Nucleic Acids Research, 1993, Vol. 21, No. 1 ***151–154***

Rotational symmetry in ribonucleotide strand requirements for binding of HIV-1 Tat protein to TAR RNA

Richard W.Barnett, Ulrike Delling, Raya Kuperman, Nahum Sonenberg[1,2] and Martin Sumner-Smith*

Allelix Biopharmaceuticals Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1, [1]Department of Biochemistry and [2]McGill Cancer Centre, McGill University, Montreal, Quebec H3G 1Y6, Canada

Received July 27, 1992; Revised and Accepted November 24, 1992

ABSTRACT

Transactivation of human immunodeficiency virus (HIV) gene expression requires binding of the viral Tat protein to a RNA hairpin-loop structure (TAR) which contains a two or three-nucleotide bulge. Tat binds in the vicinity of the bulge and the two adjacent duplex stems, recognising both specific sequence and structural features of TAR. Binding is mediated by an arginine-rich domain, placing Tat in the family of arginine-rich RNA binding proteins that includes other transactivators, virus capsid proteins and ribosome binding proteins. In order to determine what features of TAR allow Tat to bind efficiently to RNA but not DNA forms, we examined Tat binding to a series of RNA-DNA hybrids. We found that only one specific strand in each duplex stem region needs to be RNA, implying that interaction between Tat and a given stem may be solely or predominantly with one of the two strands. However, the essential strand is not the same one for each stem, suggesting a switch in the bound strand on opposing sides of the bulge.

INTRODUCTION

Transactivation of gene expression is an essential process for productive infection by human immunodeficiency virus (HIV). While several host factors have been implicated in the process, only two viral components are known to be required: the small nuclear protein Tat and a RNA hairpin-loop structure at the 5'-end of all viral transcripts (reviewed in 1–4). Tat binds to TAR in the vicinity of a two or three-nucleotide bulge in the upper stem region (5–12) (Fig 1A). The Tat-binding domain of TAR therefore has three structurally distinct regions: the bulge and the two different stem segments, one below and one above the bulge. Besides the structural features, there are TAR base sequence requirements for Tat recognition. First, an unpaired uridine at the 5'-end of the bulge (U_{23}) is critical. Second, an G-C/A-U box in the upper stem segment (G_{26}-C/A_{27}-U), immediately above the bulge, is also strongly prefered. Since the *N3* position of the essential U_{23} of the bulge cannot be methylated (13), we previously speculated that this base may interact with the G-C pair above it (14). Recently, NMR analysis has suggested an interaction with the A_{27}-U pair (16). In either case, a tertiary structure involving the bulge seems to be critical for Tat recognition.

The TAR-binding domain of Tat has been extensively examined. A stretch of nine, predominantly basic amino acids (residues 49–57, RKKRRQRRR) appears to be the principle determinant (7, 9, 10, 15). However, in contrast to the specific sequence and structural TAR features, the Tat basic domain is very tolerant of changes. In fact, an arginine nonamer (RRRRRRRRR) can effectively replace the basic domain in transactivation experiments (11,15) and synthetic peptides of this sequence preferentially recognize TAR substrates over mutant substrates in which the essential U_{23} of the bulge is replaced with a C (15). However, the arginines in the middle of this stretch (amino acids 52 or 53) are more critical than others (11, 15). Recently, it has been suggested that one of these arginines interact with two phosphates at the base of the bulge (11) and/or with the A_{27}-U pair in the upper stem (16). Thus, the tertiary structure in which the invariant U_{23} of the bulge participates presents an arginine specific site. It has been assumed that the other basic residues (arginines and lysines) interact, presumably by the formation of simple ion pairs, with certain phosphates in the bulge or stem regions to increase affinity. In support of this, we have found that progressive deletion of arginines in a mutant Tat containing the arginine homopolymer substitution reduces transactivation potential in an incremental manner (15). However, the interacting phosphates have not yet been identified. Since Tat is able to bind to RNA but not DNA forms of TAR (5), and the presentation of phosphates in the backbone of RNA duplexes differs substantially from those in DNA duplexes, we used RNA/DNA hybrids to point to critical phosphates. This paper describes a study examining the regions in TAR which must be RNA and cannot be substituted by DNA. The results imply that the critical phosphates of each of the stems may be in only one of the two strands, but that the essential RNA strand switches on opposite sides of the bulge.

* To whom correspondence should be addressed

METHODS

Oligonucleotides were synthesised by standard methods, 5′-end labelled with gamma-^{32}P-ATP and T4 polynucleotide kinase, and purified using a Sephadex G50 spun column to remove non-incorporated radiolabel (13). Oligonucleotides comprising strand 1 or strand 2 (Fig. 1B) were combined as indicated (or left uncombined in controls) and hybridised as follows: 4nM of each oligonucleotide in a total volume of 20ul of buffer [20mM Tris (pH 7.6), 3mM $Mg(OAc)_2$, 400mM NaCl, 1mM DTT] was slowly cooled from 98°C to room temperature. Binding of Tat peptide (amino acids 49−62) at a 5-fold molar excess (19nM) or recombinant full-length Tat (amino acids 1−86, ref. 17), kindly provided as a gift by Dr. Craig Rosen, at a 1,000-fold molar excess (4uM), was assayed by gel retardation as described (15). Quantitative experiments comparing the Kd for binding to TAR 27-mer by Tat 49−62 and 1−86 have demonstrated an increased affinity of about 15−40 fold for the shorter peptide (unpublished).

RESULTS

Full length TAR is typically a 59 nucleotide structure at the 5′-end of all viral transcripts (Fig. 1A). For the initial experiments reported here, we used RNA, DNA and 'mixed' derivatives of two oligoribonucleotides that we have previously reported (13) can be annealed to form a bulge-containing structure which is an effective substrate for Tat binding (Fig. 1B). Wherever DNA was employed 2′-deoxyuridine was substituted for thymidine to avoid complications from possible base-specific requirements. Binding experiments used either full-length, recombinant Tat 1−86, or a synthetic Tat peptide derived from the basic, arginine-rich, RNA-binding domain (Tat 49−62), in a gel retardation assay.

As expected, Tat peptide bound to two annealed RNA strands, but not to two DNA strands (not shown). Tat 49−62 also did not bind to annealed hybrids in which either one of the strands was entirely DNA and the other RNA (not shown), suggesting that at least a region of both strands must be RNA.

To define the required RNA regions, 'mixed' oligonucleotide strands were employed which were RNA in one half and DNA in the other. Annealed structures with one stem composed entirely of DNA, and the other of RNA, did not serve as substrates (Fig. 2A, lanes 9 and 19), showing that at least a region of each stem must be RNA. When hybrid strands were annealed to a complementary, all RNA strand, two of the four combinations were found to bind Tat 49−62 (Fig. 2A, lanes 6 and 16 cf. lanes 12 and 22). Strikingly, the results show that the essential RNA regions are on different strands in the two stems. With respect to the orientation of the strands, both must be RNA 5′ of the bulge, but can be RNA or DNA 3′ of the bulge. The essential RNA regions are therefore rotationally symmetric about the bulge.

These results show that Tat 49−62 is able to recognize certain specific DNA-containing TAR hybrids. In fact, Tat peptide recognises a substrate which has DNA in both stems, provided the essential regions are RNA (Fig. 2B, lane 10). This effect is not unique to Tat peptide. When full-length Tat was used, the same relative preferences were observed (Fig. 2C), although more quantitative analysis demonstrated that any RNA substitution tends to reduce Tat binding somewhat.

To extend the findings to a more TAR-like hairpin-loop structure, three 27-base oligonucleotides (Fig. 1C) were used.

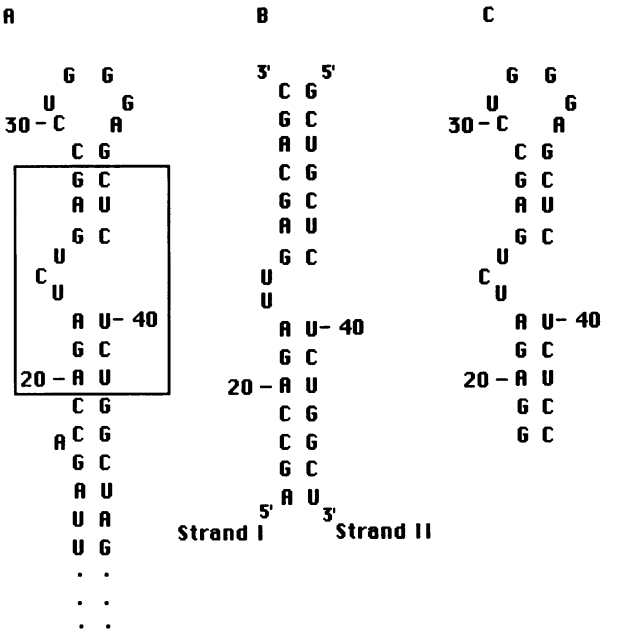

Figure 1. (**A**) The sequence and secondary structure of a typical full-length TAR. The region bound by Tat is highlighted. The substrates used in this study are shown: (**B**) The paired oligonucleotides composed of two 'strands' (strands 1 and 2 as indicated) which can be annealed to form Tat-binding substrates, and (**C**) An abbreviated 27-nucleotide hairpin-loop formed by a single oligonucleotide.

As expected, both Tat 1−86 and 49−62 bound to a full RNA structure (Fig. 2D, lanes 2 and 3), and neither bound to a 27-mer in which both critical regions were replaced with DNA (Fig. 2D, lanes 8 and 9). However, although Tat 49−62 bound to a 27-mer in which the two non-essential regions were DNA (Fig. 2D, lane 5), we were not able to show significant binding by Tat 1−86 (lane 6). Since Tat can recognize the equivalent structure without a loop (Fig. 2C, lane 11), this finding implies that an unfavourable combination of the loop and the adjacent, hybrid stem prevents Tat 1−86 recognition. This is interesting because although the loop is dispensable for *in vitro* binding by Tat 1−86, Harper and Logsdon (18) have recently suggested that Tat 1−86 interacts with components of the loop in a largely sequence-independent interaction.

To map further the requirements for RNA within the critical strands of each stem, combinations of only two or three adjacent nucleotides were substituted with deoxynucleotides. No such short substitution reduced binding to the same degree as an all DNA substitution, but in general, all had some negative effect; the effect being greater the closer the substitution was to the bulge (not shown). Therefore, there does not appear to be any single, essential ribonucleotide, but rather a general preference for RNA over DNA in one of the two strands of each stem. Extensive DNA substitutions are tolerated in the complementary strand.

DISCUSSION

The results reported here demonstrate that the Tat basic domain critically recognises only one of the two strands of each of the two stem segments, and that this recognition cannot be solely dependent on a single nucleotide position in either of the respective strands. Recognition might depend on multiple 2′-hydroxyl groups in the essential RNA regions, but it is more

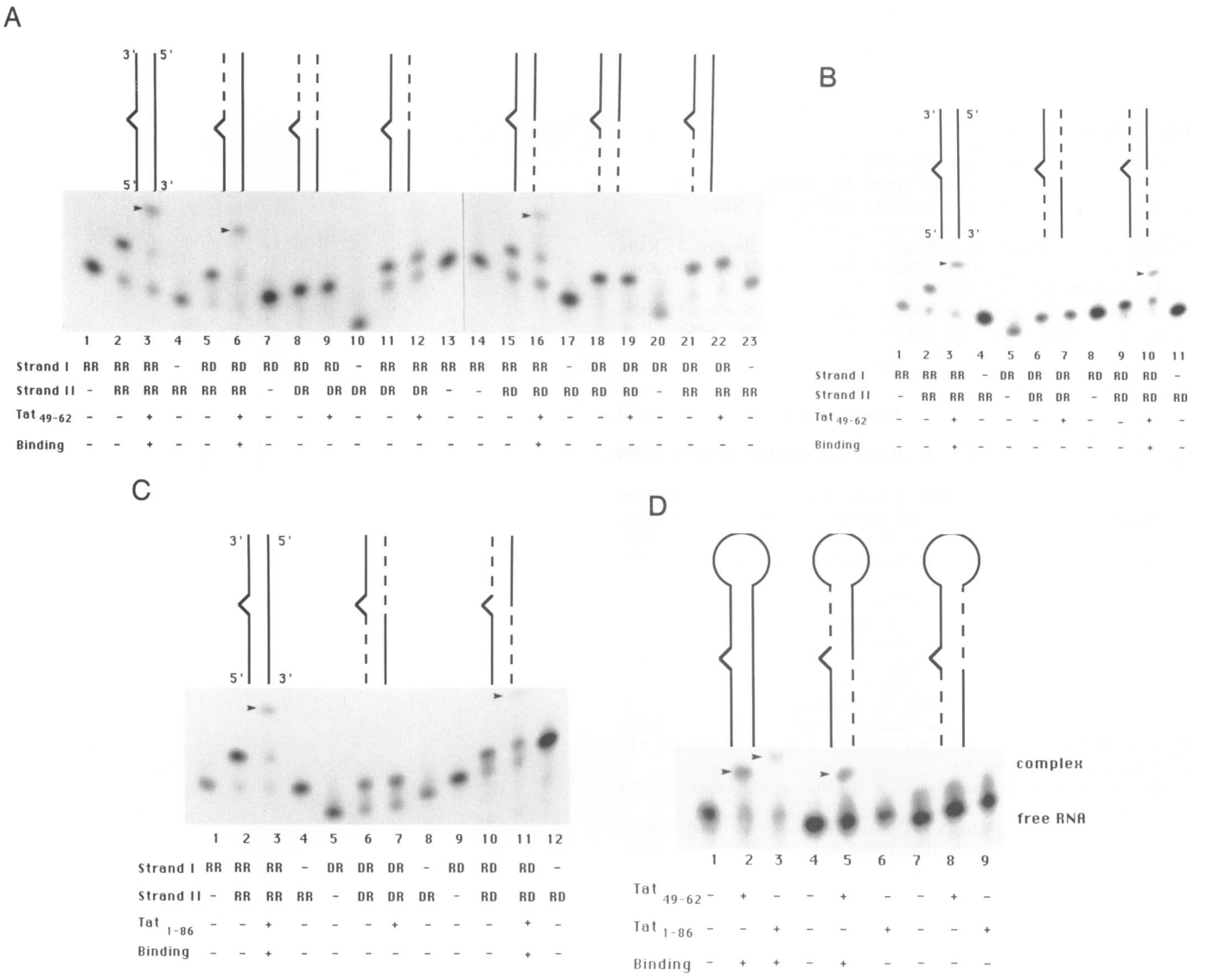

Figure 2. Binding of Tat 1–86 and 49–62 to TAR RNA-DNA hybrids. Experiments were performed with oligonucleotides composed entirely of RNA (RR), or of mixed RNA-DNA composition (RD or DR), where the order of RNA and DNA in each mixed oligonucleotide is shown in the 5′ to 3′ orientation. Oligonucleotides comprising strand 1 or strand 2 were combined as indicated (or left uncombined in controls) and hybridised as described in Methods. The expected duplex structures (A, B and C) are indicated schematically, and demonstrated to have formed by the presence of a band which migrates more slowly than either oligonucleotide alone. (**A,B**) Tat 49–62; or (**C**) full-length, recombinant Tat 1–86 were combined with the indicated duplexes. Binding of Tat to duplex results in the creation of a more slowly migrating complex (indicated by the arrowhead) and a reduction in the amount of unbound duplex. (**D**) Tat peptide (49–62 and 1–86) was assayed for binding to the indicated, labelled, single oligonucleotides prepared without pre-annealing.

likely that global, structural differences between RNA, DNA and RNA-DNA hybrids are responsible. RNA duplexes of mixed base sequence typically form A-form helices, whereas DNA duplexes usually form B-helices. NMR analysis of RNA-DNA duplexes by Chou *et al.* (19) showed that while the RNA strand adopts a classic A-helix conformation, the DNA strand is perturbed from the B-form. Our results imply that the conformation of one of the strands in a given stem must be in the A-form, and that neither B-form (as in full DNA duplexes) nor 'perturbed' forms (as in RNA-DNA hybrids) are acceptable. In contrast, the conformation of the complementary strand is not as crucial. These findings imply that Tat may interact with only one of the strands in a stem segment (i.e. that in the A-form), or at least that there is a different type of interaction between the two strands in a given stem region. Such differential strand recognition is apparent whatever mechanism is operating. However, there is an alternate explanation. Since we do not know the precise structure of the hybrid double helices, it is possible that in the case of each stem segment, only one of the two possible RNA-DNA hybrids may form an appropriate helical structure, and the other may not.

The rotational symmetry exhibited by the strand requirements implies that the bulge plays a critical role in reversing the primary strand with which Tat interacts. It will be interesting to determine if specific strand recognition of RNA duplexes is a general characteristic of the arginine-rich RNA-binding motif family.

ACKNOWLEDGEMENTS

The authors thank Craig Rosen for generously providing purified Tat protein, and Lorne Reid and Michael Ma for helpful suggestions. U.D. holds a Studentship from the Medical Research Council of Canada. This work was partly supported by grants

(to N.S.) from Health and Welfare Canada-National Health Research and Development Program and the Cancer Society (Montreal).

REFERENCES

1. Cullen, B.R. (1990) *Cell* **63**, 655–657.
2. Pavlakis, G.N. and Felber, B.K. (1990) *New Biologist* **2**, 20–31.
3. Rosen, C.A. and Pavlakis, G.N. (1990) *AIDS* **4**, 499–509.
4. Rosen, C.A. (1991) *Trends in Genetics* **7**: 9–14.
5. Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M. and Skinner, M.A. (1989) *Proc. Natl. Acad. Sci.* USA **86**, 6925–6929.
6. 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 (1990).
7. Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A. and Crothers, D.M. (1990) *Science* **249**, 1281–1285.
8. Roy, S., Parkin, N.T., Rosen, C., Itovitch, J. and Sonenberg, N. (1990) *J. Virol.* **64**, 1402–1406.
9. Roy, S., Delling, U., Chen, C.-H., Rosen, C.A. and Sonenberg, N. (1990) *Genes Dev.* **4**, 1365–1373.
10. Cordingly, 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.
11. Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D. and Frankel, A.D. (1991) *Science* **252**, 1167–1171.
12. Lazinski, D., Gradzielska, E. and Das, A.. (1989) *Cell* **59**, 207–218.
13. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L.S., Kuperman, R., Delling, U. and Sonenberg, N. (1991) *J. Virol.* **65**, 5196–5202.
14. 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.
15. Delling, U., Roy, S., Sumner-Smith, M., Barnett, R., Reid, L., Rosen, C.A. and Sonenberg, N. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6234–6238.
16. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992) *Science* **257**, 76–80.
17. Gentz, R., Chen, C.-H. and Rosen, C.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 821–824.
18. Harper, J.W. and Logsdon, N.J. (1991) *Biochemistry* **30**, 8060–8066.
19. Chou, S-H., Flynn, P., and Reid, B. (1989) *Biochemistry* **28**, 2435–2443.