human cyclin T1 protein; P-TEFb, positive transcription elongation factor complex.

10.1021/bi0159579 CCC: $22.00
Published on Web 04/24/2002

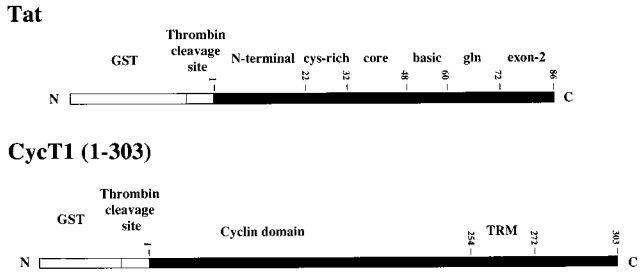

FIGURE 1: Schematic representation of GST fusion proteins showing various domains and regions of HIV-1 Tat and human cyclin T1. These fusion proteins contain an N-terminal GST moiety followed by a thrombin cleavage site. After expression, the GST fusion proteins were cleaved, purified, and characterized as described previously (*10*, *11*). The Tat−TAR recognition motif (TRM) of CycT1 consists of amino acids 254−272 at the carboxyl-terminal edge of the cyclin domain (*10*, *11*).

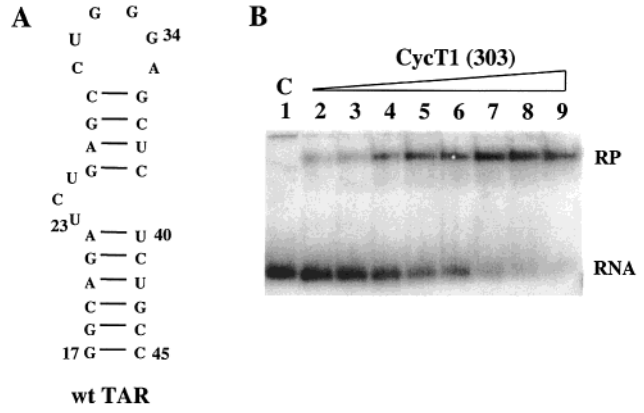

FIGURE 2: (A) Secondary structure of TAR RNA used in this study. TAR RNA spans the minimal sequences that are required for Tat responsiveness in vivo (*23*) and for in vitro binding of Tat-derived peptides (*32*). (B) Binding of recombinant wild-type Tat and CycT1 (amino acids 1−303) to HIV-1 TAR RNA. RNA−protein complex formation and analysis were performed as described in Experimental Procedures. The amount of Tat protein was fixed at 7 pmol in all lanes, and the amount of CycT1 (residues 1−303) was 0, 11, 17, 23, 28, 34, 51, 69, and 103 pmol in lanes 1−9, respectively. Lane C was a control lane without the proteins. CycT1−Tat−TAR complexes are shown as RP.

which separates two helical stem regions (Figure 2A) (*22*, *23*, *26*, *27*). The trinucleotide bulge is essential for high-affinity and specific binding of the Tat protein (*28*, *29*). The loop region is required for in vivo transactivation but is not involved in Tat binding (*29*−*33*).

While the interaction between Tat and TAR has been extensively studied and the nucleic acid structural requirements for Tat−TAR complex formation are known, we have little knowledge about the structural and functional components in TAR loop which are involved in the TAR RNA interaction with the Tat−P-TEFb. In this report, we have used a gel electrophoresis method to measure relative binding affinities of TAR RNA sequence variants as compared to the wild-type sequence for CycT1−Tat−TAR complex formation. On the basis of analysis of a comprehensive set of TAR loop mutants, we have identified the sequence and structural determinants for high-affinity CycT1−Tat−TAR ternary complex formation.

EXPERIMENTAL PROCEDURES

RNA Synthesis. Wild-type and mutant TAR RNAs were prepared by in vitro transcription (*34*, *35*). RNA containing 2-aminopurine was chemically synthesized and purchased from Dharmacon Research Inc. (Boulder, CO). For transcription reactions, all DNAs were synthesized on an Applied Biosystems ABI 392 DNA/RNA synthesizer. The template strands encode the sequences for wild-type and mutant TAR RNAs. The top strand is a short piece of DNA complementary to the 3′ end of all template DNAs having the sequence 5′TAATACGACTCACTATAG3′. The template strand of DNA was annealed to an equimolar amount of top strand DNA, and transcriptions were carried out in transcription buffer [40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 0.01% Triton X-100, and 5 mM DTT] and 4.0 mM NTPs at 37 °C for 2−4 h. For reaction mixtures (20 μL) containing 8.0 pmol of template DNA, 40−60 units of T7 polymerase (Promega) was used. Transcription reactions were stopped by adding an equal volume of sample loading buffer. RNA was purified on 20% polyacrylamide−8 M urea denaturing gels and stored in a DEPC/water mixture at −20 °C.

Enzymatically transcribed RNAs were 5′-dephosphorylated by incubation with calf intestinal alkaline phophatase (Promega) for 1 h at 37 °C in 50 mM Tris-HCl (pH 9.0), 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, and 1 mM spermidine. The RNAs were purified by multiple extractions with Tris-saturated phenol and one extraction with a 24:1 chloroform/isoamyl alcohol mixture followed by ethanol precipitation. The RNAs were 5′-end-labeled with 0.5 μM [γ-^{32}P]ATP (6000 Ci/mmol) (ICN) per 100 pmol of RNA by incubating it with 16 units of T4 polynucleotide kinase (New England Biolabs) in 70 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, and 5 mM DTT (*35*, *36*). The 5′-end-labeled RNAs were gel purified on a denaturing gel, visualized by autoradiography, eluted out of the gels, and desalted on a reverse-phase cartridge. The sequence of RNAs was determined by base hydrolysis and nuclease digestion.

Tat and Cyclin T1 Proteins. Tat (amino acids 1−86) and cyclin T1 (amino acids 1−303) were expressed in the DH5α strain of *Escherichia coli* as glutathione *S*-transferase (GST) fusion proteins. These fusion proteins consisted of an N-terminal GST moiety followed by a thrombin cleavage site (Figure 1). After expression, the GST fusion proteins were cleaved, purified, and characterized as described previously (*10*, *11*).

Electrophoretic Mobility Shift Assays. RNA−protein binding reaction mixtures contained 50 fmol of 5′-^{32}P-end-labeled RNA, 7 pmol of Tat protein, and increasing amounts of CycT1 (residues 1−303) as shown in Figure 2B. Complex formation was performed in TK buffer and incubation at 4 °C for 1.5 h. TK buffer contains 50 mM Tris-HCl (pH 7.4), 20 mM KCl, and 0.1% Triton X-100. Complexes were separated from unbound RNA by electrophoresis in non-denaturing 10% polyacrylamide gels containing 0.1% Triton X-100. Gels were run at 4 °C and 350 V for 1.5 h. The relative amounts of free and bound RNA were determined by phosphorimaging.

RESULTS

Binding of CycT1 and Tat to TAR RNA Is Highly Cooperative. To understand the thermodynamics of CycT1−Tat−TAR ternary complex formation, we synthesized TAR RNA sequences containing wild-type loop and various mutants, 5′-end-labeled the RNAs, and tested their abilities

for Tat and CycT1 binding by electrophoretic mobility shift assays. The fractional saturation of 5′-^{32}P-end-labeled TAR RNA and Tat protein was measured as a function of CycT1 (amino acids 1−303). The cyclin T1 sequence (amino acids 1−303) instead of the P-TEFb was used due to three reasons. (a) This region of CycT1 is sufficient for the formation of stable complexes with the Tat−TAR heterodimer and CDK9 (*11*, *13*−*20*). (b) The kinase subunit CDK9 does not interfere with the CycT1 specificity for Tat−TAR binding (*11*, *13*−*20*). (c) To study quantitative binding efficiencies of a ternary complex such as the Tat−TAR (wild type and mutants)−CycT1 complex, we needed to simplify the system to obtain consistent and reliable results. A typical gel of these experiments is shown in Figure 2B. In the absence of CycT1, the Tat−TAR complex was not observed under the low Tat concentrations used in these experiments. It is important to note that Tat−TAR binding is not efficient in the absence of CycT1. The electrophoretic mobility shift corresponding to the CycT1−Tat−TAR complex was characterized by comparing RNA−protein complexes containing TAR RNA with a Tat (amino acids 38−72) peptide or full-length Tat protein (data not shown). Increasing concentrations of CycT1 resulted in an increased amount of CycT1−Tat−TAR ternary complex formation. The amount of complex formation was measured by the ratio of protein-bound TAR to total RNA.

The experimental data were fitted by using the cooperative binding equation (multiple binding sites of different affinities) based on Hill's theory (*37*).

$$y = \frac{C^n \mathrm{Cap}}{K_D{}^n + C^n} \quad (1)$$

where y is the amount of the CycT1−Tat complex bound to TAR, the capacity for RNA−protein binding, Cap, is the maximum binding ability of TAR, K_D is the dissociation constant that equals the concentration of CycT1 at half the binding ability, C is the concentration of free CycT1 in the solution, and n is a measure of the cooperativity which is generally called the Hill coefficient. In real systems with positive cooperativity, n is always less than the number of maximum possible binding sites. The closer the quantity n approaches the maximum possible number of binding sites, the stronger the cooperativity is.

As shown in Figure 4, a sigmoidal curve was observed for CycT1 and Tat binding to wild-type TAR RNA. For CycT1 binding to Tat and wild-type TAR RNA, a capacity of 85.5 ± 0.5%, a Hill coefficient of 2.7, and a dissociation constant (K_D) of 2.45 ± 0.02 nM were found. On the basis of these results, we conclude that CycT1 and Tat binding to TAR RNA is highly cooperative and there are three binding sites in the RNA structure.

Nucleotides at Positions 32 and 34 in the TAR Loop Are Critical in CycT1−Tat Binding to TAR RNA. To define the role of the loop sequence of TAR RNA for CycT1−Tat−TAR complex formation, we determined the effect of each base substitution in the loop region on the affinity of the CycT1−Tat complex for TAR. The effect of each nucleotide mutation in the TAR loop was assessed by evaluating the efficiency with which mutant RNAs are bound by the CycT1−Tat complex. Protein binding capabilities of various TAR loop mutants (Figure 3) were compared to that of the

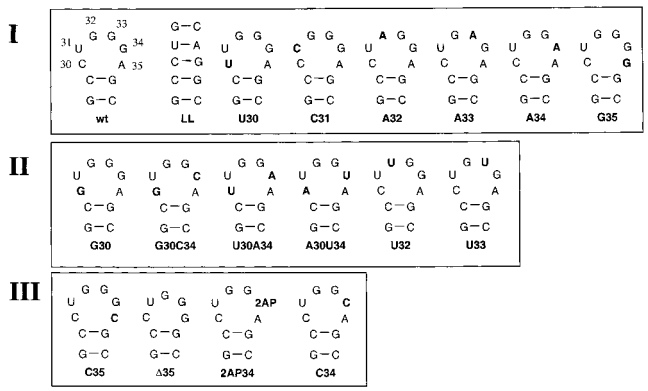

FIGURE 3: Sequence of the loop region (nucleotides 28−37) of wild-type and mutant TAR RNAs. Nucleotide mutations are shown in bold. TAR RNAs are named according to the position of mutation. There were three sets of mutant RNAs designed for our experiments. (I) Wild-type TAR and TAR without the loop residues (LL), and first set of mutants: single-base mutations from position 30 to 35. (II) C30 replaced with G30, compensatory mutations introduced to form a putative base pair between nucleotides at positions 30 and 34, and single-point mutations at position 32 and 33. (III) Mutations to define the role of chemical groups of nucleotides 34 and 35 for protein recognition. Two point mutants are C35 and C34. 2-Aminopurine and deletion of A35 are shown as 2AP34 and Δ35, respectively.

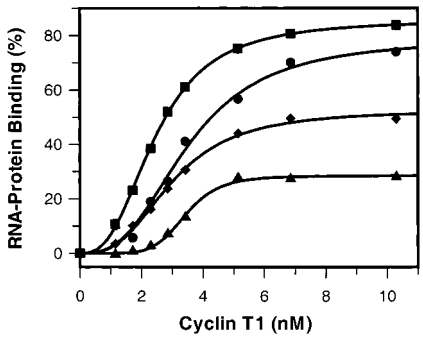

FIGURE 4: CycT1 and Tat binding to wild-type and mutant TAR RNA sequences. The efficiency of the ternary complex formation was plotted against CycT1 concentration. The experimental data were fitted by using the cooperative binding equation (eq 1) (multiple binding sites of different affinities) based on Hill's theory (*37*) as described in the text. The maximum protein binding abilities of various TAR RNA constructs were calculated from these experiments. Curve fittings for wild-type TAR RNA (■) and three mutants A33 (●), U32 (◆), and Δ35 (▲) are shown.

wild-type TAR sequence. We also synthesized TAR RNA where a loop sequence was deleted (loopless), and this RNA served as a negative control (Figure 3).

Four examples of data fitting for various TAR RNA sequences are shown in Figure 4. It is obvious that the capacities of mutant RNAs were lower than that of the wild-type TAR RNA, and the A35 deletion mutant (Δ35) RNA exhibited the lowest efficiency of binding to CycT1 and Tat. The total amount of bound RNA varied as a function of nucleotide mutation in the TAR loop. Results of these experiments are summarized in Table 1.

As an initial strategy, we replaced each nucleotide in the loop from position 30 to 35 with the corresponding purine or pyrimidine base (C → U, G → A, U → C, and A → G) and measured the capacity for the CycT1−Tat complex as described above (Figure 4). A U31 → C31 mutation in TAR exhibited CycT1−Tat binding affinities similar to that of

Table 1: Protein Binding Capacities of RNA Sequences[a]

	first set		second set		third set	
group	TAR	capacity (%)	TAR	capacity (%)	TAR	capacity (%)
A	wild type	85.5 ± 0.5	wild type	85.5 ± 0.5	wild type	85.5 ± 0.5
	U31 → C31	83 ± 6				
					A35 → C35	83 ± 5
	G33 → A33	79 ± 5				
			C30 → G30	74 ± 4		
B			G33 → U33	67 ± 5		
	A35 → G35	65 ± 4				
	C30 → U30	59 ± 1	C30·G34 → A30·U34	59 ± 6		
			C30·G34 → U30·A34	57 ± 3		
			G32 → U32	53 ± 1		
C	G32 → A32	39[b]				
					G34 → 34AP	37 ± 2
	G34 → A34	33 ± 4	C30·G34 → G30·C34	33 ± 3		
					Δ35	28.4 ± 0.6
D					G34 → C34	nd
	LL	nd	LL	nd	LL	nd

[a] Wild-type and three sets of mutant TAR RNAs arranged in four groups by decreasing efficiencies of maximal RNA−protein binding [capacity (%)]. Maximal capacities were calculated as described in the legend of Figure 4. [b] The maximal level of RNA−protein binding was calculated by measuring the level of RNA−protein complex formation at saturating protein concentrations without fitting the data to the cooperative binding model. nd represents nondetectable cooperative RNA−protein binding.

wild-type TAR RNA. Three other point mutations (C30 → U30, G33 → A33, and A35 → G35) did not significantly affect the binding affinities of the CycT1−Tat−TAR complex and exhibited a capacity of more than 50%. In contrast, changing either G34 → A34 or G32 → A32 drastically decreased the level of CycT1−Tat binding and resulted in capacities of ≈30 and ≈40%, respectively. These results indicate that single mutations at positions 30, 31, 33, and 35 lower the binding affinities but retain more than 50% capacity, while mutations at position 32 and 34 drastically decrease the level of CycT1−Tat binding to TAR RNA (Table 1). We conclude that G32 and G34 are essential for CycT1−Tat binding to TAR RNA.

The CycT1−Tat Complex Binds to a Specific TAR Loop Structure. The secondary structure of TAR RNA contains a six-nucleotide sequence, which has a flexible structure that is difficult to determine by NMR spectroscopy. What is the functional conformation of the TAR RNA loop region? Do the mutations introduced in the TAR loop sequence affect the secondary structure of the RNA that is recognized by proteins, or do they merely replace the functional groups that are essential for protein recognition? Given the loop sequence rich in G residues and a single C, we hypothesize that a C·G base pair could exist in the loop between nucleotides G32, G33, or G34 and C30. In this case, disrupting putative base pairs would completely change TAR secondary structure, thus reducing the level of CycT1−Tat−TAR complex formation. Our results show that mutations G32 → A32 and G34 → A34 have the most significant effect on the binding affinities for the protein complex, suggesting that G32 and G34 are the most likely candidates to form a base pair with C30. This notion was further supported by the observation that mutations allowing the formation of putative C30·G32 and C30·G34 base pairs have the lowest impact on protein binding (Table 1). It is interesting that substitution of C30 with U30 did not affect RNA−protein binding as was observed in the G32 → A32 and G34 → A34 mutations. These results suggested that a C·G base pair in the loop is important for protein binding and a less stable U·G base pair is sufficient for formation of a functional RNA structure.

To test our hypothesis of a C30·G32 or C30·G34 base pair in the CycT1−Tat−TAR complex, we designed a second set of TAR loop mutants (Figure 3, panel II). We first chose to investigate C30·G34 base pair possibility for two reasons. (a) G32 is close to C30 in TAR sequence and may not form a stable base pair as compared to the other G residue. (b) The G34 → A34 mutation resulted in the lowest yield of RNA−protein complex formation (30% capacity). To test the C30·G34 base pair, we introduced three different compensatory mutations to preserve a base pair in the loop (C30·G34 → G30·C34, C30·G34 → A30·U34, and C30·G34 → U30·A34) and a fourth mutation (C30 → G30) where no C·G base pair was possible within the loop. These mutant RNAs were tested for their abilities to bind the CycT1 (residues 1−303)−Tat complex, and the results are shown in Table 1. TAR mutants containing C30·G34 → A30·U34 and C30·G34 → U30·A34 putative base pairs showed more than 50% capacity; however, the C30·G34 → G30·C34 mutant formed RNA−protein complexes with only 30% capacity. Interestingly, the fourth mutant (C30 → G30) where no C·G base pair was possible within the loop showed no drastic effect on protein binding affinities (70% capacity). These results demonstrate that G34 is essential for a functional TAR loop and a substitution at this position with A or U is tolerated only if a complementary base is present at position 30 (see the Discussion).

To test the possibility of the C30·G32 base pair, we transcribed the G32 → U32 mutant and tested its efficiency for protein binding. As shown in Table 1, the G32 → U32 mutant was able to form a complex with 50% capacity. Another mutant, C30 → G30, that would also disrupt the C30·G32 base pair did not dramatically affect protein binding (70% capacity). These results indicate that the C30·G32 base pair is not likely to exist in the CycT1−Tat−TAR complex.

Substitution of a purine G33 with A33 in TAR RNA did not influence protein binding (Table 1) because 70% capacity was observed. We also tested the G33 → U33 mutation that

Table 2: RNA Functional Group Requirement for Protein Binding[a]

TAR	capacity (%)	putative base pair	O4 or O6	N7
wild type	85.5 ± 0.5	yes (C30·G34)	yes	yes
C30 → G30	74 ± 4	no	yes	yes
C30 → U30	59 ± 1	yes (U30·G34)	yes	yes
C30·G34 → U30·A34	57 ± 3	yes (U30·A34)	no	yes
C30·G34 → A30·U34	67 ± 3	yes (A30·U34)	yes	no
G34 → 34AP	37 ± 2	no	no	yes
G34 → A34	33 ± 4	no	no	yes
C30·G34 → G30·C34	33 ± 3	yes (G30·C34)	no	no
G34 → C34	nd	no	no	no
LL	nd			

[a] Wild-type and mutant TAR RNAs with substitutions at position 30 and/or at position 34 arranged by decreasing efficiencies of maximal RNA protein binding [capacity (%)]. The presence (yes) or absence (no) of putative base pairs and CO and N7 functional groups at position 34 is indicated.

showed no significant decrease in the level of protein binding, confirming that position 33 is not critical for protein recognition.

O6 and N7 Functional Groups at Nucleotide Positions 32 and 34 in the TAR Loop Are Required for the CycT1−Tat−TAR Interaction. Our results presented above show that a significant loss in the extent of RNA−protein complex formation was observed when G32 or G34 was substituted with A. As shown in Figure 3, G34 was substituted with the four natural ribonucleotides. G is the wild-type base; the U30·A34 and A30·U34 mutants showed half of the binding affinities of wild-type TAR, and the C30·G34 → G30·C34 mutant retained only 33% capacity.

We then examined the chemical nature of the ribonucleotides in TAR that could be responsible for the observed loss in protein binding affinities (Table 2). G has an O6 from a carbonyl group, which is a hydrogen bond acceptor, and electron donor nitrogen at position 7 in the purine ring. As compared to G, U and A have only one of the two above-mentioned groups (A contains N7 and U has O4). In contrast, C exposes a hydrogen bond donor amino group in place of the carbonyl moiety and does not contain N7. We reasoned that if O6 of the carbonyl group was critical for protein interactions, a TAR RNA substituted with 2-aminopurine at position 34 (G34 → 2AP34) would show lower binding affinities for the CycT1−Tat complex than wild-type TAR RNA. The nucleotide 2-aminopurine contains N7 but lacks O6 groups. We chemically synthesized TAR RNA containing 2-aminopurine at position 34 and tested for CycT1−Tat binding. Like the A34 mutant (Table 1), 34AP TAR exhibited moderate protein binding affinities (≈30% capacity), which were higher than those of the loopless TAR, indicating that the presence of N7 at position 34 restores partial interactions with proteins. Mutation at G34 with C34 that lacks both N7 and O6 resulted in a drastic loss in the level of RNA−protein complex formation, which was similar to that with loopless RNA. These data indicate that O6 and N7 at position 34 in TAR loop are essential for protein contacts.

Similarly, when G32 was substituted with A and U, modest protein binding affinities were observed (50% capacity for U and 40% capacity for A), which were higher than that of the loopless TAR, indicating that the presence of N7 or O4 at position 32 restores partial interactions with proteins. Since substitution with U32 gives considerably larger amounts of protein-binding RNA than A32 mutant, the C═O group appears to be important in this face of the TAR loop.

A Nucleotide at Position 35 Is Required, but the Identity Is Not Important for CycT1−Tat−TAR Complex Formation. During our initial studies, we found that the A35 → G35 mutation lowered the protein binding affinities (Table 1). There is a possibility that the guanosine residue in G35 could form a base pair with C30 and thus lower the binding affinity for the protein. To address this question, we tested another A35 mutant. Substitution of A35 with C35 showed 80% capacity that is similar to wild-type TAR binding affinities, indicating that the CycT1−Tat complex does not make contacts with specific functional groups of the RNA at position 35. We next determine if the presence of a nucleotide at position 35 is necessary for formation of a functional TAR loop structure. TAR RNA where A35 was deleted (Δ35) was transcribed, 5′-end-labeled, and tested for CycT1−Tat binding. As shown in Figure 4 and Table 1, Δ35 TAR poorly bound the CycT1−Tat complex. Our results show that A35 in the TAR loop acts as a spacer nucleotide that is essential for functional RNA structure and may not be involved in direct protein interactions.

DISCUSSION

By using a comprehensive set of TAR RNA mutants, we have defined essential RNA loop residues and functional groups that are required for formation of a stable CycT1−Tat−TAR ternary complex. Results of these experiments are summarized in Tables 1 and 2. To define mutant RNA function, we have divided the RNA−protein binding affinities into four groups: group A with 70−85% capacity, group B with 50−70% capacity, group C with partial protein binding, and group D with nonfunctional RNA sequences (Table 1).

Our results highlight interesting features of the TAR RNA loop and indicate that the nucleotide at position 34 plays a critical role in the RNA interactions with the protein complex. While the G34 → C34 mutation abolishes RNA−protein binding, the C30·G34 → G30·C34 mutation partially restores RNA−protein complex formation, suggesting that the presence of the G30·C34 base pair stabilizes the loop structure, allowing the CycT1−Tat complex to contact other loop residues. We further investigated the functional group requirements in the TAR loop for protein recognition (Table 2). G34 in wild-type TAR RNA has both the O6 and N7 groups exposed in the major groove side of the nucleotide, which raises the possibility that G34 could base pair with C30. However, the base pair between residues at positions 30 and 34 does not seem to be a stringent requirement for a stable interaction with the protein complex in the presence of both O6 and N7 groups at nucleotide 34 because a slight decrease in protein binding affinity is observed in the C30 → G30 mutant (which does not allow base pairing to position 34) and the C30 → U30 mutant (which replaces a putative three-hydrogen bond base pair with a weaker two-hydrogen bond base pair between U30 and G34).

The affinity of TAR RNA for the protein complex does not drop significantly in the C30·G34 → U30·A34 and C30·G34 → A30·U34 mutants which indicates that RNA is functional as long as one of the three (O6, O4, and N7) functional groups is present at position 34 and the loop

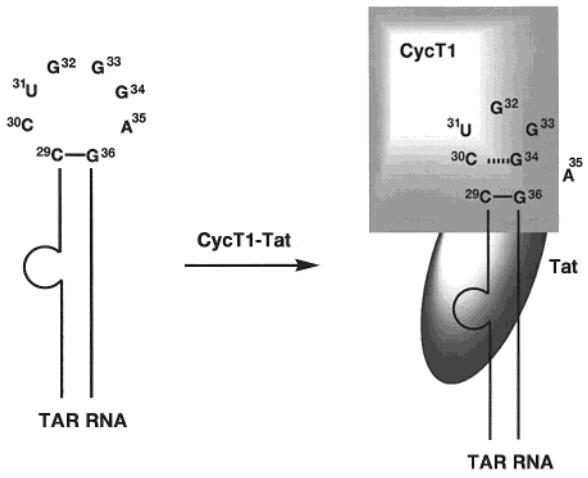

FIGURE 5: Schematic representation of the CycT1−Tat−TAR ternary complex. See the text for details.

structure is stabilized by a putative base pair formed between nucleotides 30 and 34. Interestingly, the presence of major groove functional groups does not contribute significantly when the putative base pairing of nucleotides 30 and 34 is removed because the G34 → A34 and G34 → 34AP mutants did not bind the protein complex with high affinities. On the other hand, the presence of the G30·C34 putative base pair was not sufficient for formation of a high-affinity complex with CycT1−Tat, indicating that O6 and N7 at position 34 are essential for protein recognition. Finally, deletion of O6 and N7 moieties along with deletion of the 30·34 base pair produce a mutant (G34 → C34) RNA which is not able to bind the protein complex. These results therefore indicate that major groove functional groups, O6 and N7 at position 34 in the wild-type TAR loop, are essential for protein binding, and a base pair at positions 30 and 34 further stabilizes the RNA−protein structure. It is intriguing that the O4 functional group from a U at position 34 can compensate for O6 of the G residue in protein binding, suggesting that the TAR loop has a flexible structure and specific functional groups may not be rigidly displayed at specific sites.

In the presence of a C30·G34 base pair, position 35 assumes a single-nucleotide bulge structure, which probably acts as a spacer to distort the RNA backbone and gives accessibility to the major groove at position 34 where the two critical parameters lie (O6 and N7). In fact, our results show that the identity of nucleotide 35 is not important, but a deletion of this position is not allowed.

The nucleotide at position 32 plays a key role in protein association, and our data suggest that a C═O group is involved in making contacts with the protein complex. Finally, positions 31 and 33 did not reveal a critical role, but they could be necessary for stabilization of the interaction of TAR with the CycT1−Tat complex.

CONCLUSIONS

In light of the results of our experiments presented above and previous Tat−TAR studies (*21*), we propose a model for the CycT1−Tat−TAR ternary complex (Figure 5). TAR RNA has a flexible structure, and the TAR loop adopts a number of conformations, including a C30·G34 base pair and a single-nucleotide, A35, bulge. It is well established that Tat protein binds in the trinucleotide bulge region of TAR RNA. CycT1 interacts with Tat protein, and this complex positions itself to make contacts with the G34 region of the loop sequence; G34 is critical for CycT1−Tat binding and forms a C30·G34 base pair. Two functional groups, O6 and N7, at nucleotide positions 32 and 34 in the TAR loop are essential for CycT1−Tat interactions with TAR RNA. The identities of two nucleotides, U31 and G33, are not critical, but they contribute to the stabilization of the RNA−protein complex. It is also possible that CycT1 alters the conformation of Tat and Tat contacts the loop region in the CycT1−Tat−TAR ternary complex. The presence of a single-nucleotide bulge of A35 is important for distortion of the backbone RNA structure as well as the accessibility of the functional groups in the major groove of the double-helical region. Bulge regions in the RNA helices are known to cause widening of the major groove that results in enhanced accessibility of functional groups for protein contacts (*36*, *38*). These results strongly suggest that the CycT1−Tat complex contacts functional groups in the major groove of RNA, and the A35 bulge helps to increase the accessibility of these functional groups. The CycT1−Tat interaction with TAR RNA represents another example of the flexibility and complexity of RNA structure involved in protein recognition.

It is important to note that CycT1 and Tat binding to TAR RNA is highly cooperative, and a capacity of 85.5 ± 0.5%, a Hill coefficient of 2.7, and a dissociation constant (K_D) of 2.45 ± 0.02 nM were observed. These results indicate that there are three binding sites on TAR RNA. CycT1 does not bind TAR RNA in the absence of Tat, and Tat binding to TAR, while detectable, is very inefficient in the absence of CycT1 (refs *10* and *15* and this study). It is conceivable that the CycT1−Tat heterodimer directly binds to TAR RNA in the U-rich RNA bulge region, and this binding facilitates the interactions of the CycT1−Tat complex at the other two sites in the RNA loop region. Further structural studies will be needed to elucidate the details of CycT1−Tat−TAR interactions.

In summary, our results provide new insight into RNA−protein recognition in the CycT1−Tat−TAR ternary complex, and these results could guide the design and assembly of stable CycT1−Tat−TAR complexes for future structural studies.

ACKNOWLEDGMENT

We thank Dr. Katherine Jones for human cyclin T1 and HIV-1 Tat clones and AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, for providing us various HIV-1 GST−Tat expression vectors which were made available by Dr. Andrew Rice.

REFERENCES

1. Jones, K. A. (1997) *Genes Dev. 11*, 2593−2599.
2. Cullen, B. R. (1998) *Cell 93*, 685−692.
3. Emerman, M., and Malim, M. (1998) *Science 280*, 1880−1884.
4. Jeang, K.-T., Xiao, H., and Rich, E. A. (1999) *J. Biol. Chem. 274*, 28837−28840.
5. Isel, C., and Karn, J. (1999) *J. Mol. Biol. 290*, 929−941.
6. Taube, R., Fujinaga, K., Wimmer, J., Barboric, M., and Peterlin, B. M. (1999) *Virology 264*, 245−253.

7. Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., and Price, D. H. (1997) *Genes Dev. 11*, 2622−2632.
8. Mancebo, H. S. Y., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D., and Flores, O. (1997) *Genes Dev. 11*, 2633−2644.
9. Gold, M., Yang, X., Herrmann, C., and Rice, A. (1998) *J. Virol. 72*, 4448−4453.
10. Wei, P., Garber, M. E., Fang, S.-M., Fischer, W. H., and Jones, K. A. (1998) *Cell 92*, 451−462.
11. Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P., Littman, D. R., and Jones, K. A. (1998) *Genes Dev. 12*, 3512−3527.
12. Garber, M. E., Wei, P., and Jones, K. A. (1998) *Cold Spring Harbor Symp. Quant. Biol. 63*, 371−380.
13. Bieniasz, P. D., Grdina, T. A., Bogerd, H. P., and Cullen, B. R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 7791−7796.
14. Fujinaga, K., Cujec, T., Peng, J., Garriga, J., Price, D., Grana, X., and Peterlin, B. (1998) *J. Virol. 72*, 7154−7159.
15. Bieniasz, P. D., Grdina, T. A., Bogerd, H. P., and Cullen, B. R. (1998) *EMBO J. 17*, 7056−7065.
16. Ivanov, D., Kwak, Y. T., Nee, E., Guo, J., Garcia-Martinez, L. F., and Gaynor, R. B. (1999) *J. Mol. Biol. 288*, 41−56.
17. Bieniasz, P. D., Grdina, T. A., Bogerd, H. P., and Cullen, B. R. (1999) *J. Virol. 73*, 5777−5786.
18. Wimmer, J., Fujinaga, K., Taube, R., Cujec, T., Zhu, Y., Peng, J., Price, D., and Peterlin, B. (1999) *Virology 255*, 182−189.
19. Fujinaga, K., Taube, R., Wimmer, J., Cujec, T., and Peterlin, B. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 1285−1290.
20. Chen, D., Fong, Y., and Zhou, Q. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 2728−2733.
21. Rana, T. M., and Jeang, K.-T. (1999) *Arch. Biochem. Biophys. 365*, 175−185.
22. Rosen, C. A., Sodroski, J. G., and Haseltine, W. A. (1985) *Cell 41*, 813−823.
23. Jakobovits, A., Smith, D. H., Jakobovits, E. B., and Capon, D. J. (1988) *Mol. Cell. Biol. 8*, 2555−2561.
24. Selby, M. J., Bain, E. S., Luciw, P. A., and Peterlin, B. M. (1989) *Genes Dev. 3*, 547−558.
25. Garcia, J. A., Harrich, D., Soultanakis, E., Wu, F., Zmitsuyasu, R., and Gaynor, R. B. (1989) *EMBO J. 8*, 765−778.
26. Muesing, M. A., Smith, D. H., and Capon, D. A. (1987) *Cell 48*, 691−701.
27. Berkhout, B., Silverman, R. H., and Jeang, K. T. (1989) *Cell 59*, 273−282.
28. 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. U.S.A. 86*, 6925−6929.
29. 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.
30. Feng, S., and Holland, E. C. (1988) *Nature 334*, 165−167.
31. Berkhout, B., and Jeang, K.-T. (1989) *J. Virol. 63*, 5501−5504.
32. Cordingley, M. G., La Femina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., Graham, D. J., Nguyen, T. M., Le Grow, K., Gotlib, L., Schlabach, A. J., and Colonno, R. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8985−8989.
33. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., and Sonenberg, N. (1991) *J. Virol. 65*, 5196−5202.
34. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
35. Wang, Z., and Rana, T. M. (1996) *Biochemistry 35*, 6491−6499.
36. Neenhold, H. R., and Rana, T. M. (1995) *Biochemistry 34*, 6303−6309.
37. Hill, T. L. (1985) *Cooperative theory in biochemistry: Steady-state and equilibrium systems*, Springer-Verlag, New York.
38. Weeks, K. M., and Crothers, D. M. (1993) *Science 261*, 1574−1577.

BI0159579