Protein Orientation in the Tat-TAR Complex Determined by Psoralen Photocross-linking*

(Received for publication, April 3, 1996, and in revised form, May 16, 1996)

Zhuying Wang, Xilu Wang, and Tariq M. Rana‡
From the Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey and the Molecular Biology and Biochemistry Graduate Program at Rutgers State University, Piscataway, New Jersey 08854

Replication of human immunodeficiency virus type 1 (HIV-1) requires specific interactions of Tat protein with the trans-activation responsive region (TAR) RNA, a 59-base stem-loop structure located at the 5′-end of all HIV mRNAs. We have used a new method based on psoralen photochemistry to identify a specific contact between a fragment of Tat protein (residues 38–72) and TAR RNA. We synthesized a 35-amino acid fragment containing arginine-rich RNA-binding domain of Tat (38–72), and replaced Arg⁵⁷ with Cys to introduce a unique thiol group (–SH) in our model peptide. A psoralen derivative, which can react with thiol groups, was synthesized and used for specific chemical modification of Cys⁵⁷-Tat-(38–72). We used this psoralen-Tat conjugate (psoralen-Cys⁵⁷-Tat-(38–72)) to form a specific complex with TAR RNA. Upon near-ultraviolet irradiation (360 nm), this synthetic psoralen-peptide cross-linked to a single site in the TAR RNA sequence. The RNA-protein complex was purified and the cross-link site on TAR RNA was determined by RNA sequencing, which revealed that Cys⁵⁷ of Tat is close to U₃¹ of TAR RNA. Our results provide high-resolution proximity and orientation information about Tat-TAR complex. Such psoralen-peptide conjugates provide a new class of probes for sequence-specific protein-nucleic acid interactions and could be used to selectively control gene expression or to induce site-directed mutations.

Human immunodeficiency virus type 1 (HIV-1) encodes a trans-activating regulatory protein, Tat, that is essential for trans-activation of viral gene expression (1–3). HIV-1 Tat protein acts by binding to the trans-activation-responsive region (TAR) RNA, a 59-base stem-loop structure located at the 5′-ends of all nascent HIV-1 transcripts (4–7). Upon binding to the TAR RNA sequence, Tat causes a substantial increase in transcript levels (8–10). The increased efficiency in transcription is possibly by preventing premature termination of the

*This work was supported in part by National Institutes of Health Grant AI 34785 and by the State of New Jersey Commission on Cancer Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ To whom correspondence should be addressed. Tel.: 908-235-4082; Fax: 908-235-4073; E-mail: rana@mbd.rutgers.edu.

The abbreviations used are: HIV-1, human immunodeficiency virus, type I; TAR, trans-activation responsive region; HPLC, high performance liquid chromatography.
Orientation of Tat Protein in the Tat-TAR Complex

RESULTS

Site-specific Incorporation of a Psoralen into Tat-(38-72) Sequence—The experimental strategy for site-specific psoralen conjugation of Tat-(38-72) is outlined in Fig. 2. We introduced a unique cysteine residue in the RNA-binding region of Tat at position 57 during peptide synthesis. A derivative of psoralen (8-((3-iodopropyl-1)oxy)psoralen) was synthesized and used to label the cysteine residue in Tat fragment. Psoralen-Tat conjugate was purified by HPLC and characterized by mass spectrometry. To further characterize and evaluate the binding capabilities of psoralen-peptide conjugate, we determined the dissociation constants of the Tat-(38–72) with those of the wild-type peptide (Tat-(38–72)). Equilibrium dissociation constants for psoralen-peptide complex were determined by measuring the ratios of wild-type Tat-(38–72) to TAR complexes were irradiated (lane 4). Digestion of the RNA-peptide cross-link with Proteinase K (5 units for 30 min at 37°C) resulted in an RNA species with mobility similar to TAR RNA (lane 7). Since the cross-linked RNA-peptide complex is stable to alkaline pH (9.5), high temperature (85°C), and denaturing conditions (8 M urea), we conclude that a covalent bond is formed between TAR RNA and the peptide during cross-linking reaction.

Specificity of the Cross-link Formation—Specificity of the cross-linking reaction was established by competition experiments. Cross-linking reactions were performed in a 15-μl volume containing 0.25 μM of 5'-32P-labeled TAR RNA, 1.0 μM psoralen-Tat peptide, 25 mM Tris-HCl (pH 7.4), 100 mM NaCl, and up to 1.25 μM unlabeled competitor RNA. Cross-linked products were separated by 8 M urea-20% polyacrylamide gels and visualized by phosphorimage analysis. Fig. 4 shows that cross-linking was inhibited by the addition of unlabeled wild-type TAR RNA and not by a mutant TAR RNA lacking the trinucleotide bulge. Therefore, we conclude that formation of a specific RNA-protein complex between TAR RNA and psoralen-Tat is necessary for cross-linking

Site-specific Photocross-linking of Psoralen-Cys57-Tat-(38-72) and TAR RNA Complex. TAR RNA was 5'-end labeled with 32P, and cross-linked products were resolved on 20% polyacrylamide-8 M urea gels and visualized by autoradiography. Psoralen-Cys57-Tat-(38-72) conjugate and RNA-protein cross-link are indicated by Pso-Tat and XL, respectively.

the psoralen-peptide and UV (360 nm) irradiation are required for the formation of a cross-linked RNA-protein complex (see lanes 2 and 5). Further control experiments showed that no cross-linking was observed when RNA and unmodified peptide were irradiated (lane 4). Digestion of the RNA-peptide cross-link with Proteinase K (5 units for 30 min at 37°C) resulted in an RNA species with mobility similar to TAR RNA (lane 7). Since the cross-linked RNA-peptide complex is stable to alkaline pH (9.5), high temperature (85°C), and denaturing conditions (8 M urea), we conclude that a covalent bond is formed between TAR RNA and the peptide during cross-linking reaction.

Specificity of the Cross-link Formation—Specificity of the cross-linking reaction was established by competition experiments. Cross-linking reactions were performed in a 15-μl volume containing 0.25 μM of 5'-32P-labeled TAR RNA, 1.0 μM psoralen-Tat peptide, 25 mM Tris-HCl (pH 7.4), 100 mM NaCl, and up to 1.25 μM unlabeled competitor RNA. Cross-linked products were separated by 8 M urea-20% polyacrylamide gels and visualized by phosphorimage analysis. Fig. 4 shows that cross-linking was inhibited by the addition of unlabeled wild-type TAR RNA and not by a mutant TAR RNA lacking the trinucleotide bulge. Therefore, we conclude that formation of a specific RNA-protein complex between TAR RNA and psoralen-Tat is necessary for cross-linking

Uridine 31 in TAR RNA Cross-links to Psoralen-Cys57-Tat-(38-72)—Mapping of the cross-link site on TAR RNA to single nucleotide resolution was carried out by partial RNase digestion and alkaline hydrolysis of the gel-purified RNA-protein cross-link. Fragment sizes are determined by comparison with RNA oligonucleotides of defined sequence and length generated.
by digesting RNA with RNases T1 and B. cereus. Alkaline hydrolysis of RNA and cross-linked RNA-peptide complex generates a ladder of RNA degradation products. Bands of cross-linked RNA-peptide complexes migrate slower than corresponding free RNA (lane 3 in Fig. 5, A and B). Base hydrolysis of the 5'-end-labeled cross-linked complex (lane 3 in Fig. 5A) results in an RNA ladder in which all fragments up to C30 are resolved. There is an obvious gap in the hydrolysis ladder after C30, indicating that the fragments above C30 from the 5'-end are linked to the psoralen-Tat peptide (Fig. 5A, lane 3) that is not seen with the un-cross-linked RNA (lane 2). Thus, U31 is the 5'-end cross-link site. To define 3'-end boundary of the cross-link site, we purified 3'-end-labeled RNA-protein cross-link and subjected to partial alkaline hydrolysis. Base hydrolysis ladder of 3'-end-labeled cross-link (Fig. 5B, lane 3) produces a ladder in which the fragments from the cross-linked RNA-peptide complex match those from free RNA until G32 from 3'-end. After G32, a clear gap was observed during hydrolysis of 3'-end-labeled cross-link (Fig. 5B, lane 3), while alkaline digestion of 3'-end-labeled TAR RNA resulted in a standard ladder (lane 2). This result indicates that the fragments above G32 from the 3'-end contain Tat peptide. Based on these results, we conclude that U31 of TAR RNA is the only site at which cross-linking occurs.

To further confirm the cross-link site, we transcribed a mutant TAR RNA containing G31 instead of U31 in its sequence (Fig. 1). Since psoralen reacts primarily with uridine in RNA (19), replacement of U31 with G31 in TAR RNA loop should abolish or decrease significantly the cross-link formation between psoralen-Tat and mutant TAR RNA. Results were consistent with this notion and a minor cross-link product (~1.5%) was observed when psoralen-Tat and mutant TAR RNA complex was UV irradiated (Fig. 6, lane 4). On the other hand, an RNA-protein cross-link product with high yields was obtained when RNA-protein complex containing psoralen-Tat and wild-type TAR RNA was UV irradiated (Fig. 6, lane 2). These results establish that psoralen-Cys57-Tat-(38–72) forms a single cross-link product with TAR RNA and cross-linking reaction occurs at U31 in the loop region of the RNA.

**DISCUSSION**

We have used a site-specific cross-linking strategy to determine protein orientation in Tat-TAR complex. Our results establish that Arg57 of Tat-(38–72) is close to uridine 31 in the loop region of TAR RNA.

How does Tat recognize TAR RNA? Several lines of evidence suggest that Tat protein contacts TAR RNA in a widened major groove. In a recent study from our laboratory, we used a rhodium complex, Rh(phen)2phi3+, to probe the effect of bulge bases on the major groove width in TAR RNA (21). Our studies establish two important factors involved in Tat-TAR recognition: (i) there is a correlation between major groove opening and Tat binding. At least a 2-base bulge is required for major groove widening and other conformational changes to facilitate Tat binding. This cannot be accomplished by a single base bulge. (ii) A Tat fragment (42–72) occupies the major groove of TAR RNA and abolishes access of the rhodium complex.

To determine the relative orientation of the nucleic acid and protein in the Tat-TAR complex, we have devised a new method based on psoralen photochemistry (20). We synthesized a 30-amino acid fragment containing arginine-rich RNA-binding domain of Tat-(42–72) and chemically attached a psoralen at the amino terminus. Upon near-ultraviolet irradiation (360 nm), this synthetic psoralen-peptide cross-linked to a single site in TAR RNA sequence. The RNA-protein complex was purified, and the cross-link site on TAR RNA was determined by chem-
that Arg^{52} is a likely candidate for such an interaction. However, we cannot rule out the possibility that other arginines are involved in specific interactions with bulge nucleotides. Another interesting feature of our model involves a straight helix of TAR RNA in Tat-TAR complex, which is required to fit our cross-linking data. This is consistent with a recent study by Zacharias and Hagerman (25) in which they performed transient electric birefringence measurements and showed that TAR RNA bulge introduces a bend of 50° in the absence of Mg^{2+}, which is straightened by the addition of Arg and Tat-derived peptides.

Mutational analyses have shown that sequences in the loop of TAR RNA are required for trans-activation (5, 26) and not for Tat binding (13, 17). The loop may provide the binding site for cellular factor(s) involved in trans-activation (27–30). Tat could also be involved in rearranging the loop structure that can be recognized by cellular factors. Our results show that the COOH-terminal region of RNA-binding domain of Tat is in the close proximity of U31 in TAR RNA sequence, and whether Tat directly interacts with the loop or not remains to be determined.

Fig. 7. Model for Tat-TAR recognition showing protein orientation in the RNA-protein complex. Ribbon structured TAR RNA is shown in five yellow lines and nucleotides in red. Structure of Tat protein is illustrated in a magenta color shaded ribbon. Psoralen placed at position 41 in a Tat fragment cross-linked to U42 (20), while psoralen at position 57 cross-linked to U31. Proximity of three amino acid side chains (cyan color) to the RNA nucleotides is shown: cyan and yellow, Lys^{41} and U42, respectively; cyan and green, Arg^{52} and U31, respectively. The structures of TAR RNA (24) and Tat protein (31) are based on NMR data. Structures of RNA and protein were visualized using Insight II software on an IRIS work station.