effect. The experiment was repeated twice.
8. R. Horuk, *Immunol. Today* **15**, 169 (1994).
9. D. S. Dimitrov, *Nature Med.* **2**, 640 (1996).
10. 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.
11. C. K. Lapham *et al.*, data not shown.
12. Human A2.01.CD4.401 and mouse 3T3.CD4.401 cell lines expressing tailless CD4 molecules were produced in the laboratory of D. Littman. We thank C. C. Broder for helpful discussions, C. C. Broder and E. A. Berger for providing us with vCBFY1, R. Blackburn for the generation of rabbit immune antisera to fusin and purified IgG, J. Manischewitz for cell line propagation and vaccinia virus infections, and B. Golding and K. Peden for critical review of the manuscript.

5 August 1996; accepted 9 September 1996

Tat-SF1: Cofactor for Stimulation of Transcriptional Elongation by HIV-1 Tat

Qiang Zhou and Phillip A. Sharp*

Tat may stimulate transcriptional elongation by recruitment of a complex containing Tat-SF1 and a kinase to the human immunodeficiency virus–type 1 (HIV-1) promoter through a Tat-TAR interaction. A complementary DNA for the cellular activity, Tat-SF1, has been isolated. This factor is required for Tat trans-activation and is a substrate of an associated cellular kinase. Cotransfection with the complementary DNA for Tat-SF1 specifically modulates Tat activation. Tat-SF1 contains two RNA recognition motifs and a highly acidic carboxyl-terminal half. It is distantly related to EWS and FUS/TLS, members of a family of putative transcription factors with RNA recognition motifs that are associated with sarcomas.

Tat activation of HIV-1 transcription is mechanistically different from conventional DNA sequence-specific transcription factors. Most activators affect transcription by increasing the rate of initiation, although some DNA sequence-specific transcription factors such as GAL4-VP16 stimulate both initiation and elongation (*1*). In contrast, Tat predominantly stimulates elongation (*2*). Whereas most activators interact with promoter or enhancer DNA, Tat interacts with the trans-acting responsive (TAR) RNA element (*2*). Located at the 5′ end of the nascent viral transcript, TAR forms a stem-loop structure. The specific binding of Tat to TAR is dependent on the bulge loop and immediately flanking sequences in the double-stranded RNA. Sequences in the apical loop of TAR are also important for Tat activation of transcription in vivo (*3*).

Mechanisms regulating the efficiency of elongation by RNA polymerase II have not been extensively studied. The necessity for control of elongation is highlighted by the finding that an elongation factor, Elongin, is probably the functional target of the von Hippel–Lindau tumor suppressor protein (*4*–*6*). Furthermore, regulation of elongation by Tat is essential for HIV replication. We have used the Tat trans-activation system to characterize cellular cofactors critical for Tat activation of elongation.

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed.

We have developed a partially reconstituted transcription reaction that supports a Tat-specific and TAR-dependent activation of HIV transcription (*7*) (Fig. 1A). This reaction requires a Tat-SF (Tat stimulatory factor) activity that is specific for Tat stimulation of elongation, a phosphocellulose 0.5 to 1.0 M KOAc fraction of HeLa nuclear extract (the pc-D fraction), and the purified basal factors TFIID, TFIIA, and transcription factor Sp1. Reactions with these components, but without Tat-SF activity, support activation by Sp1 and GAL4-VP16 (*7*) but not by Tat (Fig. 1A). With the inclusion of a partially purified Tat-SF fraction, Tat increased the number of transcripts elongating beyond 1000 nucleotides from an HIV-1 promoter containing the wild-type TAR element (pHIV+TAR-G400) (*7*), but not from an internal control promoter with a mutant TAR (pHIVΔTAR-G100) (Fig. 1A). The pc-D fraction contains the basal transcription factors TFIIB, TFIIE, TFIIF, TFIIH, and RNA polymerase II (*7*). Because pc-D cannot be substituted for by highly purified basal transcription factors (*8*), it probably contains other activities necessary for Tat function. With the use of this reconstituted reaction, Tat-SF was further purified (*9*).

Phosphorylation of RNA polymerase II has been implicated in regulation of the processivity of elongation (*10*). To investigate whether protein phosphorylation might be associated with Tat-SF, we examined proteins absorbed on immobilized HIV TAR RNA from a reconstituted transcription reaction in the presence of [γ-^{32}P]ATP (adenosine triphosphate) (Fig. 1B). In reactions with either the pc-D fraction or the Tat-SF fraction alone, addition of Tat did not consistently affect the phosphorylation of proteins on immobilized TAR. When both fractions were incubated together in the presence of Tat, phosphorylation of a protein of ~140 kD, termed pp140, was observed (Fig. 1B). In the absence of Tat,

Table 1. The effect of Tat-SF1 overexpression on Tat and VP16 trans-activation. Tat-SF1 gene was cloned into the mammalian expression vector pSV7d (*28*) to create pSV-Tat-SF1. pSV-Tat-SF1 or pSV7d and a reporter construct pBennCAT (*29*) containing HIV-1 LTR linked to the bacterial CAT gene (1 μg each) and an internal control plasmid pCMVβ-Gal were cotransfected into HeLa cells, either in the presence or absence of a Tat-expressing plasmid pcTat (0.3 μg) (*30*). CAT activity was measured 48 hours later as described (*31*). In control experiments, pSV-Tat-SF1 or pSV7d and the reporter construct pMyc3E1BLuc (*18*) were introduced into HeLa cells together with the plasmids pRCCMV-TFEB-VP16 (0.3 μg) expressing the TFEB-VP16 fusion protein (*18*). pMyc3E1BLuc contained the luciferase gene downstream of the adenovirus E1B promoter with three binding sites for TFEB. Reporter construct pG5E1BCAT (*19*) containing five GAL4-binding sites inserted upstream of the E1B promoter and the CAT gene was used to assay GAL4-VP16 trans-activation.

Exp.	Vector			Tat-SF1			Fold enhancement†
	−*	+	Fold act.	−	+	Fold act.	
Tat							
1	100	7,228	72.3	31.7	7,699	242.9	3.36
2	100	15,779	157.8	17.0	16,548	973.4	6.17
3	100	4,899	49.0	35.3	10,353	293.3	5.99
TFEB-VP16							
1	100	30,229	302.3	118	20,782	176.1	0.58
2	100	18,241	182.4	179	15,080	84.2	0.46
GAL4-VP16							
1	100	132,208	1,322	95.0	129,960	1,368	1.03

*CAT or luciferase activity measured in cells transfected with the empty vector plus the reporter plasmid only were normalized to a value of 100. Values shown in the second, fourth, and fifth columns were adjusted accordingly. †The fold enhancement represents the fold activation by Tat or VP16 observed in cells expressing Tat-SF1 divided by the fold activation in cells containing the empty vector.

however, only a small amount of phosphorylated pp140 was detected. Thus, the binding of a phosphorylated pp140 to TAR requires the presence of the pc-D fraction, the Tat-SF fraction, and Tat.

An intact Tat activation domain was necessary for pp140 phosphorylation on TAR. When a nonfunctional Tat mutant (K41A) (*11*), which has Lys^{41} substituted by Ala, was present in the kinase reaction (Fig. 1B), the amount of phosphorylated pp140 bound to the immobilized TAR was reduced. This was not the result of a decreased ability of K41A to interact with TAR because K41A bound to TAR as efficiently as wild-type Tat in a gel mobility-shift assay (*8*). Similar results were also obtained with another Tat mutant (TatΔC) (Fig. 1C), which lacks the cysteine-rich activation domain (amino acids 22 to 37) (*11*) and is completely defective for transcriptional activity (*8*).

The polypeptide pp140 copurified and cotitrated with Tat-SF transcriptional activity during purification. When partially purified Tat-SF was sedimented through a glycerol gradient and the first 10 fractions were analyzed in the reconstituted transcription assay for the presence of Tat-SF activity (Fig. 2A), Tat-SF activity and pp140 co-peaked. The peak of Tat-SF activity that supported Tat activation, fractions 4 and 3, corresponded to a native molecular mass of ~100 kD. Phosphorylation of pp140 in the kinase assay was also most evident in the same fractions (Fig. 2B).

Sequential incubations of the pc-D fraction, the Tat-SF fraction, and Tat with an immobilized TAR revealed a stable and direct interaction between Tat and a kinase derived from the pc-D fraction and between Tat and pp140 from the Tat-SF fraction (*8*). Therefore, columns containing immobilized Tat were used to affinity-purify Tat-SF/pp140 (*9*). Fractions eluted from either a glutathione-S-transferase (GST)–Tat column or a Tat affinity column were analyzed in reconstituted transcription assays for Tat-SF activity. Fractions enriched in Tat-SF activity supported a Tat-specific and TAR-dependent activation (Fig. 3A) and also contained pp140 as detected by the kinase assay (Fig. 3B). When analyzed by silver staining, the polypeptide profiles of these fractions were different overall, with the only common band having a mobility of 140 kD (Fig. 3C). This polypeptide was judged to be pp140 and probably a component of Tat-SF activity.

The 140-kD polypeptide was recovered from the SDS-polyacrylamide gel and subjected to digestion with lys-C, and the resulting peptides were microsequenced. Sequence analysis of six peptides indicated that pp140 was a previously uncharacterized protein. However, one of the peptides was contained in the sequence of an unidentified expressed sequence tag (EST) in the Washington University–Merck EST database. A 103–amino acid protein fragment (amino acids 387 to 489) encoded by the corresponding EST clone was expressed as a GST fusion and used to immunize rabbits for the production of polyclonal antisera. By immunoblotting, the affinity-purified antibody specifically recognized a 140-kD protein present in both HeLa nuclear extracts (*8*) and a partially purified Tat-SF fraction (Fig. 4C).

The antibody was used to test the relation between the EST clone and pp140. The Tat-SF fraction was subjected to immunodepletion with the affinity-purified antibody and then incubated together with the pc-D fraction and Tat. This mixture was subjected to immunoprecipitation with the specific antibody, and the immune complex was subsequently analyzed in a kinase reaction in the presence of [γ-^{32}P]ATP (Fig. 4A). In contrast to the control undepleted Tat-SF fraction, the depleted fraction did not contain the phosphorylated pp140. Therefore, the 140-kD protein recovered from the SDS gel and represented by the EST clone was indeed pp140, the kinase substrate.

When the Tat-SF fraction and the pc-D fraction were incubated together in the absence of Tat, followed by immunoprecipitation with the antibody to pp140 (anti-pp140), pp140 was phosphorylated by its associated kinase when the isolated immune complex was assayed in the kinase reaction (Fig. 4A). Thus, pp140 forms a stable complex with its kinase independently of Tat. The addition of Tat to the initial incubation did not change the level of phosphorylation on pp140.

A preformed complex containing pp140 and its kinase could be isolated by immunoprecipitation and detected in a kinase reac-

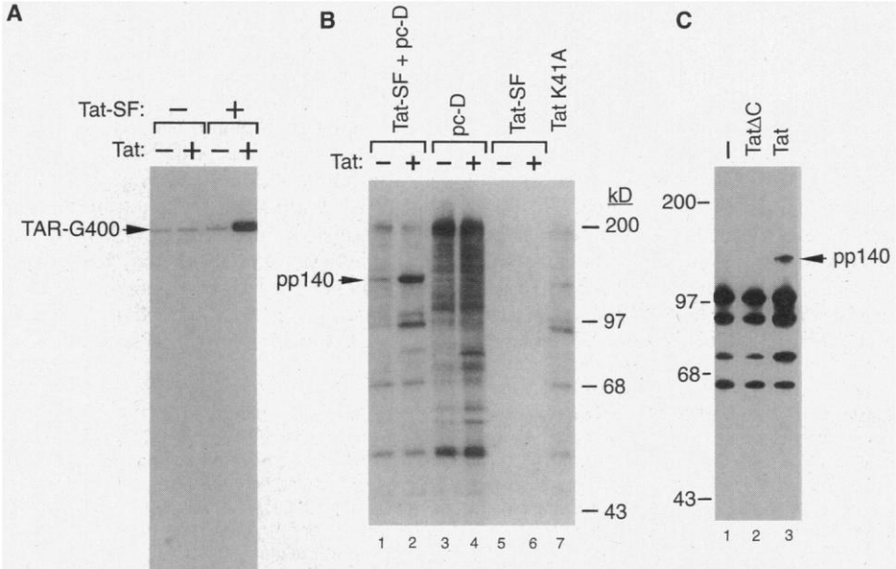

Fig. 1. (**A**) Requirement of a cellular activity, Tat-SF, for Tat activation of HIV transcriptional elongation. Reconstituted transcription reactions containing both templates pHIV+TAR-G400 and pHIVΔTAR-G100 were performed in the absence (−) or presence (+) of the 0.4 to 0.5 M KCl Q-Sepharose fraction containing Tat-SF activity as described (*7*). The pc-D fraction and purified TFIIA, TFIID, and Sp1 were present in all reactions. G-less cassettes of two different lengths were inserted into the above two templates at position +955 downstream of the HIV-1 initiation site to measure the effect of Tat on transcriptional elongation. Transcripts derived from these two templates were digested by ribonuclease T1, and the resulting 400- and 100-nucleotide G-less RNA fragments (arrows) were separated in a denaturing polyacrylamide gel. (**B**) Detection of the phosphorylated pp140 on an immobilized HIV-1 TAR RNA. Biotinylated TAR RNA (nucleotides +1 to +82) immobilized on the Paramagnetic beads was introduced into the kinase reactions containing [γ-^{32}P]ATP (10 μCi) and either the pc-D fraction alone (lanes 3 and 4) or the Tat-SF fraction alone (lanes 5 and 6), or both pc-D and Tat-SF fractions together (lanes 1, 2, and 7). Recombinant wild-type Tat (13 ng) was included in the reactions shown in lanes 2, 4, and 6. Tat mutant K41A (13 ng) was present in lane 7. After incubation for 10 min at 30°C, the TAR RNA beads were washed extensively in buffer D (*9*) containing 100 mM KCl and 0.1% NP-40, and the bound proteins were analyzed by SDS-PAGE. (**C**) The cysteine-rich activation domain of Tat is required for pp140 phosphorylation on TAR. Wild-type Tat (13 ng) or Tat mutant (TatΔC, 13 ng) lacking the cysteine-rich domain (amino acids 22 to 37) was included in the kinase reactions containing immobilized TAR. No Tat was present in the control reaction (lane 1).

tion from an unfractionated HeLa nuclear extract in the absence of Tat (Fig. 4A). This complex was stable under transcription conditions (less than 0.1 M KCl) but dissociated in washes of greater than 0.25 M KCl (8), and probably dissociated during fractionation in the purification of Tat-SF. These observations suggest that Tat is not required for the phosphorylation of pp140 by its associated kinase, but is required for the association of the phosphorylated pp140 and the kinase with TAR (Fig. 1B). Thus, Tat probably recruits a preformed complex containing pp140 and a kinase to the HIV promoter region during transcription.

To examine whether pp140 is required for Tat activation, we used anti-pp140 to immunodeplete pp140 from a partially purified fraction containing Tat-SF activity. The depleted fraction was then tested in reconstituted transcription reactions for its ability to support Tat activation (Fig. 4B). Control reactions without Tat-SF did not support Tat activation. Inclusion of the Tat-SF fraction resulted in a TAR-dependent activation by Tat. As compared with control Tat-SF fraction, depletion of the Tat-SF fraction with specific anti-pp140 immobilized on protein A–Sepharose matrix once or especially twice reduced its ability to support Tat activation. In contrast, depletion of the Tat-SF fraction with preimmune antibody either once or twice did not substantially reduce Tat activation. Similarly, depletion of the Tat-SF fraction with an unrelated antibody (monoclonal antibody 12CA5 to hemagglutinin A) had no effect on Tat activation (8). By immunoblotting (Fig. 4C), depletion with anti-pp140 efficiently removed pp140 from the Tat-SF fraction. Taken together, these experiments argue that pp140 is required for Tat-SF transcriptional activity.

Probes made from the insert of the EST clone were used to screen a λ cDNA library prepared from human HL60 cells (*12*). Seven independent plaques were isolated with overlapping inserts. The largest cDNA insert was 2.8 kb and contained a 2271–base pair open reading frame. It encoded a protein of 754 amino acids with a calculated molecular mass of 85,767 daltons (Fig. 5A), which was substantially less than the apparent molecular mass of 140 kD calculated from the mobility in an SDS gel. On the basis of several criteria, this cDNA was judged to encode full-length pp140. First, rabbit reticulocyte lysate programmed with RNA transcribed from this cDNA produced a protein with a mobility indistinguishable from that of pp140 (8). All six peptide sequences obtained from partial sequencing of pp140 were

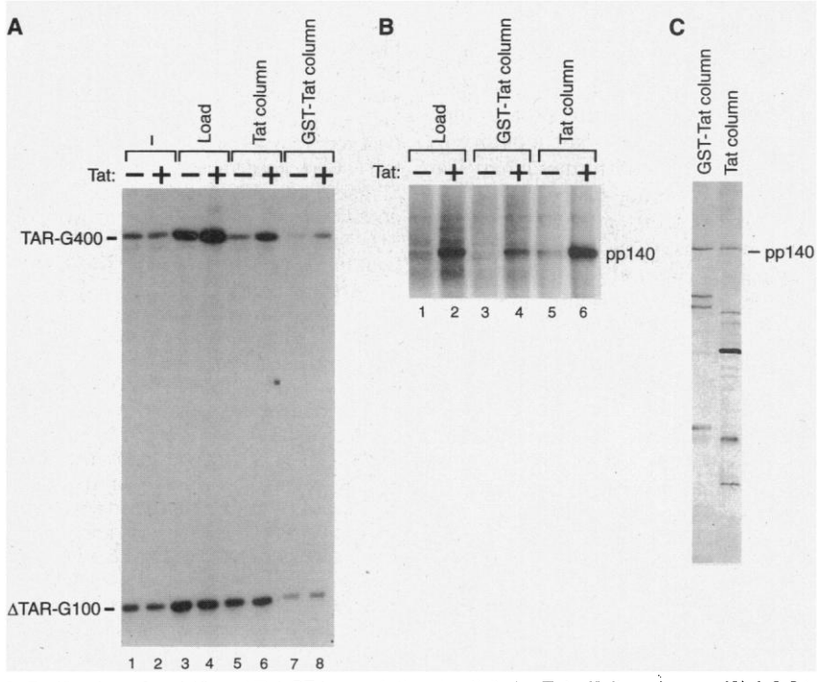

Fig. 3. Purification of pp140 and Tat-SF transcriptional activity by Tat affinity columns. (**A**) A 0.2 to 0.4 M KCl heparin Sepharose fraction (Load) containing Tat-SF activity was applied to a GST-Tat column or a Tat affinity column (*9*). Eluates from these two columns were tested for Tat-SF activity in reconstituted transcription reactions as described in Fig. 1A. (**B**) The same two fractions were tested for the presence of pp140 in a kinase assay as described in Fig. 1B. (**C**) A silver-stained SDS gel of the above two fractions is shown.

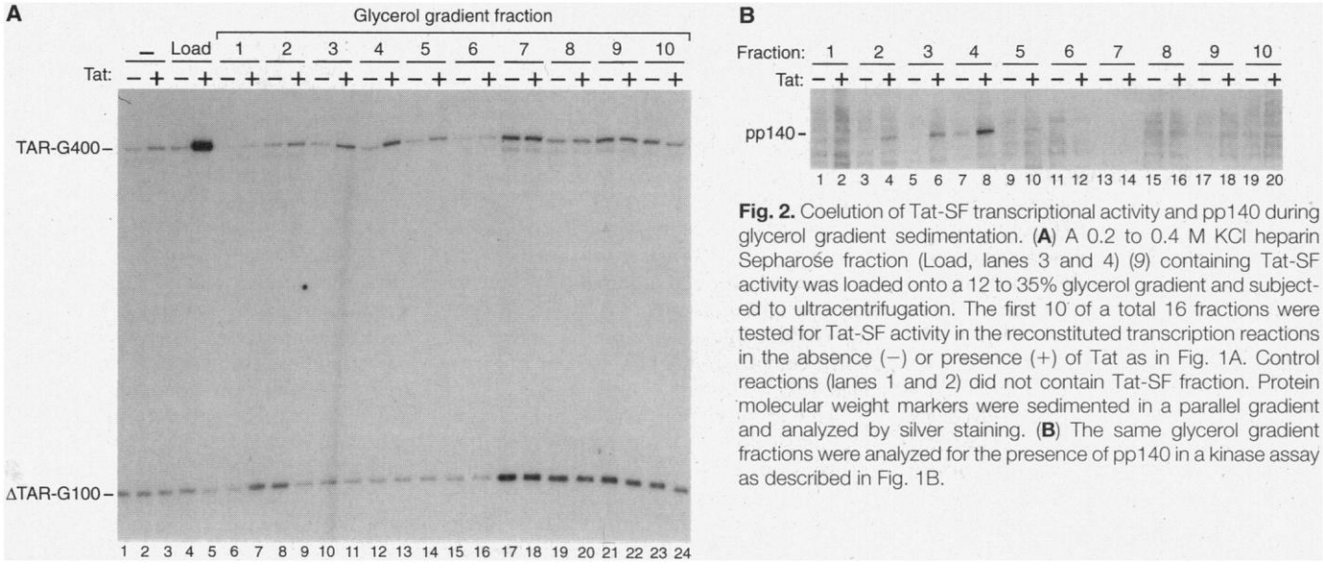

Fig. 2. Coelution of Tat-SF transcriptional activity and pp140 during glycerol gradient sedimentation. (**A**) A 0.2 to 0.4 M KCl heparin Sepharose fraction (Load, lanes 3 and 4) (*9*) containing Tat-SF activity was loaded onto a 12 to 35% glycerol gradient and subjected to ultracentrifugation. The first 10 of a total 16 fractions were tested for Tat-SF activity in the reconstituted transcription reactions in the absence (–) or presence (+) of Tat as in Fig. 1A. Control reactions (lanes 1 and 2) did not contain Tat-SF fraction. Protein molecular weight markers were sedimented in a parallel gradient and analyzed by silver staining. (**B**) The same glycerol gradient fractions were analyzed for the presence of pp140 in a kinase assay as described in Fig. 1B.

found in the predicted coding region (Fig. 5A). Finally, Northern (RNA) analysis of polyadenylated RNA isolated from several different types of human cells detected a single 3.0-kb species, a length consistent with that of the cDNA insert and adequate to encode a polypeptide of 86 kD (*8*).

Sequence analysis of the protein, named Tat-SF1, revealed that it has several distinct features (Fig. 5A). The protein can be roughly divided at position 420 into two halves. The COOH-terminal half was rich in acidic amino acids, with 48% of the last 245 amino acid residues consisting of glutamate or aspartate. The unusual acidic nature of this protein may be responsible for its aberrant mobility in an SDS gel. The COOH-terminal half also contained many serine residues that are arranged in a short peptide sequence matching consensus sites for phosphorylation by casein kinase II (*13*). Such phosphorylation would contribute additional negative charges to this region.

The NH_2-terminal half of Tat-SF1 contained two tandem RNA recognition motifs (RRMs) (*14*) that have homology to many RNA binding proteins. The first RRM of Tat-SF1 (amino acids 128 to 217, boxed in Fig. 5A) was similar in length and displayed the strongest sequence homology to the RRMs located in the COOH-terminal half of two closely related human proteins, the Ewing's sarcoma protein (EWS) (*15*) (Fig. 5B) and FUS/TLS (*16*). Furthermore, the sequence homology between Tat-SF1 and EWS or between Tat-SF1 and FUS/TLS extended beyond the two RRMs into the immediate NH_2-terminal region of Tat-SF1 (Fig. 5, A and B). Thus, Tat-SF1 is related to EWS and FUS/TLS, which are members of a class of putative transcription factors that presumably interact with RNA. Both EWS and FUS/TLS are associated with many forms of human solid tumors (*17*), such as Ewing's sarcoma (*15*) and human myxoid liposarcoma (*16*).

To investigate whether overexpression of Tat-SF1 affects the level of Tat activation in vivo, we introduced a plasmid expressing Tat-SF1 and a reporter construct containing HIV-1 long terminal repeat (LTR) linked to the bacterial chloramphenicol acetyltransferase (CAT) gene into HeLa cells either in the presence or absence of a cotransfected plasmid expressing Tat (Table 1). As a control, the effect of Tat-SF1 overexpression on transcriptional activation by the acidic activation domain VP16 in the TFEB-VP16 (*18*) and GAL4-VP16 (*19*) fusion proteins was assayed. Expression of Tat-SF1 from the transfected DNA resulted in an increase in Tat activation by an average of 5.2-fold as compared with the control HeLa cells transfected with an empty vector (Table 1). The enhanced activation mediated by Tat-SF1 was Tat-specific, because overexpression of Tat-SF1 had little, or sometimes even a slightly negative, effect on transcriptional activation by VP16. The increased fold induction by Tat resulting from Tat-SF1 overexpression was caused by a combination of a decrease in the basal level of transcription from HIV-1 LTR in the absence of Tat and a small increase in the level of Tat-activated transcription (Table 1). Because Tat-SF1 is probably a component of a protein complex that also includes a cellular kinase and perhaps other cellular components, overexpression of Tat-SF1 alone may disrupt the normal stoichiometry of the complex, resulting in a decrease in the basal level of HIV transcription. The presence of Tat could stabilize and recruit the active form of the complex to the HIV promoter to stimulate the processivity of elongation.

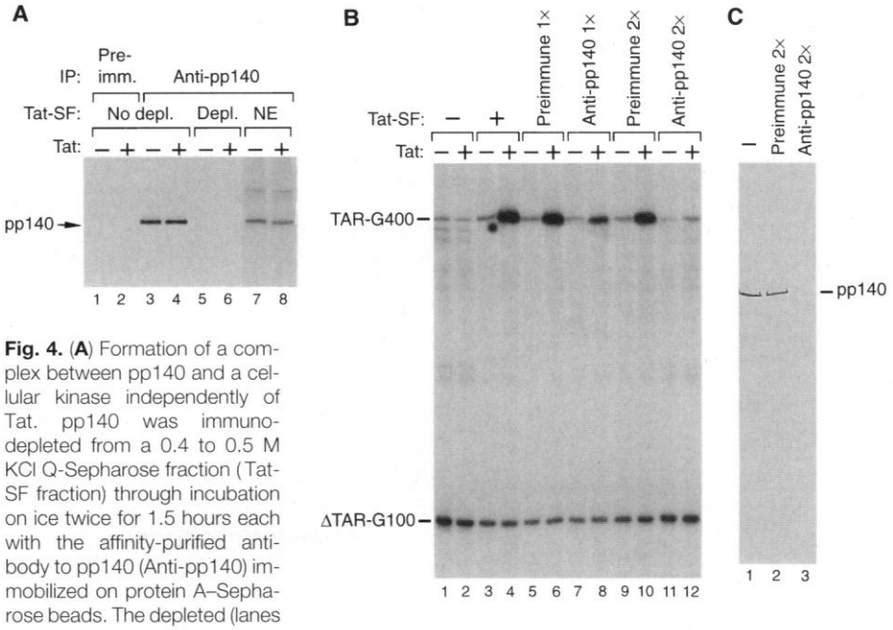

Fig. 4. (**A**) Formation of a complex between pp140 and a cellular kinase independently of Tat. pp140 was immunodepleted from a 0.4 to 0.5 M KCl Q-Sepharose fraction (Tat-SF fraction) through incubation on ice twice for 1.5 hours each with the affinity-purified antibody to pp140 (Anti-pp140) immobilized on protein A–Sepharose beads. The depleted (lanes 5 and 6) or undepleted Tat-SF fraction (lanes 1 to 4) was incubated with the pc-D fraction in the absence (−) or presence (+) of Tat, and the reaction was subjected to immunoprecipitation with anti-pp140 (lanes 3 to 6). Preimmune antibody was used in control precipitations (lanes 1 and 2). An unfractionated HeLa nuclear extract (NE) with (+) or without (−) the addition of Tat was also subjected to immunoprecipitation with the specific antibody (lanes 7 and 8). After extensive washes with buffer D containing 100 mM KCl, 0.1% NP-40, and 10 mM $MgCl_2$, the immune complex bound to protein A–Sepharose beads was incubated with [γ-^{32}P]ATP for 10 min at 30°C, washed with buffer D, and analyzed by SDS-PAGE. (**B**) Depletion of pp140 from a partially purified Tat-SF fraction inactivates Tat-SF transcriptional activity. A 0.4 to 0.5 M KCl Q-Sepharose fraction containing Tat-SF activity was subjected to immunodepletion with preimmune antibody (lanes 5, 6, 9, and 10) or anti-pp140 (lanes 7, 8, 11, and 12) immobilized on protein A–Sepharose beads as in (A). Tat-SF fraction subjected to depletion once (1×) or twice (2×), and the undepleted fraction (lanes 3 and 4), were tested in transcription reactions for Tat-SF activity as in Fig. 1A. No Tat-SF fraction was present in the control reactions (lanes 1 and 2). (**C**) Anti-pp140 removed pp140 from the Tat-SF fraction. The undepleted Tat-SF fraction and Tat-SF fraction twice depleted with the specific or preimmune antibody were analyzed by immunoblotting with anti-pp140.

The in vivo transfection and in vitro immunodepletion experiments described above strongly argue that Tat-SF1 is required for Tat activation and is at least part of the Tat-SF transcriptional activity. These results also indicate that Tat-SF1, which was identified biochemically in vitro, is relevant to processes in vivo. Immunostaining of cells of various types with antibody specific for Tat-SF1 generates a diffused nuclear pattern that excludes nucleoli (*8*). Both immunostaining and Northern blotting indicate that Tat-SF1 is expressed in a number of cell types. Thus, the cellular process mediated by Tat-SF1 might be broadly active.

A preformed complex containing Tat-SF1 and its associated kinase bound to TAR RNA weakly, probably mediated by a low-affinity interaction between the putative RNA binding domains of Tat-SF1 and TAR. The association of the complex with TAR was substantially enhanced when Tat was present in the reaction. An intact activation domain of Tat is essential in this process. Thus, Tat may activate the processivity of elongation by recruitment of the Tat-SF1-kinase complex to the HIV-1 promoter through a Tat-TAR interaction. We have not established specificity for the interaction of Tat and the Tat-SF1 complex with the TAR RNA. Because Tat specifically binds TAR and Tat-SF1 probably also

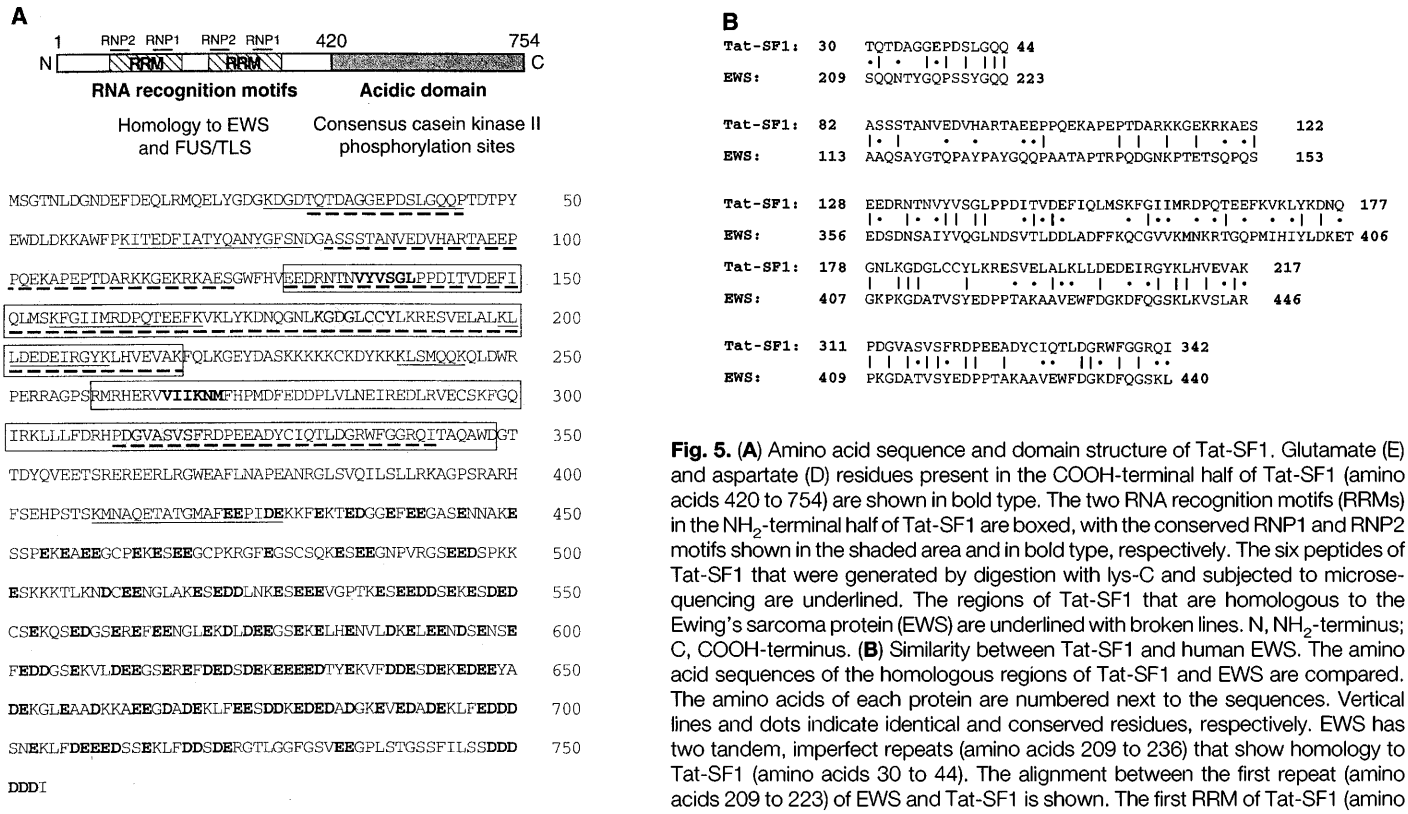

Fig. 5. (**A**) Amino acid sequence and domain structure of Tat-SF1. Glutamate (E) and aspartate (D) residues present in the COOH-terminal half of Tat-SF1 (amino acids 420 to 754) are shown in bold type. The two RNA recognition motifs (RRMs) in the NH_2-terminal half of Tat-SF1 are boxed, with the conserved RNP1 and RNP2 motifs shown in the shaded area and in bold type, respectively. The six peptides of Tat-SF1 that were generated by digestion with lys-C and subjected to microsequencing are underlined. The regions of Tat-SF1 that are homologous to the Ewing's sarcoma protein (EWS) are underlined with broken lines. N, NH_2-terminus; C, COOH-terminus. (**B**) Similarity between Tat-SF1 and human EWS. The amino acid sequences of the homologous regions of Tat-SF1 and EWS are compared. The amino acids of each protein are numbered next to the sequences. Vertical lines and dots indicate identical and conserved residues, respectively. EWS has two tandem, imperfect repeats (amino acids 209 to 236) that show homology to Tat-SF1 (amino acids 30 to 44). The alignment between the first repeat (amino acids 209 to 223) of EWS and Tat-SF1 is shown. The first RRM of Tat-SF1 (amino acids 128 to 446) is almost identical in length and is 27% identical and 52% similar in amino acid sequence to the RRM of EWS. Sequence homology similar to that observed between Tat-SF1 and EWS also exists between Tat-SF1 and human FUS/TLS (*8*), which is closely related to EWS. The RRMs of other RNA binding proteins are less homologous and show greater variations in length as revealed by the BLAST algorithm (*27*). 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.

binds RNA, perhaps with specificity, these two proteins could cooperatively mediate recruitment of the complex to TAR.

Tat activates transcription through increasing the processivity of elongation by RNA polymerase II (*20*). Activation domains of several strong transcription factors also stimulate both initiation and elongation (*1*), probably through a combination of signals. Current models suggest that polymerase elongation might be affected either by association with factors such as SII (TFIIS) (*21*), TFIIF (*22*), or Elongin (6) or by phosphorylation of the heptapeptide repeats within the COOH-terminal domain of RNA polymerase II (*10*). It remains to be determined whether Tat-SF1 or its associated kinase stimulates these interactions or affects the phosphorylation of polymerase. A cellular kinase that interacts with the activation domain of Tat has been described previously (*23*). Preliminary characterization suggests that the Tat-SF1–associated kinase differs from this previously described activity. Tat has also been reported to interact with the interferon-induced, double-stranded RNA–dependent kinase, PKR (*24*). However, none of these kinases has been shown to be required for Tat trans-activation.

Tat-SF1 is related to EWS and FUS/TLS, members of a family of putative transcription factors that may interact with RNA. When the NH_2-terminal domains of EWS and FUS/TLS are fused to particular DNA binding domains through chromosomal translocations, the chimeric proteins can promote oncogenic growth (*25*). In this context, the NH_2-terminal regions of EWS and FUS/TLS are strong activators of transcription. In contrast to the NH_2-terminal domains of EWS and FUS/TLS, which contain a series of degenerate, glutamine-rich repeats, the COOH-terminal half of Tat-SF1 is markedly acidic. Significantly, glutamine-rich domains and acidic amino acid tracts have both been associated with many potent transcriptional activators (*26*). These observations link RNA binding potential with transcriptional activators, properties similar to those of HIV Tat. Further analysis of Tat activation of the processivity of transcriptional elongation and the role of Tat-SF1 in this process will likely reveal general mechanisms of gene regulation at the stage of elongation.

REFERENCES AND NOTES

1. W. S. Blair, R. A. Fridell, B. R. Cullen, *EMBO J.* **15**, 101 (1996); J. Blau *et al.*, *Mol. Cell. Biol.* **16**, 2044 (19 → K. Yankulov, J. Blau, T. Purton, S. Roberts, D. L. Bentley, *Cell* **77**, 749 (1994).
2. B. R. Cullen, *Infect. Agents Dis.* **3**, 68 (19 → A. D. Frankel, *Curr. Opin. Genet. Dev.* **2**, 293 (19 → K. A. Jones and B. M. Peterlin, *Annu. Rev. Biochem.* **63**, 717 (19 → J. Karn and M. A. Graeble, *Trends Genet.* **8**, 365 (1992); R. B. Gaynor, *Curr. Top. Microbiol. Immunol.* **193**, 51 (1995).
→ S. Feng and E. C. Holland, *Nature* **334**, 165 (1988); J. A. Garcia *et al.*, *EMBO J.* **8**, 765 (19 → M. J. Selby, E. S. Bain, P. A. Luciw, B. M. Peterlin, *Genes Dev.* **3**, 547 (1989).
4. A. Kibel, O. Iliopoulos, J. A. De Caprio, W. G. Kaelin Jr., *Science* **269**, 1444 (1995).
5. D. R. Duan *et al.*, *ibid.*, p. 1402.
6. T. Aso, W. S. Lane, J. W. Conaway, R. C. Conaway, *ibid.*, p. 1439.
7. Q. Zhou and P. A. Sharp, *EMBO J.* **14**, 321 (1995).
8. _____, unpublished data.
9. HeLa nuclear extract in buffer D–0.1 M KCl [20 mM Hepes-KOH (pH 7.9), 20% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride] was loaded onto a phosphocellulose column preequilibrated with the same buffer. The flow-through was loaded onto a DEAE-Sepharose FF (Pharmacia) matrix column preequilibrated with buffer D–0.1 M KCl. After the column was washed with the same buffer, Tat-SF activity was eluted from the column with buffer D–0.3 M KCl. This fraction was dialyzed against buffer D–0.1 M KCl and applied to a Q-Sepharose FF (Pharmacia) matrix column preequilibrated with the same buffer. The column was washed with buffer D–0.1 M KCl, and the bound proteins were eluted with a gradient of 0.1 to 0.7 M KCl in buffer D. Fractions were analyzed for Tat-SF activity in reconstituted transcription assays and for pp140 in kinase reactions. The 0.4 to 0.5 M KCl Q-Sepharose fraction containing Tat-SF activity and pp140 was dialyzed against buffer D–0.1 M KCl and applied to a heparin Sepharose column. After the column was washed extensively with buffer

D–0.1 M KCl, Tat-SF/pp140 was eluted with increasing salt concentrations and was detected mostly in 0.2 to 0.4 M KCl fractions. These fractions were pooled, dialyzed against buffer D–0.1 M KCl, and loaded onto a glutathione Sepharose (Pharmacia) column containing GST-Tat fusion proteins. After the column was washed with buffer D–0.4 M KCl, Tat-SF/pp140 was eluted from the column with buffer D containing 1.4 M KCl. The estimated overall purification after these steps was ~3000-fold. In the experiment shown in Fig. 3, the 0.2 to 0.4 M KCl heparin Sepharose fraction containing Tat-SF activity was subjected to fractionation through an Affi-Gel 10 matrix column (Bio-Rad) containing immobilized Tat. Tat-SF activity was eluted from the column with increasing salt concentrations. The 0.6 M KCl fraction was analyzed as described in Fig. 3.

T. O'Brien, S. Hardin, A. Greenleaf, J. T. Lis, *Nature* **370**, 75 (1994); M. E. Dahmus, *Biochim. Biophys. Acta.* **1261**, 171 (1995).

11. A. P. Rice and F. Carlotti, *J. Virol.* **64**, 1864 (1990).
12. The Tat-SF/pp140 fraction eluted from the GST-Tat column was subjected to SDS–polyacrylamide gel electrophoresis (PAGE), and the pp140 polypeptide was blotted onto a nitrocellulose membrane. Approximately 15 μg of pp140 were recovered from the membrane and subjected to digestion with lys-C. Six major peptides were obtained and microsequenced. One of the peptides (KMNAQETATGMAFEEPIDE) was contained in the sequence of EST60354 in the Washington University–Merck EST database. An Xho I–Eco RI fragment corresponding to the COOH-terminus of the Tat-SF1 gene and its 3′ untranslated region was labeled and used as a probe to screen a λZipLox (Gibco BRL) cDNA library prepared from human HL60 cells. Complementary DNAs were recovered from seven independent plaques in the autonomously replicating plasmid pZL1 as instructed by the manufacturer (Gibco BRL). The largest cDNA clone containing the full-length Tat-SF1 gene was named pZL-Tat-SF1-4b and was sequenced by dideoxy-DNA sequencing with T7 DNA polymerase.
13. D. R. Marshak and D. Carroll, *Methods Enzymol.* **200**, 134 (1991).

D. J. Kenan, C. C. Query, J. D. Keene, *Trends Biochem. Sci.* **16**, 214 (1991).

O. Delattre *et al.*, *Nature* **359**, 162 (19[illegible] P. H. Sorensen *et al.*, *Nature Genet.* **6**, 146 (1994).

A. Crozat, P. Aman, N. Mandahl, D. Ron, *Nature* **363**, 640 (19[illegible] T. H. Rabbitts, A. Forster, R. Larson, P. Nathan, *Nature Genet.* **4**, 175 (1993).

17. M. Ladanyi, *Diagn. Mol. Pathol.* **4**, 162 (19[illegible] T. H. Rabbitts, *Nature* **372**, 143 (1994).
18. S. E. Harper, Y. Qiu, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8536 (1996).

J. W. Lillie and M. R. Green, *Nature* **338**, 39 (1989).

H. Kato *et al.*, *Genes Dev.* **6**, 655 (1992); R. A. Marciniak and P. A. Sharp, *EMBO J.* **10**, 4189 (1991).

M. G. Izban and D. S. Luse, *Genes Dev.* **6**, 1342 (1992); D. Wang and D. K. Hawley, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 843 (1993).

22. E. Bengal, O. Flores, A. Krauskopf, D. Reinberg, Y. Aloni, *Mol. Cell. Biol.* **11**, 1195 (19[illegible] J. Greenblatt, J. R. Nodwell, S. W. Mason, *Nature* **364**, 401 (1993).
23. C. H. Herrmann and A. P. Rice, *J. Virol.* **69**, 1612 (1995).

N. A. McMillan *et al.*, *Virology* **213**, 413 (1995).

25. W. A. May *et al.*, *Mol. Cell. Biol.* **13**, 7393 (19[illegible] H. Zinszner, R. Albalat, D. Ron, *Genes Dev.* **8**, 2513 (1994); D. D. Prasad, M. Ouchida, L. Lee, V. N. Rao, E. S. Reddy, *Oncogene* **9**, 3717 (1994).
26. P. J. Mitchell and R. Tjian, *Science* **245**, 371 (1989).
27. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
28. M. A. Truett *et al.*, *DNA* **4**, 333 (1985).
29. H. E. Gendelman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9759 (1986).

L. S. Tiley, P. H. Brown, B. R. Cullen, *Virology* **178**, 560 (1990).

31. J. R. Neumann, C. A. Morency, K. O. Russian, *BioTechniques* **5**, 444 (1987).

32. We are grateful to B. Pepinsky and Biogen for providing pure HIV Tat protein and Tat mutant TatΔC; to J. Borrow [Massachusetts Institute of Technology (MIT) Center for Cancer Research] for human cDNA libraries; and to R. Cook (MIT Biopolymers Laboratory) for peptide sequencing. We thank K. Luo, J. Borrow, and H. Kawasaki for valuable advice and discussions; and B. Blencowe, K. Cepek, G. Jones, K. Luo, and C. Query for helpful comments on the manuscript. We also thank M. Siafaca for secretarial support. Supported by grants from the National Institutes of Health (GM34277 and AI32486) to P.A.S., and partially supported by a National Cancer Institute Center core grant (CA14051). Q.Z. was supported by a postdoctoral fellowship of The Jane Coffin Childs Memorial Fund for Medical Research.

19 June 1996; accepted 23 August 1996

Accessing Genetic Information with High-Density DNA Arrays

Mark Chee, Robert Yang, Earl Hubbell, Anthony Berno, Xiaohua C. Huang, David Stern, Jim Winkler, David J. Lockhart, Macdonald S. Morris, Stephen P. A. Fodor

Rapid access to genetic information is central to the revolution taking place in molecular genetics. The simultaneous analysis of the entire human mitochondrial genome is described here. DNA arrays containing up to 135,000 probes complementary to the 16.6-kilobase human mitochondrial genome were generated by light-directed chemical synthesis. A two-color labeling scheme was developed that allows simultaneous comparison of a polymorphic target to a reference DNA or RNA. Complete hybridization patterns were revealed in a matter of minutes. Sequence polymorphisms were detected with single-base resolution and unprecedented efficiency. The methods described are generic and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability.

A central theme in modern genetics is the relation between genetic variability and phenotype. To understand genetic variation and its consequences on biological function, an enormous effort in comparative sequence analysis will need to be carried out. Conventional nucleic acid sequencing technologies make use of analytical separation techniques to resolve sequence at the single nucleotide level (*1*, *2*). However, the effort required increases linearly with the amount of sequence. In contrast, biological systems read, store, and modify genetic information by molecular recognition (*3*). Because each DNA strand carries with it the capacity to recognize a uniquely complementary sequence through base pairing, the process of recognition, or hybridization, is highly parallel, as every nucleotide in a large sequence can in principle be queried at the same time. Thus, hybridization can be used to efficiently analyze large amounts of nucleotide sequence. In one proposal, sequences are analyzed by hybridization to a set of oligonucleotides representing all possible subsequences (*4*). A second approach, used here, is hybridization to an array of oligonucleotide probes designed to match specific sequences. In this way the most informative subset of probes is used. Implementation of these concepts relies on recently developed combinatorial technologies to generate any ordered array of a large number of oligonucleotide probes (*5*).

Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051, USA.

The fundamentals of light-directed oligonucleotide array synthesis have been described (*5*, *6*). Any probe can be synthesized at any discrete, specified location in the array, and any set of probes composed of the four nucleotides can be synthesized in a maximum of $4N$ cycles, where N is the length of the longest probe in the array. For example, the entire set of $\sim 10^{12}$ 20-nucleotide oligomer probes, or any desired subset, can be synthesized in only 80 coupling cycles. The number of different probes that can be synthesized is limited only by the physical size of the array and the achievable lithographic resolution (*7*).

An array consisting of oligonucleotides complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Many different arrays can be designed for these purposes. One such design, termed a 4L tiled array, is depicted in Fig. 1A. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. By this approach, a nucleic acid target of length L can be scanned for mutations with a tiled array containing $4L$ probes. For example, to query the 16,569 base pairs (bp) of human mitochondrial DNA (mtDNA), only 66,276 probes of the possible $\sim 10^9$ 15-nucleotide oligomers need to be used.

The use of a tiled array of probes to read a target sequence is illustrated in Fig. 1C. A tiled array of 15-nucleotide oligomers varied