al of charged pyruvate groups (accounting for about one-third of the total xanthan charge), exposing neutral hydroxyls, reduces the magnitude of the high-pressure, hydration force magnitude by about a factor of 3.

Biologically, the groups that modify sugars in long-chain polysaccharides include phosphates, sulfates, carboxylates, and amino groups. Each group has its own hydration and ion-binding characteristics that modulate interactions with other molecules. Through extensive measurements of many carbohydrates, specific contributions of different groups on hydration interactions can be compiled in much the same way as assigning charge to groups for electrostatic interaction. Cells use a wide range of carbohydrate structures to accomplish a wide range of functions. The only way to establish a structure-function relation is to understand first the structure-interaction relation.

REFERENCES AND NOTES

1. D. M. LeNeveu, R. P. Rand, V. A. Parsegian, *Nature* **259**, 601 (1976); ———, D. Gingell, *Biophys. J.* **18**, 209 (1977); R. P. Rand, *Annu. Rev. Biophys. Bioeng.* **10**, 277 (1981).
2. V. A. Parsegian, R. P. Rand, N. L. Fuller, D. C. Rau, *Methods Enzymol.* **127**, 400 (1986).
3. R. P. Rand and V. A. Parsegian, *Biochim. Biophys. Acta* **988**, 351 (1989).
4. D. C. Rau, B. K. Lee, V. A. Parsegian, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2621 (1984); V. A. Parsegian, R. P. Rand, D. C. Rau, in *Physics of Complex and Supermolecular Fluids*, S. A. Safran and N. A. Clark, Eds. (Wiley, New York, 1987), pp. 115–135.
5. R. Podgornik, D. C. Rau, V. A. Parsegian, *Macromolecules* **22**, 1780 (1989); R. Podgornik and V. A. Parsegian, *ibid.* **23**, 2265 (1990).
6. C. R. Safinya *et al.*, *Phys. Rev. Lett.* **57**, 2718 (1986).
7. An Enraf Nonius fixed-anode Diffractis 601 x-ray generator equipped with a high-power, fine-focus copper tube was used for the measurement of interaxial distances. The cameras were designed and built for our particular requirements [C. P. Mudd, H. Tipton, V. A. Parsegian, D. C. Rau, *Rev. Sci. Instrum.* **58**, 2110 (1987)]. The sample cells contain both the pellet and 100 μl of the bathing PEG (or PVP) and salt solution. X-ray reflections from the polysaccharide pellets are detected by film (DEF 5, Eastman-Kodak Co.). The strong interaxial reflections can be measured either by hand or by scanning densitometer (*5*). The 14.5 Å Bragg reflection from powdered *p*-bromobenzoic acid was used to calibrate the sample to film distance (about 15 cm). The intensities of the higher order reflections that determine the interhelical packing geometry were not strong enough to confirm unambiguously an expected hexagonal packing of rods for either xanthan or schizophyllan. We nonetheless assume that the packing is hexagonal and that the interhelical spacing, D_{int} is given by $D_{int} = 2D_{Br}/\sqrt{3}$, where D_{Br} is the Bragg spacing of the x-ray reflection.
8. Polyethylene glycol, average molecular weights 8,000 and 20,000, and polyvinylpyrrolidone, average molecular weight 40,000, were purchased from Sigma and used without further purification. A tabulation of osmotic pressures for PEG and PVP solutions is given in (*2*). We use the PVP osmotic pressures determined by us from secondary osmometry against PEG rather than the pressures given by K. A. Granath [*J. Colloid Sci.* **13**, 308 (1958)].
9. For a review of xanthan structure, properties and industrial uses, see D. A. Rees, E. R. Morris, D. Thom, J. K. Madden, in *The Polysaccharides*, G. O. Aspinall, Ed. (Academic, New York, 1982), vol. 1, pp. 195–290; P. A. Sandford and J. Baird, in *The Polysaccharides*, G. O. Aspinall, Ed. (Academic, New York, 1983), vol. 2, pp. 411–490.
10. G. Holzwarth and E. B. Prestridge, *Science* **197**, 757 (1977); B. T. Stokke, A. Elgsaeter, O. Smidsrod, *Int. J. Biol. Macromol.* **8**, 217 (1986).
11. R. Moorhouse, M. D. Walkinshaw, S. Arnott, *ACS (Am. Chem. Soc.) Symp. Ser.* **45**, 90 (1977); K. Okuyama *et al.*, *ibid.* **141**, 411 (1980).
12. T. Coviello, K. Kajiwara, W. Burchard, M. Dentini, *Macromolecules* **19**, 2826 (1986).
13. K. P. Shatwell, I. W. Sutherland, S. B. Ross-Murphy, *Int. J. Biol. Macromol.* **12**, 71 (1990).
14. T. Norisuye, T. Yanaki, H. Fugita, *J. Polym. Sci. Polym. Phys. Ed.* **18**, 547 (1980); T. Sato, T. Norisuye, H. Fujita, *Macromolecules* **16**, 185 (1983).
15. T. Itou and A. Teramoto, *Macromolecules* **17**, 1419 (1984); T. Odijk, *ibid.* **19**, 2313 (1986).
16. ———, T. Matsuo, H. Suga, *Carbohydr. Res.* **160**, 243 (1987).
17. T. Yanaki, T. Norisuye, H. Fujita, *Macromolecules* **13**, 1462 (1980); T. Kashiwagi, T. Norisuye, H. Fujita, *ibid.* **14**, 1220 (1981).
18. H. Enomoto, Y. Einaga, A. Teramoto, *ibid.* **18**, 2695 (1985).
19. For an overview, see "Proceedings on Hydration Forces and Molecular Aspects of Solvation," in *Chem. Scr.* **25**, 1–120 (1985).
20. R. M. Pashley, *Adv. Colloid Interface Sci.* **16**, 57 (1982).
21. J. Widom and R. L. Baldwin, *J. Mol. Biol.* **144**, 431 (1980).
22. D. C. Rau and V. A. Parsegian, in preparation.
23. Highly purified, high molecular weight xanthan was a gift from I. Heilweil.
24. Schizophyllan was a gift from A. Teramoto (Osaka University). The viscosity-determined average molecular weight of this fractionated sample is about 2×10^5 (A. Teramoto, personal communication).
25. The xanthan measurements were originally suggested to us by W. Schowalter, R. Prud'homme, and I. Heilweil. We thank I. Heilweil, G. Holzwarth, S. B. Ross-Murphy, and F. H. Kirkpatrick for several helpful textual suggestions.

19 December 1989; accepted 19 June 1990

Fragments of the HIV-1 Tat Protein Specifically Bind TAR RNA

KEVIN M. WEEKS, CHRISTOPHE AMPE, STEVE C. SCHULTZ, THOMAS A. STEITZ, DONALD M. CROTHERS

Proteolytically produced carboxyl-terminal fragments of the human immunodeficiency virus type–1 (HIV-1) Tat protein that include a conserved region rich in arginine and lysine bind specifically to transactivation response RNA sequences (TAR). A chemically synthesized 14-residue peptide spanning the basic subdomain also recognizes TAR, identifying this subdomain as central for RNA interaction. TAR RNA forms a stable hairpin that includes a six-residue loop, a trinucleotide pyrimidine bulge, and extensive duplex structure. Competition and interference experiments show that the Tat-derived fragments bind to double-stranded RNA and interact specifically at the pyrimidine bulge and adjacent duplex of TAR.

TAT IS A POTENT TRANSACTIVATOR OF HIV-1 long terminal repeat (LTR)–linked gene expression and is essential for viral replication (*1*). The transactivator is a small protein that includes a cysteine-rich region, which may be involved in zinc-mediated dimerization (*2*), a putative activation domain (*3, 4*), and a highly basic region (Fig. 1A). The precise mechanism of Tat-mediated transactivation remains unresolved (*5, 6*) but requires TAR, which is located at the 5′ end of the untranslated leader region of all viral messenger RNAs (mRNAs), the first 57 nucleotides of which form a stable stem-loop structure in vitro (*7*) (Fig. 1E). Nucleotides spanning positions +19 to +42 are sufficient for Tat response in vivo (*8, 9*). TAR must be located immediately downstream of the transcriptional start site; inversion of the sequence such that the complementary strand is transcribed eliminates transactivation (*7, 10*). Mutations that disrupt base pairing in duplex regions adjacent to the loop in TAR interfere with transactivation; compensating mutations that restore base pairing restore Tat responsiveness (*11, 12*). Point mutations in the loop eliminate Tat response (*11*), and Tat-mediated transactivation is also lost when sequences flanking the minimal TAR element are mutated such that competing secondary structures are more stable than the native TAR hairpin (*13*). Thus both RNA structure and primary sequence affect transactivation.

One hypothesis consistent with the accumulated data is that a direct interaction between Tat and TAR mediates transactivation. The formation of a specific complex between Tat purified from *Escherichia coli*

K. M. Weeks, Department of Chemistry, Yale University, New Haven, CT.
C. Ampe, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.
S. C. Schultz, Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University, New Haven, CT.
T. A. Steitz, Department of Molecular Biophysics and Biochemistry, Department of Chemistry, and Howard Hughes Medical Institute, Yale University, New Haven, CT.
D. M. Crothers, Department of Chemistry and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.

and TAR RNA has been reported, although a large excess of protein (10^4) was required to observe partial complexation of the RNA in native gels (*14*). We now show that the basic subdomain of Tat is central for recognition of TAR RNA and that peptides containing this domain bind specifically at the pyrimidine bulge and adjacent duplex in TAR.

When a synthetic *Tat* gene is expressed in *E. coli*, the protein is only sparingly soluble, and denaturing agents are required during purification (*2, 6, 14*). Tat isolated in this way remains unfolded, readily oxidizes to yield cross-linked protein multimers, and aggregates in RNA binding assays (*2, 14*, and this report). We expressed a fusion protein containing the cyclic adenosine monophosphate (cAMP) binding domain from the *E. coli* catabolite gene activator protein (CAP) and either full-length Tat (86 residues) or a truncated form (corresponding to the first exon, 72 residues), linked by a recognition site for either the protease factor Xa or endoproteinase Glu-C (V8) (*15*). The fusion protein was purified by affinity chromatography (*16*) and has an apparent molecular mass of 26 kD. Digestion of the fusion protein with the appropriate protease gave rise to a peptide that formed a low-mobility complex in gel retardation assays with TAR RNA.

Of the three major proteolytic fragments resolved by high-performance liquid chromatography (HPLC) (Fig. 2A), one (peak 3) represents the cAMP binding domain of CAP. Material eluting in the first peak formed discrete complexes with TAR in mobility shift assays (Fig. 1, C and D); the composition and NH_2-terminal sequence of the peptide were determined. This fragment is not full-length Tat but arises from an unexpected cleavage between Gly^{48} and Arg^{49} and extends to the COOH-terminus of the fusion protein, yielding a Tat fragment of 38 residues (Tfr38, Fig. 1A). The material eluting in peak 2 was marginally soluble, readily oxidized, and probably corresponds to the cysteine-rich NH_2-terminus of Tat (Fig. 2).

Purification of a fusion protein containing the truncated form of Tat followed by proteolytic digestion produced a smaller Tat fragment (Tfr24, Fig. 1A) extending from Arg^{49} to residue 72. Fusion proteins with either Xa or V8 linkers reproducibly yielded identical proteolytic fragments corresponding to the COOH-terminus of the respective protein; cleavage is probably catalyzed by a copurified bacterial protease. Pabo and coworkers (*2*), using a variety of proteases, observed a similar proteolytic susceptibility between residues 38 and 48 of Tat. Intra-Tat cleavage is thus highly reproducible, indicative of structural flexibility or accessibility of the protease cleavage site, and is consistent with a Tat structure comprised of two domains.

Both Tfr24 and Tfr38 peptides used in this study were generated as described above and then purified to homogeneity by reversed-phase HPLC. Since both fragments specifically recognize TAR RNA and include the basic region in Tat, we chemically synthesized a 14-residue peptide (Tfr14) that spans the basic region (Fig. 1A) (*17*).

When either the proteolytic Tat fragment

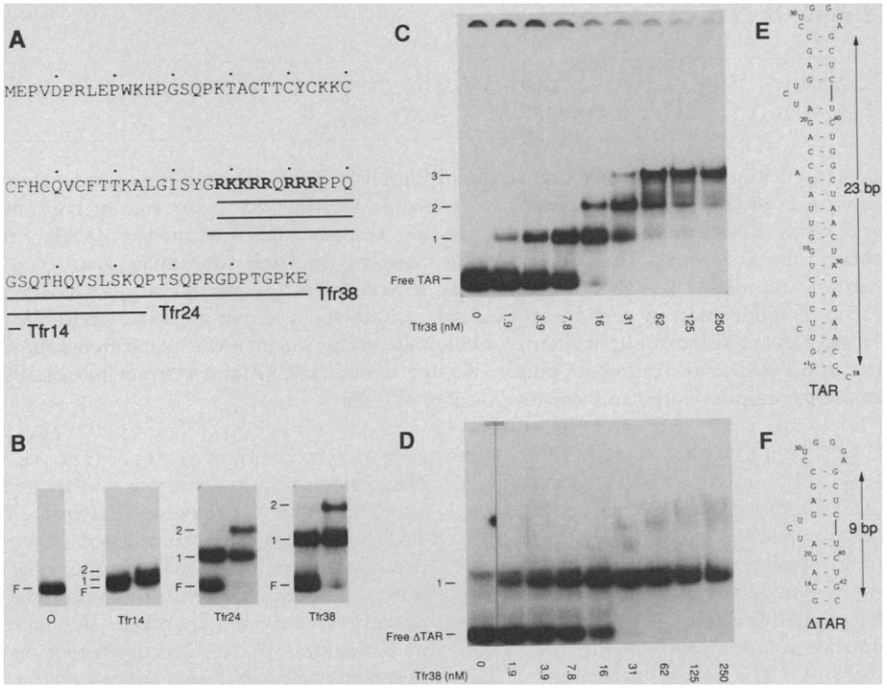

Fig. 1. Mobility shift assays of Tfr14, Tfr24, and Tfr38 with TAR. (**A**) Sequence of HIV-1 Tat protein (BRU isolate) with single-letter amino acid abbreviations. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. Peptides used in this study are indicated by dark lines. Arginine and lysine residues in the basic domain are in bold. (**B**) Relative mobilities of Tfr14, Tfr24, and Tfr38 complexed with TAR. Markers F, 1, and 2 identify free TAR and complexes containing one bound peptide and two bound peptides per RNA molecule, respectively. Radiolabeled TAR (*31*) was incubated with peptide in 10-μl reactions (10 mM tris, 70 mM NaCl, 0.2 mM EDTA, and 5% glycerol, pH 7.5) and incubated for 10 min at 22°C. Samples were loaded on a running native 10% (w/v) polyacrylamide gel (75:1 w/w acrylamide:bis) and separated by electrophoresis for 90 min in 45 mM tris-borate, 1 mM EDTA, pH 8.3 at 25 V/cm in a thermostated gel box (Hoefer) at 20°C. (**C**) Titration of Tfr38 with full-length TAR. TAR (5 nM) was incubated with the indicated concentration of peptide and assayed as in (B). Markers F, 1, 2, and 3 indicate free RNA or complexes of the RNA with the specified number of peptide molecules, respectively. (**D**) Titration of Tfr38 with ΔTAR. Reactions were separated by electrophoresis for 45 min; all other conditions were as in (B). The lower mobility band seen in the lane without peptide corresponds to a ΔTAR dimer. (**E**) Secondary structure of TAR RNA (TAR) used in this study. (**F**) Secondary structure of the minimal TAR RNA binding domain (ΔTAR). ΔTAR includes one non–wild-type base pair to facilitate transcription by T7 RNA polymerase.

Table 1. The relative binding constant K_{rel} is defined as $K_d^{TAR}/K_d^{competitor}$. (**A**) Relative affinities of Tfr38 and Tfr14 for homopolymer competitors were determined from the ratios of unlabeled competitor concentrations (in nucleotides) required to dissociate half of a radiolabeled TAR-peptide complex; calculated values for K_{rel} are reproducible to ~50%. This calculation significantly underestimates K_{rel}, because it does not take into account site degeneracy for the homopolymer competitors. (**B**) Relative dissociation constants for the TAR mutants (M28, M32, and B0) were determined from the ratios of the free and complexed wild-type and mutant RNAs (*30*); relative binding constants were quantified at ten points at Tfr38 concentrations ranging from 1 to 40 nM; 1 SD is given in parenthesis. Binding to B0 is not detectable by this assay.

Peptide	Competitor	K_{rel}
A		
Tfr38	TAR	1.0
	Poly(A) · poly(U)	0.05
	Poly(dA) · poly(dT)	0.008
	Poly(A)	0.002
	Poly(dA)	0.003
Tfr14	TAR	1.0
	Poly(A) · poly(U)	0.03
	Poly(dA) · poly(dT)	0.02
	Poly(dA)	0.003
B		
Tfr38	M32 (G → U)	1.0 (0.1)
	M28 (C → G)	0.24 (0.03)
	B0 (Δ22–24)	≤0.025

Tfr24 or Tfr38, or the chemically synthesized peptide Tfr14 was incubated with TAR RNA (Fig. 1E), electrophoresis resolved discrete complexes that have mobilities in native gels that are a function of peptide length (Fig. 1B). Titration of Tfr38 with full-sized TAR yields three discrete complexes in native gels at saturating concentrations of peptide (Fig. 1C), indicating that up to three peptide monomers, or possibly dimers, can bind the RNA. Incubation of equimolar quantities of Tfr24 and Tfr38 with TAR revealed only a single new complex with a mobility intermediate between doubly shifted Tfr24-TAR and Tfr38-TAR complexes, corresponding to a Tfr24-Tfr38-TAR complex, and confirming that each peptide binds as a monomer. Incubation of a truncated RNA molecule (ΔTAR, Fig. 1F) spanning the minimal RNA binding site (*8, 9*) with Tfr38 yields a single complex even at very high peptide concentrations, indicating that only one peptide binds this RNA (Fig. 2D). Both Tfr24 and Tfr14 also form three complexes with TAR, but only one complex with ΔTAR at saturating concentrations of peptide. We interpret these data and other results described below as an indication for both specific and weaker nonspecific components of RNA interaction by the RNA binding domain of Tat. The dissociation constant for the specific complex between one Tfr38 molecule and either TAR or ΔTAR RNA is ~5 nM. Identical experiments with Tfr24 yield a similar dissociation constant for this peptide-RNA interaction. The second binding step has an apparent dissociation constant of ~20 nM.

Preformed specific Tfr24-TAR and Tfr38-TAR complexes are stable to competition by an excess (100-fold) of unlabeled competitor [half-life ($t_{1/2}$) ≥ 2 min in 70 mM NaCl]. When a Tfr14-TAR complex is treated with an excess of unlabeled TAR and the reaction is immediately loaded (within 15 s) on a running native gel, no complex with labeled TAR is observed, suggesting that the dissociation rate for this complex is very fast. Thus, although residues COOH-terminal to the basic region are not necessary for binding to TAR, their presence contributes to the kinetic stability of the complex. These data are consistent with studies showing that although residues 5 to 58 (*18*) or residues 37 to 62 (*4*) are sufficient for transactivation, addition of residues through 72 greatly enhances transactivation in vivo (*4, 18*). We infer that extending the peptide to residue 72 may increase the kinetic stability of the RNA-protein complex leading to enhanced transactivation.

The relative migration of electrophoretic bands reported by Dingwall *et al.* (*14*) is similar to that produced by Tfr38 (Fig. 1C). Given the large excess of protein required for complex formation and the presence of non–full-length material in their protein preparations, it is possible that a degradation product of Tat rather than intact protein was actually responsible for the observed specific complexes with TAR.

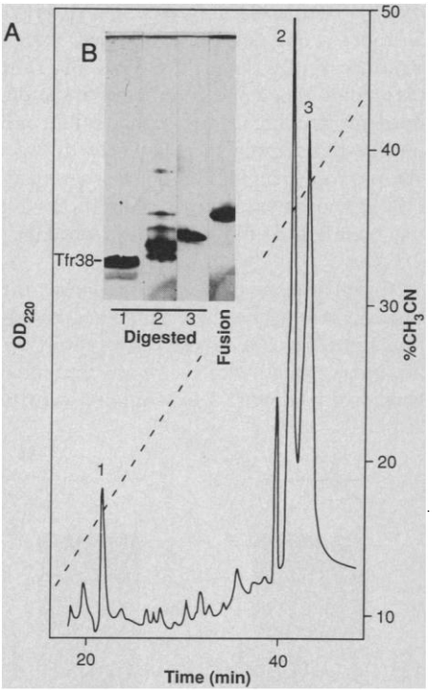

Fig. 2. Analysis of fusion protein cleavage products and purification of Tfr38. (**A**) Analytical separation of a proteolytic digest of fusion protein [full-length Tat, V8 linker (*15, 16*)]. Digestion products were resolved by C4 reversed-phase HPLC [samples were made in 0.1% trifluoroacetic acid (TFA) and eluted with a linear gradient of 0.1% TFA in acetonitrile; flow rate was 1 ml/min]. The unnumbered peak represents a minor low molecular-weight digestion product and was not further characterized. (**B**) SDS–polyacrylamide gel electrophoresis (PAGE) analysis of HPLC fractions. Lanes 1, 2, and 3 correspond to the indicated HPLC peaks (A) and represent residues 49 to 86 (Tfr38) of Tat, the NH_2-terminal domain of Tat, and the cAMP binding domain of CAP, respectively. The lane marked fusion is undigested fusion protein purified by HPLC. Samples were dissolved in 1% SDS–50 mM DTT, separated by electrophoresis (Phast-Gel System, Pharmacia) and stained with Coomassie blue. Higher molecular weight bands in lane 2 correspond to oxidized peptide multimers and co-purified undigested fusion protein. Sequence of Tfr38 (and Tfr24) was confirmed by both NH_2-terminal sequencing and compositional analysis.

To determine the specificity of Tfr38 and Tfr14 for global nucleic acid conformation, we performed competition experiments with single- and double-stranded competitors (Table 1). The binding affinity of both Tfr38 and Tfr14 is 20 to 30 times lower to double-stranded RNA [poly(A) · poly(U)] and 300 to 600 times lower to single-stranded RNA [poly(A)]. Both peptides have low affinity for single- and double-stranded DNA, including TAR DNA.

To further characterize the specificity of Tfr38 for TAR structure, we incorporated two mutations shown to eliminate transactivation in vivo (*11*) into TAR RNA transcripts extended by six 3′ nucleotides to facilitate resolution from wild-type RNA in native gels. M28 (C→G) disrupts the double-stranded stem next to the loop, and M32 (G→U) alters a loop residue (Fig. 1E). The latter mutation did not affect binding by Tfr38 in competition experiments, whereas destabilization of the double-stranded stem structure decreased binding by a factor of four (Table 1). Thus TAR sequences in which the duplex structure is interrupted may not be Tat responsive because the Tat-TAR complex is destabilized. We also constructed a mutant of ΔTAR, B0 (Δ22–24), in which the pyrimidine bulge was removed, forming a perfect 9-bp duplex. The complex between the bulgeless mutant and Tfr38 is not stable to electrophoresis; we estimate a binding constant at least 40 times lower than wild type for this mutant (Table 1). Thus the pyrimidine bulge is a significant determinant for recognition by Tfr38.

To define the RNA binding site for Tfr38 to nucleotide resolution, we used a rapid partition method (*19*) (Fig. 3), incorporating base-specific RNA chemical modification protocols (*20*). Contact sites were determined by comparing band intensities for cleavage at a given residue from RNA isolated as free probe or complexed with peptide. Interference was observed at each of ten nucleotides comprising five base pairs surrounding the trinucleotide pyrimidine bulge in TAR (Fig. 4). In the bulge, removal of the C^{23} by hydrazine treatment did not affect binding, modification of U^{22} interfered with binding, and elimination of U^{24} enhanced binding. Modification of residues in the hexanucleotide loop did not affect binding. Nucleotides U^5 to U^9 and A^{50} to A^{54} are identical to the 5 bp whose modification interferes with binding to Tfr38 (Fig. 1E). We observed no interference in this homologous region that did not contain a bulge, consistent with the B0 (Δ22–24) competition experiment.

Interference analysis thus identifies a limited region in TAR as the primary interaction site for Tfr38 in vitro (Fig. 4A). This binding site corresponds closely with a subset of the minimal TAR sequences that are sufficient and necessary for Tat response in vivo (Fig. 4B), as defined by mutational analyses (*8, 9, 11, 12*) and consistent with phylogenetic comparison (*21*), providing strong evidence that transactivation in vivo

requires a physical interaction between the COOH-terminus of Tat and a discrete region in TAR. The single hypervariable nucleotide (C^{23}) within the core binding site is the only one in which a chemical modification does not affect interaction with Tfr38. However, in addition to the boxed sequences (Fig. 4) required for interaction with Tfr38, transactivation in vivo is mediated by additional interactions at the hairpin loop in TAR, possibly with cellular factors (*22*).

Tat from HIV-1 can transactivate genes downstream of the HIV-2 LTR (*23*), suggesting that HIV-1 Tat recognizes a similar motif in both leader mRNAs. In fact, HIV-2 leader sequences can be folded into a structure similar to HIV-1 TAR (*8*). The pyrimidine bulge and 5-bp contact site defined by interference are highly conserved; outside this region, the sequences diverge. We propose that HIV-1 Tat transactivates HIV-2 gene expression via interaction with this homologous bulged duplex from HIV-2 TAR.

Several groups have suggested that the basic domain may encode a nuclear localization signal (*6, 24*). These observations are not incompatible with the results presented here, which identify RNA binding activity with this region: the basic domain may assume multiple roles in vivo. Similar highly basic subdomains have been identified in a variety of RNA binding proteins including bacterial N proteins, ribosomal proteins, RNA virus capsids, and HIV-1 Rev (*25*). We show that a short basic peptide can recognize a small bulged helix and present physical evidence for specific RNA binding by these basic arrays.

Significant precedent exists for positioning of basic peptides (*26*) and sequence-specific recognition (*27*) by proteins in the shallow and accessible minor groove of duplex RNA. In the minor groove, only the exocyclic two-amino group of guanine distinguishes A-U from G-C base pairs and possibly G-C from C-G (*28*). Inversion of two G-C base pairs (to C-G) within the minimal binding site for Tfr38 abolishes transactivation (*12*) (Fig. 4), suggesting an important role for one or both of these base pairs for specific recognition by the peptide in the minor groove. The basic region in Tat is absolutely conserved among known HIV-1 isolates (*21*); potential sites for base pair–specific interaction with duplex RNA in the minor groove involve H-bond donors, requiring the protein to contribute an H-bond

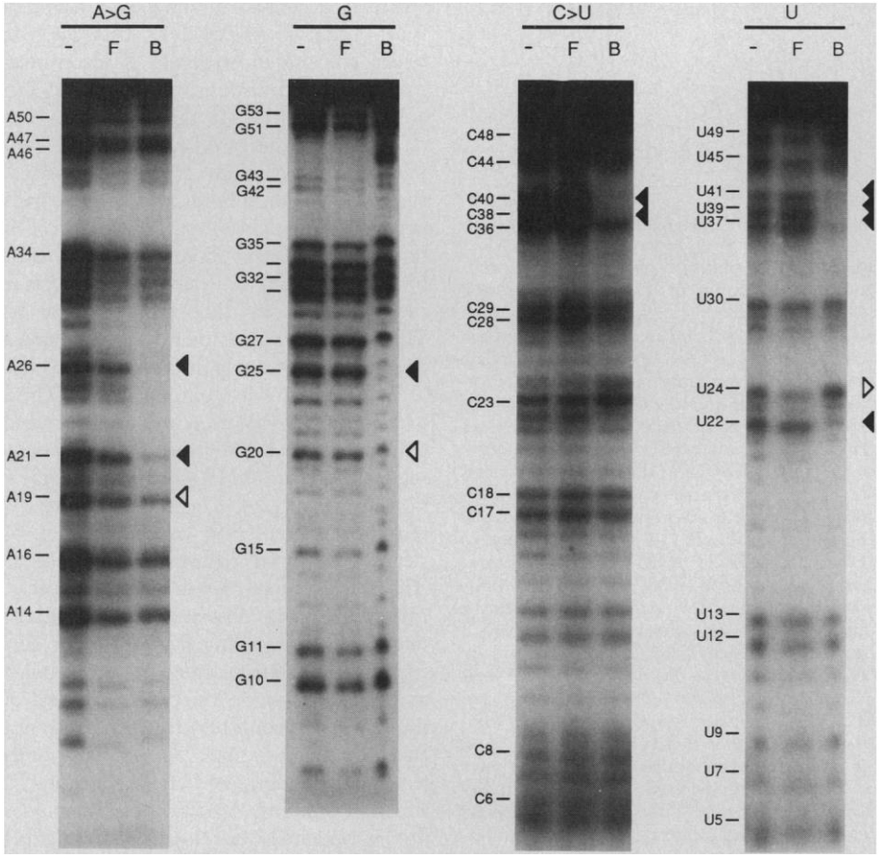

Fig. 3. Analysis of Tfr38-TAR contacts by chemical modification–based interference. The 5′ ^{32}P end-labeled TAR RNA (Fig. 1E) was modified with base-specific reagents and precipitated it once with ethanol as described (*20*). Reactions were resuspended in 7 M urea and purified by electrophoresis on denaturing gels. Full-length RNAs were eluted overnight into 600 mM sodium acetate, 1 mM EDTA, and 0.1% SDS, concentrated by ethanol precipitation, and incubated with Tfr38 sufficient for complexation of 40 to 50% of the labeled material when resolved by electrophoresis in a native gel (Fig. 1B). TAR RNA from bands corresponding to free RNA and protein-RNA complex was individually isolated from the native gel as described above, except that the elution solution contained 20 μg of proteinase K and 50 μg of transfer RNA. The RNAs were cleaved at modified nucleotides by treatment with aniline (*20*) and resolved on denaturing polyacrylamide gels. Dimethyl sulfate (G reaction) methylates guanine at N^7, forming a tertiary amine and thus introducing both a steric and electrostatic perturbation at the base pair step. Diethyl pyrocarbonate (A > G reaction) preferentially carboxyethylates N^7 and N^6 of adenine over N^7 of guanine, thereby opening the imidizole ring and introducing a significant steric perturbation in the narrow major groove of RNA. The C > U and U reactions use hydrazine attack at the pyrimidine base, leaving an abasic site (*20*). Chemical modification thus significantly perturbs local base-pair structure, precluding an interpretation of interference data in terms of major or minor groove peptide contact. RNA eluted from a control lane in which no protein was added (−) and bands corresponding to free (F) and bound (B) RNA are shown. Sites of strong and weak interference are indicated by solid and open triangles, respectively. The single site (U^{24}) at which elimination of the base enhances binding is indicated by an open triangle pointing away from the gel.

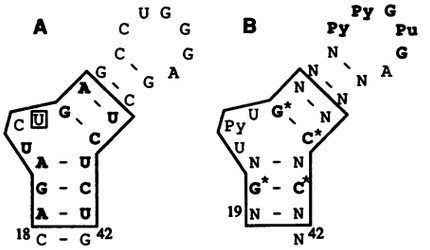

Fig. 4. Comparison of consensus HIV-1 TAR sequences involved in Tat mediated activity defined in vitro and in vivo. The binding site for Tfr38 [defined by interference (Fig. 3)] is enclosed by a heavy line. (**A**) A summary of interference analysis. Sites where residue modification interferes with Tfr38 binding are in bold type; the base U^{24} that enhances binding when it is eliminated is boxed. Modification of C^{23} does not interfere with peptide binding. (**B**) Summary of TAR elements required for transactivation of reporter genes in vivo. Positions where base pairing, but not sequence, are conserved are indicated by N (*11, 12*). Minimal binding domain defined by Jakobovits *et al.* (*8*). Two G-C base pairs where simultaneous inversion (to C-G) interferes with Tat responsiveness are indicated with asterisks (*12*). Phylogenetic consensus (*21*) suggests that mutations in double-stranded regions are conservative in the sense that they retain base pairing; at several non–base-paired positions, residue type but not identity is conserved: C-U and A-G transitions but not purine-pyrimidine transversions are tolerated, most notably at position 23, where cytosine and uridine appear nearly equally often. Where such generalization is supported by the sequence of an HIV-1 isolate, it is indicated by N, Pu, or Py, for conserved base pairing, purine, or pyrimidine, respectively.

acceptor for specific interaction. Possible acceptors within the basic domain include the side chain of glutamine and the peptide backbone (*27*), either of which may form H-bonds with one or both of the invariant G-C base pairs in the minimal Tfr38 binding domain. Another essential feature is likely to be recognition of a kinked structure (*29*) introduced by the pyrimidine bulge. Thus both sequence and structure of a limited TAR region contribute to specific recognition by the COOH-terminal domain of Tat.

The data presented here are consistent with and offer a basis for physical interpretation of a subset of the genetic data identifying elements of TAR or Tat that are required for transactivation in vivo. Our experiments delineate minimal protein and RNA elements that have pivotal roles in an interaction vital to the HIV-1 life cycle.

REFERENCES AND NOTES

1. H. Varmus, *Genes Dev.* **2**, 1055 (1988); B. R. Cullen and W. C. Greene, *Cell* **58**, 423 (1989).
2. A. D. Frankel *et al.*, *Science* **240**, 70 (1988).
3. M. Green and P. M. Loewenstein, *Cell* **55**, 1179 (1988).
4. M. Green, M. Ishino, P. M. Loewenstein, *ibid.* **58**, 215 (1989).
5. M. F. Laspia, A. P. Rice, M. B. Mathews, *ibid.* **59**, 283 (1989).
6. M. Braddock *et al.*, *ibid.* **58**, 269 (1989).
7. M. A. Muesing, D. H. Smith, D. J. Capon, *ibid.* **48**, 691 (1987).
8. A. Jakobovits *et al.*, *Mol. Cell. Biol.* **8**, 2555 (1988).
9. M. J. Selby, E. S. Bain, P. A. Luciw, B. M. Peterlin, *Genes Dev.* **3**, 3547 (1988); J. Hauber and B. R. Cullen, *J. Virol.* **62**, 673 (1988).
10. B. M. Peterlin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9734 (1986).
11. S. Feng and E. C. Holland, *Nature* **334**, 165 (1988).
12. J. A. Garcia *et al.*, *EMBO J.* **8**, 765 (1989).
13. B. Berkhout, R. H. Silverman, K-T. Jeang, *Cell* **59**, 273 (1989).
14. C. Dingwall *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6925 (1989).
15. We constructed a gene that encodes a fusion protein containing the following sequences (NH_2- to COOH-terminal): (i) the cAMP binding domain (residues 1–134) from the *E. coli* CAP protein [H. Aiba, S. Fujimoto, N. Ozaki, *Nucleic Acids Res.* **10**, 1345 (1982); D. B. McKay, I. T. Weber, T. A. Steitz, *J. Biol. Chem.* **257**, 9518 (1982)]; (ii) a six-residue linker that can be cleaved by Xa or V8; and (iii) either residues 1–86 (full length) or 1–72 (first exon) of the HIV-1 Tat protein, BRU isolate [S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alison, *Cell* **40**, 9 (1985)]. The fusion proteins were expressed in *E. coli* strain BL21 (DE3) [F. W. Studier and B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (1986)] with a T7 RNA polymerase-directed expression system analogous to that of A. H. Rosenberg *et al.* [*Gene* **56**, 125 (1987)].
16. Cells were lysed in 10 mM tris, 700 mM NaCl, 2 mM EDTA, 10 mM dithiothreitol (DTT), and pH 8. Cleared lysates were treated with 0.5% polymin P. The supernatant was loaded on a cAMP-agarose column (Sigma), washed with the lysis buffer and 500 mM K_2HPO_4, 1 mM EDTA, 5 mM DTT, pH 7.5, and 10 mM tris, 50 mM NaCl, 0.5 mM EDTA, 2 mM DTT, and pH 7.5. The fusion protein was eluted in the latter buffer containing 2 mM cAMP. Fractions containing the fusion protein were pooled and dialyzed against 10 mM tris, 100 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 0.5 mM cAMP, pH 8 (Xa constructs), or pH 7.5 (V8 constructs). Fusion proteins were digested with the appropriate protease at 37°C for 2.5 hours (enzyme:substrate ratio 1:50 for Xa, 1:100 for V8); Xa digests require 2 mM $CaCl_2$.
17. Tfr14 was synthesized by the Yale University Medical School Protein and Nucleic Acid Chemistry Facility, purified by reversed-phase HPLC, and characterized by amino acid analysis and mass spectroscopy.
18. A. D. Frankel, S. Biancolana, D. Hudson, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7397 (1989).
19. U. Siebenlist and W. Gilbert, *ibid.* **77**, 122 (1980); D. A. Peattie and W. Herr, *ibid.* **78**, 2273 (1981); D. L. Thurlow, C. Ehresmann, B. Ehresmann, *Nucleic Acids Res.* **11**, 6787 (1983).
20. D. Peattie, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1760 (1979).
21. G. Myers, A. B. Rabson, S. F. Josephs, T. F. Smith, Eds., *Human Retroviruses and AIDS* (Los Alamos National Laboratory, Los Alamos, NM, 1989).
22. R. Gaynor, E. Soultanakis, M. Kuwabara, J. Garcia, D. S. Sigman, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4858 (1988); R. A. Marciniak, M. A. Garcia-Blanco, P. A. Sharp, *ibid.* **87**, 3624 (1990).
23. M. Emerman, M. Guyader, L. Montagnier, D. Baltimore, M. A. Muesing, *EMBO J.* **6**, 3755 (1987); S. K. Arya and R. L. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9753 (1988).
24. J. Hauber, M. H. Malim, B. R. Cullen, *J. Virol.* **63**, 1181 (1989); S. Ruben *et al.*, *ibid.*, p. 1.
25. D. Lazinski *et al.*, *Cell* **59**, 207 (1989).
26. R. W. Warrant and S.-H. Kim, *Nature* **271**, 130 (1978).
27. M. A. Rould *et al.*, *Science* **246**, 1135 (1989).
28. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
29. A. Bhattacharyya, A. I. H. Murchie, D. M. J. Lilley, *Nature* **343**, 484 (1990).
30. H.-N. Liu-Johnson, M. R. Gartenberg, D. M. Crothers, *Cell* **47**, 995 (1986).
31. RNAs were synthesized enzymatically with T7 polymerase [J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **15**, 8783 (1987)]. RNA sequences are shown in Fig. 2, E and F.

32. We are indebted to M. R. Gartenberg for thoughtful and enthusiastic discussion. J. Qin and G. Sun provided technical assistance; T7 RNA polymerase was the generous gift of S. A. White. C.A. is supported by an Anna Fuller Fund Fellowship. This research was supported in part by NIH grants to T.A.S. (GM 39546) and D.M.C. (GM 21966).

2 April 1990; accepted 25 June 1990

Site-Specific Gene Expression in Vivo by Direct Gene Transfer into the Arterial Wall

ELIZABETH G. NABEL,* GREGORY PLAUTZ,* GARY J. NABEL

A recombinant β-galactosidase gene has been expressed in a specific arterial segment in vivo by direct infection with a murine amphotropic retroviral vector or by DNA transfection with the use of liposomes. Several cell types in the vessel wall were transduced, including endothelial and vascular smooth muscle cells. After retroviral infection, a recombinant reporter gene was expressed for at least 5 months, and no helper virus was detected. Recombinant gene expression achieved by direct retroviral infection or liposome-mediated DNA transfection was limited to the site of infection and was absent from liver, lung, kidney, and spleen. These results demonstrate that site-specific gene expression can be achieved by direct gene transfer in vivo and could be applied to the treatment of such human diseases as atherosclerosis or cancer.

DESPITE RECENT ADVANCES IN THE understanding of eukaryotic gene structure and regulation, a major obstacle to the therapeutic management of human disease remains the site-specific expression of genes in vivo. Although tissue-specific gene expression can be determined by enhancer or other cis-acting regulatory elements, this expression might also be achieved through the delivery of vectors to specific anatomic sites in vivo. We have previously demonstrated that a recombinant gene can be expressed in the vasculature by means of genetically modified endothelial cells implanted at specific sites on the arterial wall (*1*). Because these studies required that syngeneic cell lines be established before genetic modification, adaptation to the treatment of human disease remained cumbersome. We now report that a recombinant gene can be efficiently expressed at a specific site in vivo by direct introduction of genetic material at the time of catheterization.

The retroviral vector (*2*) was derived from the Moloney murine leukemia virus and utilized the promoter from the chicken β-actin gene to express β-galactosidase mRNA. Viral particles from the supernatant of transfected ΨCRIP retrovirus packaging cells (*3*) were concentrated by centrifugation (10^4 to 10^6 particles per milliliter) and instilled into Yucatan or outbred pig iliofemoral arteries. These animals provide an experimental model for atherosclerosis when fed a high-fat diet (*4*). After anesthesia and surgical exposure, a catheter was inserted as

Departments of Internal Medicine and Biological Chemistry, Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, MI 48109–0650.

*This work resulted from an equal contribution by the first two authors.